IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 June 2023, accepted 26 July 2023, date of publication 31 July 2023, date of current version 8 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3300224

RESEARCH ARTICLE

Denoising UWB Radar Data for Human Activity Recognition Using Convolutional Autoencoders

VIRGILE LAFONTAINE, KÉVIN BOUCHARD, (Member, IEEE), JULIEN MAíTRE, AND SÉBASTIEN GABOURY, (Senior Member, IEEE)

Laboratoire d'Intelligence Ambiante pour la Reconnaissance d'Activités (LIARA), Université du Québec à Chicoutimi, Saguenay, QC G7H 2B1, Canada

Corresponding author: Kévin Bouchard (Kevin_Bouchard@uqac.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)–Discovery Grant.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethic Committee Review Board of UQAC.

ABSTRACT Human Activity Recognition (HAR) is one of the most popular research topics thanks to its usefulness in providing targeted, meaningful assistance to older adults. Because of the aging of the population in first-world countries, it becomes increasingly important to find innovative solutions that reduce risks associated with aging-in-place policies. HAR proposes solutions that are based on Ambient Intelligence (AmI) to alleviate those risks. In this work, we exploited three UWB radars to recognize 14 activities performed by 19 participants in a prototype smart-home apartment. The main contribution of this paper is UWB radar data cleaning on a practical dataset. The UWB radar data has been filtered using an unsupervised deep convolutional autoencoder (CNN-AE) that learns background noise from the data. This filtering method is compared to the unfiltered data using a Convolutional Neural Network (CNN) classifier in a Leave-One-Subject-Out (LOSO) classification. Performances attest that the CNN-AE unsupervised filtering is efficient for HAR. In addition, we tested the generalization potential of this architecture when the dataset is comprised of a lower number of participants (1, 5, 10, and all 19 participants). Generalization in HAR is difficult as the results show the importance of data quantity and number of subjects. We obtained 69.9% top-1 accuracy when using our filtering architecture compared to 48.4% without it. To conclude, we show that an unsupervised CNN-AE can efficiently filter and generalize UWB radar data in a HAR setting while providing easier learning constraints and implementation on a practical dataset.

INDEX TERMS Activity of daily living, data filtering, data processing, deep learning, human activity recognition, unsupervised learning, UWB radars.

I. INTRODUCTION

As developed countries see a rise in life expectancy, more and more of adult life is spent after the age of 65 [1], resulting in additional strain on society as a whole. For example, adults older than 60 years of age suffer the greatest number of fatal falls, while most falls are severe enough to require medical attention [2]. This pressure is put on the healthcare system which leads to an increase in interest in gerontology technology research. More specifically, Human Activity Recognition (HAR) is getting more and more popular as technology progresses. With the rapid development of big data, 5G, the Internet of Things (IoT) and other technologies, sensors such as wearables and radars become a focus of attention as new classification models and fusion methods are developed. Non-invasive sensors are helpful to conduct HAR which helps us understand human behavior. This is of major interest as HAR can be used to evaluate the functional capacity of an individual by assessing his capabilities when performing common Activities of Daily Living (ADL) [3]. Such assessments can lead to early warnings of degrading health that are useful to caregivers or relatives. Moreover, detected anomalies can be used to alert caregivers in case of emergencies, such as the aforementioned falls. As age-related health problems such as diabetes, cardiovascular disease, Alzheimer's disease, dementia, or other chronic diseases, associated with

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva.

cognitive and physical weakening of the elderly people prevents them from living independently, smart home and HAR seem like promising solutions to the challenge of a growing aging population [4]. Indeed, the lack of specialized facilities in addition to the willingness to preserve the autonomy, independence, and self-esteem of the elderly is a strong motivation to develop HAR in automated health systems. Ambient intelligence for HAR provides detection of motion and interactions with objects and furniture inside the home (light switch activation, electrical outlets usage, etc). Recognition algorithms can then classify those activities and report anomalous behaviors. ADL performance is a crucial metric for assessing the quality of life [5] and such a recognition system could improve the standard of living when the given information is taken into account and acted upon by the caregivers.

HAR, as promising as it seems, is not without its own set of challenges. The most prominent one is the variety of sensors used, which results in different types of collected data. In addition to that, the accuracy of ADL recognition systems is inherently related to the training dataset size, participant variety, and number, as well as sufficiently diverse and balanced performed activities [6], [7]. This increases the difficulty of collecting quality data in experiments and raises the cost of providing a meaningful dataset to train models. Furthermore, special care has to be given to the validation process, as randomized train/test/validation instances separation results in all participants' data present in training and validation datasets. High inter-class variance is expected in HAR, seeing that each participant has their own way of performing activities, with tendencies to use specific movements. It is then useful to use participants-based cross-validation to assess the generalization potential of the systems [8].

In this paper, we use data provided by three wall mounted Ultra-WideBand (UWB) radars inside the Laboratoire d'Intelligence Ambiante pour la Reconnaissance d'Activités (LIARA) apartment. The UWB radar is technically an ambient sensor where collected data are richer in information than usually found on other ambient sensors-based technologies. As an example, thermal cameras are a useful option for recognizing activities of daily living, but they have limitations due to the fact that they only provide a 2D x,y image without depth information. This means that in order to avoid the occlusion of objects and subjects, thermal sensors need to be used in conjunction with RGB cameras or other equipment to facilitate 3D scene reconstruction [9]. UWB radars also alleviate the privacy concerns typically found when using cameras. They allow locating a person in a room accurately and give information on their movements (number and amplitude of movements). Such precision can be used for movement detection and activity recognition.

Our goal is to recognize 14 distinct activities performed by 19 participants. To this effect, data collected using these 3 radars have been cleaned from noise using an unsupervised denoising autoencoder filter. The use of a denoising autoencoder for the noise-cleaning process is motivated by its unsupervised properties and ease of use. When dealing with HAR datasets, it is often difficult if not outright impossible to provide clean data to feed the training process because of the noise from the transceiver circuit and low Signal to Noise ratio. Using an autoencoder allows only select limited features to partially reconstruct the data, effectively reconstructing the background noise of the data. This background noise is then subtracted from the original data, resulting in a filtered output. Using this approach, we can train this architecture without the need for data labels, greatly facilitating the data preprocessing steps. To assess the performances of this architecture on our dataset, we trained a relatively simple Convolutional Neural Network (CNN) classifier comprising only a few convolution layers. This classifier then recognizes activities with a Leave-one-subject-out (LOSO) cross-validation. Our findings show that using an unsupervised autoencoder as a denoising filter improves the mean accuracy of such a classifier by 21.6%, as the mean accuracy reaches 70.0% for the cleaned dataset whereas feeding the classifier with the original dataset results in a mean accuracy of 48.4%. In addition to this result, top scores show that using this filter gives better results when the number of instances is low in the dataset, with tested accuracy gains with as low as 3 participants in the training dataset. In summary, our contribution is as follows:

- we explored the use of an unsupervised autoencoder architecture in a data-cleaning process that is easy to train and does not require any clean data.
- we compared the performance of this denoising method with the unfiltered dataset
- we showed the generalization capabilities of such an architecture when dealing with low instances and/or subjects number, enabling the use of such a filter as an additional module in those conditions.

The remainder of the paper is organized as follows: Section II describes common HAR technologies and UWB radar noise filtering state-of-the-art. Section III presents the experimental setup and the unique set of activities performed by the participants. Section IV details the noise filtering method, including the autoencoder and classifier architectures. Section V shows and discusses our results. Finally, Section VI concludes this study.

II. RELATED WORK

In the literature, HAR is commonly used in conjunction with wearable devices. For example, smartphones and smartwatches provide a compact set of sensors (accelerometers, gyroscopes, heart-rate sensors, etc.) [10] and have the means of communicating the information to the automated health system using WiFi or Bluetooth. The precision of classifiers using this technology are usually high for a wide array of activities found in ADL, such as walking, sleeping, eating, and doing the dishes. In [11], the authors recognized four activities (ambulation, cycling, sedentary and other) from

wrist-worn accelerometer data and achieved an accuracy of 0.895. In any case, a limitation of the wearable approach is that the device has to be worn. Older adults can forget to wear the device, creating a lack of monitoring data for extended periods of time. Another approach is to use ambient sensors such as Passive Infrared (PIR) sensors, on/off switch sensors, and pressure mats [12], [13]. In [12], the authors compared Adaptive Boosting (AdaBoost) and Fuzzy C-means (FCM) classifiers for HAR from binary ambient sensors. Features were extracted from collected data using a Deep Convolutional Neural Network (DCNN). Eight activities were considered in the study: sleeping, using the toilet, using the kitchen for food or dishwashing, watching TV, eating, dressing, doing garden activities, and leaving home. The authors decided to tie the human activity to each room, resulting in activity recognition when the person was in a specific room. As an example, a person located in the dining room is probably eating a meal. Using the AdaBoost classifier, the authors reached an accuracy of 0.995. The foremost disadvantage of such an approach is the need to add sensors of numerous types proportionally to the number of activities that are to be recognized. The number of sensors can be cost-prohibitive while the use of multiple sensor types adds complexity to the data fusion process. Another approach is to exploit RGB, depth, and/or thermal cameras.

Cameras can be used to extract rich and detailed information about the resident and its environment that can be used to conduct HAR [14], [15], [16]. For example, a system capable of performing scene understanding is presented in [14]. The process includes leveraging a stereoscopic camera as part of a device network composed of vital parameters (wristband, glucose meter, etc.), home automation (smart switch, smart plugs, etc.), and physical activity tracker (video monitoring, accelerometers) devices. In this framework, video acquisition is analyzed to detect human-object interactions and activities. Specifically, the derived activities consist of an action performed on an object (e.g. sitting on a chair, to lay on a bed). Two convolutional networks are used to respectively detect an object and a human in interaction with each other after having been trained on the public COCO human interaction dataset [17].

In [15], the authors developed a monitoring approach to detect daily living activities such as eating or dressing. After analyzing the available datasets, they applied Faster R-CNN (Regions with CNN features) [18] for human localization in the video frame, I3D [19] for benign activity recognition (small movements often seen in eg. eating, drinking or washing teeth) and DeepHAR [20] for whole body movements (such as falling or walking). The resulting architecture is a low-cost vision-based monitoring expert system.

In the case of depth sensors, a recent review [21] presents a systematic discussion on the state-of-the-art using depth sensors to detect anomalous behavior, mainly fall detection. In conjunction with thermal and audio sensors, depth sensors are a meaningful non-intrusive technique in HAR. Indeed, vision-based monitoring is intrusive and raises concerns about privacy in health monitoring, especially when caregivers have access to the information.

Thus, another non-intrusive sensor found in monitoring systems is radar-based. The two commonly used radar technologies in HAR are Frequency Modulated Continuous Wave (FMCW) radars and UWB impulse radars [22]. Although FMCW radars often use micro-Doppler signatures to recognize human activities [23], the UWB radar uses ultra-short pulses (in the range of a few nanoseconds) in a wide frequency range at a low pulse power. The time of flight of emitted pulses is measured to give the precise distance between the radar and objects. A more detailed explanation is given in Section III. While this technology is non-intrusive, there is radar clutter and background noise in the received data.

Considering UWB radar data denoising techniques, the use of supervised deep learning denoising techniques for UWB radar data is common in the literature using clean ground truth supervision [24], [25], [26], [27]. Unsupervised deep learning denoising, when clean and/or labeled data are not available, is rarer. Nevertheless, several previous studies successfully leveraged deep learning architectures to denoise UWB radar data while using unsupervised or semi-supervised feature extraction [28], [29] and denoising [30].

Similarly, the authors of [24] conducted experiments where UWB radars were used in through-wall human recognition. They used a sparse autoencoder as a supervised denoising encoder and features extractor to process the radar data. The evaluation of this architecture was done using radar recordings of participants standing behind a wall. They show in their experiments that using multiple radars over a single radar, as well as using L2 regularization and dropout layers improves performances under small training data to mitigate overfitting in a small dataset condition.

In [25], the authors proposed the utilization of a Generative Adversarial Network (GAN) to denoise UWB spectrograms. They measured motion data of seven human motions: running, jumping, creeping, walking, standing, crawling, and boxing. In addition to this dataset, 1 second of slow time for each motion is simulated to produce the corresponding simulated dataset. The GAN architecture is trained in a supervised manner to differentiate between noisy-denoised outputted by a denoising generator and a real noisy-clean pair. The goal is to train a generator through the GAN architecture that can effectively denoise input data. In their study, they simulated noisy data of five decreasingly noisy Signal-To-Noise Ratios (SNR) to assess the performance of such an architecture. The generator architecture is composed of convolutional layers to extract relevant feature maps from the noisy data.

Reference [26] proposes a sparse deep autoencoder that is compared to Singular Value Decomposition (SVD) and wavelet denoising algorithms. In simulated and real experiments involving noise from a participant and walls, the use of sparsity in the CNN-AE architecture results in better

denoising, while the denoising performance of the deep autoencoder is still better than SVD and wavelet. It should be noted that the targeted noise in their study is mainly reflections from the wall and the target, with supervised training.

The authors of [27] did also use the autoencoder filtering properties to reduce the motion noise of the participant using a motionless ground truth to extract respiratory data. The autoencoder architecture is variational and takes two processed inputs calculated from the radar data frames. The training is also directed by ground truth and provides respiratory data extraction under heavy movements from the participants.

UWB radar can also be used to perform feature extraction that reduces data dimensionality while retaining information. An unsupervised feature learning architecture is developed in [28], where a K-means algorithm is performed on randomly selected patches of data frames to learn relevant features that will be used to feature-transform UWB data before SVM classification.

In [30], the authors denoised UWB data using a CNN network to compress and decompress the data for localization purposes. Similarly to our work, the sparsity and spatial correlation of the signal in the radar data frame are leveraged by stacking multiple layers of convolutions to extract the most relevant features. Deconvolutions layers are then stacked to decompress the data. This architecture is then evaluated on simulated UWB radar data, with results that demonstrate the affiliated gain in UWB denoising based on the Root Mean Square Error (RMSE) metric. Multiple simulated UWB radar datasets were created and used in training to mimic a localization scenario. In addition, results show that increasing layer size and kernel size does not necessarily improve the performance of this architecture.

Most of the work found in the literature revolves around denoising labeled data using clean and noisy pairs. In practice, labeling and creating a clean dataset is time-consuming, pricey, and difficult. Thus, CNN architectures seem interesting in their ability to extract features in unlabeled noisy data when used in unsupervised algorithms.

III. EXPERIMENTAL SETUP

This section is a description of the entire experimental setup to provide a clear understanding of all the specifics of this study. To this end, the apartment where the experiments were conducted is detailed. A brief explanation of the UWB radar is also given. To conclude this section, the activities that the participants carried out are laid out in detail.

A. APARTMENT LAYOUT AND UWB RADAR LOCATIONS

The dataset has been collected at the LIARA, which has a prototype apartment with a surface of approximately 40 square meters that aims to reproduce an entire operational apartment. This apartment is an environment to test monitoring and assistance methods. Fig. 1 presents the layout of the apartment. It is comprised of a bedroom, kitchen, and living room. More than a hundred sensors (passive infrared sensors, contact sensors, pressure-mat sensors, etc) and effectors (screens, speakers, etc) are set up to provide data collection. The objective is to monitor human activities using these data with the end goal of providing specific assistance at the right time. For example, to notify the user with speakers when the oven has been let on for an extended period of time with the help of data processing and machine learning. The use case of such an environment is to allow the elderly to keep their in-home autonomy while maintaining health monitoring and assistance.

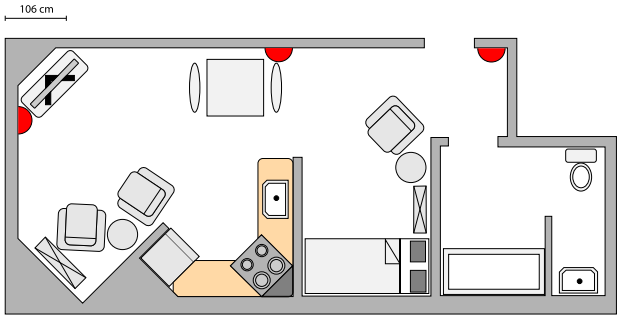

FIGURE 1. LIARA apartment layout. The 3 UWB radars are visible in red.

For this study, three wall-mounted UWB radars were used. The model used is the Xethru X4M200 from Novelda, which can be electrically supplied directly from the electrical outlets. Thus, the height at which they are mounted was selected to be around standard electrical height in Canada, approximately 36cm. The locations of the radars have been designed to provide ample coverage of the apartment's surface, with at least one radar having a direct line of sight on every possible location of the user.

B. UWB RADARS

The Xethru X4M200 uses a UWB transceiver (transmitter and receiver) operating within the 6.0-10.2GHz band. The low-frequency band that this radar uses is 6.0-8.5GHz, while the high-frequency range for the same radar is 7.25-10.20GHz. The power spectral density (PSD) of the UWB radar is very low compared to other wireless communication technologies. With its high-frequency range combined with its low PSD, this radar emits low-energy electromagnetic waves that are suitable for medical applications [31].

Due to those properties, The radar is only able to detect reflected waves over a short range of 5 to 10 meters, but can still pierce through thin walls and obstacles. In addition, the waves do not interfere much with other electromagnetic waves. The transmitter module of the transceiver continuously emits very short pulses (Impulse-radio) of about two nanoseconds. The emitted waves are then reflected by the environment, and the receiver module of the transceiver senses the reflected waves. The time of flight (ToF) is used to compute the distance between the radar and the element of the environment that reflected the wave. The very short pulse generated by the UWB radars is what gives this technology

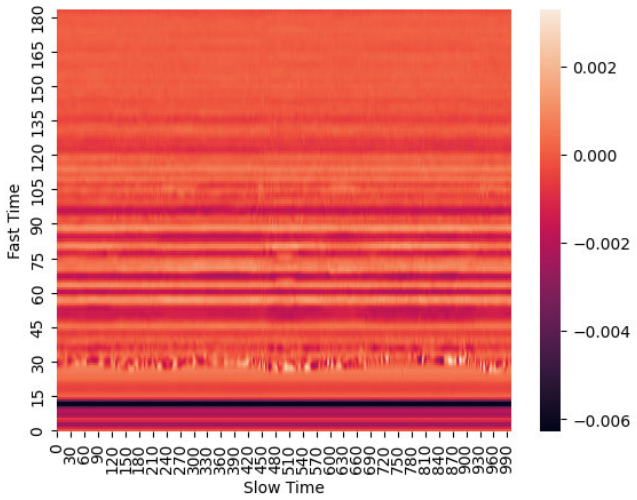

FIGURE 2. Example of a UWB radar scattering matrix of a single radar over a period of 300 frames of slow time (6 seconds) for the activity '*doing housework*'. The participant's movement is visible around the 30th bin in the fast time axis.

such precision. In our case, the UWB radar has a spatial resolution of 5.14 cm.

The data collected by the radar is a sequence of basebands. This particular model is set up to produce 50 basebands per second. A baseband is a series of values representing the amplitude of the reflections at all distances in front of the radar. In the literature, basebands can also be called a range [m], a radar frame, or fast time. In this study, a baseband has 184 values (bins) in the detection zone of [0.4-9.8m]. The radar frames start 5.14 cm before the lower range and 5.14 cm after the higher range. Collecting radar frames along one axis represents the slow time. For example, accumulating 100 radar frames at a sampling rate of 50 frames per second, the slow-time will range from 0 to 2 seconds. This collection of baseband represents a radar scattering matrix. To ensure comprehensiveness and to better understand the presented work, it is worth noting that our previous work [32] utilized the same dataset and equipment. Fig. 2 illustrates a radar scattering matrix provided by one UWB radar over 300 frames (6 seconds).

C. LIST OF HUMAN ACTIVITIES

The collected dataset is composed of 14 distinct human activities performed by 19 participants. The activities were designed to be ADLs that would often be performed in an apartment. The complete list of activities is as follows: *Brushing teeth*, *doing housework*, *drinking*, *eating*, *getting dressed and undressed*, *going to the toilet*, *putting away laundry*, *putting away dishes*, *reading a book*, *resting*, *sleeping*, *taking a shower*, *using a computer or a phone*, *washing dishes*. For each of these activities, the participants were briefed on where to be located when performing the activity inside the apartment, as well as what objects to use. As such, we attempted to reduce intra-class variability that would otherwise be significant (picture cleaning the room at different locations, with different tools such as a broom, vacuum, or participants cleaning the table and others the floor). However, the participants were allowed to do the activities in their own way (eg. moving while sleeping, the order of dishes to be washed, the posture, etc.). The goal is to have a dataset that mimics the diversity of movements different users would have when performing the same activity. It should be mentioned that all experiments have been conducted under an ethical certificate obtained through the Université du Quebec Chicoutimi human research ethic board (ID number: 2019-202), with the participant's signed consent.

IV. THE PROPOSED APPROACH

This section presents the data processing steps that have been carried out prior to model training. Then, the filtering and classifier architectures that we exploited for our HAR needs are detailed.

A. DATA PROCESSING

As seen in Section III, data are collected from three UWB radars that send timestamped radar frames at a specified sampling rate. The data collection process consists of creating for each activity three radar scattering matrices, where each radar frame is temporally aligned. The resulting scattering matrices, one for each radar, then represents the amplitude of reflected electromagnetic waves at all distances over time. The number of bins collected is 184 for each radar frame. In a preprocessing step, we cropped the start of each activity by 10 seconds and the end by 5 seconds. This was motivated by the fact that there was a delay between the start of the recording and the moment the participants were instructed to start the activity. In the same manner, the participant was instructed to stop the activity before the recording was stopped. The three scattering matrices are then stacked, effectively combining the three radar outputs in a single scattering matrice of thrice the size.

The data filtering process consists of training a deep convolutional autoencoder (CNN-AE) on all the datasets, without taking into consideration the class or the subject. The intuition is to use the restrictive nature of the CNN-AE to limit the number of features that the architecture can extract. In practice, the CNN-AE can reconstruct the background noise, without using any metadata (eg. label data, subject information, room of the activity, duration, etc.). The proposed method has the advantages of reducing the amount of necessary data preprocessing, such as labeling, as well as reducing the training time complexity that is usually found in deep learning architectures. Considering the unsupervised approach and the relatively fast training time, this cleaning step can be viewed as a fast modular architecture to be used in a preprocessing step in more complex architectures. Fig. 3 illustrates the training process, including the use of a sliding window over the slow time axis to create the training instances. The sliding window has an overlap of 90%, resulting in data augmentation to provide a sufficient instances number. The chosen shape for an instance is (500 × 552),

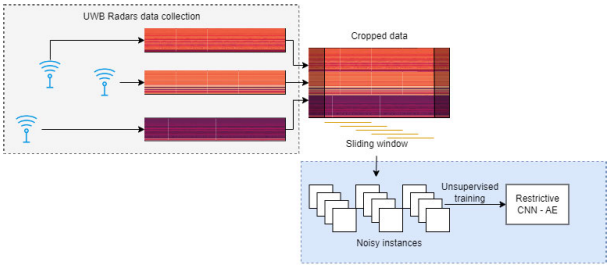

FIGURE 3. The preprocessing steps leading to the unsupervised training. An overlapping sliding window is in charge of creating the training instances.

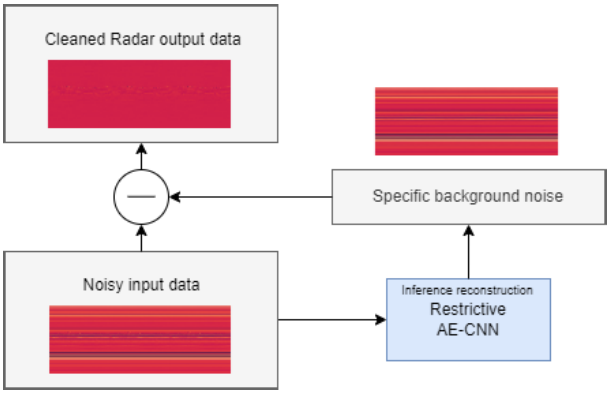

FIGURE 4. The filtering process. The denoising module reconstructs the background noise and subtracts it from the noisy input.

meaning that the duration of an instance is 10 seconds (10 times the sampling rate of 50 radar frames per second) with the three radar frames stacked on one axis (3 times 184 bins of a radar frame). The resulting total number of instances is 26391. It should be noted that using one axis per radar frame resulting in instances with the shape (500 × 184 × 3) had the effect of slowing the training phase of the CNN-AE. The instances are min-max normalized on a per-instance basis and the CNN-AE is then trained on all the created instances. The loss function used during training is the mean squared error (MSE), which drives the architecture to learn features necessary to reconstruct the instances. The training process

The trained CNN-AE is used in inference during the filtering process. Raw instances are fed for the CNN-AE to reconstruct. The partial reconstruction is made possible by the architecture, which outputs only the background noise, omitting the signal. With this noise signature available, the original data are cleaned by subtracting the reconstructed noise. This process uses the ability of the CNN-AE to encode the noise patterns and reconstruct instance-tailored noise, meaning that each instance has its specific noise reconstructed. Fig. 4 represents the filtering process, whereas Fig. 5 shows a comparison between a raw instance, its reconstructed noise, and the cleaned instance using this instance-specific reconstructed noise.

B. FILTERING EVALUATION

To evaluate the effect of the proposed data cleaning method, we chose to use a fairly simple CNN classifier to provide the basis for a comparison tool. To avoid any data contamination, two datasets were created. The first one is the raw dataset, comprised of the original data. The second dataset is the filtered data, where a sliding window has been used to filter the data using the trained CNN-AE. The two datasets are then normalized using a min-max normalization, resulting in two datasets that have the same shape, same number of possible instances, and range. Each dataset undergoes the same following training process, as summarised in Fig. 6: First, the dataset is divided into a training set, a validation set, and a test set following the LOSO-CV principle. For each of the 19 test participants, a validation participant is randomly chosen. The 17 remaining participants are then forming the training set. To give a better understanding of the size of each dataset, Table. 1 shows the number of instances of shape (500 × 184 × 3) in the training, test and validation dataset for each test participant in a single run. Next, instances are created from the training and validation sets and are fed to the CNN classifier with their activity label. To summarize, each participant is a test dataset used in cross-validation to assess the generalization potential of the architecture. At the end of the training phase of the classifier, the weights of the best-performing epoch on the validation set are saved. Its performance after training is evaluated in inference as the previously saved weights are used to classify all the instances in the test set. Evaluation metrics are then measured. This training and evaluation process is repeated 10 times to provide accurate measurements. The entire training and evaluation process is the same for the raw and filtered dataset, including the choice of the validation participant for each test participant. The evaluation metrics used for each training are accuracy and top-1 to top-14 accuracy scores. Additionnaly, to provide a comprehensive overview of the procedures employed throughout this study, we have depicted the general framework in Fig. 7.

C. MODELS ARCHITECTURES

For the CNN-AE, we stated previously that the shape of the instances is (500 × 552). This is due to the choice of using 10 seconds duration for an instance with three radar frames divided into 184 bins; one for each UWB radar. This input shape is consistent with a monochrome image, which renders the use of CNN-based deep learning architecture functional.

Fig. 8 illustrates the architecture of the CNN-AE, while Table. 2 shows the parameters of the model. We defined 2 successive convolutional layers for the encoding architecture, followed by a max pooling layer of stride 2. For the two first convolutional layers, a filter of 32 kernels of size 3 × 3 and strides of 2 has been set with a Rectified Linear Unit (ReLU) activation function. The output is then flattened and connected to a dense layer of size 800. This third layer is the constricted latent space, which is connected to the decoder. The CNN-AE decoder part has 3 deconvoluting layers set up with the same parameters as the encoding layers (filters of 32 kernels of size 3 × 3, strides of 2, and ReLU activations).

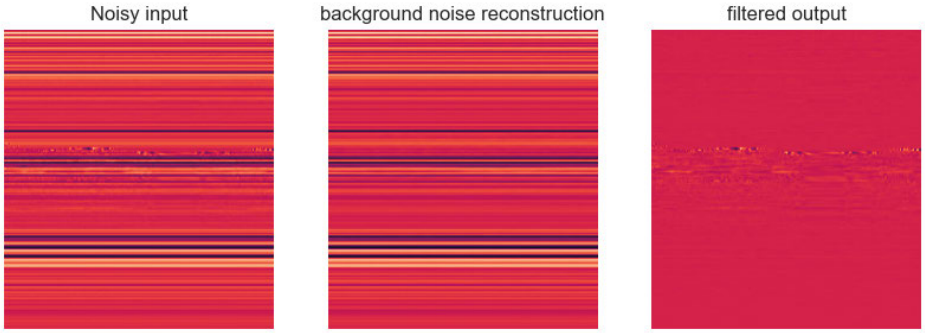

FIGURE 5. Comparison between a noisy input data frame, its background noise reconstructed by the CNN-AE, and the denoised output.

TABLE 1. Number of instances of shape (500 × 184 x 3) in used datasets for each validation participant.

	Number of instances		
Participant	Test	Validation	Training
1	1382	1380	23629
2	1389	1351	23651
3	1384	1411	23596
4	1366	1389	23636
5	1401	1382	23608
6	1382	1380	23629
7	1351	1389	23651
8	1392	1389	23610
9	1389	1384	23618
10	1380	1380	23631
11	1503	1372	23516
12	1380	1382	23629
13	1380	1382	23629
14	1372	1411	23608
15	1382	1384	23625
16	1382	1381	23628
17	1384	1382	23625
18	1411	1382	23598
19	1381	1372	23638

TABLE 2. Number of parameters for each layers of the CNN-AE.

Layer Type	Output Shape	Parameters
InputLayer	(None, 500, 552, 1)	0
ZeroPadding2D	(None, 500, 552, 1)	0
Conv2D	(None, 250, 276, 32)	320
Conv2D	(None, 125, 138, 32)	9248
MaxPooling2D	(None, 63, 69, 32)	0
Flatten	(None, 139104)	0
Dense	(None, 800)	111284000
Dense	(None, 139104)	111422304
Reshape	(None, 63, 69, 32)	0
Conv2DTranspose	(None, 126, 138, 32)	9248
Conv2DTranspose	(None, 252, 276, 32)	9248
Conv2DTranspose	(None, 504, 552, 1)	2049
SlicingOpLambda	(None, 500, 552, 1)	0
Total params:		222,736,417
Trainable params:		222,736,417
Non-trainable params:		0

The output layer is set to be the same dimensions as the input shape (500 × 522).

For the CNN classifier, Fig. 9 details the six convolutional layers that were used, while Table. 3 shows the parameters of the model. The shape of the input layer is (500 × 184 × 3), where the three radar frames are shaped in a format similar to an RGB image. Similarly to the CNN-AE input shape, this formatting allows the use of deep learning CNN-based architectures. A batch normalization layer has been inserted between almost every convolutional layer. The first convolutional layer has been set with a filter of 32 kernels of size 7 × 7 × 3, and a padding of zeroes with a ReLU activation function. With an input shape of (500 × 184 × 3), the output shape is (250 × 92 × 32). The second and third convolutional layers have been set with a filter of 64 kernels of size 5 × 5 × 32, a stride of 2, padding with zeroes, and the ReLU activation function. The output shape for each are respectively (125 × 46 × 64) and (63 × 23 × 64). The fourth, fifth, and sixth convolutional layers are set with the same parameters as the preceding layers, except for their filters of 128 kernels of size 3 × 3 ×64. The output shape of the last convolutional layer is (8 × 3 × 128). This output is then flattened in a vector of size 3072, which is then connected to two dense layers of size 200 and 100 with ReLU activation functions. The output is then finally connected to the softmax layer of 14 nodes, resulting in an architecture that can be used to classify our 14 activities.

V. RESULTS

We evaluated our denoising module on our dataset. The UWB radar frames have clutter and noise in part generated by the hardware constraint of the UWB radar, and in part generated by multiple reflections of electromagnetic waves. As seen in the raw data classification results, a classifier that uses the data without performing any denoising is significantly less accurate. The baseline classifier architecture trained on unfiltered data is evaluated using LOSO-CV. Its top-1 accuracy is

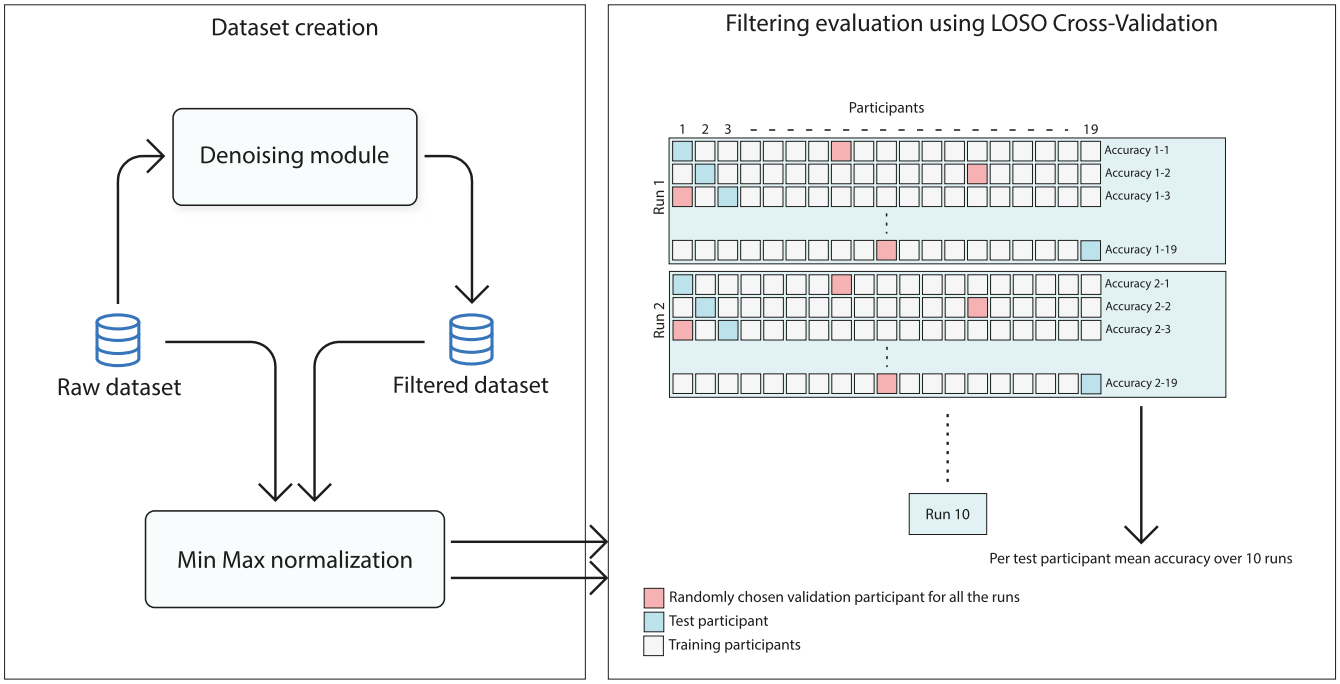

FIGURE 6. Methodology for LOSO cross-validation evaluation of the classifier's performance on the raw and filtered dataset.

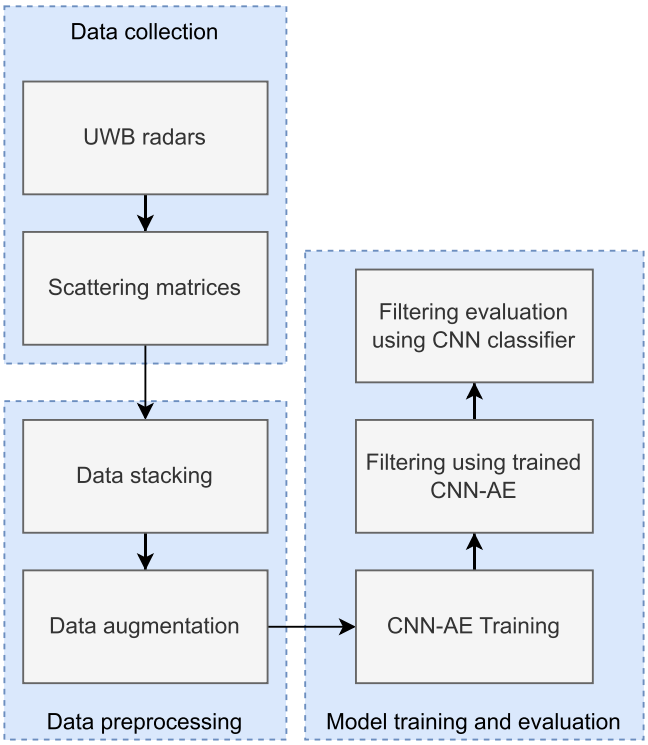

FIGURE 7. Overview of the procedures employed in this study.

48.35%, with a standard deviation of 0.115. The addition of the denoising module as a preprocessing step increases the results to an average of 69.93% (+21.58%) top-1 accuracy, with a standard deviation of 0.073 (−0.042). Fig. 10 shows the disparity in accuracy between the classifier trained on the raw dataset and the one trained on the filtered dataset for each participant, including per-participant distribution on the boxplot. For each test participant, the denoising module

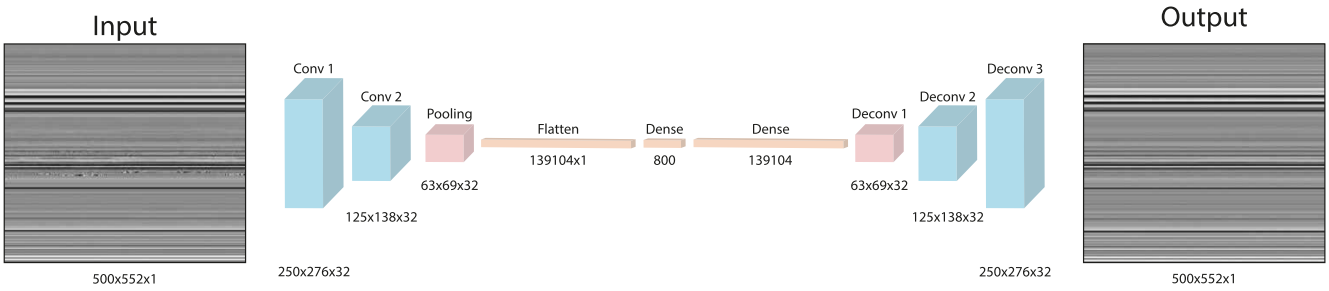

FIGURE 8. Architecture of the CNN-AE used to reconstruct background noise.

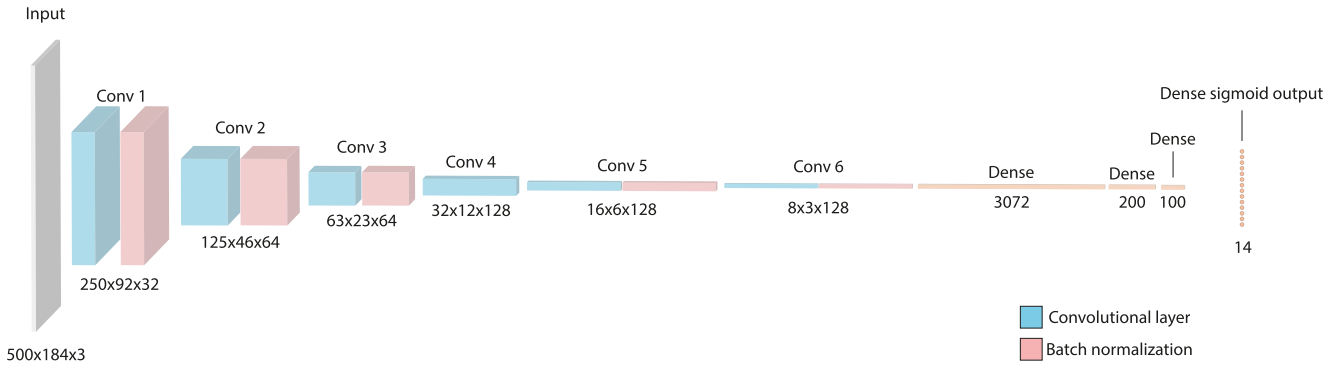

FIGURE 9. Architecture of the convolutional classifier used during evaluation.

TABLE 3. Number of parameters for each layers of the CNN classifier.

Layer Type	Output Shape	Parameters
InputLayer	(None, 500, 184, 3)	0
Conv2D	(None, 250, 92, 32)	4736
BatchNormalization	(None, 250, 92, 32)	128
Conv2D	(None, 125, 46, 64)	51264
BatchNormalization	(None, 125, 46, 64)	256
Conv2D	(None, 63, 23, 64)	102464
BatchNormalization	(None, 63, 23, 64)	256
Conv2D	(None, 32, 12, 128)	73856
Conv2D	(None, 16, 6, 128)	147584
BatchNormalization	(None, 16, 6, 128)	512
Conv2D	(None, 8, 3, 128)	147584
BatchNormalization	(None, 8, 3, 128)	512
Flatten	(None, 3072)	0
Dense	(None, 200)	614600
Dense	(None, 100)	20100
Dense	(None, 14)	1414
Total params:		1,165,266
Trainable params:		1,164,434
Non-trainable params:		832

TABLE 4. Improved classifier accuracy with CNN-AE filtered data compared to raw data.

	accuracy filtering gain (%)			
	n=5	n=10	n=15	n=19
top-1	+24.9	+27.3	+22.6	+21.6
top-2	+29.6	+26	+20	+19.4
top-3	+28.9	+25	+18.4	+18.1
top-4	+26.2	+23.7	+16.6	+16
top-5	+25	+20	+14.7	+14
top-6	+22.6	+16.9	+12.7	+11.9
top-7	+20.8	+14.1	+10.7	+9.8
top-8	+17.6	+11.4	+8.1	+7.6
top-9	+14.5	+8.7	+6.4	+6
top-10	+11.4	+6.3	+4.9	+3.9
top-11	+8	+4.5	+3.3	+2.7
top-12	+4.5	+3.3	+1.9	+1.5
top-13	+2.4	+1.5	+0.8	+0.6
top-14	+0	+0	+0	+0

consistently improves the classifier mean accuracy, often by a large margin. In addition, the accuracy spread is reduced, demonstrating that the classifier produces more constant results over several evaluation runs.

Additionally, we simulated an even lower dataset size by only selecting a few participants to assess the generalization capabilities of this denoising under very low dataset variability. To this effect, n random participants were chosen to create 3 additional subset datasets with $n \in \{5, 10, 15\}$. The chosen participants are the same for all datasets. As seen in Fig. 11, the denoising module is efficient in all cases, with a more pronounced effect the smaller the dataset. Even in an extreme case such as only using 5 participants (3 for training, 1 for test, and 1 for evaluation), the classifier that uses the denoised data still outperforms the classifier trained on the full dataset of 19 participants. This effect is summarized in Table. 4, where the accuracy gains when using the denoised dataset increase over all the top scores as the number of participants decreases. As an example, the difference in mean accuracy for the top-2 score is at +19.4% accuracy gain on the full dataset, and at +29.6% with n = 5 participants. In general, as the variability of the dataset decreases, the effect of the denoising module becomes more apparent.

Adding this CNN-AE architecture as a preprocessing step in HAR classification is an effective method to improve the

TABLE 5. Mean of classification metrics for all data subsets over all runs.

	RAW n5	RAW n10	RAW n15	RAW n19	AE n5	AE n10	AE n15	AE n19
FP	71.41	60.69	55.46	51.29	46.65	33.75	32.99	29.87
FN	71.25	59.64	55.17	51.21	46.23	33.02	31.88	28.98
TP	27.47	38.01	43.75	47.93	52.23	64.94	66.21	69.34
TN	1214.10	1222.46	1234.21	1238.50	1238.86	1249.16	1256.67	1259.91
TPR	0.28	0.39	0.44	0.48	0.53	0.66	0.67	0.71
TNR	0.94	0.95	0.96	0.96	0.96	0.97	0.97	0.98
FPR	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.02
FNR	0.72	0.61	0.56	0.52	0.47	0.34	0.33	0.29
Precision	0.28	0.39	0.44	0.48	0.53	0.66	0.67	0.70
Recall	0.28	0.39	0.44	0.48	0.53	0.66	0.67	0.71
F1-Score	0.28	0.39	0.44	0.48	0.53	0.66	0.67	0.70
Instances Number	6922	13816	20833	26391	6922	13816	20833	26391

FP=False Positive, FN=False Negative, TP=True Positive, TN=True Negative, TPR=True Positive Rate, TNR=True Negative Rate, FPR= False Positive Rate, FNR=False Negative Rate.

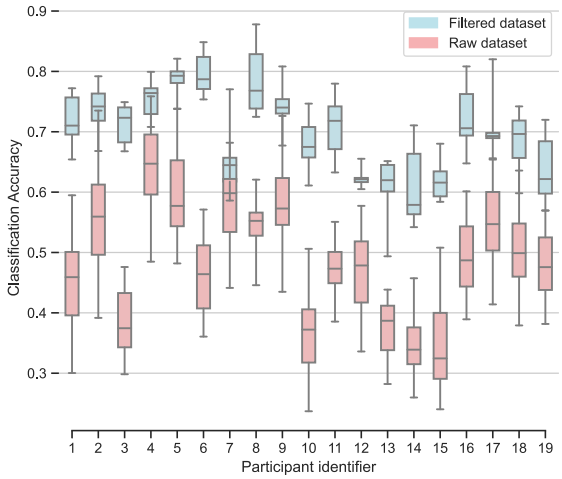

FIGURE 10. Classifier accuracy when trained on the raw dataset and filtered dataset. Distributions grouped by participants.

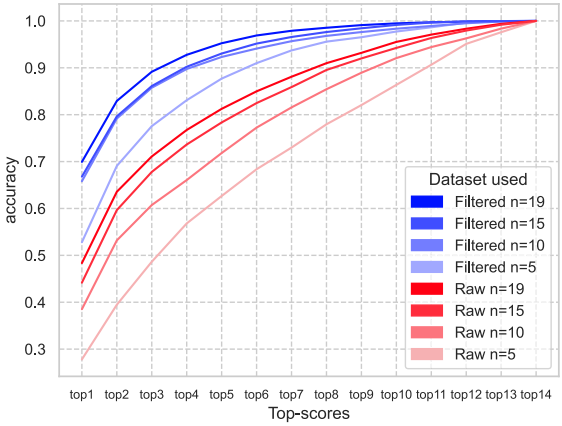

FIGURE 11. Classifier mean top accuracies when trained on the raw and filtered dataset, using a subset of n participants.

overall accuracy when labeled data is difficult to acquire. Our results show that unlabeled data can be used to improve a classifier's accuracy without the need to gather more labeled data.

Furthermore, to provide a comprehensive analysis, additional classification metrics are presented in Table. 5, where using bigger filtered datasets result in improved true positive and true negative rates, as well as reduced rates of false positives and false negatives.

VI. CONCLUSION

In this paper, we presented an unsupervised deep convolutional autoencoder used as a UWB radar scattering matrix denoiser. The restrictive part of the encoder guides the architecture to learn the specific background noise for each instance and subtracts it from the data.The training process involves using a sliding window over the slow time axis to create training instances, with an overlap of 90% for data augmentation. The instances are then min-max normalized, and the CNN-AE is trained on all instances using the mean squared error (MSE) as the loss function. This drives the architecture to learn the necessary features to reconstruct the instances.

The dataset and methodology used in this study were developed to evaluate this architecture using LOSO-CV to avoid participants' data contamination. A baseline pipeline and a denoised data pipeline were fed into a classifier, where its averaged accuracy serves as a differentiation metric to evaluate the model. Our results show that our denoising module performs under normal and low practical dataset sizes whilst also being easily trained due to its unsupervised nature.

Nevertheless, this approach has limitations. Namely, one limitation is that the dataset is collected in a controlled environment with a limited number of participants and activities. This may not fully represent the real-world scenarios and may not generalize well to different settings or people. Additionnaly, the denoising module quality may depend on the environment and not be as efficient when more noise or interferences are present. Finally, the unsupervised nature of the CNN-AE training may limit its ability to capture specific noise patterns related to different activities.

As for future work, the autoencoder could be optimized to maximize the accuracy gain. As an example, an automated hyper-tuner could use the classifier's accuracy result to fine-tune the latent space size and autoencoder architecture. Furthermore, it would be useful to investigate the incorporation of other equipment, such as thermography and

stereoscopic cameras, in conjunction with UWB radars. This approach would provide more comprehensive data for data fusion, enabling us to overcome the limitations of individual sensors, resulting in a more complete dataset.

REFERENCES

[1] *World Population Ageing 2019: Highlights*, Population Division, Dept. Econ. Social Affairs, United Nations, New York, NY, USA, 2019, p. 46.
[2] World Health Organization. (2022). *WHO Fact Sheet on Falls*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/falls
[3] R.-L. Heikkinen, "Growing older—Staying well: Ageing and physical activity in everyday life/prepared by Riitta-Liisa Heikkinen," WHO Ageing and Health Programme, 1998. [Online]. Available: https://apps.who.int/iris/handle/10665/65230
[4] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *Proc. 23th Int. Conf. Archit. Comput. Syst.*, Feb. 2010, pp. 1–10.
[5] S. Beltz, S. Gloystein, T. Litschko, S. Laag, and N. van den Berg, "Multivariate analysis of independent determinants of ADL/IADL and quality of life in the elderly," *BMC Geriatrics*, vol. 22, no. 1, p. 894, Nov. 2022, doi: 10.1186/s12877-022-03621-3.
[6] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. A. Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Appl. Sci.*, vol. 11, no. 2, pp. 1–18, Jan. 2021, doi: 10.3390/APP11020796.
[7] H. Chen, F. Xiong, D. Wu, L. Zheng, A. Peng, X. Hong, B. Tang, H. Lu, H. Shi, and H. Zheng, "Assessing impacts of data volume and data set balance in using deep learning approach to human activity recognition," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Kansas City, MO, USA, Nov. 2017, pp. 1160–1165, doi: 10.1109/BIBM.2017.8217821.
[8] D. Gholamiangonabadi, N. Kiselov, and K. Grolinger, "Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection," *IEEE Access*, vol. 8, pp. 133982–133994, 2020, doi: 10.1109/ACCESS.2020.3010715.
[9] K. Naik, T. Pandit, N. Naik, and P. Shah, "Activity recognition in residential spaces with Internet of Things devices and thermal imaging," *Sensors*, vol. 21, no. 3, p. 988, Feb. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/3/988, doi: 10.3390/s21030988.
[10] L. Köping, K. Shirahama, and M. Grzegorzek, "A general framework for sensor-based human activity recognition," *Comput. Biol. Med.*, vol. 95, pp. 248–260, Apr. 2018, doi: 10.1016/j.compbiomed.2017.12.025.
[11] A. Mannini and S. S. Intille, "Classifier personalization for activity recognition using wrist accelerometers," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1585–1594, Jul. 2019, doi: 10.1109/JBHI.2018.2869779.
[12] G. Mohmed, A. Lotfi, and A. Pourabdollah, "Employing a deep convolutional neural network for human activity recognition based on binary ambient sensor data," in *Proc. 13th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.* New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 1–7, doi: 10.1145/3389189.3397991.
[13] M. Gochoo, T.-H. Tan, S.-H. Liu, F.-R. Jean, F. S. Alnajjar, and S.-C. Huang, "Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and DCNN," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 693–702, Mar. 2019, doi: 10.1109/JBHI.2018.2833618.
[14] S.-D. Achirei, M.-C. Heghea, R.-G. Lupu, and V.-I. Manta, "Human activity recognition for assisted living based on scene understanding," *Appl. Sci.*, vol. 12, no. 21, p. 10743, Oct. 2022, doi: 10.3390/app122110743.
[15] M. Buzzelli, A. Albé, and G. Ciocca, "A vision-based system for monitoring elderly people at home," *Appl. Sci.*, vol. 10, no. 1, p. 374, Jan. 2020, doi: 10.3390/app10010374.
[16] Z. Luo, J.-T. Hsieh, N. Balachandar, S. Yeung, G. Pusiol, J. Luxenberg, G. Li, L.-J. Li, N. Downing, A. Milstein, and L. Fei-Fei, "Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring," *Proc. Mach. Learn. Res.*, vol. 85, pp. 1–18, Aug. 2018.
[17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision—ECCV 2014* (Lecture Notes in Computer Science), vol. 8693, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.
[18] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster R-CNN doing well for pedestrian detection?" in *Computer Vision—ECCV 2016* (Lecture Notes in Computer Science), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 443–457, doi: 10.1007/978-3-319-46475-6_28.
[19] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," 2017, *arXiv:1705.07750*.
[20] D. C. Luvizon, D. Picard, and H. Tabia, "2D/3D pose estimation and action recognition using multitask deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5137–5146, doi: 10.1109/CVPR.2018.00539.
[21] M. S. Momin, A. Sufian, D. Barman, P. Dutta, M. Dong, and M. Leo, "In-home older adults' activity pattern monitoring using depth sensors: A review," *Sensors*, vol. 22, no. 23, p. 9067, 2022, doi: 10.3390/s22239067.
[22] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, Jan. 2019, doi: 10.3390/rs11091068.
[23] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006, doi: 10.1109/TAES.2006.1603402.
[24] Y. Li, W. Wang, and Y. Jiang, "Through wall human detection under small samples based on deep learning algorithm," *IEEE Access*, vol. 6, pp. 65837–65844, 2018, doi: 10.1109/ACCESS.2018.2877730.
[25] D. Huang, C. Hou, Y. Yang, Y. Lang, and Q. Wang, "Micro-Doppler spectrogram denoising based on generative adversarial network," in *Proc. 48th Eur. Microw. Conf. (EuMC)*, Madrid, Spain, Sep. 2018, pp. 909–912, doi: 10.23919/EuMC.2018.8541507.
[26] S. S. Ram, S. Vishwakarma, A. Sneh, and K. Yasmeen, "Sparsity-based autoencoders for denoising cluttered radar signatures," *IET Radar, Sonar Navigat.*, vol. 15, no. 8, pp. 915–931, Aug. 2021, doi: 10.1049/rsn2.12065.
[27] T. Zheng, Z. Chen, S. Zhang, C. Cai, and J. Luo, "MoRe-Fi: Motion-robust and fine-grained respiration monitoring via deep-learning UWB radar," in *Proc. 19th ACM Conf. Embedded Netw. Sensor Syst.*, Coimbra, Portugal, Nov. 2021, pp. 111–124, doi: 10.1145/3485730.3485932.
[28] J. Yin, S. N. Tran, and Q. Zhang, "Human identification via unsupervised feature learning from UWB radar data," in *Advances in Knowledge Discovery and Data Mining* (Lecture Notes in Computer Science), D. Phung, V. S. Tseng, G. I. Webb, B. Ho, M. Ganji, and L. Rashidi, Eds. Cham, Switzerland: Springer, 2018, pp. 322–334, doi: 10.1007/978-3-319-93034-3_26.
[29] X. Li, X. Jing, and Y. He, "Unsupervised domain adaptation for human activity recognition in radar," in *Proc. IEEE Radar Conf. (RadarConf)*, Florence, Italy, Sep. 2020, pp. 1–5, doi: 10.1109/RadarConf2043947.2020.9266433.
[30] A. Abbasi and H. Liu, "Novel cascade CNN algorithm for UWB signal denoising, compressing, and ToA estimation," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 721–725, doi: 10.1109/CCWC51732.2021.9376176.
[31] M. Cavagnaro, S. Pisa, and E. Pittella, "Safety aspects of people exposed to ultra wideband radar fields," *Int. J. Antennas Propag.*, vol. 2013, Jan. 2013, Art. no. 291064, doi: 10.1155/2013/291064.
[32] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from UWB radars and deep learning," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113994.

VIRGILE LAFONTAINE received the dual M.S. degree in computer science from Université de Reims Champagne Ardenne and Université du Québec à Chicoutimi, in 2018, where he is currently pursuing the Ph.D. degree in computer science. During his M.S. studies he developed his interest in machine learning, which later prompted him to start the Ph.D. in the field of deep learning. His current work is the application of deep learning architectures on UWB radars to develop technologies for health. His research interests include data science and data mining in assistive technologies for human health.

KÉVIN BOUCHARD (Member, IEEE) received the Ph.D. degree in computer science, in 2014. He conduct a Postdoctoral Fellowship as a Project Scientist with the Center for SMART Health, University of California, Los Angeles. During this time, he co-led a deployment project of ambient technology for a Rehabilitation Center, Santa Monica. He worked on RFID localization, ambient sensing, activity recognition with UWB, and wearable technologies. He has been a Professor with Université du Québec à Chicoutimi, since August 2016. He has published more than 120 articles. He has been supervising/co-supervising more than 40 students. In addition to his academic work, he serves as the Vice President for a non-profit organization called Regroupement québécois des maladies orphelines, which supports individuals affected by rare diseases. His research interests include the exploitation of artificial intelligence and machine learning for the development of technologies for health.

JULIEN MAÎTRE received the M.Sc. degree in mechatronics from Université de Strasbourg, in 2013, and the Ph.D. degree in computer science from Université du Québec à Chicoutimi, in 2017. He is an expert in machine learning with Université du Québec à Chicoutimi, where he teaches data mining, machine learning, computer vision, and introduction to data science. His interest in the field of machine learning starts during the Ph.D. studies. His research interest includes recognize human activities from different sensors.

SÉBASTIEN GABOURY (Senior Member, IEEE) received the Ph.D. degree in mathematics from the Royal Military College of Canada, Kingston, ON, Canada, in 2012. He is currently a Full Professor and the Canada Research Chair of ambient intelligence and wearable devices for technology support in health with Université du Québec à Chicoutimi, Canada. He is the Head of the Ambient Intelligence Laboratory for the Recognition of Activities. His research is sponsored by the Canada Research Chair, the Natural Sciences and Engineering Research Council of Canada, the Quebec Research Fund on Nature and Technologies, and the Canadian Foundation for Innovation. He has published more than 200 articles and has supervised or co-supervised more than 40 graduate students. His research interests include assistive technologies for older adults, individuals with cognitive impairment, and those with neuromuscular diseases.