Classification of Traffic Signaling Motion in Automotive Applications Using FMCW Radar

Sabyasachi Biswas, *Student Member, IEEE,* Benjamin Bartlett, *Student Member, IEEE,*
John E. Ball *Senior Member, IEEE* and Ali C. Gurbuz, *Senior Member, IEEE,*

***Abstract*—Advanced driver-assisted system (ADAS) typically includes sensors such as Radar, Lidar, or Camera to make vehicles aware of their surroundings. These ADAS systems are presented to a wide variety of situations in traffic, such as upcoming collisions, lane changes, intersections, sudden changes in speed, and other common instances of driving errors. One of the key barriers to automotive autonomy is the inability of self-driving cars to navigate unstructured environments, which typically do not have any traffic lights present or operational for directing traffic. In these circumstances, it is much more common for a person to be tasked with directing vehicles, either by signaling with an appropriate sign or via gesturing. The task of interpreting human body language and gestures by autonomous vehicles in traffic directing scenarios is a great challenge. In this study, we present a new dataset collected of traffic signaling motions using millimeter-wave (mmWave) radar, camera, Lidar and motion-capture system. The dataset is based on those utilized in the US traffic system. Initial classification results from Radar microDoppler (μ-D) signature analysis using basic Convolutional Neural Networks (CNN) demonstrates that deep learning can very accurately (around 92%) classify traffic signaling motions in automotive applications.**

***Index Terms*—Micro-Doppler, autonomy, traffic gesture classification, mmWave, ADAS, CNN.**

I. Introduction

In recent times, due to the creation of reasonably priced solid-state transceivers and powerful graphics processing units (GPUs), the radio frequency (RF) sensors have gained much popularity in wide range of applications including advanced driver assisted systems (ADAS) and autonomy due to their advanced sensing capability, size and low cost. Side by side, deep leaning (DL) has also gained much popularity in many fields due to its representation capability directly from data. The use of DL has added a new dimension and raised the bar for radar-based human activity recognition (HAR) performance compared to earlier decades [1]. Some of the applications where radar has gained much popularity over the last decades include defense and security [2], UAV classification [3], ADAS [4], indoor monitoring human activity recognition [5], sign language recognition [6], and contact-less health monitoring [7].

In ADAS and autonomous vehicle (AV) systems radar is being utilized as one of the main sensing systems. AVs are currently trained to operate pristine road conditions. However, in our daily lives we face many situations where the traffic rely on human directions to navigate. This would be the case, for example, when a vehicle passing through a construction zone, in and out of school or related high traffic areas, or when automatic traffic lights are not functional. Under all these circumstances it is much more common for a person to be tasked with directing vehicles, either by signaling with an appropriate sign or via gesturing. This person could be a traffic officer or sometimes a school official, or a construction worker. If the AV were passing through any of these zones, it will have to rely on human directions to navigate and it should autonomously recognize, and classify human traffic directions.

This paper presents an initial study towards understanding the success of varying sensors including mm-wave automotive radar systems in order to recognize gestures from human traffic directors. While optical sensing have been explored in this area [8], radar has not been tested for for the problem. 12 different motions based on those highly used to direct traffic in the US traffic system are selected (See Section II-B and Fig.3 for details on the traffic gestures). First a dataset is collected in lab environment from 14 number of participants using RGB-Depth cameras, Lidars, mm-wave radars and motion-capture system that record the position of the body parts of the participant implementing the traffic directions. In this dataset single human directors using pre-defined classes of directions within the line of sight of the sensor suite are considered.

In this paper, we focus specifically on the mm-wave radar and present the effectiveness of radar for understanding of human traffic directions. At first, time-frequency domain transformation is done and micro-Doppler(μ-D) signatures have been generated from the collected radar data cubes. Afterwards, the μ-D signatures have been converted to images and classified using the developed convolutional neural network (CNN) architecture. Approximately 91% and 93% testing accuracy was achieved using RGB and gray-scale images respectively. These results are initial indications that radar is a highly effective sensor for traffic signalling motion classification.

In the upcoming sections of this paper the analysis of the dataset will be presented. The paper is organized as follows: the experimental setup and dataset is provided in Section II, mmWave FMCW radar signal model, μD spectrogram generation and the proposed CNN architecture are discussed in the Section III and Section IV presents the evaluation results of the proposed algorithm. Finally, Section V concludes the paper and discusses future works.

2023 IEEE Radar Conference (RadarConf23) | 978-1-6654-3669-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/RADARCONF2351548.2023.10149728

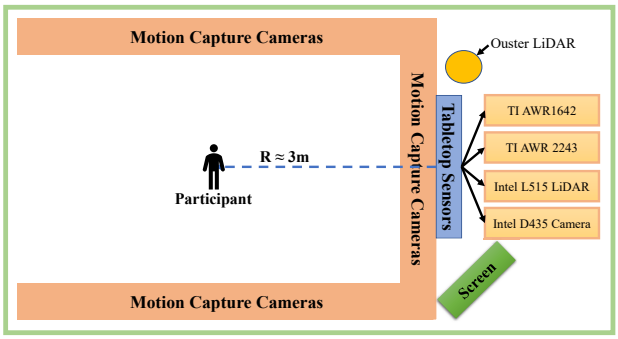


Fig. 1: Experimental Setup

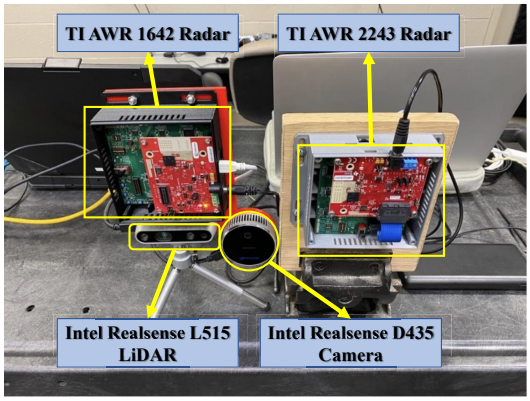


Fig. 2: Table Setup

II. Experimental Setup and Dataset

A. Experimental Setup

To perform motion analysis and classification, six distinct sensors are used in the data collection to capture both kinematic movement and visual data. Four of the six sensors used are intended for short range automotive applications. They are TI AWR2243 mmWave radar, a TI AWR1642 mmWave radar, an Intel Realsense L515 LiDAR, and an Intel Realsense D435 camera. The fifth and sixth sensors are an Ouster OS-1 360° scanning LiDAR and MotionMonitor motion capture system with Vicon motion capture cameras respectively. These six sensors are controlled in three different data collection environments. The AWR2243 is controlled using TI's mmWave Studio software for raw voltage collection; the AWR1642, L515, D435, and OS-1 sensors are controlled using the Robot Operation System (ROS) for radar scan, RGB, and point cloud information; and the MotionMonitor system collects three-dimensional positional data for key centroids on the participants' upper bodies, as well as rotation and flexion of joints and the head. Figure 1 below shows the setup for the automotive sensors, as well as the layout of the capture space of the laboratory, while Fig. 2 shows the setup for the four short ranged sensors used in the experiment.

TABLE I: AWR2243 Radar Parameters

Parameter	Value
Number of ADC Samples	256
Number of TX Channels	3
Number of RX Channels	4
Starting Frequency	77 GHz
Frequency Slope	65.998 MHz/µs
Bandwidth	3959.88 MHz
Pulse Repetition Interval (PRI)	161.29 µs
Sampling Rate	18750 kHz
RX Gain	48 dB
Periodicity	40 ms
Number of Chirp Loops per Frame	248
Number of Frames	3875
Total Time	155 sec

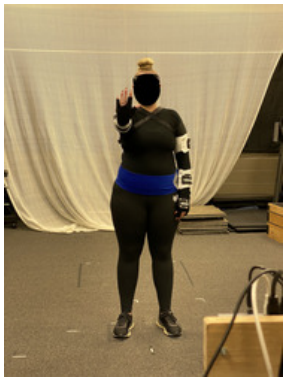
For now, only TI AWR2243 automotive FMCW radar from Texas Instruments is considered for classification. The radar operates from 77 to 81 GHz, with three transmitting (TX) channels and four receiving (RX) channels. By Using TI's mmWave Studio software, users can collect data using an attached DCA1000 EVM capture card and store it as a binary file. The device can be initialized with different parameters depending on the situation. Table I shows the parameters set for the AWR2243 radar for the experiment.

The radar was positioned on top of a table with an elevation of 1 meters prior to data collection. Participants were positioned at a distance of 3 meters in front of the radar. A computer display was placed left next to the radar, but out of the radar's field of view (FOV). The monitor provided prompts specifying the gestures that needed to be articulated. The data was collected for 155 seconds. Within this time, four different gestures have been performed with 5 repetitions each. Between each repetition there was a gap of 1 second. And after each gesture a 10 second gap was given to preview the next gesture. Each sample duration was around 5 seconds.

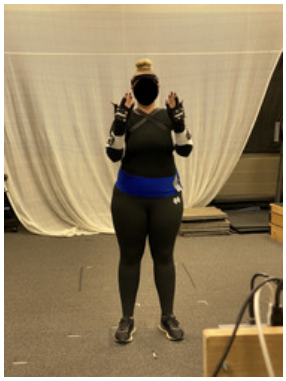
B. Dataset

For this study, a dataset consisting 12 traffic signalling motions based on US traffic system is created. These motions are designed to direct an oncoming vehicle to either stop, move from the stopping point in one of three directions, or to have traffic in any given position wait for other traffic to proceed onto the road. The gestures involve movement of not only the arms of the participants but also the hands, as well as the rotation of the head to look in specific directions. Figure 3 shows each gesture performed by a participant, along with a brief description of each gesture.

1) **Stop** – stop the oncoming vehicle; the participant moves their right hand directly in front of them, with the elbow extended (straight) and the hand open.
2) **Go** – let the oncoming vehicle move forward; the participant both hands directly in front of them, with the elbows extended; the participant then flexes (bends) both elbows, moving the hands towards the shoulders.
3) **Continue** – let oncoming vehicles already moving forward continue to move forward; the participant puts their right hand directly in front of them, with the elbow

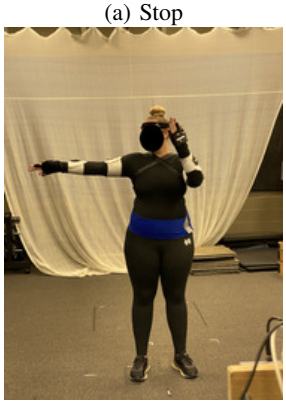


(a) Stop



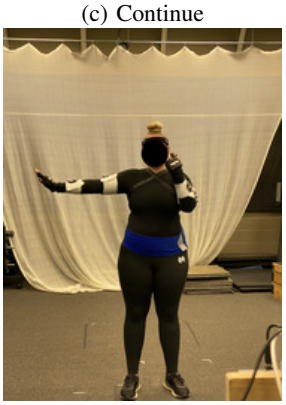
(b) Go

(c) Continue



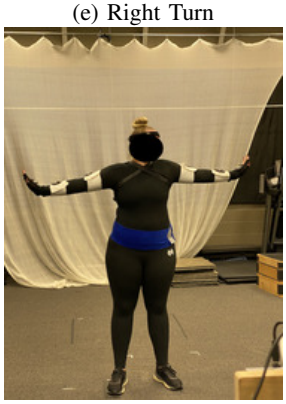
(d) Left Turn

(e) Right Turn

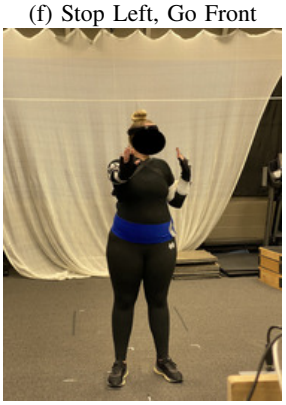


(f) Stop Left, Go Front

(g) Stop Right, Go Front

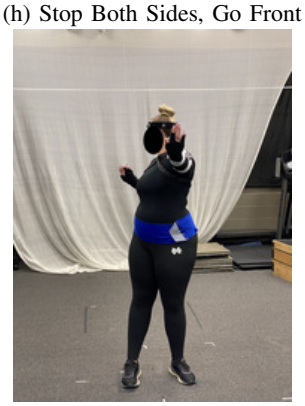


(h) Stop Both Sides, Go Front



(i) Stop Front, Go Right

(j) Stop Front, Go Left



(k) Stop Front, Go Back

(l) Stop Back, Go Front

Fig. 3: The 12 Traffic Gestures

extended; the participant then flexes the right elbow, moving the right hand towards the right shoulder.

4) **Left Turn** – direct oncoming traffic to take a left turn; the participant moves their right arm to the side (shoulder abduction) with the elbow extended, and holds the right arm in the position, pointing to the participant's right; the participant puts their left hand directly in front of them, with the elbow extended; the participant then flexes the left elbow, moving the left hand towards the left shoulder.
5) **Right Turn** – direct oncoming traffic to take a right turn; the participant moves their left arm to the side (shoulder abduction) with the elbow extended, and holds the left arm in the position, pointing to the participant's left; the participant puts their right hand directly in front of them, with the elbow extended; the participant then flexes the right elbow, moving the right hand towards the left shoulder.
6) **Stop Left, Go Front** – stop traffic on the left of the oncoming vehicle, then direct the oncoming vehicle to proceed forward; the participant moves their right arm to the side (shoulder abduction) with the elbow extended, and holds the right arm in the position, with an open hand; the participant puts their left hand directly in front of them, with the elbow extended; the participant then flexes the left elbow, moving the left hand towards the right shoulder.
7) **Stop Right, Go Front** – stop traffic to the right of the oncoming vehicle, then direct the oncoming vehicle to proceed forward; the participant moves their left arm to the side (shoulder abduction) with the elbow extended, and holds the left arm in the position, with an open hand; the participant puts their right hand directly in front of them, with the elbow extended; the participant then flexes the right elbow, moving the right hand towards the right shoulder.
8) **Stop Both Sides, Go Front** – stop traffic on both sides of the oncoming vehicle, then direct the oncoming vehicle to proceed forward; the participant raises both shoulders to the side (abduction) with the elbows extended and the hands open; the participant then puts both hands directly in front of them, with the elbows extended; the participant then flexes (bends) both elbows, moving the hands towards the shoulders.
9) **Stop Front, Go Right** – stop the oncoming vehicle, then direct traffic to the right of the oncoming vehicle to proceed forward; the participant moves their right hand directly in front of them, with the elbow extended (straight) and the hand open and keeps the right arm and hand in this position; the participant raises their left arm to the side, and quickly flexes (bends) their left elbow, bringing their hand towards their shoulder
10) **Stop Front, Go Left** – stop the oncoming vehicle, then direct traffic to the left of the oncoming vehicle to proceed forward; the participant moves their left hand directly in front of them, with the elbow extended (straight) and the hand open and keeps the left arm and hand in this position; the participant raises their right arm to the side, and quickly flexes (bends) their right elbow, bringing their hand towards their shoulder.
11) **Stop Front, Go Back** – stop the oncoming vehicle, then direct traffic opposite the oncoming vehicle to proceed forward; the participant moves their left hand directly in front of them, with the elbow extended (straight) and the hand open and keeps the left arm and hand in this position; the participant then turns their torso to the right, raises their right arm, and quickly flexes the right elbow, moving the hand towards the right shoulder.
12) **Stop Back, Go Front** – stop traffic opposite the oncoming vehicle, then direct the oncoming vehicle to proceed forward; the participant turns their torso to the right, raises their right arm, and extends the elbow with the hand open; the participant keeps the right arm in this position; the participant then moves their left hand in front of them with the elbow extended, and quickly flexes the left elbow, moving the left hand towards the left shoulder.

In the end, a total of 840 samples were collected for 12 different gestures, making it 70 samples per gesture from 14 different participants.

III. Proposed Method

A. Radar Signal Model

A Frequency modulated continuous wave (FMCW) radar system, shown in Fig. 4, consists of transmitters, receivers, mixers and analog-to-digital (ADC) converters. RF energy pulses are transmitted by a radar in various directions. When this energy strikes a target, it disperses in all directions. Only the fraction of the transmitted signal is scattered back and received by the radar receiver. The frequency of the transmitted signal in an FMCW radar changes linearly over time, typically sweeping through the whole bandwidth, allowing for the acquisition of both range and velocity measurements [9]. The chirp signal can be modeled as,

$$f_c(t) = f_{start} + \frac{B}{T_{SW}}t, \quad 0 \leq t \leq T_{SW} \tag{1}$$

where, $f_c(t)$ denotes the chirp frequency, f_{start} is the start frequency at time t = 0, B denotes the bandwidth and T_{SW} is the sweep time in seconds. The samples received by an FMCW radar is stored in a structure known as the radar data cube (RDC). The rows of RDC represent the fast-time samples, or the analog-to-digital conversion (ADC) values, and the columns represent the slow-time samples, or the chirp loops transmitted by the radar. The third dimension of RDC represents the different RX channels. Since angle estimation is not considered in our experiment, only the first channel is considered for further processing.

The intermediate frequency (IF) signal received from a target with time delay τ can be modeled as,

$$S_{IF}(t) = Aexp(2\pi(f_{start}\tau + \frac{B}{T_{SW}}t\tau - \frac{B}{2T_{SW}}\tau^2)) \tag{2}$$

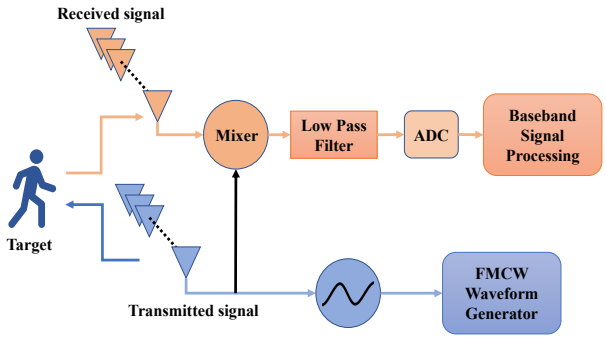


Fig. 4: Flow Diagram of an FMCW Radar [9]

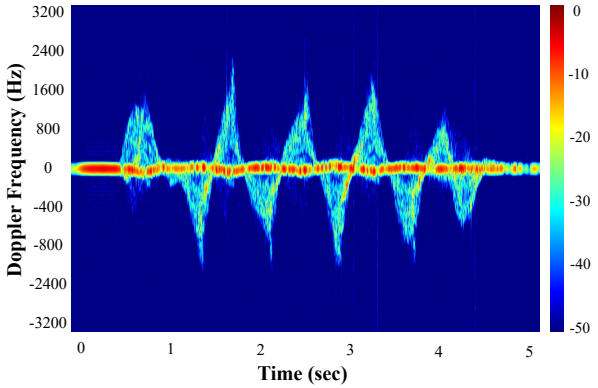


Fig. 5: A microDoppler Spectrogram example of the traffic sign 'Go'

where, the amplitude of the signal is A. After sampling, N fast time samples are produced for each pulse. The pulse number is referred to as the slow time samples. As the analog-to-digital converter (ADC) sampling interval is significantly less than the pulse repetition interval (PRI). The radar data cube (RDC) can be represented by N×P×M if the radar has M receiver channels. Different RF data representations, such as Range-Doppler (RD) videos or μ-D signatures, can be produced by Fourier processing over this RDC which will be discussed in the next section.

B. Micro-Doppler Spectrogram Generation

μ-D for a radar refers to the changes in received frequencies from a target involving small oscillating movements, such as the propellers of an aircraft spinning or a human walking with their arms and legs constantly moving. These small movements induce frequency modulations on the frequencies of a radar's signal, which is helpful in determining the kinematic properties of any given target [10]. Any radial motion that occurs toward the radar will result in a positive Doppler frequency, while radial motion away from the radar results in a negative Doppler frequency (assuming positive velocity occurs towards the radar). One way to visualize the μ-D information in a radar signal is with a spectrogram. Spectrograms are a widely used method for visualizing changes in both the time and frequency domains simultaneously. The μ-D spectrogram can be generated by using the Short-Time Fourier Transform (STFT). Equations (3) and (4) below show the discrete calculation for a signal in the time-frequency domain and computing the spectrogram from the STFT, respectively.

$$\mathrm{STFT}[x[n]]_{m,\omega} = X[m,\omega] = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \tag{3}$$

$$\mathrm{Spectrogram}[x[n]]_{m,\omega} = |X[m,\omega]|^2 \tag{4}$$

The STFT uses a FT with a windowing function rather than taking the FT of an entire signal at once. The spectrogram is the power output of the STFT. Figure 5 below shows a visualized spectrogram example of traffic sign 'Go' using an image of scaled colors. Positive Doppler frequencies are represented above the horizontal axis at 0 Hz, while negative frequencies sit below the horizontal axis.

C. The CNN Architecture Development

To classify the radar μ-D spectrogram a CNN structure has been developed. As shown is Fig. 6, the CNN structure has four convolution layers, each with kernel size 3×3, stride 1×1 and comprising of 32, 64, 64 and 128 filters respectively. A 3×3 maxpooling, batch normalization, and activation ReLU are performed after each convolutional layer. The tensor is then flattened, fed into a dense layer with a size of 512, dropped out by 0.2, and activated with ReLU. Then finally ending with the softmax classifier.

IV. Performance Analysis

To analyse the performance of the dataset, all of the spectrograms that were created were saved as 200x200 png files in both grayscale and 8 bit RGB images as we observed better performance than directly using spectrogram values (in dBm/volts). CNN input shapes are 200x200x3 and 200x200x1, respectively for RGB and grayscale images. For classification, the dataset was split into 80% training and 20% testing.

The results are shown in the Table II. From the table it is seen for both RGB and Gray-scale images the model is giving an accuracy score of more than 90%, where the RGB performed slightly better than the Gray-scale, 93.45% and 91.67% respectively. In terms of precision, recall and f1 score, the CNN (RGB) shows better performance (around 2%) than the CNN (Grayscale). The confusion matrix in Fig. 7 shows that the model performed very promisingly for most of the classes. Though the dataset were very challenging as the movements for each class varied from person to person as shown in Fig. 8, still the proposed CNN model were able to work on the dataset very effectively. This shows great promise as it can be great and necessary feature to add in the ADAS system.

V. Conclusion and Future Work

The goal of this study is to provide an initial study towards understanding the success of varying sensors including mm-

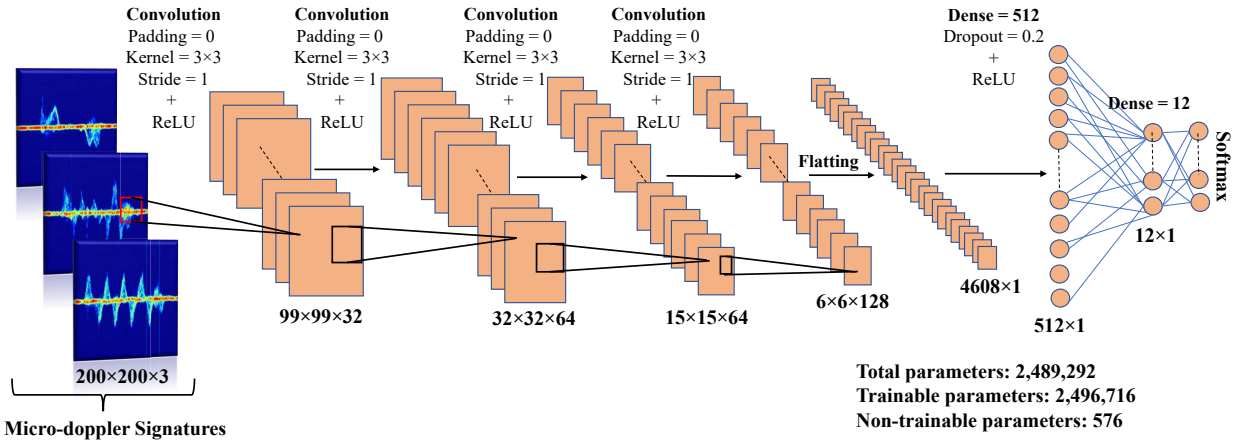


Fig. 6: The CNN architecture for traffic signalling motion classification.

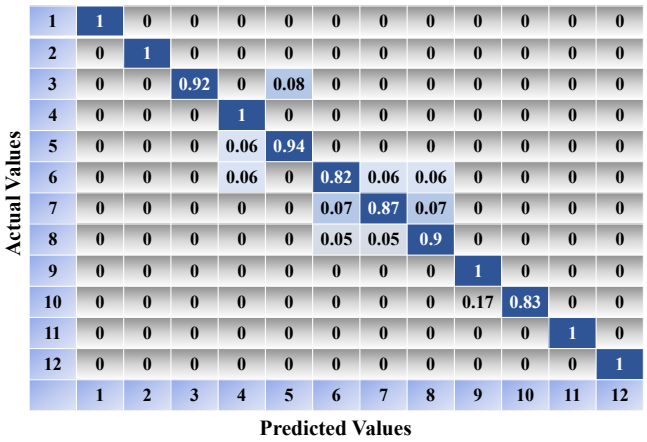


Fig. 7: Confusion matrix of CNN (rgb) for 12 class traffic gestures' data.

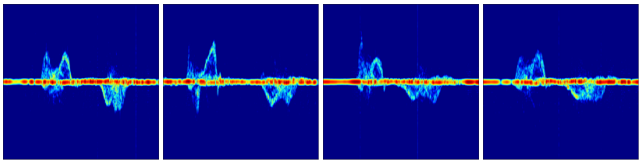


Fig. 8: Comparison of "Stop" Between Four Participants

TABLE II: Performance Comparison

Network	Testing Accuracy	Precision	Recall	F1 Score
CNN (Grayscale)	91.67	91.68	91.76	91.60
CNN (RGB)	93.45	93.75	94.06	93.70

wave automotive radar systems to recognize gestures from human traffic directors autonomously. The results of the radar data analysis show that, even with limited data samples, a learning process can accurately classify pre-defined traffic direction classes. However the initial results are from single traffic directions and in lab environment. Future work will include more realistic scenarios including fusion of radar and other automotive sensors.

VI. Acknowledgements

The data collection work was funded in part by the National Science Foundation (NSF) Awards #1931861, and #2047771. Human studies research was conducted under Mississippi State University Institutional Review Board (IRB) Protocol # IRB-21-256.

References

[1] S. Gurbuz, Ed., *Deep Neural Network Design for Radar Applications*. London: IET, 2020.

[2] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the netrad multistatic radar for micro-doppler and singular value decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933–1937, 2015.

[3] A. Huizing, M. Heiligers, B. Dekker, J. de Wit, L. Cifola, and R. Harmanny, "Deep learning for classification of mini-uavs using micro-doppler spectrograms in cognitive radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 46–56, 2019.

[4] G. Hakobyan and B. Yang, "High-performance automotive radar: A review of signal processing algorithms and modulation schemes," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 32–44, 2019.

[5] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[6] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. S. Crawford, M. M. Rahman, E. Kurtoglu, R. Aksu, T. Macks, and R. Mdrafi, "American sign language recognition using RF sensing," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3763–3775, 2021.

[7] F. Fioranelli and J. L. Kernec, "Contactless radar sensing for health monitoring," in *Engineering and Technology for Healthcare*, 2021, pp. 29–59.

[8] C. Li and S. Yang, "Traffic police gesture recognition for autonomous driving," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 1413–1418.

[9] J. J. Lin, Y. P. Li, W. C. Hsu, and T. S. Lee, "Design of an fmcw radar baseband signal processing system for automotive application," *SpringerPlus*, vol. 5, pp. 1–16, 12 2016. [Online]. Available: https://springerplus.springeropen.com/articles/10.1186/s40064-015-1583-5

[10] V. C. Chen, D. Tahmoush, and W. J. Miceli, *Radar micro-Doppler signatures*. Institution of Engineering and Technology, 2014.