Enhance micro-Doppler signatures-based human activity classification accuracy of FMCW radar using the threshold method

Nguyen Ngoc Binh[1], Pham Minh Nghia[1*], Phan Huy Anh[2], Pham Hoang Hung[3]

[1]Faculty of Radio-Electronics Engineering, Le Quy Don Technical University, 236 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam;
[2]Academy of Military Science and Technology, 17 Hoang Sam, Cau Giay, Hanoi, Vietnam;
[3]Institute of Simulation Technology, Le Quy Don Technical University, 236 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.
*Corresponding author: nghiapm2018@mta.edu.vn
Received: 05 Feb. 2024; Revised 18 Mar. 2024; Accepted 10 May 2024; Published 20 May 2024.
DOI: https://doi.org/10.54939/1859-1043.j.mst.95.2024.20-28

ABSTRACT

Nowadays, radar-based human activity classification is being widely adopted in healthcare systems due to its benefits in terms of personal privacy compliance, non-contact sensing, and being unaffected by weather conditions. This study proposes a threshold method in the pre-processing stage to improve human activity classification accuracy by determining the region of meaningful information (RMI) on the spectrogram. Initially, a mask function, which is created by a certain threshold value, is applied to the input spectrogram to highlight the RMI from the micro-Doppler (m-D) signatures. Only the highlighted RMI on the spectrogram is retained as input to the classifiers. Then, five Convolutional Neural Networks (CNNs) of varying complexity are employed to extract features, identify activities, and assess the effectiveness of the suggested approach. The experimental results demonstrate that the suggested approach has enhanced classification accuracy by up to 11% when compared to the original unprocessed dataset.

Keywords: Micro-Doppler signature; Convolutional Neural Network; Human Activity Classification.

1. INTRODUCTION

Recently, the detection and accurate identification of human activities have attracted a lot of attention, especially in the healthcare systems [1, 2]. Monitoring and classification of such actions can be achieved by the use of wearable devices [3], cameras, and advanced sensors [4]. Although wearable gadgets like smartphones or smartwatches provide significant advantages, they can be cumbersome for consumers to wear constantly. Vision-based gadgets may pose personal privacy concerns [5] and cannot operate in dark or hazy conditions. Therefore, a radar-based non-wearable device is an effective option for non-contact tracking of human movements in environments where visibility is limited.

Traditionally, radar-based human activity classification is mainly based on the m-D signatures, which are defined as the micromotions of swinging limbs, turning the body, and stretching versus the translational motion of the body [6]. A unique m-D signature that corresponds with a certain activity, as depicted in spectrogram images, has been used and extensively studied in the context of the classification of daily activities [7]. Du et al. [8] constructed a range-velocity-time 3-D model that illustrates the movements of multiple people in a multi-target situation by merging the m-D signatures with range information. In [9], a separation approach was developed to separate the m-D components from the multi-target m-D signatures. Then, a CNN with a residual dense structure was used to classify separated components. In [10], the ECM-Th-STFT separation method was used to separate the m-D signatures corresponding to the limbs from the Doppler signal of the torso to increase the accuracy of identifying daily human activities.

Moreover, the human activity classification techniques rely on both deep learning and statistical learning, which have been strongly developed recently. In [11], two conventional classifiers (Support Vector Machine (SVM), k-Nearest Neighbours (KNN)) and a well-known CNN (GoogleNet) were used to recognize activities based on m-D signatures collected by a frequency-modulated continuous wave (FMCW) radar. Although the above models can detect and classify human actions, their false alarm rates are still high. Specifically, the error rates of these methods are 21.75%, 22.85%, and 25.3%, respectively. In [12], various pre-trained models, such as AlexNet, VGGNet, and a custom-designed CNN, were explored for extracting and categorizing human activities. VGGNet, with its deep network structure, achieved the highest accuracy of 95%. However, this came at the cost of increased processing time and complexity. Despite CNNs having more layers and deeper structures that can enhance recognition accuracy, they necessitate substantial data and processing costs. To enhance the right identification rate of CNNs without requiring more input datasets and while keeping suitable computational expenses, this study proposes a threshold method in the pre-processing stage to improve human activity classification accuracy by determining the region of meaningful information (RMI) instead of all information regions of the spectrogram. Initially, a mask function, which is created by a certain threshold value, is applied to the input spectrogram to highlight the RMI from the m-D signatures. Next, only the highlighted RMI on the spectrogram is retained as input to the classifiers. Then, five CNNs of varying complexity are employed to extract features, identify activities, and assess the effectiveness of the suggested approach. The experimental results demonstrate that the suggested approach has enhanced classification accuracy by up to 11% when compared to the initial unprocessed dataset.

The rest of this paper has the following structure: Section 2 describes FMCW radar and the dataset preparation. The proposed approach is presented in section 3. Experimental and comparative results are discussed in section 4. The last section is the conclusion.

2. FMCW RADAR AND DATASET PREPARATION

2.1. Overview of the FMCW radar

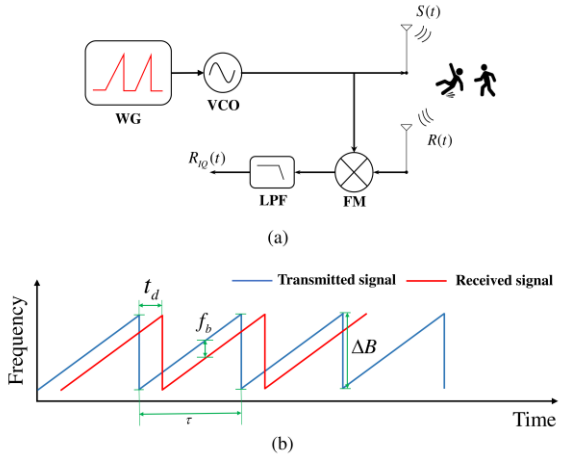


Figure 1. *Overview of the FMCW radar: (a) The FMCW radar block diagram, (b) FMCW sawtooth waveform.*

The block diagram of a basic FMCW radar system consisting of one transmitter and one receiver is detailed in Fig. 1a. FMCW transmits a sinusoid with linearly modulated frequencies over time. There are various varieties of linear frequency modulation, including sawtooth, triangle, or linear segment, of which the most basic and widely used is sawtooth modulation, which involves a succession of chirps over time (as seen in Fig. 1b).

The transmitted signal $S(t)$ is defined as [13]:

$$S(t) = A_t \cos(2\pi(f_0 t + \frac{\Delta B}{2\tau} t^2)) \quad (1)$$

where, ΔB is sweep bandwidth, τ is chirp duration, A_t is amplitude and f_0 is the carrier frequency. The reflected signal when hit the target is shown:

$$R(t) = A_r \cos(2\pi(f_0(t - t_d) + \frac{\Delta B}{2\tau}(t - t_d)^2) \quad (2)$$

where, $t_d = \frac{2D}{c}$ is the round-trip delay. D is the distance to the object and c is the speed of light.

The IF signal also known as the beat signal, following the mixer is the result of multiplying the receive and transmit signals and can be defined by:

$$R_{IQ}(t) = A_m \exp\{2\pi k t_d t - \pi k t_d^2 + 2\pi f_c t_d\} = A_m \exp\{\varphi(t)\} \quad (3)$$

in which, $k = \frac{\Delta B}{\tau}$, $A_m = \frac{A_t A_r}{2}$ and $\varphi(t)$ is a phase of the beat signal.

Next, the $R_{IQ}(t)$ signal is transformed into a 2D data matrix: the first dimension is for fast-time bins, and the second dimension is for slow-time bins. Then, the range-time matrix is obtained by applying the Fast Fourier Transform (FFT) along the fast-time direction of the matrix. A moving target indication (MTI) filter with a specific cutoff frequency is applied to the range-time matrix to suppress returns from the reflection of non-moving objects. The short-time Fourier transform (STFT) algorithm is employed by using a sequence of FFTs with short-sized windows and overlap sliding over the entire duration of the new range-time matrix to obtain a time-frequency spectrogram of the input signal. The STFT is given as follows [14]:

$$STFT_x(t, f) = \int_{-\infty}^{+\infty} x(t) w(t - \tau) \exp(-j\omega t) dt \quad (4)$$

where, $x(t)$ is the input signal and $w(t - \tau)$ is the window function. This study used window function (type: Gaussian, size: 128, and overlap: 95%). The absolute value of the complex STFT is defined as a spectrogram that contains m-D signatures and can be used to classify human activities.

2.2. Dataset preparation

The dataset was obtained using an FMCW radar sensor with a carrier frequency of 5.8 GHz (S-band) [15]. Each chirp has a bandwidth of 400 MHz and a duration of 1 millisecond. The raw signal is sampled at a sampling rate of 128 samples per chirp. It has a total of 1632 measurements divided into six groups of indoor actions, including walking, drinking, falling, sitting down, standing up, and picking.

Fig. 2 illustrates the spectrograms of the different activities in the dataset. It is evident that each activity has different unique m-D signatures, which are used to classify human activities. However, the RMI that matches the m-D signatures on the spectrogram in Fig. 2 only has a frequency change range of -100 to 100 Hz, where positive values mean movement away from the radar and negative values mean movement toward it. The rest of the spectrogram is mostly made up of background clutter and the region carrying redundant information. These obstacles would complicate the feature extraction process, leading to higher computational complexity and degrading feature robustness. Therefore, it is crucial to identify and retain only the RMI as input to the classifiers.

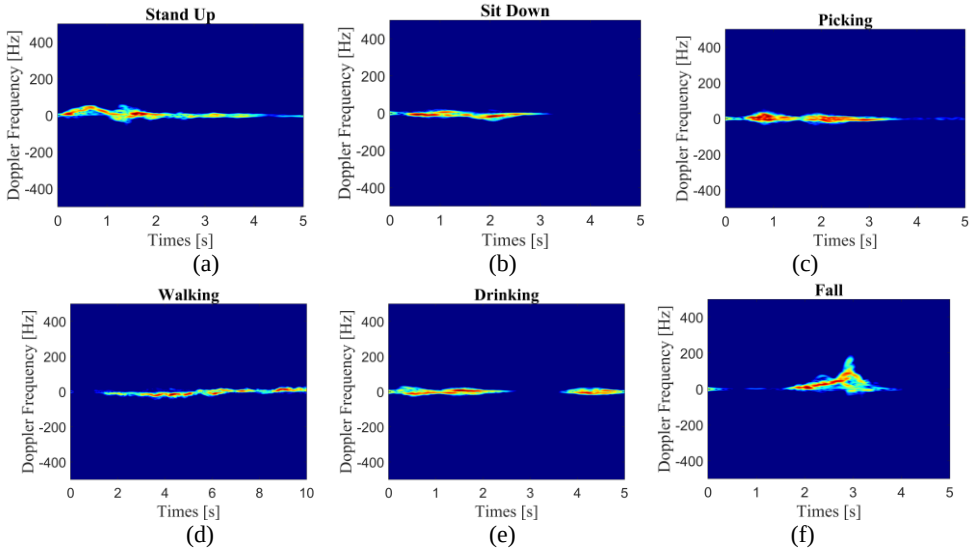


Figure 2. *The spectrogram of six actions in the original measurement dataset.*

3. THE PROPOSED APPROACH

The proposed threshold method is processed through three main steps.

Step 1: Design mask function.

$STFT_x(t,f)$ is calculated according to Eq. 4. Then, a masking function $T(t,f)$ is defined as:

$$T(t,f)=\begin{cases}1, & \left|STFT_x(t,f)\right|>ETh\\ 0, & others\end{cases} \tag{5}$$

with $Eth=Th\times avg\left\{\left|STFT_x(t,f)\right|\right\}$, where Th is the threshold that is determined empirically, $avg\{\cdot\}$ denotes the average value calculation operation, and $|\cdot|$ denotes absolute value.

Step 2: Calculate the masked spectrogram.

The obtained masked spectrogram as follows:

$$Spec_{x,masked}(t,f)=\left|STFT_x(t,f)\odot T(t,f)\right| \tag{6}$$

in here, $\odot$ represents element-wise multiplication. The power density matrix $\boldsymbol{P}_{x,masked}=Spec_{x,masked}(t,f)$, $\boldsymbol{P}_{x,masked}\in\mathbb{R}^{n\times m}$, $p(i,j)\in\boldsymbol{P}_{x,masked}$ is the point in the i^{th} frequency bin and j^{th} time bin in $\boldsymbol{P}_{x,masked}$.

Step 3: Determine the RMI on the masked spectrogram.

In each time bin index, cells with an energy value greater than 0 are selected and represent useful information on the spectrogram. $\boldsymbol{B}_{min}$ and $\boldsymbol{B}_{max}$ are the vectors, which contain the index of top cells and bottom cells in all the time bin index, respectively. Then, the top boundary (*tb*) index is the minimum value in $\boldsymbol{B}_{min}$ and the bottom boundary (*bb*) index is the maximum value in $\boldsymbol{B}_{max}$. Finally, the spectrogram contains only the RMI is represented in Eq. 7.

$$Spec_{RMI} = Spec_{x,masked}(t, f)[tb : bb] \tag{7}$$

Algorithm 1 summarizes the complete procedure of threshold method.

Algorithm 1 Threshold method

Input: $STFT_x(t, f)$

Output: $Spec_{RMI}$

Step 1: Design the masking function: $T(t, f)$

Step 2: Calculate the masked spectrogram: $\boldsymbol{P}_{x,masked} = Spec_{x,masked}(t, f)$, $\boldsymbol{P}_{x,masked} \in \mathbb{R}^{n \times m}$

Step 3: Determine the RMI on the masked spectrogram

3.1 Locating RMI, $\boldsymbol{L}_{RMI} = \{(i, j) \mid p(i, j) > 0,\ p(i, j) \in \boldsymbol{P}_{x,masked}\}$

3.2 **for** $j = 1; j \le m; j++$ **do**

3.2.1 $\boldsymbol{B}_{min}(j) = min\{i \mid (i, j) \in \boldsymbol{L}_{RMI}\}$

3.2.2 $\boldsymbol{B}_{max}(j) = max\{i \mid (i, j) \in \boldsymbol{L}_{RMI}\}$

3.3 **end for**

3.4 $tb = min\{\boldsymbol{B}_{min}\};\ bb = max\{\boldsymbol{B}_{max}\}$

3.5 $Spec_{RMI} = Spec_{x,masked}(t, f)[tb : bb]$;

4. EXPERIMENTAL AND COMPARISON RESULTS

In this section, five existing CNNs, including SimpleNet [12], ResNet [16], DINN [17], Dop-DenseNet [18], and MobileNet-V2 [19] are used to classify daily human activities on both datasets (original and processed).

The training and testing processes are carried out on a computer configured with an Intel(R) Core™ i5-12400F 2.5 GHz processor, 32 GB of RAM, and an RTX 3060Ti GPU. For training, a batch size of 32, an initial learning rate of 0.001 and 20 epochs are used with the optimizer's stochastic gradient descent. Five-fold cross-validation was used to evaluate the improvement in classification accuracy of the processed dataset compared to the original unprocessed dataset.

4.1. Overview of existing models

The parameters of five reference models are shown in table 1 (M stands for million). ResNet and MobileNet-V2, which were published in 2016 and 2018, respectively, are two popular models among the five models being compared. These two models are extensively utilized in computer vision using transfer learning because of their ability to achieve high classification accuracy on the ImageNet dataset. The two remaining networks, DINN and Dop-DenseNet, are state-of-the-art networks introduced in 2022. They are customized CNNs inspired by dense connection, which provides the re-

usability of loss features due to forward propagation during training. However, DINN is used for human activity classification, while Dop-DenseNet is used for hand gesture recognition. Finally, SimpleNet is also a self-designed network with a simple structure consisting of only a few convolutional layers and activation functions.

Table 1. *The parameters of five reference models.*

Parameters	Models				
	ResNet	Dop-DenseNet	MobileNet-V2	DINN	SimpleNet
Layers	71	49	154	129	14
Learned Parameters	11.1M	1.7M	2.2M	1.2M	14.6M

4.2. Select the threshold value

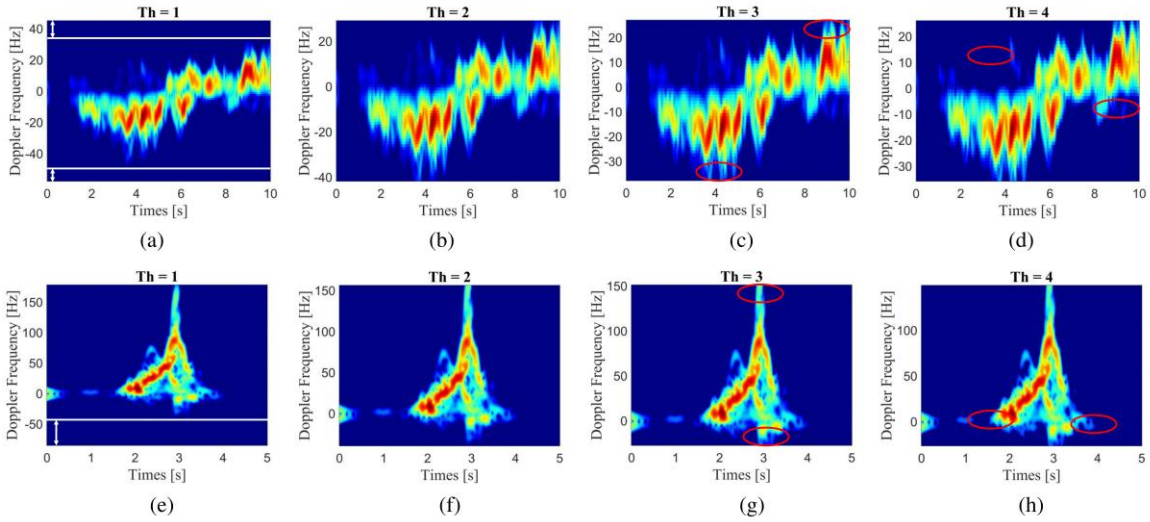


Figure 3. *The spectrogram images of walking and falling with different threshold values. Fig. 3 (a-d) shows the walking action. Fig. 3 (e-h) show the fall action.*

The threshold values, *Th*, selected in this experiment are 1, 2, 3, and 4, respectively. Fig. 3 illustrates the spectrogram images of walking and falling with various threshold values. Different threshold values produce different m-D signatures in the spectrogram. Fig. 3a,e displays the spectrogram with a threshold value of 1. Although the m-D signatures are visualized and more detailed than the unprocessed spectrogram (Fig. 2d,f), it still had some regions that did not contain meaningful information (the region outside the white border). Fig. 3(b-d) and Fig. 3(f-h) display the spectrograms processed with threshold values of 2, 3, and 4, respectively. Obviously, the region carrying redundant information has been eliminated, leaving only the region containing the interested m-D signatures in the images. However, when *Th* levels are higher, the consequence is that m-D signatures are more filtered out. Indeed, m-D signatures in the circled red areas in Fig. 3c,d,g,h will be removed in comparison to Fig. 3b,f. As a result, with a threshold value of 2, the spectrogram achieves the best balance between removing regions that contain

unmeaningful information and keeping essential m-D signatures that help improve the classification accuracy of various human activities.

Table 2. *Average Classification Accuracy (%) with varied threshold values.*

Models	Threshold values			
	Th = 1	*Th = 2*	*Th = 3*	*Th = 4*
SimpleNet	71.3	**77.7**	75.2	74
ResNet	90.5	**92.7**	92.1	91.7
DINN	89.3	**90.8**	90.2	88.7
Dop-DenseNet	89.6	**91.4**	90.8	89.9
MobileNet-V2	90.8	**93.6**	92.4	91.7

Moreover, table 2 shows the average classification accuracy of five CNN models with varying threshold values. The results demonstrate that the models achieve the highest classification accuracy with the selected threshold value of *Th = 2*. As a result, the chosen threshold value for the proposed threshold technique is 2.

4.3. The comparison results

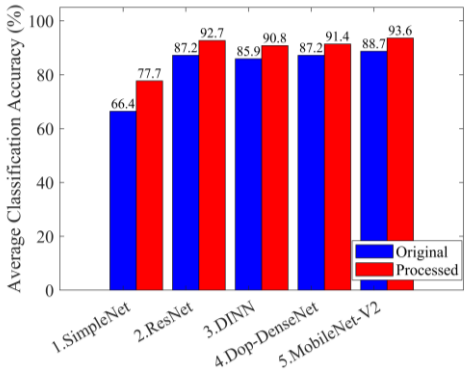
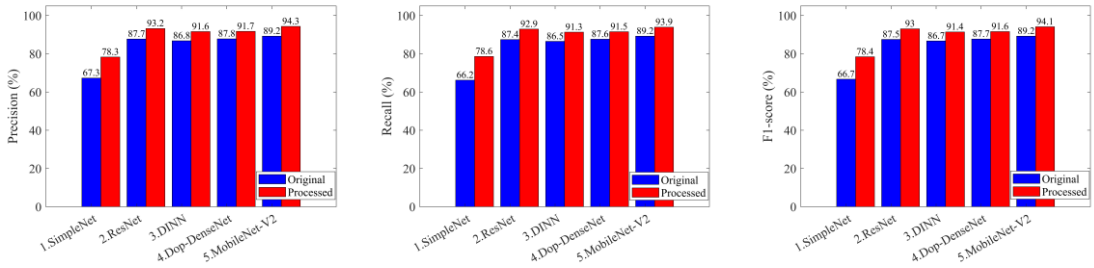


Figure 4. *Performance comparison of different classifiers.*

The results of evaluating the performance of the suggested threshold method performed on five existing CNNs are shown in Fig. 4. The SimpleNet achieves the lowest accuracy on the original unprocessed dataset, just about 66.4%, due to its simple structure. However, the accuracy of SimpleNet enhanced up to 77.7% (an increase of more than 11%) for the processed dataset. Moreover, the accuracy of the rest four networks also improved by about 5% compared to the original unprocessed dataset. This clearly demonstrates the effectiveness of the proposed method.



(a) Presision (b) Recall (d) F1-score

Figure 5. *Three metrics: precision, recall, and F1-socer of different classifiers.*

In addition, three additional important metrics: precision, recall, and F1-score, are also employed to evaluate the proposed approach effectiveness by the above models (Fig. 5). The results indicate the effectiveness of the proposed method.

5. CONCLUSIONS

In this study, the proposed threshold method is used in the pre-processing stage to improve the classification accuracy of daily human activities. The threshold values were established through experimentation to determine the interested RMI region. The experimental results demonstrate that the processed dataset by the proposed threshold method with a threshold value of 2 achieves the best balance between removing regions that contain unmeaningful information and keeping essential m-D signatures. As a result, it helps enhance classification accuracy by about 5% to 11% compared to the original unprocessed dataset.

REFERENCES

[1]. H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, *"Activities recognition and fall detection in continuous data streams using radar sensor,"* in 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), vol. 1, pp. 1-4, (2019).

[2]. J. Le Kernec et al., *"Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms,"* IEEE Signal Processing Magazine, vol. 36, no. 4, pp. 29-41, (2019).

[3]. T. R. Bennett, J. Wu, N. Kehtarnavaz, and R. Jafari, *"Inertial measurement unit-based wearable computers for assisted living applications: A signal processing perspective,"* IEEE Signal Processing Magazine, vol. 33, no. 2, pp. 28-35, (2016).

[4]. D. T. Nguyen, H. G. Hong, K. W. Kim, and K. R. Park, *"Person recognition system based on a combination of body images from visible light and thermal cameras,"* Sensors, vol. 17, no. 3, p. 605, (2017).

[5]. R. Igual, C. Medrano, and I. Plaza, *"Challenges, issues and trends in fall detection systems,"* Biomedical engineering online, vol. 12, no. 1, p. 66, (2013).

[6]. V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, *"Micro-Doppler effect in radar: phenomenon, model, and simulation study,"* IEEE Transactions on Aerospace and electronic systems, vol. 42, no. 1, pp. 2-21, (2006).

[7]. Y. Kim and H. Ling, *"Human activity classification based on micro-Doppler signatures using a support vector machine,"* IEEE transactions on geoscience and remote sensing, vol. 47, no. 5, pp. 1328-1337, (2009).

[8]. H. Du, T. Jin, M. Li, Y. Song, and Y. Dai, *"Detection of multi-people micro-motions based on range--velocity--time points,"* Electronics Letters, vol. 55, no. 23, pp. 1247-1249, (2019).

[9]. X. Huang, J. Ding, D. Liang, and L. Wen, *"Multi-person recognition using separated micro-Doppler signatures,"* IEEE Sensors Journal, vol. 20, no. 12, pp. 6605-6611, (2020).

[10]. N. Nguyen, T. Nguyen, M. Pham, and Q. Tran, *"Improving Human Activity Classification Based on Micro-Doppler Signatures Separation of FMCW Radar,"* in 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS), pp. 454-459, (2023).

[11]. S. A. Shah and F. Fioranelli, *"Human activity recognition: Preliminary results for dataset portability using FMCW radar,"* in 2019 international radar conference (RADAR), pp. 1-4, (2019).

[12]. F. A. Jibrin, A. Abdulaziz, A. S. Muhammad, A. D. Usman, and Y. Jibril, *"Indoor Human Activity Classification Based on FMCW Radar Micro-Doppler Signatures and Deep-Learning Networks,"* in 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), pp. 1-5, (2021).

[13]. V. Winkler, *"Range Doppler detection for automotive FMCW radars,"* in 2007 European Radar Conference, pp. 166-169, (2007).

[14]. J. Allen, *"Short term spectral analysis, synthesis, and modification by discrete Fourier transform,"* IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 3, pp. 235-238, (1977).

[15]. D. F. Fioranelli, D. S. A. Shah, H. Li1, A. Shrestha, D. S. Yang, and D. J. L. Kernec, *"Radar sensing for healthcare: Associate editor francesco fioranelli on the applications of radar in monitoring vital signs and recognising human activity patterns,"* Electronics Letters, vol. 55, no. 19, pp. 1022-1024, (2019).

[16]. K. He, X. Zhang, S. Ren, and J. Sun, *"Deep residual learning for image recognition,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, (2016).

[17]. N. Nguyen, M. Pham, V. Le, D. DuongQuoc, and V.-S. Doan, *"Micro-Doppler signatures based human activity classification using Dense-Inception Neural Network,"* in 2022 International Conference on Advanced Technologies for Communications (ATC), pp. 268-273, (2022).

[18]. H. Le, V.-P. Hoang, and others, *"Dop-DenseNet: Densely Convolutional Neural Network-Based Gesture Recognition Using a Micro-Doppler Radar,"* Journal of Electromagnetic Engineering and Science, vol. 22, no. 3, pp. 335-343, (2022).

[19]. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, *"Mobilenetv2: Inverted residuals and linear bottlenecks,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520, (2018).

TÓM TẮT

Nâng cao độ chính xác phân loại hoạt động của con người dựa trên các dấu hiệu micro-Doppler của radar FMCW bằng phương pháp ngưỡng

Ngày nay, phân loại hoạt động người dựa trên radar đang được ứng dụng rộng rãi trong các hệ thống chăm sóc sức khỏe dựa vào các ưu điểm về quyền riêng tư, không tiếp xúc và không bị ảnh hưởng bởi điều kiện thời tiết. Nghiên cứu này đề xuất một phương pháp ngưỡng được sử dụng trong giai đoạn tiền xử lý để cải thiện độ chính xác phân loại hoạt động người bằng cách xác định vùng thông tin có ý nghĩa (RMI) trên biểu đồ phổ. Ban đầu, một hàm mặt nạ, được tạo bởi một giá trị ngưỡng nhất định, được áp dụng cho biểu đồ phổ đầu vào để làm nổi bật RMI từ dấu hiệu micro-Doppler (m-D). Chỉ vùng RMI được đánh dấu trên biểu đồ phổ được giữ lại làm đầu vào cho các bộ phân loại. Sau đó, năm mạng nơron tích chập (CNNs) có độ phức tạp khác nhau được sử dụng để trích xuất các đặc trưng, phân loại hoạt động và đánh giá hiệu quả của phương pháp đề xuất. Kết quả thực nghiệm chứng minh rằng, phương pháp đề xuất đã nâng cao độ chính xác phân loại lên tới 11% khi so sánh với tập dữ liệu chưa được xử lý ban đầu.

Từ khoá: Dấu hiệu micro-Doppler; Mạng nơron tích chập; Phân loại hoạt động người.