



Road Users Classification Based on Bi-Frame Micro-Doppler With 24-GHz FMCW Radar

Rudi Coppola, Sajid Ahmed* and Mohamed-Slim Alouini

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

This study shows an approach for classifying road users using a 24-GHz millimeter-wave radar. The sensor transmits multiple linear frequency-modulated waves, which enable range estimation and Doppler-shift estimation of targets in the scene. We aimed to develop a solution for localization and classification, which yielded the same performance when the sensor was fixed on ground or mounted on a moving platform such as a car or quadcopter. In this proposed approach, classification was achieved using supervised learning and a set of hand crafted features independent of relative speed between the target and sensor. The proposed model is based on obtaining micro-Doppler information; only one receiver is used. Therefore, in addition to the target reflectivity, no geometrical information is used. For our study, we selected three classes: pedestrians, cyclists, and cars. We then illustrated distinctive micro-Doppler features for each class based on simulations, which we compared with real-world data. Our results confirm that a limited set of low-complexity features yields high accuracy scores when the target's trajectory does not excessively deviate from the radar's radial direction.

Keywords: FMCW, radar, micro-Doppler, machine learning, classification

OPEN ACCESS

Edited by:

Daniilo Orlando,
University Niccolò Cusano, Italy

Reviewed by:

Chengpeng Hao,
Institute of Acoustics (CAS), China
Mohammed Jahangir,
University of Birmingham,
United Kingdom

*Correspondence:

Sajid Ahmed
sajid.ahmed@kaust.edu.sa

Specialty section:

This article was submitted to
Radar Signal Processing,
a section of the journal
Frontiers in Signal Processing

Received: 28 January 2022

Accepted: 29 March 2022

Published: 20 May 2022

Citation:

Coppola R, Ahmed S and
Alouini M-S (2022) Road Users
Classification Based on Bi-Frame
Micro-Doppler With 24-GHz
FMCW Radar.
Front. Sig. Proc. 2:864538.
doi: 10.3389/frsip.2022.864538

1 INTRODUCTION

Advances in radar technology enabled the detection and classification of targets in the surroundings. This potential is valuable for multiple applications, e.g., for obstacle avoidance when operating autonomous vehicles, it is important to consider possible target maneuvers for robust path planning; therefore, a simple obstacle localization is insufficient. In general, target classification is achieved using optical sensors. The primary advantage of using radars on optical sensors is their performance invariance at night and under adverse weather conditions. Furthermore, when the wavelength of electromagnetic (EM) waves transmitted by radar is appropriately selected, it provides robustness against heavy rain, dust, and fog. For these scenarios, radar sensors can provide important support to image-based recognition systems. Radar-based classification is a relatively new field and currently is being explored using different approaches. In general, research in this field can be divided in two major branches: 1) unmodulated frequency continuous-wave (UFCW) and 2) frequency modulated continuous-wave (FMCW) radars.

With a UFCW radar, the micro-Doppler signature was used to recognize different human activities in (Kim and Ling, 2009; Zenaldin and Narayanan, 2016). The high accuracy scores achieved confirms the preciseness of the micro-Doppler features obtained. However, these approaches have specific limitations such as long observation time required to estimate frequencies and lack of range information. In UFCW radars, a long observation time is necessary to capture low-frequency components in the signal. In practice, it may require storing considerable amount of data, which is often too much for commercial micro-controllers.

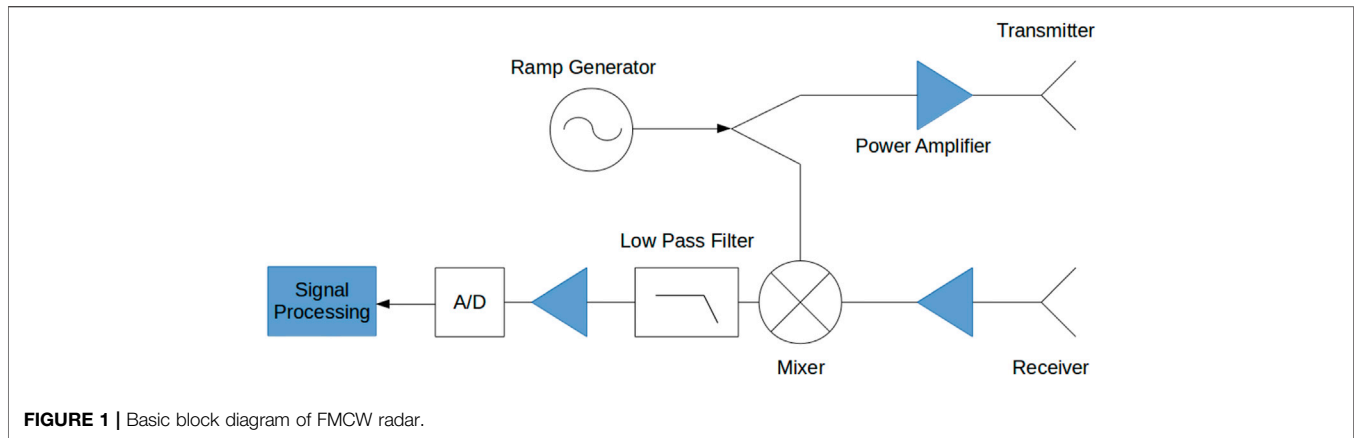


FIGURE 1 | Basic block diagram of FMCW radar.

However, the FMCW radar is of considerable interest to the automotive sector because it can provide the necessary range and help cope with the challenge of the long observation time required for low frequencies. A popular classification approach used with the FMCW radar is to pass multidimensional Fourier transformed data to a neural network architecture.

In (Perez et al., 2018), to classify three targets—namely cars, cyclists, and pedestrians—the authors used convolutional neural networks (CNNs), which extracted features from the range–Doppler–azimuth data cube. In the present study, to form a data cube, a 77-GHz 1×8 multiple-input and multiple-output (MIMO) FMCW radar system was used. The limitation of the algorithm is its high computational complexity for identifying the 3D fast-Fourier-transform (FFT). In addition to the expense of computing a 3D FFT, the proposed algorithm exploits the absolute velocity of the targets to discriminate between targets. Velocity estimation is easy for a radar on a fixed platform; however, it can be challenging for a radar on a moving platform. The CNN used in (Patel et al., 2019) to classify cars, construction barriers, motorbikes, baby carriages, bicycles, garbage containers, and stop signs is similar to that used in (Perez et al., 2018). In the present study, a 77 GHz 4×4 MIMO FMCW radar system is used to capture the target features from the range-velocity-azimuth data cube. This algorithm requires a 3D FFT; therefore, the algorithm's computational complexity is extremely high. To classify humans and vehicles in lower computational complexity, a 24 GHz 1×1 FMCW radar that depends on the range-Doppler data matrix is used in (Hyun and Jin, 2020). The algorithm proposed in this research extracts three features from the slow-time samples corresponding to a target: the complexity–scattering-point count, the scattering-point difference, and the magnitude difference rate. These features are input to a support vector machine classifier. In real life, we notice cyclists on the road along with pedestrians and cars. The performance of the algorithm significantly degrades when cyclists are added to the proposed classifier. This happens because of similarities between pedestrians and cyclists in terms of their return signal strength and micro-Doppler effects. To improve performance, additional features require to be added to the classifier.

However, we propose a low-complexity algorithm to classify the different targets in this research. The proposed algorithm uses a range–velocity data matrix to estimate the micro-Doppler

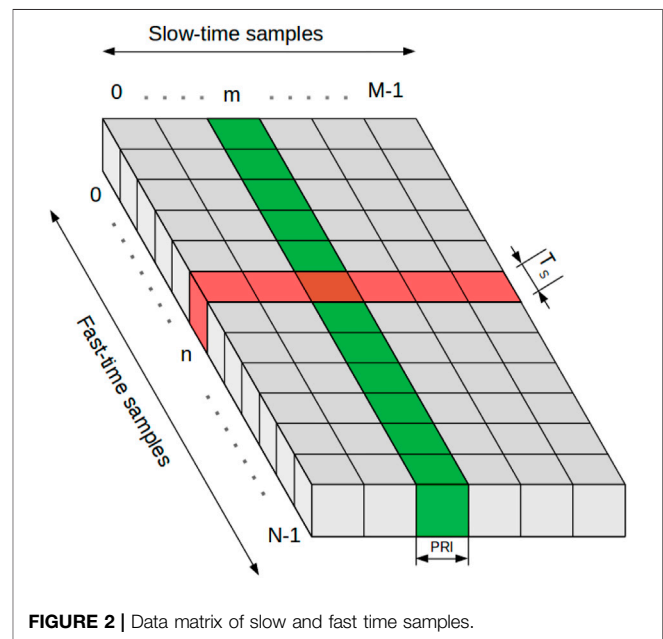


FIGURE 2 | Data matrix of slow and fast time samples.

spectrum of targets. Moreover, we design a set of simple features that extract the distinguishing characteristics of each class from the micro-Doppler spectrum of the target. We selected three targets: pedestrians, cyclists, and cars. To improve the classification of humans, cyclists, and vehicles, we adopted the extreme gradient-boosted-trees (XGBoost) classifier (Chen and Guestrin, 2016) and compared the obtained results with those of the ridge classification and K-nearest-neighbors (KNN) models. An in-depth overview of these methods can be found in (Trevor et al., 2009).

The proposed algorithm was implemented on Infineon's Position2Go module (Infineon, 2021), which is a 24 GHz 1×2 FMCW radar system. Although this module is equipped with two receivers, to ensure sufficiently fast data transmission from the device to a PC, the data of only one receiver were processed. Moreover, two receivers do not provide sufficient angular resolution to infer target geometrical information such as the

TABLE 1 | Firmware Configuration.

Firmware Configuration		
Parameter	Meaning	Value
f_s	Start Frequency	24.025 GHz
B	Bandwidth	200 MHz
T_c	Up-chirp duration	300 μ s
N_s	Samples per chirp	64
T_s	Fast-time sampling frequency	4.687 μ s
T_{PRI}	Pulse Repetition Interval	500 μ s
N_c	Chirps per frame	128
T_{FRI}	Frame Repetition Interval	200 ms

object width. The primary contributions of our proposed study are as follows:

- Unlike algorithms that compute 3D FFTs and pass the 3D feature matrix to train the classifier, our proposed algorithm computes two 1D FFTs and requires extremely few features to train the classifier. Thus, we provide a solution more suitable for hardware-constrained applications.
- Our proposed technique does not require the target's absolute velocity. Therefore, it can be used in both scenarios in which the radar is fixed and moving.
- For the designed classifier model, the trade-off between the accuracy and feature number was analyzed, which demonstrated that the gain for each newly added feature significantly slowed down after ten features were selected.
- We proposed a classifier model design for radial motion that achieves accuracies of 98.2, 97.3, and 98.1% for pedestrians, cyclists, and cars, respectively.

The remainder of this study is organised as follows: In the following section signal model is described. In **Section 3** the data sample collection and feature extraction are described. The feature extraction based on the scattering model is discussed in **Section 4** and experiment results are given in **Section 5**. Finally, we present our concluding remarks in **Section 6**.

TABLE 2 | Measurement specifications.

Measurement specifications		
Parameter	Meaning	Value
R_{max}	Maximum unambiguous range	48 m
R_{res}	Range resolution	0.75 m
V_{max}	Maximum unambiguous velocity	± 6.24 m/s
V_{res}	Velocity resolution	0.01 m/s

2 SIGNAL MODEL

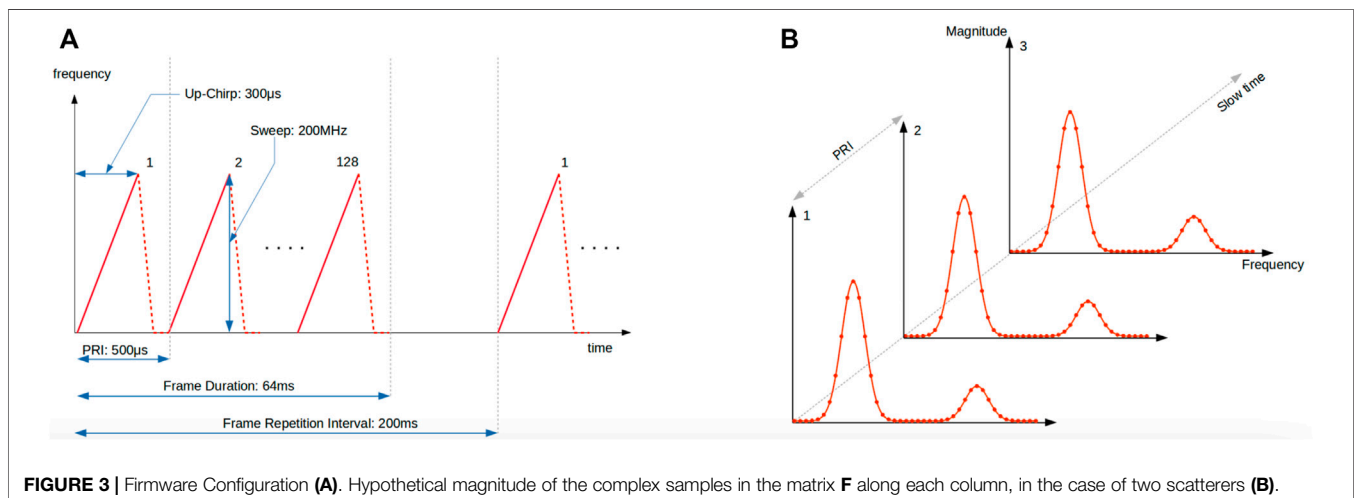
In (Richards et al., 2010; Melvin and Scheer, 2012; Melvin and Scheer, 2013; Mark and Richards, 2015), the researchers provided a rich overview of radar systems and radar signal processing. In a linear FMCW systems, the transmitted signal, called *chirp* signal, can be expressed as

$$s(t) = \exp\left(j2\pi\left(f_s t + \frac{S}{2}t^2\right)\right) \text{rect}\left(\frac{t - T/2}{T}\right), \quad (1)$$

where f_s is the starting frequency, T is the chirp period, and $S = \frac{B}{T}$ is the frequency sweep rate (or *chirp rate*), while B represents the bandwidth of the signal. Here, we do not treat the signal amplitude, we thus assume that amplitude is unity. If an object, present at distance R_o , intercepts the transmitted electromagnetic (EM) waves, a part of the signal's energy is reflected from the object. Assuming that the target is moving with the velocity v along the radial direction, its position at the time t can be described as follows: $R(t) = R_o + vt$, where R_o is the position of target at the start of chirp. Ignoring the signal's amplitude variation, the received signal $y(t)$ can be written as follows:

$$y(t) = s\left(t - \frac{2R(t)}{c}\right) = s(t - \tau(t)), \quad (2)$$

where $\tau(t) = \frac{2R(t)}{c}$ is the time delay, which is the time taken by a waveform to cover the distance from the radar to target and back.

**FIGURE 3** | Firmware Configuration (A). Hypothetical magnitude of the complex samples in the matrix **F** along each column, in the case of two scatterers (B).

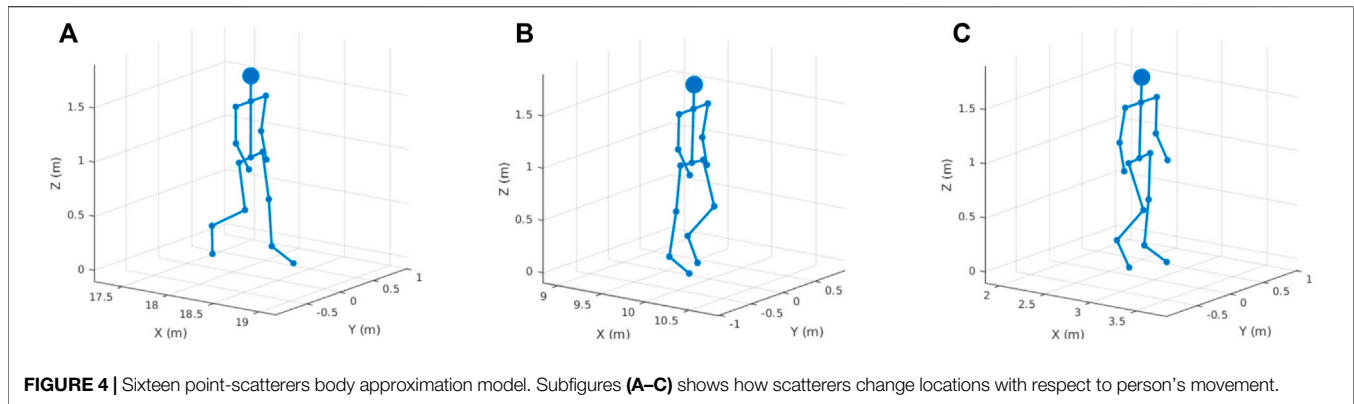


FIGURE 4 | Sixteen point-scatterers body approximation model. Subfigures (A–C) shows how scatterers change locations with respect to person's movement.

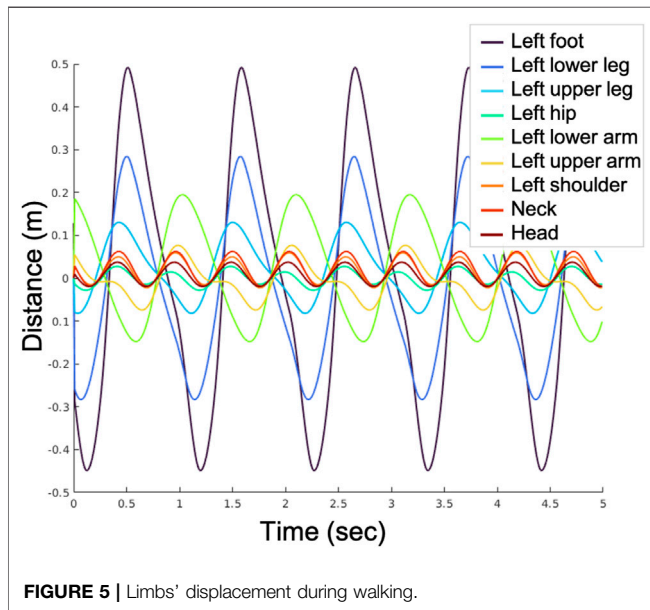


FIGURE 5 | Limbs' displacement during walking.

At the receiver, the signal is demodulated as shown in **Figure 1**; moreover, the output can be written as follows (Cooper, 1980; Mark and Richards, 2015)

$$\begin{aligned} r(t) &= (s(t) + \bar{w}(t))y^*(t) \\ &= e^{j2\pi(S\tau(t)(t-\tau(t))+f_s\tau(t)+\frac{1}{2}\tau(t)^2)} + w(t), \end{aligned} \quad (3)$$

where $\bar{w}(t)$ and $w(t)$ are the noises at the receiver before and after demodulation, respectively. Expanding **Eq. 3**, the demodulated signal can be simplified as follows

$$r(t) = e^{j\frac{4\pi R_0}{\lambda}} e^{j2\pi(S\tau_0 - f_d)t} + w(t), \quad \tau(\tau_0) \leq t \leq T \quad (4)$$

where $f_d = \frac{2v}{c}f_s$ is the Doppler frequency and $\tau_0 = \frac{2R_0}{c}$. Note that the signal in **Eq. 4** is a complex sinusoidal signal whose frequency is $S\tau_0 - f_d$ which is a combination of two frequencies and is known as beat frequency. Since both frequencies are coupled; therefore, it is difficult to separate them. However, to measure the Doppler shift with sufficient accuracy from a single chirp, an extremely long chirp period would be required, see (Mark and Richards,

2015). For this purpose, the Doppler shift can be measured by transmitting multiple equispaced chirps as follows:

$$\bar{s}(t) = \sum_{m=0}^{M-1} s(t - mT_{\text{PRI}}), \quad (5)$$

which is the M times repetition of the original chirp signal given by in **Eq. 1**. Each signal is then delayed by a multiple of T_{PRI} , which is the *pulse repetition interval*. The transmitted signal of M chirps is known as a *frame*. Within a frame, the target range at the start of m th pulse can be written as $R(mT_{\text{PRI}}) = R_0 + mvT_{\text{PRI}}$. Using the range at the start of each chirp in **Eq. 4**, the received signal after m th transmitted pulse can be written as

$$\begin{aligned} r_m(t) &= e^{-j\frac{4\pi(R_0 + mvT_{\text{PRI}})}{\lambda}} e^{j2\pi(S\tau_0 - f_d)t} + w_m(t), \\ &= e^{-j\frac{4\pi R_0}{\lambda}} e^{-j2\pi f_d m T_{\text{PRI}}} e^{j2\pi(S\tau_0 - f_d)t} + w_m(t). \end{aligned} \quad (6)$$

The continuous time domain signal in **Eq. 6** is passed to the analogue-to-digital converter and N samples are collected for each interval $I_m = [mT_{\text{PRI}}, mT_{\text{PRI}} + T]$. Therefore, in each frame MN complex samples are collected. In discrete form, these samples can be written as follows by considering the *stop-and-hop* approximation (Mark and Richards, 2015):

$$r_m(nT_s) = e^{-j\frac{4\pi R_0}{\lambda}} e^{-j2\pi f_d m T_{\text{PRI}}} e^{j2\pi(S\tau_0 - f_d)nT_s} + w_m(nT_s), \quad (7)$$

where T_s is the sampling period, $m = 0, \dots, M - 1$ and $n = 0, \dots, N - 1$.

As shown in **Figure 2**, the samples can be arranged in a 2D matrix. N samples in each column are called *fast-time samples*, and they correspond to a single transmitted chirp. The samples in rows are called *slow-time samples*, and they correspond to multiple chirps.

Figures 2, 7 shows that the value of m remains constant in each column, and the value of n remains constant in each row. Therefore, by applying FFT on fast-time samples, the beat frequency can be easily estimated. The term $e^{j2\pi f_d m T_{\text{PRI}}}$ remains unaffected during the FFT operation on the fast-time samples; therefore, the Doppler shift can be estimated by applying FFT on the slow-time samples in any row. Therefore, to estimate the range and the corresponding velocity of a target, we first apply FFT on fast-time samples. The spectrum peaks show the beat frequencies that can be used to identify the target range. By

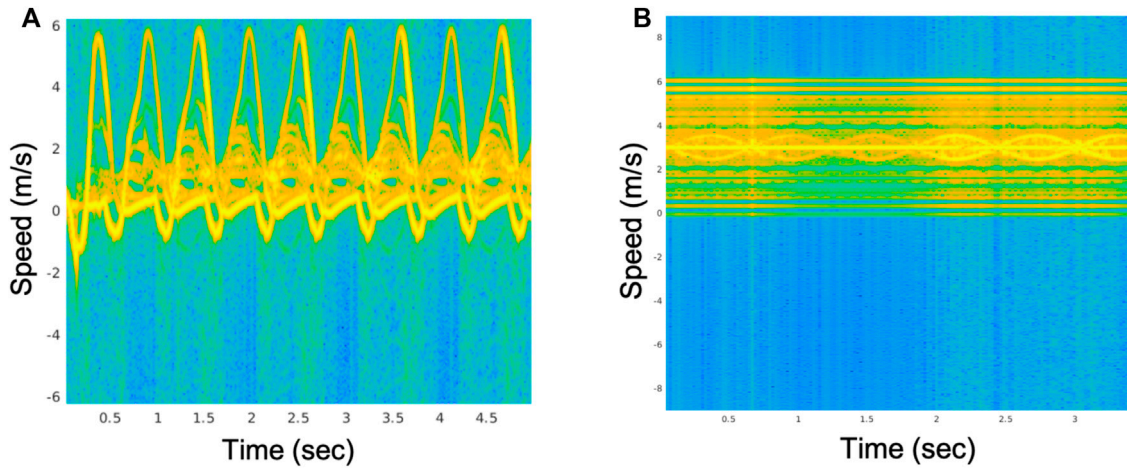


FIGURE 6 | Time-frequency analysis for a walking human (A), and a cyclist (B).

TABLE 3 | Notation of different features used throughout the paper.

All features

Acronym	Meaning
R_t	Radial distance in frame t
A	Range spectrum amplitude
SD	Standard deviation
CMD	Close micro-Doppler
BMD	Broad micro-Doppler
MTMD	Max to mean distance
HPC	High point count
MPC	Medium point count
LPC	Low point count
HP_{skew}	High point skewness
MP_{sd}	Medium point standard deviation
LP_{skew}	Low point skewness
HPKC	High peak count
MPKC	Medium peak count
LPKC	Low peak count
HPK	High peak total strength
MPK	Medium peak total strength
LPK	Low peak total strength
HPK_{skew}	High peak skewness
HPK_{sd}	High peak standard deviation
MPK_{sd}	Medium peak standard deviation
LPK_{skew}	Low peak skewness
HPCD	HPC difference
$HPKD_{sd}$	HPK_{sd} difference
$HPKD_{skew}$	HPK_{skew} difference
$MPKD_{skew}$	MPK_{skew} difference
$LPKD_{skew}$	LPK_{skew} difference
SDD	SD difference
CO	Correlation
MDR	Magnitude difference rate

applying FFT on a row corresponding to the index of a spectrum peak, the target velocity can be estimated. In most practical scenarios, where the frame duration and the velocity are low, the beat frequency can be written as follows:

$$f_b \approx S\tau_0, \quad (8)$$

which makes it easy to estimate the target range using the beat frequency. Finally, to identify the maximum system limit, we use the Rayleigh bandwidth. The Rayleigh bandwidth is $1/T$ for fast-time samples and $1/MT_{PRI}$ for slow-time samples; therefore, the range resolution ΔR , the velocity resolution Δv , the maximum unambiguous range R_{max} , and the maximum velocity v_{max} can be easily calculated as follows:

$$\Delta R = \frac{c}{2B},$$

$$\Delta v = \frac{\lambda}{2MT_{PRI}}, \quad (9)$$

$$R_{max} = \frac{Nc}{2B},$$

$$v_{max} = \pm \frac{\lambda}{4T_{PRI}}. \quad (10)$$

In the following section, the above mentioned derivations will be used to collect samples for extracting the features and estimating ranges and velocities corresponding to different targets.

3 DATA COLLECTION AND FEATURE EXTRACTION

To collect samples for the classifier, we used a commercially available radar module produced by Infineon, called Position2Go (Infineon, 2021). This module is equipped with a 24 GHz transceiver chip (BGT24MTR12) and a 32-bit microcontroller (XMC4700) for signal processing. Three micro-strip series patch antennas (one for transmission and two for reception) are printed on the board. The field-of-view of each antenna is $19^\circ \times 76^\circ$. The basic architecture of the module is similar to the one shown in Figure 1, the only difference is that Position2Go has one transmit and two receive channels. The module is described in detail in (Will et al., 2019). The module's firmware configuration settings and the specifications for this study are defined in Tables 1, 2.

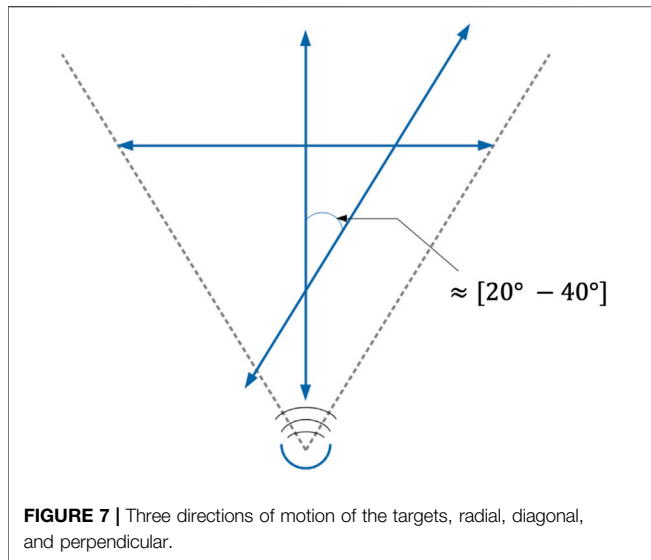


Figure 3A shows the structure of chirp transmission. Clearly, one frame contains 128 chirps each of duration $300\mu\text{sec}$, and chirp repeats itself after pulse-repetition-interval (PRI) of $500\mu\text{sec}$. At the receiver end, after the transmission of each chirp, 64 samples were collected. The transmit and receive antennas were not isolated; therefore, the transmit signal was leaked in the receiver that appeared as a low-frequency component in the demodulated signal. This component was independent of the clutter and could be completely removed by storing it in memory and then subtracting it from each column of 2D matrix when the actual experiments were performed. To obtain this signal, it was recommended to point the radar towards an absorbing material or the sky such that only the leakage power signal was received. If the radar was fixed on the ground, it would have been possible to remove the return signal of the background (or clutter). The signals caused by the leakage and background can be removed in a single step by considering a preliminary recording of the scene without the targets

$$\mathbf{r}_{\text{preliminary}} = \mathbf{r}_{\text{leakage}} + \mathbf{r}_{\text{background}}. \quad (11)$$

Subtracting this signal from the received signal in the presence of actual targets in the scene can in fact significantly improve the detection performance

$$\begin{aligned} \mathbf{r}_{\text{recordings}} &\approx \mathbf{r}_{\text{leakage}} + \mathbf{r}_{\text{background}} + \mathbf{r}_{\text{targets}} \\ \mathbf{r}_{\text{targets}} &\approx \mathbf{r}_{\text{recordings}} - \mathbf{r}_{\text{preliminary}}. \end{aligned} \quad (12)$$

The above mentioned equations are approximately valid when the background return is essentially unaffected by the presence of the targets, and the sensor is fixed on the ground.

Therefore, as shown in **Figure 2**, the collected 64 complex samples of raw data corresponding to each of 128 chirps can be used to form a 64×128 measurement matrix \mathbf{M} . After leakage and background removal, a Hamming window is applied to each column for range side-lobe reduction. To limit the straddle loss, each column of the 2D data matrix is zero-padded to $N_{\text{pad}}^R = 256$

TABLE 4 | Number of frames captured for machine learning algorithms when different targets move in radial, diagonal, and perpendicular directions.

Number of Frames			
	Radial	Diagonal	Perpendicular
Pedestrian	669	572	557
Cyclist	396	524	652
Car	310	568	875

samples. Then, each column is transformed in the frequency domain by applying FFT as follows:

$$\mathbf{F}[k, m] = \frac{1}{\sqrt{N_{\text{pad}}^R}} \sum_{n=0}^{N_{\text{pad}}^R-1} \mathbf{M}[n, m] e^{-j2\pi \frac{kn}{N_{\text{pad}}^R}}, \quad (13)$$

where $k = 0, 1, \dots, N_{\text{pad}}^R - 1$, $m = 0, 1, \dots, N_c - 1$, and $\mathbf{F}[k, m]$ shows the k th and m th element of the matrix $\mathbf{F} \in \mathbb{R}^{N_{\text{pad}}^R \times N_c}$. Depending on the number of targets, the amplitudes of each column of \mathbf{F} will look similar to amplitudes shown in **Figure 3B**. The peak values in the spectrum represent the location of moving or static targets.

To differentiate the moving targets from stationary targets, a technique known as moving target indication (MTI) filtering is used (see (Mark and Richards, 2015)). MTI involves high-pass filtering and can be implemented using the algorithm given in **Supplementary Appendix 0.1**. Stationary targets do not move from one chirp to another, and they show zero Doppler effects; they cannot be completely removed using MTI filtering alone. Therefore, to remove these targets, before MTI filtering, the mean of each row of \mathbf{F} is computed and subtracted from the respective row. We indicate with \mathbf{X} the $N_{\text{pad}}^R \times 1$ vector resulting from the overall filtering procedure. To develop an algorithm, a threshold is set and five top peaks of a column of the matrix \mathbf{X} are selected. For each peak, the corresponding indices are stored in a vector

$$\mathbf{p} = [p^1, p^2, \dots, p^5], \quad (14)$$

where $0 \leq p^i \leq N_{\text{pad}}^R - 1$. Suppose the i -th target is being located. The first condition for an index k to correspond to a spectrum peak requires the amplitude, $|\mathbf{X}[k]|$, to be greater than a fixed threshold ρ equal to -46 dB (0.005). This choice was dictated “empirically” by looking at the noise level in the recordings to minimize false-positive detections. To increase the robustness, a second condition must also be satisfied for k to be confirmed a peak index. The second condition requires that $|\mathbf{X}[k-2]| < |\mathbf{X}[k-1]| < |\mathbf{X}[k]| > |\mathbf{X}[k+1]| > |\mathbf{X}[k+2]|$. If both conditions hold then $p^i = k$.

The dimension of vector \mathbf{p} simply corresponds to the maximum number of targets that someone is interested in tracking on the same scene.

Each peak index shows the position of a moving target. The phase history Φ of each target is extracted from the corresponding row of the matrix \mathbf{F} as follows:

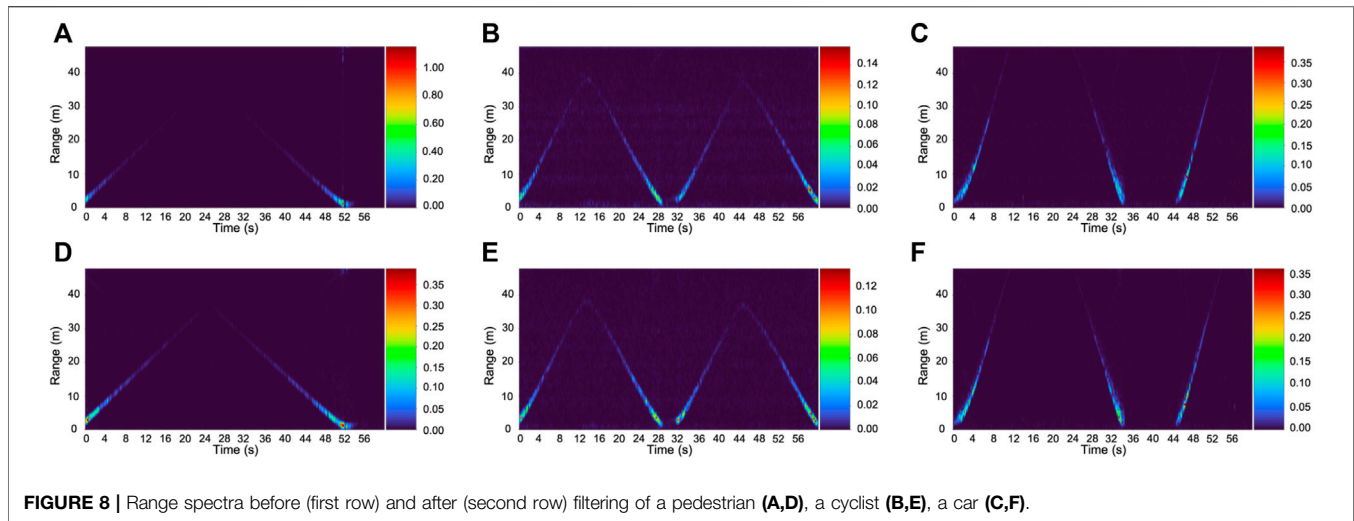


FIGURE 8 | Range spectra before (first row) and after (second row) filtering of a pedestrian (A,D), a cyclist (B,E), a car (C,F).

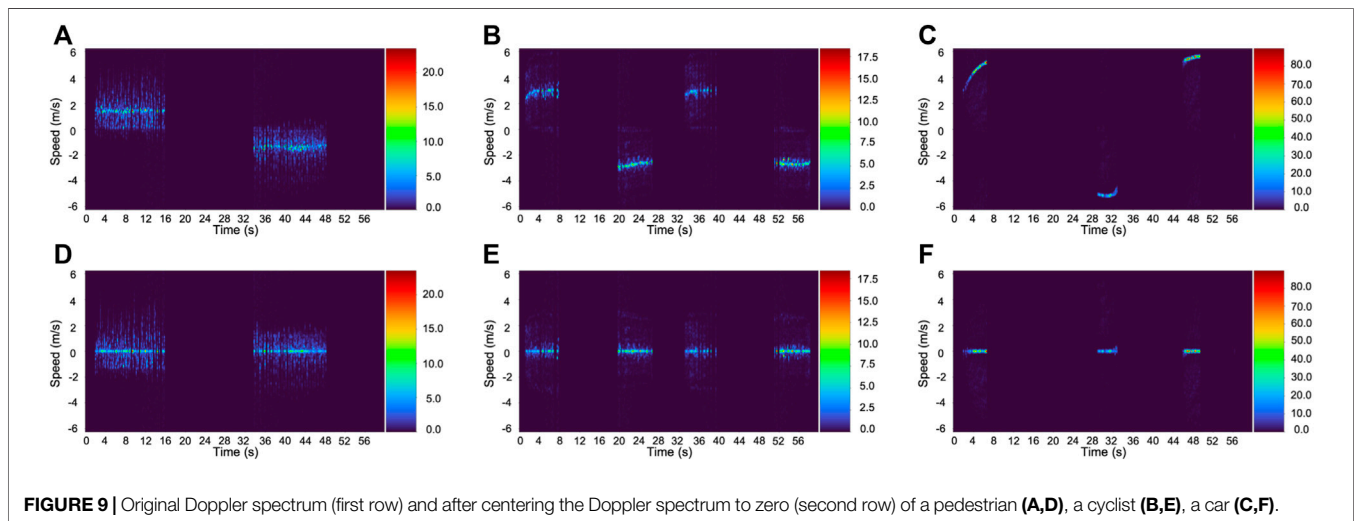


FIGURE 9 | Original Doppler spectrum (first row) and after centering the Doppler spectrum to zero (second row) of a pedestrian (A,D), a cyclist (B,E), a car (C,F).

$$\Phi^i = [\mathbf{F}[p^i, 0], \mathbf{F}[p^i, 1], \dots, \mathbf{F}[p^i, N_c - 1]]. \quad (15)$$

Each Φ^i is multiplied by a Hamming window, \mathbf{w}_H , to reduce the side-lobes and is zero-padded to N_{pad}^d samples to reduce the straddle loss effects. Finally, FFT is applied on these samples to identify the Doppler shift. Mathematically, the zero padding and Fourier transform steps can be written as follows:

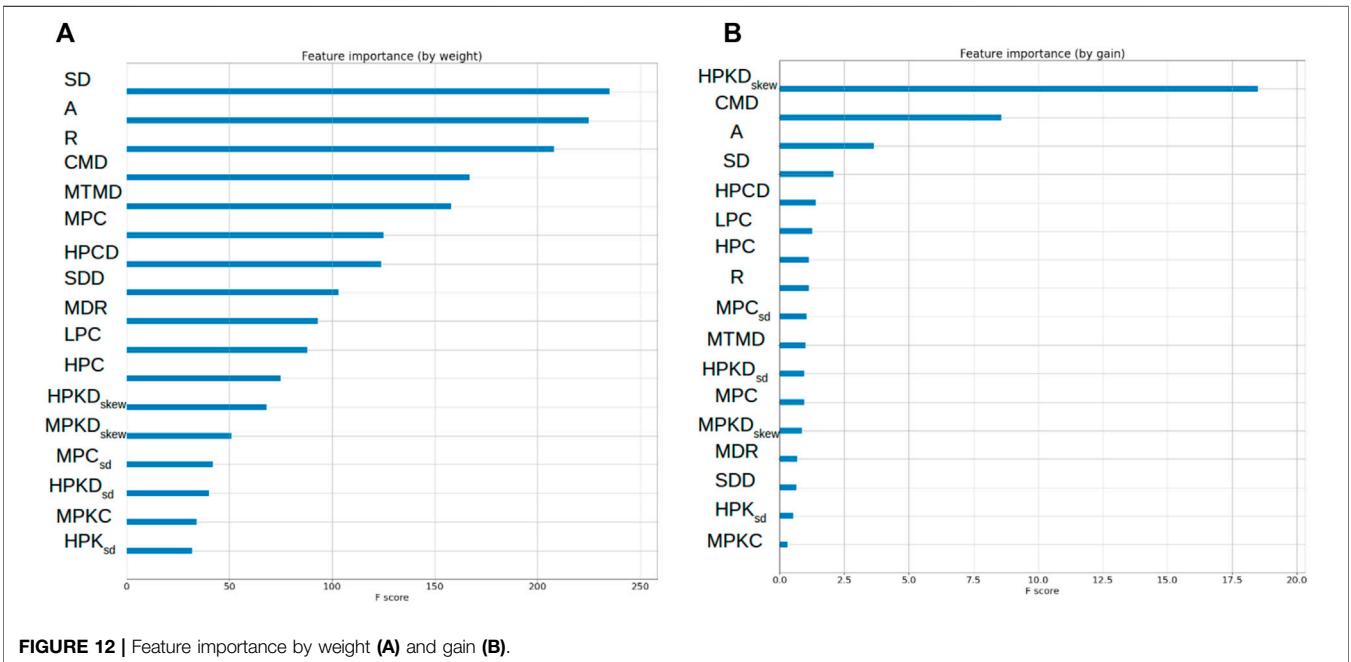
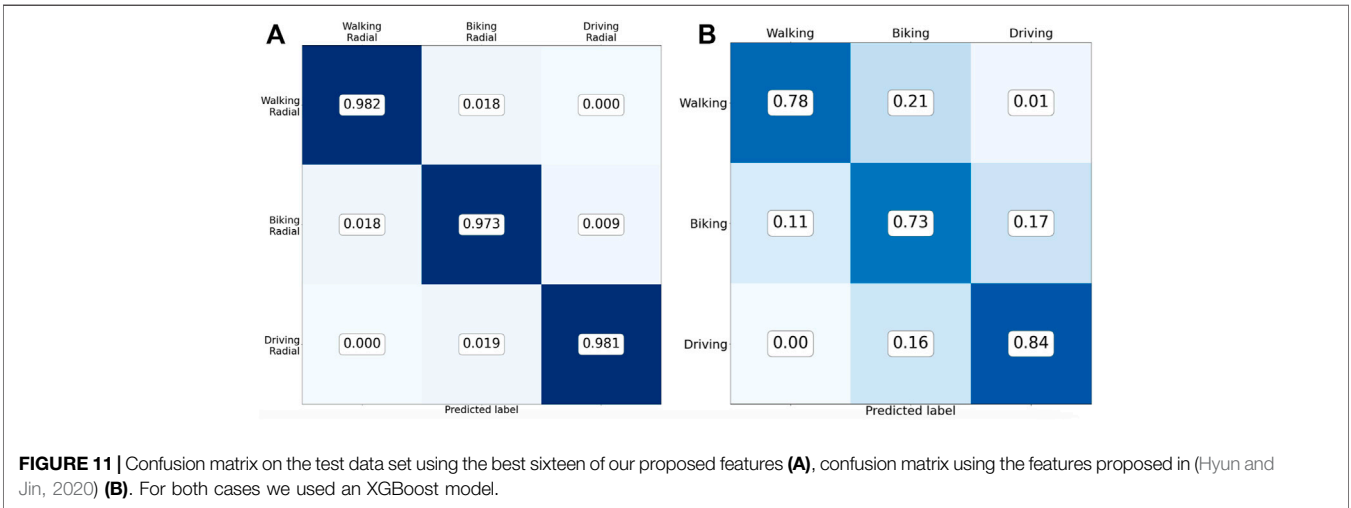
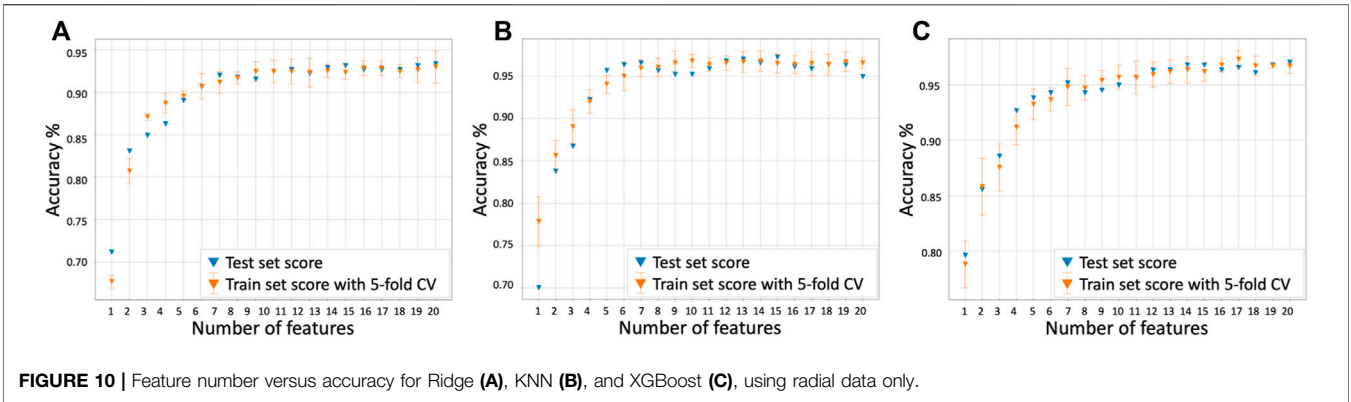
$$\begin{aligned} \bar{\Phi}^i &= [\Phi^i \circ \mathbf{w}_H, 0, 0, \dots, 0] \\ \mathbf{d}^i[k] &= \frac{1}{\sqrt{N_{pad}^d}} \sum_{n=0}^{N_{pad}^d-1} \bar{\Phi}^i[n] e^{-j2\pi \frac{kn}{N_{pad}^d}}, \end{aligned} \quad (16)$$

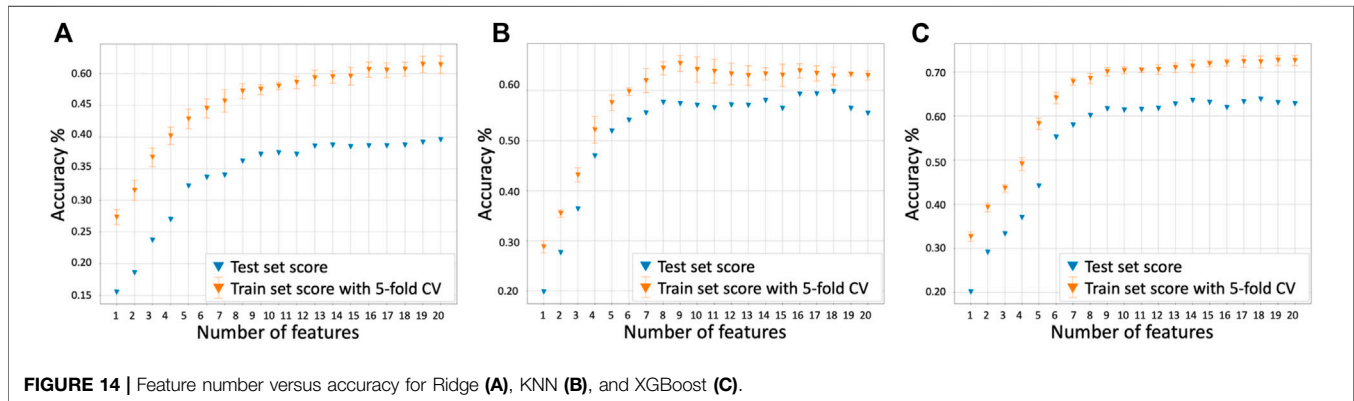
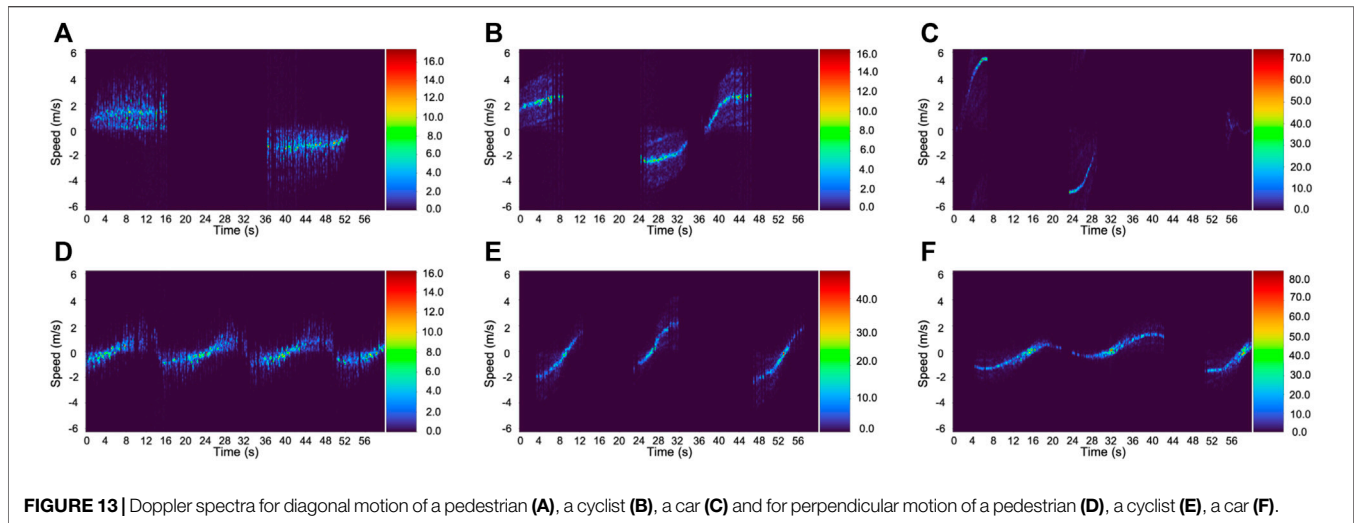
where $k = 0, 1, \dots, N_{pad}^d - 1$, and $\Phi^i \circ \mathbf{w}_H$ indicates the Hadamard product between the two vectors Φ^i and \mathbf{w}_H . The magnitude of the complex vector \mathbf{d}^i is the Doppler spectrum of the target under examination. For each target an object is created in memory, which contains the Doppler spectrum \mathbf{d}^i , the radial

distance R^i , and the average amplitude of the return signal A^i are given as follows:

$$\begin{aligned} R^i &= p^i \frac{cN_s}{2BN_{pad}^R}, \\ A^i &= \frac{1}{N_c} \sum_{m=0}^{N_c-1} |\mathbf{F}[p^i, m]|. \end{aligned} \quad (17)$$

Remark 2: This approach does not require the computation of the complete 2D FFT. Let us assume $(N)_{FFT}$ denotes the total number of operations required to compute the FFT of N elements sequence. We denote by N_c , N_R^{pad} , N_d^{pad} , and q the number of chirps per frame, the number of samples per chirp after zero-padding, the number of chirps per frame after zero-padding, and the number of targets in the scene, respectively. The 2D FFT requires $N_c(N_R^{pad})_{FFT} + N_R^{pad}(N_d^{pad})_{FFT}$ operations. In our proposed approach, we have to compute only $N_c(N_R^{pad})_{FFT} + q(N_d^{pad})_{FFT}$ operations. In our experiments, we had a single target present in the scene.





In the following section, scattering model of human and cyclist will be discussed and samples in \mathbf{d}^i s will be used to extract features.

4 SCATTERING MODEL BASED FEATURE EXTRACTION

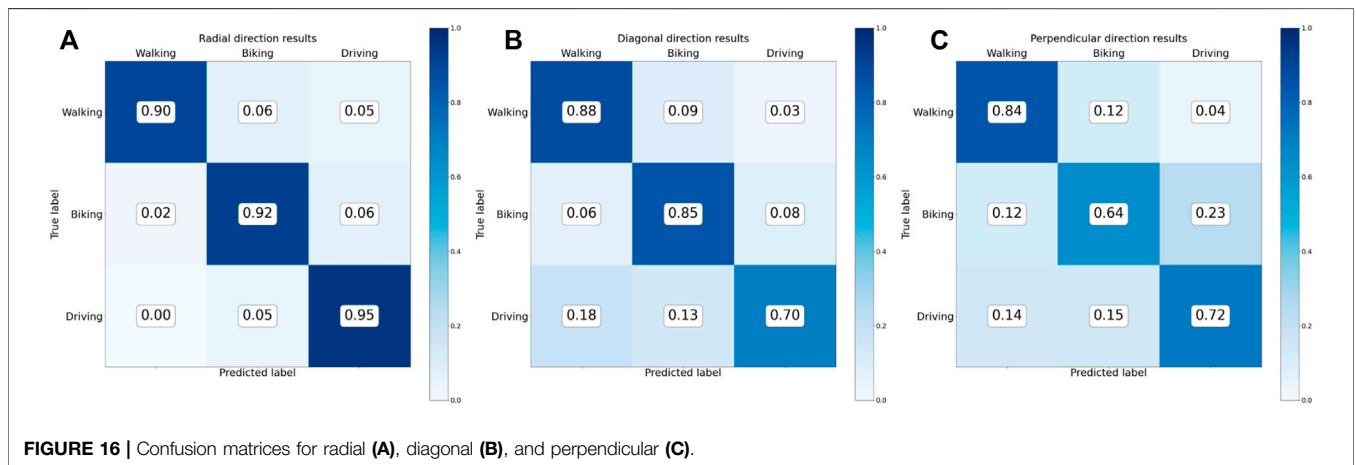
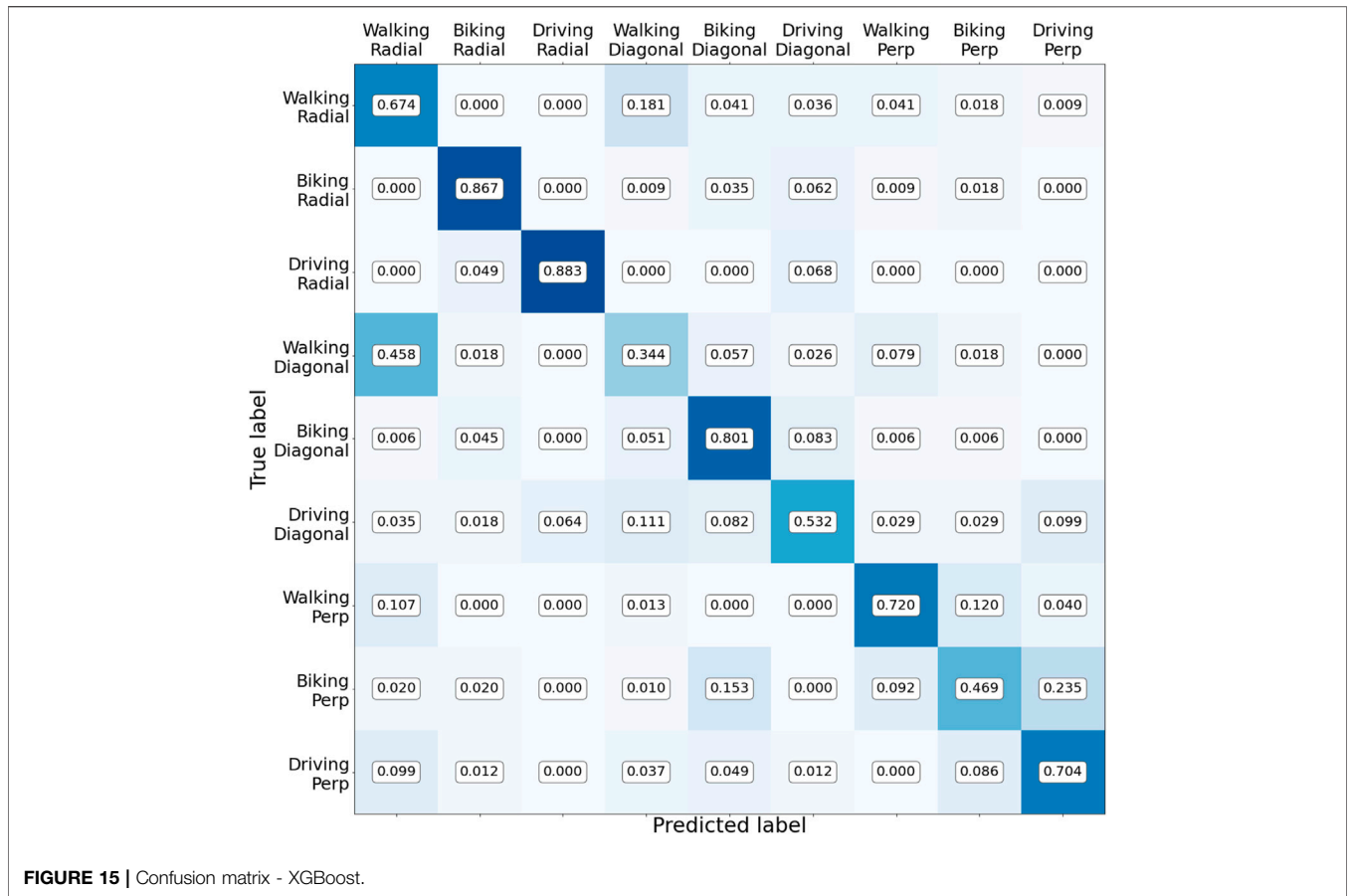
The Boulic-Thalmann is one of the example of human scattering model, which is based on bio-mechanical experimental data (Boulic et al., 1990; Melvin and Scheer, 2013). As per this model, the human echo can be approximated by the superposition of N distinct point-scatterers, each with its own dynamics.

In (Tahmoush et al., 2010), a 16 points scattering model of a 1.8-m tall person walking at 1.5 m/s, as shown in **Figure 4**, is discussed. **Figure 5** and **Figure 6A** show the movement of the corresponding simulated limbs and time-frequency spectrum, respectively. Similarly, a kinematic model of a cyclists has been developed in (Stolz et al., 2017). **Figure 6B** shows the simulated time-frequency spectrum of a cyclist traveling at 3 m/s. Here, we can see that multiple frequencies are present, in addition to the time-evolving frequencies that can be attributed to the motion of the legs and

pedals. These frequencies spread from 0 m/s to twice the cyclist's speed and are attributed to the signals returning from the wheels. The torsos and heads of the cyclists and pedestrians are responsible for the dominant return. Because of the relative motion of the various body parts with respect to the torso, multiple Doppler frequencies appear in the spectrum; this effect is known as the micro-Doppler effect (see (Chen et al., 2003)).

The spectrum of the signal's return reveals the periodicity T of limbs' kinematics, which is usually in the order of seconds. Unmodulated CW radars are best suited for time-frequency analysis; however, they require long windows to capture low frequencies. Nevertheless, FMCW radars typically transmit multiple short waveforms in a frame over a long duration. Because of the short time window, the Doppler frequencies captured during the transmission of one frame are almost constant. With sufficient velocity resolution, sufficient Doppler information could be captured for the classification.

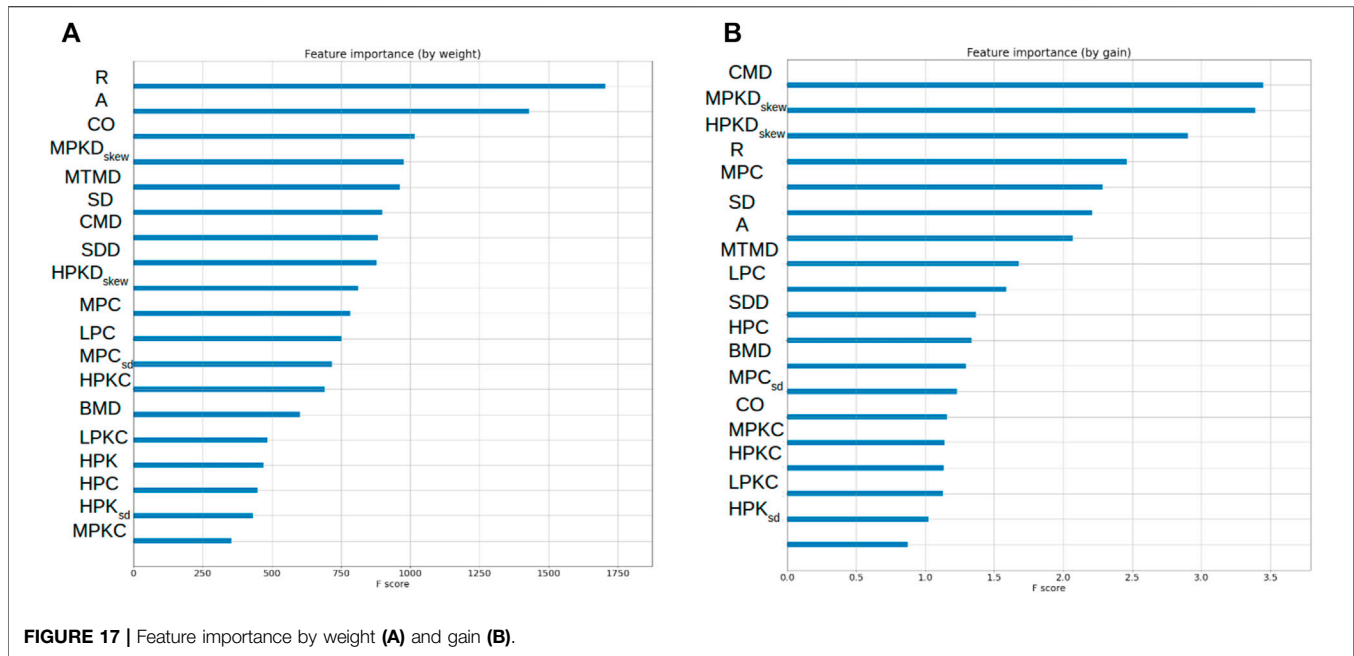
Considering these factors, we designed two subsets with low-complexity features. The first subset captures the number, strength, and distribution of peaks in the current Doppler spectrum. The second subset captures feature variations in the current and previous



Doppler spectrums. None of the features are based on the ego speed of the target, i.e., the target’s actual speed. To extract the features, we first determine the 512-point FFT of the corresponding row for each selected index p^i of the matrix F , $i = 1, 2, \dots, 5$. This spectrum specifies the Doppler frequencies associated with the target present in the p^i index location, and it is denoted by D . compactness, N_{pad}^d is denoted by N . The notation used to define different features are defined in **Table 3**. The first subset of features comprises SD, CMD,

BMD, and MTMD. The pseudo-code to identify these features is described in **Supplementary Appendix 0.2**.

The first subset includes also the following simple threshold-based features HPC, MPC, LPC, HP_{skew} , MP_{sd} , and LP_{skew} . The pseudo-code to identify these features is then described in **Supplementary Appendix 0.3**, where an approach similar to the one in (Hyun and Jin, 2020) is used. Finally, the first subset contains the following “twin” advanced threshold-based features HPKC,



MPKC, LPKC, HPK_{skew}, HPK_{sd}, MPK_{sd}, LPK_{skew}, HPK, MPK, and LPK. These features consider the number of peaks contained in the intervals defined by the thresholds, rather than merely considering the number of spectrum points. The pseudo-code used to identify these features is described in **Supplementary Appendix 0.4**. The features determined so far are based only on a single frame. To improve the prediction accuracy, the second subset of features measures the numerical variations of certain features over two subsequent frames. For the differential features defined as HPCD, HPKD_{sd}, HPKD_{skew}, MPKD_{skew}, LPKD_{skew}, SDD, CO, and MDR, the pseudo-code is described in **Supplementary Appendix 0.5**. The reason to select these features can be understood by looking at **Figure 6**. We observe that, at any instant, there are numerous strong peaks in the pedestrian's Doppler spectrum; this is captured by features like HPC, HPKC, and HPK. We can also observe that the distribution of the peaks around the pedestrian's body speed tends to be very skewed in one direction, unlike what we can observe in the cyclist's Doppler spectrum: this is captured by features like HP_{skew} and HPK_{skew}. Moreover, this skewness varies a lot from one frame to the next one: this is captured by features like HPKD_{skew}. The distribution and deviation of the peaks from the strongest frequency present in the Doppler spectrum are also captured by features like SD, CMD, BMD, and MTMD. The contribution of the wheel's motion is much lower in amplitude than the contributions of the bike's frame and cyclist's body; obviously, this characteristic can't be observed in the Doppler spectrum of a pedestrian: this is captured by features like LPC, LPKC, and LPK.

To select these features, let us assume that a set of N_o labeled observations each of feature length l are available for a

classification problem. We denote this set by \mathbf{X} . Feature selection is a procedure that identifies a column subset with cardinality $g < l$. Mathematically, if $\bar{\mathbf{X}}$ is a selected feature set, we can write the following

$$\begin{aligned} \bar{\mathbf{X}} &\subset \mathbf{X}, \\ \mathbf{X} &\in \mathbb{R}^{N_o \times l}, \\ \bar{\mathbf{X}} &\in \mathbb{R}^{N_o \times g}, g < l. \end{aligned} \quad (18)$$

Feature selection is useful to reduce the dimension of the data set; however, most importantly, it helps identify the relevant features. Viable feature selection algorithms depend on the model selected for the classification. We used *sequential feature selection* for Ridge Classification and KNNs having the radial basis function. Similarly, for classification and regression tree (CART) ensemble trees, *model-based feature selection* is used. Sequential-feature-selection (SFS) is a suboptimal recursive method to sequentially add (or subtract) relevant (or irrelevant) features (see (Ferri et al., 1994)). Forward SFS is a greedy method that maximizes a criterion function by recursively adding locally optimal features. However, MBFS is usually faster when it is possible to directly measure the importance of a feature from the model as per a given criterion.

5 EXPERIMENTAL RESULTS

Data were gathered in an area where there was ample space such that the clutter effect was minimized. The sensor was positioned on a tripod at the height of 1.3 m. We considered two pedestrians, two cyclists, and two car models. Each target was recorded while moving in three directions: radial, diagonal, and perpendicular (or azimuthal) as shown in **Figure 7**.

Data were collected during the day and at night. As shown in **Table 4**, we recorded 5,123 frames; moreover, the detection was limited between 5 and 25 m.

5.1 Radial Data

First, we provide a general overview of radar data before and after processing. **Figure 8A** shows the unfiltered range profile for a pedestrian walking in the radial direction with respect to the radar. The Fourier transform spectrum of fast-time samples for the first chirp from each frame are aligned and distributed as per the corresponding instant of transmission. From **Figure 8A**, when the pedestrian is clearly close to the radar at the time zero, the amplitude of the reflected signal is high; with the passage of time, the amplitude of the return decreases because of the pedestrian's motion away from the radar. When the pedestrian covers a distance of 40 m, the reflected signal from the pedestrian becomes negligible. After 25 s, the pedestrian takes a "U" turn and returns in the direction of the radar; therefore, the reflected signal from the pedestrian starts to increase. After 52 s, the pedestrian returns to the radar with the maximum reflected signal. **Figures 8A–C** show that at high ranges, the target peak is obscured in the nearby slowly-moving and static targets. To reduce the contribution of static and slowly moving targets, we use MTI and DC filtering; the output is then shown in **Figures 8D–F**. We see how the strength of the moving target increased compared to the static background. The gain is particularly visible for large distances. **Figures 8A,D** show the similar plots for pedestrians moving in the radial direction. **Figures 8B,E** show similar plots for cyclists moving in the radial direction, and **Figures 8C,F** show similar plots for the car moving in the radial direction.

After detecting the target, the Doppler spectrum is computed as previously described. To obtain a more informative plot, distance attenuation is compensated for by multiplying the Doppler spectrum with the square of the distance of the target. **Figures 9A–C**, show a plot of the Doppler spectrum versus the transmission time.

In **Figures 9D–F**, the Doppler spectrum's strongest component associated with the pedestrian's torso, the cyclist's torso, and the car frame was centered at 0 m/s. Despite the short frame duration supported by our sensor, the most important trait of spectrograms shown in **Figures 6A,B** were captured. The Doppler spectrum of a pedestrian typically contains multiple relevant frequencies because limbs are in motion while walking. Furthermore, the asymmetry (or *skewness*) in the distribution of these frequencies around the torso was captured. For a cyclist, the strongest Doppler components are located extremely close to the torso's Doppler. Moreover, the weak frequencies are extensively distributed because of the wheel's return. Finally, the Doppler spectrum of a car shows the return of the car's frame. Moreover, extremely weak Doppler components can be observed because of wheels. Unlike the case for a bicycle, a portion of a car's wheels is typically obscured by the bumper. Consequently, only the lowest part of the wheel is exposed, and the corresponding Doppler frequencies are asymmetrically distributed around the Doppler of the car's frame.

5.2 Radial Models

For Ridge Classification and KNNs, we depended on the Scikit-learn implementations (Pedregosa et al., 2015), whereas for the CART Tree Ensembles, XGBoost (Chen and Guestrin, 2016) was used. The selected features are extracted from the targets arranged in the data matrix $\mathbf{X} \in \mathbb{R}^{N_o \times l}$, where l is the total number of features. We consider three classes labeled by $\mathbf{y} \in \{0, 1, 2\}$ as per the class. We then analyzed the trade-off between accuracy and feature numbers. For each model, we selected a set of hyper-parameters to tune these models; up to 20 features were considered. Note that 70% data were used for training and denoted by \mathbf{X}_{train} , \mathbf{y}_{train} ; the rest 30% data were used for testing and are denoted by \mathbf{X}_{test} , \mathbf{y}_{test} . Using a model that supported only greedy feature selection algorithms, such as SFS, Algorithm 1 was applied. For each number of features n and each hyper-parameter choice h , SFS selects a column subset of \mathbf{X} as per the Forward SFS procedure described in (Ferri et al., 1994). The resulting subset was denoted by \mathbf{X}_f . The model's accuracy, a , was assessed using \mathbf{X}_f and stratified-5-fold cross-validation (CV). For each n the best tuned model was stored. A single training step shows the importance of a feature for a classifier that supports model-based feature selection (MBFS) given a model m with hyper-parameter h . Because of the data set \mathbf{X} , ModelBasedFeatureSelection returns the column indexes of \mathbf{X} ordered by importance. For example, if $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$, and the third feature is the most important, followed by the first and the second, the function would return $\mathbf{f} = [3, 1, 2]$. The results for each model are shown in **Figures 10A–C**. For reference, the score on the test set was plotted. The different classifiers have extremely similar performance trends. After considering four features, each classifier surpassed 90% accuracy. Let us consider the XGBoost classifier, which was the model with the highest accuracy on the training set, after 5-fold CV, it used 16 features and achieved an average score of 97.3%. This model was selected for testing, and yielded an accuracy of 96.5%. The confusion matrix is shown in **Figure 11A**, the importance of the features can be measured by different criterion, for example, by *weight* as shown in **Figure 12A** (i.e., the number of times a feature is used to split the tree node) or by *gain* as shown **Figure 12B** (i.e., the accuracy improvement achieved on average by the feature).

The results of our proposed approach are compared with the three feature approach proposed in (Hyun and Jin, 2020) for the human and vehicle classification. In this approach, the first feature x_1 maintains the number of Doppler spectrum points above a certain threshold. The second feature x_2 maintains the variation in x_1 from frame to frame, and the last feature x_3 keeps the echo's power fluctuation from frame to frame. Typically, the pedestrian's Doppler spectrum contains many Doppler frequencies because of the multiple body points in motion, whereas the car's Doppler spectrum typically contains a single dominant frequency. Therefore threshold-based features confirm to be extremely effective to distinguish these targets. The introduction of cyclists helped modify and expand this approach. The Doppler spectra of cyclists and pedestrians demonstrate multiple common traits. The application of features proposed in (Hyun and Jin, 2020) afforded us the scores shown in **Figure 11B**.

Algorithm 1. SFS model training

```

Result: M, A
 $N_{max} = 20;$  // Maximum feature number
 $\mathbf{H} = \mathbf{h}_{1 \times L};$  // Set of all hyper-parameters
 $\mathbf{M} = \{\}_{1 \times N_{max}};$  // Collection of  $N_{max}$  best
// trained models
 $\mathbf{A} = \mathbf{0}_{1 \times N_{max}};$  //Best accuracy scores
 $\mathbf{X} = \mathbf{X}_{train};$  // Training data
 $\mathbf{y} = \mathbf{y}_{train};$  // Training labels
 $n = l = 1;$ 
while  $n \leq N_{max}$  do
  while  $l < L$  do
     $m = \text{CreateModelInstance}(\mathbf{H}[l]);$ 
     $\mathbf{X}_f = \text{SequentialFeatureSelection}(m, \mathbf{X}, \mathbf{y}, n);$ 
     $a = \text{CrossValidation}(m, \mathbf{X}_f, \mathbf{y});$ 
    if  $a > \mathbf{A}[n]$  then
       $\mathbf{A}[n] = a;$ 
       $\text{store}(\mathbf{M}\{n\}, m);$  // Store the best
      // model in M
    end
     $l = l + 1;$ 
  end
   $n = n + 1;$ 
end

```

Comparing the confusion matrices of both approaches, we can clearly see that the results obtained by our approach offers higher accuracy than other approaches. Furthermore, graphs in **Figures 12A,B**, show ranking by importance for each feature.

Algorithm 2. MBFS model training

```

Result: M, A
 $N_{max} = 20;$  // Maximum feature number
 $\mathbf{H} = \mathbf{h}_{1 \times L};$  // Set of all hyper-parameters
 $\mathbf{A} = \{\}_{1 \times N_{max}};$  // Collection of  $N_{max}$  best
// trained models
 $\mathbf{A} = \mathbf{0}_{1 \times N_{max}};$  //Best accuracy scores
 $\mathbf{X} = \mathbf{X}_{train};$  // Training data
 $\mathbf{y} = \mathbf{y}_{train};$  // Training labels
 $n = l = 0;$ 
while  $l < L$  do
   $m = \text{CreateModelInstance}(\mathbf{H}[l]);$ 
   $m_{trained} = \text{train}(m, \mathbf{X}, \mathbf{y});$ 
   $\mathbf{f} = \text{ModelBasedFeatureSelection}(m_{trained}, \mathbf{X});$ 
  while  $n < N_{max}$  do
     $m' = \text{CreateModelInstance}(\mathbf{H}[l]);$ 
     $\mathbf{X}_f = \text{SelectColumns}(\mathbf{X}, \mathbf{f}, n);$ 
    //  $n$  most important col. of  $\mathbf{X}$ 
     $a = \text{CrossValidation}(m', \mathbf{X}_f, \mathbf{y});$ 
    if  $a > \mathbf{A}[n]$  then
       $\mathbf{A}[n] = a;$ 
       $\text{store}(\mathbf{M}\{n\}, m);$  // Store the best
      // model in M
    end
  end
   $n = n + 1;$ 
end

```

5.3 Multi-Directional Data

We extended our model to consider all three directions, as shown in **Figure 7**. **Figure 13** shows the Doppler spectra. For the diagonal motion, the primary characteristics of the pedestrian and cyclist are preserved; however, the exposure of the car's wheels to the field of view

of the radar makes the spread of weak frequencies more prominent and reduces the asymmetry, as shown in **Figure 9C**. For perpendicular motion, the Doppler information is minimal. There is a clear similarity between the frequency distributions around the dominant Doppler for the cyclist and car; in fact, we will show subsequently that these two targets can be easily confused by classifiers.

5.4 Multi-Directional Models

We followed the steps described in Algorithm 1 and Algorithm 2. In addition to target classification, we trained classifiers to predict the direction of motion; consequently, we obtain nine classes. **Figures 14A–C** shows the corresponding results of each model; these figures show that the average performance was considerably low this time. The highest accuracy on training and testing was achieved by XGBoost. Using 19 features, the classifier scored 63.0% on the test set and 72.6% (the highest) on the train set with the five-fold CV. The confusion matrix in **Figure 15** shows that the selected features yield promising results but they are insufficient to reliably predict the target *and* the direction of motion. However, let's examine the performance based on target type (pedestrian, cyclist, or car) classification without predicting the direction of motion. Here, very good results can be observed in the confusion matrices shown in **Figure 16**.

If the target is moving in the radial direction, then an average of 90% cases are correctly classified. The most challenging scenario occurs when the target moves perpendicular to the radar. From **Figure 16C**, we see that it is difficult to distinguish between the cyclist and car. However, the pedestrian was correctly classified in 84% of cases. In **Figure 17**, we show the importance of each feature as per the gain and weight.

6 CONCLUSION

This study explores the role of the commercial radar sensors for classification tasks. We demonstrated that radar technology can offer high accuracy for classifying road users using simple methods that are compatible with low-power applications. In addition to our model, the device can be directly implemented as a surveillance system or can be mounted on a vehicle while the performance levels are preserved. Our proposed algorithm has a few limitations, e.g., if the Doppler information is limited, as for targets moving along the azimuth direction, it will be necessary to consider additional complex features that might be difficult to design and interpret. In our scenarios, all targets moved; this was a necessary condition to exploit the Doppler effect. In future, the robustness of radar-based classification models to static targets can be increased if multiple transmitters and receivers are made available. In such cases, it is important to have access to geometrical information. A demonstration video of the work can be seen at <https://www.youtube.com/watch?v=AWY2Fhk7i74>.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SA and RC formulated the problem, while M-SA varified it. Most of the experiments are done by RC, SA was also participating in a few experiments. The initial draft of the paper is written by RC, while SA and M-SA finalized it.

REFERENCES

- Boulc, R., Thalmann, N. M., and Thalmann, D. (1990). A Global Human Walking Model with Real-Time Kinematic Personification. *Vis. Comput.* 6, 344–358. doi:10.1007/BF01901021
- Chen, T., and Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785
- Chen, V. C., Li, F., Ho, S.-S., and Wechsler, H. (2003). Analysis of Micro-doppler Signatures. *IEE Proc. Radar Sonar Navig.* 150, 271–276. doi:10.1049/ip-rsn:20030743
- Cooper, J. (1980). Scattering of Electromagnetic Fields by a Moving Boundary: The One-Dimensional Case. *IEEE Trans. Antennas Propagat.* 28, 791–795. doi:10.1109/TAP.1980.1142445
- Ferri, F. J., Pudil, P., Hatef, M., and Kittler, J. (1994). Comparative Study of Techniques for Large-Scale Feature Selection* *This Work Was Supported by a SERC grant GR/E 97549. The First Author Was Also Supported by a FPI grant from the Spanish MEC, PF92 73546684. *Machine Intelligence and Pattern Recognition* 16, 403–413. doi:10.1016/B978-0-444-81892-8.50040-7
- Hyun, E., and Jin, Y. (2020). Doppler-Spectrum Feature-Based Human-Vehicle Classification Scheme Using Machine Learning for an FMCW Radar Sensor. *Sensors* 20, 2001. doi:10.3390/s20072001
- [Dataset] Infineon (2021). *Demo Position2go*. Neubiberg, Germany: Infineon. Available at: <https://www.infineon.com/cms/en/product/evaluation-boards/demo-position2go/>.
- Mark, A., and Richards, P. (2015). *Fundamentals of Radar Signal Processing, Second Edition, Vol. 53*.
- Melvin, W. L., and Scheer, J. A. (2012). Principles of Modern Radar: Advanced Techniques. *Adv. Tech.* doi:10.1049/SBRA020E
- Melvin, W. L., and Scheer, J. A. (2013). Principles of Modern Radar: Volume 3: Radar Applications. *Radar Appl.* 3. doi:10.1049/SBRA503E
- Patel, K., Rambach, K., Visentin, T., Rusev, D., Pfeiffer, M., and Yang, B. (2019). “Deep Learning-Based Object Classification on Automotive Radar Spectra,” in 2019 IEEE Radar Conference, RadarConf (Boston, MA, USA: IEEE), 1–6. doi:10.1109/RADAR.2019.8835775
- Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., and Mueller, A. (2015). Scikit-learn: Machine Learning in Python. *J. Machine Learn. Res.* 19, 29–33.

FUNDING

The KAUST will pay the open access publication fees. The funds are coming from the KAUST Impact grant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frsip.2022.864538/full#supplementary-material>

- Perez, R., Schubert, F., Rasshofer, R., and Biebl, E. (2018). “Single-Frame Vulnerable Road Users Classification with a 77 GHz FMCW Radar Sensor and a Convolutional Neural Network,” in 2018 19th International Radar Symposium (IRS) (Bonn, Germany: IEEE), 1–10. doi:10.23919/IRS.2018.8448126
- Richards, M. A., Scheer, J. A., and Holm, W. A. (2010). Principles of Modern Radar. *Basic Principles*. doi:10.1049/sbra021e
- Stolz, M., Schubert, E., Meinel, F., Kunert, M., and Menzel, W. (2017). “Multi-target Reflection point Model of Cyclists for Automotive Radar,” in 2017 European Radar Conference (EURAD) (Nuremberg, Germany: IEEE), 94–97. doi:10.23919/EURAD.2017.8249155
- Tahmouh, D., Silvious, J., and Clark, J. (2010). An UGS Radar with Micro-doppler Capabilities for Wide Area Persistent Surveillance. *Radar Sensor Technol.* 7669, 766904. doi:10.1117/12.848233
- Trevor, H., Robert, T., and Jerome, F. (2009). *The Elements of Statistical Learning, Vol. 27*. Berlin, Germany: Springer.
- Will, C., Vaishnav, P., Chakraborty, A., and Santra, A. (2019). Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar. *IEEE Sensors J.* 19, 7283–7299. doi:10.1109/JSEN.2019.2914365
- Youngwook Kim, Y., and Hao Ling, H. (2009). Human Activity Classification Based on Micro-doppler Signatures Using a Support Vector Machine. *IEEE Trans. Geosci. Remote Sensing* 47, 1328–1337. doi:10.1109/TGRS.2009.2012849
- Zenaldin, M., and Narayanan, R. M. (2016). Radar Micro-doppler Based Human Activity Classification for Indoor and Outdoor Environments. *Radar Sensor Technol.* 9829, 2228397. doi:10.1117/12.2228397

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Coppola, Ahmed and Alouini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.