Advanced Human Activity Recognition through Data Augmentation and Feature Concatenation of Micro-Doppler Signatures

Original Scientific Paper

Djazila Souhila Korti
Belhadj Bouchaib University of Ain-Temouchent
Smart Structures Laboratory (SSL)
Faculty of Technology, Department of Telecommunication
Ain-Temouchent, Algeria
souhila.korti@univ-temouchent.edu.dz

Zohra Slimane
Abou Bekr Belkaid University of Tlemcen
Faculty of Technology, Department of Telecommunication
Tlemcen, Algeria
zoh_slimani@yahoo.fr

Abstract *–Developing accurate classification models for radar-based Human Activity Recognition (HAR), capable of solving real-world problems, depends heavily on the amount of available data. In this paper, we propose a simple, effective, and generalizable data augmentation strategy along with preprocessing for micro-Doppler signatures to enhance recognition performance. By leveraging the decomposition properties of the Discrete Wavelet Transform (DWT), new samples are generated with distinct characteristics that do not overlap with those of the original samples. The micro-Doppler signatures are projected onto the DWT space for the decomposition process using the Haar wavelet. The returned decomposition components are used in different configurations to generate new data. Three new samples are obtained from a single spectrogram, which increases the amount of training data without creating duplicates. Next, the augmented samples are processed using the Sobel filter. This step allows each sample to be expanded into three representations, including the gradient in the x-direction (Dx), y-direction (Dy), and both x- and y-directions (Dxy). These representations are used as input for training a three-input convolutional neural network-long short-term memory support vector machine (CNN-LSTM-SVM) model. We have assessed the feasibility of our solution by evaluating it on three datasets containing micro-Doppler signatures of human activities, including Frequency Modulated Continuous Wave (FMCW) 77 GHz, FMCW 24 GHz, and Impulse Radio Ultra-Wide Band (IR-UWB) 10 GHz datasets. Several experiments have been carried out to evaluate the model's performance with the inclusion of additional samples. The model was trained from scratch only on the augmented samples and tested on the original samples. Our augmentation approach has been thoroughly evaluated using various metrics, including accuracy, precision, recall, and F1-score. The results demonstrate a substantial improvement in the recognition rate and effectively alleviate the overfitting effect. Accuracies of 96.47%, 94.27%, and 98.18% are obtained for the FMCW 77 GHz, FMCW 24 GHz, and IR-UWB 10 GHz datasets, respectively. The findings of the study demonstrate the utility of DWT to enrich micro-Doppler training samples to improve HAR performance. Furthermore, the processing step was found to be efficient in enhancing the classification accuracy, achieving 96.78%, 96.32%, and 100% for the FMCW 77 GHz, FMCW 24 GHz, and IR-UWB 10 GHz datasets, respectively.*

Keywords*: human activity recognition, radar, micro-Doppler signature, data augmentation, preprocessing, feature concatenation*

1. INTRODUCTION

With the emergence of IoT and sensing technologies, Human Activity Recognition (HAR) has gained significant attention and found applications in various fields [1, 2]. Numerous sensing technologies have been investigated for HAR, including video devices, wearable sensors, and radars [3]. The key features of radar technology, including non-invasiveness, privacy preservation, low energy consumption, and environmental insensitivity, have made this technology very popular for HAR [4]. Non-contact radar sensors can continuously detect and monitor human activities, including gait, falls, gestures, and activities of daily living [4].

Two commonly employed radar sensors for the detection of human activities are Impulse Radio Ultra-Wide Band (IR-UWB) [5] and Frequency Modulated Continuous Wave (FMCW) [6]. Radar echoes caused by electromagnetic signals reflected from different parts of the body contain valuable information about human motion, known as the micro-Doppler effect. The Micro-Doppler signatures encompass frequency components derived from the translational motion of the body or the vibration and rotation of its non-rigid parts. These distinctive features can be directly exploited to identify various human activities [7-9].

Due to the ability of deep feature self-learning, Deep Neural Networks (DNNs) have been successfully applied for HAR based on micro-Doppler signatures [10]. DNNs mainly focus on automatic feature learning, which neither requires manual intervention nor relies on prior knowledge. Recent studies have shown that hybrid models like CNN-LSTM [8, 11, 12] are very useful and can improve performance over individual networks in identifying human activities. Although DNNs have been shown to be effective for micro-Doppler recognition, their performance is sometimes lacking due to data sparsity [13, 14]. DNNs are known to be data-intensive, requiring large amounts of labeled training samples to obtain satisfactory results [15]. Unfortunately, the requirement for a large amount of radar data is difficult to meet. Data acquisition and annotation remain complex, expensive, and time-consuming tasks.

Data augmentation is an effective technique used to increase the size and diversity of the training set by generating new synthetic samples from the original data. Several approaches in computer vision have been proposed, including geometric transformations and photometric transformations [16, 17]. However, in some cases, radar data cannot be augmented by performing transformations as it might distort the signature patterns [18, 19]. Advanced approaches, including Generative Adversarial Network (GAN), have emerged as a popular class of modern deep learning models for synthetically generating image data [20]. Various works on GAN extensions, such as Deep Convolution GANs (DCGANs), Auxiliary Classifier GAN (AC-GAN), CycleGANs, and Progressively-Growing GANs [21, 22], have been adopted in the radar area and have shown a great ability to mimic complex real-world data. However, the fidelity of the generated data is not guaranteed. Moreover, training GANs is very challenging and requires a lot of effort for implementation. Another alternative workaround in situations where large training data are difficult to access is by using transfer learning (TF) and Domain Adaptation (DA) [23, 24]. These approaches involve training models on a large dataset and then fine-tuning their weights on a small target dataset. However, the required accuracy cannot be met only by directly transferring the features. The models may exhibit unpredictable performance when a mismatch occurs between the source and target training content.

This work aims to develop an automated HAR system that achieves both lower computational requirements and high classification accuracy with small micro-Doppler datasets, with two main objectives in mind. The first objective is to propose a simple and efficient strategy for generating micro-Doppler spectrograms used for training classification algorithms. The main contribution of this research lies in the creation of new samples derived from the original dataset with non-overlapping features. Our focus is on developing an augmentation process that ensures the preservation of essential features in the micro-Doppler signatures without introducing unintended distortions. Additionally, we strive to establish a robust and adaptable augmentation strategy that can be applied to various datasets, rather than relying on a specific one. In contrast to previous studies [13], our approach stands out through its exclusive utilization of image manipulation techniques directly applied to micro-Doppler signatures. Instead of resorting to conventional image transformations [25], we propose the application of Discrete Wavelet Transform (DWT) as a novel alternative. This unique approach sets our work apart and contributes to the advancement in the context of data augmentation strategies.

The original micro-Doppler spectrograms are projected onto the DWT subspace. From this projection, spectrograms are decomposed, resulting in different sub-band images. Next, the decomposition components returned by the DWT process are used in different configuration to generate new samples. From a single spectrogram, three samples are generated, increasing the number of training data without creating duplicates. This approach makes it possible to enhance human motion characteristics to improve the accuracy of the final classifier without adding prior knowledge or re-acquiring data. The second objective is to extend our previous work [26] based on Hand Gesture Recognition (HGR) to the HAR application. Our previous contribution consisted of extending a single sample to three representations by extracting low-level feature images using the Sobel filter.

The generated low-level feature images are then used as independent input for the model. We used a lightweight CNN-LSTM-SVM to classify hand gestures using IR-UWB. Compared to existing models, our approach provides simplicity with significantly high performance.

The remainder of the paper is organized as follows: Section 2 briefly reviews scholarly works related to data augmentation for Micro-Doppler signatures. Section 3 describes the proposed approach. Section 4 provides the evaluation datasets and implementation details. Section 5 presents the experimental results and comparative analysis. The discussion is presented in Section 6, and Section 7 concludes the paper.

2. RELATED WORKS

In order to significantly reduce the cost and effort of data measurement, several data augmentation strategies have been proposed. Researchers in the radar-based HAR field have attempted to use Transfer Learning (TL) and Domain Adaptation (DA) methods [27, 28]. These greatly reduce the dependence of models on large training samples, improve recognition accuracy, and convergence speed. Several DNN models pre-trained on the ImageNet dataset have been used to classify limited micro-Doppler signatures for human activities [29, 30]. Du *et al*. [29] presented a transfer-learned residual network to classify human activity based on micro-Doppler spectrograms. The performance of the model was evaluated on the CMU Mocap dataset. Du *et al*. [30] proposed using pretrained VGG-19 on the ImageNet dataset for the classification of micro-Doppler measurements. The experimental results demonstrate that transfer-learned VGG-19 outperforms the trained model from scratch and gives a reduction in the number of parameters and computing operations. However, the required accuracy cannot be met only by directly transferring the features. The characteristics of radar signatures differ widely from those of optical images.

The models may exhibit unpredictable performance when a mismatch occurs between the source and target training content. Another approach for dealing with low sample support is the simulation of micro-Doppler signatures using motion capture [31]. This approach enables generating a large number of micro-Doppler signatures with different postures and motion speeds of the human model. Moreover, different technical aspects of the radar are supported, such as frequency, angle, and location of the radar. However, the main drawback of this approach is that the generated data is too clean and perfect, while in real scenarios the data is affected by various environmental factors, sensor parameters, and target characteristics. For example, variations introduced by the surrounding environment such as obstructions caused by walls, objects, or movements that are not related to the target. The gap between the real world and the simulation can considerably deteriorates the performance of the models. To tackle this problem, GAN has been proposed as a means to generate highly realistic simulated images [32, 33]. An early effort at applying GAN to synthesize new micro-Doppler signatures was first proposed in [34] for walking gaits at different speeds. A similar approach has been proposed in [35] to generate different human actions, extended to other movements than simply walking. Gurbuz *et al.* [13] used a GAN to synthesize micro-Doppler signatures collected from three radars for cross-frequency training. Results show an increase in the overall classification accuracies. Erol *et al.* [36] used ACGAN to generate more diverse and crisp synthetic micro-Doppler signatures. The results showed the effectiveness of synthetic ACGAN data adapted to different detection locations and environments. Zhong *et al.* [37] proposed using DCGAN to achieve data augmentation of the micro-Doppler samples set. Experimental studies have shown that the combination of GAN and CNN can achieve effective recognition. However, the experimental simulation is set up in an ideal environment without other target interference. A major drawback of the synthetic data generated by GAN is that its fidelity is not guaranteed. Although GAN has demonstrated its ability to generate realistic synthetic data, their fidelity is not guaranteed. Micro-Doppler signatures generated by GAN have been found to correspond to kinematically impossible behaviors or to classes of motions different from those expected [38, 39].

3. PROPOSED METHODOLOGY

The proposed methodology comprises three steps: data augmentation, data preprocessing, and classification. In the data augmentation step, DWT is applied to the original samples for the decomposition process. The obtained components are manipulated and used in different configurations to generate new samples to enrich the training set. In the preprocessing step, the Sobel filter is applied to each sample to expand the dataset and generate low-level image features to be used as independent input for the classification model. In the classification phase, a three-input CNN-LSTM-SVM is applied to classify the processed data into the corresponding activity.

The schematic diagram of the proposed methodology is depicted in Fig. 1.

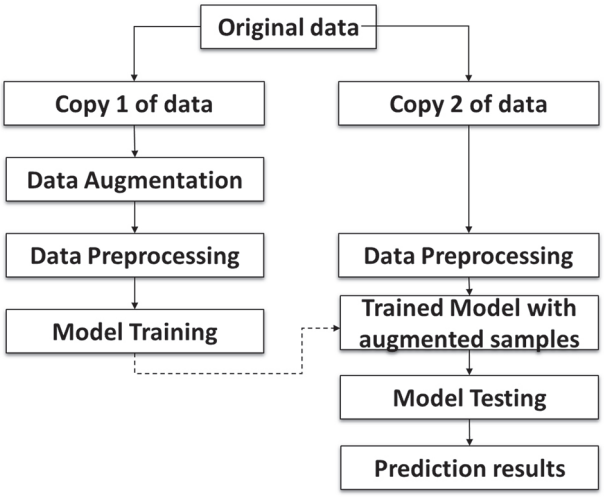


Fig. 1. Schematic diagram of the proposed methodology

3.1. DWT AUGMENTATION

In this work, we focus on the generation of 2D micro-Doppler samples to enrich the training set. We propose to use DWT with the aim of investigating its notable feature that allows the decomposition of an image into components. The DWT scheme is shown in Fig. 2.

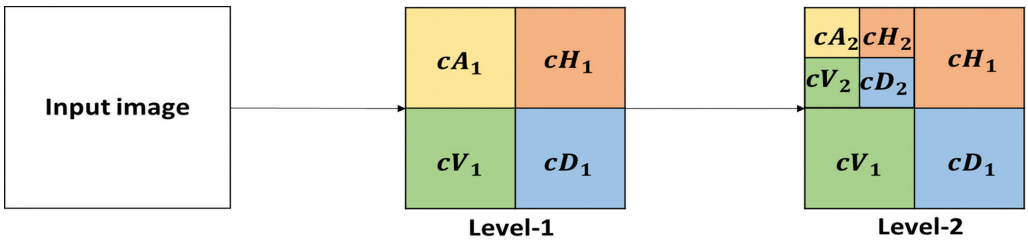


Fig. 2. 2D discrete wavelet decomposition process

3.1. DWT AUGMENTATION

In this work, we focus on the generation of 2D micro-Doppler samples to enrich the training set. We propose to use DWT with the aim of investigating its notable feature that allows the decomposition of an image into components. The DWT scheme is shown in Fig. 2. The DWT applies two filter banks, the low-pass filter, and the high-pass filter, to hierarchically decompose the input image into sub-bands, resulting in a single-level high and low-frequency parts. The low-frequency part is further divided into high and low-frequency parts for further levels of decomposition. The decomposition process is achieved by means of small waves called wavelets, of variable frequency and limited duration. There are many types of wavelet functions, including Haar, Daubechies, Symlets, and Coifflets [40]. Due to its low computing requirements, the Haar transformation has been primarily used for image processing and pattern recognition and is adopted in this work. By employing the DWT with the Haar wavelet, we can effectively decompose the micro-Doppler signatures into different frequency sub-bands, enabling further manipulation to enrich the training set.

Starting with the original image of size NxN, the Micro-Doppler signatures augmentation process is as follows:

- The low-pass and high-pass filters are applied to each row of the image and then sampled by a factor of 2, giving two half-images. One with scaling coefficients and the other with wavelet coefficients. Both images correspond to half the line width of the original image (N/2xN).
- The low-pass and high-pass filters are applied in the column direction of the image generated by the first step. This results in four quarter sub-bands representing different frequency ranges and spatial orientations within the original image. The quarter sub-bands refer to the single-level decomposition coefficients including approximation cA1, diagonal cD1, horizontal cH1, and vertical coefficients of size N/2 x N/2 as shown in Fig. 3. We consider a single-level decomposition to be an appropriate size, as additional levels of decomposition may result in the loss of useful information.
- The important information is concentrated in the approximation coefficient cA1. The horizontal cH1, vertical cV1, and diagonal cD1 coefficients can easily be perturbed by noises. Therefore, we propose to inject each of these coefficients on the cA1 and sum the pixels between the two images as shown in Fig. 3. This way three new images are generated as follows:

cA1 + cD1
cA1 + cV1
cA1 + cH1

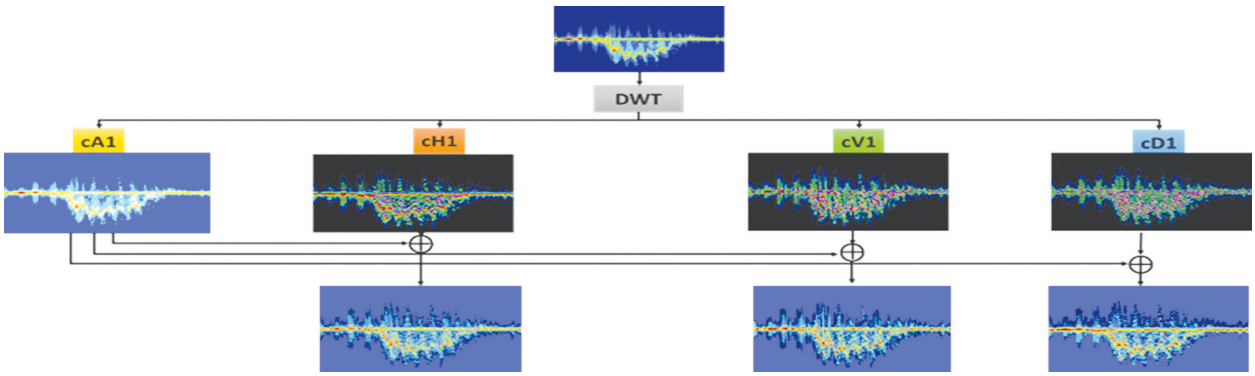


Fig. 3. Proposed augmentation process

3.2. PREPROCESSING

For the preprocessing phase, we adopt the proposed approach in [26]. This method is designed to improve the image content in order to extract and learn discriminative features. It makes use of the Sobel filter to generate low-level image features to be used as independent inputs for the model. These features include the image gradient in the x-direction (Dx), y-direction (Dy), and both x and y-directions (Dxy). The samples are first binarized and then extended using the Sobel

filters. The latter uses two directional filters to convolute the input image, respectively, to obtain Dx, Dy and Dxy. More details about the data preprocessing can be found in [26].

3.3. CLASSIFICATION

Micro-Doppler signatures are, by nature, variable time series data. They represent the time-varying velocity of different parts of the human body. Therefore, it is necessary to involve a process that exploits the temporal correlations of spatial features rather than simply detecting the global geometric shape.

For the classification stage, the three-input CNN-LSTM-SVM model [26] is considered for activity classification. The results from our previous work have demonstrated the effectiveness of combining CNN and LSTM for automatic spatiotemporal feature learning. The spatiotemporal feature extraction is carried out by the Three-input CNN-LSTM. The CNN part consists of three branches with similar layer configurations. Each of the Dx, Dy and, Dxy images is processed in a separate branch by performing multiple convolution operations to extract spatial features. The three CNN branches operate in parallel, and their outputs are combined for further processing by the LSTM. The concatenated features are reshaped and provided as input to the LSTM for temporal feature extraction. The LSTM captures and memorizes how the features extracted by the CNN layers change over time. The output of the LSTM is put into vector form and fed into the multiclass SVM. The SVM uses the OneVsRest strategy and gives the prediction result.

4. MATERIALS AND METHOD

4.1. DATASET

We evaluate the proposed approach on the Multi-Frequency Sensor Network Human Activity Database proposed by Gurbuz *et al.* [13]. The database consists of three datasets acquired from three synchronized radar sensors operating in monostatic mode. The sensors include the 77 GHz FMCW radar IWR1443 from Texas Instruments, the 24 GHz FMCW radar from Ancortek, and the 10 GHz XeThru X4 UWB pulse radar from XeThru. Each dataset comprises 11 classes of human activities, namely WLKT (walking towards the radar), WALKA (walking away from the radar), PICK (picking up an object from the ground), BEND (bending over), SIT (sitting on a chair), KNEEL (kneeling), CRWL (crawling towards the radar), LIMP (limping with a stiff right leg), WTOES (walking on both toes), SHTEPS (walking with small steps), and SCSSR (walking with scissors). Data acquisition involved six participants of various ages, heights, and weights. Each activity is performed 10 times by a participant, resulting in 60 signatures per class per sensor. The micro-Doppler signatures from all three radar sensors for all eleven activities are shown in Fig. 4.

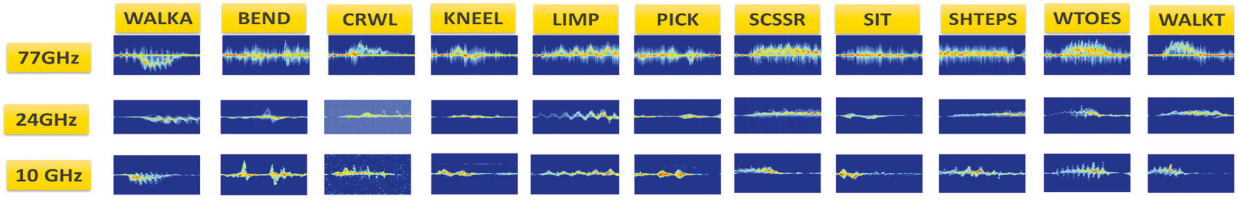


Fig. 4. Micro-Doppler signatures for each human activity class/radar

4.2. IMPLEMENTATION

The implementation of the proposed augmentation technique is performed using Matlab R2021 on a machine running an environment with an Intel (R) Core (TM) i5 2.40 GHz CPU, 16GB of RAM, 1TB of hard disk, and Windows 10. To consistently visualize the impact of the augmentation strategy on the model's generalization ability, we divided the datasets into three subsets: training, validation, and test. The validation set is used as an intermediary checkpoint, allowing us to assess the model's behavior during the training phase. When using the original datasets with a restricted number of samples, the validation set serves the purpose of identifying overfitting. Furthermore, when utilizing augmented samples, the validation set serves a dual purpose. Firstly, it enables us to ensure that the model effectively learns from the augmented samples. Secondly, it assists in ensuring that the augmented samples closely match the characteristics of the test data. The operations of training, validation, and testing were conducted on Google Collaboratory.

To evaluate our proposal in the experimental analysis, four experiments are performed for each dataset.

- **Experiment 1**: use the original datasets without augmentation or preprocessing. Each dataset is split into 80% for training, 10% for validation, and 10% for test. The same sample is simultaneously provided to all three CNN branches. To facilitate a concise comparison, we utilize the same test set across all experiments. To achieve this, we employ the random seed parameter to ensure consistent partitioning.
- **Experiment 2**: use the preprocessed original datasets without augmentation. Each dataset is split into 80% for training, 10% for validation, and 10% for test similar to experiment 1. Each CNN branch is fed with Dx, Dy, and Dxy, respectively.

- **Experiment 3**: use the augmented samples for training and the original samples for testing. The augmented samples are binarized and split into 90% for train and 10% for validation. Two tests are realized. Test (a): Use only a 10% portion of the original dataset as the test set, as mentioned in Experiment 1. Test (b): To prevent biased evaluation, we test the model on the entire original dataset, including all samples.
- **Experiment 4**: use the processed augmented samples for training and the processed original samples for testing. The model is trained using extended augmented images, where each CNN branch is fed with Dx, Dy, and Dxy, respectively. For the test, we follow the same procedure as Experiment 3. The test process of Experiment 3 is replicated.

Note that the model for all experiments is trained from scratch for 100 epochs with a batch size of 16 using the Adam optimizer with a learning rate set to 0.001. All the default configurations of the model are left intact as mentioned in [26]. Except for the LSTM layer, the number of units is changed to 300. For the SVM classifier, the number of binary classifiers is set to 11 corresponding to the number of class activities. We use the same SVM hyperparameters as mentioned in [26]. By keeping hyperparameters constant over all datasets for all experiments, we can demonstrate that the mitigation of overfitting is attributed to data augmentation rather than hyperparameter tuning.

4.3. EVALUATION METHOD

To evaluate the classification performance of the model, five assessment measures are used, including accuracy, precision, recall, F1-score, and the confusion matrix. These metrics are calculated based on the number of true positives (T_P), true negatives (T_N), false positives (F_P), and false negatives (F_N) using the following equations:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (1)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (2)$$

$$Recall = \frac{T_P}{T_P + F_N} \quad (3)$$

$$F1\text{-}score = \frac{2T_P}{2T_P + F_P + F_N} \quad (4)$$

5. EXPERIMENTAL RESULTS

The results of each experiment within the proposed framework will be provided in this section. Each subsection will display the results separately for each dataset.

5.1. 77 GHZ DATASET

The results are presented in Table 1. The confusion matrix and the classification report obtained from the combination of augmentation and preprocessing applied to the test data are depicted in Fig. 5 (A) and Fig. 6 (A), respectively.

Table 1. Comparative classification performance on the 77 GHz dataset

Experiment		Train Acc %	Val Acc %	Test Acc %	Precision %	Recall %	F1-score %
Exp 1		100	93.32	92.36	92.58	92.33	92.30
Exp 2		100	95.44	93.93	95.12	93.94	93.82
Exp 3	(a)	100	98.01	95.45	97.47	95.45	94.73
	(b)	100	98.01	96.47	96.61	96.42	96.40
Exp 4	(a)	100	98.21	98.48	98.99	98.18	98.45
	(b)	100	98.21	96.78	96.97	96.71	96.67

5.2. 10 GHZ DATASET

The results are presented in Table 2. The confusion matrix and the classification report obtained from the combination of augmentation and preprocessing applied to the test data are depicted in Fig. 5 (B) and Fig. 6 (B), respectively.

Table 2. Comparative classification performance on the 10 GHz dataset

Experiment		Train Acc %	Val Acc %	Test Acc %	Precision %	Recall %	F1-score %
Exp 1		100	88.29	81.33	86.58	81.82	80.33
Exp 2		100	91.52	83.33	88.71	83.33	82.80
Exp 3	(a)	100	99.48	100	100	100	100
	(b)	100	99.48	98.18	98.23	98.15	98.17
Exp 4	(a)	100	100	100	100	100	100
	(b)	100	100	100	100	100	100

5.3. 24 GHZ DATASET

The results are presented in Table 3. The confusion matrix and the classification report obtained from the combination of augmentation and preprocessing applied to the test data are depicted in Fig. 5 (C) and Fig. 6 (C), respectively.

Table 3. Comparative classification performance on the 24 GHz dataset

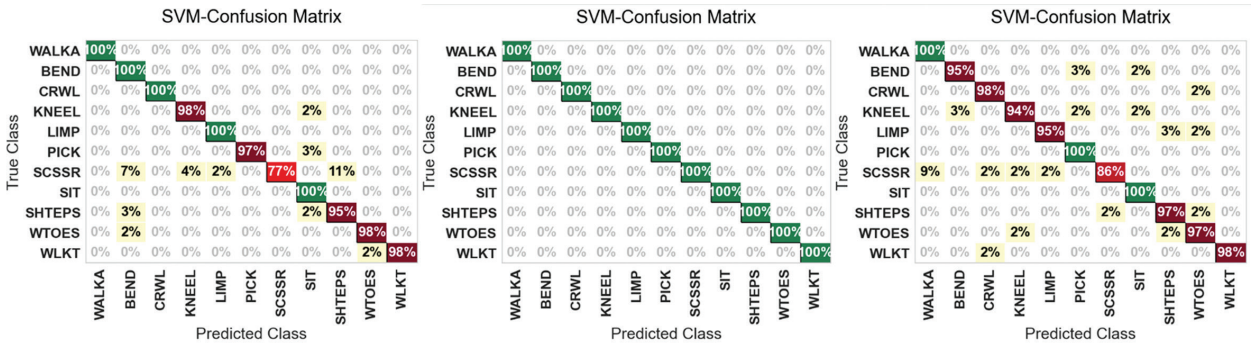
Experiment		Train Acc %	Val Acc %	Test Acc %	Precision %	Recall %	F1-score %
Exp 1		100	86.44	84.84	85.12	87.86	85.25
Exp 2		100	88.41	86.63	86.49	86.64	85.87
Exp 3	(a)	100	95.91	92.42	92.10	94.60	92.81
	(b)	100	95.91	94.27	94.50	94.14	94.14
Exp 4	(a)	100	98.09	98.18	98.18	98.18	97.98
	(b)	100	98.09	96.32	96.40	96.30	96.30

5.4. COMPARISON TO STATE-OF-THE-ART APPROACHES

Our proposed model results can be directly compared with those of Gurbuz *et al*. [13] and Vishwakarma *et al*. [25] as they also used the same 77 GHz dataset. The results are shown in Table 4.

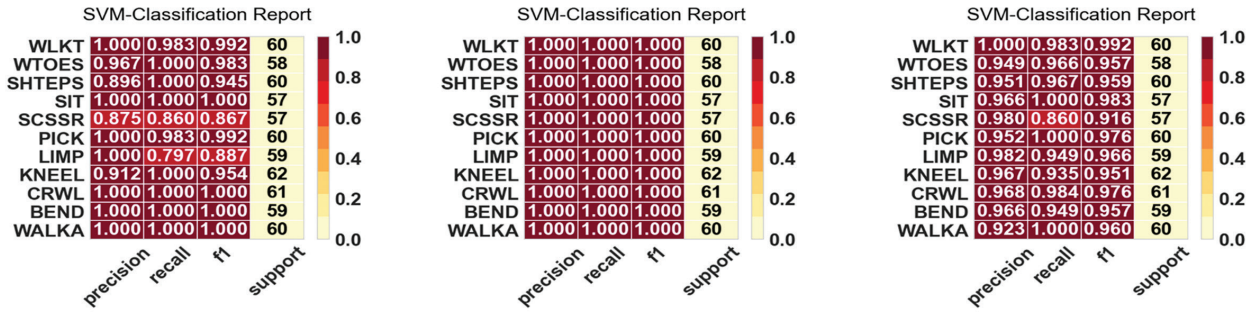
Table 4. Comparative classification performance on the 77 GHz dataset with state of art methods.

Reference	Model	Accuracy
Gurbuz et al. [13]	CAE	85.40%
Vishwakarma et al. [25]	Modified Alexnet	96.44%
Our approach	Three input-CNN-LSTM-SVM	96.78%



(a) (b) (c)

Fig. 5. Confusion matrix exp 4 (b) : (a) 77 Ghz, (b) 10 GHz, (c) 24 GHz, dataset



(a) (b) (c)

Fig. 6. Classification report exp 4 (b) : (a) 77 Ghz, (b) 10 GHz, (c) 24 GHz, dataset.

6. DISCUSSION

The results of Experiment 1 in Tables 1, 2, and 3 demonstrate that the model is suffering from overfitting. During the training phase, the model exhibits exceptional accuracy, reaching 100% on all three datasets. Nevertheless, when validated/tested on new, unseen data, the model's accuracy drops significantly and remains notably lower than its training accuracy. This discrepancy between training and validation/test accuracy is due to the limited size of the training set, which causes a lack of generalization. With restricted training data, the model is unable to capture all possible patterns and variations in the data. As a result, it tends to memorize rather than learn generalizable patterns.

From Experiment 2 in Tables 1, 2, and 3, we observe that using preprocessed data helps to enhance the model's accuracy on all datasets. An improvement rate in the test accuracy of 1.52%, 2%, and 2.2% is noticed for the 77 GHz, 10 GHz, and 24 GHz datasets, respectively. The preprocessing stage plays a crucial role in enhancing the content of images and facilitating the extraction and concatenation of additional features. Despite these benefits, the model's generalization ability remains limited. This is due to insufficient exposure to variations necessary for learning robust features that can effectively generalize to new data.

As observed from Experiment 3, the inclusion of the augmentation process during the training phase enhanced robustness and improved the model's performance. After conducting test (a) from Experiment 3, as shown in Tables 1, 2, and 3, a significant improvement in accuracy across all three datasets is observed. This noteworthy progress is primarily attributed to introducing higher variability among training samples. This variation enables the model to effectively discern patterns and overcome the limitations associated with generalization. Consequently, the model becomes better at making accurate predictions on unseen data, reducing the detrimental effects of overfitting that were apparent in the results of Experiments 1 and 2. The analysis of test (b) results from Experiment 3, as presented in Tables 1, 2, and 3, highlights the comparative performance with the results of test (a). The model achieved an accuracy of 96.43%, 98.18%, and 94.27% on the 77 GHz, 10 GHz, and 24 GHz datasets, respectively. This indicates that the model is well-generalized and has the potential to perform effectively on large new unseen data. This further

substantiates the reliability of the proposed augmentation approach. However, when comparing the model's performance on test (a) and test (b) from Experiment 3, employing a larger test set yielded superior accuracy for both the 77 GHz and 24 GHz datasets. This difference in accuracy suggests that using only a fraction of samples, often obtained through random splitting, might introduce greater variability. This issue arises from the possibility that the selected instances may not accurately represent the entire dataset, leading to biased evaluations. To address this concern, it is more appropriate to consider evaluating the model on a larger number of test samples, promoting a fair assessment.

Following Experiment 4, further improvement is achieved by combining data augmentation and preprocessing. The classification reports depicted in Fig.6 shows consistent results in terms of precision, recall, and F1-score for all three datasets. The model achieved accuracies of 96.78%, 100%, and 96.32% on the 77 GHz, 10 GHz, and 24 Ghz datasets, respectively. The analysis of the confusion matrices (A) and (C) in Fig. 5, corresponding to the 77 GHz and 24 GHz datasets respectively, reveals slight overfitting. The model encounters difficulty in accurately distinguishing between different activities within the various classes. This challenge can be attributed to the presence of highly similar features shared among these classes. The existence of common characteristics leads to confusion and misclassification, contributing to the model's limitations in differentiation. We believe that a review of the model's architecture and parameters could contribute to enhancing its ability to discern between activities. The classification report (B) presented in Fig. 6 indicates that the model performs better on the 10 GHz dataset in terms of classification metrics. Moreover, upon examining the confusion matrix (B) depicted in Fig. 5, it becomes apparent that the number and percentage of false positives and false negatives in the test set are zero. This achievement gains even greater significance as the dataset includes more discriminating patterns, enabling the model to effectively distinguish between different activities. This advantage stems from the high resolution of the IR-UWB, which allows for the detection of subtle motion patterns. As a result, the model achieves superior accuracy and the capacity to capture unique patterns associated with distinct activities and individuals.

To justify the relevance of the proposed approach, a comparative analysis of the performances with those reported in the literature using the same dataset is carried out. The results are reported in Table 4.

Gurbuz *et al.* [13] proposed to use GAN-synthesized data from the 77 GHz dataset to train a Convolutional AutoEncoder (CAE). Their model achieved a modest accuracy of 85.40%. The authors attributed this suboptimal performance to a mismatch between the distributions of synthetic and real data. In contrast, our model provided an improvement rate of 11.38%, reaching an accuracy of 96.78%. Through our approach, we successfully mitigated the mismatch problem by introducing an augmentation strategy that fosters better generalization on unseen data. Vishwarkarma *et al*. [25] proposed to artificially add Additive White Gaussian Noise (AWGN) to the 77 GHz dataset samples and use them to train a modified AlexNet with an attention mechanism. Compared to their model, which reached an accuracy of 96.44%, ours provided a slight improvement of 0.34%. In contrast to our approach, which makes use only of augmented samples, the authors in [25] utilized a combination of original and augmented samples for training purposes. Introducing original samples likely contributed to their improved performance as these samples share the same distribution as the test samples. Additionally, the authors employed a complex architecture with millions of parameters, while our model consists of a simple and lightweight structure maintained with 635397 trainable parameters. Furthermore, it is important to acknowledge that their approach has solely been evaluated on the 77 GHz dataset, and therefore, its generalizability to other datasets remains uncertain.

In conclusion, the proposed approach has been demonstrated to enhance data efficiency during the training process. The generated augmented samples exhibit significant attributes, as affirmed by the outcomes of Experiment 3. These results provide robust validation of the efficacy of the augmentation mechanism. As a result, we can infer that the predictive performance of the model, trained with augmented images using wavelet decomposition, showcases favorable characteristics pertaining to generalization and relevance.

7. CONCLUSION

The significance of this paper lies in offering a practical and effective solution to tackle the scarcity of micro-Doppler signatures for HAR. Our DWT-based augmentation strategy along with preprocessing mitigates the need for extensive data collection and complex models, making it feasible to achieve high performance. Future work in this research aims to explore the impact of using different wavelet basis functions on classification performance and investigate the design of a new architecture to further enhance the classification performance.

8. REFERENCES

[1] G. Diraco, G. Rescio, P. Siciliano, A. Leone, "Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing", Sensors, Vol. 23, No. 11, 2023, p. 5281.

[2] I. Ullmann, R. G. Guendel, N. C. Kruse, F. Fioranelli, A. Yarovoy, "A Survey on Radar-Based Continuous Human Activity Recognition", IEEE Journal of Microwaves, Vol. 3, No. 3, 2023, pp. 938-950.

[3] M. M. Islam, S. Nooruddin, F. Karray, G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects", Computers in Biology and Medicine, Vol. 149, 2022, p. 106060.

[4] A. Dey, S. Rajan, G. Xiao, J. Lu, "Radar-based Human Activity Recognition: Is it Ready for Aging in Place?", TechRxiv, 2023.

[5] J. Maitre, S. Bouchard, K. Gaboury, "Data filtering and deep learning for enhanced human activity recognition from UWB radars", Journal of Ambient Intelligence and Humanized Computing, Vol. 14, No. 6, 2023, pp. 7845-7856.

[6] Z. Wang, A. Ren, Q. Zhang, A. Zahid, Q. H. Abbasi, "Recognition of Approximate Motions of Human Based on Micro-Doppler Features", IEEE Sensors Journal, Vol. 23, No. 11, 2023, pp. 12388-12397.

[7] S. Hassan, X. Wang, S. Ishtiaq, N. Ullah, A. Mohammad, A. Noorwali, "Human Activity Classification Based on Dual Micro-Motion Signatures Using Interferometric Radar", Remote Sensing, Vol. 15, No. 7, 2023, pp. 1–21.

[8] F. Luo, E. Bodanese, S. Khan, K. Wu, "Spectro-temporal modelling for human activity recognition using a radar sensor network", IEEE Transactions on Geoscience and Remote Sensing, Vol. 61, 2023, pp. 1-13.

[9] L. Jiang, M. Wu, L. Che, X. Xu, Y. Mu, Y. Wu, "Continuous Human Motion Recognition Based on FMCW Radar and Transformer", Journal of Sensors, Vol. 2023, 2023.

[10] X. Li, Y. He, X. Jing, "A survey of deep learning-based human activity recognition in radar", Remote Sensing, Vol. 11, No. 9, 2019, p. 1068.

[11] S. Huan, L. Wu, M. Zhang, Z. Wang, C. Yang, "Radar Human Activity Recognition with an Attention-Based Deep Learning Network", Sensors, Vol. 23, No. 6, 2023, p. 3185.

[12] H. Zhou, Y. Zhao, Y. Liu, S. Lu, X. An, Q. Liu, "Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System", Sensors, Vol. 23, No. 10, 2023, p. 4750.

[13] S. Z. Gurbuz, M. M. Rahman, E. Kurtoglu, T. Macks, F. Fioranelli, "Cross-frequency training with adversarial learning for radar micro-Doppler signature classification (Rising Researcher)", Radar Sensor Technology XXIV, Vol. 11408, 2020, pp. 58-68.

[14] M. Mahbubur Rahman, S. Z. Gurbuz, "Multi-Frequency RF Sensor Data Adaptation for Motion Recognition with Multi-Modal Deep Learning", Proceedings of the IEEE Radar Conference, Atlanta, GA, USA, 07-14 May 2021, pp. 1-6.

[15] L. Alzubaidi et al. "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications", Journal of Big Data, Vol. 10, No. 1, 2023, p. 46.

[16] G. Zhou, Y. Chen, C. Chien, "On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks", BMC Medical Informatics and Decision Making, Vol. 22, No. 1, 2022, p. 226.

[17] X. Hao, L. Liu, R. Yang, L. Yin, L. Zhang, X. Li, "A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition", Remote Sensing, Vol. 15, No. 3, 2023. p. 827.

[18] S. Yang et al. "The Human Activity Radar Challenge: benchmarking based on the Radar signatures of human activities dataset from Glasgow University", IEEE Journal of Biomedical and Health Informatics, Vol. 27, No. 4, 2023, pp. 1813-1824.

[19] R. Bravin, L. Nanni, A. Loreggia, S. Brahnam, M. Paci, "Varied Image Data Augmentation Methods for Building Ensemble", IEEE Access, Vol. 11, 2023, pp. 8810-8823.

[20] Y. Yang, Y. Zhang, C. Song, B. Li, Y. Lang, "Omnidirectional Spectrogram Generation for Radar-based Omnidirectional Human Activity Recognition", IEEE Transactions on Geoscience and Remote Sensing, Vol. 61, 2023, pp. 1-13.

[21] A. Oubara, F. Wu, A. Amamra, G. Yang, "Survey on Remote Sensing Data Augmentation: Advances, Challenges, and Future Perspectives", Proceedings of the 5th International Conference on Computing Systems and Applications, Algiers, Algeria, 17-18 May 2022, pp. 95-104.

[22] D. Bathe, K. Patil, N. Sanjay, "Leveraging Potential of Deep Learning for Remote Sensing Data", Proceedings of the International Conference on

Intelligent Systems and Human Machine Collaboration, Maharashtra, India, 8-9 July 2022, pp. 129–145.

[23] O. Pavliuk, M. Mishchuk, C. Strauss, "Transfer Learning Approach for Human Activity Recognition Based on Continuous Wavelet Transform", Algorithms, Vol. 16, No. 2, 2023, p. 77.

[24] A. Alkasimi, A. V. Pham, C. Gardner, B. Funsten," Geolocation tracking for human identification and activity recognition using radar deep transfer learning", IET Radar, Sonar & Navigation, Vol. 17, No. 6, 2023, pp. 955-966.

[25] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. R. Adve, K. Chetty, "Attention-enhanced Alexnet for improved radar micro-Doppler signature classification", IET Radar, Sonar & Navigation, Vol. 17, No. 4, 2023, pp. 652-664.

[26] D. S. Korti, Z. Slimane, K. Lakhdari, "Enhancing Dynamic Hand Gesture Recognition using Feature Concatenation via Multi-Input Hybrid Model", International journal of electrical and computer engineering systems, Vol. 14, No. 5, 2023, pp. 535–546.

[27] Y. Lang, Q. Wang, Y. Yang, C. Hou, D. Huang, W. Xiang, "Unsupervised Domain Adaptation for Micro-Doppler Human Motion Classification via Feature Fusion", IEEE Geoscience and Remote Sensing Letters, Vol. 16, No. 3, 2018, pp. 392–396.

[28] H. Du, T. Jin, Y. Song, Y. Dai, "Unsupervised Adversarial Domain Adaptation for Micro-Doppler Based Human Activity Classification", IEEE Geoscience and Remote Sensing Letters, Vol. 17, No. 1, 2019, pp. 62-66.

[29] H. Du, Y. He, T. Jin, "Transfer Learning for Human Activities Classification Using Micro-Doppler Spectrograms", Proceedings of the IEEE International Conference on Computational Electromagnetics, Chengdu, China, 26-28 March 2018, pp. 1-3

[30] H. Du, T. Jin, Y. Song, Y. Dai, M. Li, "Efficient human activity classification via sparsity-driven transfer learning," IET Radar, Sonar & Navigation, Vol. 13, No. 10, 2019, pp. 1741-1746.

[31] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, M. G. Amin, "Diversified radar micro-Doppler simulations as training data for deep residual neural networks", Proceedings of the IEEE Radar Conference, Oklahoma City, USA, 23-27 April 2018, pp. 612-617.

[32] Y. Lang, C. Hou, H. Ji, Y. Yang, "A Dual Generation Adversarial Network for Human Motion Detection Using Micro-Doppler Signatures", IEEE Sensors Journal, Vol. 21, No. 16, 2021, pp. 17995-18003.

[33] L. Qu, Y. Wang, T. Yang, Y. Sun, "Human Activity Recognition Based on WRGAN-GP-Synthesized Micro-Doppler Spectrograms", IEEE Sensors Journal, Vol. 22, No. 9, 2022, pp. 8960-8973.

[34] X. Shi, Y. Li, F. Zhou, L. Liu, "Human Activity Recognition Based on Deep Learning Method", Proceedings of the IEEE International Conference on Radar, Brisbane, QLD, Australia, 27-31 August 2018, pp. 1-5.

[35] I. Alnujaim, S. S. Ram, D. Oh, Y. Kim, "Synthesis of Micro-Doppler Signatures of Human Activities from Different Aspect Angles Using Generative Adversarial Networks", IEEE Access, Vol. 9, 2021, pp. 46422-46429.

[36] B. Erol, S. Z. Gurbuz, M. G. Amin, "Motion Classification Using Kinematically Sifted ACGAN-Synthesized Radar Micro-Doppler Signatures", IEEE Transactions on Aerospace and Electronic Systems, Vol. 56, No. 4, 2020, pp. 3197-3213.

[37] Y. J. Zhong, Q. S. Li, "Human Motion Recognition in Small Sample Scenarios Based on GAN and CNN Models", Progress in electromagnetics research M, Vol. 113, 2022, pp. 101-113.

[38] B. Erol, S. Z. Gurbuz, M. G. Amin, "Synthesis of micro-doppler signatures for abnormal gait using multi-branch discriminator with embedded kinematics", Proceedings of the IEEE International Conference on Radar, Washington DC, USA, 28-30 April 2020, pp. 175-179.

[39] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, S. Z. Gurbuz, "Effect of Kinematics and Fluency in Adversarial Synthetic Data Generation for ASL Recognition with RF Sensors", IEEE Transactions on Aerospace and Electronic Systems, Vol. 58, No. 4, 2022, pp. 2732-2745.

[40] E. Guariglia, R. C. Guido, G. J. P. Dalalana, "From Wavelet Analysis to Fractional Calculus: A Review", Mathematics, Vol. 11, No. 7, 2023, pp. 1-12.