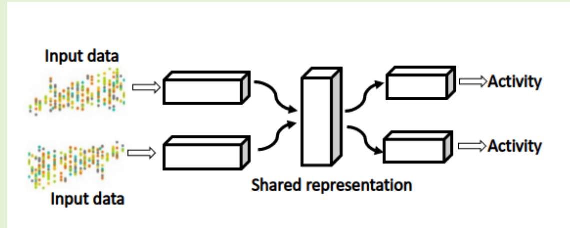
# Cross-Domain Activity Recognition Using Shared Representation in Sensor Data

Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei

***Abstract*— Existing models based on sensor data for human activity recognition are reporting state-of-the-art performances. Most of these models are conducted based on single-domain learning in which for each domain a model is required to be trained. However, the generation of adequate labelled data and a learning model for each domain separately is often time-consuming and computationally expensive. Moreover, the deployment of multiple domain-wise models is not scalable as it obscures domain distinctions, introduces extra computational costs, and limits the usefulness of training data. To mitigate this, we propose a multi-domain learning network to transfer knowledge across different but related domains and alleviate isolated learning paradigms using a shared representation. The proposed network consists of two identical causal convolutional sub-networks that are projected to a shared representation followed by a linear attention mechanism. The proposed network can be trained using the full training dataset of the source domain and a dataset of restricted size of the target training domain to reduce the need of large labelled training datasets. The network processes the source and target domains jointly to learn powerful and mutually complementary features to boost the performance in both domains. The proposed multi-domain learning network on six real-world sensor activity datasets outperforms the existing methods by applying only 50% of the labelled data. This confirms the efficacy of the proposed approach as a generic model to learn human activities from different but related domains in a joint effort, to reduce the number of required models and thus improve system efficiency.**




***Index Terms*— Activity Recognition, Cross-domain learning, Deep Learning, Temporal Evaluation, Sensor Data**

## I. INTRODUCTION

Human activity recognition (HAR) is the problem of correctly predicting human activities based on a series of sensor readings that capture environment interaction and human movements. HAR systems are widely used to stimulate different applications including healthcare monitoring, eldercare, smart education [1] and robot vision [2]. HAR is often a supervised learning task that relies on labelled data to train a model. Acquiring enough amount of labelled data for training is difficult and expensive, if not impossible, and usually requires significant manual efforts from domain experts [3]. Moreover, generally supervised learning methods work based on the assumptions that the data of the training and testing phases of a model must lie in the identical feature space, have the same label space and have the same underlying distribution. Further, the performance of models often decreases when the training and testing data have different feature spaces and underlying distributions [3]. Hence, the above assumption lead to the major challenges of machine learning methods. In HAR, the same set of binary sensors are used to generate training and testing data in order to have the same feature space. Moreover, the participants should have similar preferences or habits in both training and testing data to render the same underlying distribution. Finally, the set of activities in the training and testing data are the same to make the same label space.

Several transfer learning methods have been suggested to relax the assumptions of supervised learning with maintaining reasonable accuracy. Transfer learning is about transferring knowledge across different but related domains to reduce the need for labelled data, reduce the training time and improve the accuracy of the target task [3], [4]. Transfer learning refers to various strategies [4] which include multi-domain learning [5], [6], self-taught learning, [7], [8], domain adaptation [9], multi-task learning [10], sharing of knowledge and representations [11]. In this paper, transfer learning is used to make HAR systems more robust, adaptable and effectively reuse the knowledge by learning a model across multiple domains. Developed transfer learning models for HAR systems have improved the recognition score, but most of the transfer learning methods have certain limitations and this will be deferred to Section 2 for further discussion. Multi-domain learning (MDL) as a transfer learning technique refers to sharing information about the same task (e.g., activity recognition) across different contextual domains (datasets) [6]. MDL uses a shared representation to learn a task from different domains

Manuscript received XX XX, XXXX; accepted XX XX, XXXX. Date of publication XX XX, XXXX; date of current version XXXX XX, X, (Corresponding author: Rebeen Ali Hamad.)

Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo are with the Department of Computer and Information Sciences, Northumbria University Newcastle upon Tyne NE1 8ST, UK; (e-mail: rebeen.hamad, longzhi.yang, wailok.woo@northumbria.ac.uk)

Bo Wei is with the School of Computing and Communications, Lancaster University, Lancaster, LA1 4WA; (e-mail: bo.wei@lancaster.ac.uk)

(datasets) in parallel, thereby learned features in a domain can improve learning in other domains [5], [12]–[14]. An essential transfer learning problem is how to reduce the divergence between the source and target domains with keeping the data properties [12].

This paper contributes to mitigating the differences between the source and target domains using a shared representation to transfer knowledge across different but related human activity domains. We propose a multi-domain learning approach to jointly learn activity recognition based on different but related domains rather than learning each domain in isolation. The proposed method is a deep neural network that consists of two identical sub-networks. The sub-networks have the same configuration in terms of the number of layers, hyper-parameters and weights. Parameter updating is mirrored using a shared layer across both sub-networks. The proposed network consists of two 1D causal convolutional neural networks (CNN) layers for each of the sub-networks and is followed by a shared fully connected layer. The shared layer projects learned features to a common latent space to transfer knowledge between the source and target domains. The proposed network accumulates sight of correlation amongst different but related human activity domains where common knowledge can be shared to enhance the performance of activity recognition across several related domains. A linear attention layer that increases focus on the important time steps is appended to the shared layer. The proposed method simultaneously processes two related domains: source and target domains. The benefit of our proposed method is firstly to reduce the need and effort for labelled data of the target domain. The proposed network uses the training data of the target domain with restricted size and the full training data of the source domain, yet provided better performance than using the full training data in a single domain setting. Secondly, the proposed multi-domain learning network reduces the training time by rendering a generic model for related domains compared to fitting a model for each domain separately. The effectiveness of transfer learning using the proposed cross-domain learning network is evaluated by the performance of the existing single-domain learning (SDL) models on the target datasets. The proposed MLD network outperforms the existing state-of-the-art SDL methods that are trained directly on the full target domain data. Moreover, the results demonstrate that the proposed MDL network can improve the performance of the source domain in addition to the performance of the target domain.

In summary, the contributions of this paper are listed below:

i. Proposing a cross-domain deep learning network to transfer knowledge across different but related domains to improve accuracy and reduce the need for labelled data.
ii. Proposing linear attention mechanism within the proposed network supports extracting fine-grained features of human activities.
iii. Causal convolution is used within the proposed network to stop information flow from future to past to preserve the ordering of the time steps.

The rest of the paper is structured as follows. The related works are reviewed in Section 2. The proposed network is described in Section 3. Section 4 presents experiments and evaluations for the proposed network. Finally, the conclusion of this paper is presented in Section 5.

## II. Related works

Conventional machine learning methods have been proposed for HAR, such as a k-nearest neighbour, hidden Markov models, decision trees (C4.5 algorithm), naive-Bayes and support vector Machine [15]. These algorithms have been effective in activity recognition. However, they heavily rely on traditional manual feature engineering which is limited by both domain knowledge experts and time constraints. Hence the power of the automated feature extraction approach, i.e. deep learning, has attracted increasing interest from various research areas such as cognitive assistance and context-aware systems [16], [17] and human activity recognition [18]–[20]. Different techniques based on deep learning have been proposed in activity recognition on various benchmark datasets. Most of the systems of activity recognition based on deep learning report state-of-the-art performances [21], [22]. A deep convolutional LSTM is proposed for multimodal wearable activity recognition [23]. A deep learning model is suggested for activity recognition based on CNN to systematically extract and learn features from raw input data [24]. Deep learning approaches in the form of ensemble learning are also used for activity recognition using wearable sensors [25]. Joint learning of temporal deep learning models is proposed to improve activity recognition and handle imbalanced class problems for human activity recognition [26]. Dilated causal convolution based on the self-attention technique is proposed to improve and accelerate human activity recognition [27]. However, these approaches for HAR operate based on a strong assumption that the data of training and testing have the same distribution [3]. Besides, several methods based on CNN are proposed for HAR from wearable sensor data. An attention-based CNN method is proposed for HAR to process weakly labelled human activities data [28]. A layer-wise CNN with a local loss for HAR systems is proposed [22]. Moreover, a selective CNN kernel mechanism is introduced for HAR from wearable sensor data [29]. Besides, a shallow CNN to consider cross-channel interaction within the human activities is proposed [30]. A triplet cross-dimension attention for HAR is proposed to capture the interaction between three dimensions, i.e., sensor, temporal and channel [31]. These methods have reported state-of-the-art performance, however, they have only validated by the wearable sensors data which are less imbalanced. Moreover, often the proposed models of HAR systems have focused on observations of only one domain, which needs a large amount of labelled data to train a model for HAR [5].

Self-attention [32] mechanism is used in most existing state-of-the-art models to focus more on the most relevant time steps and enhance the performance of HAR systems. The self-attention mechanism is appended to dilated causal convolution (DCC+MSA) to improve HAR [27] using data captured from wearable sensors and smart homes. The self-attention mechanism is also used for DeepConvLSTM to improve HAR using data from wearable sensors [33]. Moreover, HAR+Attention based on self-attention block is used to improve HAR from wearable sensor data [34]. Despite the effectiveness of the self-attention mechanism in improving HAR, this mechanism suffers from quadratic computational complexity which delays the learning process and occupies more memory. Furthermore, these methods only process a single domain. To remedy these limitations, we propose an MDL network with a linear attention mechanism to improve HAR and reduce the need for labelled data using a cross-domain learning representation layer.

In deep learning, a domain refers to a dataset where its examples are drawn from the same distribution [35]. Multiple datasets with different data distributions can be used to target the same problem by fusing multi-domain features. MDL aims to perform a task (e.g., activity recognition) across different but related domains simultaneously. MDL has become an interesting problem since the success of deep learning approaches based on large-scale training data. MDL as a transfer learning technique addresses the problem of learning a task from different but related domains that share some commonalities on their input-output mapping functions [6]. MDL learning optimizes multiple objectives jointly and simultaneously to improve the performance of multiple domains by exploiting commonalities and differences across domains using a shared representation. Moreover, multi-domain learning reduces the sampling burden required to improve generalization by learning multiple similar domains as opposed to learning each domain in isolation [6], [13]. Similar domains using a shared representation transfer knowledge to each other during the learning process to mitigate the negative effect of data scarcity [14], [36]. In contrast to an SDL that only uses domain-specific data to independently learn a model for each domain, MDL leverages data of all domains and shares knowledge among domains leading to better prediction performance and model generalization [37]. Figure 1 shows the architecture of SDL and MDL. There is limited research for activity recognition using MDL particularly based on sensor data. MDL is proposed between two domains of human activities using transfer learning from Web search [5]. These works yet firstly relied on classical statistical features engineering. Second, Web search is used to obtain relevant Web pages for the activities, and then techniques of information retrieval are employed to further process the extracted Web pages. MDL methods are proposed for HAR as reported in [38], [39], but the methods are only validated using wearable sensors data.

Self-attention mechanism [32], as the key component of the state-of-the-art transformer structure, has been broadly used to enhance the capability of temporal models to extract fine-grained features for HAR [27], [28], [33], [34], [40]. Self-attention mechanism has three learned matrix: queries $Q \in R^{N\times D_k}$, keys $K \in R^{M\times D_k}$, and values $V \in R^{M\times D_v}$, where $N$ and $M$ are the lengths of the queries and keys (or values), $D_k$ and $D_v$ are the dimensions of keys (or queries) and values. Self-attention mechanism as the scaled dot-product attention used by transformer model is show in Equation 1.

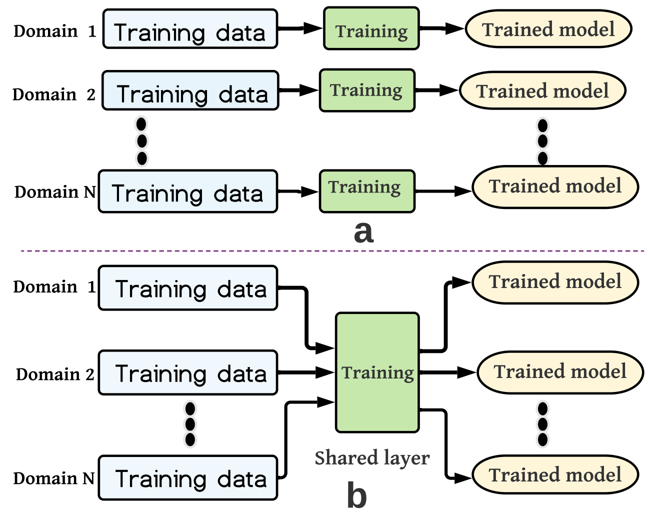



Fig. 1: Comparison of two systematic Architecture : (a) Single Domain learning (SDL) and (b) Multi-domain learning (MDL), denotes model learning in either SDL or MDL architecture.

$$\boldsymbol{Attention}(\boldsymbol{Q}, \mathrm{K}, \boldsymbol{V}) = softmax(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^T}{\sqrt{D_k}})\boldsymbol{V} \qquad (1)$$

The dot-products of $Q$ and $K$ are scaled by $\sqrt{D_k}$ to mitigate the softmax gradient vanishing problem. Often the softmax $(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^T}{\sqrt{D_k}})$ is called the attention matrix. The computation complexity of the self-attention softmax function $(\boldsymbol{Q}\cdot\boldsymbol{K}^T)\boldsymbol{V}$ is quadratic with respect to the length of an input data sequence which increases the training time and adds more parameters. This is the limitation of the self-attention mechanism which is addressed in Section III-C by proposing a linear attention mechanism.

In this work, we propose an MDL network to share knowledge between activity recognition domains, improve accuracy and reduce the need for labelled data. We compare the proposed network with several existing models.

## III. Proposed network

The proposed MDL network aims to improve the performance of HAR systems, decrease the learning time and reduce the number of learned models. The distinctive characteristics of the proposed network are: (1) the proposed network simultaneously processes different but related datasets for HAR and provides a recognition performance for each of the datasets; (2) the proposed network preserves the ordering of temporal sequential input data and avoids information flow from future time steps to past time steps using causal convolution; (3) the network uses a linear attention mechanism to focus more on the most pertinent information in the sequence input. The following subsections provide details about the proposed network.

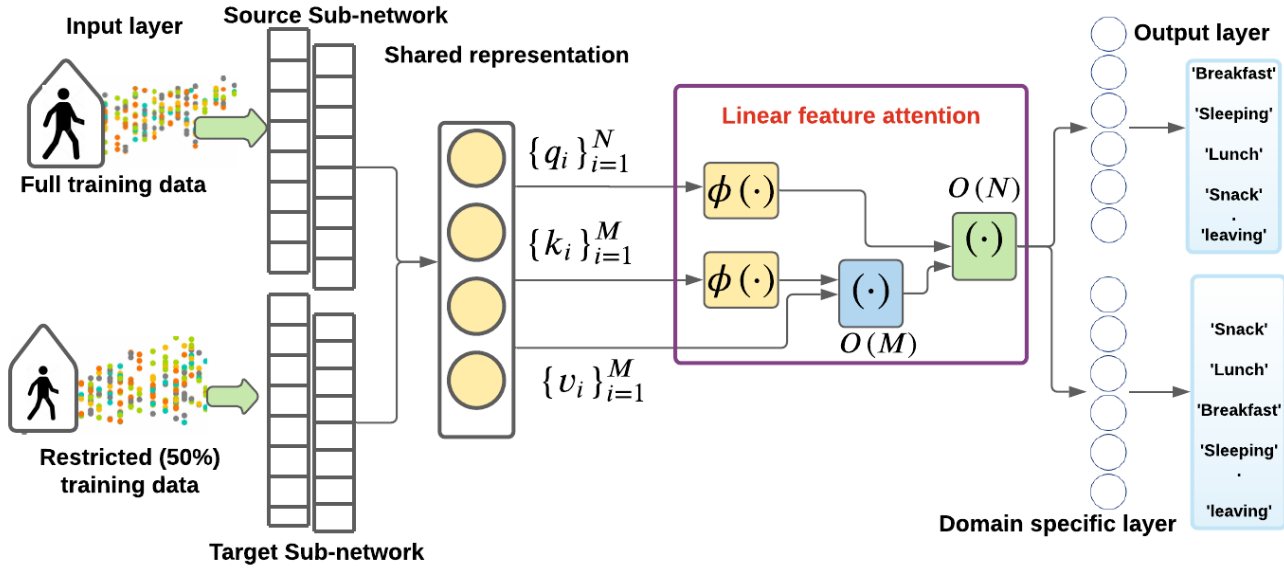



Fig. 2: An overview of proposed network

### A. Proposed MDL

One of the major reasons for the misrecognition of human activities is the unavailability of complementary features which yield semantic information about human activities. In different but related domains, the complementary features are present with various intensities and scales [41]. We propose an MDL network to learn generic and robust feature representations from multiple domains that outperform state-of-the-arts of SDL on all of the domains by large margins. Suppose $X$ and $Y$ are the features and the label spaces respectively. A joint probability distribution P(X,Y), represents a defined domain on $X \times Y$. $P_k$(X) and $P_k$(X,Y) denote the marginal distribution and joint distribution of the datasets in the $k$-th domain [42]. Every single dataset is associated with a sample $D_k$={ $x_i, y_i$ $\}_{i=1}^{L_k}$ where $L_k$ is the sample size of the $k$-th domain. Given $N$ related domains $P_{d=1}^N$(X, Y) and their corresponding datasets $D_k$={ $x_i, y_i\}_{i=1}^{L_k}$ from multiple human activity domains, the goal of the proposed MDL model is to learn a robust and multi-branch model $f : X_k \rightarrow Y_k$, $k$ = {1, 2, . . . , N} and to perform HAR on all the domains in parallel. The proposed model is able to learn features from multiple domains using a shared representation and inference of all the domains. Fusing multi-domain features enable the proposed network to learn better features from multiple datasets for HAR rather than learning each domain in isolation. The shared representation of MDL aims to project learned features to shared feature space and to transfer knowledge between the domains. The proposed network shares data between two domains: source and target domains. Besides, the proposed network can be extended for sharing knowledge among many domains with having different but related domains data.

### B. Architecture of Proposed Network

The structure of the proposed network consists of two identical sub-networks which accept distinct inputs but are joined by a shared representation. Each of the sub-networks consists of two layers of causal 1D CNN followed by a fully connected shared layer. Then the linear attention mechanism is appended to the shared layer followed by a domain-specific layer for each of the domains. The feature maps are fed into the softmax output layer for HAR. The softmax activation function is only used for the output layer which provides a probability of each class and thus the sum of these probabilities is bound to be one. The softmax activation function is commonly used in the output layer for multiclass classification of neural networks. In the softmax activation function, increasing the output value of one activity makes the other activities go down, hence the proposed network will output the activity with the highest probability value. Figure 2 shows the structure of the proposed network. Causal 1D CNN processes temporal sequential inputs independently and performs operations in parallel with avoiding information leakage from future time steps to past time steps of the sequence input data [27], [43], [44]. The causal convolution preserves the ordering of time steps and controls the proposed network to predict an activity at time $t$ based on only the information from time $t$ and earlier in the input sequence [27], [45]. The number of the filters of the convolutional layers is 64 for each of the sub-networks and the kernel size of the convolutional layers which specify the length of the convolution window is equal to 3. The stride of the convolutional filters that control how the filter convolves around the temporal sequential inputs is equal to 1 in order to shift the convolutions one unit at a time. The Adam optimizer and cross-entropy loss function are used in the proposed network.

The first advantage of the proposed network is to reduce the need and effort for producing a large set of labelled data. Thereby the network uses the full training data of the source domain and a restricted size of the training data of the target domain. Yet the proposed network can achieve better performance than using the full training data of the target domain in a single domain setting. Secondly, the proposed method reduces the training time by training a generic model for two domains compared to fitting a model for each domain separately. The proposed network enriches the diversity of the training data due to the domain discrepancies using the shared representation. The diversity of the training data from the proposed network ensures that the training process can provide more discriminative features to the model. Parameter updating is mirrored using a shared representation between the sub-networks. The proposed network improves recognition performance for both domains by transferring knowledge across the domains.

The shared layer within the proposed network shares a layer between the domains to exploit discriminating and different features from both domains to render a robust and mutually complementary feature. The complementary features enrich the learning by bringing different features from each of the domains and each sub-network within the proposed network improves the earlier layers of the other sub-network to boost the performance of the proposed network. The shared layer supports jointly optimizing the sub-networks and increases the functionality of the sub-network to gain more insight into the domains to improve recognition accuracy. The shared representation reduces the number of parameters by half which can mitigate overfitting [46]. The effectiveness of transfer learning using the proposed approach is evaluated by the performance of the SDL model on the target domain using full training data. Algorithm 1 shows further information about how the layers of the proposed network are stacked.

### C. Linear Attention Mechanism

The self-attention technique scales quadratically with the length of the input temporal data and adds more weight parameters to methods that increase learning time and requires more memory. To address this problem, random feature attention (RFA) is proposed as a linear attention mechanism [47] that uses random feature methods to approximate the softmax function. RFA disentangle the softmax $(\boldsymbol{QK}^T)$ into $\boldsymbol{QK}^T$ and compute $\boldsymbol{QK}^TV$ in reverse order $\boldsymbol{Q}(\boldsymbol{K}^TV)$ leading to a linear time and space attention. RFA approximate or replace the unnormalized attention matrix exp $(\boldsymbol{QK}^T)$ with $\phi(\boldsymbol{Q})$ $\phi(\boldsymbol{K})$ where $\phi$ is a feature map that is applied in row-wise manner. Therefore, RFA linearly computes unnormalized attention matrix by $\phi(\boldsymbol{Q})(\phi(\boldsymbol{K})^TV))$, as illustrated in Figure 3. The vector form of the self-attention mechanism is shown in Equation 2. RFA vector form is shown in Equation 3. RFA as a linear attention mechanism is adopted for our proposed network to increase the capability of our proposed network in extracting fine-grained features of human activities.

$$\boldsymbol{Att}(\boldsymbol{Q}_t, \{\boldsymbol{K}_i\}, \{\boldsymbol{V}_i\}) = \sum_i \frac{exp(\boldsymbol{Q}_t \cdot \boldsymbol{K}_i/\mathcal{T})}{\sum_j exp(\boldsymbol{Q}_t \cdot \boldsymbol{K}_j/\mathcal{T})} \boldsymbol{V}_i^\top \quad (2)$$

$$RFA(\boldsymbol{Q}_t, \{\boldsymbol{K}_i\}, \{\boldsymbol{V}_i\}) = \frac{\phi(\boldsymbol{Q}_t)^\top \sum_i \phi(\boldsymbol{K}_i) \otimes \boldsymbol{V}_i}{\phi(\boldsymbol{Q}_t) \cdot \sum_j \phi(\boldsymbol{K}_j)} \quad (3)$$

**Algorithm 1** Cross-domain learning by shared representation

**Input:** $Source\_domain, Target\_domain$
First causal CNN layer for each of the sub-networks
Layer normalization and Dropout layer
Second causal CNN layer for each of the sub-networks
Layer normalization and Dropout layer
Fully connected shared layer to join both sub-networks
Layer normalization and Dropout layer
Apply Linear attention
Fully Connected layer for each of the sub-network
**Output:** Soft-Max Layer to classify activities

## IV. Evaluation of Sensor-Based Activity Recognition Using Cross-Domain Learning

In this section, the proposed network is evaluated based on human activities that are collected from smart home sensors and wearable sensors. Human activity datasets followed by the used deep learning models in this study are reported. Then the experimental results are discussed. The Google Colab that provides a single 12GB NVIDIA Tesla K80 GPU is used for the experiments.

TABLE 1: Details of the smart environment datasets

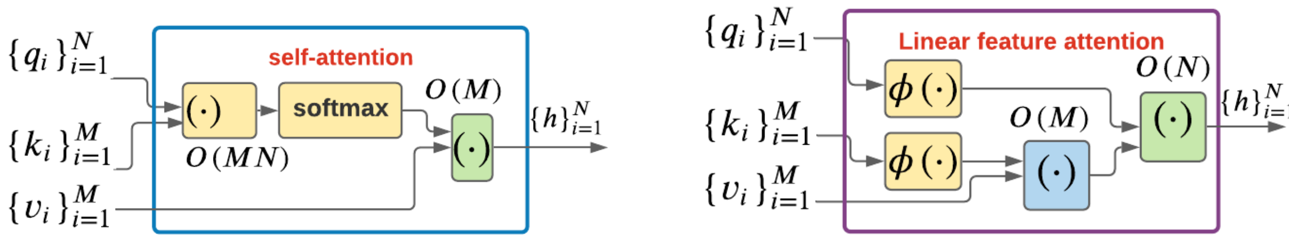
| Smart Homes | Setting | Activities | Rooms | Sensors | Duration (days) |
|---|---|---|---|---|---|
| **Ordonez A** | Home | 10 | 4 | 12 | 14 |
| **Ordonez B** | Home | 11 | 5 | 12 | 21 |
| **Kastern A** | Apartment | 10 | 3 | 14 | 25 |
| **Kastern B** | Apartment | 13 | 2 | 23 | 14 |

TABLE 2: Details of human activities in the Ordonez homes

| Activity | Home A | Home B |
|---|---|---|
| **Leaving** | 1,664 | 5,268 |
| **Breakfast** | 120 | 309 |
| **Spare Time/ TV** | 8,555 | 8,984 |
| **Dinner** | - | 120 |
| **Sleeping** | 7,866 | 10,763 |
| **Snack** | 6 | 408 |
| **Toileting** | 138 | 167 |
| **Lunch** | 315 | 395 |
| **Grooming** | 98 | 427 |
| **Showering** | 96 | 75 |
| **Total** | 20,456 | 30,427 |

TABLE 3: Number of activities in wearable wireless identification and sensing datasets

| **Activity** | **RoomSet1** | **RoomSet2** |
|---|---|---|
| **Ambulating** | 1,956 | 335 |
| **Lying** | 30,983 | 20,537 |
| **Sit on bed** | 15,162 | 1,244 |
| **Sit on chair** | 4,381 | 530 |
| **Total** | 52,482 | 22,646 |




(a) Self-attention. (b) Linearized attention.

Fig. 3: Complexity of standard self-attention and linearized self-attention.

TABLE 4: Details of human activities in the Kasteren datasets

| Activity | Home B | Activity | Home A |
|---|---|---|---|
| **Eat_dinner** | 46 | **Get_drink** | 21 |
| **Brush_teeth** | 25 | **Prepare_Breakfast** | 59 |
| **Get_a_drink** | 6 | **Get_snack** | 24 |
| **Get_dressed** | 27 | **Go_to_bed** | 11,599 |
| **Leaving_the_house** | 12,223 | **Prepare_Dinner** | 325 |
| **Go_to_bed** | 6,050 | **Leave_house** | 19,693 |
| **Prepare_brunch** | 82 | **Take_shower** | 221 |
| **Eat_brunch** | 132 | **Brush_teeth** | 21 |
| **Prepare_dinner** | 87 | **Use_toilet** | 154 |
| **Wash_dishes** | 25 | | |
| **Use_toilet** | 39 | | |
| **Take_shower** | 109 | | |
| **Total** | 38,900 | **Total** | 40,005 |

## A. Datasets

*1) Smart Home Sensor Datasets:* Four HAR datasets based on smart homes are used to perform experiments. Embedded sensors were employed to capture information about human activities in smart homes. Different binary sensors are used to record human daily activities such as sleeping, cooking, and eating [21]. Pressure sensors are used on couches and beds to record the user's presence. Human movements are recorded using different sensors such as passive infrared (PIR), reed switches, and float sensors. Table 1 shows the details of the Ordóñez smart homes A and B [48] regarding the number of rooms in the homes and sensors, duration, sensors, and frequency of activities. Human activities were collected using equipped binary sensors in 14 days from smart home A and 21 days in smart home B. Table 2 shows the details of human activities from Ordóñez homes A and B.

Table 1 also shows the details of the Kastern smart homes A and B [49] regarding the number of rooms in the home and sensors, duration, and activities. In Kasteren home A, 10 human activities were recorded using 14 sensors in 25 days over 40,005 minutes for a 26 years old resident. In Kasteren home B, 13 human activities were recorded using 22 binary sensors in 14 days over 38,900 minutes for a 28 years old resident. Table 4 shows the frequency of activities from Kasteren home A and B datasets. These are single occupancy homes.

*2) Wearable sensor datasets:* Two HAR datasets based on wearable sensor are also used to conduct experiments [50], [51]. These datasets are recorded for ambulatory monitoring of 14 senior patients from 78 to 78 ±4.9 years old. The sensors were worn to record information about four activities: i)ambulating ; ii) lying ; iii) sit on chair; iv) sit on bed. Two clinical room configuration (*Roomset1* and *Roomset2*) were used to record these two datasets. Table 3 shows the distribution of the recorded human activities from both datasets.

*3) Generating Input Data from Collected Smart Homes:* The raw smart home datasets are preprocessed to make the input datasets. The raw sensor readings are segmented into one-minute window size $\Delta$t=1. The smart home raw data consist of the start and end times of the sensor readings and provide information about the place, location and type of the sensors. Multiple and incremental FTW as a feature generating technique is employed to build the input model datasets from the raw sensor readings. The FTW is used as a suitable technique to generate the input datasets for HAR application and the details of FTW are described in [21], [26], [52]–[54]. FTW has exposed that it can extract signal sensors from a short to a long duration of daily routine activities such as snacks or sleeping from sensor raw readings. [52]. Moreover, temporal models have rendered better performance for HAR applications when the FTW is employed to build the input datasets [21], [52]. Algorithm 2 shows the process of generating input datasets from raw sensors data using FTW. The system usually performs near real time as FTW is used.

**Algorithm 2** Generating input datasets by FTWs

**Input:** $Raw\_data$ — Home A and B are the input raw data
$FTWs \leftarrow Fibonacci$ — values from Fibonanci Sequence
$Sensor\_intervals \leftarrow Raw\_data$ — sensor intervals data
**for** $ftw \leftarrow FTWs$ **do**
  **for** $sen\_intv \leftarrow Sensor\_intervals$ **do**
    apply $ftw$ on $sen\_intv$
  **end for**
  $feature\_vector \leftarrow max(ftw)$
**end for**
$input\_dataset \leftarrow feature\_vector$
**Output:** $input\_dataset$

*4) hyperparameters Optimization:* Hyperparameter tuning is a powerful optimization procedure to reach maximum effectiveness and to render accurate deep learning models. After conducting an extensive trial and error process, 0.0001 as the

TABLE 5: F1-score results of the proposed network from Ordonez home A and B datasets: source domain uses all its training data, target domain uses 50% of its training data

| Activity | **Proposed MDL network** | | | | **Existing SDL methods** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **Home A** | | | **Home B** | | |
| | Home A source domain | Home B target domain | Home B source domain | Home A target domain | Deep ConvLSTM | HAR+ Attention | DCC+ MSA | Deep ConvLSTM | HAR+ Attention | DCC+ MSA |
| **Sleeping** | 99.77 | **98.06** | 99.71 | **98.74** | 97.53 | 97.11 | 97.63 | 96.37 | 95.42 | 98.30 |
| **Breakfast** | 88.59 | **83.76** | 85.11 | **86.96** | 83.11 | 84.51 | 85.71 | 74.87 | 75.39 | 76.87 |
| **Lunch** | 98.42 | **99.23** | 99.72 | **97.09** | 95.44 | 94.39 | 96.93 | 95.21 | 95.31 | 99.63 |
| **Grooming** | 85.00 | **90.33** | 92.98 | **82.14** | 75.32 | 80.00 | 80.01 | 85.33 | 87.91 | 88.87 |
| **Spare Time** | 99.16 | **84.29** | 86.15 | **98.65** | 96.83 | 97.21 | 98.57 | 78.21 | 79.32 | 81.48 |
| **Leaving** | 98.05 | **96.20** | 98.00 | **96.36** | 95.29 | 95.51 | 98.79 | 89.79 | 92.09 | 93.33 |
| **Snack** | 86.59 | **80.91** | 83.32 | **85.31** | 70.74 | 82.63 | 84.82 | 76.16 | 76.31 | 78.59 |
| **Showering** | 95.54 | **84.11** | 86.19 | **97.34** | 80.65 | 86.89 | 93.84 | 79.43 | 79.12 | 82.84 |
| **Toileting** | 83.85 | **87.63** | 89.57 | **81.29** | 69.89 | 77.25 | 79.76 | 83.56 | 83.24 | 86.11 |
| **Dinner** | | **89.11** | 91.70 | | | | | 86.19 | 86.49 | 89.45 |
| **Average** | **92.77** | **89.36** | **91.24** | **91.54** | 84.97 | 88.55 | 90.78 | 84.51 | 85.06 | 87.34 |

learning rate, a 20% dropout rate after each learning layer with 64 as the batch size are used for the proposed network to converge. Early stopping is used as a regularization technique to determine the number of epochs and to prevent overfitting by stopping the training when the validation error of the proposed network starts increasing. The 20% dropout rate as a regularization technique after each learning layer is used to further avoid overfitting [55]. To normalize the feature maps across the batches, batch normalization is utilised after each learning layer [56] for making the proposed network faster and more stable during training. It was noted that by the original paper of batch normalization, the batch normalization may reduce the need for dropout technique [56], however, the later successful networks such as GAN [57] and Inception V4 [58] still utilize both batch normalization and dropout techniques, which improves the recognition performance. Therefore, we use the same strategy (using batch normalization and both dropout) in our paper.

## *B. Results*

The experimental results and findings of the proposed MDL network are exposed and discussed. In this section, a comparative analysis between the proposed MDL network and other state-of-the-art methods is presented. The state-of-the-art methods employed in this analysi are DCC+MSA [27] DeepConvLSTM [33] and HAR+Attention [34]. F1-score as a performance metric is used to evaluate the results of the experiments. The F1-score ($2\frac{\text{precision}\times\text{recall}}{\text{prscecision}+\text{recall}}$) is the weighted average of recall ($\frac{TP}{TP+FN}$) and precision ($\frac{TP}{TP+FP}$) [52]. F1-score as a suitable metric is broadly conducted in HAR applications due to the imbalance nature of HAR [13], [21], [52], [54], [59].

The leave-one-day-out cross-validation is used for the evaluation of the smart home datasets as it is commonly used for HAR [21], [23], [52], [53]. The collected data for a single day of the human activities are used for the testing set and collected data for the rest of the days are used for the training set. This process is repeated until the collected data from all the days are involved in the training set and the testing set. Regarding the wearable sensor data, we used k-fold cross-validation because information about collection dates are not provided in the wearable sensors data. The average F-score is computed from the results of the cross-validation as done in the following research.

The effectiveness of the proposed MDL network is evaluated by comparing with the existing SDL state-of-the-art methods. The proposed network processes full training data of a source domain while using only 50% training data of a target domain. The proposed MDL network using 50% training data of a target domain can outperform the SDL method on the same datasets with all the training data. Hence the proposed MDL network can reduce the need for labelled data using cross-domain learning by jointly training several domains. The results of the source domain in addition to the target domain are also improved compared to the existing state-of-the-art methods. The results of the target domains based on the proposed MDL network are highlighted in the tables to be easily compared with the results of the current methods. Furthermore, we performed an ablation study to show the performance of the proposed network where the training target domain data is 50%, 25% or 10%.

*1) Results from Ordonez Environments' datasets :* Table 5 shows results of the proposed MDL network and existing methods from Ordonez home A and B datasets. The results of the proposed MDL network is compared with the existing SDL state-of-the-art methods. The proposed MDL network has achieved better results compared to the results obtained by the state-of-the-art methods. The results of the target smart home domains with 50% of their training data based on the proposed MDL network are significantly improved compared to their results obtained by the SDL approach with full training data. In addition to the improvement of the target domains, the results of the source domains are also improved. This is revealed that the proposed MDL network transfers knowledge between both the source and target human activity domains using the shared representation. The shared layer exposes different features from source and target domains to render strong mutual complementary features. Complementarity in the proposed MDL using a shared representation boosts the recognition performance. This is because the model in the MDL approach delivers distinctive features from both source and target human activity domains to enrich the training and each model refines the earlier layers of the other model and reduces their weaknesses. The joint optimization of the MDL

TABLE 6: F1-score results of the proposed network from Kasteren home A and B datasets: source domain uses all its training data, target domain uses 50% of its training data

| Activity | Proposed MDL network | | | | Existing SDL methods | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Home A | | | Home B | | |
| | Home A source domain | Home B target domain | Home B source domain | Home A target domain | Deep ConvLSTM | HAR+ Attention | DCC+ MSA | Deep ConvLSTM | HAR+ Attention | DCC+ MSA |
| **Get_Snack** | 75.29 | | | **71.73** | 57.22 | 58.71 | 63.69 | | | |
| **Prepare_breakfast** | 86.46 | | | **83.72** | 76.97 | 79.54 | 83.32 | | | |
| **Brush_teeth** | 60.35 | **58.98** | 60.52 | **56.96** | 43.59 | 52.22 | 54.44 | 42.89 | 47.82 | 51.18 |
| **Get_drink** | 70.87 | **62.86** | 65.93 | **67.61** | 59.33 | 59.54 | 66.92 | 44.15 | 44.75 | 53.00 |
| **Go_to_bed** | 89.06 | **99.62** | 99.81 | **87.42** | 80.16 | 81.76 | 86.54 | 96.32 | 94.48 | 99.73 |
| **Leave_house** | 88.95 | **96.97** | 97.19 | **85.98** | 80.02 | 82.19 | 84.28 | 92.98 | 93.21 | 96.39 |
| **Prepare_Dinner** | 97.91 | **97.63** | 99.51 | **95.92** | 89.56 | 91.87 | 95.42 | 96.21 | 95.31 | 97.51 |
| **Take_shower** | 92.75 | **84.92** | 86.49 | **90.26** | 85.13 | 89.23 | 89.11 | 81.95 | 82.13 | 83.13 |
| **Use_toilet** | 75.79 | **65.03** | 69.06 | **72.82** | 67.82 | 69.34 | 71.97 | 56.13 | 54.19 | 62.18 |
| **Eat_brunch** | | **94.1** | 96.21 | | | | | 90.93 | 91.11 | 95.92 |
| **Eat_dinner** | | **88.06** | 91.68 | | | | | 86.79 | 86.29 | 90.02 |
| **Prepare_brunch** | | **85.26** | 89.85 | | | | | 85.62 | 84.29 | 88.10 |
| **Get_dressed** | | **48.32** | 50.54 | | | | | 31.79 | 41.11 | 42.63 |
| **Wash_dishes** | | **81.63** | 84.92 | | | | | 75.38 | 77.25 | 82.36 |
| **Average** | **81.93** | **80.28** | **82.45** | **79.50** | 71.09 | 73.82 | 77.29 | 73.42 | 74.32 | 78.51 |

TABLE 7: F1-score results of the proposed network from Wearable sensors Roomset 1 and Roomset 2: source domain uses all its training data, target domain uses 50% of its training data

| Activity | Proposed MDL network | | | | Existing SDL methods | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Roomset 1 | | | Roomset 2 | | |
| | Roomset 1 source domain | Roomset 2 target domain | Roomset 2 source domain | Roomset 1 target domain | Deep ConvLSTM | HAR+ Attention | DCC+ MSA | Deep ConvLSTM | HAR+ Attention | DCC+ MSA |
| **Ambulating** | 98.79 | **93.90** | 95.23 | **97.73** | 93.67 | 95.22 | 97.63 | 87.11 | 89.79 | 91.93 |
| **Sit on bed** | 99.30 | **98.61** | 99.35 | **99.46** | 95.31 | 96.25 | 99.90 | 97.52 | 99.79 | 99.95 |
| **Lying** | 98.86 | **95.70** | 96.37 | **97.71** | 94.12 | 95.21 | 97.70 | 89.32 | 94.85 | 95.32 |
| **Sit on chair** | 75.31 | **74.31** | 76.89 | **73.45** | 60.52 | 70.11 | 72.84 | 68.87 | 69.87 | 72.31 |
| **Average** | **93.06** | **90.63** | **91.96** | **92.08** | 85.90 | 89.19 | 92.02 | 83.40 | 85.70 | 88.57 |

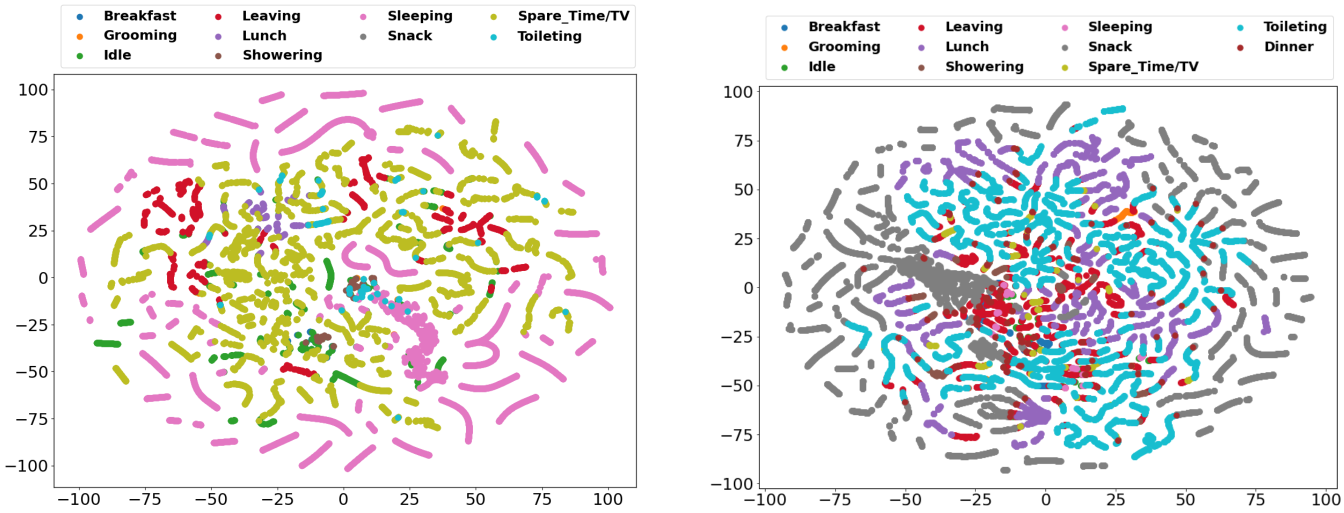
approach enhances the functionality of the proposed network to gain more insight into the source and target domain features to increase the recognition result score. Figure 4 shows the t-SNE map for human activity recognition from Ordonez smart home A and B datasets, while Figure 5 shows a t-SNE map for human activity recognition from the proposed MDL network without the last softmax layer. The plot shows how the proposed network properly distinguishes the activities. Furthermore, Figure 6 shows the loss and accuracy of training and validation with early stopping technique in Ordonez Home A and B. Besides, Figure 7 shows the confusion matrix one of the runs of the proposed network from the Ordonez Home A and B datasets.

*2) Results from Kasteren Datasets :* Table 6 presents the outcomes of the proposed MDL network compared to existing methods from the Kasteren smart home A and B datasets. The results of the proposed MDL network is also compared with the existing state-of-the-art methods. The common activities from source and target domains have been shaded in the table. The results of the source domains using the full training data and target domains using 50% of their training data based on the proposed MDL network have been improved compared with the results obtained by the existing methods. The shared representation within the proposed network adjusts learning of the network by exploiting commonalities and differences across domains which improves performance. Some human activities of Home A are not available in the Home B dataset and also some Home B activities are not available in the Home A. The values of missing activities are not counted in the average of the results in Tables 5 and 6.

*3) Results from Wearable Sensors Datasets :* Table 7 shows the results of the proposed MDL network and existing methods from the wearable sensors datasets. The results of source and target domains are improved using the proposed MDL network compared to the existing state-of-the-art methods. The results indicate the effectiveness of the proposed MDL network for sharing knowledge between source and target domains using the shared representation.

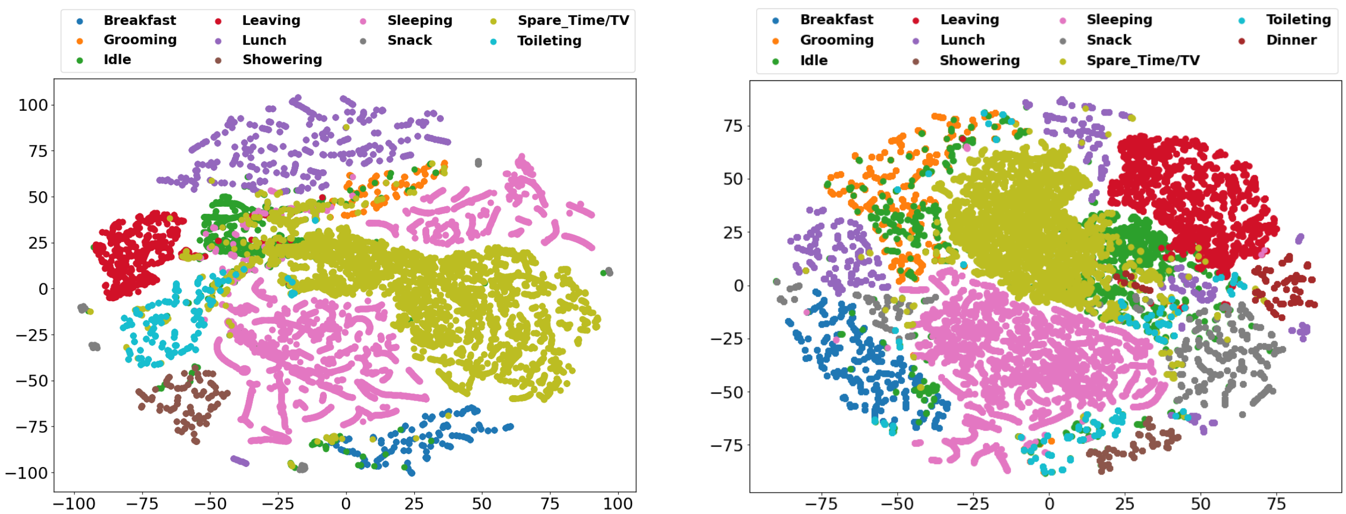
*4) Ablation study:* We performed ablation studies to show the performance of the proposed network where the percentage of the used training data of the target domains are 50%, 25%, or 10%. The results show that the 50% of the training data is the best configuration and renders better results as shown in Table 8. Moreover, the performance of the proposed network based on 25% or 10% of the training target domain data is compared with the state-of-the-art methods when the training data is 25% or 10%. Figure 8 shows the performances of the proposed network are yet better compared to the performances achieved by the existing state-of-the-art methods. This is because the proposed network uses a shared representation to transfer knowledge across learning processes between the source and target domains. Furthermore, by sharing knowledge across learning processes between the source and target domains, knowledge from one of the domains is distilled to facilitate the learning of the other domain.

Further ablation experiments are performed to show the effectiveness of the causal convolution and linear attention of the proposed network. Table 9 shows the results of the



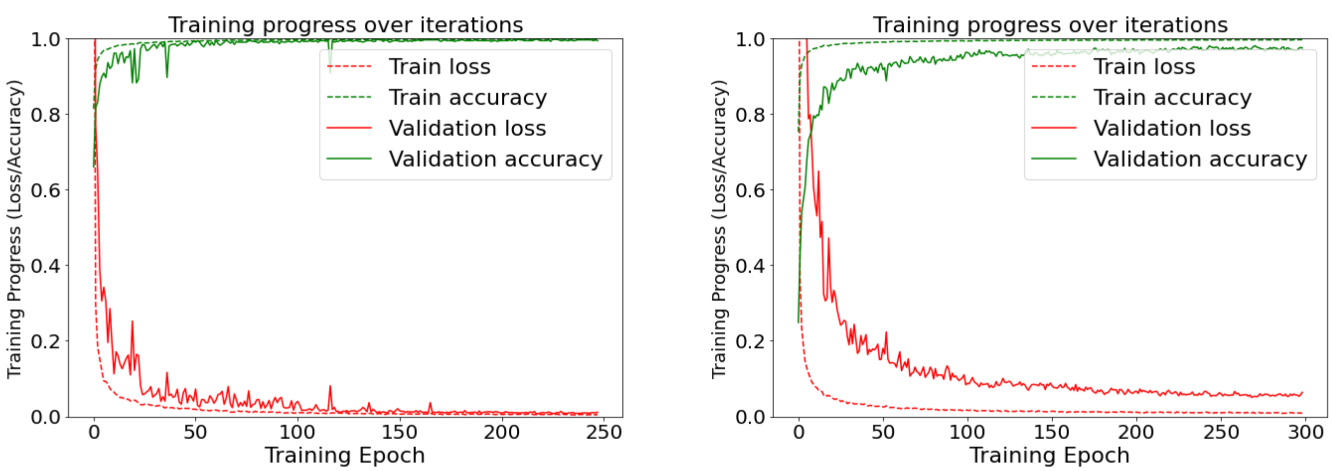

(a) Ordonez Home A (b) Ordonez Home B

Fig. 4: t-SNE map for human activity from input datasets




(a) Ordonez Home A (b) Ordonez Home B

Fig. 5: t-SNE map for human activity from proposed MDL

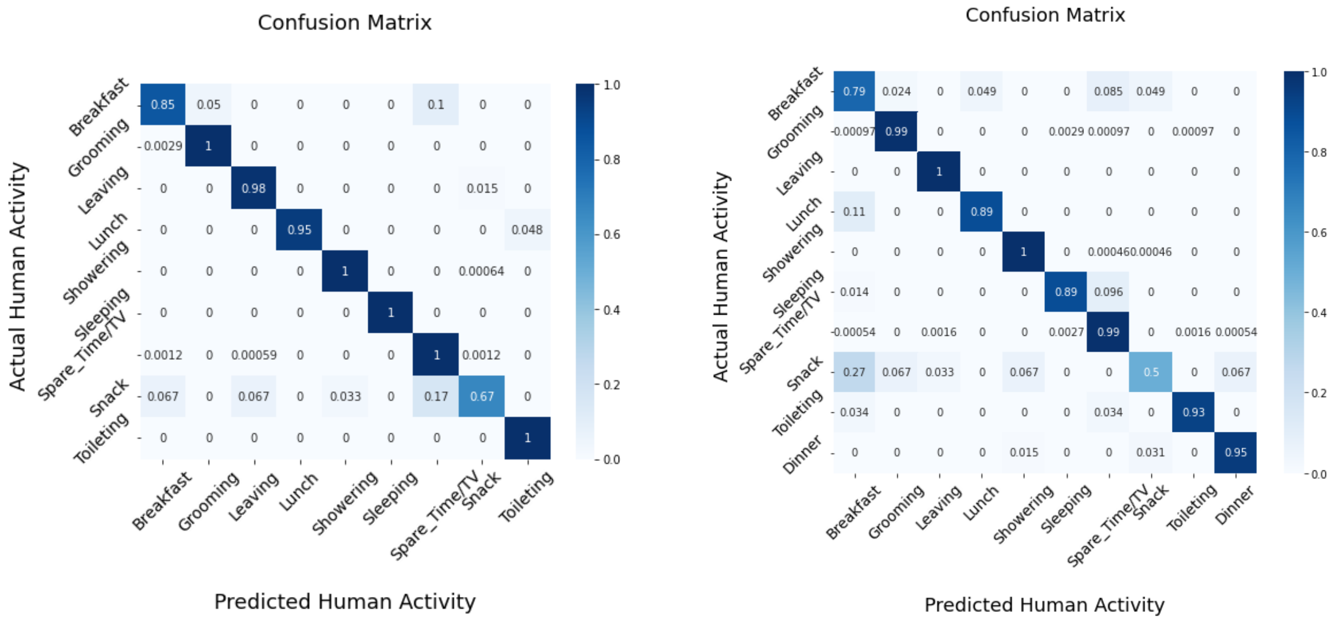



(a) Ordonez Home A (b) Ordonez Home B

Fig. 6: Loss and accuracy of training and validation with early stopping technique

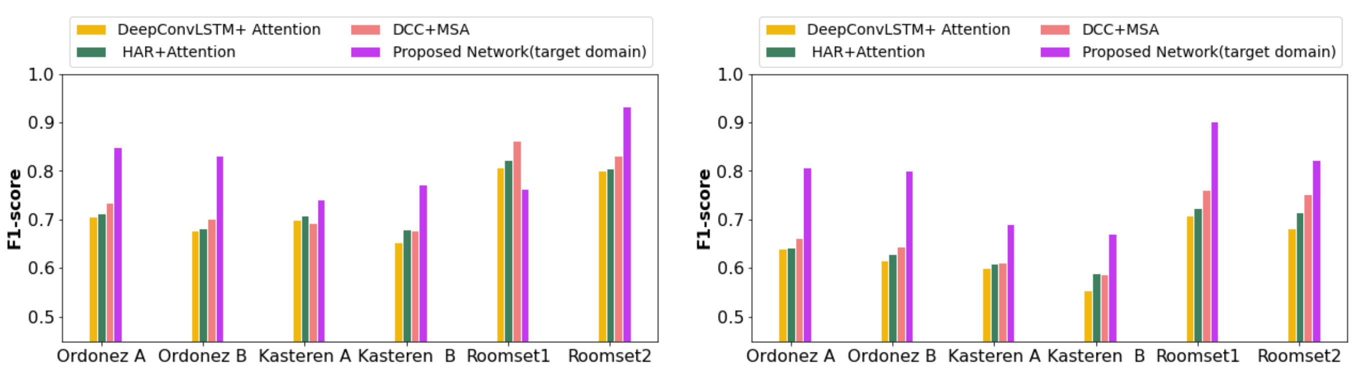
TABLE 8: F1-score results of the proposed network where target domain uses 50%, 25% and 10% of its training data

| Datasets | 50% target data | | 25% target data | | 10% target data | | 50% target data | | 25% target data | | 10% target data | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | source domain | target domain | source domain | target domain | source domain | target domain | source domain | target domain | source domain | target domain | source domain | target domain |
| | Home A | Home B | Home A | Home B | Home A | Home B | Home B | Home A | Home B | Home A | Home B | Home A |
| **Ordonez** | 92.77 | 89.36 | 88.67 | 83.21 | 87.81 | 80.11 | 91.24 | 91.54 | 89.54 | 84.85 | 85.13 | 80.04 |
| **Kasteren** | 81.93 | 80.28 | 80.45 | 77.50 | 77.12 | 73.53 | 82.11 | 79.98 | 79.45 | 74.11 | 77.87 | 69.54 |
| | Roomset 1 | Roomset 2 | Roomset 1 | Roomset 2 | Roomset 1 | Roomset 2 | Roomset 2 | Roomset 1 | Roomset 2 | Roomset 1 | Roomset 2 | Roomset 1 |
| **Wearable sensors** | 93.06 | 90.63 | 90.28 | 85.21 | 86.88 | 82.23 | 91.96 | 92.08 | 89.15 | 86.17 | 86.76 | 80.18 |




(a) Ordonez Home A (b) Ordonez Home B

Fig. 7: Confusion matrix




(a) 25% training data (b) 10% training data

Fig. 8: F1-score results of the proposed network and state-of-the-art methods with 25% or 10% of training target domain data

TABLE 9: Results of F1-score of ablation studies of the proposed network

| Datasets | Source domain | | | | | Target domain | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Just CNN** | **Without causal** | **Without linear attention** | **with Self-attention** | **proposed Network** | **Just CNN** | **Without causal** | **Without linear attention** | **with Self-attention** | **proposed Network** |
| **Ordonez Home A** | 86.59 | 88.02 | 89.38 | 90.11 | **92.77** | 85.31 | 86.03 | 86.24 | 88.27 | **91.54** |
| **Ordonez Home B** | 84.92 | 85.46 | 88.13 | 88.67 | **91.24** | 85.71 | 86.52 | 87.59 | 88.13 | **89.36** |
| **Kastern Home A** | 76.34 | 77.51 | 78.00 | 78.82 | **81.93** | 74.29 | 76.18 | 77.04 | 78.05 | **79.50** |
| **Kastern Home B** | 76.28 | 77.15 | 79.22 | 80.19 | **82.45** | 73.03 | 75.88 | 77.73 | 77.92 | **80.28** |
| **Wearable RoomSet1** | 86.16 | 88.74 | 90.34 | 90.47 | **93.06** | 85.62 | 86.39 | 88.75 | 89.67 | **92.08** |
| **Wearable RoomSet2** | 86.12 | 86.82 | 88.56 | 90.00 | **91.96** | 86.11 | 86.48 | 88.73 | 89.27 | **90.63** |

proposed network without using the causal convolution and linear attention. The results of the proposed network with just using CNN without causal and attention mechanisms. The results show that the proposed method is affected by each of the causal convolution and the linear attention mechanism. For instance, the proposed network obtained the F-score of 88.59 without using causal convolution and obtained the F-score of 89.38 without using linear attention for the Ordonez home A as a source domain. Moreover, the proposed network obtained the F-score of 86.59 with only using general CNN without using causal and attention mechanism. This indicates that the causal convolution has a higher contribution compared to the linear attention. However, the linear attention mechanism still can contribute more than the self-attention mechanism as shown in the results of Table 9. The results indicate that the network obtained a minimum F-score based on only using CNN and not using causal and attention mechanism. Moreover, the proposed network obtained the F-score of 91.54 for the Ordonez home A as the target domain with the linear attention while the proposed network based on the self-attention obtained the F-score of 88.27. The results of the ablation experiments indicate that the causal convolution and linear attention mechanism have an effective contribution to the proposed network.

## V. Conclusion

The paper proposes an MDL network to further improve the accuracy of human activity recognition from smart home and wearable sensors data compared to the existing methods. The proposed MDL network learns on two different but related sensor generated domains using a shared representation. The shared representation transfers knowledge across the domains to make strong mutually complementary features that improve the recognition rate and mitigate the data scarcity problems. The proposed MDL network uses the source domain with its full training data and the target domain with 50% of its training data to reduce the requirement of large labelled training dataset. The proposed MDL network is evaluated based on six human activity datasets. Extensive experiments are conducted to make a comparative analysis for the results that are obtained by the proposed MDL network and existing methods. The experimental results on the source and target domains achieved based on the proposed MDL network are better than those achieved by the existing methods. The effectiveness of the proposed MDL network is validated by the results on the target domains even though only 50% of the training data is used in the learning compared the existing methods which use the full training data. In addition to improving the results of the target domains, the results of the source domains are also improved using the proposed network compared to the existing state-of-the-art models. Moreover, an ablation study is performed to show the performance of the proposed network when the target data is reduced to 25% or 10% which are also compared with the existing methods. The results indicate that the proposed network with 25% or 10% for the target data renders better performance against the state-of-the-art methods based on the 25% or 10% of the training data. The requirement of the full labelled data of the source domain is the limitation of the proposed network. Hence, future work will investigate the elimination of a significant amount of labelled data from the source and target domains using unsupervised cross-domain learning.

## References

[1] Xin Cheng, Lei Zhang, Yin Tang, Yue Liu, Hao Wu, and Jun He. Real-time human activity recognition using conditionally parametrized convolutions on mobile and wearable devices. *IEEE Sensors Journal*, 22(6):5889–5901, 2022.

[2] Tingting Liu, Hai Liu, You-Fu Li, Zengzhao Chen, Zhaoli Zhang, and Sannyuya Liu. Flexible ftir spectral imaging enhancement for industrial robot infrared vision sensing. *IEEE Transactions on Industrial Informatics*, 16(1):544–554, 2019.

[3] Diane Cook, Kyle D Feuz, and Narayanan C Krishnan. Transfer learning for activity recognition: A survey. *Knowledge and information systems*, 36(3):537–556, 2013.

[4] Xiaoye Qian, Huan Chen, Yi Cai, Kuo-Chung Chu, Wenyao Xu, and Ming-Chun Huang. Transfer learning model knowledge across multi-sensors locations over body sensor network. *IEEE Sensors Journal*, 2022.

[5] Derek Hao Hu, Vincent Wenchen Zheng, and Qiang Yang. Cross-domain activity recognition via transfer learning. *Pervasive and Mobile Computing*, 7(3):344–358, 2011.

[6] Yongxin Yang and Timothy M Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014.

[7] Uday Kamath, John Liu, and James Whitaker. Transfer learning: Scenarios, self-taught learning, and multitask learning. In *Deep Learning for NLP and Speech Recognition*, pages 463–493. Springer, 2019.

[8] Guanyu Cai, Yuqin Wang, Lianghua He, and MengChu Zhou. Unsupervised domain adaptation with adversarial residual transform networks. *IEEE transactions on neural networks and learning systems*, 31(8):3073–3086, 2019.

[9] Qi Kang, SiYa Yao, MengChu Zhou, Kai Zhang, and Abdullah Abusorrah. Effective visual domain adaptation via generative adversarial distribution matching. *IEEE Transactions on neural networks and learning systems*, 32(9):3919–3929, 2020.

[10] Trevor Standley, Amir R Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? *arXiv preprint arXiv:1905.07553*, 2019.

[11] Carlo D'Eramo, Davide Tateo, Andrea Bonarini, Marcello Restelli, and Jan Peters. Sharing knowledge in multi-task deep reinforcement learning. In *International Conference on Learning Representations*, 2019.

[12] Lixin Duan, Ivor W Tsang, and Dong Xu. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):465–479, 2012.

[13] Yunsheng Li and Nuno Vasconcelos. Efficient multi-domain learning by covariance normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5424–5433, 2019.

[14] Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F Wu, and Steve J Altschuler. Multi-domain adversarial learning. *arXiv preprint arXiv:1903.09239*, 2019.

[15] Alvina Anjum and Muhammad U Ilyas. Activity recognition using smartphone sensors. In *2013 ieee 10th consumer communications and networking conference (ccnc)*, pages 914–919. IEEE, 2013.

[16] Md Zia Uddin and Mohammad Mehedi Hassan. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. *IEEE Sensors Journal*, 19(19):8413–8419, 2018.

[17] Yusra Asim, Muhammad Awais Azam, Muhammad Ehatisham-ul Haq, Usman Naeem, and Asra Khalid. Context-aware human activity recognition (cahar) in-the-wild using smartphone accelerometer. *IEEE Sensors Journal*, 20(8):4361–4371, 2020.

[18] Sidrah Liaqat, Kia Dashtipour, Syed Aziz Shah, Ali Rizwan, Abdullah Alhumaidi Alotaibi, Turke Althobaiti, Kamran Arshad, Khaled Assaleh, and Naeem Ramzan. Novel ensemble algorithm for multiple activity recognition in elderly people exploiting ubiquitous sensing devices. *IEEE Sensors Journal*, 21(16):18214–18221, 2021.

[19] Mitchell Webber and Raul Fernandez Rojas. Human activity recognition with accelerometer and gyroscope: A data fusion approach. *IEEE Sensors Journal*, 21(15):16979–16989, 2021.

[20] Davide Buffelli and Fabio Vandin. Attention-based deep learning framework for human activity recognition with user adaptation. *IEEE Sensors Journal*, 21(12):13474–13483, 2021.

[21] R. A. Hamad, A. G. Salguero, M. Bouguelia, M. Espinilla, and J. M. Quero. Efficient activity recognition in smart homes using delayed fuzzy temporal windows on binary sensors. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1, 2019.

[22] Qi Teng, Kun Wang, Lei Zhang, and Jun He. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition. *IEEE Sensors Journal*, 20(13):7265–7274, 2020.

[23] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[24] E Ramanujam, Thinagaran Perumal, and S Padmavathi. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 2021.

[25] Yu Guan and Thomas Plötz. Ensembles of deep lstm learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):1–28, 2017.

[26] Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Applied Sciences*, 10(15):5293, 2020.

[27] Rebeen Ali Hamad, Masashi Kimura, Longzhi Yang, Wai Lok Woo, and Bo Wei. Dilated causal convolution with multi-head self attention for sensor human activity recognition. *Neural Computing and Applications*, pages 1–18, 2021.

[28] Kun Wang, Jun He, and Lei Zhang. Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sensors Journal*, 19(17):7598–7604, 2019.

[29] Wenbin Gao, Lei Zhang, Wenbo Huang, Fuhong Min, Jun He, and Aiguo Song. Deep neural networks for sensor-based human activity recognition using selective kernel convolution. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

[30] Wenbo Huang, Lei Zhang, Wenbin Gao, Fuhong Min, and Jun He. Shallow convolutional neural networks for human activity recognition using wearable sensors. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021.

[31] Yin Tang, Lei Zhang, Qi Teng, Fuhong Min, and Aiguo Song. Triple cross-domain attention on human activity recognition using wearable sensors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022.

[32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[33] Satya P Singh, Madan Kumar Sharma, Aimé Lay-Ekuakille, Deepak Gangwar, and Sukrit Gupta. Deep convlstm with self-attention for human activity decoding using wearable sensors. *IEEE Sensors Journal*, 21(6):8575–8582, 2020.

[34] Saif Mahmud, M. T. H. Tonmoy, Kishor Kumar Bhaumik, A. M. Rahman, M. A. Amin, M. Shoyaib, Muhammad Asif Hossain Khan, and A. Ali. Human activity recognition from wearable sensor data using self-attention. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain*, 2020.

[35] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1249–1258, 2016.

[36] Yitao Cai and Xiaojun Wan. Multi-domain sentiment classification based on domain-aware embedding and attention. In *IJCAI*, pages 4904–4910, 2019.

[37] Prachi Garg, Rohit Saluja, Vineeth N Balasubramanian, Chetan Arora, Anbumani Subramanian, and CV Jawahar. Multi-domain incremental learning for semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 761–771, 2022.

[38] Jindong Wang, Vincent W Zheng, Yiqiang Chen, and Meiyu Huang. Deep transfer learning for cross-domain activity recognition. In *proceedings of the 3rd International Conference on Crowd Science and Engineering*, pages 1–8, 2018.

[39] Jindong Wang, Yiqiang Chen, Lisha Hu, Xiaohui Peng, and S Yu Philip. Stratified transfer learning for cross-domain activity recognition. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE, 2018.

[40] Carlos Betancourt, Wen-Hui Chen, and Chi-Wei Kuan. Self-attention networks for human activity recognition using wearable devices. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1194–1199. IEEE, 2020.

[41] Zeeshan Ahmad and Naimul Mefraz Khan. Multidomain multimodal fusion for human action recognition using inertial sensors. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 429–434. IEEE, 2019.

[42] Yajing Liu, Xinmei Tian, Ya Li, Zhiwei Xiong, and Feng Wu. Compact feature learning for multi-domain image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7193–7201, 2019.

[43] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.

[44] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[45] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[46] Roman Pflugfelder. An in-depth analysis of visual tracking with siamese neural networks. *arXiv preprint arXiv:1707.00569*, 2017.

[47] Hao Peng, Nikolaos Pappas, Dani Yogatama, Roy Schwartz, Noah A Smith, and Lingpeng Kong. Random feature attention. *arXiv preprint arXiv:2103.02143*, 2021.

[48] Fco Ordóñez, Paula De Toledo, Araceli Sanchis, et al. Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors*, 13(5):5460–5477, 2013.

[49] Tim LM van Kasteren, Gwenn Englebienne, and Ben JA Kröse. Human activity recognition from wireless sensor network data: Benchmark and software. In *Activity recognition in pervasive intelligent environments*, pages 165–186. Springer, 2011.

[50] Roberto L Shinmoto Torres, Damith C Ranasinghe, Qinfeng Shi, and Alanson P Sample. Sensor enabled wearable rfid technology for mitigating the risk of falls near beds. In *2013 IEEE International Conference on RFID (RFID)*, pages 191–198. IEEE, 2013.

[51] Asanga Wickramasinghe and Damith C Ranasinghe. Recognising activities in real time using body worn passive sensors with sparse data streams: To interpolate or not to interpolate? In *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services on 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 21–30, 2016.

[52] Javier Medina-Quero, Shuai Zhang, Chris Nugent, and M Espinilla. Ensemble classifier of long short-term memory with fuzzy temporal windows on binary sensors for activity recognition. *Expert Systems with Applications*, 114:441–453, 2018.

[53] Javier Medina Quero, Claire Orr, Shuai Zang, Chris Nugent, Alberto Salguero, and Macarena Espinilla. Real-time recognition of interleaved activities based on ensemble classifier of long short-term memory with fuzzy temporal windows. In *Multidisciplinary digital publishing institute proceedings*, volume 2, page 1225, 2018.

[54] Rebeen Ali Hamad, Masashi Kimura, and Jens Lundström. Efficacy of imbalanced data handling methods on deep learning for smart homes environments. *SN Computer Science*, 1(4):1–10, 2020.

[55] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[56] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[57] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[58] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[59] TL Kasteren, Gwenn Englebienne, and BJ Kröse. An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing*, 14(6):489–498, 2010.