

Received July 24, 2020, accepted August 5, 2020, date of publication August 18, 2020, date of current version September 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017681

# Deep Human Activity Recognition With Localisation of Wearable Sensors

ISAH A. LAWAL<sup>1</sup> AND SOPHIA BANO<sup>2</sup>

<sup>1</sup>Faculty of Applied Computing and Technology (FACT), Noroff University College, 4608 Kristiansand, Norway

<sup>2</sup>Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London W1W 7TY, U.K.

Corresponding author: Isah A. Lawal (isah.lawal@noroff.no)

This work was supported in part by the Noroff University College, under the FACT Research Grant.

**ABSTRACT** Automatic recognition of human activities using wearable sensors remains a challenging problem due to high variability in inter-person gait and movements. Moreover, finding the best on-body location for a wearable sensor is also critical though it provides valuable context information that can be used for accurate recognition. This article addresses the problem of classifying motion signals generated by multiple wearable sensors for the recognition of human activity and localisation of the wearable sensors. Unlike existing methods that used the raw accelerometer and gyroscope signals for extracting time and frequency-based features for activity inference, we propose to create frequency images for the raw signals and show this representation to be more robust. The frequency image sequences are generated from the accelerometer and gyroscope signals from seven different body parts. These frequency images serve as the input to our proposed two-stream Convolutional Neural Networks (CNN) for predicting the human activity and the location of the sensor generating the activity signal. We show that the complementary information collected by both accelerometer and gyroscope sensors can be leveraged to develop an effective classifier that can accurately predict the performed human activity. We evaluate the performance of the proposed method using the cross-subjects approach and show that it achieves an impressive F1-score of 0.90 on a publicly available real-world human activity dataset. This performance is superior to that reported by another state-of-the-art method on the same dataset. Moreover, we also experimented with the datasets from different body locations to predict the best position for the underlying task. We show that shin and waist are the best places on the body for placing sensors and this could help other researchers to collect higher quality activity data. We plan to publicly release the generated frequency images from all sensor positions and activities and our implementation code with the publication.

**INDEX TERMS** Human activity recognition, deep learning, sensor localisation, wearable sensors.

## I. INTRODUCTION

The ubiquity and functionality of wearable devices such as smartphones, smartwatches, and fitness wristbands equipped with motion sensors (e.g. accelerometer and gyroscope) create new opportunities for continuous monitoring of human physical activities [1]. Since many human activities can be reliably recognised based on the motion information, the automatic and accurate classification of motion signals generated by the motion sensors can facilitate the development of an effective automated human activity recogniser (HAR) for human-centred monitoring systems [2]. The importance of HAR in sectors such as healthcare, fitness, sports, and

entertainment cannot be overemphasised [3]. For example, HAR systems are used to monitor human activities to aid medical diagnosis and assisting patients with impaired physical mobility [4]. Similarly, HAR systems are been incorporated in many home entertainment products such as the Microsoft Kinect for the recognition of hand gestures and body movements to enhance gaming experience [5].

Recently people (either for health or personal reasons) have adopted the habit of carrying two or more wearable devices such as smartwatch and smartphone. While the complementary motion information gathered by these multiple sensors can be combined to improve the accuracy of the activity recogniser, the detection of the on-body position of the sensors is important because the quality of automatic activity recognition depends largely on the position of sensor

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

providing the motion signals. This article deals with the accurate recognition of both human activities and position of the wearable device generating the motion information. We explore the idea of converting raw motion signals into frequency-based image sequences [6], [7] and developing a cooperative two-stream convolutional neural network for the prediction of the human actions and sensor location. The contribution of this article includes:

- The design of a monolithic two-stream Convolutional Neural Network (CNN) for predicting both human actions and the different sensor locations referred to as Deep Human Activity and Location Recognition (DHALR). The device localisation allows examining the impact of the position information on the accuracy of the activity recognition.
- To the best of our knowledge, we contribute the first approach for simultaneous recognition of both human activity and sensor location using frequency images.
- Extensive experimentation that shows the effectiveness of using the combination of complementary motion information from multiple devices for improving the recognition of activities in a real-world setting.

The rest of the paper is organised as follows: Section 2 provides the related work, Section 3 discusses the methodology, Section 4 presents the dataset, Section 5 discusses the experimental setup and Section 6 presents the discussion of results. Section 7 concludes the paper.

## II. RELATED WORK

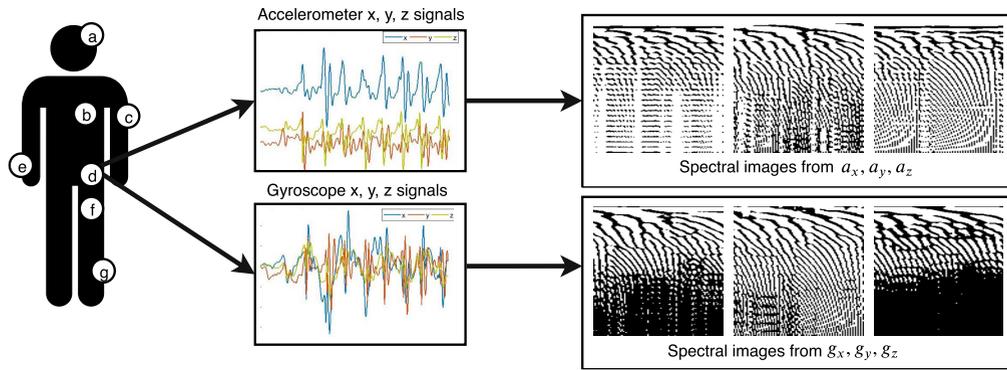
The problem of human activity recognition using wearable sensors involves the characterisation of the body parts motion using sensory data [8]. The motion data usually comprises of the physical acceleration and orientations of movable body parts, measured using accelerometer and gyroscope, respectively [9]. Machine learning methods such as Support Vector Machine (SVM) [10], Random Forest (RF) [11], Long Short-Term Memory Network (LSTM) [12] and Convolutional Neural Networks (CNNs) [6] have been used to develop the characterisation model. Ortiz Jorge [13] characterised the motion data obtained with the sensors in a waist-mounted smartphone to recognise six human activities. The method employed SVM to analyse the hand-crafted features develop from the motion data. The method accurately recognised dynamic activities such as walking and climbing but missed recognising most of the static actions like sitting and standing. Catal et al. [14] employed an ensemble approach that combined multiple classifiers to improve the accuracy of human activity recognition. The method also used hand-crafted features estimated from raw acceleration data. Inoue *et al.* [15] on the other hand, avoided the costly feature engineering process mentioned in the previous methods, by directly using the raw accelerometer data as input to train a deep recurrent neural network for human activity recognition. The study recorded an improved recognition performance and lower learning time. Nair et al. [16] also proposed a method

**TABLE 1. Main characteristics of the existing human activity recognition methods and their comparison with our proposed DHALR method.**

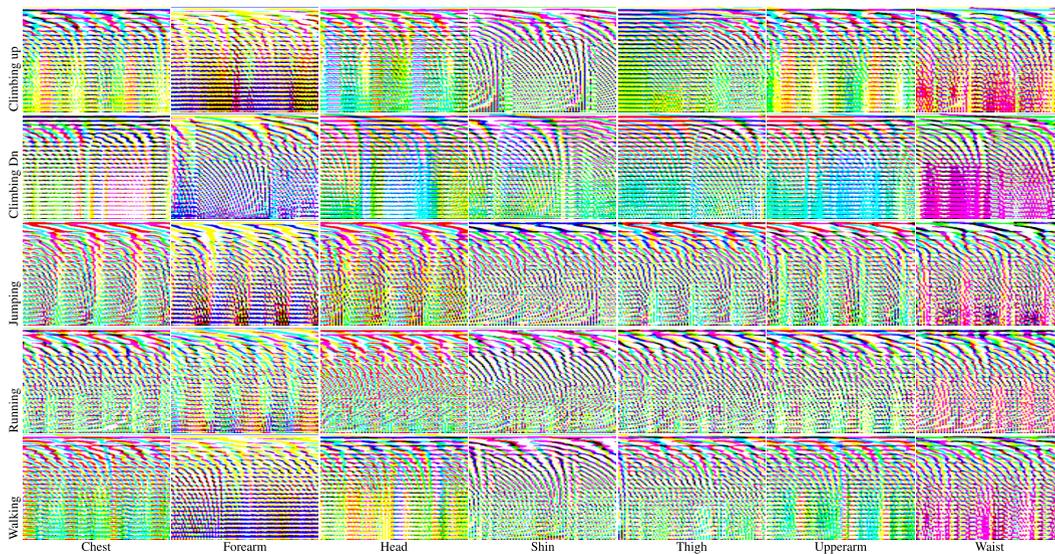
Properties	[14]	[15]	[7]	[18]	[6]	[16]	[13]	[11]	DHALR
Incorporates at least 2-sensor readings					✓		✓	✓	✓
Records a substantial recognition accuracy	✓	✓	✓	✓	✓	✓	✓	✓	✓
Incorporate on-body device localisation				✓				✓	✓
Requires no costly feature engineering		✓	✓		✓	✓			✓
Employs cross-subject performance evaluation					✓			✓	✓
Simultaneous activity & location recognition									✓

that used temporal CNN for recognising human activities from raw motion signals acquired using smartphone sensors. Lawal and Bano [6] proposed a CNN-based model for recognising human activities. In contrast to previously mentioned approaches, the method [6] used two sets of frequency image sequences generated from the raw accelerometer and gyroscope signals, respectively as inputs. The method [6] trained two independent CNN models, one for each set of the image sequences, and then combine two CNN models outputs to recognise the human activities. Similarly to [6], Jiang and Yin [7] used deep CNN to recognise human activities by converting the raw acceleration signals into signal images and providing these images as inputs.

All the human activity recognition studies mentioned above have been conducted using a single wearable device without any consideration about the device location on the user's body. But the position information of the wearable device can facilitate in improving the accuracy of the activity recognition [17]. Kunze *et al.* [18] proposed a method for classifying patterns of sensor readings to recognise the walking activity, and then analyse the characteristics of the walking motion to localise the sensor position. A drawback of this method is that changes in the sensor position cannot be detected unless the device wearer is in motion. Sztyler *et al.* [11] proposed a method for analysing the motion data obtained from several wearable devices using random forest classifier. The method also incorporated a technique to detect the position of the wearable device producing the motion signal. The method achieved high accuracy but required a costly feature engineering process. In this article, we extend our previous work in [6] by proposing a technique that trains two-stream CNN using frequency-based activity images developed from accelerometer and gyroscope motion data, to perform human activity recognition. Unlike [6], in the present work, we develop a strategy to simultaneously predict both the human activity and locations of the wearable device producing the activity signal. Moreover, while in [6] we used motion data from a single waist-mounted wearable device for the evaluation of the HAR, in this work, however, we use a much larger dataset consisting of motion data obtained from seven wearable devices positioned in seven different parts of the body including the chest, forearm, head, shin, thigh, upper-arm, and waist. Table 1 compares the main characteristics of the state-of-the-art human activity recognition methods with those of our proposed approach.



**FIGURE 1.** Illustration of the set-up for our activity image generation. We generate frequency-based activity images (right) from tri-axial accelerometer and gyroscope signals (centre). We collected synchronised motion data from the following locations (left): a- head, b- chest, c- arm, d- waist, e- wrist, f- thigh and g- shin.



**FIGURE 2.** Samples of the frequency images obtained from the tri-axial accelerometer signals for different human activities from seven different on-body locations. Note for each row, the same activity results in different motion signals from different locations.

### III. METHODOLOGY

The proposed DHALR method consists of two main parts, namely, activity image generation and classifier modelling. The tri-axial accelerometer and gyroscope signals are converted into activity images (Sec. III-A) which forms the input to our two-stream CNN classification network (Sec. III-B).

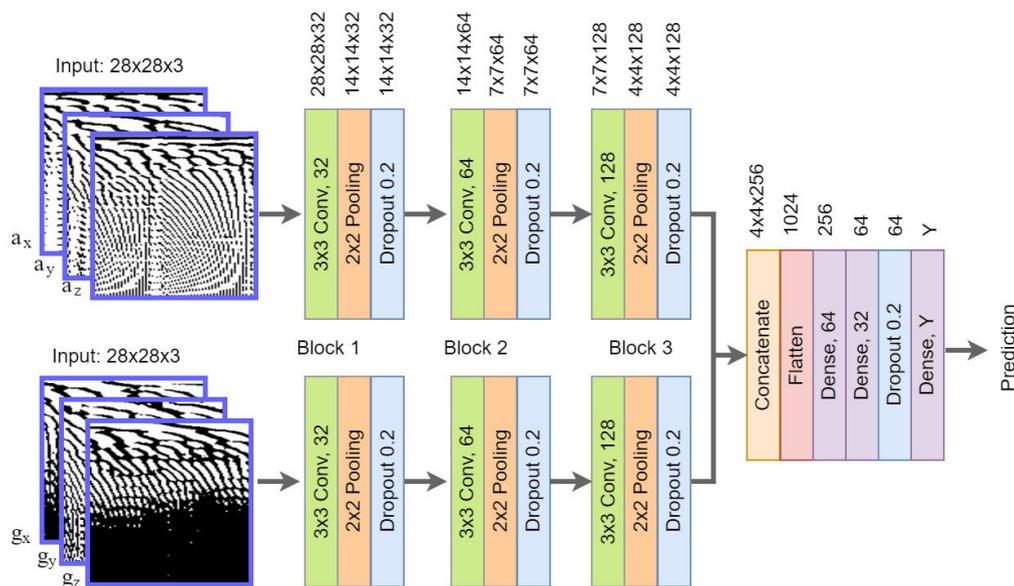
#### A. ACTIVITY IMAGE GENERATION

Frequency-based features are shown to be more effective compared to time-based features [19] for HAR. Therefore, we created frequency (activity) images from the raw tri-axial accelerometer and gyroscope signals by applying Short-time Fourier Transform (STFT) using a window size of one second with an overlap of 0.5 seconds. STFT is commonly used to determine the frequency content in local sections of a signal that continuously changes over time. We used Matlab Spectrogram function for obtaining the frequency images. A window size of one second is most effective in HAR as it can cover one cycle of most of the repetitive dynamic

activities (running, climbing, jumping and walking) [13]. A frequency image from each tri-axial signal is created by applying the STFT to each 1-dimensional signal followed by concatenating the three images to obtain a three-channel image. These are then resized to  $28 \times 28 \times 3$  to be used as input to our CNN model. The generated activity images and implementation code are published online for research purposes.<sup>1</sup> Figure 1 shows the setup for collecting the signals from the accelerometer and gyroscope and converting them to their respective frequency images.

In this article, we used the multi-sensor multi-modal human activity dataset from [11], described briefly in section IV. We used accelerometer and gyroscope data from all seven different sensor mounting locations (as shown in Fig. 1). Figure 2 shows some samples of the frequency images generated using the accelerometer signals obtained for the five dynamic activities and all sensor positions.

<sup>1</sup>Activity Images for HAR [online]. <https://isahalawal.github.io/IsahALawal/datasets/>



**FIGURE 3.** The proposed cooperative two-stream CNN architecture which takes as input both accelerometer and gyroscope frequency images and predicts the activity labels or location labels or both activity and location labels.

**B. CLASSIFIER MODELLING**

CNN is a type of deep neural network that is commonly used for analysing imaging data [20]. CNN-based methods are shown to be more robust than hand-crafted feature classification methods [21]. Unlike natural image classification problem, frequency images are low-resolution simplistic images with less natural texture information. These images encode signals and need relatively fewer convolutional layers for distinguishing between activities and locations. Therefore, we designed a simplified two-stream VGG-like [22] architecture for human activity and location recognition. The proposed DHALR network architecture is shown in Figure. 3. The network takes the two  $28 \times 28 \times 3$  dimensional tri-axial accelerometer and tri-axial gyroscope frequency images as input and predicts the activity or sensor location or both activity and sensor location  $Y$ . We use both accelerometer and gyroscope images as input because the related works on sensor-based human activity recognition [6] showed that combining the motion information from both accelerometer and gyroscope improves the recognition accuracy.

The proposed DHALR consists of three cascaded convolutional blocks, where block 1 is composed of 32,  $3 \times 3$  filters, block 2 is composed of 64,  $3 \times 3$  filters and block 3 is composed of 128,  $3 \times 3$  filters. Each convolution is followed by the  $2 \times 2$  max-pooling and dropout. The convolutional and max-pooling layers are used to learn the local spatial structure in the training images. The outputs of block 3 from the two streams are concatenated, flattened and passed through two fully connected layers, followed by a dropout and final dense layer (with softmax) equal to the number of output predictions required. The fully-connected layers help to integrate global information from across the images and to accurately

classify the human activity or/and sensor location. Dropout is a regularisation term added to avoid over-fitting during training [23]. The dropout helps to deactivate some of the nodes in the network at random during training, which helps in improving its generalisation capability. We used *ADAM* optimiser with a learning rate of 0.01 to train the network, because of its good performance in deep neural network learning [24].

**IV. DATASET DESCRIPTION**

To evaluate the performance of our proposed approach, we use the RealWorld Human Activity Recognition (RWHAR) dataset presented in [11].<sup>2</sup> Table 2 summarises the main characteristics of the original dataset. The dataset consists of motion signals from seven different body parts including chest, forearm, head, shin, thigh, upper arm, and waist (as indicated in Figure 1), that are gathered by using seven wearable devices (mainly smartphones and smartwatches) attached to the said positions. Each of the wearable devices contains six different sensors which include accelerometer, gyroscope, GPS, light, magnetometer, and audio, that were used to collect the motion signals. Fifteen people (8 male and 7 females) participated in the data collection process and each participant adorned with the seven synchronised wearable devices were instructed to perform 8 different activities which include climbing stairs down and up, jumping, lying, standing, sitting, running/jogging, and walking for approximately ten minutes (except for jumping which was performed for only 1.7 minutes due to exhaustive nature of the activity).

<sup>2</sup>RealWorld HAR dataset [online]. [http://sensor.informatik.uni-mannheim.de/#dataset\\_realworld](http://sensor.informatik.uni-mannheim.de/#dataset_realworld) [Last accessed 18.07.2020]

**TABLE 2.** Summary of the main characteristics of the RealWorld Human Activity Recognition (RWHAR) dataset from [11] which is used for analysing and validating our proposed DHALR method.

Property	Description
Number of activities	8
Number of sensors	6
Number of wearable devices	7
Number of participants	15
Duration of activities	10mins per activity
Types of wearable device	Smartphone and Smartwatches
Types of sensors	Accelerometer, Gyroscope, GPS, Light, Magnetometer, Mic
Types of activities	Climbing, Jumping, Lying, Standing, Sitting, Running, Walking

**TABLE 3.** Summary of the frequency (activity) images dataset created from RWHAR dataset. Total number of samples obtained for each activity and each sensor position for the 15 participants are reported. the images.

Activity	Position	Chest	Forearm	Head	Shin	Thigh	Upperarm	Waist	Total
		acc	15,180	15,180	15,180	15,180	15,180	15,180	15,180
Climbing up	gyro	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
	acc	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
Climbing down	gyro	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
	acc	2,520	2,520	2,520	2,520	2,520	2,520	2,520	17,640
Jumping	gyro	2,520	2,520	2,520	2,520	2,520	2,520	2,520	17,640
	acc	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
Running	gyro	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
	acc	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
Walking	gyro	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
	acc	15,180	15,180	15,180	15,180	15,180	15,180	15,180	106,260
<b>Total</b>	-	126,480	126,480	126,480	126,480	126,480	126,480	126,480	<b>885,360</b>

During the activities, the readings from both accelerometer and gyroscope sensors were sampled at 50Hz. We use the accelerometer and gyroscope data to develop frequency-based activity images for the different activities for our experiments as discussed in Sec. III-A. We generated a total of 885,360 frequency-based activity images for five dynamic activities over all the sensor positions. Table 3 shows the distribution of the activity images. For each activity, we obtain 15,180 frequency images from each sensor position and sensor type (accelerometer or gyroscope), except for jumping where we obtained 2,520 images. In total, we obtained 855,360 frequency images across all dynamics activities and sensor positions. We plan to publicly release these activity images to support benchmarking and future researches in this area.

## V. EXPERIMENTAL SETUP

We develop the proposed DHALR using Tensorflow, an open-source machine learning library produced by Google [25]. We evaluate the performance of the proposed method using a cross-subject validation approach, whereby we train the human activity recogniser with activity data obtained from 12 specific individuals in the dataset and then evaluate it with data from other 3 different people that were not present in the training set. We measure the accuracy of the DHALR on the evaluation set using precision, recall, and F1-score performance metrics. The F1-score  $\in [0, 1]$ , gives an estimation of the accuracy of the DHALR by computing the harmonic mean of the precision and recall scores. An F1-score that is close to 1 is desirable as it indicates a high recognition

performance. We adopted these metrics as they are the standard measures used for estimating the goodness of pattern recognition models [26].

We compare the performance of the DHALR against other classical CNN architectures such as LeNet5 [27] and ResNet50 [28]. We also compare the best results of the DHALR with those reported by Sztyler *et al.* [11], and Lawal and Bano [6]; other state-of-the-art methods using the same dataset. All our experiments were conducted on a PC having the following specifications: AMD Fx-8370, 8-core processor @ 4.0GHz, 32GB of RAM, Nvidia GeForce GTX1050 6GB GPU, and Microsoft Windows 10 operating system.

## VI. RESULTS AND DISCUSSION

We conducted six extensive experiments such that each experiment was designed with specific goals including finding the best sensor location for activity recognition, comparison with existing activity recognition methods and validation of the robustness of the proposed DHALR against other CNN-based methods. The specifications of each experiment are summarised in Table 4 and discussed below are the obtained results.

### A. EXPERIMENT 1: SENSOR POSITION INFERENCE INDEPENDENT OF THE ACTIVITY

This experiment aims to demonstrate the ability of the proposed DHALR to predict the correct position of the sensors based on the pattern in the activity data generated. Thus, we train the DHALR with the activity data (both

**TABLE 4.** Details of the six experiments performed for the evaluation of the proposed DHALR method. For each experiment, its goal, network input and predictions are mentioned along with the Figures and (or) Tables numbers displaying the results.

Exp.	Goal	Input	Prediction	Result
1	To recognise sensor position independent of the activity	All activity data from seven sensor positions	Labels of seven sensor positions from previously unseen activity images	Table 5, Figure 4
2	To recognise activity independent of sensor location	All activity data from seven sensor positions	Labels of five dynamic activities from previously unseen activity images	Table 6, Figure 5
3	To shows activity recognition against the sensor position	All activity data from individual sensor position	Labels of five dynamic activities from previously unseen activity images	Table 7
4	To show recognition improvement by using complementary activity data	Activity data from the two best sensor positions	Labels of five dynamic activities from previously unseen activity images	Table 8, Figure 6
5	To recognise activity and sensor position simultaneously	All activity data from seven sensor positions	Labels of thirty-four different combinations of activity and position from previously unseen activity images	Figure 7
6	To compare the robustness against other CNN design and existing methods	All activity data from the best sensor position	Labels of five dynamic activities from previously unseen activity images	Table 9, Table 10

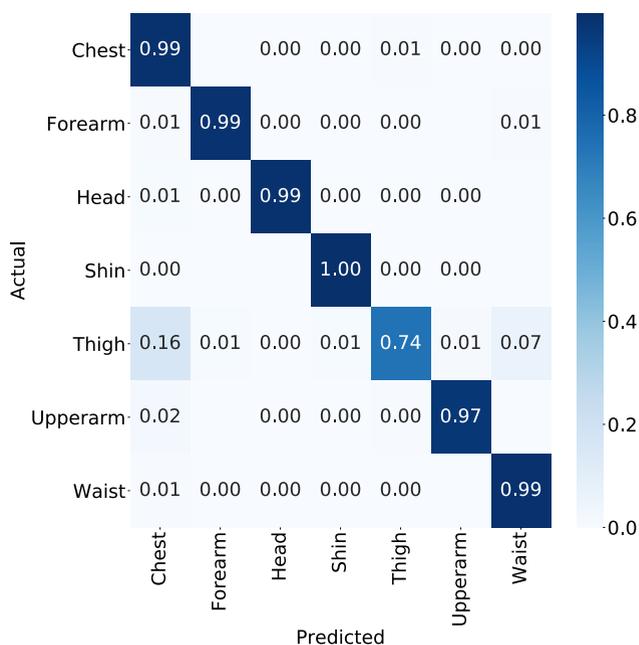
**TABLE 5.** Experiment 1: DHALR results in predicting the sensor positions independent of the activity.

Sensor position	Precision	Recall	F1-score
Chest	0.83	0.99	0.90
Forearm	0.99	0.99	0.99
Head	0.99	0.99	0.99
Shin	0.98	1.00	0.99
Thigh	0.98	0.74	0.84
Upperarm	0.99	0.97	0.98
Waist	0.93	0.99	0.96
Average	0.96	0.95	0.95

accelerometer and gyroscope) from all the seven sensor positions and evaluate it using the evaluation set. We record the performance of the DHALR in terms of the achieved precision, recall and F1-scores. Table 5 shows the results obtained by the DHALR in predicting the different positions of the sensors. The DHALR achieves an impressive F1-score of at most 0.99 for most of the sensor positions except for thigh, where it obtains an F1-score of 0.84. Figure 4 shows the confusion matrix of the DHALR prediction. The values in the diagonal indicate the accuracy of the prediction, while the values below and above the diagonal in the figure show the error incurred. Overall, these results show that the patterns of the activity data produced by the sensors positioned on the seven body parts are distinctly different and can easily be differentiated.

**B. EXPERIMENT 2: ACTIVITY INFERENCE INDEPENDENT OF THE SENSOR POSITION**

We perform this experiment to evaluate the ability of the DHALR to predict human activities independent of the position of the sensor producing the activity data. We train the DHALR with the activity data generated by all the seven wearable sensors combined, and we deploy it to predict the following activities: climbing up/down, Jumping, Running and Walking. Table 6 shows the results on

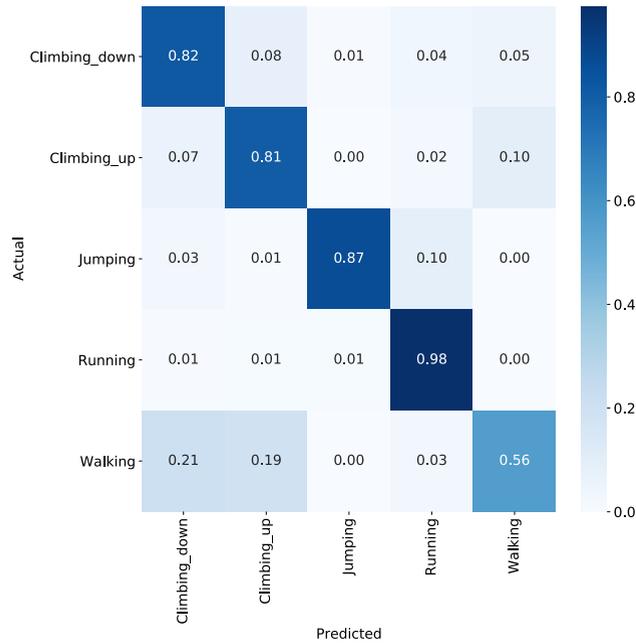


**FIGURE 4.** Experiment 1: Confusion matrix showing the DHALR sensor position prediction. The values in the diagonal indicate the accuracy (recall) of the prediction.

the evaluation set. The DHALR achieved a significantly high F1-score of 0.95 and 0.89 for running and jumping activities compared to climbing down, climbing up and walking activities that resulted in the F1-scores of 0.75, 0.71 and 0.68, respectively, irrespective of the sensor position. The confusion matrix of the results is shown in Figure 5, which highlights cases where some activities are not correctly recognised. For example, walking is erroneously confused with climbing down/up activities. This is partly because some of the participants walking and climbing gaits are very similar, thus generating similar motion signals that are difficult to differentiate. We showed in Sec. VI-E, that by using complementary activity data, the confusion between walking and climbing activities can be improved.

**TABLE 6. Experiment 2: DHALR results in predicting the activities independent of the sensor positions.**

Activity	Precision	Recall	F1-score
Climbing down	0.63	0.82	0.71
Climbing up	0.70	0.81	0.75
Jumping	0.90	0.87	0.89
Running	0.92	0.98	0.95
Walking	0.85	0.56	0.68
Average	0.80	0.81	0.80



**FIGURE 5. Experiment 2: Confusion matrix of the DHALR results using all the sensor position data.**

**C. EXPERIMENT 3: ACTIVITY RECOGNITION AGAINST SENSOR POSITION**

Next, we are interested in finding the best sensor position for accurately recognising the various activities. Therefore, we conducted this experiment to understand how the different sensor positions affect the DHALR accuracy. We train the DHALR with all the activity data from the seven sensor positions, one sensor position at a time. We evaluate and record the performance of the DHALR in predicting the activities in the evaluation set. Table 7 shows the F1-scores of the DHALR for all the activities against the different sensor positions. We observe that the DHALR performance for each activity varies across all the sensor positions, which indicates that there is no single optimal sensor position for all of the activities. However, we noted that when the sensor is positioned on the waist or shin, the DHALR perform much better for all the activities with a mean F1-score of 0.86 and 0.88, respectively. Thus we can consider the shin and Waist, as the best sensor positions for predicting dynamic activities. Also, we observe from Table 7 that the activity recognition

performance for the thigh sensor is particularly low which is in line with our findings of experiment 1 (section VI-A). We investigate the cause of this low score by viewing the videos of the data collection setup for all the participants. We discover that unlike the other six on-body devices, the one marked as the thigh is loosely placed in the front pocket of the participants’ trouser. Thus, during the execution of the physical activities, the wavering movement of the device can cause the embedded accelerometer and gyroscope sensors to generate erroneous motion signals that are different from the real signals depicting the actual activities been performed.

**D. EXPERIMENT 4: PERFORMANCE IMPROVEMENT BY COMPLIMENTARY ACTIVITY DATA**

In this experiment, we combine the activity data from the shin and waist-mounted sensors to train the DHALR. We perform this experiment to show that by using complementary activity data from the best sensor positions (discussed in section VI-C), the recognition accuracy of the DHALR can be improved. We evaluate the performance of the trained DHALR using the evaluation set and record the obtained F1-scores. The DHALR achieved an improved performance with a mean F1-score of 0.90 for all the five activities predicted. Table 8 compares the performances of the DHALR when trained with activity data from all the seven sensors independent of their position, from waist-mounted sensor only, and from both waist and shin mounted sensors. Note that in most existing human activity recognition [13], waist is considered an ideal position as it is closer to the centre of mass of the human body. We observed from Table 8 that jointly using activity data from waist and shin mounted sensors increases the recognition accuracy (F1-score) from 0.80 to 0.90; an impressive 10% improvement. Moreover, combining the activity data from both sensors also helps to provide additional discriminatory information about closely related activities such as walking and climbing, thereby aiding the DHALR to reduce the confusion between these two activities. This reduction in the confusion of the walking and climbing activities by the DHALR can be seen by comparing the improved confusion matrix in Figure 6 with that of Figure 5.

**E. EXPERIMENT 5: SIMULTANEOUS ACTIVITY AND SENSOR POSITION RECOGNITION**

We perform this experiment to evaluate the ability of the DHALR to simultaneously predict both the activity and the position of the sensor producing the activity signal. Given the five activities and seven sensor positions, the DHALR is expected to predict thirty-four different combinations of both activity and sensor positions. From a pattern recognition perspective, this is a difficult multi-label problem. Thus we train the DHALR with the activity data from all the sensor positions, whereby each of the training data is assigned two labels i.e. the activity the data is depicting and position of the sensor producing the data. We evaluate the performance of the trained DHALR using the evaluation set. The evaluation results show that the DHALR achieves a mean precision,

TABLE 7. Experiment 3: DHALR F1-scores in predicting five activities from seven sensor positions.

Activity	Chest	Forearm	Head	Shin	Thigh	Upper arm	Waist
Climbing down	0.69	0.72	0.69	0.78	0.47	0.68	0.81
Climbing up	0.67	0.72	0.73	0.88	0.59	0.74	0.86
Jumping	0.84	0.88	0.90	0.91	0.36	0.83	0.91
Running	0.94	0.95	0.94	0.97	0.83	0.94	0.91
Walking	0.65	0.68	0.64	0.84	0.35	0.56	0.81
Average	0.76	0.79	0.78	<b>0.88</b>	0.52	0.75	<b>0.86</b>

TABLE 8. Experiment 4: DHALR activity recognition results for different combinations of sensor data.

Activity data	Metric	Climbing down	Climbing up	Jumping	Running	Walking	Average
All sensors	Precision	0.63	0.70	0.90	0.92	0.85	0.80
	Recall	0.82	0.81	0.87	0.98	0.56	0.81
	F1-Score	0.71	0.75	0.86	0.95	0.68	0.80
Waist only	Precision	0.80	0.83	0.94	0.85	0.93	0.87
	Recall	0.82	0.90	0.88	0.98	0.71	0.86
	F1-score	0.81	0.86	0.91	0.91	0.81	0.86
Waist + Shin	Precision	0.91	0.86	0.94	0.88	0.92	0.90
	Recall	0.85	0.87	0.97	0.99	0.84	0.90
	F1-score	<b>0.88</b>	<b>0.87</b>	<b>0.94</b>	<b>0.93</b>	<b>0.88</b>	<b>0.90</b>

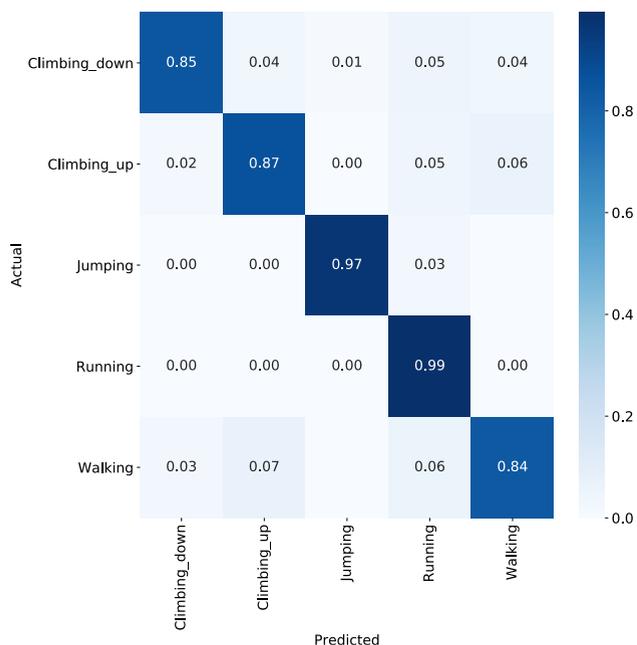


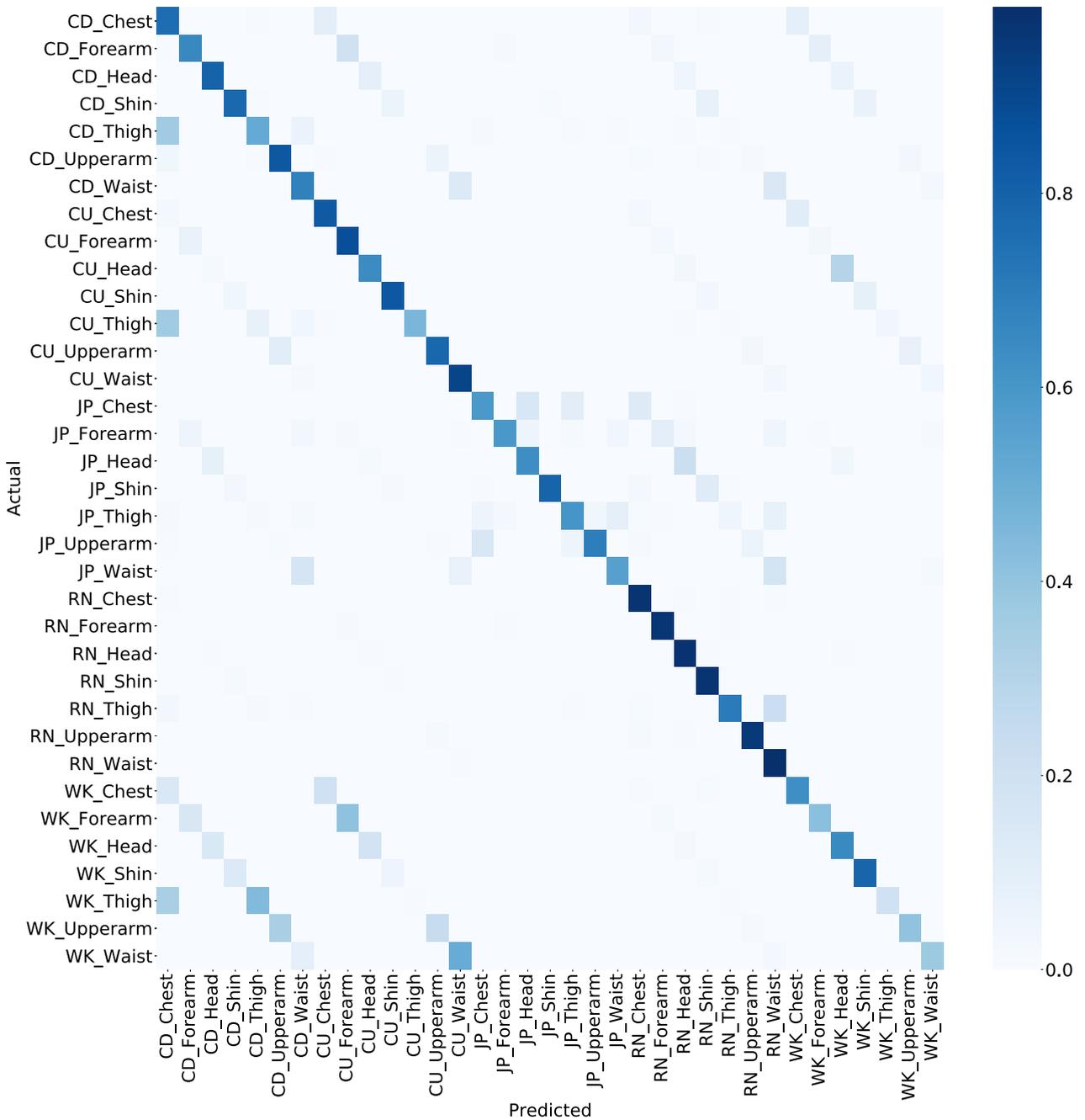
FIGURE 6. Experiment 4: Confusion matrix of the results of the DHALR trained with activity data generated from both shin and waist mounted sensors. These results show significant improvement compared to when training is done using all sensor position data as reported in Figure 5.

recall and F1-score of 0.77, 0.72 and 0.71, respectively. This is an encouraging result considering the difficult nature of the problem. Figure 7 shows the confusion matrix of the thirty-four combinations of both activity and sensor positions. The intensity of the colours in the diagonal of the confusion

matrix represents the level of the accuracy of the prediction. We can observe that the DHALR correctly predicted most of the activities with the corresponding positions of the sensors generating the activity data. We also observe some instances where the DHALR err, which include walking and climbing activities where the activity data is generated by the sensor positioned in the thigh. For example, Walking\_Thigh is wrongly classified as ClimbingDown\_Thigh. This particular case is not unexpected as we have shown in the previous experiments that the thigh is not a suitable position for activity recognition.

F. EXPERIMENT 6: ROBUSTNESS COMPARISON AGAINST OTHER CNN-BASED MODELS AND EXISTING METHODS

We performed this experiment to compare the robustness of the proposed DHALR against two classical CNN architectures commonly used for natural image classification. Specifically we implement LeNet5 [27] and ResNet50 [28], a shallow and deep CNN architectures, respectively. We train both CNNs using the same activity data from shin mounted sensor. Table 9 shows the performance of all the methods on the evaluation set. The DHALR performed much better compared to the other two methods with a mean F1-score of 0.88. The LeNet5 achieves a mean F1-score of 0.79, while the ResNet50 obtain an average F1-score of 0.73. These results support our assertion that unlike natural image classification problem, frequency images which encode activity signals can be accurately recognised using CNN architecture (like the proposed DHALR) which incorporates relatively fewer convolutional layers.



**FIGURE 7. Experiment 5: Confusion matrix of the DHALR results showing the prediction of both activity and sensor position. Note the acronyms used, e.g. CD\_Chest means climbing down activity with the sensor positioned in the chest. CD:Climbing Down, CU:Climbing Up, JP:Jumping, RN:Running, and WK:Walking.**

Finally, we also compare the performance of the DHALR with [11]: another state-of-the-art method. We chose to compare with this method because the authors reported their evaluation on the same RWHAR dataset. Table 10 shows the comparison of DHALR best results with those reported in [11] and our previous work [6], respectively. The table clearly shows that the proposed DHALR with a mean F1-score of 0.94 for position recognition and 0.90 for activity recognition achieves superior performance compared to [11] which

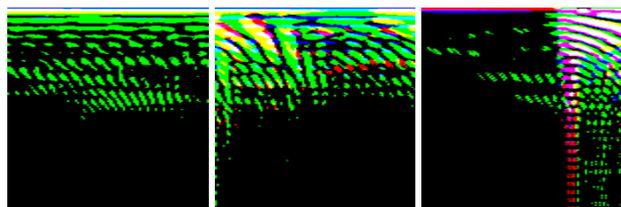
reported a mean F1-score of 0.89 and 0.87 respectively, and [6] which obtained an average F1-score of 0.87 for activity recognition. Unlike [11], the DHALR can effectively classify the various activities and device positions, due to the transformation of the raw tri-axial accelerometer and gyroscope motion readings into frequency images that encode the activity signals, and the use of a two-stream CNN classifier to capture the intrinsic similarities among the activity images.

**TABLE 9.** Experiment 6: Evaluation results of different CNN methods using activity data from shin-mounted sensor.

Method	Metric	Activity recognition
LeNet5 [27]	Precision	0.79
	Recall	0.81
	F1-score	0.79
ResNet50 [28]	Precision	0.81
	Recall	0.72
	F1-score	0.73
DHALR (proposed)	Precision	0.87
	Recall	0.88
	F1-score	<b>0.88</b>

**TABLE 10.** Experiment 6: Evaluation results of different methods for position and activity recognition.

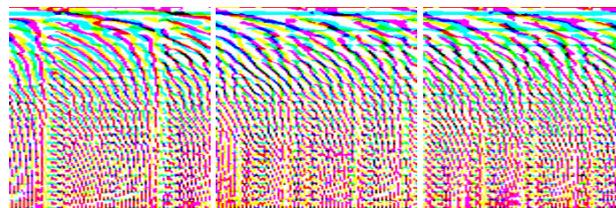
Method	Metric	Position recognition	Activity recognition
Sztylet et al. [11]	Precision	0.89	0.88
	Recall	0.89	0.86
	F1-score	0.89	0.87
Lawal and Bano [6]	Precision	-	0.88
	Recall	-	0.87
	F1-score	-	0.87
DHALR (proposed)	Precision	0.95	0.90
	Recall	0.94	0.90
	F1-score	<b>0.94</b>	<b>0.90</b>

**FIGURE 8.** Samples of the corrupted activity data for climbing down, climbing up and walking generated from thigh-mounted sensor.

### G. SOURCES OF RECOGNITION ERROR

We observed that some activities are more difficult to differentiate and/or recognise, due to the similarities in the manner the activities are performed. Specifically, walking and climbing up/down are often confused in our experiments. We studied the instances where some of these problems occur and examine the affected activity data in the evaluation set. The prediction errors can be attributed partly to the following reasons

- Errors due to corrupted activity data. Figure 8 shows samples of corrupted activity data which the proposed DHALR misrecognised. Such samples occur at the start or/and end of an activity. These erroneous data are difficult for the DHALR because they are corrupted. The solution could be to discard them from both the training and evaluation set accordingly since they lack information that will improve the recogniser.
- Errors due to the similarity in the manner closely related activities are performed. Figure 9 shows samples of the activity data for walking and climbing activities with very similar movement patterns. This type of error can

**FIGURE 9.** Samples of the activity data for climbing down, climbing up and walking showing very similar motion patterns.

be mitigated by using additional motion information from a complementary sensor as input during the training of the recogniser as validated in Section VI-D.

### VII. CONCLUSION

We proposed a novel method for human activity and sensor location recognition by proposing a two-stream convolutional neural network. We used frequency-based activity images from both accelerometer and gyroscope sensors mounted on several body locations as input to our network. The network jointly encoded both accelerometer and gyroscope frequency images, concatenated the two feature maps and predicted either the activity or location or both activity and location. We evaluate the performance of the proposed method using real-world human activity dataset, and the experimental results show that the proposed DHALR is robust compared to other activity recognition methods and CNN-based networks (commonly used in natural image classification). Unlike existing HAR methods, which mainly rely on single (waist) sensor information for activity inference, we showed that the shin position is more accurate than the waist. Moreover, combining complementary information from both waist and shin data helped in further improving the activity recognition accuracy.

### REFERENCES

- [1] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, 2015, pp. 127–140.
- [2] S. Choi, "Understanding people with human activities and social interactions for human-centered computing," *Hum.-Centric Comput. Inf. Sci.*, vol. 6, no. 1, pp. 1–9, Dec. 2016.
- [3] A. Malaisé, P. Maurice, F. Colas, F. Charpillet, and S. Ivaldi, "Activity recognition with multiple wearable sensors for industrial applications," in *Proc. 11th Int. Conf. Adv. Comput.-Hum. Interact.*, 2018, pp. 1–7.
- [4] P. Kumari, L. Mathew, and P. Syal, "Increasing trend of wearables and multimodal interface for human activity monitoring: A review," *Biosensors Bioelectron.*, vol. 90, pp. 298–307, Apr. 2017.
- [5] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8553–8562, Oct. 2019.
- [6] I. A. Lawal and S. Bano, "Deep human activity recognition using wearable sensors," in *Proc. 12th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Jun. 2019, pp. 45–48.
- [7] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015, pp. 1307–1310.

- [8] A. Jordao, A. C. Nazare, J. S. de Souza, and W. R. Schwartz, "Human activity recognition based on wearable sensor data: A standardisation of the state-of-the-art," *CoRR*, 2018. [Online]. Available: <http://arxiv.org/abs/1806.05226>
- [9] G. Ogbuabor and R. La, "Human activity recognition for healthcare using smartphones," in *Proc. 10th Int. Conf. Mach. Learn. Comput.*, 2018, pp. 41–46.
- [10] I. A. Lawal, F. Poiesi, D. Anguita, and A. Cavallaro, "Support vector motion clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2395–2408, Nov. 2017.
- [11] T. Sztyley, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervasive Mobile Comput.*, vol. 38, pp. 281–295, Jul. 2017.
- [12] O. S. Eyobu and D. Han, "Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," *Sensors*, vol. 18, no. 9, p. 2892, Aug. 2018.
- [13] J. L. R. Ortiz, *Smartphone-Based Human Activity Recognition* (Springer Theses). Cham, Switzerland: Springer, 2015.
- [14] C. Catal, S. Tufekci, E. Pirmitt, and G. Kocabag, "On the use of ensemble of classifiers for accelerometer-based activity recognition," *Appl. Soft Comput.*, vol. 37, pp. 1018–1022, Dec. 2015.
- [15] M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artif. Life Robot.*, vol. 23, no. 2, pp. 173–185, Jun. 2018.
- [16] N. Nair, C. Thomas, and D. B. Jayagopi, "Human activity recognition using temporal convolutional network," in *Proc. 5th Int. Workshop Sensor-Based Activity Recognit. Interact.*, Sep. 2018, pp. 1–8.
- [17] D. Coskun, O. D. Incel, and A. Ozgovde, "Phone position/placement detection using accelerometer: Impact on activity recognition," in *Proc. IEEE 10th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Apr. 2015, pp. 1–6.
- [18] K. Kunze, P. Lukowicz, H. Junker, and G. Tröster, "Where am i: Recognizing on-body positions of wearable sensors," in *Location- and Context-Awareness*, T. Strang and C. Linnhoff-Popien, Eds. Berlin, Germany: Springer, 2005, pp. 264–275.
- [19] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 871–879, Mar. 2009.
- [20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [21] L. Nanni, S. Ghidoni, and S. Brahmam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognit.*, vol. 71, pp. 158–172, Nov. 2017.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
- [23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [24] J. B. Diederik and P. Kingma, "Adam: A method for stochastic optimisation," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–13.
- [25] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. USENIX Conf. Operating Syst. Design Implement.*, 2016, pp. 265–283.
- [26] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, early access, Aug. 21, 2018, doi: [10.1016/j.aci.2018.08.003](https://doi.org/10.1016/j.aci.2018.08.003).
- [27] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>



**ISAH A. LAWAL** received the bachelor's degree in electrical electronic engineering from Abubakar Tafawa Balewa University (ATBU), Bauchi, Nigeria, the master's degree in computer engineering from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, and the joint Ph.D. degree in interactive and cognitive environments from the Queen Mary University of London, U.K., under the Erasmus Mundus Fellowship, and the University of Genoa, Italy,

in 2015. After, he worked as a Research Assistant with KFUPM. He is currently an Associate Professor and the Program Leader of applied data science with the Noroff School of Technology and Digital Media Kristiansand, Norway. Previously, he worked as an Assistant Professor with Jouf University, Saudi Arabia, where he taught artificial intelligence and smart system courses. His research interests include incremental learning methods for the characterisation of sequential data and deep learning techniques for human activity recognition.



**SOPHIA BANO** received the bachelor's degree in mechatronics engineering and the master's degree in electrical engineering degree from the National University of Sciences and Technology (NUST), Pakistan, the M.Sc. degree from the Computer Vision and Robotics (VIBOT) Programme run jointly by Heriot-Watt University, U.K., the University of Girona, Spain, and the University of Burgundy, France, under the Erasmus Mundus Scholarship, and the joint Ph.D. degree in interactive and cognitive environments under the Erasmus Mundus Fellowship,

from the Queen Mary University of London, U.K., and the Technical University of Catalonia, Spain, in 2015. She did an internship at Imperial College London, U.K. She is currently a Research Scientist with the Surgical Robot Vision Group, Wellcome / EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London. Previously, she worked as a Postdoctoral Research Assistant with the University of Dundee, under the U.K. EPSRC project titled augmenting communication using environmental data to drive language prediction (ACE-LP). Her research interests include devising computer vision and deep learning techniques for multi-modal sequential data analysis.

• • •