Bi-LSTM Network for Multimodal Continuous Human Activity Recognition and Fall Detection

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

***Abstract*—This paper presents a framework based on multi-layer bi-LSTM network (bidirectional Long Short-Term Memory) for multimodal sensor fusion to sense and classify daily activities' patterns and high-risk events such as falls. The data collected in this work are continuous activity streams from FMCW radar and three wearable inertial sensors on the wrist, waist, and ankle. Each activity has a variable duration in the data stream so that the transitions between activities can happen at random times within the stream, without resorting to conventional fixed-duration snapshots.**

The proposed bi-LSTM implements soft feature fusion between wearable sensors and radar data, as well as two robust hard-fusion methods using the confusion matrices of both sensors. A novel hybrid fusion scheme is then proposed to combine soft and hard fusion to push the classification performances to approximately 96% accuracy in identifying continuous activities and fall events. These fusion schemes implemented with the proposed bi-LSTM network are compared with conventional sliding window approach, and all are validated with realistic "leaving one participant out" (L1PO) method (i.e. testing subjects unknown to the classifier). The developed hybrid-fusion approach is capable of stabilizing the classification performance among different participants in terms of reducing accuracy variance of up to 18.1% and increasing minimum, worst-case accuracy up to 16.2%.

***Index Terms*—Radar sensing, multimodal sensing, Recurrent Neural Network, human activity recognition, continuous activity pattern, hybrid fusion**

I. Introduction

In many Western countries and China, the increasingly aging population causes additional challenges to provide healthcare for managing multiple chronical conditions (multimorbidity) and provide timely support in case of critical events such as a fall [1], [2]. Falls represent a leading cause for injuries and discomfort, both at a physical and psychological level, with life expectancy after the event correlated with the time to receive help. Beyond the detection of critical events, the continued analysis of daily routines and activity patterns is also significant to identify possible changes and anomalies that may be related to worsening health conditions. These might go unnoticed by the subjects themselves until the symptoms are too severe to require hospitalization and acute treatment.

To enable this continued and personalized healthcare monitoring in home environment, different sensing technologies have been suggested in recent years, in the context of human activity recognition and fall detection [3]. These include wearable sensors [4]–[6], image and video sensors [7], [8], ambient sensors [3], and radio frequency and radar sensing [9]–[12]. In particular, radar has attracted considerable interest in the sensing research community thanks to its contactless capabilities (whereby the end-user does not need to wear, carry, or interact with any additional device, which can help for acceptance and compliance), and to its lack of plain images or videos to be recorded (which can help for potential issues of privacy). Hence, numerous studies in the literature have investigated the use of radar sensing for human activities classification, personnel recognition, and presence sensing, even in through-the-wall conditions [10]–[16]. The radar information can be represented in a 3D space, containing range (physical distance), time, and velocity (measured through the Doppler effect), sometimes referred to as "radar cube" [17]. Among these different domains of radar information, micro-Doppler is typically used, exploiting the small modulations on the received radar signal caused by "micro-motion" of individual body parts (e.g., limbs, torso, head) [18].

The radar information, particularly micro-Doppler data, can be degraded during the tangentially movement of target to the radar line-of-sight or out of the antenna beam. Therefore, the use of additional radar nodes (multistatic/distributed radar) or additional heterogeneous sensors avoid any data degradation in a multimodal framework [19]–[22]. This enables to exploit the complementary advantages of different sensing modalities, combine information at the most relevant level (e.g., at the signal, feature, or decision level), and capitalize on a plurality of sensors that are widely available in modern and smart living environments.

Neural network-based classification methods, in particular, CNN (Convolutional Neural Network) and autoencoders have attracted a lot of attention and showed to generally outperform conventional classifiers in terms of classification accuracy, at the price of additional training complexity. H. Sadreazami et al. [23] proposed a CNN-based Capsule network to identify the fall accidents through ultra-wide band radar and the results indicated that it over performed SVM with different kernel

The authors acknowledge support of the UK EPSRC grant EP/R041679/1. The authors are grateful to the volunteers who helped with the data collection, in particular Dr Lun Ma, Chang'an University, and to Dr D. Ciuonzo at University of Naples for discussions and ideas on data fusion techniques.

All authors are with the James Watt School of Engineering, University of Glasgow, G12 8QQ, Glasgow, UK (e-mail: francesco.fioranelli@glasgow.ac.uk).

functions and regular CNN. F. Luo's paper [24] also pointed that SVM plus 2-D PCA (Principle Component Analysis) and CNN (RadarNet) showed better performance than other classifiers in activity classification, subject recognition and outdoor localization. S. Gurbuz et al. [25] compared the walking pattern recognition of three different radar sensors and one sonar using a broad range of features. However, these research works focused on binary or small number of classes' problem, and the human motions they collected and analyzed is a X-s observation rather than longer, continuous sequence data.

In this paper, we propose a novel framework for information fusion of radar and wearable inertial sensors data based on a bidirectional Long Short-Term Memory (bi-LSTM) neural network. This expands our previous research in [22], [26] by considering more challenging and realistic continuous activities, i.e., activities that are performed one after another with random duration and transitions, rather than fixed-length, separated, snapshots for each activity. In this case, radar and wearable sensors data are processed within the recurrent LSTM neural network as a continuous-time series, the sequence of data, rather than as individual images, as typically done in convolutional neural networks. Rather than considering radar data [27] or wearable sensors data [28] in isolation, we investigate soft and hard fusion schemes for the data, as well as a novel hybrid approach that is shown to increase the overall accuracy while reducing the variance of the results across different test subjects.

The remainder of this paper is organized as follows: Section II introduces the experimental setup and describes the continuous data collection. Section III discusses the data pre-processing, feature extraction, and selection. Section IV presents the results of the considered classification approaches. Finally, section V summarizes the paper and draws conclusions touching on future work.

II. Experimental Setup and Data Collection

The data was collected with 15 male and 1 female participants aged 21-35 years in a common room at the University of Glasgow, as shown in Fig. 1. The participants were asked to simulate daily activities in the activity zone in front of the radar sensors, approximately 3m × 2.2m.

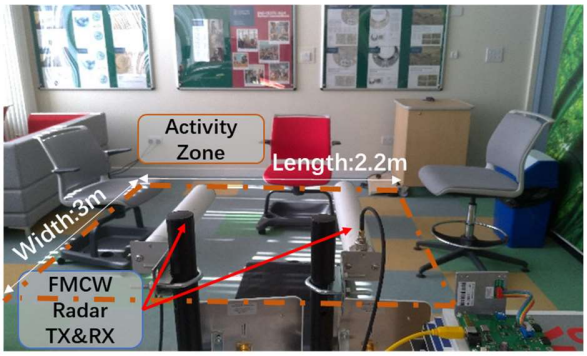

Fig. 1. View of the experimental setup for recording data: common room at the University of Glasgow, with furniture and clutter nearby.

An FMCW radar system and three wearable sensors were used to record the activity data; notably, the FMCW radar operating at 5.8GHz with 400MHz instantaneous bandwidth and 1ms duration and three wearable IMUs (Inertial Measurement Units) placed on the participants' bodies with sampling rate at 50Hz. Each IMU is comprised of one tri-axial accelerometer, gyroscope, and magnetometer; it can provide nine degrees of freedom by simultaneously reading the target experienced acceleration, angular speed, and magnetic field variation. The three sensors were placed on the wrist, waist, and ankle of the subject with a flexible strap.

The data collection trigger of each IMU sensor is synchronized through a bespoke Wi-Fi router to ensure simultaneous data collection of the three sensors; the radar data were also collected simultaneously, with a manual alignment of the radar trigger with respect to the wearable sensors using a MATLAB script. The data include six human activities, namely walking (A1), sitting on a chair (A2), standing up (A3), bending to pick up an object (A4), drinking a glass of water (A5) and simulating a frontal fall (A6). These activities are shown in the sketches in Fig. 2. The top row presents separate activities, collected as separate files with fixed durations and breaks in between. The following three rows show the continuous activities, performed one after the other with random duration and transitions. The overall duration of a single sequence of activities was 35 seconds, and three different types of sequences were considered, with a different order of the six activities, as shown in Fig. 2, to manage the classification of different transitions.

For each of the 16 participants, the 3 different sequences of continuous activities were recorded, for a total of 48 radar recordings. The data for each recording have 28 degrees of freedom, as there are 3 IMUs (wrist, waist, ankle) with 3 sensors (accelerometer, gyroscope, and magnetometer), each with 3 axis (X-Y-Z) data, plus the radar data.

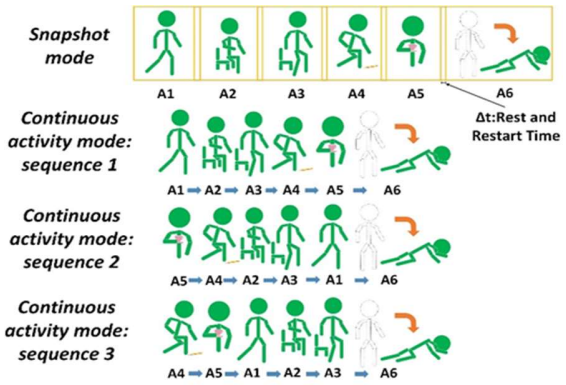

Fig. 2. Sketch of the human activities recorded - top: snapshot mode, bottom: continuous activity mode from sequence 1 to 3.

III. Data Preprocessing and Feature Extraction

A. Radar Signal Processing

The motivation of the pre-processing is to generate low noise data for feature extraction and further classification. For radar, three steps were followed. First, a FFT (Fast Fourier Transform) was applied to the raw data matrix to convert it into Range-Time domain. After that, a sharp notch filter was utilized to remove background static clutter such as furniture and walls,

assuming cut-off frequencies of ±0.0075Hz. This procedure also highlighted the range bins containing the target signature, on which Short-Time Fourier Transform (STFT) was applied with a 0.2s Hamming window and 95% overlapping to generate micro-Doppler signatures.

B. Inertial Sensor Data Processing

Some DC components exist in the inertial sensor data, for instance, the gravity effect when the accelerometer tries to measure the acceleration and the earth original magnetic field strength on top of the magnetic field strength changes due to the activities. Apart from this, noise from outside vibration and tilting could also influence the performance of inertial sensors. To address this, a simple bandpass filter was used to mitigate the DC components and human-induced vibrations. Prior to the filtering, a FFT was required to plot the spectrum for selecting the right cut-off frequencies, which were set at 0.1Hz for the lower frequency band and 25Hz for the higher band.

C. Feature Extraction

TABLE I LIST OF RADAR FEATURES

Radar features	*Number of features*
Mean, standard deviation, skewness, and kurtosis of the centroid of the Doppler spectrogram	4
Mean, standard deviation, skewness, and kurtosis of the bandwidth of the Doppler spectrogram	4
Two-dimensional mean, standard deviation, skewness, and kurtosis of the whole segment of the spectrogram	4
Mean and standard deviation of the first left and right eigenvector of the SVD decomposition of the spectrogram	4
Sum of pixels of the entire left and right matrices	2
Mean of the diagonal of the left and right matrices	2

TABLE II LIST OF INERTIAL SENSOR FEATURES

Time domain	**#**	**Frequency domain**	**#**
Mean	3	Spectral Power	9
Variance	3	Coefficients Sum	3
Standard Deviation	3	Spectral Entropy	3
Skewness	3		
Kurtosis	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		
Autocorrelation(Mean,STD)	3		
Cross Correlation(Mean,STD)	3		
Average of the absolute value of each axis	1		
Number of features	**43**	**Number of features**	**15**

For the radar data, features successfully used in our previous work were used [22], [26]and are listed in Table I. These include moments and statistical descriptors extracted from the centroid and bandwidth of the spectrogram matrix, as well as from its SVD (Singular Value Decomposition) form.

For the inertial sensors, 58 features listed in Table II were used. These can be divided into time and frequency domains [29], where the former includes various statistical moments of the time series of wearable data, and the latter considers power spectral densities in different bands and cross-correlation across data from different axes (e.g., X and Y accelerometer).

D. Feature Selection

Fusing features from the radar sensor and the three wearable sensors yield 194 features (58 x 3 for the wearables, plus 20 for the radar). To decrease the computational load and select only the most relevant and informative features, feature selection is applied [30]. In particular, in this paper, we use Sequential Backward Selection (SBS) in conjunction with a quadratic SVM (Support Vector Machine) classifier, whereby the selection process starts with all available features and progressively drops some until this yields a performance improvement. A threshold of 50 remaining features was chosen as hard stop criterion for the algorithm, approximately 25% of the initial set. In Fig. 3 the classification accuracy as a function of the number of features dropped in the SBS process is shown. Approximately 93% accuracy is achieved by selecting 57 features (i.e., dropping 137 features), where 46 features are from IMU, and radar contributes the remaining 11. The process yields an increase in accuracy of about 3.1% compared to using all features, and 60% saving in computational time.

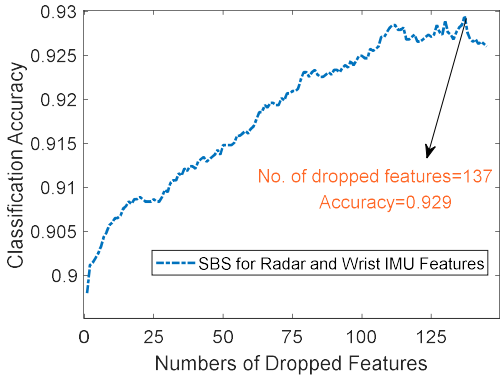

Fig. 3. Sequential backward selection for the continuous activity stream.

IV. CLASSIFICATION RESULTS FOR CONTINUOUS ACTIVITY PATTERN

A. Sliding Window Method

To process continuous data, a conventional sliding window method can be used to divide them into shorter segments for feature extraction and classification. For this, window sizes between 2s to 5s with a 0.5s interval and overlap from 0% to 90% were tested, considering the feature fusion of FMCW radar plus each IMU sensor (wrist, waist, and ankle) separately. The highest overlapping was set at 90% not to increase too much the number of segments over which feature extraction and classification had to be performed. “Leave one participant out” (L1PO) cross-validation method was used, whereby the SVM classifier was trained with data from 15 participants and tested with data from the 16th unknown subject, with the process repeated 16 times for all available people and data, and averaged. This cross-validation method is much closer to the real-world application scenario compared to traditional ‘Hold out’ or ‘k fold’ partition methods because there is no opportunity for the classifier to be trained with data from the actual end-user. The results are reported in the heat maps in Fig.

4. For the radar-only case, the accuracy reaches 83.82% with a window size equal to 4s and 90% overlap, whereas the fusion of radar and wrist IMU features yields the best performance out of the three different combinations. The improvement is approximately 6% when using a 3.5s window and the same overlapping factor. Generally, it appears that the classification accuracy is proportional to the overlapping factor, and the optimal point is typically with a medium-sized window.

Fig. 5 and 6 show the confusion matrices for the cases of the radar only, and radar plus wrist IMU fusion, where the column elements represent the predicted class and the row elements represent the true class. For the FMCW radar, the main misclassification is between 'A4' and 'A5', so that over 15% of the activity 'picking up an object' have been misclassified to 'drinking water,' and vice versa. Besides that, the activity 'A2' and 'A3' do not have high sensitivity, and some minor errors occur between the most significant class 'A6' (fall) and other activities. Fig. 6 shows the improvement obtained by combining the IMU on the wrist and radar at the feature level. The sensitivity of all the classes shows an improvement from 1.2% to 21.7%, and the misclassifications between activities are reduced to a lower level. However, there is still scope for potential improvement, such as the classification rate in class 'A2' and 'A4', and the false alarms in 'simulating fall A6'.

B. Bi-LSTM-based Deep Neural Network

LSTM [11], [12] is a type of recurrent neural network (RNN) known for the capability of modeling time-series of data. The basic component of the architecture is a simple LSTM cell [31] containing three gates, namely, input gate, forgot gate, and output gate. The "input gate" decides which information to remember, and the "forget gate" selects the information to drop. The "output gate" is a process to evaluate which input in the memory could become the output. The LSTM network can provide a prediction at each time unit, potentially generating predictions as the sensors are sampling and generating the data over time.

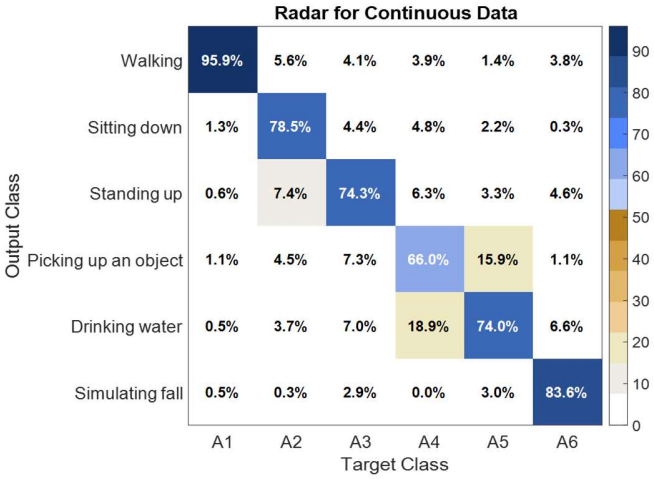

Fig. 5. Confusion matrix of radar for a continuous activity pattern using a sliding window.

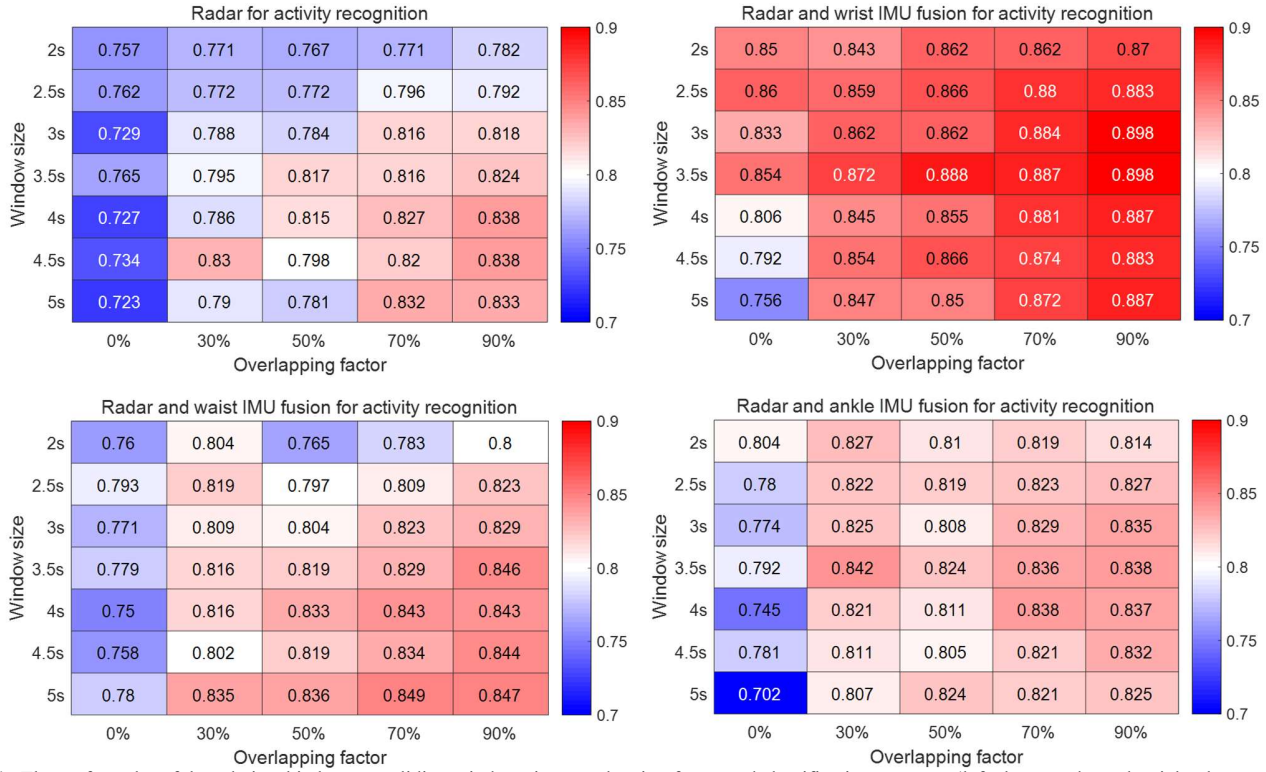

Fig. 4. The surface plot of the relationship between sliding window size, overlapping factor and classification accuracy (left above: radar-only, right above: radar and wrist IMU fusion, left below: radar and waist IMU fusion, right below: radar and ankle IMU fusion).

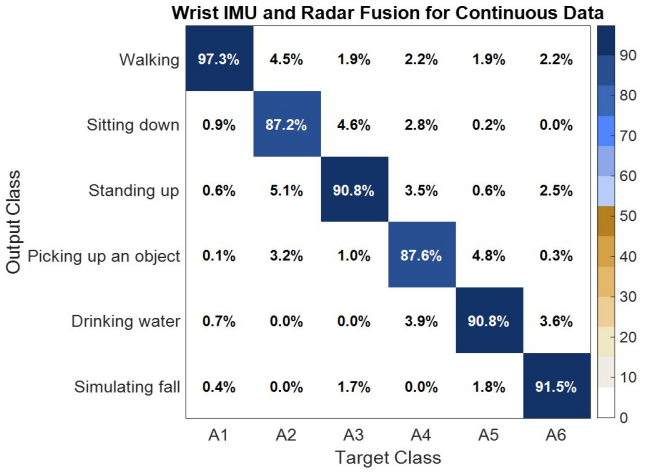

Fig. 6. Confusion matrix of wrist IMU and radar fusion for continuous activity stream using a sliding window.

sequence input layer, multiple bi-LSTM layers and one soft-max layer. It takes sequence data (e.g. continuous speech signals, real-time human motions) as inputs, the input dimension is $x \times \tau$, where x is the sensor data with different degrees of freedom, and τ is the time bin (in our case, 1741-time bins from 35s human motions data). The output dense layer, also referred as a soft-max function, turns the output vector from the network into an equal-length probability matrix. The class yielding the highest probability is chosen to be the prediction label. This numbers in this matrix are also known as scores or confidence levels of the classifier. In our application, this type of network is potentially more effective than regular RNNs to deal with different orders of the same activities in a given data frame to classify.

To address the limitation of the conventional sliding window method and increase the classification performance, a double layer Bi-LSTM network architecture is utilized to classify continuous activities patterns and validated with L1PO method.

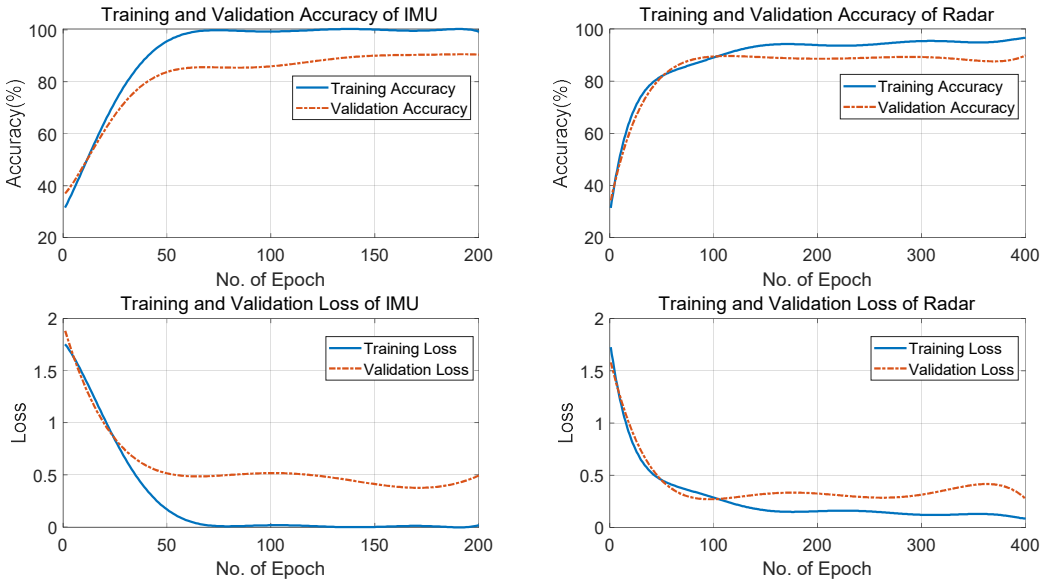

Fig. 7. Training progress of IMU and radar using double-layer bi-LSTM (top-left: training and validation accuracy of inertial sensor, top-right: training and validation accuracy of radar, bottom-left: training and validation loss calculated by cross entropy for inertial sensor, bottom-right: training and validation loss for radar.

TABLE III THE HYPER-PARAMETERS OF LSTM NETWORK TRAINING

Hyper-parameters	Radar	Wearable Sensors
Training: Validation: Test	14:1:1	14:1:1
SGD Optimizer	Adam	Adam
Decay	0.9	0.9
Initial Learning rate	1e-3	1e-3
Learning rate drop period	200	100
Number of input dimension	8	9
Number of bi-LSTM layers	2	2
Number of dropout layers	2	2
Dropout probability	0.5	0.5
Training epochs	400	200
Validation frequency	Once per epoch	Once per epoch

In terms of the time-dependent classification task, both past and future input features for a specific period can be useful information. Hence, the bi-LSTM layers proposed by Graves [32], [33] are chosen, as they can learn the backward and forward long-term dependencies between small timestamps of the data sequence. Bi-LSTM network is comprised of a

In the previous section, we have shown that the IMU on the wrist provides the most accurate results when used with radar. Thus, this is selected, and the other two IMUs (waist and ankle) discarded. A validation dataset for the Bi-LSTM networks is also selected to support the training process. This contains all the sequences of activities from one participant, and it is used to search optimal hyper-parameters and to fine-tune after the initial training. To follow L1PO approach, the data from 14 subjects are used for training and the data from the remaining 16th subject for testing, with the process repeated for all subjects (training to testing ratio 14:1). The other hyper-parameters are listed in Table III for the radar and the wearable network, respectively. The initial training rate is fixed at 0.001, and the learning rate automatically drops to 10% of the original value when the training iterations reach half of the total. Each bi-LSTM layer is followed by a dropout layer with 0.5 dropping rate for preventing overfitting.

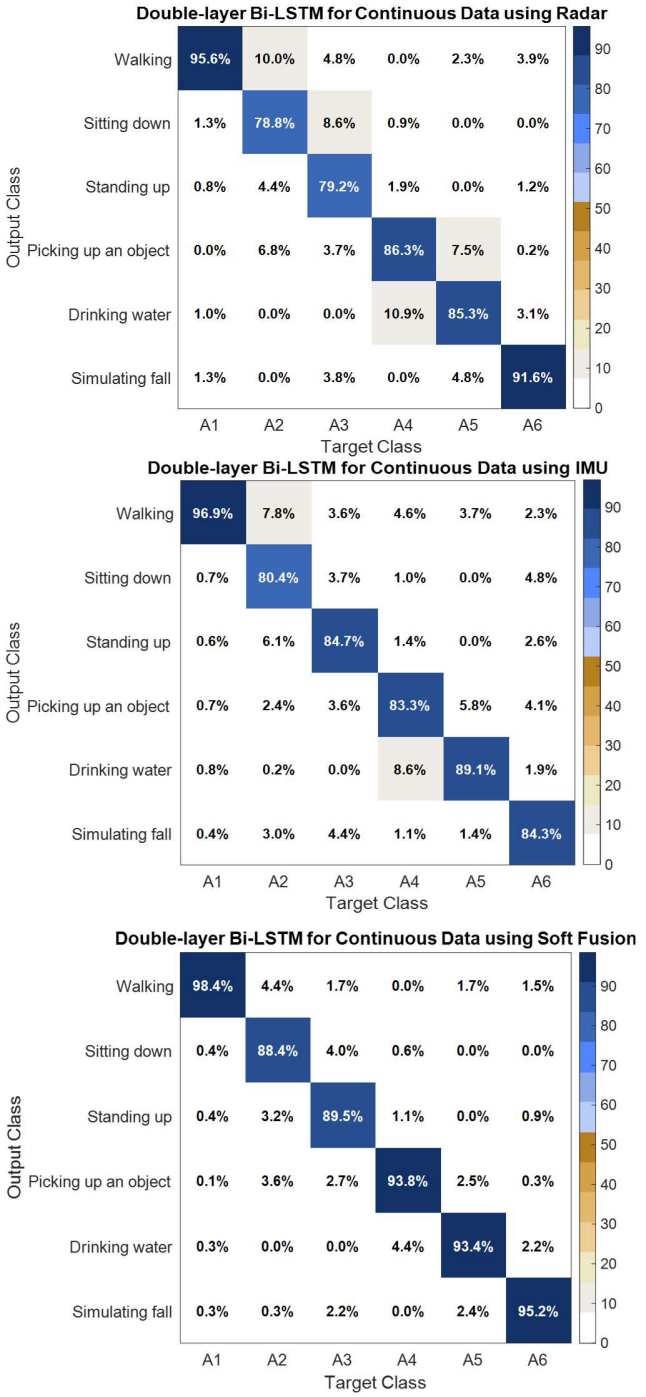

Fig. 8. (a). The confusion matrix for a bi-LSTM network using radar (b).using wrist IMU (c). using equal weight fusion.

The features used as inputs of the proposed Bi-LSTM are metrics extracted from the original radar spectrograms and wearable data as a function of time. For the radar sensor, Doppler centroid and bandwidth are considered, together with upper and lower envelopes, mean, standard deviation, skewness, and kurtosis calculated for each time bin of the spectrogram [34], [35]. For the Wrist IMU, the data include 9 features with the X, Y, and Z axes data of accelerometer, gyroscope, and magnetometer. Fig. 7 shows the training and validation accuracy and loss curves as a function of epochs. The network processing IMU data converged more rapidly (about 200 epochs) than the corresponding network processing radar data (400 epochs). The validation accuracy achieved 90% in both cases, whereas the IMU training accuracy was almost 100% and radar one was around 98%.

Fig. 8 (a)-(b) present the confusion matrices of using radar and wearable IMU separately, where the average accuracy for 'leaving one participant out' method is 88.9% and 89.1%, for radar and wrist IMU respectively. The main misclassification for radar is between 'picking up an object' and 'drinking water,' whereas 10% of 'sitting' patterns are misclassified to 'walking' and 8.6% 'standing' are misclassified to 'sitting.' For the IMU network, approximately 15.7% of fall events are not correctly detected, and about 10.3% of other activities trigger false alarms. It appears that the current activity can be often confused with the last and next activities, as a consequence of using a bi-directional LSTM based network where the prediction is influenced by the "memory" of previous and successive events. This causes an offset in the sequence of predicted activities with respect to the ground truth, or in other words, an offset prediction of the position of the transition between activities.

C. *Soft Fusion with Radar and Wearable IMU using LSTM*

Fusion at decision level between the results of the radar and wrist IMU networks is then considered to improve results further. Each classifier yields a scoring matrix as output of the Softmax layer in terms of posterior probabilities to characterize the confidence level that the network chooses a specific class as the correct output class. Soft fusion [26], [36] is a process to generate the new prediction label by incorporating the scores from separate sensors, in this case, radar and wrist IMU. The following Eq. 1 summarizes the weighted combination of the two sensors' scores mathematically, where W_R denotes the weight on radar and W_I is the weight related to the inertial sensor. Radar and wrist IMU are set initially to the same weight (W_R=W_I). $S_R(\tau, c)$ is the radar score matrix for prediction corresponding to time bin τ and class c, whereas $S_I(\tau, c)$ and $S_F(\tau, c)$ are for the inertial sensor and fusion respectively.

$$S_F(\tau,c) = W_R \cdot S_R(\tau,c) + W_I \cdot S_I(\tau,c) \quad (1)$$

Soft fusion benefits from a low computational load and can still provide significant improvement. In our case, as shown in the confusion matrix in Fig. 8(c), the average classification accuracy increases to 94.7%, compared to using wrist IMU or radar individually, with an improvement of approximately 5.5%. The performance of each class is boosted by the fusion, especially the fall detection rate, where the gain is around 10.9%. Apart from that, the misclassified events between two neighbour classes are in general reduced. Hence, soft fusion appears to increase the capability of recognizing the transitions between activities.

D. *Hard Fusion with Radar and Wearable IMU using LSTM*

Hard fusion [37] uses the prediction results from radar and wrist IMU directly, rather than combining and weighting their confidence levels. Typical hard fusion approaches include majority voting (MV), recall combiners (RC), and Naïve Bayes (NBC) combiners [26], [37]–[39], where majority voting works well only when there is an odd number of classes to avoid decision clashes. Recall combiner is an optimal combiner,

where the possibility of a certain class being selected as the true class $P(C_k|d)$ is derived by the Eq. (2). The sensitivity of the class of interest is separated from the confusion matrix, and the remaining classes are considered as one 'united class'. Meanwhile, it is assumed that the misclassification is shared equally among those remaining classes. To conclude, the output probability of RC depends on the sensitivity or recall level of each classifier.

$$P(C_k \mid d) = P(C_k) \cdot \prod_{m \in M_+^k} p_{mk} \cdot \prod_{m \in M_-^k} \frac{1 - p_{mk}}{C - 1} \quad (2)$$

Assume that a classifier ensemble contains N distinct classifiers and the number of class to be identified is equal to C. k is an integer index between 1 to C to indicate the class of interest (e.g., C_1 is class 1). $P(C_k)$ denotes the number of classifiers which support class C_k as the true class in terms of a supporting rate. M_+^k represents the classifier ensemble which support class C_k, whereas M_-^k denotes the classifier ensemble which support other classes, m is the classifier ID. If the classifier supports class C_k, then the output will be the product of $P(C_k)$ and the confusion matrix element p_{mk} (classifier m, row and, column k). Hence in this case, p_{mk} is the recall of this class, otherwise, $P(C_k)$ is multiplied with the shared misclassification $\frac{1-p_{mk}}{C-1}$.

RC has a prominent limitation, i.e., the fact that the misclassification probability is divided equally to each class, whereas in real testing scenarios, the misclassification always varies for each class. In other words, the contributions of different classes to the total classification error are not equal. To address this, a robust Naïve Bayes combiner is exploited to consider the actual misclassification rates for each class. The output possibility of every class is associated with the recall of the interested class and the misclassification between the class chosen by the classifier and the class of interest; it is shown in Eq. 3.

$$P(C_k \mid d) = P(C_k) \cdot \prod_{m=1}^{N} p_{m,C_m,k} \quad (3)$$

The output is the product of the classifier supporting rate and the element of the confusion matrix (classifier m, row C_m and column k), where C_m refers to the prediction label of classifier m.

Theoretically, RC and NBC are all optimal combiners. The performance of the RC is proportional to the number of classes and number of classifiers, whereas the gain of NBC is not as high as RC. Besides that, NBC is not suitable for high noise level data, and the computational intensity in terms of the number of parameters per observation for NBC ($N*C^2+C$) is much higher than RC ($N*C+N$) [37].

E. Proposed Soft and Hard Fusion Integration Method

In addition to the soft and hard fusion schemes described in the previous two sections, a novel approach is proposed and used in this paper to leverage the strengths of both. This method uses the soft fusion results as an 'additional classifier' for the basic architecture of the recall and Naïve Bayes combiners. Furthermore, the classification results of weighted soft fusion are used to implement more "virtual classifiers" for the hard fusion, where the information ratio of radar and IMU in these classifiers is varied from 1:0.1 to 0.1:1. The "virtual classifiers" are used to leverage performance advantages of different fusion ratios with data from the original two classifiers (for radar and wrist IMU), but saving training and computation time and effort that new real classifiers would require.

Table IV summarizes the number of classifiers used together as an ensemble at the input of the hard fusion. The conventional hard fusion using the predictions from the radar and wrist IMU classifiers has length 2. The proposed method A adds the soft fusion results as the third classifier, hence increasing L to 3. The proposed methods B and C includes additional "virtual classifiers" by adding more soft fusion results calculated with different ratios of weights for the two sensors. In the former case (B), the step in changing this ratio is 0.2. Hence, 10 additional classifiers are added for a total of 13. In the latter case (C), the step in changing the soft fusion ratio is 0.1. Hence, 18 additional classifiers are added for a total of 21 in the hybrid fusion approach. The number of joint classifiers is listed in Table IV. The ratio needs to be chosen carefully, attempting to reach an optimal balance between covering all the necessary fusion ratios to leverage on information from radar and wrist IMU, but without generating too many additional classifiers that may not be providing useful information. The values of 0.1 and 0.2 were selected through a series of empirical tests.

Fig. 9 shows the average classification performance with respect to the number of different classifiers used as inputs of the hard fusion, as shown in Table IV; note that the X axis is in logarithmic scale. The results are generated using 'leaving one participant out' (L1PO) cross validation. It can be seen that the Naïve Bayes combiner outperforms the recall combiner for all cases, even if the difference in absolute terms is small. The optimal fusion result is the proposed method B (see table IV) with NB combiner, yielding approximately 95.8% accuracy. There is no further gain in increasing the number of classifiers beyond 13, but the most significant improvement is obtained when adding the soft fusion to the classifier ensemble (i.e., number of classifiers for hard fusion increased from 2 to 3), with approximately +0.94% for NBC and +1.2% for RC in terms of accuracy increase.

TABLE IV NUMBER OF CLASSIFIERS USED AS INPUT OF THE PROPOSED HARD FUSION SCHEME

Classifier ensemble length	Inputs of the combiner
L=2 (Normal hard fusion)	Radar, wrist IMU
L=3 (Proposed method A)	Radar, wrist IMU, normal soft fusion
L=13 (Proposed method B)	Radar, wrist IMU, normal soft fusion, weighted soft fusion with 10 different ratios
L=21 (Proposed method C)	Radar, wrist IMU, normal soft fusion, weighted soft fusion with 18 different ratios

The confusion matrix for the best fusion approach is shown in Fig. 10. Compared to the equal-weighted soft fusion, whose confusion matrix was shown in Fig. 10, the proposed fusion B yields an improvement in accuracy of about 2.9% and 2.7% for classes 'A2' and 'A3', whereas the sensitivity of fall detection also increases by 0.7%.

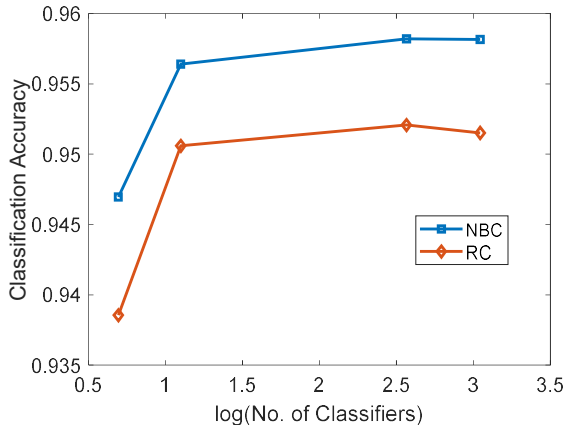

Fig. 9. Number of classifiers versus classification accuracy for the proposed hybrid fusion scheme.

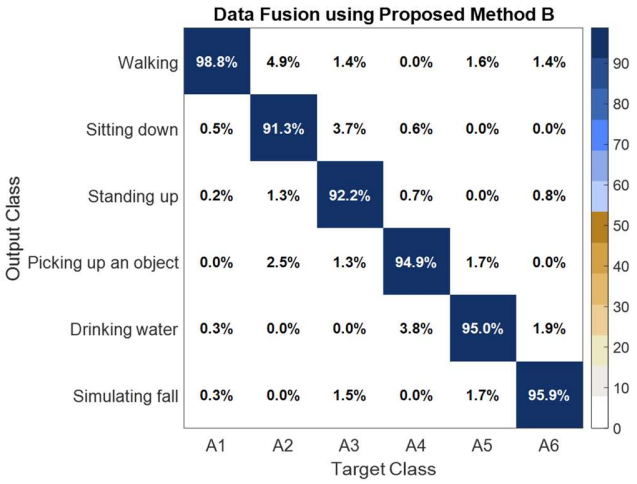

Fig. 10. The confusion matrix of proposed hybrid fusion method B.

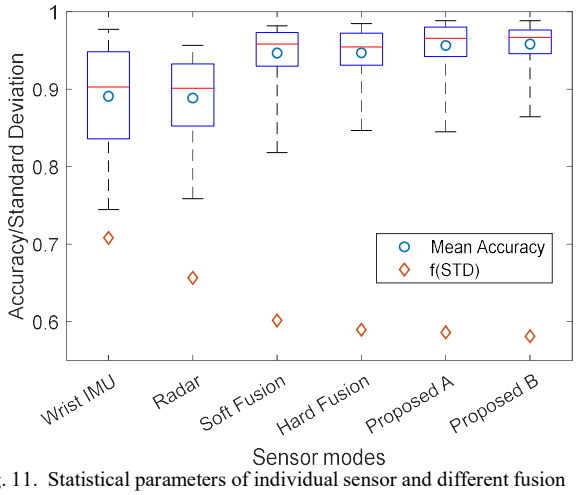

Fig. 11. Statistical parameters of individual sensor and different fusion methods.

To further analyze the performance of the proposed approaches, Fig. 11 shows statistics of the classification results, namely the mean, median, 25th and 75th percentiles, maximum, minimum, and standard deviation across all the "leave one person out" classifications. Note that the standard deviation values have been linearly transformed by f(STD) to make their values comparable to the other metrics for easier visualization discussed in the previous section, but here it can be noted that the minimum values, as the worst-case scenarios across different subjects, are also increased from 74% to approximately 86%. Equally, the distance between 25th and 75th percentiles and the standard deviation across cases of different subjects also decreased with the proposed hybrid fusion, showing that the classification performances become more stable and robust across participants "unknown to the

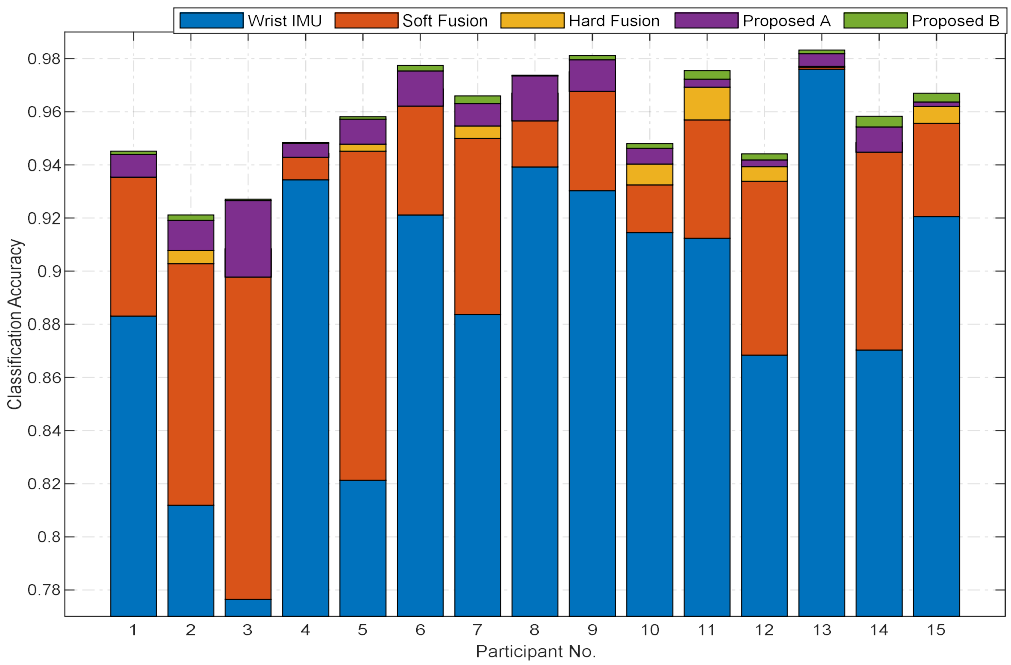

Fig. 12. The accuracy improvement for each participant with different fusion methods.

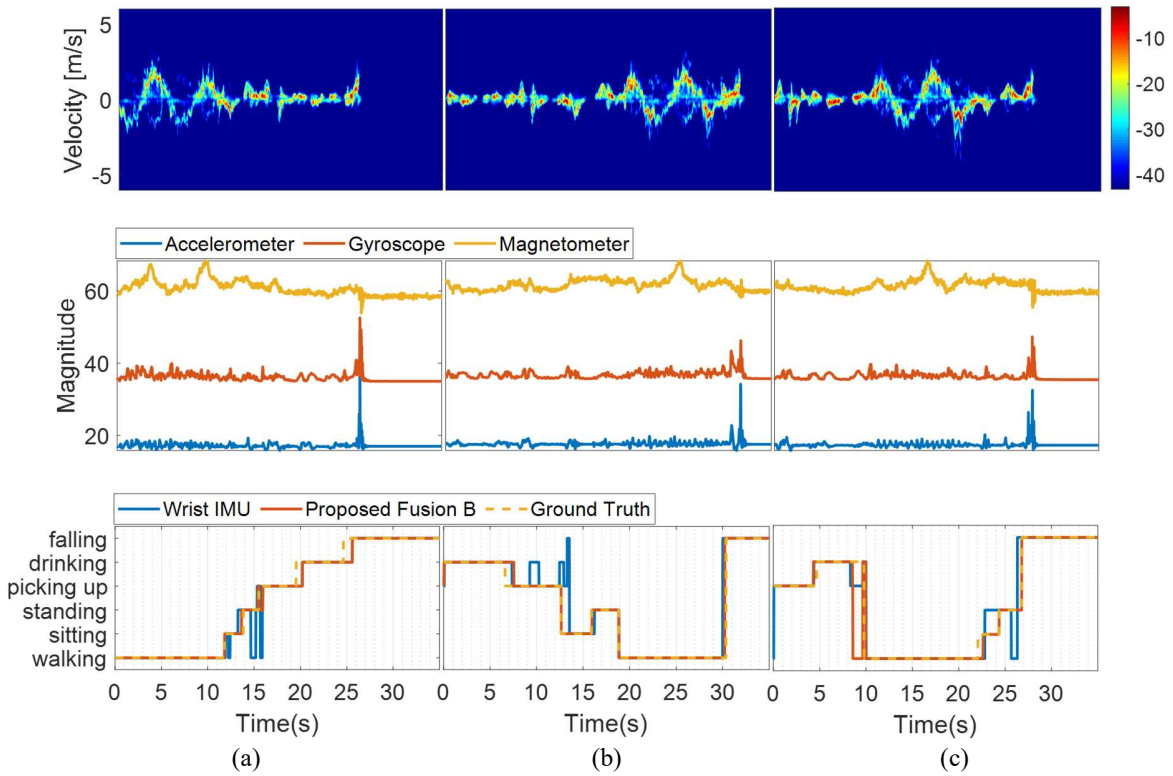

Fig. 13. The activity pattern 'tracking' (first row: radar Doppler spectrogram for activity sequence 1, 2 and 3 in (a), (b) and (c), second row: absolute value of X, Y and Z axis of the inertial sensor data, third row: prediction label of using IMU-only and proposed fusion B, as well as ground truth.

classifier." This is because soft fusion with different information weights is capable of recognizing distinctive activities, whereas the proposed hard fusion scheme maximizes the overall 'fusion gain' by exploiting the available information in the perspective of a probability combiner. (essentially a multiplication by 3 and then plus 0.55). All proposed fusion schemes increase the average accuracy as

Fig. 12 shows the improvement in classification accuracy for each participant, from the baseline case of only using a wearable sensors (blue), to the different proposed fusion schemes. The proposed scheme B appears to the best information fusion method, where 13 out of 15 participants obtain a further performance improvement upon the proposed fusion A. The results displayed in Fig. 12 confirms the observations made for Fig. 11, as to the overall reduction of accuracy variability (standard deviation) across participants thanks to the proposed fusion schemes.

Finally, Fig. 13 presents an example of input data for three sequences of continuous activities performed by one participant and the corresponding "activity tracking" provided by the developed classifiers. The top row shows the radar spectrograms with the amplitude displayed in logarithmic scale. The middle row shows the corresponding data for the wearable, as the absolute value (out of the X-Y-Z tri-axial information) for the accelerometer, gyroscope, and magnetometer. A spike due to the sudden change corresponding to the final fall activity can be seen in all three cases. The bottom row shows the "activity tracking" provided by the wrist IMU only (blue) vs. the best fusion scheme with wearable and radar (red), compared with the ground truth (dashed yellow). The proposed fusion scheme appears to correct the majority of the misclassification events occurring when only one sensor is used.

V. Conclusion

This paper discussed a framework based on multi-layer bi-directional LSTMs to implement multimodal fusion for sensing and to classify human activities. Continuous sequences of activities with random transitions were considered in this work, rather than conventional separated activity data with a fixed duration. Bi-LSTM networks allowed to avoid manual segmentation and limitations of using sliding windows, while at the same time enabling the implementation of information fusion schemes, in particular, a novel hybrid approach of soft-hard combination fusion. The bi-LSTM framework and the proposed fusion schemes were validated on data from FMCW radar and wearable sensors, corresponding to sequences of six human activities performed by 16 participants. Leave one person out validation approach was followed throughout, to demonstrate the approaches' performances when dealing with data of subjects "unknown to the classifier." The proposed hybrid approach is shown to yield an average classification accuracy of approximately 96% while improving performances and robustness across all participants (an increase of minimum value accuracy and a reduction of standard deviation).

Future work will seek to validate the method in a wider cohort of participants and activities, including a larger set of measurement environments, aspect angles with respect to the radar, and span of age and physical conditions of the participants. In terms of the implementation of the neural networks, deeper architectures can be considered with more data collected, as well as customization to the structure and hyper-parameters to avoid overfitting while managing the diversity of data for each participant and scenario. The format

of the input data also has scope for further work, considering, for example, radar data from the range-time domain as complementary or alternative to spectrograms, and other sensing modalities if available. Besides that, testing the classification model with different sensors (e.g. training with radar and three IMUs and evaluate with only radar data or cross frequency testing on different radar dataset) is very worth to explore in terms of evaluating the capability of the classifier under more complex condition and for cross-modality learning. Furthermore, particular interest is in the implementation of the hybrid algorithms on embedded platforms and in real-time, moving towards more realistic deployment conditions.

REFERENCES

[1] W. H. O. Ageing and L. C. Unit, "WHO global report on falls prevention in older age," *World Heal. Organ.*, 2008.

[2] NIHR Dissemination Centre, "HELP AT HOME Use of assistive technology for older people," *Natl. Inst. Heal. Res.*, pp. 3–6, 2018.

[3] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andrès, "From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems," *IEEE Sens. J.*, vol. 17, no. 3, pp. 812–822, 2017.

[4] 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, 2016.

[5] C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst. Man, Cybern. A Syst. Humans*, vol. 41, no. 3, pp. 569–573, 2011.

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

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

[8] D. Wu *et al.*, "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1583–1597, 2016.

[9] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 511–526, 2017.

[10] C. Ding *et al.*, "Continuous Human Motion Recognition With a Dynamic Range-Doppler Trajectory Method Based on FMCW Radar," *IEEE Trans. Geosci. Remote Sens.*, 2019.

[11] S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, 2019.

[12] J. Le Kernec *et al.*, "Radar Signal Processing for Sensing in Assisted Living: The challenges associated with real-time implementation of emerging algorithms," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 29–41, 2019.

[13] P.-H. Chen, M. C. Shastry, C.-P. Lai, and R. M. Narayanan, "A portable real-time digital noise radar system for through-the-wall imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 4123–4134, 2012.

[14] Y. Wang, Q. Liu, and A. E. Fathy, "CW and pulse–Doppler radar processing based on FPGA for human sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3097–3107, 2012.

[15] B. Vandersmissen *et al.*, "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, 2018.

[16] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, 2009.

[17] B. Erol and M. G. Amin, "Radar Data Cube Processing for Human Activity Recognition Using Multi Subspace Learning," *IEEE Trans. Aerosp. Electron. Syst.*, 2019.

[18] V. C. Chen, W. J. Miceli, and D. Tahmoush, *Radar micro-Doppler signatures: processing and applications*. The Institution of Engineering and Technology, 2014.

[19] 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, 2015.

[20] 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, 2017.

[21] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, 2013.

[22] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Magnetic and Radar Sensing for Multimodal Remote Health Monitoring," *IEEE Sens. J.*, p. 1, 2018.

[23] H. Sadreazami, M. Bolic, and S. Rajan, "CapsFall: Fall Detection Using Ultra-Wideband Radar and Capsule Network," *IEEE Access*, vol. 7, pp. 55336–55343, 2019.

[24] F. Luo, S. Poslad, and E. Bodanese, "Human Activity Detection and Coarse Localization Outdoors Using Micro-Doppler Signatures," *IEEE Sens. J.*, vol. 19, no. 18, pp. 8079–8094, 2019.

[25] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar Navig.*, vol. 11, no. 1, pp. 107–115, 2017.

[26] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A Multisensory Approach for Remote Health Monitoring of Older People," *IEEE J. Electromagn. RF Microwaves Med. Biol.*, vol. 2, no. 2, pp. 102–108, 2018.

[27] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digit. Signal Process.*, vol. 87, pp. 125–131, 2019.

[28] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[29] 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, 2016.

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

[31] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv Prepr. arXiv1508.01991*, 2015.

[32] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural networks*, vol. 18, no. 5–6, pp. 602–610, 2005.

[33] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 6645–6649.

[34] F. Fioranelli, M. Ritchie, and H. Griffiths, "Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler," *IET Radar, Sonar Navig.*, vol. 10, no. 9, pp. 1702–1710, 2016.

[35] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson, "Knowledge exploitation for human micro-Doppler classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2125–2129, 2015.

[36] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sens. J.*, vol. 16, no. 3, pp. 773–781, 2016.

[37] L. Kuncheva and J. Rodríguez, *A weighted voting framework for classifiers ensembles*, vol. 38. 2014.

[38] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-Garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inf. Fusion*, vol. 46, pp. 147–170, 2019.

[39] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Multi-classification approaches for classifying mobile app traffic," *J. Netw. Comput. Appl.*, vol. 103, pp. 131–145, 2018.