Range Aware Recognition of Hand Gestures Using FMCW Radar

Iida, Yosuke / 飯田, 耀介

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
19
(開始ページ / Start Page)
1
(終了ページ / End Page)
8
(発行年 / Year)
2024-03-24
(URL)
https://doi.org/10.15002/00030606

Range Aware Recognition of Hand Gestures Using FMCW Radar

Yosuke Iida
Graduation School of Computer and Information Sciences
Hosei University, Tokyo, Japan
yosuke.iida.5m@stu.hosei.ac.jp

***Abstract*—This paper introduces a hand gesture classification implementation that combines FMCW radar and deep learning. Unlike previous work, this research implements a range-aware methodology to automatically locate the hand range bin enabling more accurate classification. Our proposal is based on the observations made about the impact of the subject chest in the Time-Range map and the signature of the hand movement in the Range-Doppler map. Then, three distinct methods are proposed for hand range bin localization. Method I and II take advantage of the first observation by locating the chest and using it as a reference to select the hand bin. On the other hand, Method III exploits the second observation about the Doppler signature of the hand movement to directly select the hand bin. Both subject-dependent and independent evaluations are performed to classify six hand gestures and investigate the impact of several schemes including the type of input data, the use of a CNN-only and CNN-LSTM layer, an increased range of up to 150 cm, three types of angles, and presence of nearby person scenarios. The evaluation showed an average accuracy of 98.72% for subject-independent and 98.69% for subject-independent, with automatic hand position detection even at distances of up to 150 cm.**

Keywords—FMCW radar, range bins, range-aware, Doppler, hand gesture, deep learning.

I. INTRODUCTION

In the past, gesture recognition primarily depended on visual, infrared-based systems, and wearable devices that required physical contact. However, these conventional methods have their drawbacks, especially when used in different environmental conditions and lighting situations, due to their invasive nature. This led to the rise of Radio Frequency (RF) technology including various types of radars such as Continuous Wave (CW), Frequency Modulated Continuous Wave (FMCW), and Ultra-wideband (UWB) in addition to Wi-Fi, as a promising research avenue to address the limitations of traditional methods. Among these, FMCW is used in this paper due to its flexibility and features including Multiple-Input Multiple-Output (MIMO) antennas, range, angle, and velocity that can be achieved.

Numerous studies have explored radar technology for hand gesture recognition, particularly the FMCW type. Table I presents a comprehensive overview of the most recent research, highlighting various radar platforms, parameters, and classification methodologies. [1] used a 24 GHz FMCW radar to classify four distinct gestures through a custom CNN model, achieving 99.9% accuracy under subject-dependent conditions. The work in [2-8] used 60 GHz-based radar, while [9-18] adopted a 77 GHz radar. Among these studies, [2] classified five gestures using Range-Doppler images and a CNN-LSTM model, reporting 94.34% accuracy. Some research combined multiple radar data types such as TR, TD, and RD maps for classification, demonstrating the feasibility of using radar technology in diverse applications, such as controlling a Nursing Bed with a CRNN-based model to classify six gestures [3]. Furthermore, a real-time-based system [6] integrated radar data with Transformer encoder-based classification, achieving 93.95% accuracy for various hand gestures within a range of 10 to 20 cm.

However, these studies exhibit several limitations. The precise localization of the hand's range bin within radar data remains a challenge due to radar signal sensitivity to environmental reflections. Most of these studies bypassed this challenge, relying on fixed-range solutions or limiting the radar's Field of View (FoV) to the hand. This approach, while valuable for assessing feasibility, is impractical in real-world scenarios. These studies tend to concentrate their efforts on relatively short ranges, typically within the confines of 50 cm [19]. This narrow scope overlooks the broader contextual factors that come into play when a user interacts with a radar system at varying distances from the radar source.

To tackle these issues, the two basic observations are based on. First, the chest signature emerges as the most prominent and identifiable element [20] in the radar signal compared to other body parts for an idle subject, serving as a reliable point of reference for pinpointing hand location. Second, hand movements generate unique Doppler signatures, allowing for effective differentiation from other bodily signals through the analysis of range and Doppler information.

Based on these observations, three methods are proposed to achieve a range-aware implementation that can automatically detect the hand location. The first two methods first locate the chest and then estimate the hand's relative position, while the third method leverages the Doppler signature of hand movement. All three methods produce Time-Doppler matrices and spectrogram images for CNN-based classification. One of the goals in this research is to assess and figure out the most effective method, considering factors like data type, deep learning model, radar distance/angle, and nearby person impact. The results indicate that Method III with a base CNN achieves the highest classification accuracy, with 98.72% in subject-dependent evaluations and 98.69% in subject-independent scenarios. The main contributions of this research include the following.

- *Range-aware automatic selection of hand-bin*: Proposed three methods for detecting accurate hand positions using range and Doppler data from FMCW radar.
- *Combinational optimization of the methods, data types, and classification models*: Classified six gestures based on the best combination achieving 98.69% in the range of 30 to 150 cm, which is farther than conventional work.
- *Extensive evaluation of the proposed methods and impact factors*: Evaluated the performance and limitations of the proposed method through different distances, angles, and the presence of a nearby person.

Supervisor: Prof. Jianhua Ma

TABLE I. RELATED WORK ON FMCW RADAR-BASED HAND GESTURE CLASSIFICATION

Ref.	Freq. (GHz)	Frame (ms)	Num. of Chirps	Resolution		Radar Data Used	Algorithm Used	Num. of Gestures	Eval.	Acc. (%)
				Range	*Velocity*					
[1], 2022	24	–	–	–	–	TD	CNN	4	SD	99.9
[2], 2018	60	100	16	2.1 cm	244 cm/s	RD	CNN & LSTM	5	SI	94.3
[3], 2021	60	24	32	3 cm	16 cm/s	RD	CNN & LSTM	6	SI	94.3
[4], 2021	60	200	16	–	–	2-axis TR	EMA	2	–	–
[5], 2021	60	200	–	–	–	RD	CNN	11	SI	78.9
[6], 2022	60	–	16	2.5 cm	122 cm/s	RD	Transformer	20	SD	94.0
[7], 2022	60	20	256	–	3.9 cm/s	3-axis TR	CNN	5	SI	89.1
[8], 2023	60	100	128	5 cm	–	TR, TD, TA	CNN & LSTM	12	SI	> 60
[9], 2020	77	25	64	4.8 cm	15.4 cm/s	TR, TD, TA	CNN	8	SI	84.6
[10], 2020	77	14	128	15 cm	8.1 cm/s	TD	CNN	8	SI	90.5
[11], 2020	77	10	–	4.7 cm	36 cm/s	RD	CNN & HMM	10	SI	97.4
[12], 2021	77	20	32	4 cm	22 cm/s	2-axis TR, TD	MobileNetV2	3	SD	> 90
[13], 2021	77	55	-	3.8 cm	–	RD	EVL-NN	5	SI	89
[14], 2021	77	41	256	3.8 cm	4.8 cm/s	TD	CNN	6	SI	85.3
[15], 2022	77	32	32	–	–	TR	CNN	5	SD	98.3
[16], 2022	77	30	–	3.9 cm	13 cm/s	2-axis TR, TD	LSTM	12	SD	94.3
[17], 2022	77	20	128	–	–	TR, TD, TA, DA, RA	CNN	6	SI	92.0
[18], 2023	77	40	128	3.9 cm	–	TR, TD	CNN & Transformer	6	SD	99.2

"–": Not mentioned, **T**: Time, **R**: Range, **D**: Doppler, **A**: Angle, **SD**: Subject-dependent, **SI**: Subject-independent

The remainder of this paper is structured as follows. Section II delves into the details of the proposed range-aware hand gesture classification method. The experimental setup and the dataset details are explained in Section III, while the evaluation results are reported in Section IV. The conclusion and future work are outlined in the last section.

II. HAND RANGE-AWARE GESTURE RECOGNITION

This section offers a comprehensive overview of our range-aware hand gesture classification pipeline as illustrated in Fig. 1. The pipeline has four main steps: raw data collection and pre-processing, hand-range estimation, data formatting and preparation, and hand gesture classification.

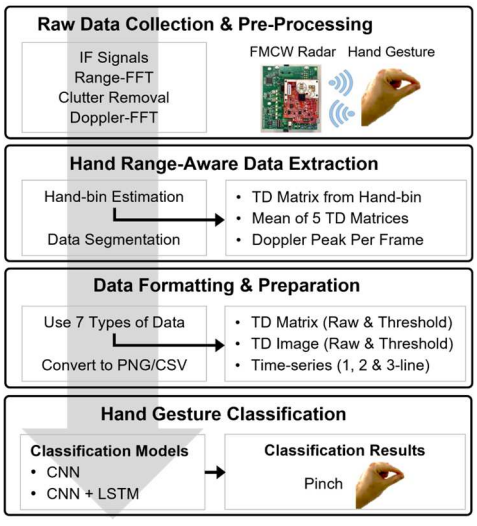

Fig. 1. Range-aware hand gesture classification pipeline.

A. Raw Data Collection and Pre-processing

The first task in the proposed pipeline is collecting data using radar and performing the different pre-processing steps required. The FMCW radar adopted in this research is the Texas Instruments (TI) board operating at 60 GHz frequency (IWR6843ODS) [21] coupled with a DCA1000 acquisition card [22] that helps obtain the raw data with higher transfer bandwidth. The raw data obtained from the FMCW radar and its processing flow are shown in Fig. 2.

The radar generates a sinusoidal signal called "chirp" whose frequency increases linearly with time. This signal is emitted via the transmitting (TX) antenna and subsequently received by the receiving (RX) antenna. An intermediate frequency (IF) signal is derived from the difference between the transmitted TX and the received RX signals. The IF signal is then converted into a digital format using an Analog-to-Digital Converter (ADC). Subsequently, a Fast Fourier Transform (FFT) is applied to extract range information by separating the reflections from each chirp into a series of distances called "range bins" where each range bin is a multiple of 7.8cm (radar range resolution). This operation is commonly referred to as 1DFFT or Range-FFT. Each range-bin data can be aligned according to the temporal data obtained from each chirp inside a radar frame (128 chirps per frame), to form the Time-Range (TR) map where each frame is 25ms long. Furthermore, a clutter removal procedure is executed to suppress noise from static objects by removing the mean value of the signal, leaving only the signature from moving targets.

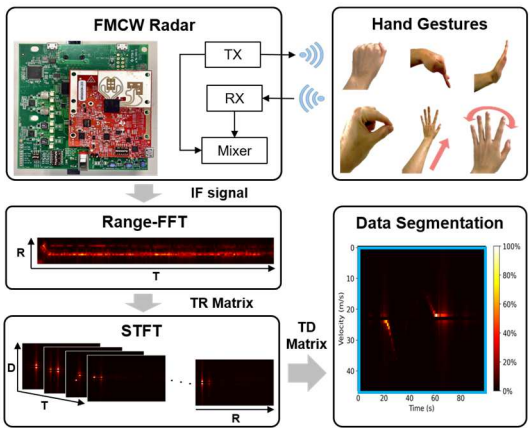

Fig. 2. Raw data collection and pre-processing flow.

Doppler-FFT is then applied to the TR sequence to construct the Time-Range-Doppler (TRD) matrix using Short-Time Fourier Transform (STFT). Subsequently, the Time-Doppler (TD) map is generated by selecting a particular range bin where the hand gestures are located. The precise estimation of hand range is very important since it has a direct impact on the quality of the signal, which may further influence the accuracy of the hand gesture classification.

B. Data Extraction

To estimate a correct hand range, i.e., the bin where the hand gesture is located, three different methods are proposed. Since a chest is always with displacement due to breathing and heartbeat that cause a relatively strong signal in radar data [23] [24], Methods I and II use this information to first locate the chest range bin as the strongest signal in the TR data. Then they temporarily select the 5th bin before the chest as the initial range bin for the hand. This temporary bin is selected when considering the bin size (7.8cm) and the typical distance between the extended arm and the chest (≈ 40cm).

As depicted in Fig. 3, Method I will determine the exact hand-bin by further selecting the strongest signal among the 5 bins centered around the temporary hand-bin and forms the final TD matrix data based on the determined bin. This range is chosen based on the possible distance covered when extending and retracting the hand which can be up to 15 cm (±2 range bins) before and after the temporary hand range bin.

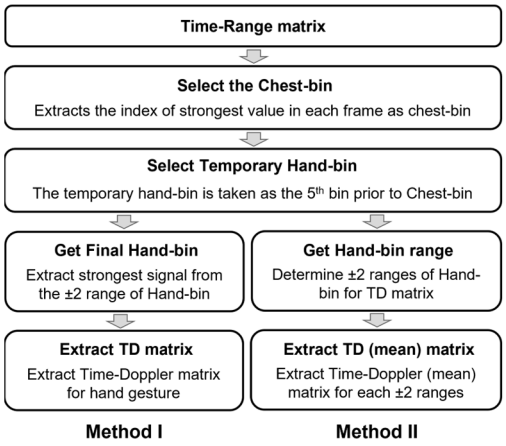

Fig. 3. Raw data collection and pre-processing flow.

On the other hand, instead of selecting the strongest bin, Method II generates the TD matrix data for all 5 bins surrounding the temporary hand bin and then takes the resulting mean matrix as the data used for classification. The pseudocodes for Method I and II are shown in Algorithm 1 and 2 respectively, while the sample data of TR and TD are given in Fig. 4.

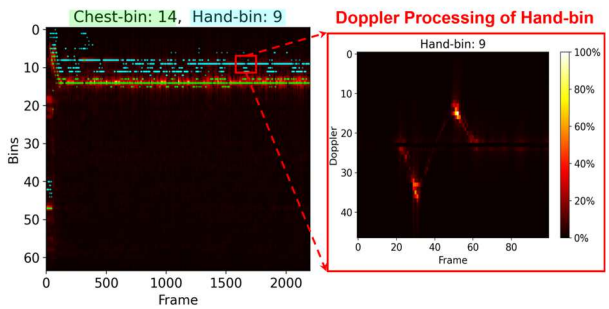

Fig. 4. Example of the output from Method I and II: TR (*left*), TD (*right*).

Algorithm 1 of Method I

list *Bins* ← [ ]
for *R* **in** *TR* **do**
  Chest-bin ← **Argsort** (*R*) [−1]
  Hand-bin_tmp ← *Chest-bin* − 5
  Hand-bin ← **Argsort** (*R* [*Hand-bin_tmp* − 2, *Hand-bin_tmp* + 2]) [−1]
  Add *Hand-bin* **to** *Bins*
end for
Hand-bin ← **Mode** (*Bins*)
Get *TD* **from** *Hand-bin*

Algorithm 2 of Method II

list *Bins* ← [ ]
for *R* **in** *TR* **do**
  Chest-bin ← **Argsort** (R) [−1]
  Add *Chest-bin* **to** *Bins*
end for
Hand-bin_tmp ← **Mode** (*Bins*) − 5
Hand-bin_range ← *R* [*Hand-bin_tmp* − 2, *Hand-bin_tmp* + 2]
Get 5 *TDs* **from** *Hand-bin_range*
TD ← **Mean** (5 *TDs*)

Unlike Methods I and II, Method III relies on detecting the hand location based on the nearest detected movement in the Range-Doppler (RD) data [25]. As shown in Fig. 5, this is achieved by first applying a max filter with a neighborhood of 3 to extract the peak values from each RD matrix followed by a threshold mask of 10 (determined empirically) to further highlight the relevant peaks by reducing the other values to 0.

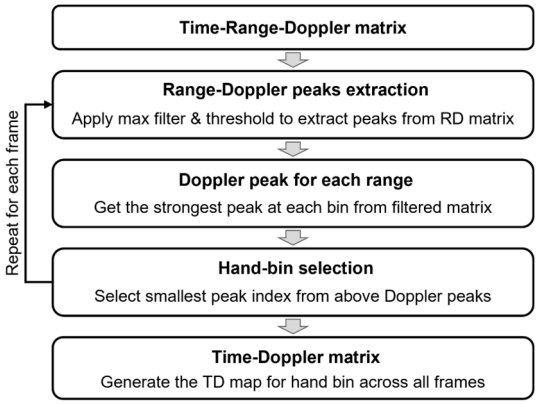

Fig. 5. Movement-based hand range detection and data extraction.

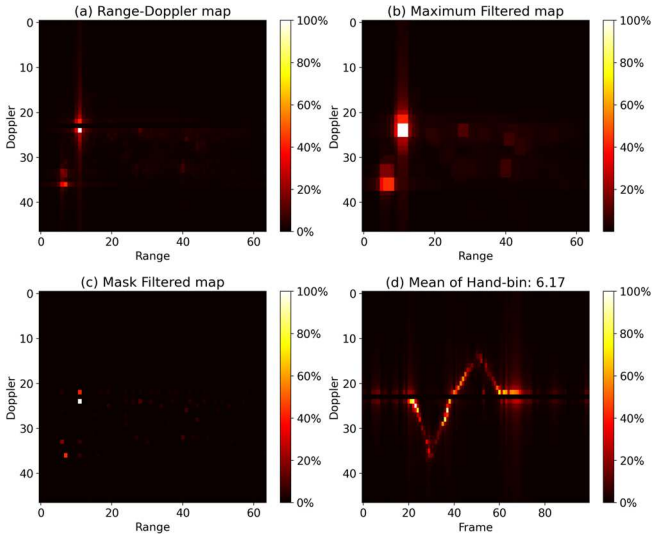

Fig. 6. Resulting maps at each step of Method III.

Next, the strongest Doppler peaks are extracted from the filtered results of each range bin, and the smallest bin index among these peaks is taken as the hand location. This process

is repeated for each frame in the data. Finally, the set of hand locations across all frames is used to generate a TD map for the hand gesture. Algorithm 3 summarizes the implementation steps of Method III. Visualizing the results as an image, it can be seen that the combination of Doppler information between frames effectively captures hand gesture features, as shown in Fig. 6.

Algorithm 3 of Method III

list *Dopplers* ← [ ]
for *RD* **in** *TRD* **do**
  RD_MF ← **Maximum Filter** (*RD*, 3)
  RD_Mask ← **np.where** (*RD*, *RD_MF*, *RD*, 0)
  if *RD_Mask* ≤ 10 **then**
    RD_Mask ← 0
  end if
  list *Doppler_peaks* ← [ ]
  for *R* **in** *RD_Mask* **do**
    Add *Doppler_peaks* **to Max** (*R*)
  end for
  Hand-bin ← **np.where** (*Doppler_peaks* ≠ 0) [0]
  Get *D* **form** *Hand-bin*
  Add *D* **to** *Dopplers*
end for
Get *TD* **from** *Dopplers*

C. Data Formatting

Seven distinct data formats are derived from the Time-Doppler (TD) data generated by the three proposed methods explained previously. Fig. 7 shows the four main visual representations of these data formats. The raw data format involves the use of the resulting TD matrix in a numerical format. This is done by saving the result in its original form within a CSV file and treating it as raw data. The threshold data format introduces a threshold mechanism, set at three times the average value of the entire matrix (set empirically), with values below this threshold set to zero. The modified matrix is then saved as threshold data in another CSV file.

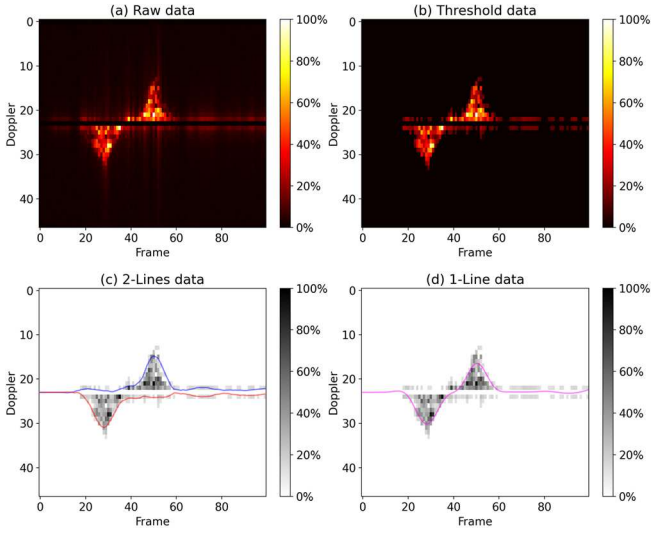

Fig. 7. Four main distinct data formats.

The above raw and threshold data are also represented visually as (a) and (b), respectively, and are treated as image data. In these formats, the original matrices are rendered into spectrograms, where the intensity of color is used to convey signal strength. This approach provides a different perspective on the data, for observing signal variations and patterns that may not be so clear in the numerical representations.

The time series data traces the edges of the gesture features in the positive and negative directions in each frame based on the threshold matrix data shown in (b). The resulting lines are smoothed using a Savitzky-Golay filter to obtain a 2-line format shown in (c). The 1-line format is obtained from (d) the difference between these two lines, and the 3-line format from the combination of these three lines.

D. Hand Gesture Classification

Two CNN-based models are investigated to classify the different hand gestures from the collected data. Fig. 8 provides a visual representation of the Convolutional, Dense, and LSTM layers employed in these models, along with their respective parameters and structural arrangements.

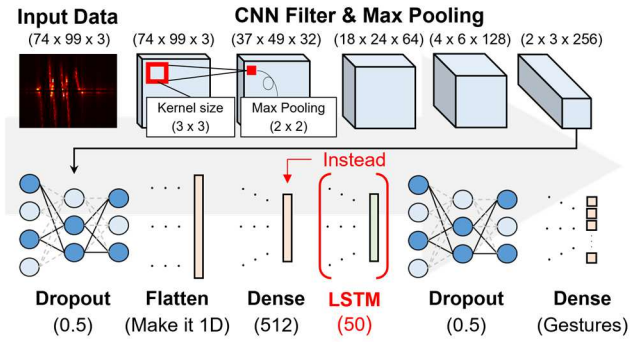

Fig. 8. CNN-only and CNN-LSTM architectures adopted.

The input data for both models depend on the type of data, with a structure of 100 x 47 for matrix data, 99 x 74 x 3 for image data, and 100 x N (number of lines) for line data. The first model is based on a basic CNN architecture comprising five convolutional layers interspersed with five Max pooling layers. Max pooling is implemented to reduce the size of the feature map in the convolutional layer, retaining only robust features. To prevent overfitting during training, a dropout layer with a 0.5 ratio is incorporated. Subsequently, dense layers are used for classification, relying on the features extracted in the previous convolutional part of the model. The final dense layer, representing the model's output, is sized according to the number of gestures learned (six gestures).

The second model is similar to the first one but replaces one of the dense layers with an LSTM layer as highlighted in red in Fig. 8. Both models use the Adam optimizer with a learning rate of 0.001, a batch size of 32, and 500 epochs for training. An early stopping condition is implemented, ensuring that the training halts if the loss value fails to decrease over several epochs and saves the model in the state with the lowest loss value.

III. Experimental Setup and Hand Gesture Dataset

This section presents details about the experiment scene and data collection protocol. In addition, the hand gestures considered in this study, the dataset collected, and their use for training and evaluation of the models are explained.

A. Experimental Setup and Protocol

The hand gestures implemented in this work were selected to satisfy an intention for simple control of contactless hand gesture-based applications. As shown in Fig. 9, the six hand gestures from big gestures like push to tiny gestures like pinch were used in this study.

The experimental procedures were carried out in the laboratory involving fourteen subjects. Fig. 10 illustrates the setting of the range experiment scene, in which the radar device mentioned in Section II (A) was used. In each experiment, the subject is asked to sit in front of the radar at

a specific distance. As shown in Fig. 11, the subject performs the same gesture 10 times following the beep sounds. The experiments for all six gestures are repeated once for each distance, namely 30, 60, 90, 120, and 150cm away from the radar. In addition, during the experimental measurement, only one subject is in the range to avoid the detection of other moving objects. This beep is also used in the experiments in terms of angle and nearby person impact described below.

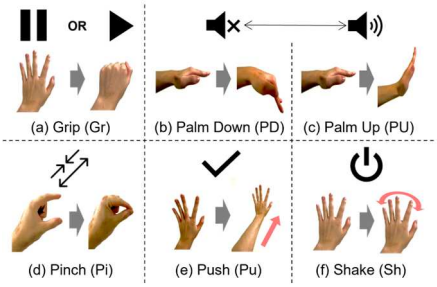

Fig. 9. Hand gesture types used in the experiment.

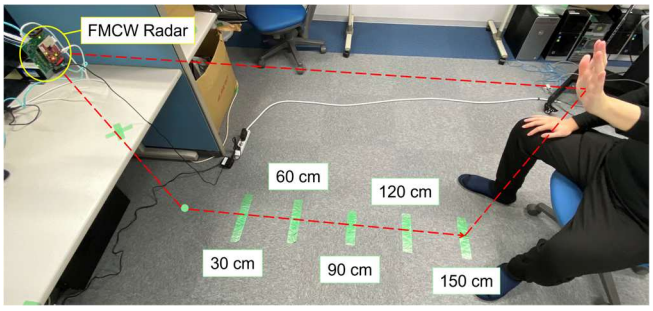

Fig. 10. Experimental scene of range conditions.

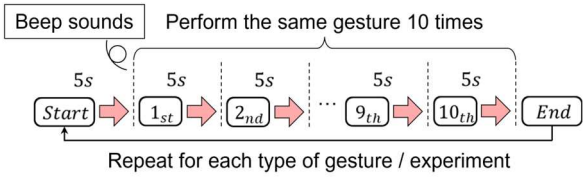

Fig. 11. Hand gesture data collection protocol.

In the angle variation experiment, contrary to the range condition, the hand gesture is performed from a position other than directly in front of the radar. Fig. 12 illustrates three distinct angular conditions: (i) Horizontal (45 degrees to the right from the radar), (ii) Vertical (upper side from radar), and (iii) Diagonal (combination of horizontal and vertical). The vertical angle may vary based on the physical characteristics of the subject. Under all the above angular conditions, the hand-to-radar distance is approximately 60 cm. This experiment was composed of three types of gestures with representative characteristics: small movement Pinch, large movement Push, and fast movement Shake. These angular conditions and gesture sets are aimed at evaluating the effectiveness of the proposed method.

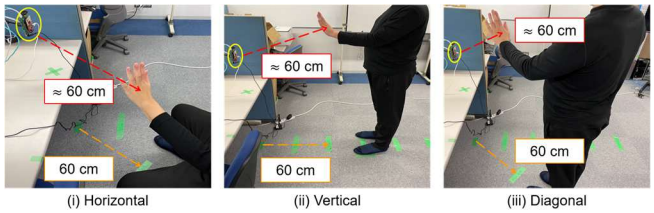

Fig. 12. Experimental scene of angle conditions.

In the experiments of nearby person conditions, a person other than the subject performing the gesture was used to investigate the impact of dynamic signals from the nearby person. As shown in Fig. 13, the subject executing the gesture was positioned 60 cm in front of the radar, while (i) another person sat next to the gesturing subject, and (ii) a person moved horizontally with the radar positioned behind the gesturing subject. To indicate the detailed distances, in the experiment (i), the distance between both persons is approximately 30 cm, and the distance between the radar and the sitting persons is about 120 cm. On the other hand, in the experiment (ii), the walking person is approximately 180 cm from the radar and approximately 40 cm from the subject performing the gesture. The gestures used are the same as in the angle condition experiment.

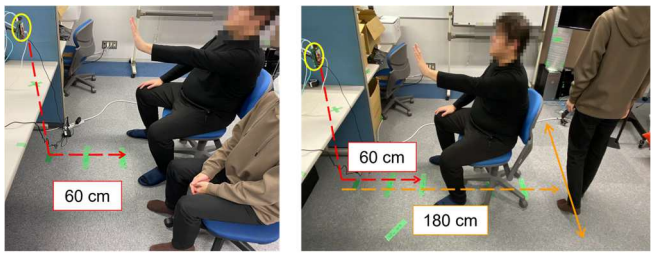

Fig. 13. Experimental scene of nearby person conditions.

A laptop PC is connected to the radar through an Ethernet cable to collect the data, and save it into a binary file with TI mmWave Studio software [26]. Subsequently, the binary file is parsed by MATLAB while all other data processing, model training, and evaluation are performed using a custom Python code developed at our laboratory. The deep learning models were trained on a desktop PC with a 6 CPU core AMD Ryzen 5, NVIDIA RTX3060 GPU, and 16 GB RAM for higher computational speed.

B. Hand Gesture Dataset and Data Partitioning

Considering the range experiment for fourteen subjects (they did not participate in all experiments), six different gestures repeated ten times, across three different distances, our dataset has a total of 3600 samples (600 samples for each gesture). The image data was saved in PNG format while the matrix data was saved as CSV files organized according to the six gesture classes. In the case of training for subject-dependent evaluation, the data are separated using the split ratio of 8:1:1 while applying a 10-fold cross-validation process. On the other hand, for subject-independent, Leave-One-Out Cross-Validation (LOOCV) is adopted to set the data from ten subjects for training and the remaining subject data for testing.

A total of five subjects participated in the angles and nearby person experiment, performing three different gestures. Three different angles and two multiple-person scenarios were conducted, and the data set included 150 samples (50 samples per gesture) for each experiment. This dataset is utilized as test data, employing a model trained on the dataset collected in the range condition experiments. To evaluate the generality of the training model, each subject is removed from the range condition experiment, which is treated as the training data set, so that both data sets do not contain data from the same subject.

IV. Results and Discussions

This section summarizes the hand gesture classification results from the CNN model compared with the combination of CNN and LSTM for each of the three proposed hand-bin estimation methods. Furthermore, in-depth discussions about the impact of data type, change in distance, subject-dependent vs independent, angle, and nearby person performance based on the results are given.

A. Impact of Data Type and Model

Table II shows the classification accuracies and their averages for each combination of model, data partitioning method, proposed method, and data format on matrix data. When using the matrix data type, it is clear that using the raw data with Method II produced the best results for both subject-dependent and subject-independent assessments. This may be due to the thresholding matrix data deleting important features of the gesture. In addition, since Method II uses the average map of five TD maps, the features may have been fully used in the learning process.

In addition, a slight decrease in accuracy was observed when transitioning from subject-dependent to subject-independent evaluations. This discrepancy can be attributed to the model training approach, where subject-dependent models are trained on data from all subjects. In contrast, subject-independent models are tested on subjects not included in the training dataset. Furthermore, the CNN-only model is more accurate than the combination model of CNN and LSTM in all conditions.

TABLE II. Accuracy Result for Matrix Data (Rate %).

Model		*CNN-only*		*CNN+LSTM*		**Ave.**
Data		*Raw*	*Threshold*	*Raw*	*Threshold*	
SD	*M I*	94.41	93.91	93.83	93.02	93.79
	M II	**96.75**	96.52	96.3	96.19	96.44
	M III	96.58	95.97	96.16	95.52	96.06
Ave.		95.91	95.47	95.43	94.91	95.43
SI	*M I*	92.5	91.63	92.13	90.47	91.68
	M II	**96.02**	95.55	95.36	95.0	95.48
	M III	95.83	95.5	94.47	93.66	94.86
Ave.		94.78	94.23	93.99	93.04	94.01

SD: Subject-dependent, **SI**: Subject-independent, **M**: Method

On the other hand, the combination of Method III and the CNN-only model in the case of image data generated the best results for both subject-dependent and independent recognition as shown in Table III. Although there is a drop in accuracy when considering subject-independent, the results are still over 97.4% which is considered good. This can be explained by the additional features that the model can learn from the rich image-based data compared to matrix data. Overall, image data demonstrated better performance than matrix data and line data for both subject-dependent and subject-independent classification with the combination of Method III and a CNN-only model.

For the line data, accuracy improves with increasing each additional line, and even more so for the CNN+LSTM model. This is because the line data is time series, and the effect of LSTM, which is suitable for this, is evident. Despite the smallest input size data format, the line data achieved 97.69% accuracy for the combination of Method III, CNN and LSTM model, and three lines data format on the subject-independent. The accuracy is already as good as 3-line when 2-line is used, and this is thought to be due to tracking hand Doppler, where features extend in the positive and negative directions of the TD map.

TABLE III. Accuracy Result for Image Data (Rate %).

Model		*CNN-only*		*CNN+LSTM*		**Ave.**
Data		*Raw*	*Threshold*	*Raw*	*Threshold*	
SD	*M I*	98.38	98.27	97.75	95.94	97.58
	M II	98.63	98.49	98.27	97.22	98.15
	M III	**98.72**	98.63	98.38	98.05	98.44
Ave.		98.58	98.46	98.13	97.07	98.06
SI	*M I*	98.33	97.91	97.47	95.13	97.21
	M II	98.58	98.08	97.69	96.14	97.62
	M III	**98.69**	98.58	98.11	97.41	98.2
Ave.		98.53	98.19	97.76	96.23	97.68

SD: Subject-dependent, **SI**: Subject-independent, **M**: Method

TABLE IV. Accuracy Result for Line Data (Rate %).

Model		*CNN-only*			*CNN+LSTM*			**Ave.**
Data		*1 L*	*2 L*	*3 L*	*1 L*	*2 L*	*3 L*	
SD	*M I*	82.72	94.19	94.99	91.52	95.91	96.58	92.65
	M II	86.58	96.0	96.22	93.94	96.72	97.61	94.51
	M III	90.27	96.27	96.88	94.8	97.63	**97.72**	95.59
Ave.		86.52	95.49	96.03	93.42	96.75	97.3	94.25
SI	*M I*	81.86	94.08	94.13	91.36	95.88	96.52	92.3
	M II	85.83	94.86	95.3	91.77	96.55	97.44	93.62
	M III	89.91	95.97	96.83	93.63	97.36	**97.69**	95.23
Ave.		85.87	94.97	95.42	92.25	96.6	97.22	93.72

SD: Subject-dependent, **SI**: Subject-independent, **M**: Method, **L**: Line

B. Impact of Hand Distance

Fig. 14 summarizes the classification performance of hand gestures at different distances: 30, 60, 90, 120, and 150 cm. To make this assessment, threshold image data are used in conjunction with Method III and a CNN-only model due to this combination previously showing the best results. In the case of subject-dependent evaluations, the model displayed high accuracy. It consistently exceeded 97% across all distances, confirming its reliability. Notably, it performed better at 30, 60, 90, and 120 cm but slightly less accurately at 150 cm, which was expected since the radar signal naturally weakens as the subject moves farther away, making it harder to detect gesture features.

However, in the case of subject-independent (training data does not include test users) evaluations, the accuracy was slightly lower than subject-dependent with a randomly split data set. At 150 cm, the accuracy was the lowest, but even then, it remained at a respectable accuracy of 97.5%. These results indicate that despite the challenges of subject-independent scenarios and increased distance, our model remains reliable and effective.

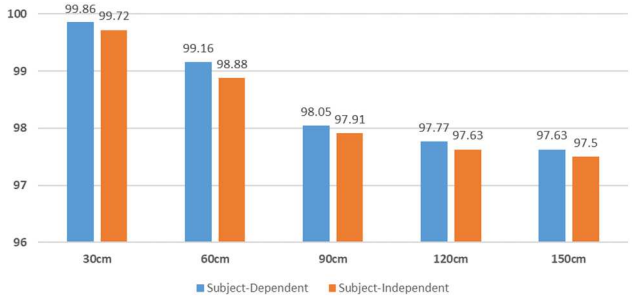

Fig. 14. Classification accuracy by distance.

C. *Impact of Gesture Type*

To check the impact of hand gesture type on performance, the classification accuracy is shown in Fig. 15 for the subject-dependent (a) and subject-independent (b) validation using image data, Method III, and a CNN-only model (the best combination shown previously). From (a), the classification model exhibited good reliability in distinguishing various hand gestures in a subject-dependent evaluation, achieving a classification rate of 98% and higher for nearly all gestures. Likewise, subject-independent (b) provides further evidence of our model's reliability and accuracy, with all gestures achieving over 97% accuracy.

The main misclassifications were between PD and PU, and from Ps to Pi. These may be attributed to the similarity of PD and PU gestures due to the symmetrical signature in the radar Doppler images, and Ps has some data on weak signals at distances of 90 cm or more. It is also possible that the strength of the signal varies with the speed at which the gesture is performed. This underscores the importance of including angle range maps in the future to address this limitation.

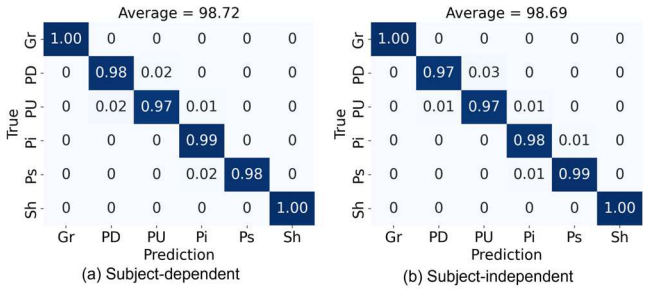

Fig. 15. Hand gesture classification (*subject-dependent vs independent*).

D. *Impact of Angle*

Figure 16 shows (a) the average accuracy of gesture classification at each angle, (b) the confusion matrix for the diagonal angle with the lowest accuracy, (c) the Push gesture in front of the radar, and (d) the Push gesture at the upper right angle from the radar, which contains noise. Normal is a condition in which there are no horizontal and vertical angles, just the radar and the subject facing each other, and the distance between them is 60 cm. Despite cross-validation to ensure that each subject participating in the collection of angle data was not included in the training set of range data, Normal achieved 100% accuracy for all of them. However, some subjects were less accurate for horizontal and vertical angles, and accuracy was even lower for angles in both angles.

The comparison of push gestures shows that (d) contains noise other than gesture features not found in (c). This noise might be attributed to the subject performing the gesture while standing, causing reflected signals from body parts other than the chest to have a greater impact compared to the seated state. In particular, the noise at the diagonal angle (d) was a noticeably larger feature than at the non-normal angle (c), which may have been a wrong learning process. In addition, in the experimental setup, this feature appeared due to the body being closer to the front of the radar than to the gesturing left hand. Therefore, the TD map must be trimmed so that only gesture features remain, and the state during non-gestures must be learned.

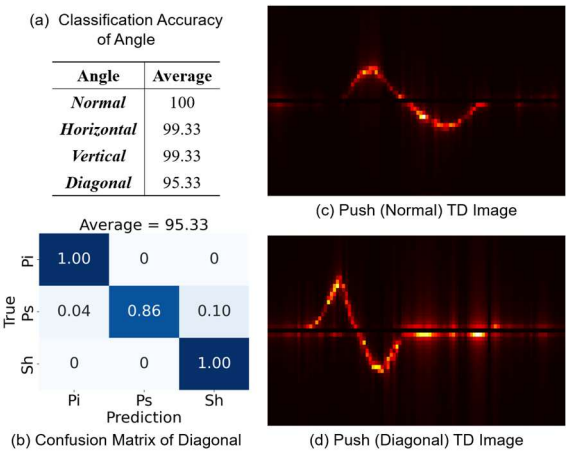

Fig. 16. Angle results and impact of decreased accuracy.

E. *Impact of Nearby Person*

This experiment is to investigate the impact of gestures in experiments with two people who are all within radar sensing range, including scenarios with a person sitting next to the subject or walking behind the subject. Fig. 17 shows the average classification accuracy per scenario in (a), the confusion matrix for the "walking person" experiment with the lowest accuracy (b), and comparison examples of the TR map for the Pinch gesture that affected accuracy for (c) Sitting and (d) Walking. While sufficient accuracy was obtained in the experiment with a person sitting next to the subject, the accuracy was significantly decreased in all subject test sets in the experiment with a person walking behind the subject. Among them, there was a high rate of misclassification for Pinch gestures with small motions and weak reflection signals.

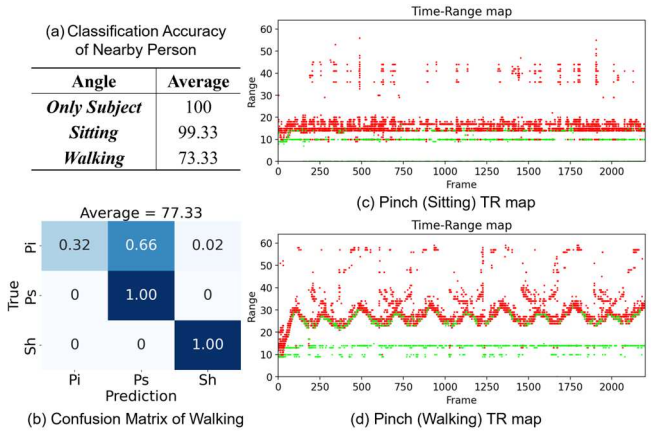

Fig. 17. Nearby person results and impact of decreased accuracy.

Comparing the TR maps, in scenario (c) a person walking is detected in the range of 20 to 35 of the range value, with scattered green dots indicating detection of hand position, and green dots are not detected stably at 9 or 10 of the range value expected at the hand detection position. This indicates that when people are next to each other, the effect is small due to their proximity to each other, while when people are far from each other (and also when people are continuously moving behind the subject), they are affected by stronger reflected signals. Thus, the accuracy of hand position detection may depend on the type of gesture and the distance between people, as well as the interference of large and/or fast-moving objects.

V. Conclusion and Future Work

In this paper, a range-aware hand gesture recognition system based on FMCW radar and deep learning is presented. The key contributions of this research include the proposals of three novel methods to automatically detect the hand range bin using radar data. Additionally, the classification of six hand gestures using a CNN-based model from a distance of up to 150 cm was achieved. The evaluation results showed that the combination of Method III with base CNN proved to be the best among the three proposed methods. This research achieves an average accuracy of 98.72% in the subject-dependent scenario and 98.69% in the subject-independent scenario surpassing the previously similar work [14]. In the angular condition, at a fixed distance of 60 cm, 99% was achieved at 45 degrees horizontally and vertically and 95% at diagonal angles. On the other hand, in the scenario with a nearby person, the result was good if the person was close to the other person, but the performance was affected if the person was far away from the subject and kept moving.

While the results show promise, there are some limitations and areas for improvement in future research. To enhance the model's accuracy and enable additional capabilities, it will be very useful to include angle information in combination with range and Doppler to improve our range awareness methodology and help distinguish similar gestures. Another important aspect is to increase the dataset by training more subjects and gesture types. It also needs to be robust to the effects of any distance or angle range, two or more subjects, or other dynamic interfering objects. In the future, the model can be further enhanced and more practical by training it to recognize special gestures for specific applications. This is a logical next step based on the results achieved by this research so far. Furthermore, the ability to recognize gestures from further away and to extract data only from subjects who are gesturing among multiple persons could open up new possibilities and make the technology even more versatile.

References

[1] H. Le, V. Hoang, V. Doan, and D. Le, "Dop-DenseNet: Densely Convolutional Neural Network-Based Gesture Recognition Using a Micro-Doppler Radar", Journal of Electromagnetic Engineering and Science, vol. 22, issue. 3, p335-343, May 2022.

[2] S. Hazra and A. Santra, "Robust Gesture Recognition Using Millimetric-Wave Radar System", IEEE Sensors Letters, vol. 2, issue. 4, pp. 1-4, November 2018.

[3] F. Du, Y. Zhao, H. Wang, and Q. Cao, "Hand Gesture Controlled Nursing Bed Using FMCW Radar", Journal of Physics: Conference Series, vol. 1948 (The 2021 2nd International Conference on Internet of Things, Artificial Intelligence and Mechanical Automation), May 2021.

[4] P. Nallabolu, L. Zhang, and C. Li, "Human Presence Sensing and Gesture Recognition for Smart Home Applications with Moving and Stationary Clutter Suppression Using a 60-GHz Digital Beamforming FMCW Radar", IEEE Access, vol. 9, pp. 72857-72866, May 2021.

[5] M. Scherer, M. Magno, J. Erb, P. Mayer, M. Eggimann, and L. Benini, "TinyRadarNN: Combining Spatial and Temporal Convolutional Neural Networks for Embedded Gesture Recognition with Short Range Radars", IEEE Internet of Things Journal, vol. 8, issue. 13, pp. 10336-10346, July 2021.

[6] J. Choi, C. Park, and J. Kim, "FMCW Radar-Based Real-Time Hand Gesture Recognition System Capable of Out-of-Distribution Detection", IEEE Access, vol. 10, pp. 87425-87434, August 2022.

[7] Z. Xia and F. Xu, "Time-Space Dimension Reduction of Millimeter-Wave Radar Point-Clouds for Smart-Home Hand-Gesture Recognition", IEEE Sensors Journal, vol. 22, issue. 5, pp. 4425-4437, March 2022.

[8] P. Zhao, C. Lu, B. Wang, N. Trigoni, and A. Markham, "CubeLearn: End-to-End Learning for Human Motion Recognition from Raw mmWave Radar Signals", IEEE Internet of Things Journal, vol. 10, issue. 12, June 2023.

[9] Z. Xia, Y. Luomei, C. Zhou, and F. Xu, "Multidimensional Feature Representation and Learning for Robust Hand-Gesture Recognition on Commercial Millimeter-Wave Radar", IEEE Transactions on Geoscience and Remote Sensing, vol. 59, issue. 6, pp. 4749-4764, July 2020.

[10] N. Kern, M. Steiner, R. Lorenzin, and C. Waldschmidt, "Robust Doppler-Based Gesture Recognition with Incoherent Automotive Radar Sensor Networks", IEEE Sensors Letters, vol. 4, issue. 11, pp. 1-4, October 2020.

[11] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Lu, 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, issue. 4, pp. 1-28, December 2020.

[12] L. Santana, A. Rocha, A. Guimaraes, I. Oliveira, J. Fernandes, S. Silva, and A. Teixeira, "Radar-Based Gesture Recognition Towards Supporting Communication in Aphasia: The Bedroom Scenario", International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, pp. 500-506, February 2022.

[13] H. Liu, A. Zhou, Z. Dong, Y. Sun, J. Zhang, L. Liu, H. Ma, J. Liu, and N. Yang, "M-Gesture: Person-Independent Real-Time In-Air Gesture Recognition Using Commodity Millimeter Wave Radar", IEEE Internet of Things Journal, vol. 9, issue. 5, pp. 3397-3415, July 2021.

[14] W. Jiang, Y. Ren, Y. Liu, Z. Wang, and X. Wang, "Recognition of Dynamic Hand Gesture Based on Mm-Wave FMCW Radar Micro-Doppler Signatures", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4905-4909, June 2021.

[15] B. Jin, Y. Peng, X. Kuang, Z. Zhang, Z. Lian, and B. Wang, "Robust Dynamic Hand Gesture Recognition Based on Millimeter Wave Radar Using Atten-TsNN", IEEE Sensors Journal, vol. 22, issue. 11, pp. 10861-10869, June 2022.

[16] P. Grobely and A. Narbudowicz, "MM-Wave Radar-Based Recognition of Multiple Hand Gestures Using Long Short-Term Memory (LSTM) Neural Network", Electronics 2022, vol. 11, issue. 5, no. 787, March 2022.

[17] Z. Liu, H. Liu, and C. Ma, "A Robust Hand Gesture Sensing and Recognition Based on Dual-Flow Fusion with FMCW Radar", IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, October 2022.

[18] B. Jin, X. Ma, Z. Zhang, Z. Lian, and B. Wang, "Interference-Robust Millimeter-Wave Radar-Based Dynamic Hand Gesture Recognition Using 2D CNN-Transformer Networks", IEEE Internet of Things Journal (Early Access), July 2023.

[19] S. Ahmed, K. Kallu, S. Ahmed, and S. Cho, "Hand Gestures Recognition Using Radar Sensors for Human-Computer-Interaction: A Review", Remote Sensing, vol. 13, issue. 3, no. 527, February 2021.

[20] B. Walid and J. Ma, "Accuracy Assessment and Improvement of FMCW Radar-based Vital Signs Monitoring under Practical Scenarios", The 20th IEEE International Conference on Dependable, Autonomic & Secure Computing (DASC), pp. 1-6, September 2022.

[21] Texas Instruments, "IWR6843ISK-ODS", Accessed on: January 2023. [Online]. Available: https:// www.ti.com/tool/IWR6843ISK-ODS.

[22] Texas Instruments. "DCA1000EVM". Accessed on January 2023. [Online]. Available: https:// www.ti.com/tool/DCA1000EVM.

[23] G. Sacco, E. Piuzzi, E. Pittella, and S. Pisa, "An FMCW Radar for Localization and Vital Signs Measurement for Different Chest Orientations", Sensors, vol. 20, issue. 12, June 2020.

[24] M. Mercuri, P. Russo, M. Glassee, I. Castro, E. Greef, M. Rykunov, M. Bauduin, A. Bourdoux, I. Ocket, F. Crupi, and T. Torfs, "Automatic Radar-based 2-D Localization Exploiting Vital Signs Signatures", Scientific Reports, vol. 12, no. 1, p. 7651, May 2022.

[25] Y. Jhaung, Y. Lin, C. Zha, J. Leu, and M. Koppen, "Implementing a Hand Gesture Recognition System Based on Range-Doppler Map", Sensors, vol. 22, issue. 11, June 2022.

[26] Texas Instruments, "MMWAVE-STUDIO", Accessed on January 2023. [Online]. Available: https:// www.ti.com/tool/MMWAVE-STUDI.