IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

Ultra-Wideband Radar-Based Activity Recognition Using Deep Learning

FARZAN M. NOORI[1], MD ZIA UDDIN[2],(SENIOR MEMBER, IEEE), AND JIM TORRESSEN[1,3],(SENIOR MEMBER, IEEE)
[1]Department of Informatics, University of Oslo, Norway
[2]Software and Service Innovation, SINTEF Digital, Oslo, Norway
[3]RITMO Centre, University of Oslo, Norway

Corresponding author: Farzan M. Noori (e-mail: farzanmn@ifi.uio.no).

ABSTRACT With recent advances in the field of sensing, it has become possible to build better assistive technologies. This enables the strengthening of eldercare with regard to daily routines and the provision of personalised care to users. For instance, it is possible to detect a person's behaviour based on wearable or ambient sensors; however, it is difficult for users to wear devices 24/7, as they would have to be recharged regularly because of their energy consumption. Similarly, although cameras have been widely used as ambient sensors, they carry the risk of breaching users' privacy. This paper presents a novel sensing approach based on deep learning for human activity recognition using a non-wearable ultra-wideband (UWB) radar sensor. UWB sensors protect privacy better than RGB cameras because they do not collect visual data. In this study, UWB sensors were mounted on a mobile robot to monitor and observe subjects from a specific distance (namely, 1.5–2.0 m). Initially, data were collected in a lab environment for five different human activities. Subsequently, the data were used to train a model using the state-of-the-art deep learning approach, namely long short-term memory (LSTM). Conventional training approaches were also tested to validate the superiority of LSTM. As a UWB sensor collects many data points in a single frame, enhanced discriminant analysis was used to reduce the dimensions of the features through application of principal component analysis to the raw dataset, followed by linear discriminant analysis. The enhanced discriminant features were fed into the LSTMs. Finally, the trained model was tested using new inputs. The proposed LSTM-based activity recognition approach performed better than conventional approaches, with an accuracy of 99.6%. We applied 5-fold cross-validation to test our approach. We also validated our approach on publically available dataset. The proposed method can be applied in many prominent fields, including human–robot interaction for various practical applications, such as mobile robots for eldercare.

INDEX TERMS Human Activity Recognition, LSTM, LDA, PCA, XeThru UWB Sensor,

I. INTRODUCTION

ACCORDING to a 2017 report by the Department of Economic and Social Affairs in the United Nations, the population of older adults is increasing more rapidly than other age groups [1]. In 2015, one out of eight people worldwide were aged 60 years or older. By 2050, the number of older adults is expected to reach nearly 2.1 billion. A major challenge in working with an ageing population is the effective delivery of healthcare services [2]. Moreover, healthcare for older adults is a matter of great concern for their relatives. This is particularly true when older adults are alone at home as they are at a risk of being affected by unforeseen circumstances, such as falls. Recently, independent living among older adults has become a significant challenge from

both social and economic perspectives. Therefore, assisting older adults with their well-being and autonomy has become a research topic of great interest [3].

Understanding the current state and context of users is crucial for assisting in their everyday lives. Human movement has been actively studied using distinguished ambient sensors [4], [5]. Previously, video-based sensors have been used for human activity recognition and fall detection [6]. However, video-based sensors often face challenges in their use owing to privacy issues. In contrast, a non-contact ambient sensor that has no such privacy issues is the XeThru ultra-wideband (UWB) radar [7]–[9]. Hence, sensors, such as UWB radars, can be used for general robot navigation sensing and emergency analysis based on human body movement

while preserving privacy, particularly for older adults living independently.

Sensors with wireless communication ability make human-machine interaction robots suitable for human behaviour and vital sign analysis [9]. Many researchers have explored ambient sensors for monitoring human behaviour and health status [10]–[15]. For instance, *CASAS* adopted machine learning tools for user behaviour analysis [10]. *GatorTech* is an earlier research project wherein many ambient sensors were used to provide user services, such as voice and behaviour recognition [11]. Zhang et al. [12] proposed an assisted living environment to help prolong the time for which older adults could live in their homes. *SWEET-HOME*, a French project, aimed at developing an assisted living technology mainly based on audio analysis [13]. Billias et al. [14] adopted ambient sensors, such as cameras and microphones, to analyse the daily activities of older users. Recently, *Lio* [15], a personal robot assistant developed by *F&P Robotics*, was introduced with a multifunctional arm. The robot could assist patients autonomously and provide several healthcare functions. During the long COVID-19 pandemic, additional functions, such as disinfection operations and remote detection of elevated body temperature, were performed by *Lio*.

Human activity recognition (HAR) and emergency detection have made significant progress in recent years through machine learning techniques [16]. Most previous HAR studies have relied on hand-crafted features, which are sometimes difficult to distinguish with sufficient accuracy to classify activities [17]. Conventional pattern recognition techniques, such as K-nearest neighbour (KNN) [18], support vector machines (SVMs) [19], artificial neural networks (ANNs) [20], and random forest (RF) [21], perform well in HAR and emergency detection. Meanwhile, in recent years, we have witnessed an incredible growth in machine learning research enabled by advancements in deep learning approaches. Deep learning has resulted in remarkable performance in many research areas, such as computer vision [22], business analytics, and natural language processing [23]. Recently, convolutional neural networks (CNNs) have shown significant improvement in classifying human activities [24]. Yang et al. [25] built a CNN-based architecture that could analyse multi-channel time-series data. A unified layer was introduced to merge multiple channels prior to classification. Moreover, the CNN-based multi-channel time-series architecture is task-dependent and is characterised by a higher discrimination accuracy for classifying human activities. Previous research has shown the use of UWB sensors to recognize human activities [26], [27]. Singh et al. [28] proposed a framework for HAR using point clouds generated by mmWave radar. Their activities were related to exercise. Sharma et al. [29] introduced a channel impulse response based HAR system which can recognize sitting, standing and lying positions.

It has been observed that the deep network structure in deep learning is more suitable than traditional machine learning approaches for supervised and incremental learning [30]. Thus, deep learning is an ideal approach for analysing
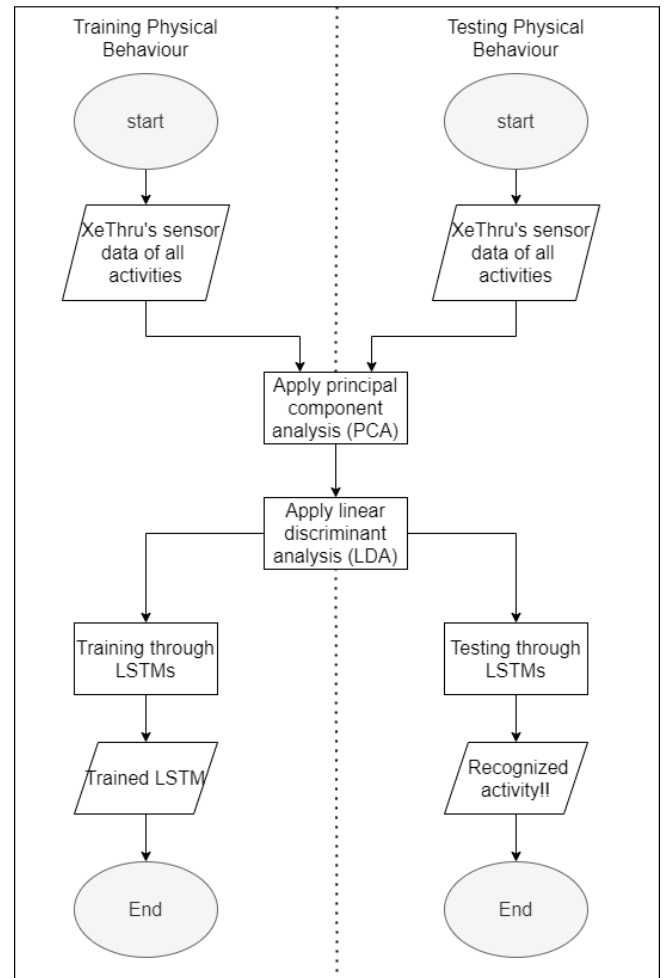

FIGURE 1: Basic flows of the proposed enhanced discriminant features–based activity recognition system using LSTM.

human behaviour and health status using data from newly introduced sensors, such as XeThru UWB radars [8], [9], [31]. A recurrent neural network (RNN) is one of the most popular deep learning techniques for time-series data [30], [32]. An RNN is adopted to decode time-sequential data for modelling various events, such as an emergency due to an unusual heart rate. Therefore, a special type of RNN, namely long short-term memory (LSTM), is proposed in this study to classify the XeThru UWB radar data. The performance of conventional approaches, such as SVM, AdaBoost, multilayer perceptron (MLP), quadratic discriminant analysis (QDA), KNN, RF, and decision trees (DTs), have also been evaluated and compared with that of LSTM. Furthermore, as XeThru UWB radar sensors collect a large number of data points in a single frame, principal component analysis (PCA) and linear discriminant analysis (LDA) are introduced for dimensionality reduction. Figure 1 illustrates the basic flows of the proposed system in two steps, namely training (left-hand side) and testing (right-hand side).

In this study, we investigate whether a XeThru UWB radar sensor can recognise different complex human activities.

Thus, the contributions of this work are two-fold:

- A XeThru UWB radar sensor is used with a novel LSTM-based approach to classify activities.
- Enhanced discriminant analysis (EDA), combining PCA with LDA, is proposed to reduce data dimensionality and extract significant features before feeding them into the classifiers.

METHODOLOGY
SUBJECTS

This study aimed to classify five different activities using a UWB sensor. Overall, 13 participants were included in the study with six of them being female subjects. All of them were normal healthy people. The ages of the participants vary from 22 – 50 years. All subjects voluntarily participated in the experiments, and written consent was obtained from all subjects before participation. The experiments and data collection were approved in advance by the Norwegian Centre for Research Data (NSD). All experiments were performed in accordance with relevant guidelines and regulations.

DATA ACQUISITION FROM UWB RADAR

The UWB radar has been used for imaging in sensing through walls, [33]–[35], detecting humans [36], [37], assisting in public security [38], and recognising moving subjects [39], [40]. In the current study, we used XeThru X4, a compact impulse-radio UWB radar system on a chip, as shown in Figure 2. The radar is configurable and provides developers with a high degree of freedom to develop new applications, ranging from basic presence detection to vital sign analysis. The pulse transmitted by the radar can be configured within two bands, namely the lower and upper bands. The lower pulse generator enables transmission within the 6.00–8.50 GHz band, whereas the higher pulse generator enables transmission within the 7.25–10.20 GHz band. To capture the reflected energy, the radar applies a high-speed sampler with a sampling rate of 23.328 GS/s, which can sample up to 1,536 samples [8]. The distance from the radar to an object is called the slant range, which can be determined by

$$R = \frac{CxT}{2},\qquad(1)$$

where C is the speed of light and T is the time required for signal reflection. The divisor 2 is used because the radar signal travels to the target and then travels the same distance back to the radar. The radar system is dependent on waveform design in several ways. The range resolution is proportional to the bandwidth, and the signal-to-noise ratio (SNR) of the output signal is directly proportional to the waveform energy. In contrast, the signal wavelength affects the radial velocity resolution [41]. Because of its short duration, good spectrum coverage, and ease of implementation in CMOS, a frequency-shifted Gaussian pulse can be considered as an excellent candidate for a UWB. The frequency-shifted Gaussian pulse can be determined by
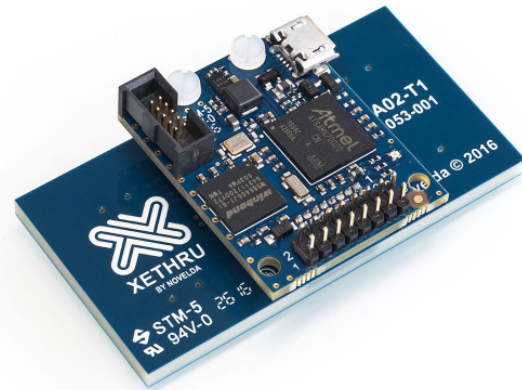

FIGURE 2: Ultra-wideband Xethru X4 sensor [8].

$$g(t) = p(t)\cos(\omega_c t) = V_{TX}\exp\left(-\frac{t^2}{2\tau^2}\right)\cos(\omega_c t),\quad(2)$$

where $p(t)$ denotes the Gaussian pulse envelope and τ determines the $-10dB$ bandwidth.

$$\tau = \left(2\pi f_B (\log_{10}(e))^{1/2}\right)^{-1}.\qquad(3)$$

The pulse amplitude V_{TX} is dependent on the regulatory limits for the peak and average output power [42]. For swept-threshold (ST) sampling, the sweep time is dependent on the number of pulses (n_{pulses}) and the pulse repetition frequency (PRF).

$$t_{\text{sweep}} = \frac{n_{\text{pulses}}}{\text{PRF}}.\qquad(4)$$

The matched filter radar equations are used to obtain the SNR for an ideal pulse-based radar receiver (RX) from a target, given its range R and radar cross-section (RCS) σ_{RCS} [43]:

$$\text{SNR}_{\text{ideal}} = \frac{P_t \cdot t_p \cdot n_{\text{pulses}} \cdot G^2 \cdot \sigma_{\text{RCS}} \cdot \lambda^2}{k_B \cdot T_0 \cdot F \cdot (4\pi)^3 \cdot R^4},\qquad(5)$$

where G is the antenna gain, k_B is Boltzmann's constant, λ is the wavelength, T_0 is the temperature in Kelvin, P_t is the transmitted pulse power, F is the RX noise factor, and t_p is the pulse duration. The transmitted pulse is approximated as a rectangular windowed pulse with length t_p such that the energy of the pulse E_p equals

$$E_p = P_t \cdot t_p = \frac{V_{\text{TX}}^2}{2Z_L} \cdot t_p.\qquad(6)$$

In an ST-based pulse radar, (5) becomes

$$\text{SNR}_{\text{ST}} = \frac{P_t \cdot t_p \cdot n_{\text{pulses}} \cdot G^2 \cdot \sigma_{\text{RCS}} \cdot \lambda^2 \cdot G_{\text{ST}}}{k_B \cdot T_0 \cdot F \cdot (4\pi)^3 \cdot R^4 \cdot n_{\text{steps}}},\qquad(7)$$
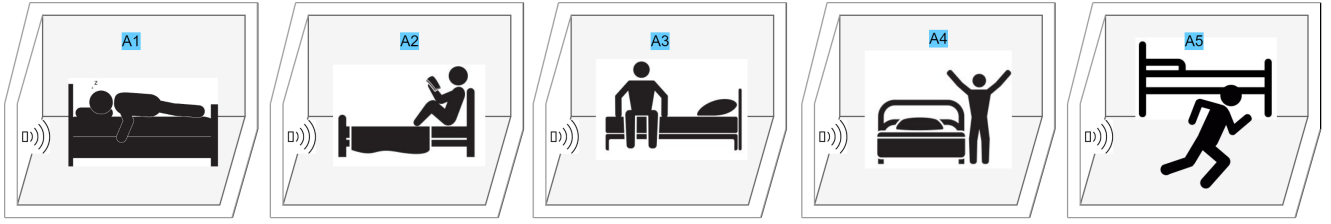
FIGURE 3: Schematic setup of the five indoor activities in the experiment.

FIGURE 4: An A3 posture and the robot with a UWB sensor.

where n_{steps} is the number of steps in the threshold sweep that causes an SNR loss and G_{ST} is the SNR gain obtained from multiple threshold levels covering the noise region around the signal value, which results in an integration effect. The largest signal in the sampled frame can be used to determine the threshold sweep range, which occurs at the minimum distance, in addition to the noise. Assuming $5 \cdot sigma_{noise}$ is sufficient as the maximum noise voltage, the number of threshold steps required is

$$n_{\text{steps}} \geq \frac{2 \cdot (V_{\text{RX}} \cdot A_{\text{FE}} + 5 \cdot \sigma_{\text{noise}})}{q} - 1, \quad (8)$$

where A_{FE} is the voltage gain of the RX front end, and the received pulse amplitude V_{RX} is given by

$$V_{\text{RX}} = V_{\text{TX}} \cdot \sqrt{\frac{G^2 \cdot \sigma_{\text{RCS}} \cdot \lambda^2}{(4\pi)^3 \cdot R^4}}. \quad (9)$$

The ST-SNR loss can be expressed as

$$L_{\text{ST}} = 10 \cdot \log_{10}((V_{\text{sweep}} + q)/(1.772 \cdot \sigma_{\text{noise}})). \quad (10)$$

The major advantage of using ST rather than a multi-bit system is that it only requires a 1-bit quantiser, which can increase the inherent linearity of the system and simplify the design. Moreover, with no reduction in the SNR, ST can be operated over longer consecutive ranges.

In our study, we used a UWB impulse radar, which can sense vital sign data, such as presence, breathing patterns, and movement from people who are either sitting on a bed or walking around. The UWB signal returns to obtain the required echo matrix with frames corresponding to each range bin, as shown below [44]:

$$\mathbf{M} = \{r^{(x)}[y] \ : \ x = 1, 2, .., X; y = 1, 2.., Y\}, \quad (11)$$

where x and y denote the range and time, respectively, while X is the extent of the range and Y is the extent of the time span of the data. Hence, the total size of the data is XY. The frame M further passes through PCA, which is discussed in more detail in the $Dataset and Metrics$ section.

DATASET AND METRICS

The obtained dataset comprised five normal activities: lying ($A1$), sitting on the bed with the legs on the bed ($A2$), sitting on the bed with the legs on the floor ($A3$), standing ($A4$), and walking ($A5$). An illustration of each activity is given in Figure 3. Figure 4 shows an A3 posture wherein a robot with a UWB sensor monitors a subject.

The possible classification outcomes are based on true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). To measure the performance, we used the following metrics.

Accuracy is the ratio of the number of correctly predicted observations to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

Precision is the ratio of the number of correctly predicted positive observations to the total number of predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

Recall is the ratio of the number of correctly predicted positive observations to the total number of observations in the actual class.

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

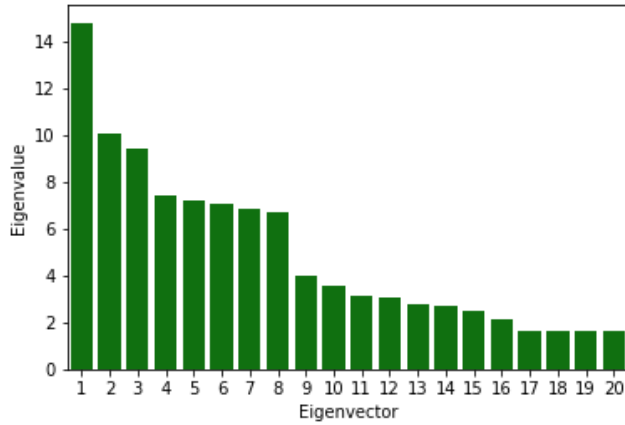F1-score is an overall measure of the accuracy of the model combining precision and recall.

IEEE *Access*

FIGURE 5: Top 20 eigenvalues corresponding to the eigenvectors.

FIGURE 6: 3D scatter plot of the features after applying PCA.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \qquad (15)$$

FEATURE EXTRACTION USING PRINCIPAL COMPONENT ANALYSIS (PCA)

We apply Gaussian kernel–based PCA to the input data to approximate the original data with fewer dimensions. PCA focuses on the direction of the maximum covariance in the new feature space. It reduces the dimensions by focusing mainly on the essential variations in the data. Because the data are nonlinear, a Gaussian kernel–based PCA was used [45]. The covariance matrix of the data is defined as

$$K = \frac{1}{N} \sum_{i=1}^{N} (\nabla(M_i).\nabla(M_i)^T), \qquad (16)$$

$$\nabla(M_i) = \gamma(M_i) - \bar{\gamma}, \qquad (17)$$

$$\bar{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \gamma(M_i), \qquad (18)$$

where $\bar{\gamma}$ is a Gaussian kernel and N is the total number of events in the activity period. Eigenvalue decomposition can be applied as

$$K = E^T \alpha E, \qquad (19)$$

where E represents the principal components, α represents the eigenvalues, and K is the diagonal matrix of the eigenvalues. Then, the features for an event can be represented by projection of the principal components as

$$P = M E_m^T. \qquad (20)$$

The size of the matrix E becomes $t \times m$, where t is the dimension of each vector, m is the number of principal components to be considered, and K is an $m \times m$ diagonal matrix. Moreover, E reflects the original coordinate system onto the eigenvectors. The eigenvector corresponding to the
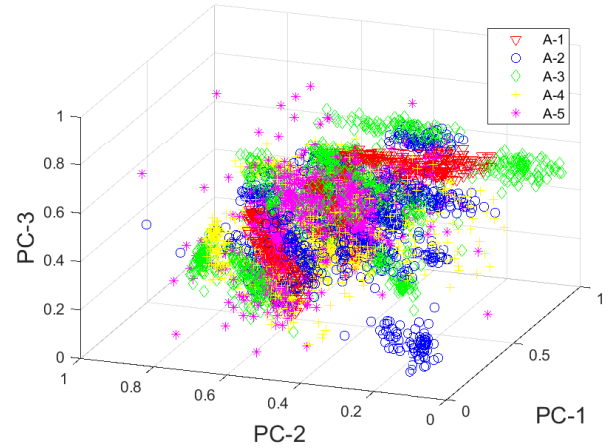
largest eigenvalue indicates the axis of largest variance, and the eigenvector corresponding to the next largest eigenvalue indicates the axis that is orthogonal to the first indicating the second largest variance, and so on. Typically, eigenvalues close to zero have negligible variance and can thus be excluded. Hence, the m eigenvectors corresponding to certain large eigenvalues can be used to define the subspace. Figure 5 shows the top 20 eigenvalues corresponding to the first 20 eigenvectors.

PCA is a second-order statistics-based method of analysis that represents global information [46]. Applying PCA to human activities produces global features that represent frequently moving parts of the human body engaged in various activities [47], [48]. Because we apply EDA, which is a combination of PCA and LDA, for dimensionality reduction, the principal components extracted through PCA are passed through LDA.

FEATURE EXTRACTION USING LDA

LDA is popular in supervised classification approaches. It creates hyperplanes to separate the different classes. The hyperplanes maximise the separation between classes and minimise the intra-class variance. LDA, which is known to extract the best features and reduce the dimensionality of the data [49], projects the input data in a lower-dimensional space. The equations below define the within-class S_W and between-class S_B scattering comparison.

$$S_B = \sum_{i=1}^{C} J_i (n_i - n)(n_i - n)^T \qquad (21)$$

$$S_W = \sum_{i=1}^{C} \sum_{m_k \in C_i} J_i (n_k - n_i)(n_k - n_i)^T, \qquad (22)$$

where J_i is the number of vectors in the ith class C_i, c is the number of classes (number of activities), n_i is the mean of class c_i, and m_k is the vector of a specific class.
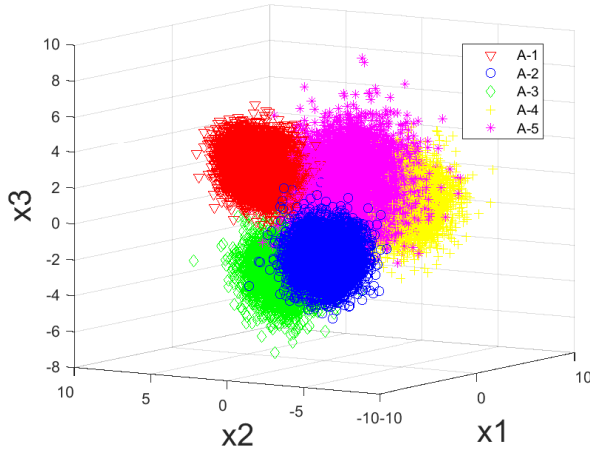
FIGURE 7: 3D scatter plot of the features after applying LDA.

The optimal discrimination matrix is selected by maximising the ratio of the determinant of the between- and within-class scatter matrices as

$$D_{opt} = argmax \frac{|D^T S_B D|}{|D^T S_W D|}, \qquad (23)$$

where D_{opt} is the set of discriminant vectors of S_W and S_B corresponding to the $(c-1)$ largest generalised eigenvalues λ, and can be obtained by solving (24):

$$S_B d_i = \lambda_i S_W d_i, i = 1, 2, ..., (c-1), \qquad (24)$$

where the rank of S_B is $(c-1)$ or less; hence, the upper bound value of t is $(c-1)$. The PCA features of the different activities are projected onto the LDA features as

$$L = P D_{opt}^T. \qquad (25)$$

Thus, the EDA features (i.e. L) are obtained to apply machine learning algorithms for training and testing of activities.

CLASSIFICATION

We applied different classification algorithms to the dataset for comparative and performance analyses. SVM [50], introduced by Vapnik, uses support vectors. It has been widely used in HAR systems owing to its high classification performance [51], [52]. It creates hyperplanes to maximise the margins between classes. By minimising the cost function, the optimal solution can be obtained, namely, the solution that maximises the distance between the hyperplane and the nearest training point. Herein, a nonlinear multiclass SVM with a *sigmoid* kernel was used. Sigmoid is used as it's a popular kernel. However, we also implemented RBF and Gaussian kernel, but it didn't improve the results remarkably.

Adaptive boosting, known as AdaBoost [53], is used primarily for ensemble learning or meta-learning. It applies an iterative approach to learn from the mistakes of classifiers to

improve their performance. AdaBoost has been widely used in HAR by researchers [54]. MLP [55] is also known as a feed-forward ANN. It consists of multiple hidden layer in addition to the input and output layers. With the help of error backpropagation, it can be trained to classify data that are not linearly separable. MLP has been used in ambient assisted living to recognise poses and to monitor dangerous situations [56]. QDA is closely related to LDA, but it does not assume that the covariance of each class is identical [57].

KNN is the simplest classification technique used for machine learning. The KNN algorithm determines the points from the training data that are close enough to be considered when selecting the class to predict a new observation [58]. The RF [59] method is used for both classification and regression problems. It generates multiple DTs based on the random selection of variables and data, and recognises dependent variables based on the DTs. RF has been widely used to recognise different human activities [60], [61]. In this study, 10 DTs were used to explore the classes.

A decision support tool that utilises a model of decisions or tree-like graphs and their possible consequences, including the utility and probability of event outcomes, is called a decision tree. A decision tree is a well-known classifier used in machine learning. Its structure is similar to a flowchart in which each internal node represents a test of an attribute, such as the probability a coin flip producing heads or tails. Each branch corresponds to a possible test outcome, and each leaf node corresponds to the class label. The decision is taken after applying all the features. The classification rules are based on the paths from the root to the leaf [62].

Recurrent neural networks

The events are represented based on time-sequential data from the sensor. A machine learning model able to encode time-sequential data is suitable for our purpose. Hence, RNNs were used in this study. They are one of the most widely applied deep learning methods for modelling events underlying time-sequential data. An RNN typically consists of recurrent relations within the model's hidden units, which connect its history (i.e. memory) to the present. RNNs often face vanishing gradient problems that cause challenges in processing long-term information. This phenomenon is known as long-term dependency. LSTM, proposed by Hochreiter and Schmidhuber [63], can solve the vanishing gradient problem typical of RNNs. RNNs and LSTMs have performed well in various fields, such as handwriting and speech recognition [64]. Figure 8 shows a sample deep RNN consisting of N LSTM units. Each LSTM block consists of an input gate I, a forget gate F, and an output gate U. The input gate I is expressed as

$$I_t = s\left(Y_{LI} L_t + Y_{HI} H_{t-1} + a_I\right). \qquad (26)$$

The input gate Y represents the weight matrix, bias a, and sigmoid function s. The forget gate F is expressed as follows:
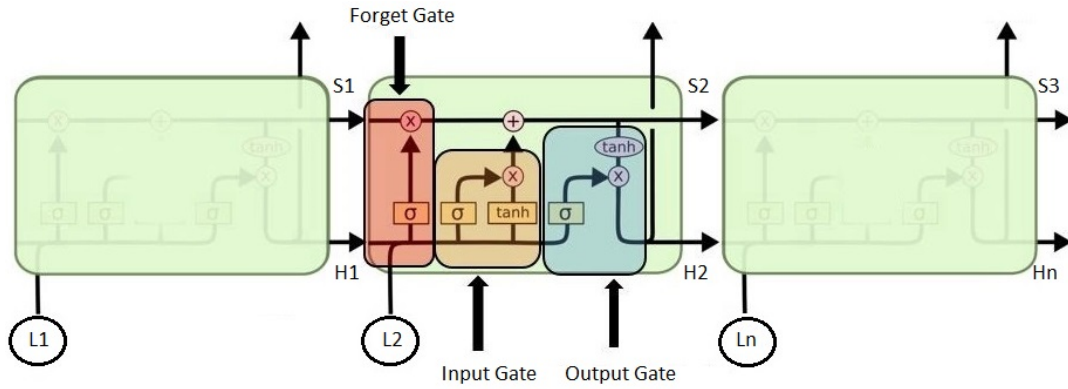
IEEE *Access*

FIGURE 8: Basic architecture of an RNN.

$$F_t = s\left(Y_{LF}L_t + Y_{HF}H_{t-1} + a_F\right). \quad (27)$$

The long-term memory is stored by a cell in a state vector B, which can be represented as

$$B_t = FB_{t-1} + I\tanh\left(Y_{LB}L_t + Y_{HB}H_{t-1} + a_B\right). \quad (28)$$

The output gate O represents the output as

$$U_t = s\left(Y_{LU}L_t + Y_{HU}H_{t-1} + a_U\right). \quad (29)$$

The hidden state H is represented as

$$H_t = U_t\tanh\left(B_t\right). \quad (30)$$

The final output N can be determined by

$$N_i = \text{softmax}\left(Y_U H_i + a_U\right), \quad (31)$$

where i represents the LSTM number and H hidden states.

We have used four hidden layers stacked LSTM. The first three layers have 100 memory units (or smart neurons) followed by 50 memory units in the next layer. Finally, because this is a classification problem, we use a dense output layer with a softmax activation function to make predictions for five classes. Our model has 847,055 trainable parameters.

RESULTS AND DISCUSSION

In this section, we describe the experiments performed on the XeThru UWB sensor dataset to recognise various human activities. The dataset consisted of data from 13 participants, with a total of 65,000 samples for the five activities. Ten radar frames per second were used, with 1,535 data points in each frame. Therefore, the data size for each sample was $10 \times 1,535$. In total, we used nine different classification approaches, with 80% of the total mixed up dataset used for training all models and 20% used for testing. Moreover, 10% of the training dataset was used for validation.

For performance measures, the accuracy, precision, recall, and F1-scores were evaluated for each classifier. The

accuracy of each classifier is listed in Table 2. All classifiers were tested using a built-in Python library, scikit-learn. Two conventional classifiers, SVM and AdaBoost, performed poorly in classifying activities, with accuracies of less than 50% as shown in Figure 10a and 10b, respectively. Neither classifier could distinguish between the standing and walking activities. MLP performed only slightly better than these classifiers, with an accuracy of 51.3%. This classifier could detect the lying posture well, as shown in Figure 10c.

The QDA and KNN classifiers performed moderately well and classified all activities with an accuracy of 75%. Figure 10d and 10e shows the confusion matrices of QDA and KNN. QDA primarily misclassified the activities when the subjects were sitting on the bed with their legs on the bed or floor. RF and DT were the only two conventional approaches with accuracies of approximately 90%. The confusion matrices of these two classifiers are shown in Figure 10f and 10g, respectively.

Finally, we tested the dataset using a state-of-the-art LSTM, built using Keras [65], with Tensorflow as the back end. It outperformed all other classifiers by achieving an excellent accuracy of 98%. Figure 10h shows the confusion matrices, while graphs of the model's accuracy and the model's loss are shown in Figure 11.

As we had 1,535 data points in each frame, EDA was introduced to reduce the dimensionality. In the EDA, a PCA was introduced to reduce the dimensions. The performance was insufficient when using PCA alone, as shown in the scatter plot in Figure 6; thus, LDA was introduced. The combination of PCA and LDA, which we refer to as enhanced discriminant analysis (EDA), outperformed all other approaches. The 3D plot of the features extracted after application of LDA is shown in the scatter plot in Figure 7. The EDA with LSTM training model performed better than the LSTM model alone and demonstrated excellent performance in all activities, as shown in Figure 10i. The trained model and loss graphs are shown in Figure 12. The overall accuracies of all nine classifiers that were implemented are listed in Table 2. The precision, recall, and F1-scores of the top four classifiers alone are presented in Table 1. Furthermore, to validate our
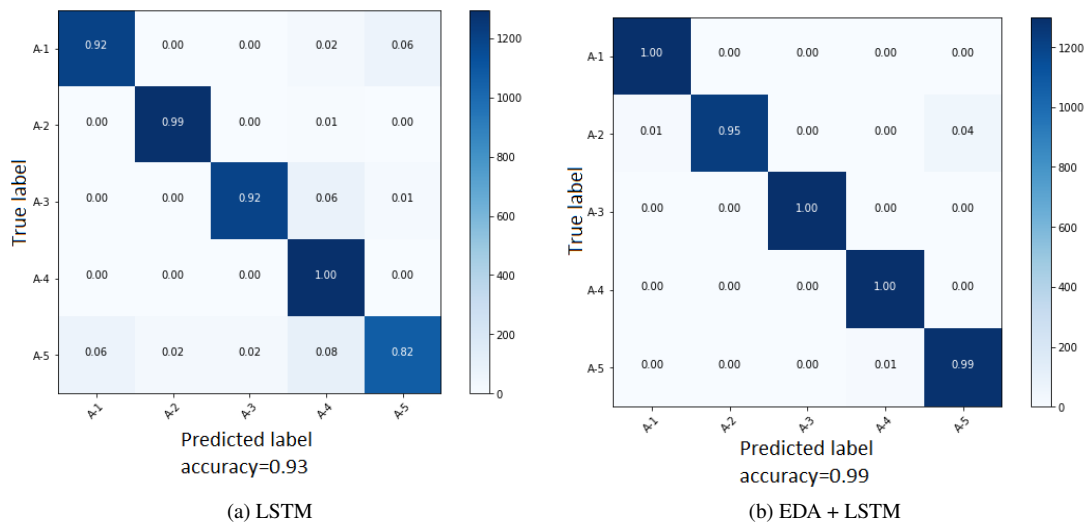
(a) LSTM

(b) EDA + LSTM

FIGURE 9: Confusion matrix of 5-fold cross-validation.

Approach	Activity	Precision	Recall	F1-score
Random Forest	Lying	0.95	0.97	0.96
	Sitting on the bed while the legs are also on the bed	0.88	0.91	0.89
	Sitting on the bed while the legs are on the floor	0.85	0.86	0.86
	Standing	0.81	0.81	0.81
	Walking	0.86	0.81	0.83
Decision Trees	Lying	0.98	0.99	0.98
	Sitting on the bed while the legs are also on the bed	0.94	0.95	0.94
	Sitting on the bed while the legs are on the floor	0.89	0.90	0.90
	Standing	0.89	0.86	0.88
	Walking	0.84	0.85	0.85
LSTM	Lying	0.97	0.90	0.93
	Sitting on the bed while the legs are also on the bed	0.99	0.99	0.99
	Sitting on the bed while the legs are on the floor	0.89	1.00	0.94
	Standing	0.97	0.99	0.98
	Walking	0.99	0.94	0.96
Enhanced Discriminant Analysis (EDA) + LSTM	Lying	0.92	0.99	0.95
	Sitting on the bed while the legs are also on the bed	0.99	0.99	0.99
	Sitting on the bed while the legs are on the floor	1.00	0.99	0.99
	Standing	1.00	0.99	0.99
	Walking	0.99	0.93	0.96

TABLE 1: Classification performances of top four classifiers.

Classifier	Accuracy (%)
SVM	34.7
AdaBoost	47.3
Multilayer Perceptron	51.3
Quadratic Discriminant Analysis	74.4
KNN	76.5
Random Forest	87.1
Decision Trees	91.0
LSTM	98.0
EDA + LSTM	**99.6**

TABLE 2: Accuracies of all classifiers.

approach, 5-fold cross-validation was performed on LSTM and EDA with LSTM for HAR. The confusion matrices of both approaches are shown in Figure 9.

Afterwards, we continued our experiments using leave-one-subject-out validation. Initially, we took 5% data out from the testing subjects, mixed it with the training data, and obtained an average accuracy of 86%. Figure 13 shows the

average confusion matrix of our results. In the next experiments, we applied no data taken from the testing subjects, i.e., leave one whole subject out and obtained the average accuracy of 66%, as shown in Figure 14. Given the lower performance, we are planning to work on improving the subject generic validation results in the future by improving the model. However, during the experiments with similar data selections, we experienced that none of the traditional approaches yielded accuracies more than 60%.

Singh et al. [28] introduced exercise-based activities by using mmWave radar. Bouchard et al. [66] introduced 15 daily life activities based on ten males. No female participated in their study. Moreover, the age range was between 22 to 39. In our case, we have quite a diverse age range, and almost half of the participants were females to avoid bias in the study. Sharma et al. [29] proposed a channel impulse response based activity recognition system, but their activities were limited, i.e. standing, sitting and lying. While
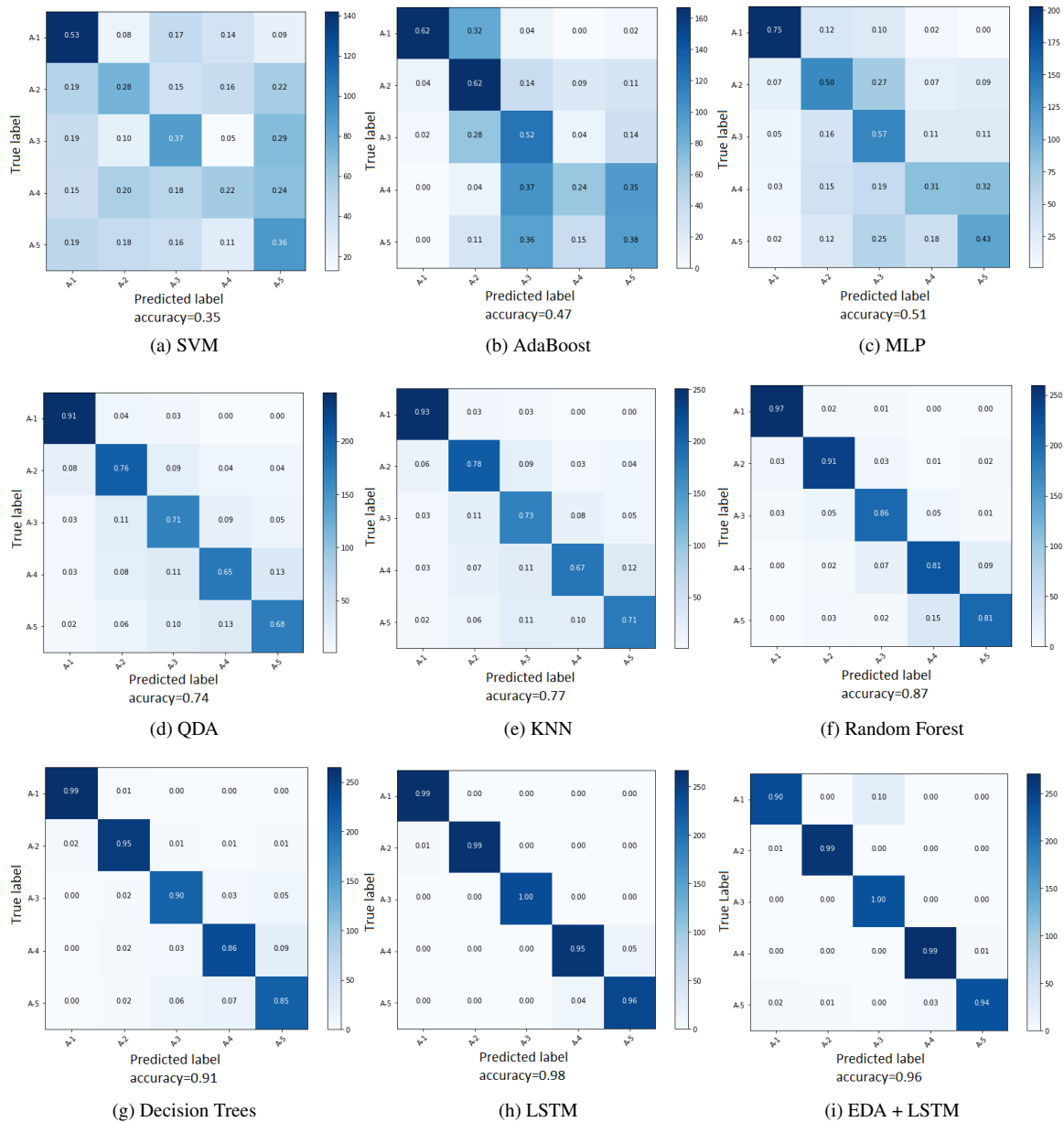
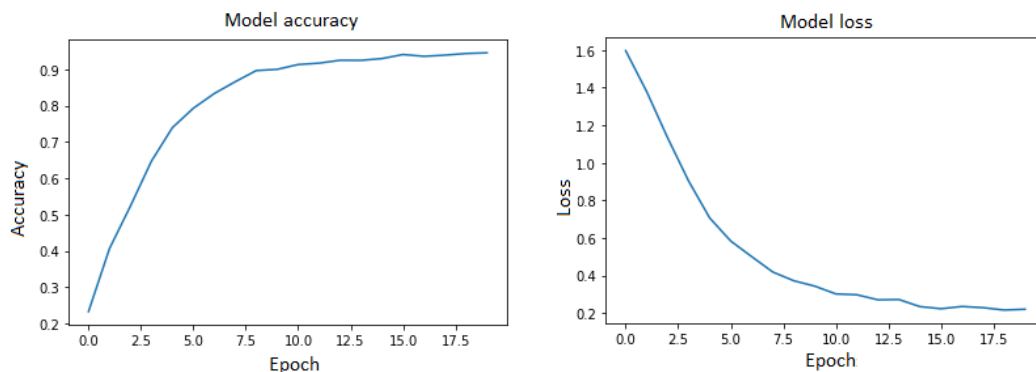FIGURE 10: Confusion Matrices of all classifiers.

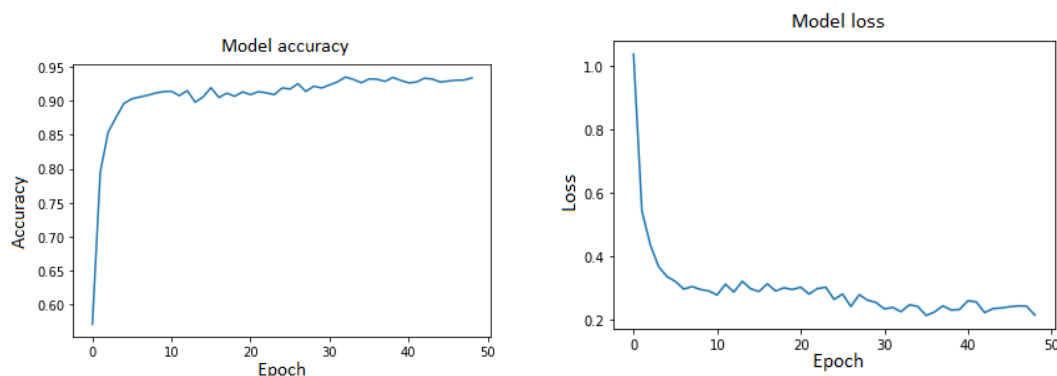FIGURE 11: Accuracy and loss graphs of LSTM.

FIGURE 12: Accuracy and loss graphs of proposed approach, EDA + LSTM.
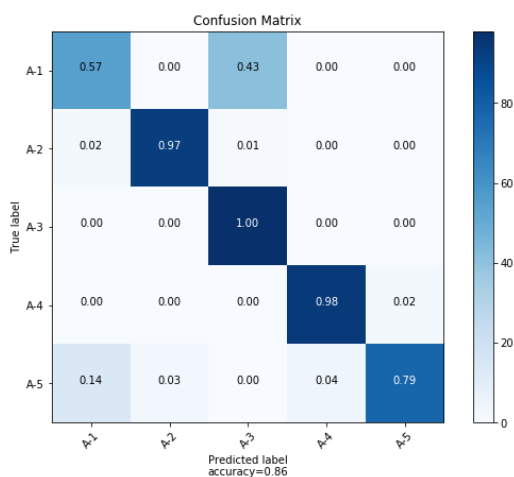

FIGURE 13: Confusion matrix of partial-leave-one-person-out validation.
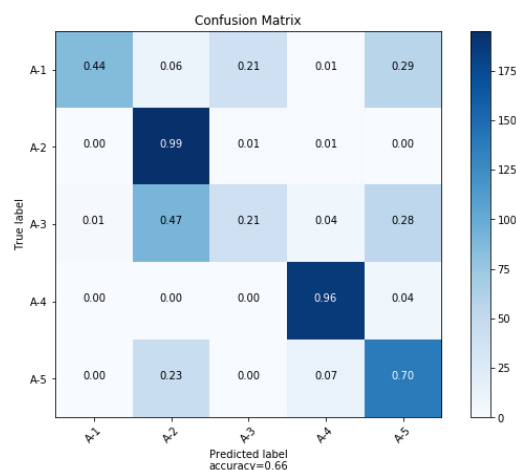

FIGURE 14: Confusion Matrix of leave-one-person-out validation.

our activities also cater to a different sitting posture, i.e. either legs are on the bed or the floor. Furthermore, we also introduced an EDA based feature extraction approach. We also tested our approach on a Ahmed et al. [67] dataset (as they used a similar UWB sensor), which showed promising results and outperformed their results. Figure 15 shows the confusion matrix and a–l show the twelve different dynamic hand gestures as proposed in [67]. To the best of the author's knowledge, there is no current study based on UWB sensors that focused on enhanced discriminant features analysis before feeding the data into the machine learning algorithms.

CONCLUSION

In this study, a novel approach was proposed for HAR using a UWB sensor and state-of-the-art deep learning models. The proposed approach is beneficial for older adults because it is difficult for them to wear actigraphy devices 24/7 or to be monitored through an RGB camera, which could breach their privacy. In this work, a UWB sensor was mounted on a robot at a certain distance. Because the UWB sensor has several features, EDA was used to reduce the dimensions

of the features before feeding them into the deep learning model. The results were compared with those of conventional approaches. The proposed approach was found to perform significantly better, with an accuracy of 99.6%. Moreover, 5-fold cross-validation was performed for generalisation of the system. Furthermore, we implemented our approach on publically available dataset and got better results.

In the future, we intend to perform more complex experiments in real-time environments. Furthermore, we plan on extending the algorithms by introducing the heart rate into the monitoring system to detect emergencies. Moreover, we also need to acquire data from older people since the approach focuses on eldercare. Finally, we can also introduce multiple UWB sensors in the apartment or other thermal/depth-based sensors to localize the exact position of persons. In that context, different sensor fusion strategies might be explored.
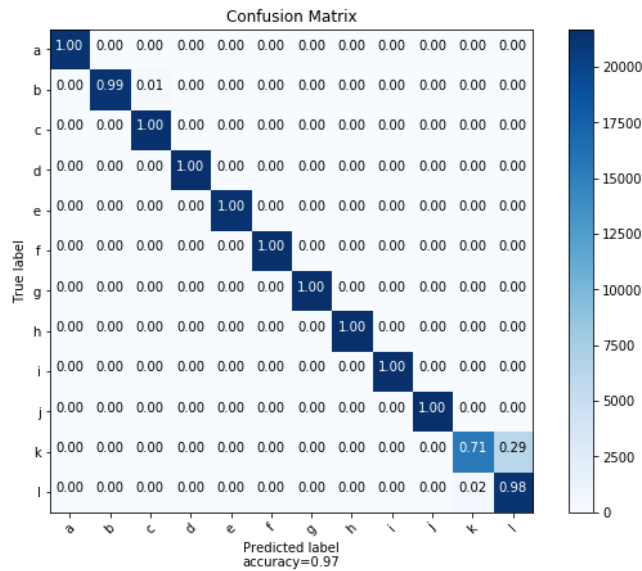
ACKNOWLEDGMENT

FIGURE 15: Our approach classifying twelve dynamic hand gestures based on [67] dataset.

under Grant Agreement No. 288285, and through its Centres of Excellence scheme, Project No. 262762. In Figure 4, the first author shows the experimental environment for better understanding of the setups. He was not a subject in the experiments though, to avoid the biasness of the study.

REFERENCES

[1] Department of Economic and Social Affairs, Population Division (2017). World Population Ageing. United Nations, 2017.
[2] A. Sharkey and N. Sharkey, "Granny and the robots: ethical issues in robot care for the elderly," Ethics and information technology, vol. 14, no. 1, pp. 27–40, 2012.
[3] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," Sensors, vol. 15, no. 12, pp. 31 314–31 338, 2015.
[4] S. K. Tasoulis, C. N. Doukas, V. P. Plagianakos, and I. Maglogiannis, "Statistical data mining of streaming motion data for activity and fall recognition in assistive environments," Neurocomputing, vol. 107, pp. 87–96, 2013.
[5] E. Gambi, G. Temperini, R. Galassi, L. Senigagliesi, and A. De Santis, "ADL recognition through machine learning algorithms on IoT air quality sensor dataset," IEEE Sensors Journal, vol. 20, no. 22, pp. 13 562–13 570, 2020.
[6] N. Jaouedi, F. J. Perales, J. M. Buades, N. Boujnah, and M. S. Bouhlel, "Prediction of human activities based on a new structure of skeleton features and deep learning model," Sensors, vol. 20, no. 17, p. 4944, 2020.
[7] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, "Continuous human activity classification from FMCW radar with Bi-LSTM networks," IEEE Sensors Journal, vol. 20, no. 22, pp. 13 607–13 619, 2020.
[8] N. Andersen, K. Granhaug, J. A. Michaelsen, S. Bagga, H. A. Hjortland, M. R. Knutsen, T. S. Lande, and D. T. Wisland, "A 118-mw pulse-based radar SoC in 55-nm CMOS for non-contact human vital signs detection," IEEE Journal of Solid-State Circuits, vol. 52, no. 12, pp. 3421–3433, 2017.
[9] D. T. Wisland, K. Granhaug, J. R. Pleym, N. Andersen, S. Støa, and H. A. Hjortland, "Remote monitoring of vital signs using a CMOS UWB radar transceiver," in 2016 14th IEEE International New Circuits and Systems Conference (NEWCAS). IEEE, 2016, pp. 1–4.
[10] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," Computer, vol. 46, no. 7, pp. 62–69, 2012.
[11] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The gator tech smart house: A programmable pervasive space," Computer, vol. 38, no. 3, pp. 50–60, 2005.

[12] Q. Zhang, Y. Su, and P. Yu, "Assisting an elderly with early dementia using wireless sensors data in smarter safer home," in International Conference on Informatics and Semiotics in Organisations. Springer, 2014, pp. 398–404.
[13] Q. Ni, A. B. Garcia Hernando, D. la Cruz, and I. Pau, "The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," Sensors, vol. 15, no. 5, pp. 11 312–11 362, 2015.
[14] A. S. Billis, E. I. Papageorgiou, C. Frantzidis, E. I. Konstantinidis, and P. D. Bamidis, "Towards a hierarchically-structured decision support tool for improving seniors' independent living: the USEFIL decision support system," in Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments, 2013, pp. 1–4.
[15] J. Mišeikis, P. Caroni, P. Duchamp, A. Gasser, R. Marko, N. Mišeikienė, F. Zwilling, C. de Castelbajac, L. Eicher, M. Früh et al., "Lio-a personal robot assistant for human-robot interaction and care applications," IEEE Robotics and Automation Letters, vol. 5, no. 4, pp. 5339–5346, 2020.
[16] C. Taramasco, Y. Lazo, T. Rodenas, P. Fuentes, F. Martínez, and J. Demongeot, "System design for emergency alert triggered by falls using convolutional neural networks," Journal of Medical Systems, vol. 44, no. 2, p. 50, 2020.
[17] M. A. Khan, M. Sharif, T. Akram, M. Raza, T. Saba, and A. Rehman, "Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition," Applied Soft Computing, vol. 87, p. 105986, 2020.
[18] D. W. Aha, Lazy learning. Springer Science & Business Media, 2013.
[19] C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data mining and knowledge discovery, vol. 2, no. 2, pp. 121–167, 1998.
[20] G. P. Zhang, "Neural networks for classification: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 4, pp. 451–462, 2000.
[21] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.
[22] V. Jacquot, Z. Ying, and G. Kreiman, "Can deep learning recognize subtle human activities?" in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14 244–14 253.
[23] J. Tao and X. Fang, "Toward multi-label sentiment analysis: a transfer learning based approach," Journal of Big Data, vol. 7, no. 1, pp. 1–26, 2020.
[24] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multi-fusion representations and CNNs," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 16, no. 2, pp. 1–19, 2020.
[25] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition." in Ijcai, vol. 15. Buenos Aires, Argentina, 2015, pp. 3995–4001.
[26] C. Ding, L. Zhang, C. Gu, L. Bai, Z. Liao, H. Hong, Y. Li, and X. Zhu, "Non-contact human motion recognition based on uwb radar," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 8, no. 2, pp. 306–315, 2018.
[27] L. Cheng, A. Zhao, K. Wang, H. Li, Y. Wang, and R. Chang, "Activity recognition and localization based on uwb indoor positioning system and machine learning," in 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2020, pp. 0528–0533.
[28] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems, 2019, pp. 51–56.
[29] S. Sharma, H. Mohammadmoradi, M. Heydariaan, and O. Gnawali, "Device-free activity recognition using ultra-wideband radios," in 2019 International Conference on Computing, Networking and Communications (ICNC). IEEE, 2019, pp. 1029–1033.
[30] M. Z. Uddin, M. M. Hassan, A. Alsanad, and C. Savaglio, "A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare," Information Fusion, vol. 55, pp. 105–115, 2020.
[31] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," Remote Sensing, vol. 11, no. 9, p. 1068, 2019.
[32] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," Nature communications, vol. 11, no. 1, pp. 1–11, 2020.

[33] Y. Wang, Q. Liu, and A. E. Fathy, "Simultaneous localization and respiration detection of multiple people using low cost UWB biometric pulse doppler radar sensor," in 2012 IEEE/MTT-S International Microwave Symposium Digest. IEEE, 2012, pp. 1–3.

[34] ——, "CW and pulse–doppler radar processing based on FPGA for human sensing applications," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 5, pp. 3097–3107, 2012.

[35] X. Liang, J. Deng, H. Zhang, and T. A. Gulliver, "Ultra-wideband impulse radar through-wall detection of vital signs," Scientific reports, vol. 8, no. 1, pp. 1–21, 2018.

[36] M. Mercuri, P. J. Soh, L. Boccia, D. Schreurs, G. A. Vandenbosch, P. Leroux, and G. Amendola, "Optimized SFCW radar sensor aiming at fall detection in a real room environment," in 2013 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems. IEEE, 2013, pp. 4–6.

[37] G. Wang, C. Gu, T. Inoue, and C. Li, "A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring," IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 11, pp. 2812–2822, 2014.

[38] W. Hu, Z. Zhao, Y. Wang, H. Zhang, and F. Lin, "Noncontact accurate measurement of cardiopulmonary activity using a compact quadrature doppler radar sensor," IEEE Transactions on Biomedical Engineering, vol. 61, no. 3, pp. 725–735, 2013.

[39] M. Mercuri, P. J. Soh, G. Pandey, P. Karsmakers, G. A. Vandenbosch, P. Leroux, and D. Schreurs, "Analysis of an indoor biomedical radar-based system for health monitoring," IEEE Transactions on Microwave Theory and Techniques, vol. 61, no. 5, pp. 2061–2068, 2013.

[40] A. Muqaibel and A. Safaai-Jazi, "Characterization of wall dispersive and attenuative effects on UWB radar signals," Journal of the Franklin Institute, vol. 345, no. 6, pp. 640–658, 2008.

[41] D. K. Barton, Radar system analysis and modeling. Artech House, 2004.

[42] G. Kolumbán, T. Krébesz, and F. C. Lau, "Feasibility of UWB radio: Impulse radio versus chaos-based approach," in Proceedings of 2010 IEEE International Symposium on Circuits and Systems. IEEE, 2010, pp. 2450–2453.

[43] N. Levanon, "Radar principles," wi, 1988.

[44] A. Sarkar and D. Ghosh, "Through-wall heartbeat frequency detection using ultra-wideband impulse radar," in 2019 International Conference on Range Technology (ICORT). IEEE, 2019, pp. 1–5.

[45] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Deep learning based nonlinear principal component analysis for industrial process fault detection," in 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 1237–1243.

[46] M. Z. Uddin, J. Lee, and T.-S. Kim, "Independent shape component-based human activity recognition via hidden markov model," Applied Intelligence, vol. 33, no. 2, pp. 193–206, 2010.

[47] ——, "An enhanced independent component-based human facial expression recognition from video," IEEE Transactions on Consumer Electronics, vol. 55, no. 4, pp. 2216–2224, 2009.

[48] A. Jalal, M. Z. Uddin, and T.-S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," IEEE Transactions on Consumer Electronics, vol. 58, no. 3, pp. 863–871, 2012.

[49] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Transactions on pattern analysis and machine intelligence, vol. 19, no. 7, pp. 711–720, 1997.

[50] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[51] D. A. Adama, A. Lotfi, C. Langensiepen, K. Lee, and P. Trindade, "Human activity learning for assistive robotics using a classifier ensemble," Soft Computing, vol. 22, no. 21, pp. 7027–7039, 2018.

[52] B. M. Abidine, L. Fergani, B. Fergani, and M. Oussalah, "The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition," Pattern Analysis and Applications, vol. 21, no. 1, pp. 119–138, 2018.

[53] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of computer and system sciences, vol. 55, no. 1, pp. 119–139, 1997.

[54] A. Subasi, D. H. Dammas, R. D. Alghamdi, R. A. Makawi, E. A. Albiety, T. Brahimi, and A. Sarirete, "Sensor based human activity recognition using adaboost ensemble classifier," Procedia Computer Science, vol. 140, pp. 104–111, 2018.

[55] T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, 2009.

[56] B. M. V. Guerra, S. Ramat, G. Beltrami, and M. Schmid, "Automatic pose recognition for monitoring dangerous situations in ambient-assisted living," Frontiers in Bioengineering and Biotechnology, vol. 8, p. 415, 2020.

[57] A. Ghosh, R. SahaRay, S. Chakrabarty, and S. Bhadra, "Robust generalised quadratic discriminant analysis," arXiv preprint arXiv:2004.06568, 2020.

[58] U. Khandelwal, A. Fan, D. Jurafsky, L. Zettlemoyer, and M. Lewis, "Nearest neighbor machine translation," arXiv preprint arXiv:2010.00710, 2020.

[59] T. K. Ho, "Random decision forests," in Proceedings of 3rd international conference on document analysis and recognition, vol. 1. IEEE, 1995, pp. 278–282.

[60] F. M. Noori, B. Wallace, M. Z. Uddin, and J. Torresen, "A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network," in Scandinavian Conference on Image Analysis. Springer, 2019, pp. 299–310.

[61] D. Ren, N. Aubert-Kato, E. Anzai, Y. Ohta, and J. Tripette, "Random forest algorithms for recognizing daily life activities using plantar pressure information: a smart-shoe study," PeerJ, vol. 8, p. e10170, 2020.

[62] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," Pattern Recognition, vol. 43, no. 10, pp. 3605–3620, 2010.

[63] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[64] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in International conference on machine learning, 2014, pp. 1764–1772.

[65] F. Chollet et al., "keras," 2015.

[66] K. Bouchard, J. Maitre, C. Bertuglia, and S. Gaboury, "Activity recognition in smart homes using uwb radars," Procedia Computer Science, vol. 170, pp. 10–17, 2020.

[67] S. Ahmed, D. Wang, J. Park, and S. H. Cho, "Uwb-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors," Scientific Data, vol. 8, no. 1, pp. 1–9, 2021.

FARZAN M. NOORI received his B.E. and M.S. degrees in Mechatronics Engineering from Air University. He was a Lab Engineer in Department of Mechatronics Engineering from 2014 to 2017. In 2017, He Joined the University of Coimbra as a Research Assistant.

Since 2018, he is working as a PhD fellow under the Robotics and Intelligent Systems (ROBIN) research group at Dept. of Informatics, University of Oslo, Norway. His research interests include computer vision, image processing, sensor fusion, healthcare and brain-computer interface. Contact him at farzanmn@ifi.uio.no and find more information at https://sites.google.com/view/fmnoori/

MD ZIA UDDIN received his Bachelor degree in Computer Science and Engineering in August of 2004 and right after that he joined as a Lecturer in Computer Science and Engineering department of International Islamic University Chittagong of Bangladesh where he worked till February of 2007. Then, he started his Ph.D. degree in Biomedical Engineering in March of 2007 at Kyung Hee University of South Korea and obtained his Ph.D. in February 2011. After his Ph.D., he joined the Electronics Engineering Department of Inha University of South Korea as an Assistant Professor in March of 2011 and worked there till August of 2013. Right after that, he joined Computer Education department of Sungkyunkwan University (SKKU) of South Korea as an Assistant Professor and worked there till December, 2016. From January of 2017 to September 2019, he worked as a post-doctoral research fellow under Robotics and Intelligent Systems (ROBIN) research group at Dept. of Informatics, University of Oslo, Norway. Since October 2019 to till now, he is working as a research scientist in Software and Service Innovation Group, SINTEF Digital, Oslo, Norway.

His research fields are mainly focused on data & feature analysis from various sources (sensors and others), physical/mental healthcare, human machine/computer/robot interaction, pattern recognition, deep learning, artificial intelligence etc. His research outcomes have been published in famous journals such as Information Fusion, IEEE transactions on consumer electronics, Future Generation Computer Systems, etc.

JIM TORRESEN received his M.Sc. and Dr.ing. (Ph.D) degrees in computer architecture and design from the Norwegian University of Science and Technology, University of Trondheim in 1991 and 1996, respectively. He has been employed as a senior hardware designer at NERA Telecommunications (1996-1998) and at Navia Aviation (1998-1999). Since 1999, he has been a professor at the Department of Informatics at the University of Oslo (associate professor 1999-2005). He is also a principal investigator in the RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion.

He has been a visiting researcher at Kyoto University, Japan for one year (1993-1994), four months at Electrotechnical laboratory, Tsukuba, Japan (1997 and 2000) and one year at Cornell University, US (2010-2011). His research interests include artificial intelligence, ethical aspects of AI and robotics, machine learning, robotics, and applying this to complex real-world applications which have resulted in more than 200 scientific peer-reviewed papers in international journals, books and conference proceedings. He is a member of the Norwegian Academy of Technological Sciences (NTVA) and the National Committee for Research Ethics in Science and Technology (NENT). He is a senior member of IEEE. Contact him at jimtoer@ifi.uio.no and find more information at http://jimtoer.no/

. . .