mmAssist : Passive Monitoring of Driver's Attentiveness Using mmWave Sensors

Argha Sen*, Anirban Das†, Prasenjit Karmakar‡, Sandip Chakraborty§
Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur
Email: *arghasen10@gmail.com, †anirbanfuture@gmail.com, ‡prasenjitkarmakar52282@gmail.com, §sandipc@cse.iitkgp.ac.in

***Abstract*—Continuous monitoring of driver attentiveness inside a car has been of significant importance for quite some time. However, the state-of-the-art techniques are primarily inclined toward image-based data, which is invasive and, therefore, could pose challenges in the pervasive adoption of such a system. This work proposes a novel approach for continuous driver attentiveness monitoring, leveraging millimeter Wave (mmWave) sensing to address that. The sensing infrastructure is compact, lightweight, and bears the exclusive potential to be adopted in a pervasive manner due to the continuously increasing popularity of mmWave hardware with 5G technology. We study the driver's attention as a multi-class problem and address that using Range Doppler information from an mmWave radar. We evaluate the proposed methodologies in a lab and a real-world driving scenario. Within the lab-based setup, we achieved an accuracy of 88%, whereas, in the real-world system, we could achieve an accuracy of up to 79% while monitoring the driver's activities associated with driving attentiveness. The source code is publicly available in GitHub [1].**

***Index Terms*—Driver attentiveness, mmWave sensing, Range Doppler.**

I. Introduction

Statistically, driver factors, including driver errors, have been the most common cause of road accidents [2]. As a result, over the past years, road safety has evolved as a serious concern for the current generation of the intelligent transportation system. A general user survey conducted in the American context [3] clearly shows that more than 90% of the individuals wish to have a service alerting them about dangerous driving situations. The rest of the globe is no different in this. These concerns have resulted in carmakers integrating assistance for driver attention in their products. Toyota's *Driver Attention Monitor*, Mercedes's *Attention Assist*, Volvo's *Driver Alert Control*, etc. [4] are the results of such initiatives. However, the working principles of these systems and the specific factors and behaviors being addressed are not always identical, even though the broader objective remains the same – estimating whether the driver is attentive while driving. Research initiatives have explored diverse technical territories to study drivers' attentiveness in this direction. Whereas some works [5] assess the vehicle's orientation within the environment, others observe the behavioral and physiological measures [6] of the driver while estimating the attentiveness. The latter has been proven to be more promising and has been adopted by most of the recent works [6]–[8]. However, when observed keenly, the direct behavioral observation of the driver primarily relies on image or video sources captured through a camera. Instinctively, individuals (not only the drivers, but the passengers as well) have low preference [9] over their images to be captured due to the inherent privacy concerns, making its pervasive acceptance a major challenge. Another hindrance to the camera-based approach is that it requires costly resources for capturing and processing the data, thus elevating the overall cost of implementation. With that motivation, in this work, we explore *millimeter wave (mmWave)* based sensing to monitor driver's attentiveness.

Recent years have seen a drastic evolution of mmWave technology. The emergence of the necessary hardware and the availability of unlicensed frequency bands have throttled the adoption of mmWave technologies across several domains. With the shift of communication infrastructure towards 5G, a huge number of consumer devices will have integrated mmWave hardware. Apart from the use for communication, mmWave's tendency to be affected by surrounding objects has resulted in various COTS (commercial off-the-shelf) products that are geared towards sensing-related applications. Texas Instruments, Aura Intelligent Systems, Inc., etc., manufacture off-the-shelf products which leverage mmWave technology for sensing the surroundings. Interestingly, mmWave can penetrate objects such as clothes, thus making it suitable to deal with surfaces light cannot reach. In recent times, mmWave-based radars have been used to address a diverse range of problems such as human recognition [10], gesture recognition [11], vital sign monitoring [12] and the likes, which justifies its potential.

Motivated by these factors, this work proposes *mmAssist* , a minimally invasive driver's attentiveness monitoring system using an off-the-shelf mmWave radar. It captures different drivers' activities associated with the attentiveness of the driver. The proposed technique has multiple advantages over state-of-the-art image-based approaches: 1) It significantly lowers individuals' privacy concerns by avoiding capturing images inside a car. 2) It can also work under different background conditions of the environment, such as the lighting condition of the car, background acoustic noises, etc. 3) Being passive monitoring, it even avoids arguably invasive measures such as the use of wearables [13].

Despite the aforementioned advantages, several hurdles and challenges must be overcome to reliably sense the driver's attentiveness using the mmWave modality. **First**, we must associate the driver's attentiveness with different activity signatures. For example, yawning or talking activities can be mapped to the inattentiveness of a driver. So associating attentiveness with driver's activities is the first challenge. The **second** challenge is differentiating the driver from other users in the vicinity or within the sensor's field of view. As in a real driving scenario, we can have multiple passengers inside a car which can easily impact the reflected mmWave signals. The **Third** and the final challenge is the annotation of the mmWave data. Recognizing driver's activities using any supervised learning model needs labeled activities along with the mmWave data. However, annotating these diverse human activities for different users is time-consuming and can have label-jitters and human errors.

Henceforth, we design *mmAssist* with the notion of addressing these challenges. Thus, the contributions of this work can be summarized as follows:

Contribution 1: We conduct a survey as a pilot study on user preference to justify the need for an alternative to existing state-of-the-art approaches. From the survey's outcome, it is evident that users wish to avoid being continuously monitored by a video camera.

Contribution 2: We leverage a single COTS mmWave radar to design a novel approach to monitor driver attentiveness, which to the best of our knowledge, has not been used earlier to address the problem. While doing so, multiple sub-problems are addressed, such as the isolation of the driver from the rest of the passengers, detection of head orientation, etc. Therefore, we also shape methodologies to address those individual problems.

Contribution 3: With the support of a team of volunteers, we have collected a comprehensive *in-car* dataset of driver activities with an mmWave radar installed inside a car. The dataset comprises around 11 hours of driving data in addition to 4 hours of data from an *in-lab* setup with a simulated environment. These come with annotations for the necessary events.

Contribution 4: Through in-lab and real-world experiments, we analyze the performance of *mmAssist* , which achieves a weighted F1-Score of 88% on *in-lab* setup, and 79% on real-world deployment. The source code is publicly available in GitHub. [1].

II. Related work

The problem of monitoring drivers' attentiveness has evolved through a number of stages. Earlier works [14], [15] primarily investigated the vehicle's orientation in terms of its position, velocity, acceleration, etc. Irregular behaviors [16] while driving has also been explored to assess attentiveness. Usually, these abnormal driving patterns are affected by several factors, such as drowsiness, fatigue, etc., and have been exploited by the abovementioned approaches. However, driving habits may also be impacted by real-world uncontrollable factors such as traffic conditions, weather, etc., and therefore may not bear significant signatures for driver inattentiveness. Thus, the researchers gradually adopted computer vision-based approaches [17], [18] over these non-visual feature-based approaches for addressing the problem. Typically, the driver's behavior is captured using a camera to identify irregular patterns of improper driving. These image-based approaches involved different imaging sources such as RGB cameras [19], infrared cameras [20] as well as thermal visions [21] to capture images within the car, especially of the driver. The image-based approaches target various facial features [22] such as eye blinks, eye movements, yawning, and facial expressions. Some works [23] also target other movement signatures such as hand movement, mobile phone usage, etc. Incidentally, facial features have also been fused [24] with data of car orientation patterns to address the same problem. However, as stated earlier, visual data possess privacy concerns in many aspects and pose deployment challenges. The other direction of attentiveness-monitoring leverage the signals collected by a set of body-mounted sensors. The rapid advent of MEMS (Micro Electro Mechanical Systems) technologies has gifted several wearable sensors [25], which could capture critical signals from a human body relevant to a driver's attentiveness. In this direction, different data sources, such as EEG (Electroencephalogram) [26] signals, heart rate, etc., have been explored in the recent past. However, a significant limitation with the wearables is that the models trained with the data from one age group cannot perform effectively in a different age group [27]. Although the invasive footprint of many such wearables [28] is minimal, they do restrict an individual's day-to-day movement. For instance, a simple ignorance to carry the wearable could jeopardize the complete monitoring system. Moreover, a user might not consistently be forced to wear a device, such as an EEG sensor, and could be reluctant to wear one regularly. In such a case, the whole system fails. In addition, deploying many such sensing devices, especially on a human body on a regular basis, could be a tedious task.

Lastly, researchers have also explored the domain of acoustics. In this direction, Doppler shifts from acoustic signals [29], [30] as well as FMCW chirps [31] were leveraged to detect abnormal body movements of the driver. The captured patterns are then analyzed to predict improper driving scenarios. However, it is worth noting that environmental noise plays a pivotal role [32] in acoustic-based sensing. Moreover, location, orientation, and privacy concerns could also be significant with acoustic sensing. Finally, when devices such as smartphones are involved in acoustic sensing, it is worth noting that different devices have different sensitivity patterns [33] for audio signals. Therefore, there exists a need for a solution that is minimally invasive, easily deployable, and does not restrict the subjects from their day-to-day kinematics.

III. Problem formulation

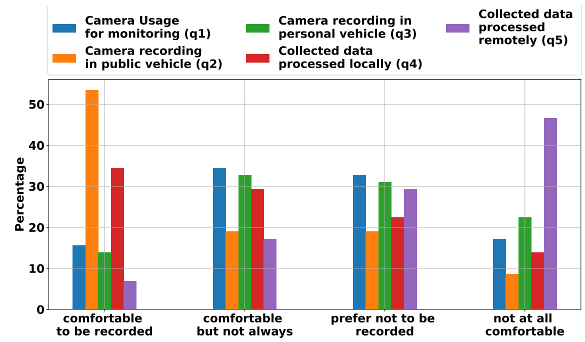
We first conduct an online survey to gauge the requirement of a non-privacy-invasive sensing modality over existing approaches for driver attentiveness monitoring. This section first narrates the survey questionnaire and its Outcome. The succeeding subsections state the problem and identify the challenges that must be addressed for designing the solution.

A. Pilot survey

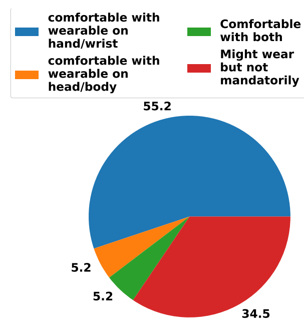
In the mid of August 2022, we conducted an online survey[1]. The objective is to gain insights into the following two questions: (1) What is the general users' view on using a camera to monitor passengers? and (2) What is the general users' perspective on the Usage of wearables and body-mounted sensors while driving? The survey is formulated for individuals in general who do not necessarily have any technical knowledge about how a driver attentiveness monitoring system works. The questionnaire first raises a query, *q1*, for an overall perspective on using a camera for continuous monitoring inside a car. The following two queries, *q2 and q3*, document whether the individual is comfortable with a camera recording inside a public vehicle and a personal vehicle, respectively. The succeeding queries, *q4 and q5* capture whether the user prefers data to be processed locally or transmitted remotely, respectively. The opinions on each of the five queries are measured through a four-point Likert scale. The neutral option has been ignored, assuming the surveyees might move on without carefully considering the question. Each measure was interpreted with textual meaning for the surveyees. The final query, *q6*, assesses a surveyee's perspective as a driver on the usage of wearables on a wrist or using sensors such as EEG (Electroencephalogram) mounted on the head or body. Fifty-eight people from different demographic backgrounds participated in the survey. The Outcome exposes some exciting patterns in peoples' preferences. About 33% of the people have responded to q1 stating that they are alright being recorded but not always, whereas 35% prefer not to be recorded. Only 16% expressed their comfortableness with the camera. The details of the responses can be seen in Fig. 1(a). On observing the responses for q2 and q3, we can see that most people who number more than 54% are only comfortable being continuously monitored in case it is a public vehicle. Whereas, in a personal vehicle, this number drops to just 10%. Within a personal vehicle, the majority of the people (35.4%) do not wish to be monitored all the time. With q4, about 66% of people were comfortable being recorded if the data is processed/stored in a device they own, of which half still wish not to be recorded all the time. However, with q5, 50% of the respondents strongly disliked the idea that data storage and processing is done in a remote system or cloud. Fig. 1(a) detail q4 and q5 further. With q6, the people preferring wearables such as smartwatches or magnetic tags on hand/wrist over sensors attached to their head/body marked the majority with above 52% (see Fig. 1(b)). However, a significant 35.4% wish that any such sensor should not be mandatory during the driving period, which is essential to address the problem. This clearly indicates a reluctance to leverage such devices to monitor a driver.

The survey clearly reflects some critical impressions of the general population. For one, people are fairly comfortable with continuous video monitoring and would avoid being in the midst of it. In addition, people would prefer wearables that can be worn easily, such as on hands and wrists, and even prefer not to be restricted by them.

[1]https://forms.gle/4fC8u3NK9gYvLMw66

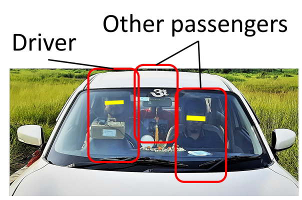


(a)

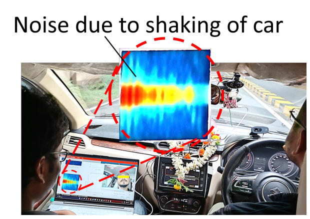


(b)

Fig. 1. Outcome of the pilot survey on the Usage of - (a) camera for monitoring (q1-q5) (b) wearables for monitoring (q6).



(a)



(b)

Fig. 2. (a) Presence of multiple individuals affects the Range Doppler (b) Range Doppler variation due to the shaking of the car.

B. Problem Statement

The objective of the work is *to formulate a minimally invasive driver attentiveness monitoring system leveraging mmWave technology.* We aim to identify a set of states of the driver, which eventually define if the driver is focused on driving.

C. Constituent sub-problems

We now state the research problems identified that need to be addressed to fulfill the primary objective.

Sub-problem 1: Isolation of the driver.
An mmWave radar senses the dynamics of bodies/objects within the environment, including individuals and other

moving entities apart from the driver. For instance, the passengers, as shown in the scenario of Fig. 2(a), also may lie within the radar's field of view. However, the movements of these bodies are irrelevant to our problem and, therefore, must be obliterated by identifying them. Thus, the first challenge is to isolate the driver from the rest of the human subjects.

Sub-problem 2: Pattern annotation.
Whenever we aim to annotate a set of data for *training purpose*, a significant amount of time and labor needs to be invested. Even more importantly, human error during annotation could again mold the prediction model's performance. Therefore, an automated and efficient annotation technique could be significantly beneficial both in terms of time, reliability, and affordability.

Sub-problem 3: Detection of inattentiveness.
The final challenge corresponds to capturing the patterns of inattentiveness. Several driver activities, such as the driver looking in another direction or the driver talking, indicate that the driver lacks adequate focus on driving. Although the existing literature is an essential source of a typical set of such activities, the challenge lies in capturing signatures specific to these activities using mmWave sensing, especially in a noisy scenario (refer Fig. 2(b)) as can be found inside a car. These signatures, in turn, facilitate the automated assessment of driver inattentiveness.

With the problem specifications in hand, we now move ahead to state the relevant preliminaries.

IV. Preliminaries

In this section, we first introduce the essential factors involved with an mmWave radar which are relevant to capture the driver's signature specific to attentiveness.

The primary working principle of an mmWave radar involves the transmission of short wavelength (1-10 mm) electromagnetic waves. Currently, the frequency bands of mmWave radars are mainly divided into three categories: 24 GHz frequency band, 60-64 GHz frequency band, and 77–81 GHz frequency band. The radar used in this paper works at 60-64 GHz. Because of this wide frequency band (4GHz), it has a high range resolution ($\approx$ 4 cm) and is suitable for distinguishing different actions of the human body.

Frequency Modulated Continuous Wave (FMCW) is the most used modulation method for these automotive mmWave radars. FMCW radar system mainly includes a transceiver antenna, RF front-end, modulation signal, a signal processing module, etc. The radar receives the reflection from an object to estimate the *range, velocity, and angle* of the object.

FMCW radar uses a linear *'chirp'* or swept frequency transmission. When receiving the signal reflected by an object in the path, the radar performs a dechirp operation by mixing the transmitted signal with the reflected signal, producing an Intermediate Frequency (IF) signal. The principle of ranging and speed measurement of the FMCW radar is based on signal processing on this IF signal.

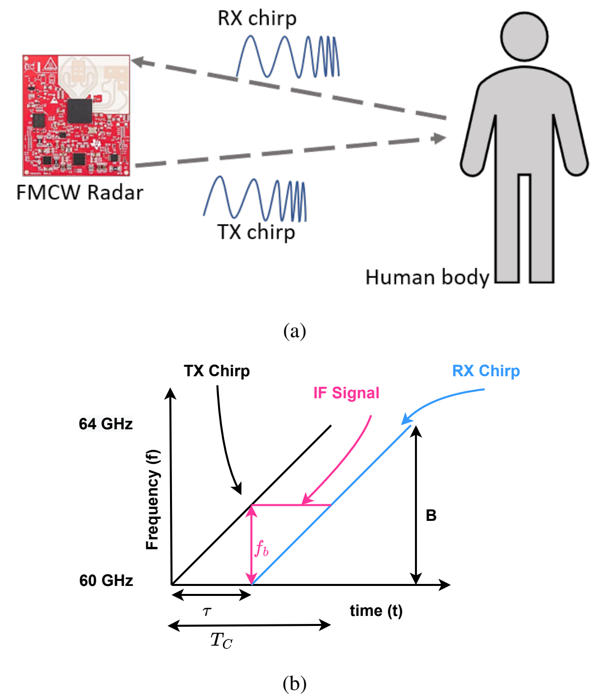


Fig. 3. Working principle of an FMCW Radar - (a) radar function, (b) chirp frequency vs time.

A. The Principle of Ranging of FMCW Radar

FMCW radar generates continuous waves in the form of periodic signals, where the frequency increases linearly with time. Each of these signal forms is referred to as *chirps*. There is a certain frequency difference between the signal reflected by the object and the transmitted signal. The distance information between the target and the radar can be obtained by measuring this frequency difference [34].

As shown in Fig. 3(b), suppose we have a transmitted chirp (TX chirp) with transmission time T_C. Due to an object present at a distance of d from the radar, we have a received chirp (RX chirp) after a time delay of τ. The IF signal generated by mixing the TX and RX chirp has a beat frequency, say f_b. The slope (S) of this FMCW chirp can be represented as,

$$S = \frac{B}{T_C} = \frac{f_b}{\tau}. \tag{1}$$

Also, the time delay τ can be derived as,

$$\tau = \frac{2d}{c} \tag{2}$$

where d is the distance of the detected object and c is the speed of light. Thus, the distance of the detected object can be given as,

$$d = \frac{c}{2} \cdot \frac{T_C}{B} \cdot f_b \tag{3}$$

To sum up, the detection range of the FMCW radar is determined by Equation 3, and the key parameter is the frequency difference (f_b) between TX and RX chirp. Now to determine this beat frequency, a Fast Fourier Transform (FFT)

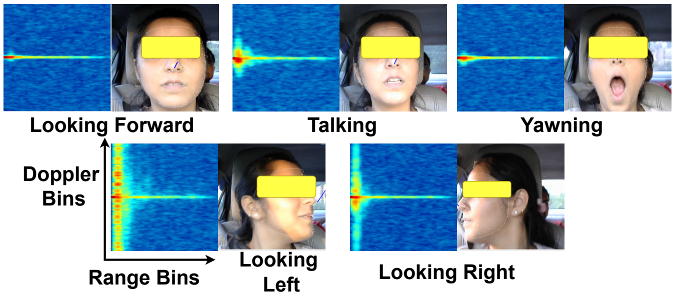


Fig. 4. Variation in range-doppler heatmap with different driver activities under *in-lab* setup

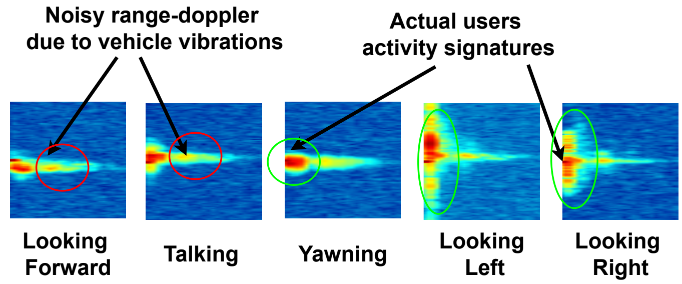


Fig. 5. Variation in range-doppler heatmap with different driver activities under driving scenarios

is conducted on the IF signal (f_b). This range-FFT produces frequency peaks at locations where the user is present. And based on equation 3, the corresponding range is calculated.

B. The Principle of Speed Measurement of FMCW Radar

In order to measure velocity, an FMCW radar transmits N number of chirps separated by a transmission time of T_C. On applying range-FFT (as discussed in Sec. IV-A), the objects present at different locations are captured. If the object (say a user) is moving, then the range-FFT corresponding to each chirp will have peaks in the same location but with a different phase. If the user is moving with a speed of v, the measured phase difference between two successive RX chirps corresponding to a motion of vT_C can be given as,

$$\Delta\phi = \frac{4\pi v T_C}{\lambda} \tag{4}$$

A second FFT, called *Doppler-FFT*, is performed on these N phasors to determine the movement or velocity of the object. Thus, by including both the range information and the doppler information, we can capture the things within the field of view and their movement. This information is captured in the form of a 2D matrix called *range-doppler*. Next, we will discuss how this *range-doppler* information is being used for capturing the driver's behavior.

V. System Design

To capture a driver's attentiveness, in this work, we primarily relied on *range-doppler* based data. We observe this heatmap information is the correct choice for capturing different driver's activity signatures. As shown in Fig. 4 with varying orientations of the driver's head or mouth motion, we observe a significant variation in the range-doppler heatmap. Thus based on these observations, we finally define a set of activity classes relevant for determining a driver's attentiveness.

A. Associated Classes

We assess the driver's inattentiveness primarily from two perspectives: 1) If the driver is diverted from driving the vehicle. This is characterised by three *classes*: C_1, when the driver is *looking Right*; C_2, when the driver is *looking left*; and C_3, when the driver is *talking*. And 2) If the driver is feeling drowsiness and therefore is fatigued. We characterise this using *class*, C_4, when the driver is *yawning*. Apart from these, we also consider the scenario when the driver is attentive. We characterize this by considering a fifth class, C_5 when the driver is *looking forward*. Based on the associated classes, we capture the corresponding range-doppler heatmap.

B. Data pre-processing and augmentation

As shown in Fig. 4 the range-Doppler is in a picture-like form. Its abscissa is the range, the ordinate is the Doppler speed, and the value in the picture is the thermal value. We export the range-Doppler data from each frame separately, where each frame is a 2D array of size 128x64. Here 128 is the number of doppler bins, and 64 is the number of range bins. In the *in-lab* setup as shown in Fig. 4 we observe that the doppler values in the higher range bins are invariant due to lack of movement, or it captures the movement of unintended objects. Also for the *in-car* setup as shown in Fig. 5, we observe these unintended doppler variation in the higher range bins because of vehicle vibrations, other users present inside the car. Hence, we only took doppler values up to 2m range bins. Due to the sparsity of the range-doppler, we resize it to 48x48, reducing the computation in subsequent steps without significant loss in fidelity. The activities (i.e., looking left, looking right, yawning, etc.) span over a short time and, therefore, have a temporal impact on the shift in range-doppler values. We stack multiple consequent frames of the range-doppler to improve the data quality and achieve a 2D multichannel array with a size of 48x48x4. The primary reason behind stacking together 4 range-doppler frames is that range-doppler is captured at 2 fps, and each activity shows distinguishable range-doppler patterns with the first 2 seconds. Next, to make the framework robust against the user's distance from the sensing setup, we augment the collected dataset by randomly shifting the stacked range-doppler frames in the range direction, which simulates the user's distance change with respect to the mmWave radar without changing their activity labels.

C. Model Architecture

The range-doppler frames have different spatial patterns for each activity, which persists throughout that activity. Therefore, to exploit such Spatio-temporal correlation, we have used a 2D Convolutional Neural Network architecture (2D-CNN) as shown in Fig. 6. Convolution 2D operation takes into account

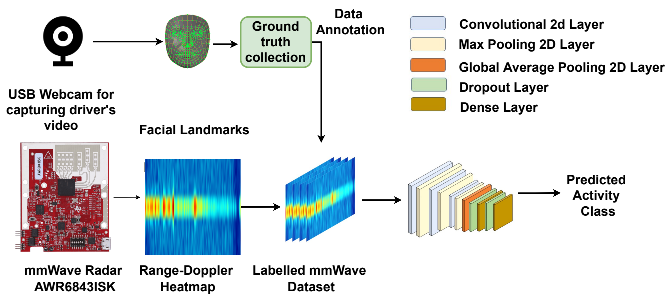


Fig. 6. System Overview

TABLE I
2D-CNN ARCHITECTURE

CNN layer	Parameters			Dimensions
	Kernel	Channels	Drop rate	
Input layer	–	4	–	48 × 48 × 4
conv1	5 × 5	32	–	48 × 48 × 32
maxpool1	3 × 3	–	–	16 × 16 × 32
conv2	3 × 3	64	–	16 × 16 × 64
maxpool2	2 × 2	–	–	8 × 8 × 64
conv3	3 × 3	96	–	8 × 8 × 96
maxpool3	2 × 2	–	–	4 × 4 × 96
G-avg pool	–	–	–	1 × 96
dropout1	–	–	20%	1 × 96
dense1	–	32	–	1 × 32
dropout2	–	–	10%	1 × 32
softmax	–	5	–	1 × 5

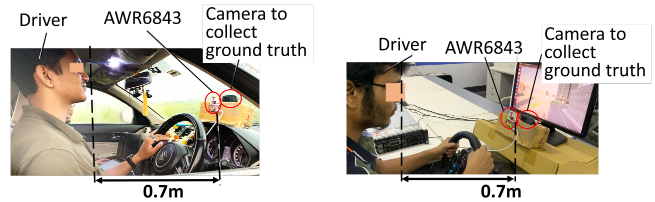
the dependency of neighboring values within all possible kxk regions at each range-doppler frame and the temporal relationship of past 4 such frames. Further, it computes several cross-channel feature maps which are useful for subsequent layers of the 2D-CNN. In the convolution architecture, we have used three 2D Convolutional Layers with 'same' padding and Relu activation alongside 2D Max Pooling Layers. Next, a Global Average Pooling Layer is added to extract the average spatial activation across the entire feature map. Finally, we add two successive Dropout and Dense layers, where the dropout rate is kept as 20% and 10% for the two Dropout layers, respectively, to prevent over-fitting. The last layer has 5 neurons with softmax activation to output a joint probability distribution over the five activity classes. More details on the 2D-CNN network architecture is shown in TABLE I. We evaluate our 2D-CNN model with two other state-of-the-art classifiers in Section VI-C.

VI. IMPLEMENTATION AND EVALUATION

We deploy the experimental setup in two environments: 1) an *in-lab* setup, which can be characterized by comparatively less noise and is primarily aimed to validate *mmAssist* and 2) an *in-car* set up to study the practical scenario of driver monitoring. In either of the cases, the components used and the relative placements of the devices are identical.

A. Hardware setup and deployment

In this work, we have used AWR6843ISK [35], an FMCW-based mmWave radar from Texas Instruments. The radar works in the frequency range of 60-64GHz with a range resolution of approximately 4 cm, which is adequate for measuring the activities of interest listed in section V-A. The radar's maximum range is up to a distance of 10 meters, with a Field-of-View of $-70°$ to $+70°$ on azimuthal and elevation planes. This Field-of-View is adequate for monitoring the driver from within the car. In the *in-car* setup, the radar is placed on the dashboard of a 5-seater sedan car, facing the driver, as shown in Fig. 7(a). The driver, in this case, is sitting at a distance of 0.7 meters from the radar. However, he/she can adjust his/her seat and hence the distance from the radar, as long as he/she is within its vicinity. The same applies for the *in-lab* setup as shown in Fig. 7(b), which involves a display monitor and steering wheel[2] based simulation setup mimicking driving scenarios. In either of the cases, a Logitech USB camera[3] is placed next to the radar to capture the necessary ground truth information from the driver. The camera and the radar are connected to a PC using USB ports, where the data is collected. The various parameter configurations of the radar are shown in TABLE II. As seen from the table, the radar generates 128 chirps every frame. 2 frames are collected by the radar per second, which is adequate to measure the movements of concern and, at the same time, functions at the optimal amount of baud-rate (921600 bps) for transferring through the connected USB port. The collected data is stored in the connected PC with the timestamps provided by the PC, thereby maintaining time synchronization between the sensed information and the camera capture.



(a) (b)

Fig. 7. Data collection Setup for - (a) *in-car* and (b) *in-lab* scenarios.

TABLE II
RADAR CONFIGURATION

Parameters	Value
Start Frequency (GHz)	60
Scene Classifier	Best Velocity Resolution
Range Resolution (m)	0.083
Maximum Unambiguous Range (m)	9.06
Maximum Radial Velocity (m/s)	0.32
Radial Velocity Resolution (m/s)	0.01
Frames per Second	2
Number of chirps per frame	128
Number of ADC Samples	256

[2] https://www.logitechg.com/en-in/products/driving/driving-force-racing-wheel.941-000143.html
[3] https://www.logitech.com/en-in/products/webcams.html

B. Data collection and annotation

The data is collected from four volunteers at different times of the day. In the *in-car* setup, different times of the day observe different traffic scenarios. The driving environment ranges from both within the campus to the suburbs nearby. These road environments have significance, as the campus roads are much smoother than the suburbs. The volunteers have different ages, gender, and heights, and most importantly, different driving patterns. In the *in-lab* setup, the volunteers can sit in front of the setup and drive using the Carla [36] simulator. In the *in-car* setup, the volunteers were instructed to adjust their seats as per their will and drive as they please. However, to attain a dataset rich with adequate distribution of events, the volunteers were instructed to simulate the scenarios specified in section V-A in either setup. The *in-car* setup also observes the presence of passengers in the co-driver seat as well as behind the driver. The passengers were also free to do their natural movements, as seen in a car.

The data is collected in continuous streams, each for six to eight minutes, and stored in the data collection PC. Using mediapipe [37] framework, facial landmarks of the driver are collected for annotation of the driver activity classes (See Section V-A), which removes the manual annotation process as done in previous works, such as [7]. It is a *Human-in-Loop* annotation process as we need to provide only the threshold hyperparameters such as left and right head orientation thresholds as well as mouth opening thresholds for yawning and talking activities. However, the camera frame rate is 30FPS, whereas the mmWave data has a frame rate of 2 FPS. Thus, we down-sample the ground truth data to 2 FPS by taking the most repeating activity class within 15 video frames. Finally, based on matching the timestamp, mmWave range-doppler data is annotated with mediapipe generated labels. Now we discuss the model architecture used for predicting driver activities using mmWave data.

C. Baselines

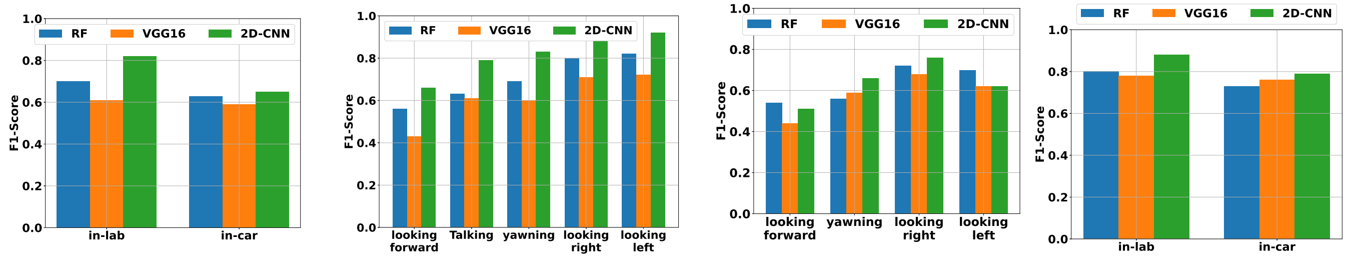
We have used the software package based on Python 3.8.12, TensorFlow v2.9.1, and Scikit-learn v1.0.2 for implementing the 2D-CNN-based driver activity prediction model alongside other two state-of-the-art models: Random Forest (RF) and VGG-16 [38]. The RF model's features are engineered to take the range-doppler's min, max, mean and standard deviation within a kernel size of 16x16 for all 4 range-doppler 2D array of size 48x48, resulting $(\frac{48}{16})^2$x4x4 = 144 features. We have taken the Random Search Cross Validation approach [39] to search for the best hyperparameters for the RF classifier within a wide range of values for each hyperparameter, performing K-Fold cross-validation with each combination of the hyperparameter values. Based on that, we have selected 240 estimators with a max depth of 50. On the other hand, we have used VGG-16 [38] network, pre-trained on the ImageNet [40] dataset to do transfer learning. This transfer learning approach helps in reducing the feature extraction part from high dimensional image-like input data, as all the trained convolutional layers in VGG-16 are used as feature extractors and do not require retraining. On top of this base VGG-16 model, we have added and trained 2D-Global Average Pooling and successive Dropout and Dense layers, as done in the 2D-CNN for classifying driver activities. The models are trained with a train-test split of 70%-30% and a validation split of 20% from the training set.

D. Evaluation

In this section, we evaluate the performance of *mmAssist* by measuring the activity classification's accuracy under different datasets and subjects using 2D-CNN, RF, and VGG-16 classifiers.

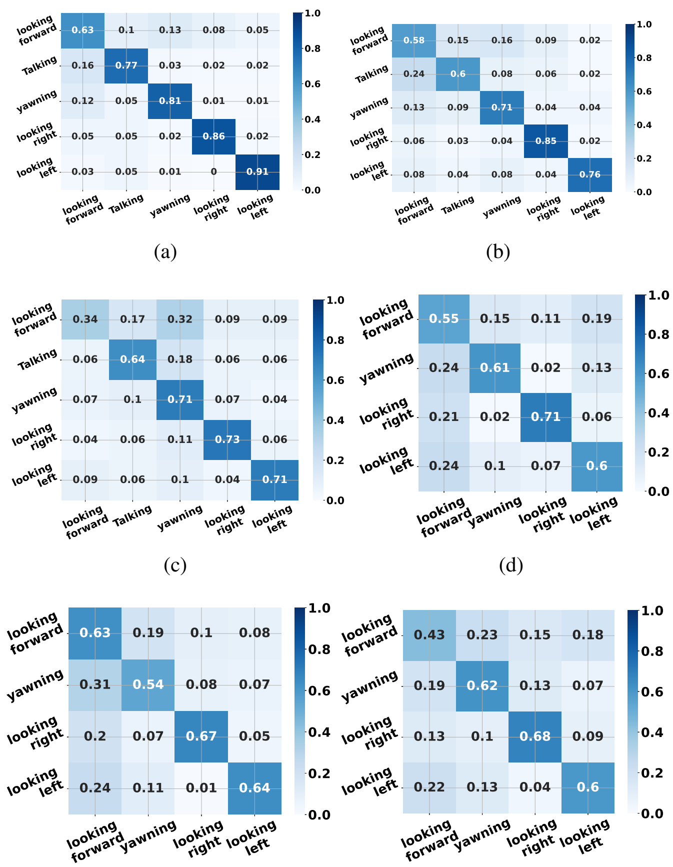
Fig. 8(a) shows the overall weighted F1-Score for driver activity classification under the *in-lab* and *in-car* datasets. In both, the datasets F1-Score of 2D-CNN based classifier is higher compared to the RF and VGG-16 based classifiers. The primary reason behind poor performance with VGG-16 based classifier is that this model is pre-trained with the imagenet dataset and doesn't fit well in engineering features from a 2D range-doppler heatmap, which is mostly sparse across the range bin axis except for the range bin where the driver is present. Although for the RF classifier, the features are passed across a window size of 16x16, this window can determine the users' position and doppler variation over the entire range-doppler heatmap. Thus we observe a slightly better performance for the RF-based classifier in comparison to VGG-16. On the other hand 2D-CNN has more freedom in engineering the spatio-temporal features as the features are not hand-engineered like RF (min, max, mean, standard deviation) and also a lower inference time because of lower number of convolutional layers (compared to VGG-16). The overall performance of *in-lab* dataset is higher in comparison to the *in-car* dataset. This is because, in real driving, unexpected situations interfere with the range-doppler variation (e.g., frequent stops, vehicle vibration, road humps, etc.). This leads to unpredictable variation in the range-doppler which might overlap with the variation due to user's activities. Thus, *in-car* setup gives a slightly lower accuracy in comparison to *in-lab* setup.

To understand the classification performance in more detail, in Fig. 8(b) and 8(c) we show the F1-Score for different activities. As observed in Fig. 8(b), the F1-Score for looking left, looking right, and yawning is higher than 80% for the 2D-CNN model. However, the accuracy is lower 77% and 63% in the case of talking and looking forward respectively. This is primarily because the range-doppler information captured using mmWave modality can accurately estimate the sudden variation or change within the Field-of-View. Looking left or right and even yawning activities occur in a small-time window, and thus these events cause immediate phase shifts in the reflected mmWave signal. And thus, using the range-doppler heatmap, these activities can be classified by the learning model. However, talking and looking forward are rather continuous activities that occur for a longer time duration, where the phase shifts don't occur in a sudden manner like the other three activities. In Fig. 9(a), 9(b),



(a) (b) (c) (d)

Fig. 8. The weighted F1-Score attained in different experimental scenarios - (a) overall activity detection, across different activities in (b) *in-lab* (c) *in-car*, and (d) attentiveness detection.



(a) (b) (c) (d) (e) (f)

Fig. 9. Activity detection confusion matrix for 2D-CNN, RF and VGG-16 on - (a)(b)(c) *in-lab* and (d)(e)(f) *in-car* datasets, respectively.

9(c) we report the confusion matrix for the three classifiers. From the confusion matrix also, it is clear that there exists a slight miss-classification between continuous time activities looking forward and talking. Based on this observation, we have removed talking from the *in-car* dataset and reported the weighted F1-Score for all the other four driver's activities. As shown in Fig. 8(c), the test accuracy for *in-car* setup is lower compared to *in-lab* dataset as the doppler shifts occur not only because of the driver but also because of the vehicle vibrations and road conditions. Also, the confusion matrix for the three classifiers under *in-car dataset* is shown in Fig. 9(d), 9(e), 9(f).

Finally, to infer driver's attentiveness report, we have considered two final classes, namely, attentive and inattentive. We merged looking left, looking right, talking and yawning together as the inattentive class and looking forward as the attentive class. Finally, we trained all the three classifiers with this modification and report the weighted F1-Score in Fig. 8(d). As shown in Fig. 8(d), for the 2D-CNN classifier, we achieve a weighted F1-Score of 88% for the *in-lab* dataset and 79% for the real-world *in-car* dataset.

Thus, from our evaluation study, we may infer that using mmWave modality, we can indeed pervasively sense the key activity features needed for drivers' attentiveness. However, detecting continuous macro activity, such as looking forward or talking, remains challenging. Also, indoor activity sensing is much easier than in-vehicle activity monitoring, as small-scale vibrations can also impact the doppler information.

VII. Conclusion and Future Works

Monitoring driver attentiveness is a significant problem of interest that has encouraged researchers to explore diverse domains. In this direction, we propose a minimally invasive approach leveraging a single off-the-shelf mmWave sensor, which is expected to be a pervasive technology with the evident paradigm shift of wireless communications towards 5G. In this proposed idea called, *mmAssist* we formulate three different models for predicting lack of attention on the part of the driver. We validate *mmAssist* first with an in-lab setup and then study its applicability in a real-world scenario. We strongly believe that the minimal invasive footprint of *mmAssist* could be a significant reason for it to mark its place as a practical and prominent commercial solution. In future work, we wish to explore a possible rating mechanism for the driver on a suitable scale to measure driving quality. Also in real-world driving we observe a slight lower accuracy in *mmAssist* due to vehicle vibrations or road conditions etc. In our future work, we wish to address these challenges. For this, we would also like to integrate image from outside the car and vehicle IMU data as an additional modality and make the system more robust and even more practical.

References

[1] A. Sen, "mmassist." [Online]. Available: https://github.com/arghasen10/mmAssist

[2] W. Wierwille, R. Hanowski, J. Hankey, C. Kieliszewski, S. Lee, A. Medina, A. Keisler, and T. Dingus, "Identification and evaluation of driver errors: Overview and recommendations," 2002.

[3] Otonomo and E. Research, "What american drivers think about connected car data and privacy," Otonomo and Edison Research, Herzeliya Pituach, Israel, Tech. Rep., August 2018.

[4] H. Laouz, S. Ayad, and L. S. Terrissa, "Literature review on driver's drowsiness and fatigue detection," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE, 2020, pp. 1–7.

[5] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Intelligent driver drowsiness detection system using uncorrelated fuzzy locality preserving analysis," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 4608–4614.

[6] H. Yang, L. Liu, W. Min, X. Yang, and X. Xiong, "Driver yawning detection based on subtle facial action recognition," *IEEE Transactions on Multimedia*, vol. 23, pp. 572–583, 2020.

[7] I. Dua, A. U. Nambi, C. Jawahar, and V. Padmanabhan, "Autorate: How attentive is the driver?" in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–8.

[8] I. Janveja, A. Nambi, S. Bannur, S. Gupta, and V. Padmanabhan, "Insight: monitoring the state of the driver in low-light using smartphones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.

[9] P. Aditya, R. Sen, P. Druschel, S. Joon Oh, R. Benenson, M. Fritz, B. Schiele, B. Bhattacharjee, and T. T. Wu, "I-pic: A platform for privacy-compliant image capture," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 235–248.

[10] M. Z. Ozturk, C. Wu, B. Wang, and K. R. Liu, "Gaitcube: Deep data cube learning for human recognition with millimeter-wave radio," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 546–557, 2021.

[11] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 4, no. 4, pp. 1–28, 2020.

[12] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 211–220.

[13] H. Huang, H. Chen, and S. Lin, "Magtrack: Enabling safe driving monitoring with wearable magnetics," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 326–339.

[14] T. Ersal, H. J. Fuller, O. Tsimhoni, J. L. Stein, and H. K. Fathy, "Model-based analysis and classification of driver distraction under secondary tasks," *IEEE transactions on intelligent transportation systems*, vol. 11, no. 3, pp. 692–701, 2010.

[15] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 2010, pp. 1–8.

[16] F. Friedrichs, M. Miksch, and B. Yang, "Estimation of lane data-based features by odometric vehicle data for driver state monitoring," in *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2010, pp. 611–616.

[17] B. Cyganek and S. Gruszczyński, "Hybrid computer vision system for drivers' eye recognition and fatigue monitoring," *Neurocomputing*, vol. 126, pp. 78–94, 2014.

[18] G. Borghi, M. Venturelli, R. Vezzani, and R. Cucchiara, "Poseidon: Face-from-depth for driver pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4661–4670.

[19] I. Dua, T. A. John, R. Gupta, and C. Jawahar, "Dgaze: Driver gaze mapping on road," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5946–5953.

[20] C. Yan, Y. Wang, and Z. Zhang, "Robust real-time multi-user pupil detection and tracking under various illumination and large-scale head motion," *Computer Vision and Image Understanding*, vol. 115, no. 8, pp. 1223–1238, 2011.

[21] S. Kajiwara, "Driver-condition detection using a thermal imaging camera and neural networks," *International journal of automotive technology*, vol. 22, no. 6, pp. 1505–1515, 2021.

[22] K. Das, S. Sharak, K. Riani, M. Abouelenien, M. Burzo, and M. Papakostas, "Multimodal detection of drivers drowsiness and distraction," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, 2021, pp. 416–424.

[23] T. Hoang Ngan Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides, "Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 46–53.

[24] Z. Li, S. Bao, I. V. Kolmanovsky, and X. Yin, "Visual-manual distraction detection using driving performance indicators with naturalistic driving data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2528–2535, 2017.

[25] N. E. Haouij, J.-M. Poggi, S. Sevestre-Ghalila, R. Ghozi, and M. Jaïdane, "Affectiveroad system and database to assess driver's attention," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, ser. SAC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 800–803.

[26] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of eeg and emg signals for wireless biometrics," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[27] T. Kundinger, P. K. Yalavarthi, A. Riener, P. Wintersberger, and C. Schartmüller, "Feasibility of smart wearables for driver drowsiness detection and its potential among different age groups," *International Journal of Pervasive Computing and Communications*, 2020.

[28] Y. Cao, F. Li, X. Liu, S. Yang, and Y. Wang, "Towards reliable driver drowsiness detection leveraging wearables," *ACM Transactions on Sensor Networks (TOSN)*, 2022.

[29] X. Xu, H. Gao, J. Yu, Y. Chen, Y. Zhu, G. Xue, and M. Li, "Er: Early recognition of inattentive driving leveraging audio devices on smartphones," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[30] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D 3-guard: Acoustic-based drowsy driving detection using smartphones," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1225–1233.

[31] H. Jiang, J. Hu, D. Liu, J. Xiong, and M. Cai, "Driversonar: Fine-grained dangerous driving detection using active sonar," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–22, 2021.

[32] Y. Bai, L. Lu, J. Cheng, J. Liu, Y. Chen, and J. Yu, "Acoustic-based sensing and applications: A survey," *Computer Networks*, vol. 181, p. 107447, 2020.

[33] V. Hromadová, P. Kasák, R. Jarina, and P. Brída, "Frequency response of smartphones at the upper limit of the audible range," in *2022 ELEKTRO (ELEKTRO)*. IEEE, 2022, pp. 1–5.

[34] C. Iovescu and S. Rao, "The fundamentals of millimeter wave radar sensors," *Texas Instruments*, 2020.

[35] T. Instruments, "User's guide 60ghz mmwave sensor evms," *online], https://www.ti.com/lit/pdf/swru546*.

[36] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[37] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[39] W. Koehrsen, "Hyperparameter tuning the random forest in python," *Towards Data Science*, 2018.

[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.