Magnetic and Radar Sensing for Multimodal Remote Health Monitoring

Haobo Li, Aman Shrestha, *Student Member, IEEE*, Hadi Heidari, *Senior Member, IEEE*, Julien Le Kernec, *Senior Member, IEEE*, and Francesco Fioranelli, *Member, IEEE*

***Abstract*— With the increased life expectancy and rise in health conditions related to aging, there is a need for new technologies that can routinely monitor vulnerable people, identify their daily pattern of activities and any anomaly or critical events such as falls. This paper aims to evaluate magnetic and radar sensors as suitable technologies for remote health monitoring purpose, both individually and fusing their information. After experiments and collecting data from 20 volunteers, numerical features has been extracted in both time and frequency domains. In order to analyze and verify the validation of fusion method for different classifiers, a support vector machine with a quadratic kernel, and an artificial neural network with one and multiple hidden layers have been implemented. Furthermore, for both classifiers, feature selection has been performed to obtain salient features. Using this technique along with fusion, both classifiers can detect 10 different activities with an accuracy rate of approximately 96%. In cases where the user is unknown to the classifier, an accuracy of approximately 92% is maintained.**

***Index Terms*— Magnetic sensor, radar sensing, assisted living, feature selection, neural networks, machine learning.**

I. Introduction

LONGER life expectancy in recent years has generated a growing incidence of multi-morbidity (i.e. the simultaneous presence of multiple chronic conditions) and occurrence of critical events such as falls or strokes in the older population. This is a serious issue for the wellbeing of the people involved and their families, as well as posing a series of societal challenges in terms of effective and affordable healthcare provision. In particular, it has been shown that timely medical assistance after fall events can significantly reduce their short and long-term consequences, and that regular monitoring in vulnerable population can enable early detection and treatment of potentially serious health conditions [1]–[3].

A variety of sensing technologies have been investigated to address these issues and enable detection of critical events (falls) and automatic monitoring of daily activities. These include acoustic sensors, Passive Infrared (PIR), pressure sensors on floors, video and depth cameras, wearable sensors that can comprise accelerometer, gyroscope, magnetic sensors, and Radio-Frequency (RF) sensors that can employ passive or active radar principles [4]–[11]. These different sensing approaches have disadvantages and advantages, not only related to the technical implementation or complexity of the sensors and systems themselves, but also to the perception, feeling, and required compliance at the end-users' side, specifically for the people being monitored [12]. Metrics to validate and compare the different technologies include sensitivity and specificity in classifying the different activities, computational complexity, cost of the sensor, detection ranges, false alarm rates, requirements for the users (e.g. wearables need to be worn at all times to be effective), and management of privacy or data integrity issues [12]. These can be significant, especially when sensors are installed in private areas of the home environment such as bedrooms or bathrooms.

In this paper, we consider two types of sensors, namely magnetic sensor and radar, and investigate their performance in the context of daily activity recognition and fall detection, when using different algorithms, including neural networks, for feature extraction and classification. It was reported in our previous work [13], that magnetometer and radar produced many of the most robust features when searching for optimal feature combinations by applying sequential forward selection (SFS) on the feature set including those extracted from all the inertial sensors and radar. Magnetometers are easily miniaturised and integrated in common wearable sensors and objects [14], [15]. They enable to measure small changes in the magnetic field induced by the presence of human personnel, from relatively large variations due to daily activities as in our work, to the very small variations induced by musculoskeletal and cerebral activity [16]. However, such magnetic sensors require the users to wear or carry them and manage their battery life. This may be an issue for older people less familiar with technology, especially when physical frailty is associated to cognitive impairments [17]. Radar sensing has been proposed recently in the context of ambient assisted living, and their capabilities and performance are still being validated in realistic scenarios [18], [19]. Their attractiveness arises from enabling contactless monitoring, with no reliance on end-user's compliance, insensitivity to light and noise conditions in the environment, and their expected perception as less invasive than optical sensors, as no plain videos of the monitored people are collected.

Manuscript received August 23, 2018; revised September 21, 2018; accepted September 24, 2018. Date of publication October 8, 2018; date of current version September 18, 2019. The work of A. Shrestha was supported by Doctoral Training Award. The work of F. Fioranelli was supported by the U.K. Engineering and Physical Science Research Council EPSRC under Grant INSHEP EP/R041679/1. The associate editor coordinating the review of this paper and approving it for publication was Prof. Mehmet Yuce. *(Corresponding author: Francesco Fioranelli.)*

The authors are with the School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: hadi.heidari@glasgow.ac.uk; francesco.fioranelli@glasgow.ac.uk).

Digital Object Identifier 10.1109/JSEN.2018.2872894

In this paper, we expand our previous work in to consider a newer, larger database of sensors' signatures collected involving 20 volunteers aged 22-32 years. Although still limited, this appears to be in the top 3% in terms of number of subjects compared with some of the works on wearable for human motion analysis in the literature [13], [20]. Magnetic sensors have been often used jointly with accelerometer and gyroscopes for fall detection [7], [21], [22], or not considered in favour of using only data from the two aforementioned inertial units [23]–[26]. It is therefore interesting to investigate what classification performance can be obtained with only data from magnetic sensors, thus reducing the computational complexity and potentially battery consumption by processing data from a single wearable sensor. Radar sensors were also used for the data collection, to collect simultaneous recordings together with the wearable sensor. The combination of these heterogeneous sensing technologies and the fusion of their information can address the performance limitations of each individual sensor, and improve the classification achieved for the overall system.

Fusing information from different sensors within wearables (accelerometer, gyroscope, and magnetic sensor) has been suggested in the literature [7], [21], [22], [27], but their joint use with radar sensing and experimental validation is, to the best of our knowledge, a novel research area. Furthermore, rather than using more complex signal level fusion methods based on Kalman filtering or Vector Observation [7], [27], simpler feature level methods to combine information from magnetic sensors and radar sensors are proposed. The Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers are implemented to verify that the fusion method is valid for different classifiers. Furthermore, the influence of the ANN architecture in terms of the number of neurons in hidden layers against overall accuracy is presented.

The rest of this paper presents the experimental scenario, data collection and pre-processing in Section II. This is followed in Section III by the feature extraction methodologies while Section IV describes the implemented classifiers. Finally, Section V discusses the results obtained, with conclusions and future work highlighted in section VI.

II. Experimental Setup and Data Collection

The data analyzed in this work were collected at the University of Glasgow with a group of 20 volunteers aged between 22 and 32 years. The ten activities are described in [20], and include: walking, walking while carrying an object with both hands, sitting on and standing up from a chair, bending to pick up a pen and to tie shoelaces, standing while drinking and answering a phone call, simulating a frontal fall onto a mat, and crouching to check under an imaginary bed and coming back up. A pictorial representation of these activities and their recording length are given in Fig. 1, along with a concept figure representing the idea of the overall fusion of the two sensors, and photos of the radar system with antennas and the wearable at the wrist of one of the volunteers. These activities were purposely chosen to be similar in pairs and to include potential classes that may be misclassified with

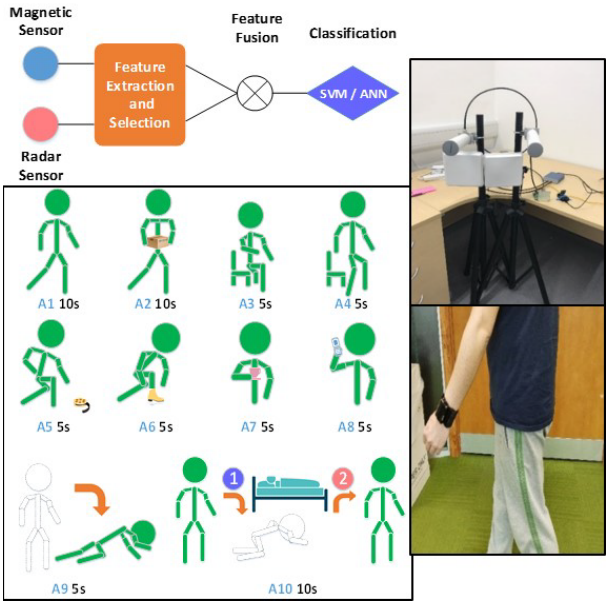


Fig. 1. Conceptual schematic of using magnetic sensor and radar together and pictorial representation of the 10 activities classified. Pictures of the radar system and its antenna, and the wearable at the wrist of the subject are also provided.

falls, in particular those like sitting and bending down that present body movement and acceleration towards the floor. This will test the robustness of the proposed classification approach. Falls are particularly critical to be identified reliably, with low false alarms and low missed detections, as the long lie time following falls can have very severe effects on the health of the person affected [1]–[3]. Each of the 10 activities was recorded for 3 repetitions for each of the 20 volunteers, generating a dataset of 600 readings in total from each sensor. Each snapshot recording of activities 1, 2, and 10 had duration 10s; all the other 5s.

The sensors included a nine DOF (Degrees of Freedom) wearable Inertial Measurement Unit (IMU) produced by X-IO Technologies, a Frequency Modulated Continuous Wave (FMCW) radar operating at 5.8 GHz, and a Continuous Wave (CW) radar system operating at 24 GHz. The wearable sensor was placed with a bracelet on the wrist of the dominant hand of the participants while recording the data. In this work, we focus only on data from the Hall-effect based magnetometer sensor Bosch BMM150 within the wearable. This can measure magnetic fields in the range of ± 1300 μT with approximately 0.3 μT resolution. When the subjects move the magnetometer while performing activities, different electric voltages are produced because of the Hall current effect and recorded by the sensor [15]. The different activities are expected to produce different sequences of variations of the magnetic field measured, and therefore different patterns of recorded voltages that can be analyzed for classification. The voltage signals related to the magnetic field values along the X, Y, and Z axis are then digitised at 20 Hz (default value, whereas those for the accelerometer and gyroscope are digitised at 400 Hz by default). The operating current of the wearable at normal mode is 500μA, whereas the

supply voltage is approximately 3.3V, hence, the total power consumption is in the range of 1.65mW. As mentioned in the introduction, magnetometers enable to measure small fluctuations in the magnetic field due to human presence and activities, even when these may generate small values of velocity or acceleration, not easily captured by the other inertial sensors (or by the radar). Furthermore, including additional signal processing on the accelerometer and gyroscope data (storing, filtering, feature extraction and classification) would require additional computational resources and related power consumption, in addition to the normal consumptions of the sensors themselves. Hence, the focus of our work here to attempt to maximise classification performance with a limited number of sensors.

The Ancortek SDR 580-B FMCW radar transmits linear chirp signals with 400 MHz instantaneous bandwidth at 1 kHz PRF (Pulse Repetition Frequency). Each backscattered chirp signal is received and sampled at 128 KHz. The carrier frequency of the radar is 5.8 GHz, in C-band. The transmitted power of radar system is specified to be approximately 100 mW, with the gain of the Yagi antennas equal to approximately 17 dB. The CW radar (RF-beam K-MC1) transmits approximately 18 dBm EIRP (Effective Isotropic Radiated Power) of power at 24 GHz and the down-converted digitized signal is sampled at 44 kHz. The FMCW operates with a single transmitter and receiver antenna whereas the CW radar has a micro-strip transceiver with transmit and receive capability within the antenna module.

The operational transmit powers for both FMCW and CW systems are in the order of 10dBm which is within the range of wireless LAN and Wi-Fi routers. Both radars are powered by USB, so their consumption is expected to be limited to USB specifications, but both operate with 2.5W. When assessing autonomy and lifetime of the overall classification system, both radars would be connected to a computer for normal operation and recording, so their power consumption is less of an important factor compared to the wearable sensors, for which is desirable to consider only one single sensor, at a low recording and transmit rate. For both radar sensors, the activities' signatures are well within the unambiguous Doppler range, thus avoiding any degradation for aliasing. Only data collected from the 5.8 GHz radar sensor are considered in this work as they can provide both range and Doppler information as opposed to the CW radar (only Doppler information). The data collection was performed in the laboratory of the Communication, Sensing and Imaging group at the University of Glasgow, an office-like space with pieces of furniture and an open area in the middle.

The radar sensors were placed on a wooden table at approximately 80 cm height, pointing to the open area where the subjects were performing the different activities at a distance of approximately 2 m (with this changing slightly as the subjects were performing the various activities). The chosen antenna height allows to keep the torso of the human subjects in the middle of the antenna beam to maximise the received power. The distance was limited by the dimension of the available measurement space, but it is considered a good approximation of a normal indoor space size, where the radar could be deployed on the ceiling or at the walls for monitoring purposes.

The signals collected by the magnetic sensor were pre-processed through a band pass filter, with lower and upper cut-off frequencies of respectively 0.4 Hz to remove bias noise and 10 Hz to remove vibration noise from the spectrum.

The FMCW data were processed according to the usual range-Doppler processing, based on double Fast Fourier Transform (FFT) on the raw data recorded by the radar, as shown in [28] and [29]. By stacking the received radar chirps in a matrix and performing a FFT across fast-time on them, Range-Time-Intensity plots can be generated. A second FFT can be performed across radar chirps to generate range-Doppler maps to identify range bins containing targets' signatures. The complex signal at those range bins is extracted and notch-filtered to remove the contribution near 0 Hz from static targets such as walls and furniture. This signal is then processed through Short Time Fourier Transform (STFT) to characterise their micro-Doppler signatures and generate spectrograms, in this case with a Hamming window length of 0.3 ms and 95% overlap. Spectrograms are Doppler vs time patterns, describing the movement of different body parts over time, such as torso and limbs [30], [31].

Fig. 2 illustrates the signals for four different daily activities in terms of radar micro-Doppler signature (spectrogram) and raw signals of magnetic sensor. Positive Doppler components are associated to movements towards the radar, and negative components to those away from the radar. This can be seen in the example of walking back and forth activity, where the main contribution from the torso (in red colour) alternates between positive and negative values as the person walks towards and away from the radar. In the case of falling (towards the radar in this case), there is a strong and sudden velocity component. Phone call and drinking water are performed while the subjects were mostly stationary, and moving just one arm and slightly their torso, hence the signature is overall more concentrated around the 0 Hz component.

As the spectrogram is normalised with respect to the strongest reflection in each figure, when the distance between the radar and the target changes, the difference in the signature would predominantly be in the increased or decreased level of background noise visible in the spectrogram. This would be an issue to compensate for in outdoor environments, but less relevant for the case of indoor measurements over a few meters as in our work.

It is less easy to empirically infer patterns in the magnetic sensor's data, but for example a sudden and brisk transition can be seen for the falling event at about 1.5 s, similar to what is shown by the radar spectrogram.

One issue with using radar spectrograms is the dependence of the micro-Doppler intensity on the cosine angle between the movement velocity and the radar line-of-sight. This weakened signature may not contribute to relevant information for classification. The magnetometer is agnostic to this problem, and it would continue providing reliable data about the movement of the subject even in case of unfavourable aspect angles for the radar sensor; hence, there is scope to fuse information from these two sensing modalities.

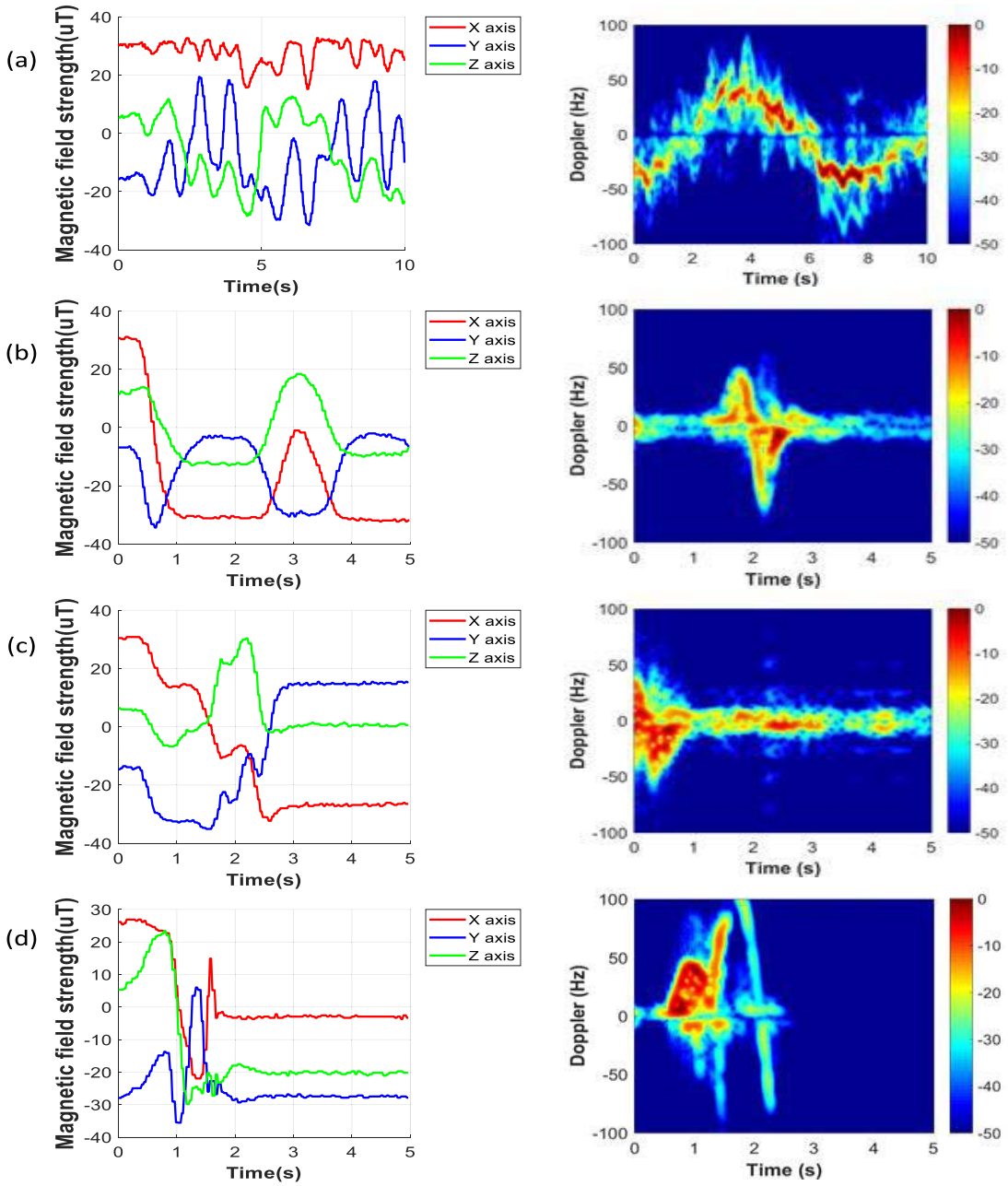


Fig. 2. Examples of magnetic sensor X, Y, and Z axis signal and corresponding FMCW radar spectrograms for four activities: (a) falling, (b) walking back and forth, (c) taking a phone call, and (d) drinking water.

III. Feature Extraction

Sixty-four numerical features were extracted from the magnetic sensor's data along the 3 axes X, Y and Z, looking at both the time and frequency domain [27], [32]–[34]. These are summarized in Table I. Time-domain features (such as mean, variance, and higher order statistical moments like skewness and kurtosis) are extracted to evaluate the deviation of the signal, whereas cross correlation-based features show great potential in classifying activities with significant change on signal magnitude along two dimensions. Frequency-domain (spectral) features are extracted to capture the energy distribution of the signal and include the magnitude of the Power Spectral Density (PSD) at three different frequency bands, namely, 0.5-1 Hz, 1-5 Hz and 5-10 Hz, the sum of Fourier Transform coefficients, and the spectral entropy based on the Power Density Function normalized to between 0 and 1 of the PSD. For the radar sensor, the 24 extracted features included a range of features from the spectrogram, its image texture and the derivative cadence velocity profile. Singular Value Decomposition based spectral and temporal projections reduce the information within the spectrogram to the first few vectors of U and V matrices [35], [36]. The statistical moments of these indicate the amount of information/activity in the overall time and frequency bins of the spectrogram. Doppler centroid and bandwidth [30] are comparative measures which define the central mass of the micro Doppler signature and its

TABLE I
FEATURE TABLE FOR MAGNETIC SENSOR

Time domain	#	Frequency domain	#
Norm of XYZ	1	Spectral Power	9
Mean	3	Coefficients Sum	3
Standard Deviation	3	Spectral Entropy	3
Autocorrelation(Mean,STD)	6		
Cross Correlation(Mean,STD)	6		
Variance	3		
RMS* (Root Mean Square)	3		
MAD (Median Absolute Deviation)	3		
Inter-quadrature Range	3		
Range	3		
Minimum	3		
25th percentiles	3		
75th percentiles	3		
Skewness	3		
Kurtosis	3		
Number of features	**49**	**Number of features**	**15**

Doppler spread. Image texture-based features were entropy and the skewness of the grey levels of the spectrogram. These utilize the information within the texture of the image and give a metric for the total information content and total texture shift respectively [37]. Energy curves of the Doppler and time bins of the spectrogram take the coefficients of the bins and their subsequent moments enumerate the energy within a given time/frequency band. The cadence velocity profile [38] is the Fourier transform of the spectrogram across time which gives time-localized information within the spectrogram. Step repetition frequency (the central component) along with upper and lower values of the shape spectrum of the profile were used as features.

Considering both sensors, the feature extraction step generates a *600* by *88* matrix, denoting the number of observations and the number of features respectively.

IV. CLASSIFICATION METHODS

Two classifiers, a Quadratic-kernel SVM and an ANN [39] with one hidden layer, were considered as classifiers to discriminate different activities. SVM uses features to generate a hyperplane margin based on the distribution of features of a certain class [40]. It has been compared with weaker learning algorithms in [13] for indoor activity classification and is historically known to be a robust classifier. The ANN design used in this work is a multilayer perceptron (MLP), which is comprised of one input layer, one or more hidden layers and then fully connected to several output neurons. All the training and testing procedures are implemented in MATLAB using the pre-installed functions within the Statistics and Machine Learning and Neural Network toolboxes.

The dataset was stochastically divided into two parts, with 70% data for training and 30% data for testing on a per class basis. The per class basis was set for stratification in the test set to prevent class imbalance. By using this deterministic approach unwanted bias in the results is minimised, which would occur in cases of imbalance between classes in the training and test sets. This process is repeated 10 times for each test, and the average results across all the repetitions are presented.

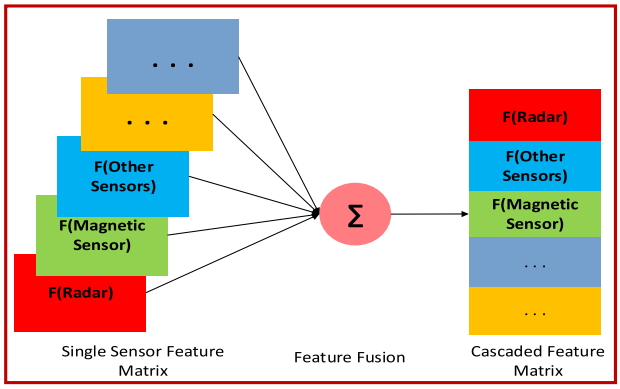


Fig. 3. Graphical representation of feature level fusion scheme.

A. Feature Selection

In applications such as this where the pool of features is large and diverse, there can be redundancies in the feature set which lower the overall classification performance. To prevent this, features subsets which perform well can be generated through the method of feature selection. Sequential Feature Selection (SFS) is one such method, which starts with an empty set of features and adds them by searching the feature space for salient features and selecting those improving the overall accuracy [41]. For SVM classifiers, feature selection has been shown to increase accuracy by 6 to 7% depending on the sensor being used [19]. Feature selection has the further benefit of reducing the overall number of features required for the optimal set; meaning that for deployment, computational loads would be minimized for optimal results.

B. Feature Level Fusion

Another method used to improve classifier is feature level fusion [31], which involves the concatenation of the feature vectors from different sensors to create a new diverse features pool, as graphically illustrated in Fig.3. Although increasing the size of the features pool is likely to introduce redundancies, SFS can then be used to minimize this effect and get a reduced set containing the most suitable features from different sensors.

C. Classifier Performance Metrics

To evaluate the performance of the classifiers, metrics can be derived from the correct classification and misclassifications events in confusion matrices, as shown in the binary example in table II (simple but relevant in the case of fall vs non-fall classification). The diagonal elements indicate the correct identification for a given class of interest (e.g. A). This is represented as a True Positive and when the other class which is not of interest is classified correctly (e.g. B), whereas a false negative indicates a 'missed detection' and a false positive a 'false alarm'. Typical metrics for classification performance are listed in [42], notably, sensitivity or recall, specificity and precision. These metrics have been calculated for the 10-class case described in this paper. Further metrics to characterize the overall classification without yielding class by class including overall classification accuracy and F-measure can also be considered.

TABLE II
CONFUSION MATRIX EXAMPLE FOR CLASSIFICATION METRICS

True\Predicted	A	B
A	True Positive	False Negative
B	False Positive	True Negative

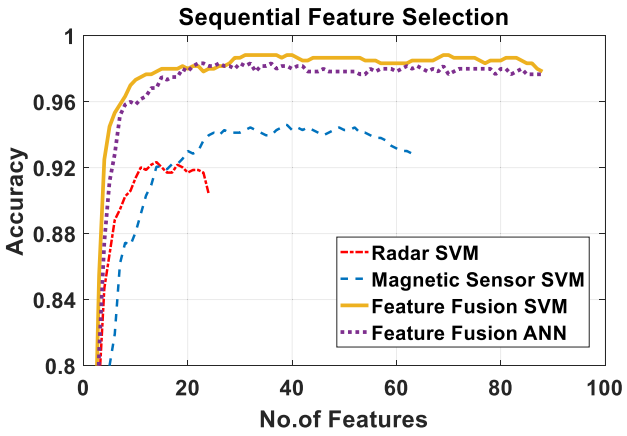


Fig. 4. Accuracy when applying SFS for different sensors and classifiers.

V. CLASSIFICATION RESULTS

This section presents the classification results. Initially, we evaluate the role of SFS and discuss the features selected by the algorithm, then we will use individual sensor for activity classification with SVM followed by comparison of fusion and neural network approaches.

Fig. 4 summarizes the results in terms of classification accuracy for different sensors and classifiers when selecting different number of features through the SFS algorithm. The accuracy profiles as a function of number of features become stable after using approximately 30 features for the magnetic sensor individually, and 10 features for the radar. With those features, the average accuracy is in the order of 93-94% for the magnetic sensor and 92% for the radar sensor. Adding extra features will not bring any significant benefit to the classification and may actually reduce the overall performance if all the available features are used, as shown in Fig. 4.

Feature selection with feature fusion allows combining the most suitable features from both radar and the magnetic sensors. The resulting accuracy outperforms the cases of both sensors used individually. There is a correspondence between the most suitable features selected for each sensor individually, and those selected on the fused set of features. Some of the best features selected have been listed in Table III.

Fig. 4 shows also the accuracy as a function of the number of features used as input to a single hidden layer ANN with 50 neurons, assuming a fused pool of features is used. Results appear to be very similar for SVM and ANN in this case, with a more detailed discussion in terms of metrics including computational time provided later on in this section.

TABLE III
ROBUST FEATURES FOR RADAR AND MAGNETOMETER

Magnetic Sensor	Radar
STD of auto-correlation of axis	Mean Centroid
RMS of axis	Mean of principle V vector
Mean of cross-correlation between axis	Mean Doppler bandwidth
STD of cross-correlation between axis	Standard deviation of Centroid
Gated Spectral Power Y, Z	Entropy of the image texture
Min of axis	Std. of Doppler bandwidth
STD of auto-correlation Z	Std. of EBC
Mean of auto-correlation X, Z	Step repetition Frequency
Spectral Entropy Y	Minimum Cadence Velocity
75th percentiles Z	Max Cadence Velocity
Norm of X, Y, Z	
Mean X and Mean Z	
STD Y and STD Z	
Inter-quadrature Range Z	
Sum of FFT Z	
25th percentile X	
Variance Y	
Range Y	
Total 30	**Total 10**

A. Support Vector Machine (SVM) With SFS

Table IV to VI presents the confusion matrices when SVM classifier is used with feature selection, respectively for the magnetic sensor and the radar individually, and then with feature fusion. A colour code is used with green on the diagonal elements, and yellow (below 10%) and orange (above 10%) to highlight misclassification events.

From table IV, the magnetic sensor yields high classification sensitivity and specificity for most of the class; however, for the A3 'sitting down' and A5 'picking up an object', the classifier did not perform as well as for the other classes. Low specificity also occurs for A9 'fall', with multiple false alarms from different activities, which appears to suggest this approach and sensor are not optimal for fall detection. As highlighted earlier, as falls have severe consequences on the health of the people monitored, the reliable detection of falls remains a critical requirement. Furthermore, it is paramount that there are no false alarms when it comes to fall detection, as these would undermine the confidence of the end-users (vulnerable people and their carers) in this technology. From Table V, radar performs well with respect to class A9 'fall' with less false alarms. However, there is considerable confusion between similar activities across the wider set of classes considered (for example A1 and A2 with walking and walking while carrying an object). In Table VI the results for feature fusion are shown. Compared with the individual use of the magnetic sensor, false alarms for the 'fall' class are removed, at the price of a slightly increased missed detection with a specific "confusing" activity, A6 'tying shoelaces'. In general, the other activities are classified with accuracy above 94%, without any instances of misclassification above 10% (no orange elements). The results seem to suggest that different sensors and related features can contribute to different tasks for assisted living, with some sensors and some feature combinations more suitable for recognition of a set of daily activities, and others more suitable for fall detection.

TABLE IV
CONFUSION MATRIX FOR MAGNETIC SENSOR SVM

%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	**100**	0	0	0	0	0	0	0	0	0
A2	0	**100**	0	0	0	0	0	0	0	0
A3	0	0	**79.3**	4.8	0	11.1	0	0	4.8	0
A4	0	0	0	**93.7**	6.3	0	0	0	1.6	0
A5	0	0	15.9	0	**84.1**	0	0	0	3.2	0
A6	0	0	0	0	0	**100**	0	0	0	0
A7	0	0	0	0	0	0	**90.4**	4.8	4.8	0
A8	0	0	0	0	0	0	0	**88.9**	11.1	0
A9	0	0	0	0	4.8	0	0	0	**95.2**	0
A10	0	0	0	0	0	0	0	0	0	**100**

TABLE V
CONFUSION MATRIX FOR RADAR SVM

%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	**88.1**	11.9	0	0	0	0	0	0	0	0
A2	5.8	**94.2**	0	0	0	0	0	0	0	0
A3	0	0	**95**	0	2.5	0	0	1.7	0	0.8
A4	0	0	0	**93.6**	2.2	2.8	1.4	0	0	0
A5	0	0	2.8	0	**88.3**	0	7.8	1.1	0	0
A6	0	0	0	1.1	0.6	**95.3**	0.5	2.5	0	0
A7	0	0	0	0.3	1.7	0.6	**83.6**	13.8	0	0
A8	0	0	0.3	0	3	1.7	8.9	**86.1**	0	0
A9	0	0	0	0.6	0	3.3	0	0	**95.8**	0.3
A10	0	0	0	0	1.4	2.5	0	0	0	**96.1**

TABLE VI
CONFUSION MATRIX FOR FUSION SVM

%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	**100**	0	0	0	0	0	0	0	0	0
A2	0	**100**	0	0	0	0	0	0	0	0
A3	0	0	**93.9**	0	6.1	0	0	0	0	0
A4	0	0	0	**97.8**	2.2	0	0	0	0	0
A5	0	0	2.2	0	**97.8**	0	0	0	0	0
A6	0	0	0	0.6	1.1	**98.3**	0	0	0	0
A7	0	0	0	0	0	0	**95.6**	4.4	0	0
A8	0	0	0	0.6	1.1	1.7	2.2	**94.4**	0	0
A9	0	0	0	0	0	6.1	0	0	**93.9**	0
A10	0	0	1.1	0	0.6	0.5	0	0	0	**97.8**

This could be exploited by hierarchical and adaptive classification schemes as future work.

B. Artificial Neural Network (ANN)

The validation accuracy produced using ANN with a single hidden layer and between 1 and 50 neurons utilizing different sensors is illustrated in Fig. 5. Results obtained using fusion of magnetic and radar sensor outperform those obtained using each sensor individually, and the fusion result for accuracy is on average 96%, fairly close to that obtained using SVM. It is interesting to notice that the plateau of the accuracy is reached when over 10 neurons are used for the fusion cases, i.e. at least one neuron or more for each output class. When the sensors are used individually, the accuracy pattern increases in a more gradual manner with the number of neurons. For both radar and magnetic sensor, it appears that the plateau of accuracy is reached when the number of neurons in the hidden layer is comparable to the optimal number of features selected during SFS (approximately 14 features for radar and

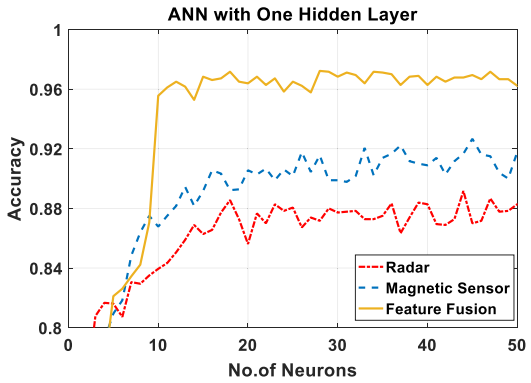


Fig. 5. One-layer ANN Accuracy with different sensors and number of neurons.

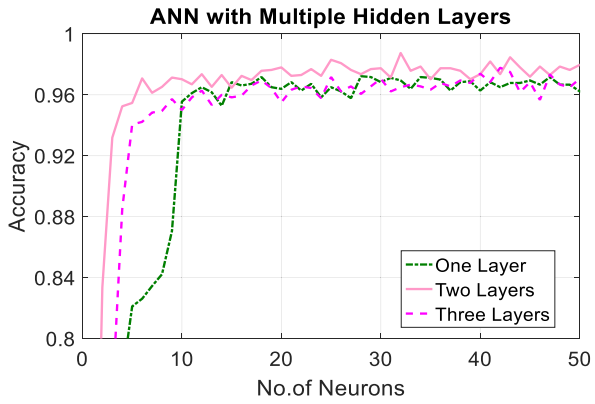


Fig. 6. ANN accuracy when using multiple hidden layers.

34 features for magnetic sensor). This may suggest that the ANN can select automatically relevant information from the feature space and limit the effect of redundancies. There is further evidence of this in Table IX where the performance measures are summarized. The accuracy for SFS SVM with fusion and ANN with fusion but without SFS are similar, as if the ANN performs internally a form of feature selection.

Fig 6 shows that there is little difference in accuracy when using multiple hidden layers for the ANN. In this case, the number of neurons are varied in the last hidden layer and kept constant at 50 for the other layers. Increasing the number of layers has a proportional increase in the computational load and time required for processing, but there appears to be no discernible difference (about 0.8%) between using one or two hidden layers aside from the fact that less neurons are required to reach the optimal point.

Training and testing performance over the epochs of the ANN with 1 hidden layer and 50 neurons is evaluated in Table VII, where the training and testing accuracy increase steadily until 40 epochs after which the model seems to over fit, probably due to the small data input set. For our case, limiting the number of epochs to 40 will reduce the computational cost for training and still provide optimal results.

Class-wise comparison of the confusion matrices for the SVM (Table VI) and the one-layer ANN (shown in Table VIII) suggests that there is a slight improvement in the classification

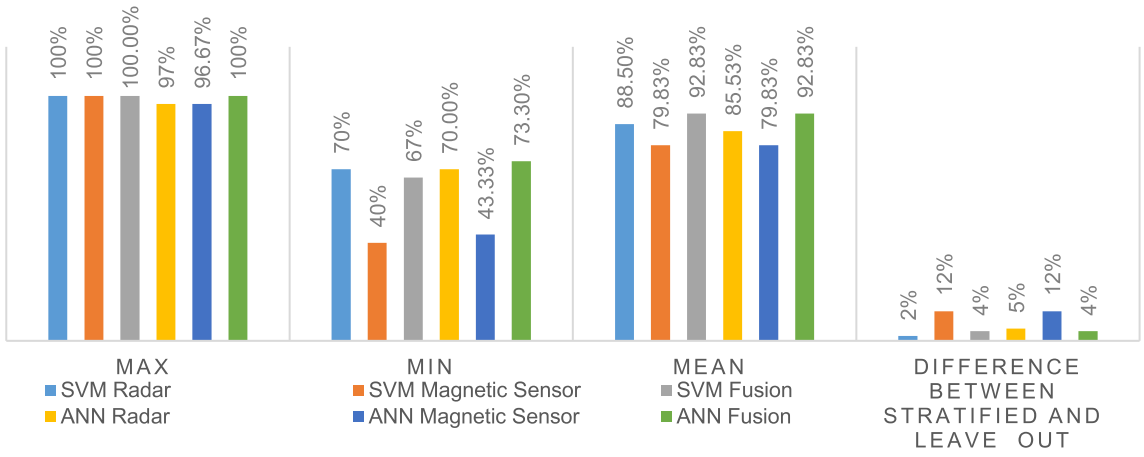


Fig. 7. Comparison of max, min, and mean accuracy when classification tests done with "leave one subject out" approach.

TABLE VII

TRAINING ACCURACY OVER EPOCHS FOR ANN WITH 1 HIDDEN LAYER AND 50 NEURONS

Epochs	Training Accuracy (%)	Testing Accuracy (%)
10	94.6	82.5
20	100	92.9
30	100	96
40	100	98.4
50	100	94.4

TABLE VIII

CONFUSION MATRIX FOR FUSION ANN WITH ONE HIDDEN LAYER

%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	**100**	0	0	0	0	0	0	0	0	0
A2	0	**100**	0	0	0	0	0	0	0	0
A3	0	0	**93.1**	0	2.8	2.8	0.8	0.3	0	0.3
A4	0	0	0	**97.2**	1.4	1.4	0	0	0	0
A5	0	0	0.3	0	**98.6**	0	0	1.1	0	0
A6	0	0	0	0.8	0	**100**	0	0	0	0
A7	0	0	0	0	1.4	0	**95**	3.6	0	0
A8	0	0	0	0.3	0.8	1.1	4.2	**94.2**	0	0
A9	0	0	0	1.1	0	3.6	0.8	0	**94.4**	0
A10	0	0	0.3	3.1	1.9	0	0	0	0.3	**94.7**

of A9 'fall', whereas the performance across the whole set of activities appear to be fairly similar.

The summary of the classification metrics is given in Table IX, considering the different combinations of sensors and classification approaches. Note that the ANN considered here had one-hidden layer and 50 neurons. Overall, the results appear to confirm the previous trends observed in Fig. 4 and 5, with fusion providing a significant improvement in terms of classification performance. In terms of computational load, there is a trade-off between time and number of features employed, where an ANN with one-hidden layer takes around 2-3 times longer than SVM to finalize the training and testing across the full data set.

C. Leave One Subject Out

In real-life applications, the classifcation model will not have data from the test subject, therefore it would be

TABLE IX

PERFORMANCE COMPARISON OF DIFFERENT SENSOR METHODS: AVERAGE OF ALL 10 CLASSES -* INDICATES USE OF SFS

Classifier/Sensor	*Precision (%)*	*Recall (%)*	*Specificity (%)*	*F-measure (%)*
SVM (Radar) *	91.4	91.7	99.1	91.4
SVM (Mag) *	92.3	91.7	99.1	92.0
SVM (Fusion)	96.8	96.6	99.6	96.7
SVM(Fusion) *	97.8	98.3	99.9	98.1
ANN (Radar)	89.6	89	98.8	89.3
ANN (Mag)	92.3	92.1	99.1	92.2
ANN (Fusion)	97.4	97.2	99.7	97.3
ANN(Fusion} *	97.4	97.9	99.9	97.7

interesting to observe the classifier's ability to identify activities from unknown subjects. For the leave one subject out, observations from one specific participant was selected as the testing set and the remaining participants were used for training. This was repeated until all of the participants were tested upon, then the cumulative classifcation results were calculated. These results are displayed in Fig. 7, in terms of 'min', 'max', and 'mean' accuracy. The 'max' and 'min' variables represent the best and worst-case scenario for an individual participant under test, and the 'mean' is the average across all the 20 participants. The 'difference' variable is the delta between the accuracy from the stratified test in table IX and this approach of "leave one subject out". The results show a significant variability in accuracy with specific subject, with the extreme case of individual magnetic sensor for both ANN and SVM, which yields accuracy of approximately 40%. Radar performs robustly with both classifiers, as the mean results are only 2 to 4% lower than the results from the stratified set. The differences are more prominent for the magnetic sensor as the accuracy is 12% lower for both classifiers with this sensor. Feature fusion helps greatly in this scenario, as it enables to recover this loss and increase the accuracy by 12%.

The minimum value for each of the sensor-classifier combinations displays the challenging issue of activity classification for users for whom no data was available at training. It can be seen as the lower bound of the classification performance for a

TABLE X
STATISTICAL PARAMETERS OF LEAVE ONE SUBJECT OUT TEST

	SVM Radar	*ANN Radar*	*SVM Magnetic Sensor*	*ANN Magnetic Sensor*	*SVM Fusion*	*ANN Fusion*
Lower quartiles	0.83	0.8	0.73	0.7	0.9	0.9
Upper quartiles	0.95	0.93	0.9	0.87	1	1
Variance	0.0073	0.0052	0.02	0.0139	0.0071	0.0075

new subject. This is in some cases significantly smaller (even 53%) than the expected accuracy from the stratified testing, for which the subject is not completely unknown. Even after fusion and feature selection, we see that the difference between this lower bound of performance and the expected rate from the stratified testing is approximately 24%. Fusion helps the magnetic sensor in this case as the deltas between the two testing methods is minimized from 12% to 4%. Therefore, feature fusion is necessary for ambient activity monitoring as the additional degree of freedom provides a significant benefit.

Upon closer inspection, it appeared that the minimum performance case is related to the data of one specific subject, who moved in a different style compared with the average of the other subjects, at a slower and more relaxed pace. This could happen, as subjects were not strictly instructed to perform activities with a predefined speed and style, to account for as much realism as possible even in this controlled experiment. In a sense, the testing data from this subject can be considered outliers compared with the training data from the other subjects. This is shown in Table X, which emphasises that the minimum figure is an outlier by comparing the lower and upper quartiles of the classification accuracy for different situations. Radar has lower variance and interquartile range in classification, whereas for magnetic sensor both are larger. Through fusion of two sensors' data at feature level, the interquartile range is reduced to 0.1 while the variance remains at the same level as for radar. A closer difference between the accuracy in the two quartiles is seen as a confirmation that only a limited set of samples, namely those for one specific subject, scored a low accuracy compared with the other subjects' data. Comparing different classifiers, ANN performs better than SVM when we consider the variance when radar and magnetometer are used individually. But for the fusion case, their performances are very similar with all the statistical parameters for both classifiers.

VI. CONCLUSIONS

For assisted living applications such as activity monitoring and fall detection, inertial sensors including magnetometer have been widely used, whereas radar has only recently attracted interest in this area. In order to overcome shortcomings of a single sensing technology sensor fusion methods have demonstrated advantages in combining information from heterogeneous sensors. In this paper, two different classification methods, one 'classic' Quadratic-kernel based SVM and one 'popular' multilayer feed-forward ANN have been selected to validate the performance of our SFS fusion technique compared to using radar and magnetic sensor individually. Results show that by using feature selection, we benefit from an improved classifier accuracy and reduced false alarms, while simultaneously reducing the computational load through the removal of redundant features. In this case, for both magnetic and radar sensors individually this accuracy improvement was approximately 2%. More significant improvements were achieved by using feature fusion, namely approximately 6% for radar and 5% for magnetic sensor. Additionally, tests with ANNs show a slight increase of accuracy; a closer look into the results show that this effect comes from increase in the detection of certain classes (such as sitting down, tying shoelaces and, picking up an object) rather than across all of them. Furthermore, using more than one layer does not appear to provide an increase, despite having higher computational costs.

These preliminary results have demonstrated the potential of multi-sensory fusion in human activity recognition for assisted living. For future work, more participants with a greater diversity in gender, age and body types will be collected, with a drive to also perform measurements in different environments and with different deployment geometries of the radar sensor. Furthermore, additional features (such as spectrogram-based features from magnetic sensor, and different time-frequency transforms such as Cohen class and/or wavelet features from radar) with potential greater saliency should be extracted. The apparent relationship between salient features generated by sequential forward selection needs to be explored further to understand what makes a good feature set, and how different features combine to enhance accuracy in order to engineer and select better hand-crafted features. Training and testing with more complex neural networks (e.g. Recurrent Neural Networks and Long Short-Term Memory Networks) will be explored with a larger database. Their implementation in real-time on hardware platforms while maintaining the offline accuracy of human activity recognition with integrated magnetic and radar sensing system are still outstanding challenges to be addressed in future work.

REFERENCES

[1] W. H. O. Ageing and L. C. Unit, *WHO Global Report on Falls Prevention in Older Age*. Geneva, Switzerland: World Health Organization, 2008.

[2] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *J. Amer. Geriatrics Soc.*, vol. 66, no. 4, pp. 693–698, Apr. 2018.

[3] M. Terroso, N. Rosa, A. T. Marques, and R. Simoes, "Physical consequences of falls in the elderly: A literature review from 1995 to 2010," *Eur. Rev. Aging Phys. Activity*, vol. 11, no. 1, p. 134, 2014.

[4] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

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

[6] K. Chaccour, R. Darazi, A. H. El Hajjam, and E. Andrès, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors J.*, vol. 17, no. 3, pp. 812–822, Feb. 2017.

[7] I. H. López-Nava and A. Muñoz-Meléndez, "Wearable inertial sensors for human motion analysis: A review," *IEEE Sensors J.*, vol. 16, no. 22, pp. 7821–7834, Nov. 2016.

[8] F. Erden, S. Velipasalar, A. E. Cetin, and A. Z. Alkar, "Sensors in assisted living: A survey of signal and image processing methods," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 36–44, Mar. 2016.

[9] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "Radar and RGB-depth sensors for fall detection: A review," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3585–3604, Jun. 2017.

[10] H. Heidari *et al.*, "CMOS vertical Hall magnetic sensors on flexible substrate," *IEEE Sensors J.*, vol. 16, no. 24, pp. 8736–8743, Dec. 2016.

[11] L. Montanini, A. Del Campo, D. Perla, S. Spinsante, and E. Gambi, "A footwear-based methodology for fall detection," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1233–1242, Feb. 2018.

[12] *Report to Congress: Aging Services Technology Study*, U.S. Dept. Health Human Service, Washington, DC, USA, 2012.

[13] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE J. Electromagn., RF Microw. Med. Biol.*, vol. 2, no. 2, pp. 102–108, Jun. 2018.

[14] V. Nabaei, R. Chandrawati, and H. Heidari, "Magnetic biosensors: Modelling and simulation," *Biosensors Bioelectron.*, vol. 103, pp. 69–86, Apr. 2017.

[15] H. Heidari, N. Wacker, S. Roy, and R. Dahiya, "Towards bendable CMOS magnetic sensors," in *Proc. 11th Conf. Ph.D. Res. Microelectron. Electron. (PRIME)*, Jun./Jul. 2015, pp. 314–317.

[16] E. Boto *et al.*, "Moving magnetoencephalography towards real-world applications with a wearable system," *Nature*, vol. 555, no. 7698, pp. 657–661, 2018.

[17] J. Fleming and C. Brayne, "Inability to get up after falling, subsequent time on floor, and summoning help: Prospective cohort study in people over 90," *BMJ*, vol. 337, p. a2227, Nov. 2008.

[18] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, "Extracting gait velocity and stride length from surrounding radio signals," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 2116–2126.

[19] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 71–80, Mar. 2016.

[20] H. Li *et al.*, "Multisensor data fusion for human activities classification and fall detection," in *Proc. IEEE SENSORS*, Glasgow, U.K., Oct./Nov. 2017, pp. 1–3.

[21] P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, "A high reliability wearable device for elderly fall detection," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4544–4553, Aug. 2015.

[22] P. Pierleoni *et al.*, "A wearable fall detector for elderly people based on ahrs and barometric sensor," *IEEE Sensors J.*, vol. 16, no. 17, pp. 6733–6744, Sep. 2016.

[23] B. Andò, S. Baglio, C. O. Lombardo, and V. Marletta, "A multisensor data-fusion approach for ADL and Fall classification," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 1960–1967, Sep. 2016.

[24] B. Andò, S. Baglio, C. O. Lombardo, and V. Marletta, "An event polarized paradigm for ADL detection in AAL context," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1814–1825, Jul. 2015.

[25] S. S. Khan and B. Taati, "Detecting unseen falls from wearable devices using channel-wise ensemble of autoencoders," *Expert Syst. Appl.*, vol. 87, pp. 280–290, Nov. 2017.

[26] D. De, P. Bharti, S. K. Das, and S. Chellappan, "Multimodal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Comput.*, vol. 19, no. 5, pp. 26–35, Sep./Oct. 2015.

[27] R. C. King, E. Villeneuve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin, "Application of data fusion techniques and technologies for wearable health monitoring," *Med. Eng. Phys.*, vol. 42, pp. 1–12, Apr. 2017.

[28] D. E. Barrick, "FM/CW radar signals and digital processing," Nat. Ocean. Atmos. Admin. Wave Propag. Lab, Boulder, CO, USA, Tech. Rep. AD-774 829, 1973.

[29] S. Salous, *Radio Propagation Measurement and Channel Modelling*. Hoboken, NJ, USA: Wiley, 2013.

[30] F. Fioranelli, M. Ritchie, S. Z. Gürbüz, and H. Griffiths, "Feature diversity for optimized human micro-Doppler classification using multistatic radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 640–654, Apr. 2017.

[31] D. Tahmoush, "Review of micro-Doppler signatures," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1140–1146, Dec. 2015.

[32] S. Chernbumroong, A. S. Atkins, and H. Yu, "Activity classification using a single wrist-worn accelerometer," in *Proc. 5th Int. Conf. Softw. Knowl. Inf., Ind. Manage. Appl. (SKIMA)*, Sep. 2011, pp. 1–6.

[33] H. S. AlZubi, S. Gerrard-Longworth, W. Al-Nuaimy, Y. Goulermas, and S. Preece, "Human activity classification using a single accelerometer," in *Proc. 14th UK Workshop Comput. Intell. (UKCI)*, Sep. 2014, pp. 1–6.

[34] J. Zhu, R. San-Segundo, and J. M. Pardo, "Feature extraction for robust physical activity recognition," *Hum.-Centric Comput. Inf. Sci.*, vol. 7, no. 1, p. 16, Dec. 2017.

[35] J. J. M. de Wit, R. I. A. Harmanny, and P. Molchanov, "Radar micro-Doppler feature extraction using the singular value decomposition," in *Proc. Int. Radar Conf.*, Oct. 2014, pp. 1–6.

[36] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of unarmed/armed personnel using the NetRAD multistatic radar for micro-Doppler and singular value decomposition features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1933–1937, Sep. 2015.

[37] X. Shi, F. Zhou, L. Liu, B. Zhao, and Z. Zhang, "Textural feature extraction based on timeâŁ"frequency spectrograms of humans and vehicles," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1251–1259, Dec. 2015.

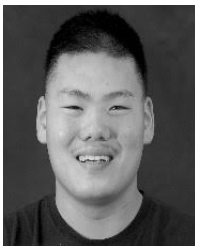
[38] R. Ricci and A. Balleri, "Recognition of humans based on radar micro-Doppler shape spectrum features," *IET Radar, Sonar Navigat.*, vol. 9, no. 9, pp. 1216–1223, Dec. 2015.

[39] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural Network Design*, vol. 20. Boston, MA, USA: PWS, 1996.

[40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[41] S. Z. Gürbüz, B. Erol, B. Çagliyan, and B. Tekeli, "Operational assessment and adaptive selection of micro-Doppler features," *IET Radar Sonar Navigat.*, vol. 9, no. 9, pp. 1196–1204, Dec. 2015.

[42] R. M. Gibson, A. Amira, N. Ramzan, P. Casaseca-de-la-Higuera, and Z. Pervez, "Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic," *Appl. Soft Comput.*, vol. 39, pp. 94–103, Feb. 2016.



Haobo Li received the B.Eng. degree in electrical and electronic engineering from Northumbria University, Newcastle-upon-Tyne, U.K., in 2015, and the M.Sc. degree in communication and signal processing from the University of Newcastle in 2016. He is currently pursuing the Ph.D. degree with the School of Engineering, University of Glasgow, Glasgow, U.K. He is also involved in information fusion of multiple sensing technologies for assisted living applications and gesture recognition.

Aman Shrestha received the B.Eng. degree in electronic and electrical engineering from the University of Birmingham, Birmingham, U.K. He is currently pursuing the Ph.D. degree with the School of Engineering, University of Glasgow, Glasgow, U.K., with a focus on the application of radar systems and radar signal processing for assisted living, in particular activity recognition and fall detection.

Hadi Heidari (SM'18) received the Ph.D. degree in microelectronics from the University of Pavia, Italy, with a focus on integrated CMOS magnetic sensory microsystems. He held post-doctoral position with the University of Glasgow. He joined the Glasgow College UESTC in 2016. He was a Research Visitor with the University of Macau, China, and McGill University, Canada. He is currently an Assistant Professor (Lecturer) with the School of Engineering and the Lead of the Microelectronics Lab (meLAB) with the University of Glasgow, U.K. He has authored or co-authored over 80 peer-reviewed publications in international journals or conference proceedings and acts as a reviewer for several journals and conferences. He has been a recipient of a number of awards, including the Rewards for Excellence prize from UofG (2018), the IEEE CASS Scholarship (NGCAS'17), and Best Paper Awards from the ISSCC'16, ISCAS'14, and PRIME'14.

Julien Le Kernec (SM'17) received the B.Eng. and M.Eng. degrees in electronic engineering from the Cork Institute of Technology, Ireland, in 2004 and 2006, respectively, and the Ph.D. degree in electronic engineering from University Pierre and Marie Curie, France, in 2011. He held a post-doctoral position with the Kuang-Chi Institute of Advanced Technology, Shenzhen, China, from 2011 to 2012. He was a Lecturer with the Department of Electrical and Electronic Engineering, The University of Nottingham, Ningbo, China, from 2012 to 2016. He is currently a Lecturer with the School of Engineering with the Systems, Power and Energy Group, University of Glasgow. His research interest includes radar system design, software defined radio/radar, signal processing, and health applications.

Francesco Fioranelli (M'16) received the Ph.D. degree from Durham University, U.K., in 2014. He was a Research Associate with University College London. He has been an Assistant Professor with the University of Glasgow since 2016. He is a member of the Communication, Sensing and Imaging Research Group, where he leads the research stream in radar systems and applications. His research interest includes multistatic and distributed radar sensing, the application of machine learning to radar data for classification in healthcare and security, characterization of radar signatures of vehicles and UAVs, wind farms, and sea clutter.