Identification, Activity, and Biometric Classification using Radar-based Sensing

Le Nguyen*, Constantino Álvarez Casado*, Olli Silvén*, Miguel Bordallo López*†
*Center for Machine Vision and Signal Analysis, University of Oulu
†VTT Technical Research Centre of Finland
Oulu, Finland
{le.nguyen, constantino.alvarezcasado, olli.silven, miguel.bordallo}@oulu.fi

***Abstract*—We explore the possibility of leveraging radar-based sensing systems to analyze vital signs for classification, user identification, and regression tasks. Specifically, we extract time-domain and frequency-domain features from distance, respiration, and pulse signals obtained by filtering radio-frequency signals. Our Random Forest classification models are trained on these features to recognize scenarios in which the radar data were collected, categorize individuals into age groups, and classify human activities. For classification, we achieved up to 94.7% of accuracy when distinguishing apnea and normal breathing in the lying position. We then show the feasibility of identifying individuals in a small group using vital signs, which can support model fine-tuning with data acquired from new users. Furthermore, we used a Random Forest regression model to estimate the Body Mass Index, height, and weight of subjects. These classification, identification, and regression models benefit smart systems that can simultaneously identify users, recognize their behaviours, and extract their vital signs from radar sensors.**

***Index Terms*—radar, vital signs, user identification, signal processing, classification**

I. Introduction

We are surrounded by radio-frequency (RF) signals originating from electronic devices with wireless communication capability (e.g. WiFi). The presence of persons along with their body movements considerably impacts wireless signals in both amplitude and phase. This effect can be analyzed to develop a sensing modality for human monitoring systems: RF sensing, which is mostly based on the Received Signal Strength Indicator (RSSI), Channel State Information (CSI), Frequency Shift for Frequency Modulated Carrier Wave (FMCW), and Doppler Shift [1]. The analysis of RF signals enables device-free human sensing systems that offer localization, identification, activity recognition, and vital sign measurement. Although these systems have achieved significant successes with both commodity hardware (e.g., WiFi cards) and specialized equipment (e.g. radar sensors), there still exist open problems that are actively investigated, such as: dynamic environments, human movement, and multi-user sensing.

In dynamic environments, wireless signals are usually disturbed by the presence of other wireless devices and the reflection of obstacles. This leads to the development of signal processing and machine learning methods to improve the reliability of the measurements. In human movements, the classification and estimation of human actions are challenging, especially when the activities are dynamic and performed at different speeds. Finally, multi-user sensing is a challenging task due to the high variation in the received signals, which could result in a significant degradation in the performance of the systems.

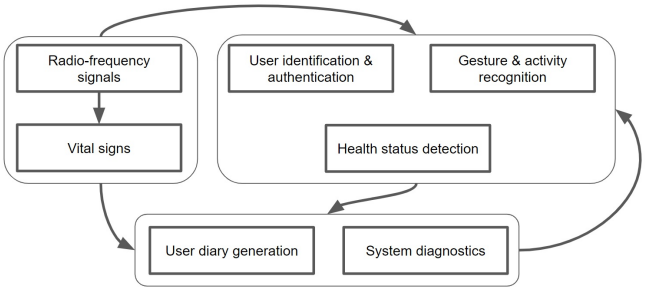


Fig. 1. Modules of a vital sign sensing and analysis system based on radio-frequency signals

In this work, we propose to process the RF signals for both vital sign measurement and context recognition, including biometrics, activities, and identification, among others. The benefit of this approach is two fold: (i) it provides insights of vital signs and the corresponding contexts (such as in smart-home systems [2]), and (ii) it facilitates a logging mechanism, which can help us to detect any issues during the sensing (such as artifacts caused by body movements) and to assess the quality of estimated vital signs. Figure 1 illustrates the modules in a smart system that processes radio-frequency signals to extract vital signs and provides applications (e.g., user identification and activity recognition) to users. Such a system can also monitor its functions and users' routines. We vision a system that can leverage communication signals to simultaneously perform multiple tasks: identifying users, estimating vital signs, recognizing users' activities, monitoring its own operations, etc.

We leverage different biosignals extracted from radar data to classify the different scenarios depicted during data collection, the age group of the subjects, and their activities (see Section V-A). We evaluated Random Forest classifiers [3] on three radar datasets, Lying [4], Sitting [5], and Children [6],

This research has been supported by the Academy of Finland 6G Flagship program under Grant 346208 and PROFI5 HiDyn under Grant 32629, and the InSecTT project, which is funded under the European ECSEL Joint Undertaking (JU) program under grant agreement No 876038.

which were originally utilized to evaluate the measurement of vital signs.

These classification and regression models can be integrated into smart-home systems to simultaneously measure vital signs and recognize different user behaviors and characteristics. Compared to camera-equipped systems, the RF-based approach is less intrusive since it does not directly access and analyze the visual data of the users. Some RF-based systems can work in non-line-of-sight scenarios depending on the penetration characteristics of radio waves [7]. Compared to wearable sensors, an RF-based system does not require users to attach any device [8] to their bodies. Hence, it eliminates the issue of forgetting to carry or charge the sensors. Furthermore, the RF-based sensing capability can be fused into the communication channel. Our work allows the implementation of multiple tasks through analyzing the RF signals: estimating vital signs, recognizing users' activities, identifying users, characterizing the user's biometrics, and other contextual information. The results of these tasks are aggregated to perform further analysis. For example, we can generate a diary of individuals or monitor the system's operations.

In the experiments, we demonstrated the feasibility of identifying individuals, especially for small groups, e.g., 4-6 subjects (see Section V-C). The identification module is essential in systems that monitor multiple users in a shared space without relying on wearable devices. In addition, we trained a Random Forest regression model to estimate the Body Mass Index (BMI) [1] of subjects using radar-sensing vital signs (see Section V-D).

II. Related Work

RF-based sensing facilitates contactless solutions to implement human monitoring systems, including localization, identification, activity recognition, and vital sign measurement. Compared to other modalities, such as cameras or wearable sensors, RF-based sensing can operate in non-line-of-sight scenarios [7], and it does not require on-body devices [8].

The environment and humans can influence radio signals via reflection, refraction, diffraction, and absorption. By constructing an RSSI model in an area covered by a MICA2 sensor network, Zhang *et al.* [9] localized a single moving object that did not carry any RF devices. In a more challenging setting with multiple objects, Xu *et al.* [10] introduced a sequential algorithm to detect, count, and track them.

The characteristics of wireless signals propagating around the human body depend on several factors, including body composition and movements [11]. Zhang *et al.* [12] analyzed unique perturbations in the WiFi CSI (at 2.4 GHz and 5 GHz frequency bands), which were caused by the diversity of human gait. They extracted features from the spectrogram representation to identify users. The effect of anthropometric measurements (e.g., height and weight) on signal reflection was also leveraged to build a device-free person re-identification method [13]. This system utilized an FMCW radio operating at frequencies between 5.4 and 7.2 GHz and could work in the presence of occlusions and low-light scenarios. Lin *et al.* [14] utilized a continuous-wave radar to implement a cardiac motion-sensing system. They extracted geometric and non-volitional features of the cardiac motion from the radar signals to develop a continuous authentication method.

In wireless communication, CSI describes signal propagation with the effects of time delay, amplitude attenuation, and phase shift. Based on human interference in the wireless signal, Han *et al.* [15] implemented a device-free fall detection system that found and classified abnormal CSI series. To classify human activities, Sigg *et al.* [8] extracted time and frequency domain features from RSSI of ambient RF signals (e.g., 82.5 MHz FM Radio).

The implementation of RF-based vital sign measurement methods is based mainly on the reflection of the signal from a stationary person [16]. Small changes in the wireless signals correlate with subtle movements in the human body caused by breathing and cardiac activity. Various radio hardware, from commodity WiFi to mmWave sensors, has been used for measuring vital signs such as breathing volume [17], respiration and heart rates [18], and cardiac motion [14]. These systems assumed the scenario of one stationary human subject at a close distance from the transmitter and receiver. Usually, human presence is detected before vital signs are estimated [19]. Monitoring multiple persons is another challenge of these RF-based sensing systems. Ahmad *et al.* [20] leveraged range-gating and beamforming and multiple receiving channels to separate objects. In addition, body movements or speaking may influence the estimation of vital signs [6]. The context (e.g., body orientation, movements, and antenna positioning) can influence the operations of RF-based vital sign sensing systems [21]. Hence, when extracting vital signs, we propose classifying human activities and scenarios in which sensing data is collected to understand these systems' operations better.

III. Description of the Datasets

We perform experiments on three different publically available datasets, containing both radar signals and synchronized reference signals acquired with medical devices. Although the initial objective of these datasets is to extract vital parameters, all of them are stratified based in different biometric characteristics and scenarios, a feature that could be exploited to perform different classification and regression tasks. The description of the datasets is depicted below.

A. Lying dataset

We called the first dataset from Schellenberger *et al.* [4], the *Lying* dataset because each subject was positioned on a tilt table. There were 30 healthy subjects, including 16 females and 14 males. The radar system was a six-port interferometer. The RF signal was generated by a Keysight PSG Analog Signal Generator E8257D at a frequency of 24.17 GHz. The reference system was the Task Force Monitor 3040i

[1]Body mass index - BMI: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

from CNSystems Medizintechnik GmbH, which could collect electrocardiogram (ECG), impedance cardiography (ICG), and continuous blood pressure (BP).

The radar data was collected from the subjects in four scenarios: Resting when the participants breathed calmly, Valsalva-maneuvre with pauses, apnea when the subjects hold their breath as long as possible, Tilt-up when the table was raised, and Tilt-down which was the opposite of Tilt-up. The tilt-up and tilt-down scenarios triggered the autonomic nervous system of the subjects; therefore, they might affect the heart and respiration rates.

B. Children dataset

The second dataset was provided by Yoo *et al.*'s [5]. It contains data from 50 children (24 boys and 26 girls) aged less than 13 years. Hence, we called it *Children* dataset.

In this dataset, each participant was sitting in a chair, and if the participant was under six years old, a child car seat was used. The radar system was equipped with a Texas Instruments IWR6843 mmWave sensor, operating at 60.25 GHz. A Nihon Kohden BSM6501K patient monitor system was utilized as the reference sensor, which provided heart rate and respiration rate data in the corresponding signal waveforms. In addition to the measurement of vital signs, they used the data to perform age group classification (biometric information): children under three years old in Group 1 (Under-3), children aged from three to less than six years old in Group 2 (3-To-Under-6), children aged six to less than nine years old in Group 3 (6-To-Under-9), and children aged nine to less than 13 years old in Group 4 (9-To-Under-13).

C. Sitting dataset

The third dataset was collected by Shi et al.'s [6], and we called it *Sitting* dataset. It utilized the same 24-GHz radar system as the first one [4].

This dataset includes 11 sitting subjects (seven male and four female) that were measured in different scenarios (normally breathing, apnea, breathing after sport, and breathing while speaking) and in various measurement positions (e.g., at the carotid, on the back, and several frontal positions on the thorax). The reference sensors included a digital stethoscope, an electrocardiography (ECG) monitor, and a respiration sensor. Figure 2 visualizes data from one participant in the dataset. We used the radar data of this dataset and its stratification of users to perform activity recognition.

IV. Methodology

We extracted features from physiological signals acquired with radar sensors. These features are composed of statistical, morphological, physiological, and frequency-domain representations. These features are subsequently used to feed machine learning-based classification and regression models to infer certain information about the users. Figure 3 shows the pipeline to implement the tasks in this work.

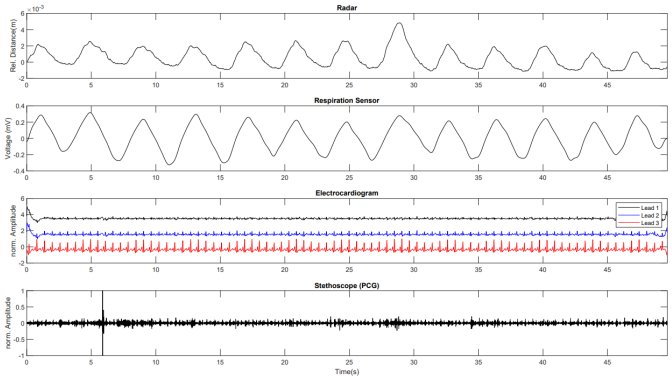


Fig. 2. Processed radar data from one person in the *Sitting* dataset [6], reproduced from [6]

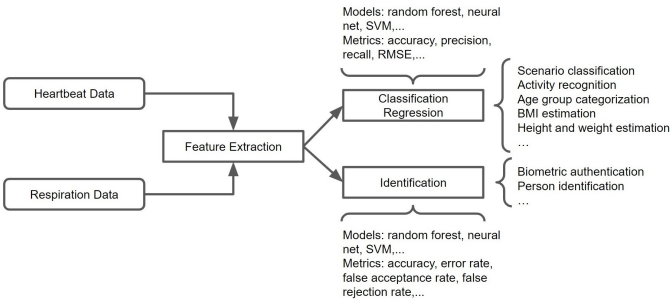


Fig. 3. A general process of utilizing radar-based vital sign data to implement classification, regression, and identification models

A. Feature extraction

The distance radar data is preprocessed in the original datasets to obtain respiration and heart signals by accommodating it to different frequency bands. The original datasets have used appropriate band-pass filters of 0.05 Hz ... 1.7 Hz for the respiration rate range, 0.7 Hz ... 15 Hz for the pulse frequency range, and 16 Hz ... 80 Hz for the heart sound frequency range when available. This detailed process is described in each corresponding dataset specifications [4] [5] [6].

Before feature extraction, we further filter the heart signals to match the typical frequency bands of the heart pulse. We filter the signal using a Butterworth band-pass filter of order 5, using a 0.65-4.0 Hz band. All three signals (distance radar, respiration, and filtered heart) are then segmented into sliding windows of 30 seconds with a displacement step of 0.25 seconds between consecutive windows.

From each window from all three signals, in order to form a high-dimensional feature vector, we extracted a total of 39 features (for a total 117) including statistical, morphological and physiological features, both in frequency and time domains. In particular, the statistical features include the *mean, min, max, std, dynamic range* and four *percentiles* (10, 25, 75 and 90) [22]. The fractal analysis features include the *Katz fractal dimension, Higuchi fractal dimension* and *detrended fluctuation analysis* of the entire window, and the mean of the three fractal analysis features computed in sub-windows of 2 seconds of the whole window. The entropy analysis features

include *permutation entropy, spectral entropy, approximate entropy, sample entropy, Hjorth mobility and complexity* and *number of zero-crossings* of the entire window. The heart and heart rate variability (HRV) related features include heart rate (HR), breathing rate (BR), interbeat interval (IBI), differences between R-R intervals (pNN20, pNN50), Poincare analysis, frequency domain components (Very-Low-Frequency, Low-Frequency, High-Frequency, ratio of Low-Frequency to High-Frequency), the standard deviation of N-N intervals (SDNN), among others [23]. To compute them, we use the *Numpy Python* library to compute the statistical features, the *AntroPy Python* package, a software tool for computing the complexity of time-series, to extract both fractal and entropy features [24] and two Python libraries, namely *Neurokit2* [25] and *HeartPy* [26], to compute HRV related features.

B. Classification and regression models

We utilized Random Forest classification models [3] to recognize scenarios, classify activities, and categorize age groups. Random Forest is an ensemble learning method whose outputs are the combined results of a set of decision or regression trees. Each tree is constructed from a subset of samples during the training time, and the training algorithm works on a subset of features. Random forest models are also used for the identification of users. A Random Forest regression model is a modification of the random forest classification model that is based on an ensemble of regression trees. It outputs the average prediction from all individual trees. Random Forest Regressors are used in all three regression tasks.

The input of a trained Random Forest model RF is a feature vector $v \in \mathbb{R}^k$ described in Section IV-A, where k is the number of features. In this work, depending on the tasks, the output of RF classifiers can be a scenario $s \in \{\text{Resting, Valsava, Apnea, TiltUp, TiltDown}\}$ [4], an activity $a \in \{\text{Normal, Apnea, Sport, Speech}\}$ [6], an age group $g \in \{\text{Under-3, 3-To-Under-6, 6-To-Under-9, 9-To-Under-13}\}$ [5]. For the identification task, the output of the model is the identity of an individual i. For the regression tasks, the outputs are the estimation the Body Mass Index of human body: $\text{BMI} = \frac{w}{h^2}$, and individual values for the weight in kilograms w, and the height in meters h.

V. Experiments and Results

According to the dataset protocols described before, we evaluated the classification, identification, and regression tasks in all publicly-available datasets of radar signals. We performed the experiments in the *Leave-One-Group-of-Subjects-Out (LOGSO)* setting. We randomly selected 80% (approximately) of the users for training and used the remaining data for testing.

Specifically, in the *Lying* dataset [4], 24 subjects were in the training set for scenario classification, 40 subjects in the *Children* dataset [5] were in the training set for age group categorization, and 9 subjects in the *Sitting* dataset [6] were in the training set for activity recognition. We repeated the experiment five times for each dataset, i.e., we used 5-fold cross-validation and reported the mean accuracy and, in some cases, the variation between folds.

A. Classification of scenarios

The scenario classification tasks can support applications such as recognition of positioning for smart-homes [2] or patient monitoring [27]. The scenario classification task was performed on the *Lying* [4] dataset.

The labels are scenarios $s \in \{\text{Resting, Valsava, Apnea, TiltUp, TiltDown}\}$. Since this dataset contains more diverse subjects, 30 subjects at different age ranges, we have experimented with different class divisions. First, we show the classification results for 5 classes, depicting all scenarios. The results of this classification task can be seen in Figure 4.

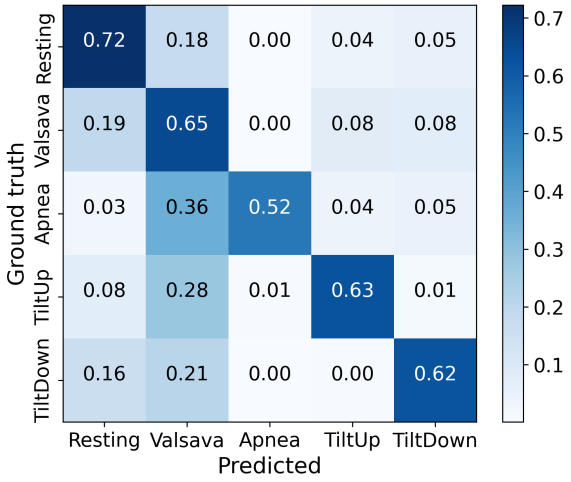


Fig. 4. Scenario classification for 5-class configuration [4]

In addition, we perform two binary classification tasks. In the first one, we categorized the scenarios into normal and abnormal cases. The results of the classification can be seen in Figure 5. In addition, we classify the apnea scenario that depicts an abnormal breathing that can occur spontaneously, against all normal scenarios. We show these classification results in Figure 6. The results show that we can recognize different scenarios with a moderate but practical accuracy of about 65.5%. When detecting more critical scenarios, such as the binary classification of normal vs. abnormal respiration, the accuracy rises to nearly 80%. For the most vital case, recognizing the irregular apnea respiration compared to other normal scenarios, the classification accuracy is almost 95%. This result is acceptable for most of the possible use cases.

B. Other classification tasks

In addition to breathing scenarios, we perform two complementary classification tasks: age grouping on the *Children* dataset [5] and activity recognition on the *Sitting* dataset [6].

For the *Children* dataset [5], we have classified them into different age groups. The results can be seen in Figure 7). The labels were the age groups $g \in \{\text{Under-3, 3-To-Under-6, 6-To-Under-9, 9-To-Under-13}\}$. Children under three years of age were in Group 1 (Under-3).

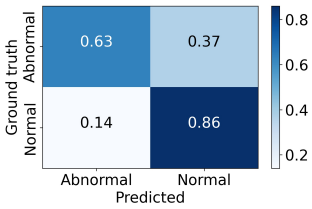


Fig. 5. Scenario classification for Normal vs Abnormal scenarios [4]

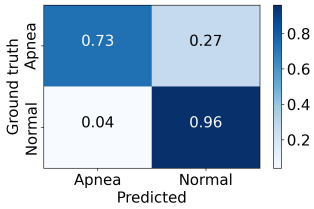


Fig. 6. Scenario classification for Apnea vs Normal scenarios [4]

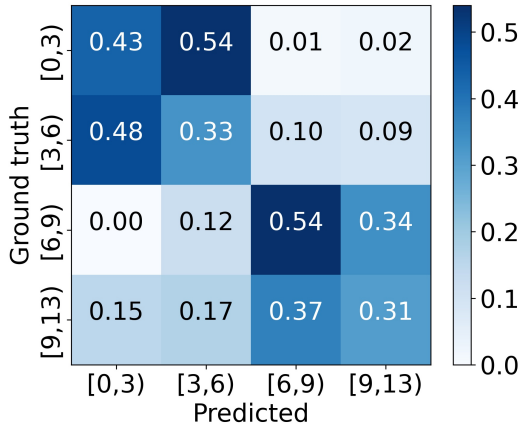


Fig. 7. Age group classification [5]

Children aged from three to less than six years old were in Group 2 ((3-To-Under-6). Children aged six to less than nine years old were in Group 3 (6-To-Under-9). The remaining were in Group 4 (9-To-Under-13).

The activity recognition task was experimented with the *Sitting* dataset [6] dataset (see Figure 8). In this datasets, the depicted activities are $a \in \{\text{Normal}, \text{Apnea}, \text{Sport}, \text{Speech}\}$. The classification results can be seen in Figure 8.

The results show that different classification tasks are possible with moderated accuracy. Although they could be useful for certain use cases, further study to obtain better classification results is still needed. However, it is known that

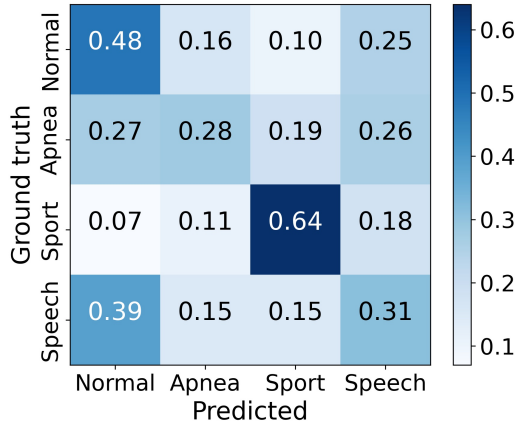


Fig. 8. Activity classification [6]

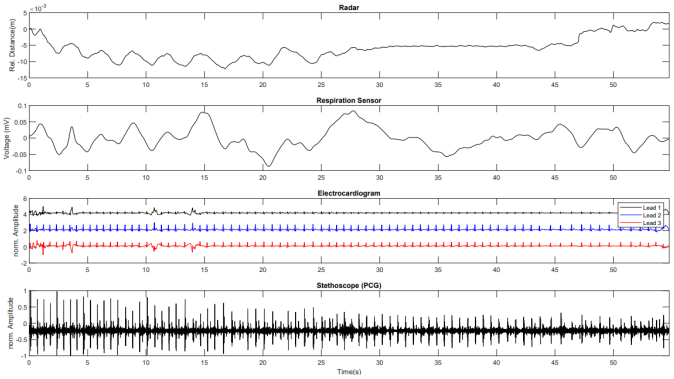


Fig. 9. Annotation issue in the *Sitting* dataset [6]: the apnea segment may contain normal breathing signals

the annotation quality can influence the accuracy of the classification models. We observed that some segments in the *Sitting* dataset [6] were labeled as *Apnea* but contained *Normal-breathing* signals. For example, we visualized such a segment in Figure 9: the first half was similar to normal respiration while the second half was the actual apnea. We argue that some of these annotation errors would explain the comparatively lower accuracies for these tasks.

C. Identification

The identification task aims to recognize individuals in a group, which is useful when multiple users share a common place (e.g., an apartment or office [2]). We used the Random Forest classifier to identify the users in the group. We split the sequence data into two parts in the middle for each subject: the first half for training and the second for testing. Figure 10 summarize our results.

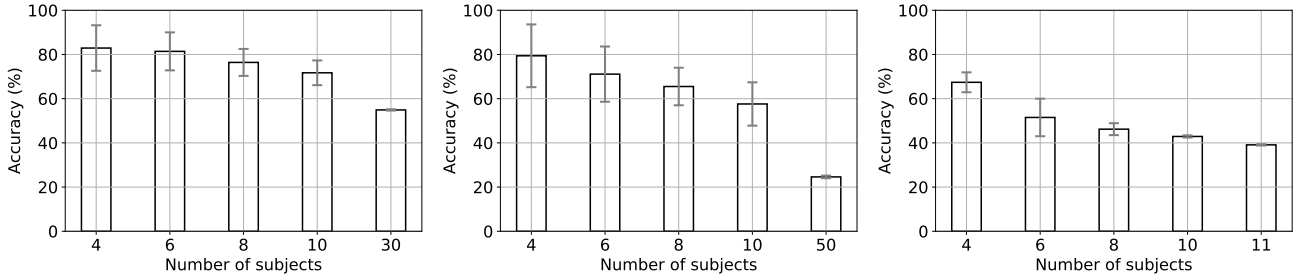
Following the procedure of Zhang *et al.* [12], we evaluated the identification task on all three datasets and varied the number of subjects to explore the limit of the identification models. On the Schellenberger *et al.*'s *Lying* dataset [4] (see Figure 10a), we varied the number of subjects to be identified $n \in \{4, 6, 8, 10, 30\}$. Note that $m = 30$ is the total number of subjects in the dataset [4].

Similarly, we varied the number of subjects to be identified $n \in \{4, 6, 8, 10, 50\}$ in the *Children* dataset [5] (see Figure 10b), in which $m = 50$ is the total number of subjects. Lastly, Figure 10c displayed the identification result on the *Sitting* dataset [6]. In this case, we varied the number of subjects to be identified $n \in \{4, 6, 8, 10, 11\}$, since there were only $m = 11$ participants.

We conclude that it is feasible to identify users using radar data in a small group (e.g., 4-6 users). As expected, the task becomes more challenging when there are more users. One possible direction to tackle the challenge is to develop specific models for a group of users based on their body characteristics (e.g., age, height, and weight).

D. Estimation of BMI values

Human body characteristics, such as the BMI, correlate with the micro-Doppler signature of the subject [28], and it is



(a) *Lying* dataset [4] (b) *Children* dataset [5] (c) *Sitting* dataset [6]

Fig. 10. Identification accuracy when the number of subjects is varied

TABLE I
METRICS OF THE ESTIMATED BMI VALUES

	MAE	MAPE	RMSE
Lying [4]: 30 subjects	3.75 (1.04)	0.16 (0.04)	4.82 (1.3)
Children [5]: 50 subjects	2.16 (0.4)	0.13 (0.02)	2.8 (0.6)
Sitting [6]: 11 subjects	3.67 (1.2)	0.18 (0.07)	4.23 (1.13)

expected that a radar signal should be able to discern it to some extent. In this context, we trained a Random Forest regressor to predict the BMI values of the subjects in all three datasets in a *Leave-One-Group-of-Subjects-Out (LOGSO)* setting. The model was trained with 80% of the subjects and the remaining data was used for testing. We repeated the experiment five times and collected the predicted BMI values of all subjects, in a typical 5-fold validation scheme.

Table I summarizes the metrics of the regression model, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE), while the numbers in parentheses are the standard deviation.

In our results, we obtain an MAE for the *Children* dataset, whose subjects were all healthy, which is 2.16. This is in line with recent results in similar datasets that are calculated using different modalities, such as facial images with an MAE of 2.23 [29], or whole-body images with an MAE of 1.66 [30]. Results with wearable motion sensors in normal persons also show similar errors, ranging from MAEs of 1.701 to 2.181 [31]. We visualized the mean predicted BMI values for all subjects of the three datasets in Figure 11, Figure 12, and Figure 13.

VI. ADDITIONAL EXPERIMENTS

In addition to classification, identification, and regression tasks, we have studied the impact of using different window sizes and machine learning models on the total accuracy. We perform these complementary studies in the classification tasks of the *Lying* dataset [4].

1) Experiments on Window Size: Since we extracted features from each sliding window of the radar data, the window size can influence the results. Table II shows the results with varying window sizes. Since the *Lying* dataset is unbalanced (e.g., apnea samples are fewer than others), we also calculated the balanced accuracy [32].

TABLE II
SCENARIO CLASSIFICATION ACCURACY IN VARIOUS SETTINGS [4]

Settings	Window (s)	Accuracy	Acc. (balanced)
5 classes	8	61.9% (4.9%)	57.5% (5.1%)
5 classes	15	62.8% (4.7%)	59.1% (4.9%)
5 classes	30	**65.5% (5.1%)**	62.8% (4.6%)
Apnea vs Normal	8	93.1% (2.7%)	75.2% (4.1%)
Apnea vs Normal	15	93.4% (3.3%)	79.0% (3.1%)
Apnea vs Normal	30	**94.7% (3.1%)**	84.7% (2.6%)
Abnormal vs Normal	8	72.5% (2.8%)	64.8% (2.2%)
Abnormal vs Normal	15	74.0% (3.6%)	67.5% (1.2%)
Abnormal vs Normal	30	**79.9% (5.4%)**	74.5% (3.1%)

The optimal result was obtained with the window size of 30 seconds. However using shorter windows such as, e.g., 8 seconds, does not degrade the performance in a very significant manner.

2) Experiments on different classification algorithms: For comparative purposes, we performed the scenario classification on 23 different machine learning models, as implemented in scikit-learn [2]. We summarized the experimental results in Table III. From the results, it can be seen that the Random Forest Classifier model [3] achieved the highest accuracy (65.5%) for the 5-class configuration [4], while its training time is still reasonable.

A. Regression of height and weight

We performed additional experiments to estimate the height and weight of subjects in the *Lying* dataset [4], using a Random Forest regression model. BMI, height, and weight are important measurements to assess health conditions. We trained the regression model on 24 subjects and estimated the height and weight of the remaining ones. Then, we calculated the BMI values from the estimated height and weight. The error metrics are showed in Table IV. The calculated BMI (computed from the estimated height and weight) and the estimated BMI (directly regressed from vital signs) were compared to the ground-truth BMI.

VII. CONCLUSION

We implemented user identification, scenario and activity classification, and biometric feature regression using physiological signals collected with radar sensors [4] [5] [6]. We

[2]https://scikit-learn.org/stable/

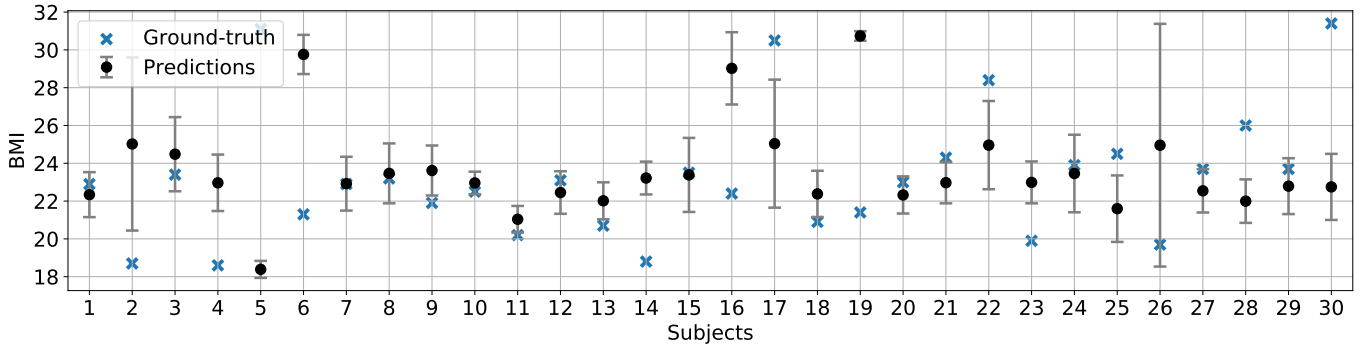


Fig. 11. Predicted BMI values of the *Lying* dataset [4]

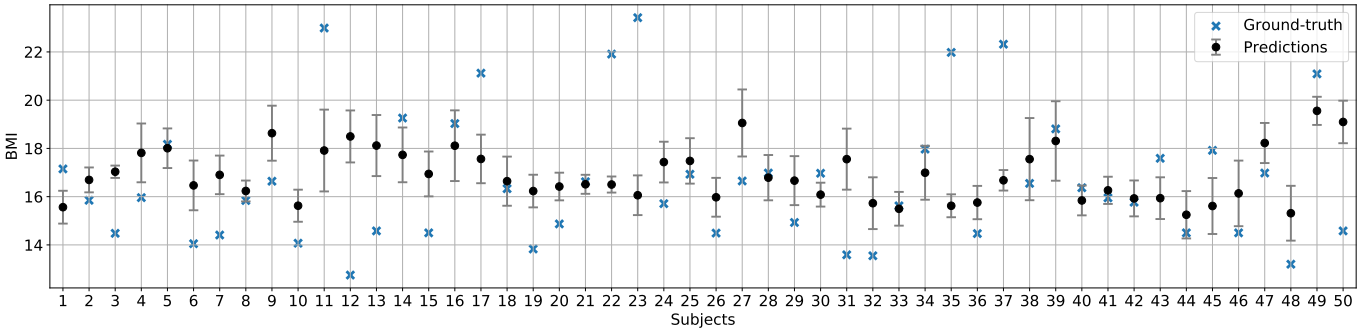


Fig. 12. Predicted BMI values of the *Children* dataset [5]

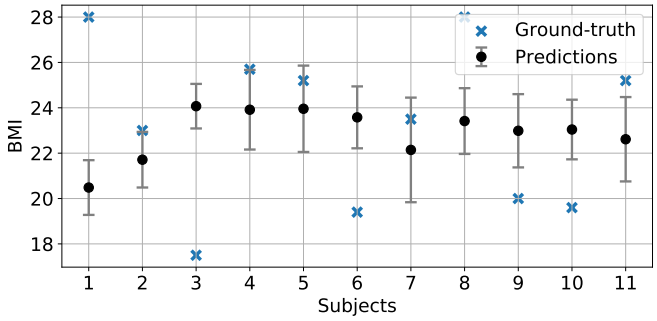


Fig. 13. Predicted BMI values of the *Sitting* dataset [6]

leveraged both time-domain and frequency-domain, morphological, physiological, and statistical features (117 features) with the widely used Random Forest models.

For scenario and activity classification, we show that using *Leave-One-Group-of-Subjects-Out (LOGSO)* validation, we were able to recognize among the five scenarios with a reasonable accuracy, which increased for critical cases such as recognizing abnormal vs. normal respiration.

For user identification, we showed that we could identify an individual using only radar data in controlled environments. We offer the results for 4, 6, 8, 10, and the total number of subjects in each dataset. Although the accuracy decreases when the number of individuals increases, the result is useful with reasonably-sized groups.

We argue that training or fine-tuning of subject-specific models is still needed for accurate context classification. In this context, users' identification and body characteristics estimation could support the collection of new data and the creation of personalized models. In particular, we showed the

TABLE III
SCENARIO CLASSIFICATION ACCURACY IN VARIOUS ALGORITHMS [4]

Classifiers	Acc	Time (s)
AdaBoostClassifier	0.46	27.05
BaggingClassifier	0.47	37.25
BernoulliNB	0.37	0.32
CalibratedClassifierCV	0.50	383.88
DecisionTreeClassifier	0.43	6.21
DummyClassifier	0.34	0.2
ExtraTreeClassifier	0.39	0.28
ExtraTreesClassifier	0.53	7.65
GaussianNB	0.30	0.34
KNeighborsClassifier	0.40	13.30
LGBMClassifier	0.52	2.66
LinearDiscriminantAnalysis	0.49	1.18
LinearSVC	0.50	103.03
LogisticRegression	0.49	2.32
NearestCentroid	0.34	0.26
PassiveAggressiveClassifier	0.40	1.01
Perceptron	0.40	0.85
QuadraticDiscriminantAnalysis	0.31	0.48
RandomForestClassifier	**0.65**	**36.97**
RidgeClassifier	0.48	0.30
RidgeClassifierCV	0.48	0.70
SGDClassifier	0.48	3.85
SVC	0.48	272.90
XGBClassifier	0.54	53.38

TABLE IV
ESTIMATION OF HEIGHT & WEIGHT AND CALCULATION OF BMI [4]

	Height (cm)	Weight (kg)	Calculated BMI	Estimated BMI
MAE	9.15 (4.54)	14.93 (8.0)	3.72 (2.56)	3.75 (1.04)
MAPE	0.05 (0.02)	0.22 (0.15)	0.16 (0.10)	0.16 (0.04)
RMSE	9.80 (4.47)	15.76 (7.6)	4.02 (2.47)	4.82 (1.3)

potential of Random Forest models to identify users in a small group and estimate their BMI values and other biometric traits.

Our results show that the classification of the context using only radar signals is feasible, even when using datasets not directly intended for these tasks that are prone to annotation problems. Future studies should focus on incorporating better machine learning models and preprocessing schemes, possibly more tolerant to label noise.

ACKNOWLEDGEMENTS

This research has been supported by the Academy of Finland 6G Flagship program under Grant 346208 and PROFI5 HiDyn program under Grant 32629, and by the InSecTT project under the European ECSEL Joint Undertaking (JU) program grant No 876038.

REFERENCES

[1] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Comm. Surveys & Tutorials*, 2020.

[2] C.-Y. Hsu, R. Hristov, G.-H. Lee, M. Zhao, and D. Katabi, "Enabling identification and behavioral sensing in homes using radio reflections," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[3] L. Breiman, "Random forests," *Machine learning*, 2001.

[4] S. Schellenberger, K. Shi, T. Steigleder, A. Malessa, F. Michler, L. Hameyer, N. Neumann, F. Lurz, R. Weigel, C. Ostgathe, *et al.*, "A dataset of clinically recorded radar vital signs with synchronised reference sensor signals," *Scientific data*, 2020.

[5] S. Yoo, S. Ahmed, S. Kang, D. Hwang, J. Lee, J. Son, and S. H. Cho, "Radar recorded child vital sign public dataset and deep learning-based age group classification framework for vehicular application," *Sensors*, 2021.

[6] K. Shi, S. Schellenberger, C. Will, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelpin, "A dataset of radar-recorded heart sounds and vital signs including synchronised reference sensor signals," *Scientific data*, 2020.

[7] F. Adib and D. Katabi, "See through walls with wifi!," in *Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)*, 2013.

[8] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Trans. on Mobile Computing*, 2014.

[9] D. Zhang, J. Ma, Q. Chen, and L. M. Ni, "An rf-based system for tracking transceiver-free objects," in *IEEE International Conference on Pervasive Computing and Communications*, 2007.

[10] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *International Conference on Information Processing in Sensor Networks*, 2013.

[11] Q. Xu, Y. Chen, B. Wang, and K. J. R. Liu, "Radio biometrics: Human recognition through a wall," *IEEE Transactions on Information Forensics and Security*, 2017.

[12] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2016.

[13] L. Fan, T. Li, Y. Yuan, and D. Katabi, "In-home daily-life captioning using radio signals," in *European Conference on Computer Vision*, 2020.

[14] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac scan: A non-contact and continuous heart-based user authentication system," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017.

[15] C. Han, K. Wu, Y. Wang, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," in *IEEE Conference on Computer Communications*, 2014.

[16] J. Lin, "Noninvasive microwave measurement of respiration," *Proceedings of the IEEE*, 1975.

[17] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *Annual IEEE International Conference on Computer Communications*, 2016.

[18] D. T. Petkie, C. Benton, and E. Bryan, "Millimeter wave radar for remote measurement of vital signs," in *2009 IEEE Radar Conference*, 2009.

[19] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, "Monitoring vital signs using millimeter wave," in *Proceedings of the 17th ACM Inter. Symposium on Mobile Ad Hoc Networking and Computing*, 2016.

[20] A. Ahmad, J. C. Roh, D. Wang, and A. Dubey, "Vital signs monitoring of multiple people using a fmcw millimeter-wave sensor," in *2018 IEEE Radar Conference (RadarConf18)*, 2018.

[21] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: Do user location and body orientation matter?," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016.

[22] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, 2020.

[23] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, 2017.

[24] R. Vallat and M. Walker, "An open-source, high-performance tool for automated sleep staging," *eLife*, vol. 10, 10 2021.

[25] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, pp. 1689–1696, feb 2021.

[26] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, 2019.

[27] Z. Kabelac, C. G. Tarolli, C. Snyder, B. Feldman, A. Glidden, C.-Y. Hsu, R. Hristov, E. R. Dorsey, and D. Katabi, "Passive monitoring at home: a pilot study in parkinson disease," *Digital biomarkers*, 2019.

[28] S. Abdulatif, F. Aziz, K. Armanious, B. Kleiner, B. Yang, and U. Schneider, "Person identification and body mass index: A deep learning-based study on micro-dopplers," in *IEEE Radar Conference (RadarConf)*, 2019.

[29] M. Jiang, G. Guo, and G. Mu, "Visual bmi estimation from face images using a label distribution based method," *Computer Vision and Image Understanding*, 2020.

[30] A. Pantanowitz, E. Cohen, P. Gradidge, N. Crowther, V. Aharonson, B. Rosman, and D. M. Rubin, "Estimation of body mass index from photographs using deep convolutional neural networks," *Informatics in Medicine Unlocked*, 2021.

[31] Y. Yao, L. Song, and J. Ye, "Motion-to-bmi: Using motion sensors to predict the body mass index of smartphone users," *Sensors*, 2020.

[32] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *International Conference on Pattern Recognition*, 2010.