Deep Human Activity Recognition Using Wearable Sensors

Isah A. Lawal
Department of Applied Data Science
Noroff University College, Kristiansand, Norway
Isah.Lawal@noroff.no

Sophia Bano
Wellcome/EPSRC Centre for Interventional and Surgical
Sciences, Department of Computer Science,
University College London, United Kingdom

ABSTRACT

This paper addresses the problem of classifying motion signals acquired via wearable sensors for the recognition of human activity. Automatic and accurate classification of motion signals is important in facilitating the development of an effective automated health monitoring system for the elderlies. Thus, we gathered hip motion signals from two different waist mounted sensors and for each individual sensor, we converted the motion signal into spectral image sequence. We use these images as inputs to independently train two Convolutional Neural Networks (CNN), one for each of the generated image sequences from the two sensors. The outputs of the trained CNNs are then fused together to predict the final class of the human activity. We evaluate the performance of the proposed method using the cross-subjects testing approach. Our method achieves recognition accuracy (F1 score) of 0.87 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.

CCS CONCEPTS

• **Computing methodologies** → **Activity recognition and understanding**; **Neural networks**; **Ensemble methods**.

KEYWORDS

Human Activity Recognition, Convolutional neural network, Ensemble method

1 INTRODUCTION

Human-centred monitoring systems are devices equipped with low-powered and low-cost sensors which are mounted on the human (e.g. wearable) or outside (e.g. cameras), for the collection of useful data for human activities and behaviour analysis [12]. Often these devices come with a wireless unit for data transmission to external storage and/or analysis unit. The common area of application of

these systems includes home monitoring for elderly or patients in medicine [6], among others. In this paper, we focus on the analysis of the signals obtained via waist-mounted wearable sensors as this is commonly used to monitor the health status of older patients with movement assistive devices [1]. These sensors usually generate complex hips motion signals which are difficult to interpret without expert intervention. To automatically interpret the sensor readings in order to infer the kind of human activity that is been performed by a user, a computationally efficient modelling technique that will provide a meaningful characterisation of the sensor data is required.

Recently, machine learning methods including Support Vector Machines (SVM) [7] and Random Forest [11], have been proposed for the characterisation and automated interpretation of the sensor data. However, these methods usually require a hand-crafted features extraction by an expert which often affects the recognition accuracy [12]. Convolutional Neural Networks (CNN) [4] on the other hand, is a form of deep neural network learning paradigm that has the ability to learn complex patterns in data (including images) without the need for prior feature extraction by an expert [12]. Inspired by the success of CNN in related applications, we explore deep CNN as an alternative approach for recognising human activities including climbing jumping, lying, running, sitting, standing and walking using activity images generated from the signals obtained via waist-mounted sensor devices. For brevity, we will refer to our proposed approach as Deep Human Activity Recognition (DHAR) henceforth.

2 RELATED WORK

The presences of sensors such as accelerometer and gyroscope in many wearable devices have made it possible to collect body parts motion data and to recognise various activities performed by humans [11]. The accelerometer measures the physical acceleration of movable body parts, while the gyroscope is used to measure their orientations. The signals from both sensors are commonly used for Human Activity Recognition (HAR) to differentiate among very similar human activities [6]. Different classifiers have been proposed for the analysis of the motion signals and classification of the human activities. Ortiz Jorge [7] proposed the use of a waist-mounted smartphone for HAR. The method employed features obtained by computing the mean, correlation, and frequency skewness of the raw signals generated by the accelerometer and gyroscope embedded in the smartphone. These features were used to train an SVM to classify six human activities. Although the method was able to accurately classify dynamic activities (e.g. walking and climbing), some static actions were misclassified. Sztyler et al. [11] proposed a method for HAR whereby the position of the wearable devices on the human body can change depending on user preference. The method combine frequency and gravity based features and used random forest classifier in order to determine the orientation of
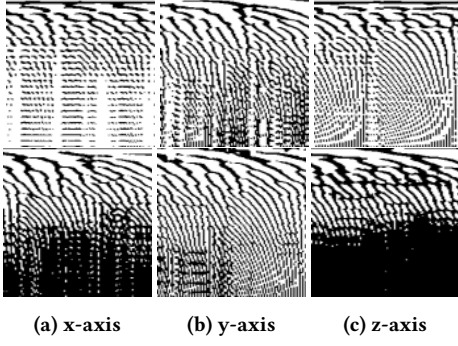
(a) x-axis   (b) y-axis   (c) z-axis

Figure 1: Samples of the activity images generated from tri-axial accelerometer (top) and gyroscope (bottom) signals.

Figure 2: Block diagram of the proposed DHAR. Sensor #1: Accelerometer and Sensor #2: Gyroscope

the device and to identify the human activity. Nair et al. [5] on the other hand, employed a temporal CNN to directly classify human activities on smartphone. The advantage of this method is that it takes as input the raw sensor data without the need for the costly pre-processing phase, which in effect, reduced the learning time and improved the classification performance. Similarly to the Nair et al. [5] method, our proposed method (DHAR) employs CNN as a classifier for the activity classification but in contrast, we used activity image sequences that are generated from the accelerometer and gyroscope signals as inputs. We trained two independent CNN, one for each of the image sequence from the two sensors and combined their outputs to finally classify the human activities.

3 METHODOLOGY

3.1 Frequency Image Generation

For the DHAR modelling, we used the multi-sensor human activity dataset described in section 4.1. We used only the data from the waist mounted sensors as this is the most widely used position in HAR [7] and besides, our current aim is to show the effectiveness of deep CNN learning for HAR. Frequency-based features are shown to be more effective compared to time-based features [8]. Hence, we apply 512-points Fast Fourier Transform (FFT) to each signal of the tri-axial accelerometer and gyroscope readings using a windows size of 1 second with an overlap of 0.5 seconds. Window size of 1 second have been shown to be most effective in HAR as it can cover one cycle of most of the repetitive dynamic activities (running, climbing, jumping and walking) [7]. We then convert the results of the FFT to 28 x 28 images, one for each axis of the tri-axial accelerometer and gyroscope signals, respectively. Figure 1 shows samples of the images for climbing down activity. We used these image sequences as inputs to train the proposed DHAR model.

3.2 DHAR Modelling

Our proposed DHAR architecture is inspired by VGG-like network [9] except that for DHAR relatively fewer convolutional layers are enough for distinguishing different activities. We independently train two-set of five-layered CNN, using the activity images generated previously from the accelerometer and gyroscope signals. We use this approach because of the expectation that the resulting diversity in the decision-making process of the two CNN would
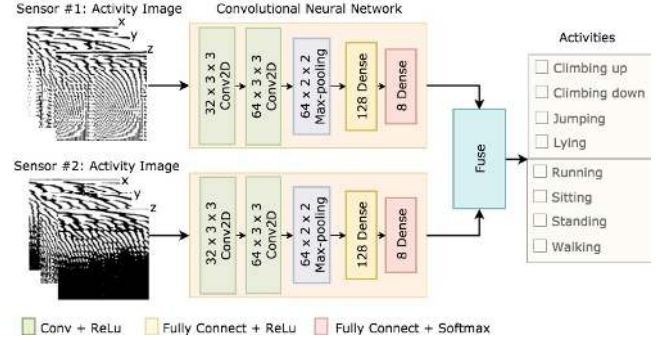
boost the generalisation performance of the DHAR and improve the reliability of the activity classification. Each CNN model consists of two convolutional layers, composed of 32, 3 x 3 and 64, 3 x 3 filters and Rectified Linear Unit (ReLU) activation function, followed by a max-pooling layer composed of a 2 x 2 filter and two fully-connected layers of 128 and 8 neurons (as shown in Figure 2). Inputs to the network are the two sets of activity images fed independently to the two CNNs and output are the 8 classes of the human activities in the training data. While the convolutional and max-pooling layers are used to learn the local spatial structure in the training images, the fully-connected layers, on the other hand, help to integrate global information from across the images and to accurately classify the human activity. The continuous outputs of the two CNN models are combined (fuse) together using soft voting technique and the class with the highest vote is chosen as the final output of the DHAR. In order to reduce the risk of overfitting, we used dropout during the training process [10]. The dropout helps to deactivate some of the neurons in the CNN at random during training, thereby improving the generalisation ability of the CNN. We used *ADAM* optimiser to train the CNN, because of its good performance in deep neural network learning [2].

4 EXPERIMENTS AND DISCUSSION

4.1 Dataset

We use the publicly available RealWorld Human Activity Recognition (RWHAR) dataset [11][1] for the experimentation. The dataset is developed from motion signals acquired from six sensors (accelerometer, gyroscope, GPS, light, magnetometer, and audio) at seven body positions (chest, forearm, head, shin, thigh, upper arm, and waist) by attaching seven devices (smartphones and smartwatches) at the said locations. These devices were synchronised using the network provider time service and sensors like accelerometer and gyroscope were sampled at $50Hz$. Fifteen individuals were involved in the data gathering study and each individual performed eight activities (climbing stairs down and up, jumping, lying, standing, sitting, running/jogging, and walking). Each activity was performed approximately for 10 minutes except for jumping that was performed for approximately 1.7 minutes. Out of the 15 subjects

---

[1]RealWorld HAR dataset [online]. http://sensor.informatik.uni-mannheim.de/#dataset_realworld [Last accessed 08.03.2019]
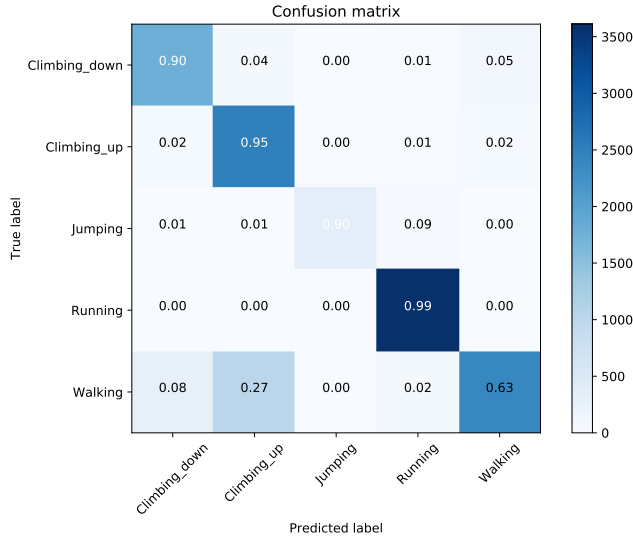
Figure 3: Confusion matrix of the recognised dynamic activities, showing cases where the DHAR misclassified some of the activities.

involved, 8 were male and 7 were female. Further details about this dataset can be found in [11].

4.2 Experimental Setup

To facilitate the efficient development and training of the DHAR model, we used an open source machine learning library produced by Google known as *TensorFlow*. We evaluate the proposed DHAR using the dataset described in section 4.1. We used the activity data consisting of 201328 image sequences from 12 individuals (person #1 to #12) for training and reserved the remaining data consisting of 47872 activity images from the other 3 different individuals (i.e. person #13 to #15) for testing, in order to evaluate the robustness of the proposed DHAR on cross-subjects test cases. During the experiments, we used cross-validation to select the optimal learning rate and batch size for the CNN and sets the dropout to 0.5. We assessed the performance of the DHAR on the test data using the following metrics: precision, recall and F1 score [3]. While the precision score measures the exactness of the DHAR results, the recall, on the other hand, gives an indication of the completeness of the DHAR outputs. The F1 score estimates the accuracy of the DHAR by computing the harmonic mean of the precision and recall scores [3].

4.3 Discussion

The experimental results for 5 dynamic activities (i.e. climbing down, climbing up, jumping, running and walking) are presented in Table 1. The table shows that the DHAR achieves an impressive F1 score of 0.93 and 0.98 for jumping and running activities, respectively, and a low score of 0.76 for the walking activity. Overall, the DHAR achieves F1 score of 0.87 on average for the cross-subjects test cases. This score is superior to the 0.78 reported by Sztyler

Table 1: Performance of the proposed DHAR on the cross-subjects test data for the five dynamic activities.

| Activity	Precision Score	Recall Score	F1 Score
Climbing down	0.83	0.90	0.86
Climbing up	0.69	0.95	0.80
Jumping	0.96	0.90	0.93
Running	0.96	0.99	0.98
Walking	0.94	0.63	0.76
Average	**0.88**	**0.87**	**0.87**

et al. [11] which employs information from several on-body sensors to recognise the same dynamic activities using the same dataset. Figure 3 shows the confusion matrix of the results, highlighting those dynamic activities that are misclassified by the DHAR. We observed that in some cases, walking is confused with climbing activity. We investigate the cause of this confusion by viewing the videos of the activities performed by the 3 persons (person #13-#15) in the test set. We discover that the DHAR erroneously confused walking with climbing activities, because some people in the test cases perform both activities in a very similar way, thus generating the same patterns of motion signals which are difficult to differentiate. Specifically, we observed in the activity video that person #15 performed the walking and climbing tasks in a very controlled and robot-like manner, differently from the realistic way person #13 and #14 performed the same activities in the video. Figure 4 shows the confusion matrix of the results obtained with the data from person #15 only. The figure shows how the walking and climbing activities are significantly overlapped, thus affecting the overall results shown in Table 1. In the future, this problem can be addressed by using another Inertial Measurement Unit (IMU) on a different body location in order to gather complementary motion signal with additional discriminatory information that can be used to differentiate between similar and closely related human activities.

We also investigate and compare the performance of the DHAR trained separately with only the data from an accelerometer, gyroscope and the combination of both, respectively. Figure 5 shows the mean F1 scores for the DHAR on the three cases investigated. As expected, the DHAR trained with the combined data from the two different sensors (as shown in Figure 2) achieves the highest F1 score of 0.870 ±0.090, followed by the DHAR trained with data from the accelerometer only, which obtained a score of 0.850 ±0.092. The lowest F1 score of 0.721 ±0.150 was achieved by the DHAR trained with only the data from the gyroscope. These results show that though gyroscope can provide information about the orientation of a movable body part and accelerometer can measure the body translation, the combination of data from both sensors is required to accurately recognise dynamic activities.

Table 2 shows the experimental results for all the 8 activities including the static and dynamic activities combined. The DHAR achieved the overall F1 score of 0.78 on the cross-subjects test cases. The table also shows that lying and running activities have the highest F1 scores of 0.99 and 0.93, respectively, while sitting and standing have the lowest scores of 0.47 and 0.55, respectively. The low F1 scores for sitting and standing activities could be attributed partly to the window size used during the FFT transformation of the
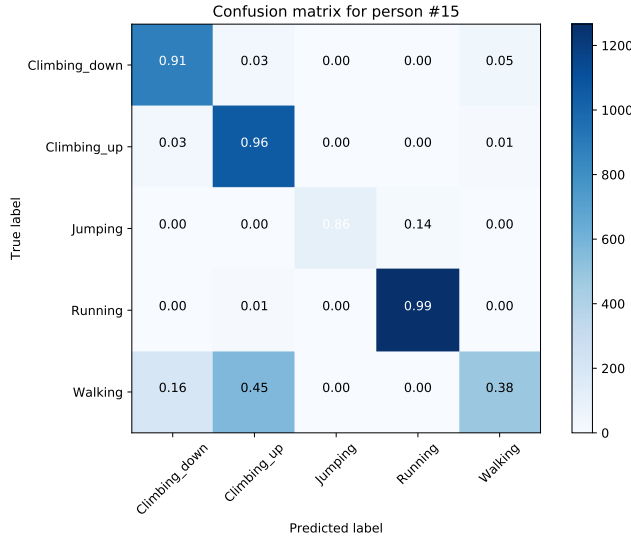
Figure 4: Confusion matrix of the data from person #15 showing overlapping of walking and climbing activities. This particular individual performed the walking and climbing activities in a very similar and controlled manner differently from other participants in the same video, hence affecting the overall performance.
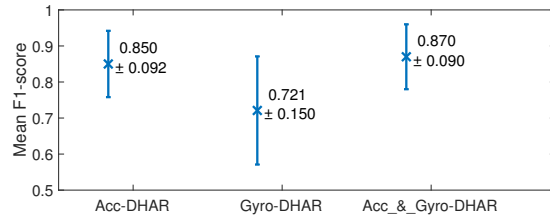
Figure 5: Comparison of the performance of the DHAR trained separately with data from only an accelerometer (Acc-DHAR), gyroscope (Gyro-DHAR) and the combination of both (Acc_&_Gyro-DHAR).

raw sensor signals to activity images. Although a window size of 1 second is enough to cover one cycle of repetitive dynamic activities such as running, however, for static activities such as sitting and standing, a longer window period might be required in order to capture useful information representing the postural transitions between the activities that will help to differentiate them apart [7]. Future work will investigate the effect of different window sizes on the performance of DHAR in terms of accuracy and computational cost, and the use of recurrent neural networks that can capture temporal aspect as well.

5 CONCLUSION

In this paper, we showed the effectiveness of deep CNN for cross-subjects human activity recognition using wearable sensors. We achieved this by transforming raw hip motion signals obtained from waist-mounted accelerometer and gyroscope sensors to activity

Table 2: Performance of the proposed DHAR on all 8 human activities including the static and dynamic activities.

Activity	Precision Score	Recall Score	F1 Score
Climbing down	0.79	0.90	0.84
Climbing up	0.71	0.95	0.81
Jumping	0.88	0.87	0.88
Lying	0.98	0.99	0.99
Running	0.89	0.98	0.93
Sitting	0.56	0.41	0.47
Standing	0.50	0.60	0.55
Walking	0.95	0.65	0.77
Average	**0.78**	**0.79**	**0.78**

images, and using these images to independently train two sets of CNN. The outputs of the two CNN are ensemble together to predict different types of human activities including climbing down, climbing up, jumping, lying, running, sitting, standing and walking. We validated the performance of the proposed DHAR method on publicly available real-world HAR dataset and showed that the DHAR achieves an overall F1 score of 0.87 for dynamic activities and 0.78 for both static and dynamic activities. This performance is comparable to another state-of-the-art method which uses multiple on-body sensors for motion information retrieval and hand-crafted features for HAR. In addition to the applications demonstrated in this paper, the proposed DHAR could be used for other non-invasive human-centred monitoring systems that use cameras to directly capture images depicting human activities.

REFERENCES

[1] Aimée K. Bright and Lynne Coventry. 2013. Assistive Technology for Older Adults: Psychological and Socio-emotional Design Requirements. In *Proc. of the International Conference on Pervasive Technologies Related to Assistive Environments*. 1–4.
[2] Jimmy Ba Diederik P. Kingma. 2015. Adam: A Method for Stochastic Optimisation. In *Proc. of the International Conference on Learning Representations*. 1–13.
[3] Cyril Goutte and Eric Gaussier. 2005. A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation. In *Proc. of the European Conference on Advances in Information Retrieval Research*. 345–359.
[4] Andrey Ignatov. 2018. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing* 62 (2018), 915 – 922.
[5] Nitin Nair, Chinchu Thomas, and Dinesh Babu Jayagopi. 2018. Human Activity Recognition Using Temporal Convolutional Network. In *Proc. of the International Workshop on Sensor-based Activity Recognition and Interaction*. 1–8.
[6] Godwin Ogbuabor and Robert La. 2018. Human Activity Recognition for Healthcare Using Smartphones. In *Proc. of the International Conference on Machine Learning and Computing*. 41–46.
[7] Jorge Luis Reyes Ortiz. 2015. *Smartphone-Based Human Activity Recognition*. Springer International Publishing.
[8] Stephen J. Preece, John Y. Goulermas, Laurence P. J. Kenney, and David Howard. 2009. A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data. *IEEE Transactions on Biomedical Engineering* 56, 3 (2009), 871–879.
[9] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proc. of the International Conference on Learning Representations*.
[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15, 1 (2014), 1929–1958.
[11] Timo Sztyler, Heiner Stuckenschmidt, and Wolfgang Petrich. 2017. Position-aware activity recognition with wearable devices. *Pervasive and Mobile Computing* 38, 2 (2017), 281 – 295.
[12] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2018. Deep learning for sensor-based activity recognition: A Survey. *Pattern Recognition Letters* (2018).