World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:16, No:8, 2022

Human Fall Detection by FMCW Radar Based on Time-Varying Range-Doppler Features

Xiang Yu, Chuntao Feng, Lu Yang, Meiyang Song, Wenhao Zhou

Abstract**—**The existing two-dimensional micro-Doppler features extraction ignores the correlation information between the spatial and temporal dimension features. For the range-Doppler map, the time dimension is introduced, and a frequency modulation continuous wave (FMCW) radar human fall detection algorithm based on time-varying range-Doppler features is proposed. Firstly, the range-Doppler sequence maps are generated from the echo signals of the continuous motion of the human body collected by the radar. Then the three-dimensional data cube composed of multiple frames of range-Doppler maps is input into the three-dimensional Convolutional Neural Network (3D CNN). The spatial and temporal features of time-varying range-Doppler are extracted by the convolution layer and pool layer at the same time. Finally, the extracted spatial and temporal features are input into the fully connected layer for classification. The experimental results show that the proposed fall detection algorithm has a detection accuracy of 95.66%.

Keywords**—**FMCW radar, fall detection, 3D CNN, time-varying range-Doppler features.

I. Introduction

In recent years, the aging of the global population is becoming increasingly serious, and falls are the second leading cause of accidental or non-accidental injuries among the elderly. Fall detection has gradually become a research hotspot in the medical field. The World Health Organization (WHO) reports that 42% of the elderly over 70 years old fall at least once a year [1]. Therefore, it is of great significance to accurately detect falls of the elderly for their personal safety.

Fall detection using FMCW radar mainly includes three parts: radar echo signal processing, feature extraction, and classification. The key to fall detection is to obtain micro-Doppler features corresponding to human movements from radar sampling data. At present, traditional machine learning and deep learning are mainly used for radar fall detection. The machine learning methods mainly use statistical theory to extract shallow features from radar echo signals and then select classifiers according to the extracted feature information. Support Vector Machine (SVM) is commonly used as a classifier. Principal Component Analysis (PCA) was used to extract spatial features from range-doppler maps of human movements, and then the extracted features are classified using SVM [2]. Micro-Doppler features are extracted by constructing micro-Doppler maps frame by frame by projecting the distance Doppler map to the velocity dimension, and finally classifying them with a Bayesian hyperparametric optimized SVM [3]. The above traditional feature extraction method requires a lot of manual parameter adjustment, and the system model has some problems such as poor scalability. The rise of deep learning has attracted widespread attention in the field of radar human fall detection. Deep learning uses CNN to extract the deep features of signals. In [4], the original micro-Doppler spectrogram generated by the human action radar echo signal was directly input into CNN for fall detection for the first time, but this method has time-frequency window limitations when extracting features from the micro-Doppler image. The size is difficult to control and cannot be flexibly changed. Larger windows will lead to low time resolution, and smaller windows will lead to low-frequency resolution. Bidirectional Long Short-Term Memory (Bi-LSTM) is used to learn time-series information of continuous actions for fall detection [5], but the detection accuracy of this method is only 91%, which needs to be improved. The fusion network of 1D CNN and LSTM learns the time-series signal features of falling actions and classifies them [6]. In this method, time series signals contained relatively single feature information, resulting in insufficient feature extraction and other problems.

Aiming at the problems of the above deep learning-based fall detection methods, this paper proposes a fall detection method for FMCW radar based on time-varying range-Doppler features. The range-Doppler map is obtained by the two-dimensional fast Fourier transform (FFT) of the human motion echo signals collected by the radar, and the data cube formed by the multi-frame range-Doppler maps corresponding to a single action is 3D CNN. 3D CNN extracts spatiotemporal features along three dimensions, and finally classifies them through a classifier to improve the detection accuracy by optimizing the performance of feature extraction.

II. Radar Echo Data Preprocessing

A. FMCW Radar Signal Model

FMCW radar collects movement information of the human body by transmitting linear frequency modulated continuous wave, frequency-modulated periodic signal is [7]:

$$X_T(t) = A_T e^{-j2\pi(f_c t+\int_0^t \frac{B}{T_c}\tau d\tau)} \tag{1}$$

where A_T represents the transmit signal power, f_c represents the center frequency, B represents the bandwidth and T_c is the continuous period of one chirp. The echo signal received by the radar is:

Xiang Yu, Chuntao Feng, Lu Yang, Meiyang Song, and Wenhao Zhou are with School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China (e-mail: 1066367135@qq.com).

$$X_R(t) = A_R e^{-j2\pi\left\{f_c(t-\Delta t)+\int_0^t\left[\frac{B(\tau-\Delta t)}{T_c}+\Delta f_d\right]d\tau\right\}} \quad (2)$$

where A_R, Δt and Δf_d represent the received signal power, time delay, and Doppler frequency shift, respectively. The intermediate frequency signal obtained by mixing the transmitted signal and the received signal is:

$$X_{IF}(t) = X_T(t)X_R(t) \approx A_T A_R e^{j2\pi(f_c\Delta t+(f_I-\Delta f_d)t)} \quad (3)$$

where f_I represents the frequency of the intermediate frequency signal at a time,

$$f_I = \frac{B}{T_c}\Delta t \quad (4)$$

Sampling the intermediate frequency (IF) signal, the FMCW radar echo data can be expressed in the form of a radar sampling data matrix, and the mth sampling data of IF signal in the nth repeated period are:

$$S_I(n,m) = A_I e^{j2\pi\frac{f_I(n)-\Delta f_d(m)}{f_s}} \quad (5)$$

where A_I represents the amplitude of the IF signal and f_s represents the sampling rate.

B. Range-Doppler Map

Human motion modulates the radar signal, resulting in the Doppler effect, in which body movements produce a large amount of micro-Doppler effects. In order to extract the micro-Doppler feature information of human motion, a range-Doppler map is introduced, and the distance and velocity information corresponding to the scattering points of the human body can be obtained in the range-Doppler map.

By performing FFT along the fast time dimension (single chirp) and slow time dimension (between chirp and chirp), the range-Doppler maps corresponding to human movement can be obtained. When FFT is performed on radar data along the fast time dimension, the sweep period of a single chirp is very short, so the influence of Doppler frequency can be ignored. The range information of the target can be obtained by extracting the abscissa frequency corresponding to the spectral peak of each frame data.

$$f_{mB} \approx f_{sB} = \frac{2 f_I R}{C T_m} \quad (6)$$

$$R = \frac{C T_m}{2 f_c} \cdot f_{sB} \quad (7)$$

where f_{mB} and f_{sB} represent the IF signal frequency of the target in the moving and stationary states, respectively, and C is the speed of light. When FFT is performed on radar data along the slow time dimension, the frequency effect caused by the speed of the target cannot be ignored, so the Doppler frequency can be obtained in the slow time dimension.

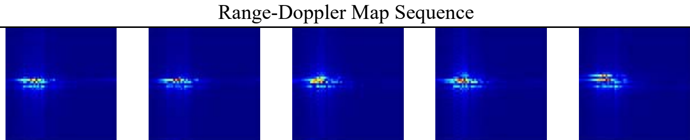
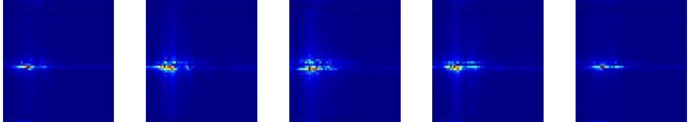
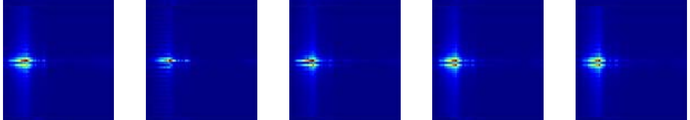
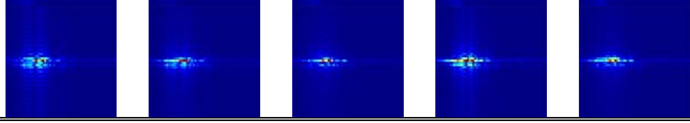
$$f_d = \frac{2 f_I v}{C} \quad (8)$$

Thus, the target velocity can be obtained as:

$$v = \frac{f_d C}{2 f_I} \quad (9)$$

In this paper, the range-Doppler sequences obtained from the publicly available radar human motion echo dataset from the University of Glasgow [8] are shown in Table I after passing through an MTI filter to remove static interference and then performing a 2D FFT. The radar equipment of this data set operates in C-band (5.8 GHz), with a bandwidth of 400 MHz and a duration of 1 ms of chirp signal.

TABLE I
RANGE-DOPPLER SEQUENCE

Action	Range-Doppler Map Sequence
falling	
sitting	
standing	
walking	

III. Fall Detection Based on Time-Varying Range-Doppler Features

Falls are events that evolve dynamically over time, and their corresponding range-Doppler sequences contain a large amount of motion information. The existing feature extraction based on a two-dimensional range-Doppler map adopts the method of extracting spatial features and temporal features respectively, ignoring the correlation information between the spatial and temporal dimension feature information. In response to this problem, this paper proposes a fall detection method for FMCW radar based on time-varying range-Doppler features, which uses the efficient feature extraction capability of 3D CNN to extract the time-varying features of actions from the time-varying range-Doppler map, thereby get richer feature information.

A. 3D CNN

3D CNN [9] adds a temporal dimension to 2D CNN, which can simultaneously extract spatial and temporal features. 3D CNN uses a three-dimensional convolution kernel to check the three-dimensional data sliding along the spatio-temporal dimension to extract the spatio-temporal features of the motion data. Multiple consecutive frame maps sequentially extract the time-series features through the convolution layer. Each feature map in the convolution layer is connected to the previous layer, to obtain the motion information of the target. 3D convolution is used to extract human motion micro-Doppler features, which not only preserves the timing information between time-varying range-Doppler map frames but also depicts the time-space correspondence.

B. Fall Detection Method Based on Time-Varying Range-Doppler Features

In order to automatically extract features from the time-varying range-Doppler maps, a 3D CNN consisting of convolutional layers and fully connected (FC) layers was constructed. Fig. 1 shows the proposed 3D CNN network model architecture, and Table II gives the kernel size and output shape of each layer. The key properties of this network model are the inclusion of different processing units of convolution, pooling, activation, and normalization.

Firstly, the multi-frame range Doppler map corresponding to a single action is input into the CNN, the first 3D convolutional layer c_1, 32 kernels $\{k_j^{c_1}\}_{j=1}^{32}$ of size 3 x 3 x 3 are convolved with an input image of stride 1, then a bias value is added and an activation function is applied to the output, resulting in a feature diagram Z^{c_1} with a depth of 32. The activation function adopts the activation function of Leaky Rectified Linear Unit (Leaky ReLU), and its calculation method is

$$f(x)=\begin{cases} x & x>0 \\ leak*x & x\le 0 \end{cases} \tag{10}$$

where *leak* represents a decimal greater than 0. Each unit in a convolutional layer is connected to a local block in the feature diagram of the previous layer through a set of convolution kernels. The feature extracted by the three-dimensional convolution kernel is still a three-dimensional data, and the feature extracted by the 3D CNN is specifically expressed as

$$v_{ij}^{xyz}=F\left(\sum_m\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}+b_{ij}\right) \tag{11}$$

where v_{ij}^{xyz} is the eigenvalue of the jth feature map pixel (x,y,z) of the ith layer, F is the activation function, m is the number of feature maps in the (i-1)th layer, P_i, Q_i and R_i are the spatial and temporal dimensions of the 3D convolution kernel of the ith layer, w_{ijm}^{pqr} is the weight of the convolution kernel connected at (p,q,r) of the mth feature map of the previous layer, b_{ij} is the bias difference of the jth feature map of the ith layer.

In order to obtain distortion-invariant features and remove redundant non-feature information, a non-overlapping 3D max-pooling (MP) p_1 of size 3 x 3 x 3 is used to perform pooling by selecting the maximum value of convolution features in neighboring neurons located in the previous convolutional layer. The specific sampling of the three-dimensional maximum downsampling in the region $S_1\times S_2\times S_3$ is

$$x_{abc}=\max_{0\le i\le S_1,0\le j\le S_2,0\le k\le S_3}(x_{a\times s+i,b\times t+j,c\times r+k}) \tag{12}$$

where x_{abc} is the three-dimensional sampling output value at point (a,b,c), $x_{a\times s+i,b\times t+j,c\times r+k}$ is the three-dimensional input value at point $(a\times s+i,b\times t+j,c\times r+k)$, and s, t, and r are the sliding steps in three directions, respectively.

Since the spatial size of the feature map decreases after the pooling operation, it is necessary to increase the depth by adding more filters. Given this, in the next convolutional layer c_2, the feature diagrams from the previous pooling layer $Z^{c_1p_1}$ are convolved with 64 kernels $\{k_j^{c_2}\}_{j=1}^{64}$ of size 3 x 3 x 3 and stride 1. A bias value is added and a Leaky ReLU activation function is applied to the output, resulting in a feature diagram Z^{c_2} of depth 64, and then a second MP layer p_2 of size 3 x 3 x 3 is applied. Convolutional layers capture low-level features, while higher layers extract high-level features by combining low-level features. Furthermore, the interaction between neurons in each convolutional layer is sparse, and there is a weight-sharing mechanism. In the proposed network, the convolutional layer is followed by two FC layers and an output layer for class prediction, the number of neurons in both FC layers is 512, and the output of the second convolutional layer $Z^{c_2p_2}$ is flattened into a 1D vector, resulting in the parameters being concatenated and used as input to the first FC layer. The neurons in each FC layer are fully connected to all neurons in the previous layer. Their activations can be computed by applying matrix multiplication and bias offset, followed by the Leaky ReLU activation function of the first two layers and the softmax function of the output layer. The softmax function is given by:

$$h^r=\frac{\exp(Z^r)}{\sum_{v=1}^{2}\exp(Z^v)} \quad r=1,2,3,4 \tag{13}$$

where Z^r is the rth score of the output layer, and h^r represents the output of the softmax function, that is, the probability of the predicted classes.

The entire network was trained by the back-propagation algorithm [10] to update weights iteratively and minimize them using cost functions. Adam optimizer is used in the learning process, which combines the characteristics of AdaGrad and RMSProp algorithm to enable the optimization algorithm to deal with sparse gradients [11] and has been proved to provide fast convergence speed. By optimizing the model, feature learning and classification can promote each other, and the learned features have stronger classification ability for the final classification task.

In the proposed 3D CNN network model, hyperparameter selection is accomplished by using different hyperparameter adjustment models and selecting models according to verification accuracy, namely grid search optimization technology. The hyperparameters of the proposed network are the learning rate, which determines the magnitude of the weight change in the event of misclassification during the training phase, and the dropout regularization factor, which limits the network's adaptation to training data and thus avoids overfitting. In order to train a more efficient 3D CNN model, a batch normalization (BN) layer is added after the convolutional layer. The BN layer uses the mean and standard deviation of the mini-batch to continuously optimize the output of the network, making the output value of the entire network more stable and enhancing the generalization and representation capabilities of the network model.

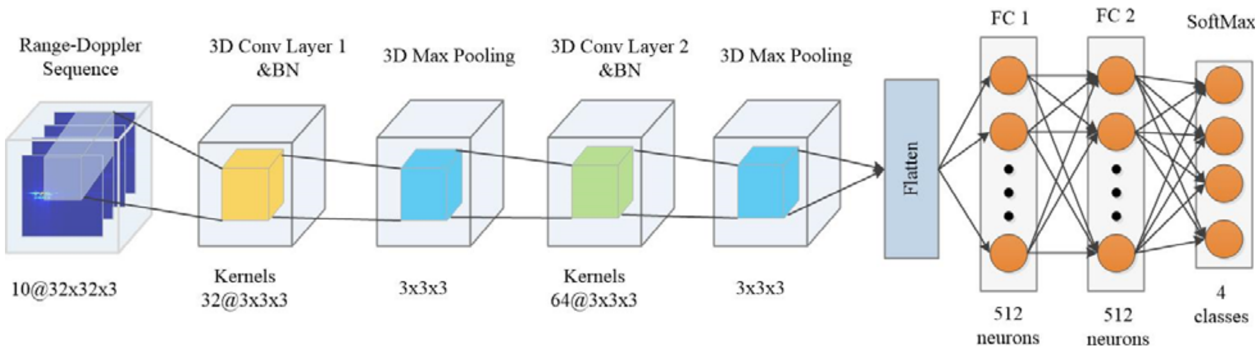


Fig. 1 Fall detection network model diagram

TABLE II
3D CNN NETWORK PARAMETERS

Layer	Parameter	Value
Input Layer	Input Size	32 x 32 x 10 x 3
3D Conv Layer1	Kernel Size	3 x 3 x 3
	Of filters	32
Max Pooling Layer	Kernel Size	3 x 3 x 3
3D Conv Layer2	Kernel Size	3 x 3 x 3
	Of filters	64
Max Pooling Layer	Kernel Size	3 x 3 x 3
FC Layer1	Of neurons	512
FC Layer2	Of neurons	512
FC Layer3	Of neurons	4
	Classifier	Softmax
Output Layer	Of outputs	4

IV. EXPERIMENTAL VERIFICATION

A. Evaluation Metric

To evaluate the performance of the proposed fall detection system, five evaluation metrics, precision, precision, recall, F1-score, and confusion matrix, are used, which are defined as:

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{14}$$

$$precision = \frac{TP}{TP + FP} \tag{15}$$

$$recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 - score = \frac{2(precision \times recall)}{precision + recall} \tag{17}$$

TP: true positive (correct classification), TN: true negative (correct rejection), FP: false positive (misclassification), and FN: false negative (missing detection). The confusion matrix, consisting of the four terms defined above (TP, TN, FP and FN) in tabular form, shows the performance indicators for individual activities.

B. 3D CNN Model Training and Testing

The radar human motion dataset published by the University of Glasgow is composed of 83 volunteers of different ages, heights, and weights, which indicate that the dataset has good generalizability. The actions in this paper include four human actions: standing, sitting, walking, and falling. The radar echo data are preprocessed to obtain the range-Doppler sequence maps corresponding to different actions, as shown in Table I in Section II. For a single action, 10 frames of range-Doppler maps are used as a sample, the corresponding time step is 1 s, and the number of samples corresponding to each action in the network model training is 1050, a total of 4200 groups of samples. The range-Doppler sequence maps of different actions are uniformly processed into a size of 32 x 32, and randomly divided according to action categories to obtain five subsets with similar numbers. Each training data set is divided into four subsets as the training set, the other subset is used as the test set, and 5-fold cross-validation is performed.

The network model training is based on the Tensorflow deep learning framework using python language, using the Adam optimizer, the initial learning rate is set to 0.0001, the loss function uses the cross-entropy loss function, the batch size is set to 32, and the epoch of each training iteration is set to 200.

C. Fall Detection Results

Table III and Fig. 2 show the results of four types of action tests using the 3D CNN model. Table III is the confusion matrix for four types of action classification. The column elements represent the real action categories, the row elements represent the predicted categories of each action, and the values in the matrix represent the predicted percentages. From the confusion matrix, it can be seen that the detection accuracy of the fall detection algorithm proposed in this paper can reach 95.66%. Except for falls, the detection accuracy of other actions is above 80%. Fig. 2 is the ROC curve diagram of four types of action detection. The area in the figure represents the area covered by the ROC curve, and the larger the area covered indicates the better classification effect of the model. As can be seen from Fig. 2, the ROC curve corresponding to each action is close to the upper left corner of the coordinate axis in the figure. The areas of the macro and micro ROC curves are 96% and 97%, respectively, and the area of the four classes of actions reaches more than 95%, indicating that the model has good detection and classification ability.

TABLE III
CONFUSION MATRIX

Action	Falling	Standing	Sitting	Walking
Falling	95.66%	3.21%	0.57%	0.56%
Standing	0.50%	82.00%	11.50%	6.00%
Sitting	0.00%	14.43%	81.96%	3.61%
Walking	0.42%	0.00%	0.83%	98.75%

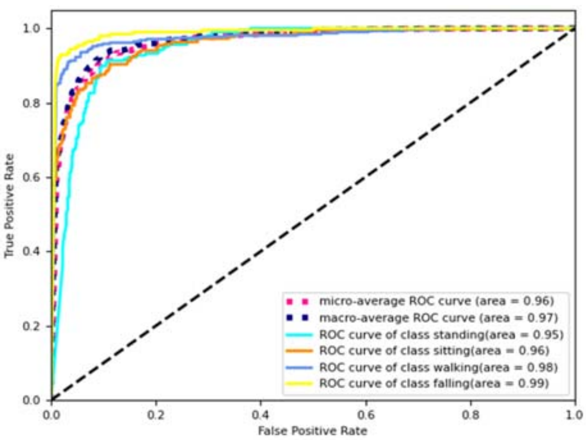


Fig. 2 ROC curve

For time-varying range-Doppler features, time is one of the dimensions we can control. In order to study the effect of action duration on the accuracy of model detection, we trained and tested actions with durations of 2.5 s and 5 s respectively, and obtained the test results of 2.5 s as shown in Table IV and Fig. 3. The accuracy of fall detection is 95.98%. The 5 s test results are shown in Table V and Fig. 4, and the fall detection accuracy is 97.37%. Comparing the detection results with durations of 1 s, 2.5 s, and 5 s, it can be seen from Table VI that the longer the action lasts, the more feature information it contains, and the higher the detection accuracy of the model.

TABLE IV
CONFUSION MATRIX CORRESPONDING TO 2.5 S

Action	Falling	Standing	Sitting	Walking
Falling	95.98%	2.19%	0.15%	1.68%
Standing	0.33%	91.67%	8.00%	0.00%
Sitting	0.00%	4.93%	92.54%	2.53%
Walking	0.23%	0.45%	0.19%	99.13%

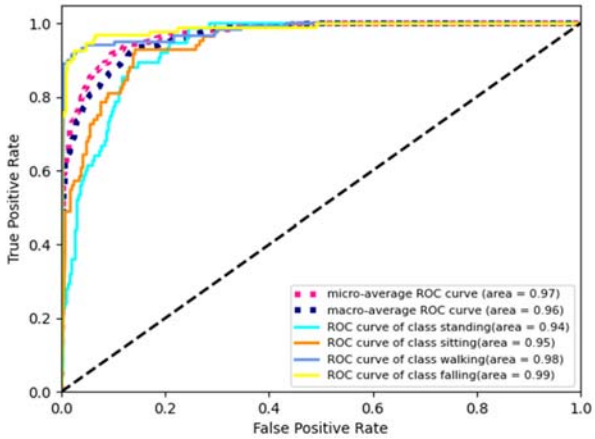


Fig. 3 ROC curve corresponding to 2.5 s

TABLE V
CONFUSION MATRIX CORRESPONDING TO 5 S

Action	Falling	Standing	Sitting	Walking
Falling	97.37%	0.00%	0.00%	2.63%
Standing	0.30%	93.02%	4.35%	2.33%
Sitting	0.00%	5.74%	93.26%	1.00%
Walking	0.00%	0.00%	0.00%	100%

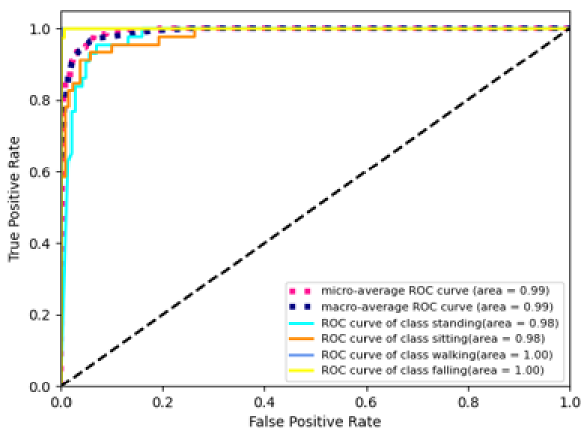


Fig. 4 ROC curve corresponding to 5 s

TABLE VI
EVALUATION INDICATORS CORRESPONDING TO DIFFERENT TIME PERIODS

	Accuracy	Precision	Recall	F1-score
1 s	95.66%	0.9905	0.9566	0.9733
2.5 s	95.98%	0.9942	0.9598	0.9767
5 s	97.37%	0.9969	0.9737	0.9852

D. Algorithm Comparison

To further analyze the performance of the proposed method, it is compared with fall detection algorithms of different network frameworks. The results in Table VII show that the detection accuracy of the fall detection method proposed in this paper is about 4 percentage points higher than that of other methods, and it can better detect falls. In addition, the network model proposed in this paper has a simple framework and is composed of the most basic CNN. It can achieve better fall detection by only two layers of three-dimensional convolution and pooling. Compared with other complex network structures, it has a smaller workload.

TABLE VII
ALGORITHM COMPARISON TABLE

Model type	Accuracy	Precision	Recall	F1-score
CNN	91.35%	0.9909	0.9135	0.9506
Bi-LSTM [5]	91%	0.9546	0.9100	0.9318
CNN+LSTM [6]	92.15%%	0.8300	0.9600	0.8903
3D CNN	95.66%	0.9905	0.9566	0.9733

V. CONCLUSION

In order to make full use of the correlation information between the spatial and temporal features of micro-Doppler and improve the accuracy of fall detection, this paper introduces the time dimension information into the range-Doppler map of human motion radar and proposes a human fall detection method based on the time-varying range-Doppler feature of FMCW radar. The experimental results show that the fall detection accuracy of the method proposed in this paper is 95.66%. Subsequent improvements will be made in the complexity of the algorithm and the real-time performance of the network model.

REFERENCES

[1] I. Alujaim, I. Park and Y. Kim, "Human Motion Detection Using Planar Array FMCW Radar Through 3D Point Clouds," 2020 14th European Conference on Antennas and Propagation (EuCAP), 2020, pp. 1-3.
[2] Liubing Jiang, Guangmeng Wei, and Li Che, “Radar human action recognition method based on convolutional neural network,” in Computer Applications and Software, vol. 36, no. 11, pp. 168-174+234, 2019.
[3] Chenxu Ding, Yuanhui Zhang, and Zhetao Sun, “Human complex motion recognition based on FMCW radar,” in Radar Science and Technology, vol. 18, no. 6, pp. 584-590, Jan. 2020.
[4] Y. Kim and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 8-12, Jan. 2016.
[5] A. Shrestha, H. Li, J. Le Kernec and F. Fioranelli, "Continuous Human Activity Classification From FMCW Radar With Bi-LSTM Networks," in IEEE Sensors Journal, vol. 20, no. 22, pp. 13607-13619, 2020.
[6] J. Maitre, K. Bouchard and S. Gaboury, "Fall Detection with UWB Radars and CNN-LSTM Architecture," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 4, pp. 1273-1283, April 2021.
[7] Lili Zhang, Bo Liu, Lele Qu and Yuxuan Liu, “FMCW radar human action recognition based on feature fusion convolutional neural network,” in Telecommunications Technology, vol. 62, no. 2, pp. 147-154, 2022.
[8] F. Fioranelli, Shah S A, Li H, “Radar sensing for healthcare,” in Electronics Letters, vol. 55, no. 19, pp. 1022-1024, 2019.
[9] T. Wang, J. Li and M. Zhang, “An enhanced 3DCNN-ConvLSTM for spatiotemporal multimedia data analysis,” in Concurrency and Computation: Practice and Experience, vol. 33, no. 2, pp. e5302, 2021.
[10] Hinton G E, Osindero S, Teh Y W. A fast-learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
[11] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv, 2014, 14(12):1-13.