Artificial Neural Network for Human Activity Recognition by Use of Smart Insoles

Luigi D'Arco[a], Haiying Wang[a], Graham McCalmont[a], XianQi Lan[b] and Huiru Zheng[a]

[a]*School of Computing, Ulster University, Belfast, Antrim, UK*
[b]*Beitto Ltd., Fuzhou, China*
[c]*Corresponding author*

Abstract

Human Activity Recognition (HAR) is an area with high interest. It can be used, for example, for medical purposes, for rehabilitation, for monitoring in sports as well as for prevention and monitoring of the elderly. The most used devices for this purpose are wearable devices. These devices have small sizes and can integrate different sensors inside. The only problem with these devices is that when using multiple devices at the same time, it can create an annoyance for the user. In addition, processing and understanding the data provided by such devices are sometimes not immediate and there is no standard rule that can be followed. In this study, a solution to HAR is presented by integrating a pair of smart insoles as a non-hindering device for the user and as a secondary purpose, there is the creation of a pipeline that can be simply recreated and extended to multiple subjects as needed. Smart insoles integrate pressure and inertia sensors inside. Alongside this device, an Artificial Neural Network model is developed to autonomously extract salient information directly from the raw data. Three subjects were included in the study. Each of them completed a series of activities among a well-defined set (fast walking, normal walking, slow walking, sitting, standing, downstairs, upstairs, and sit to stand). The results obtained with this solution achieved an average accuracy of 99.47%.

Keywords

Artificial Neural Network, Human Activity Recognition, Smart insole

1. Introduction

The Human Activity Recognition (HAR) has been a theme that has always fascinated the scientific community, which has spent many years looking for solutions suitable for everyday use. One of the objectives of HAR is to reveal information about a user's behaviour so that computing systems can help them with their daily tasks more effectively. Initially, the leading technology involved for HAR was computer vision. Computer vision made it possible, through the use of capture devices or video devices, to analyse the fundamental patterns and gestures of some actions. Different works can be found in literature [1, 2, 3], however, they all share the constraint of being prepared for a given environment, be it small or large, and cannot be easily adapted to different scenarios. Returning to the main goal of integrating this technology into

CERC 2021: Collaborative European Research Conference, September 09–10, 2021, Cork, Ireland
darco-l@ulster.ac.uk (L. D'Arco); hy.wang@ulster.ac.uk (H. Wang); mccalmont-g@ulster.ac.uk (G. McCalmont); allan@beitto.com (X. Lan); h.zheng@ulster.ac.uk (H. Zheng)
0000-0001-7179-8281 (L. D'Arco); 0000-0001-8358-9065 (H. Wang); 0000-0001-7648-8709 (H. Zheng)

everyday life settings, efforts to find new solutions have been shifted from computer vision to systems that could be worn by users (wearable systems). Wearable systems allow not only to be integrated into multiple environments without having to worry about external constraints but at the same time reduce the overall dimensions and work continuity factors over time.

The study of HAR using wearable devices has made it possible to extend the possible applications of HAR to scenarios that may previously have been unthinkable [4] and by the spread of ubiquitous devices, the integration of HAR in daily life has been greatly simplified so that today we can find several studies in the literature that address it [5, 6, 7]. HAR has been found to be useful in a variety of real-world situations, such as rehabilitation [8], sport monitoring [9] as well as prevention for the elderly [10]. Despite the numerous applications and benefits of wearable systems, there is always the area covered by these devices that is too large for pervasive use. In addition to the aforementioned problem, it must be considered that the processing of sensor's data is not as immediate as images video because they are difficult to understand for the human and there is no technique that could be applied for all the sensor's data. Indeed, the majority of the studies use feature extraction techniques or even statistical analysis to make the data more readable and meaningful before processing. With the growth of Machine Learning and Neural Networks, the processing of data is becoming more simple and it does not require so depth knowledge of mathematics to process that data. Neural Networks provides an intrinsic function of auto discovering of the most important features providing the scientist with a way to process directly the raw data.

The purpose of this work is to develop a HAR system that can be used by users without any restrictions, as well as to establish a pipeline that can be easily replicated and expanded to include as many subjects as desired. A smart insole is employed as a device for the HAR to achieve this goal because it decreases the user's encumbrance as well as the work necessary for installation in a new settlement. Pressure sensors and inertia sensors have been integrated into the smart insole. An Artificial Neural Network (ANN) is the engine that recognises the activities and allows raw data to be processed directly.

The rest of the paper is structured as follow: in Section 2 the existing solutions are analysed, the study details and the approach chosen are presented in Section 3 followed by the results obtained and the discussion in Section 4, and in Section 5 the overall paper will be summarised and the future work highlighted.

2. Related Work

Using smart insoles for HAR is certainly not new, and multiple studies can be found in the literature which can be divided into two main categories according to the way the HAR is carried out: threshold algorithms, machine learning algorithms.

Threshold algorithms base their functioning on the study of the data collected by the subjects and on the basis of observations on the change of such data in relation to the activity carried out, minimum thresholds are defined which, if exceeded, indicate the activity performed. Moufawad el Achkar et al. [11] introduce instrumented shoes that can collect movement and foot loading data unobtrusively during daily living. The instrumented shoes include a 3D accelerometer, a 3D gyroscope, a 3D magnetometer, a barometric sensor and a force sensing

insole comprised of 8 pressure cells. Ten elderly subjects were recruited for data collection, and asked to complete a predefined track to mimic physical activities of daily life (level walking, sit-to-stand, sitting, standing, uphill, downhill, upstairs, downstairs and elevator use). The data were sampled at 200 Hz, then segmented at 5 seconds with 2.5 seconds overlap. The classification relies on an expert-based hypothesis, the foot loading, orientation, and elevation can be used to classify postural transitions, locomotion, and walking type. The resulting classifier has sensitivity and precision for sitting, standing, and walking that are greater than 95%. Stair ascending had the lowest precision (89%) and sensitivity (79%) whereas elevator up and down had the lowest precision (89%) and sensitivity (79%) respectively (78%). The overall accuracy rate was 97.41%.

Machine learning algorithms try to imitate human learning and consequently process data incrementally, improving its accuracy during execution through the use of statistical methods uncovering key insights within data. De Pinho et al. [12] propose a machine learning HAR classifier based on a foot-based wearable device. The wearable device in question comprises two components: a plantar with pressure sensors and a microcontroller equipped with a 3D accelerometer, 3D gyroscope, 3D magnetometer and a barometric sensor. Eleven volunteers participated in the experiment from which 2 hours of feet posture and movement data were gathered. The activities that the classifier identifies are: walking straight, walking slope up and down, ascending and descending stairs and sitting. The data collected were segmented at 0.3 seconds and instead of processing sensors raw data, the features were extracted, in particular the standard deviation, variance, minimum, maximum and average values are used. Random Forest was used as classified and validated by the means of a Leave-one-out Cross-Validation strategy. The average accuracy of the classifier was 93.34%.

Identifying the most important features to extract requires in-depth knowledge of the dataset and requires several statistical studies on the data so that the identified features can best represent the data collection. To overcome these problems, several studies use deep learning techniques instead of machine learning as they differ from the latter for their intrinsic ability to filter data and then select only the most important information, thus avoiding the feature extraction phase. Pham et al. [13] propose a deep convolutional neural network (CNN) for detecting a set of low-level activities including running, walking, standing, jumping, kicking, cycling. They used a 3D accelerometer sensor integrated into a pair of shoes. The data were sampled at a frequency of 50 Hz and segmented with a sliding window approach of 2 seconds and 50% overlap between two consecutive windows. Ten subjects were included in the study which were requested to perform each activity from 10 to 30 minutes. The CNN was evaluated using a 10 Fold Cross Validation reaching relatively high performances, on average 93.41% precision and 93.16% recall. Zhang et al. [14] design a HAR system based on an insole-based IMU sensor driven by a subject-independent algorithm. The IMU sensors involved are a 3D acceleration and a 3D gyroscope, and the data are sampled at a frequency of 100 Hz. In the data preprocessing section, raw IMU sensor data is separated into two parts: one part is manually labelled with activity labels, while the other half is left unlabelled, The data are segmented in slices of 2 seconds and then the subject-specific information are removed from them. The deep learning model uses the labelled data to train but also can take advantage of unlabelled data in the training process which can improve the recognition performance. Eight volunteers are asked to perform 5 basic daily activities: standing, laying, walking downstairs, walking, running. The

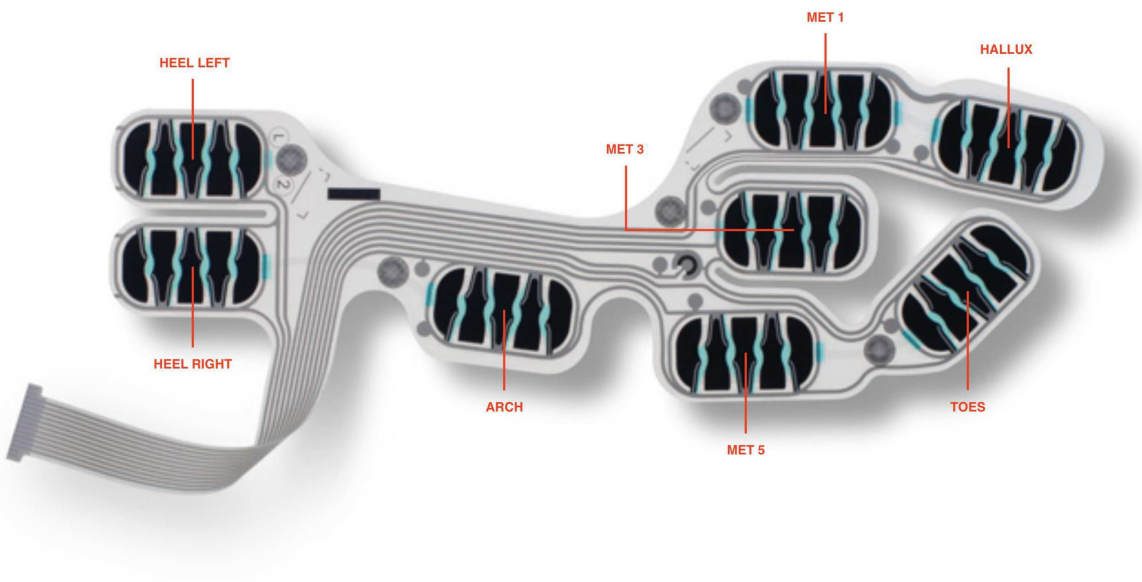


Figure 1: Pressure cells on the IEE Smart Foot Sensor.

model was evaluated with a Leave-One-Out approach reaching an overall accuracy of 98.92%. Paydarfar et al. [15] present a HAR system based on piezoresistor-based instrumented shoes and a Recurrent Neural Network (RNN). The hardware consists of a pair of sneakers, with an onboard microcontroller that is connected to 3 piezoresistor sensors located at the calcaneus, metatarsals, and phalanges. Twenty healthy subjects are engaged in the experiment. The activities collected are: walking, standing, balancing on the left foot, balancing on the right foot, toe-up, and ascending stairs. Each activity is performed by each subject for 45 to 120 seconds and the sampling frequency was set to 50 Hz. The data are segmented in slices of 1 second but in a way that each slice differ from the previous just by 1 timestep. The RNN was validate using a Leave One Out strategy reaching an accuracy of 87.0 ± 8.9%.

3. Method

3.1. Sensors and Data Collection

In this study a smart insole kit has been used, the ActiSense Kit provided by the IEE Luxembourg S.A. Two IEE Smart Foot Sensors and two ActiSense electronics (ActiSense ECU) are included in the ActiSense kit; the former is a pressure-cell insole and the latter is a device made up of various IMU sensors. The IEE Smart Footwear sensor is available in eight different sizes, ranging from 28 to 47 (EU) in which are located eight high dynamic pressure cells, at positions: left heel, right heel, arch, met 1, met 3, met 5, hallux, and toes (see Fig. 1). The ActiSense electronic includes a 3-axis accelerometer, a 3-axis gyroscope and a 3-axis magnetometer.

The data was collected on three healthy volunteers (3 males, ages 24-45). The IEE ActiSense Kit was worn by each participant, and two feet measurements were taken. The data was col-

lected via an Android software provided by the equipment maker, which took advantage of a Bluetooth link between the smartphone and the kit. The sampling frequency was set to 200 Hz, and each participant performed various activities in a predefined set (Downstairs, Fast Walking, Normal Walking, Sit to Stand, Sitting, Slow Walking, Standing, Upstairs) without supervision or restriction. In total, 120 minutes of recording were collected.

3.2. Data Distribution and Over-Sampling Technique

The data collected from the subjects are in form of multivariate time series. Due to the multiple sensors involved and the capability of the same sensors to provide multiple values is possible to summarise the data collected as described in Eq. 1.

$$\sum_{i=1}^{k} s_i = (d^1, d^2, ..., d^t) \tag{1}$$

where k represents the number of sensors involved, d^i the multiple values (such as for accelerometer that typically has three axes).

A single measurement at time t (sample) is not sufficient for the representation of an activity performed by a subject, so we need to group together multiple consecutive samples that are likely to contain information about one activity (segment). A segment, hence, can be represented through the Eq. 2.

$$w_i = (t_1, t_2) \tag{2}$$

where t_1 and t_2 represent, respectively, the starting and ending time within the time series. The segment usually are referred as "windows" and the difference between t_1 and t_2 as "window size".

In this study, the window size is set to 2 seconds, as it was noted that this number is repeatedly chosen in the literature when there are conditions that agree with ours, for instance, sample rate, type of data and subjects included.

After data segmentation one of the most important analysis to perform is to assess the data distribution in relation to the classes. If the data show that one or more classes have a large number of samples in contrast to other classes that have few, what is encountered is a class imbalance problem. The problem of class imbalance is relevant since it has been demonstrated to produce a major bottleneck in the performance that can be achieved using traditional learning methods that presume a balanced class distribution [16]. As shown in Fig. 2 the number of elements in the classes: downstairs, sit_to_stand, sitting, standing and upstairs, are much less than the other classes resulting in an unbalanced dataset. To overcome this problem two possible approaches are available: assign distinct costs to train examples [17], re-sample the original dataset, either by over-sampling the minority class and/or under-sampling the majority class. The approach chosen is the latter, in particular, the Synthetic Minority Over-sampling Technique (SMOTE) [18] was used. Starting with the minority classes, SMOTE develops new synthetic examples. A representative from the minority class is picked at random. The k closest neighbours of the example are picked at random depending on the quantity of over-sampling

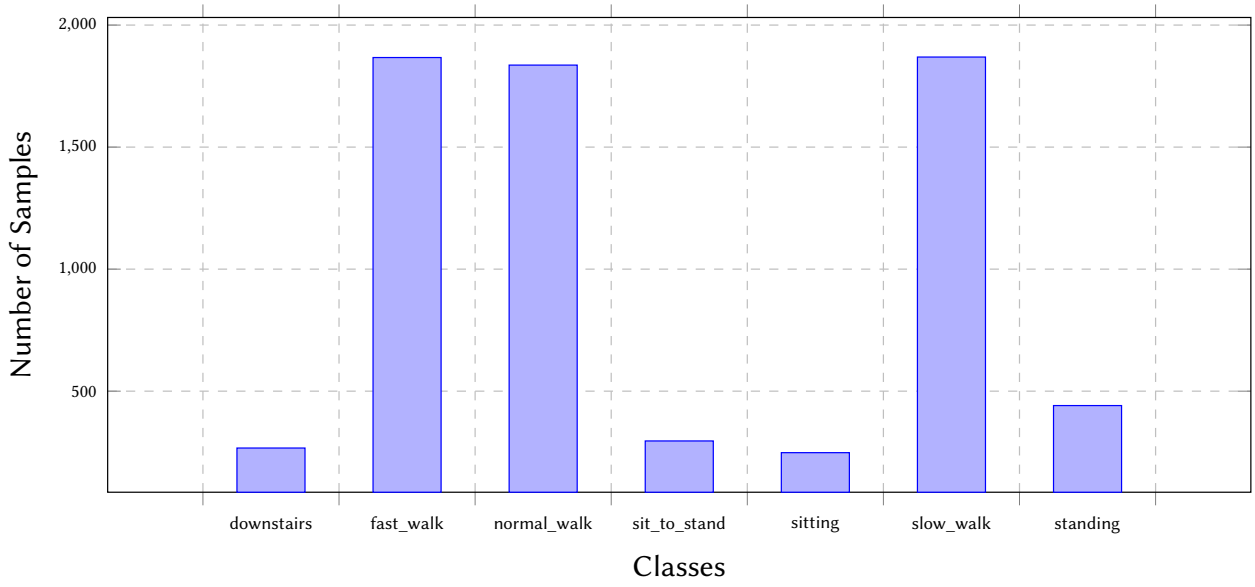


Figure 2: Samples distribution of the dataset after data segmentation with window size of 2 seconds.

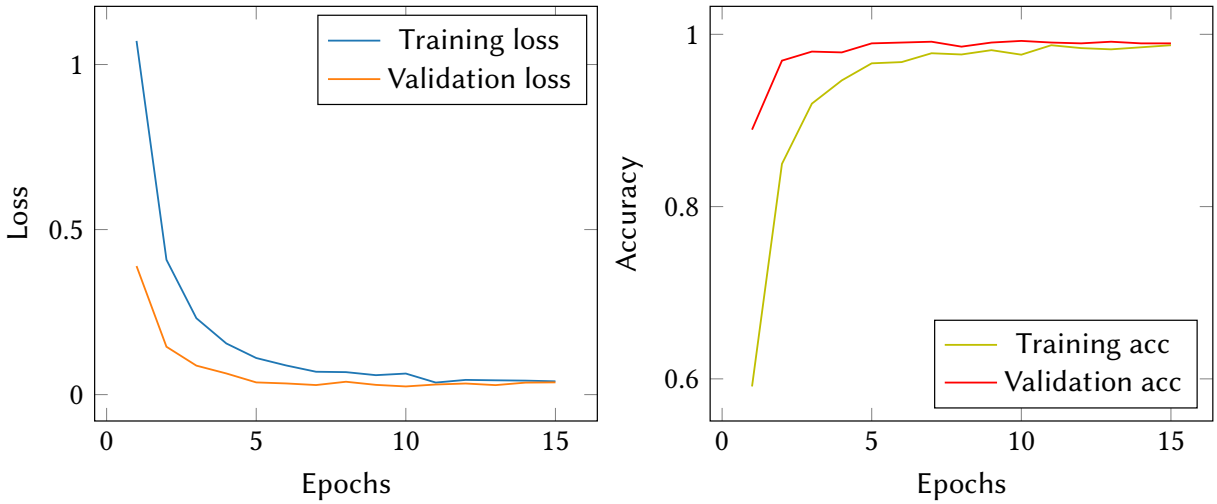
Table 1
Layers that compose the architecture of the ANN.

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 13600)	0
dense (Fully Connected)	(None, 256)	3481856
dropout (Dropout)	(None, 256)	0
relu (Activation)	(None, 256)	0
dense_1 (Fully Connected)	(None, 128)	32896
dropout_1 (Dropout)	(None, 128)	0
relu (Activation)	(None, 128)	0
dense_2 (Fully Connected)	(None, 8)	1032
softmax (Activation)	(None, 8)	0

necessary. Multiply the difference between the example and the neighbour by a random number between 0 and 1, then add the result to the example causing the selection of a random point along the line segment between two specific examples. This strategy forces the minority class's decision-making region to become more general.

3.3. Model Architecture

Human activities by their nature are difficult to recognise as they can be influenced by several factors. Each person differs from others as well as from himself in different circumstances. The large number of devices that can be used also affects the recognition itself. All this has led to a missing common definition of Human Activity and therefore having to analyse the data of each study in a different way [19]. Data analysis and core information extraction (feature extraction) take time and effort. Neural Networks has become a popular choice because they



(a) (b)

Figure 3: ANN model training history according to training and validation datasets. (a) Loss (b) Accuracy.

can overcome this problem.

In this study an Artificial Neural Network (ANN) is used to discriminate activities based only on raw data provided by the smart insole. The ANN tries to mimic the behaviour of the human brain. An ANN is therefore composed of three or more layers that are interconnected. The first layer consists of input neurons. Those neurons send data to the deeper layers, which in turn will send the final output data to the last output layer.

The layers that build up the ANN are shown in Table 1. The ANN takes as input the previously separated windows which have a size of (400, 34). Since the layers of our architecture cannot process a two-dimensional sample, the first layer of the architecture is a Flatten. A Flatten layer converts the input data in a column-wise shape to feed into the next layers. The Fully Connected layer consists of the weights and biases along with neurons and all the inputs are connected to every activation unit of the next layer. As a neural network learns, the weights of neurons are tuned providing some specialisation. Neighbouring neurons, on the other hand, begin to rely on this specialisation, which, if carried too far, can result in a weak model that is overly specialised to the training data. For this reason, a Dropout Layer is required to prevent overfitting by dropping out units in the neural network. Each neuron can be dropped with a probability p or kept with a probability $1 - p$ [20]. The Activation functions define how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network. Two types of activation functions are used: Rectified Linear Activation (ReLU) and Softmax.

The dataset was split for the training purpose, train and test sets were created. The ANN was trained for 15 epochs and the batch size set to 32. Moreover, the train set was split into two parts (train and validation, respectively 80% and 20%) for the objective of evaluating if during training overfitting or underfitting occurred. The training history of the model is shown in Fig. 3.

Table 2
Performances of the ANN model against the testing dataset.

	Precision	Recall	F1-Score
downstairs	0.9930	0.9965	0.9947
fast_walk	0.9811	0.9924	0.9867
normal_walk	0.9894	0.9824	0.9859
sit_to_stand	0.9966	1.0000	0.9983
sitting	1.0000	1.0000	1.0000
slow_walk	1.0000	0.9891	0.9945
standing	1.0000	1.0000	1.0000
upstairs	0.9963	0.9963	0.9963
Accuracy			0.9947
Macro avg	0.9945	0.9946	0.9946
Weighted avg	0.9947	0.9947	0.9947

4. Results and Discussion

The ANN model was trained with the goal of recognising 8 different basic activities: *downstairs, fast walk, normal walk, sit to stand, sitting, slow walk, standing and upstairs.*

To evaluate the performance of the classifier four metrics were used: accuracy, precision, recall and f1-score. The accuracy is the number of correctly predicted data points out of all the data points, in other words, it measures how often the algorithm classifies a data point correctly (see Eq. 3).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

The precision is the number of correct positive results divided by the number of positive results predicted by the classifier (see Eq. 4).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

The recall is the number of correct positive results divided by the number of all relevant samples (see Eq. 5).

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

The F1_Score is the Harmonic Mean between precision and recall. It tells you how precise your classifier is, as well as how robust it is (see Eq. 6).

$$F1\_Score = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

The overall accuracy of the model on the test set is 99.47%. Analysing the Table 2 is possible to notice that the classes *sitting* and *standing* can be clearly identified against others with precision and recall that are 100%. The worst class recognised is the *normal_walk* with an F1-Score of 98.59%.

Table 3
Confusion matrix of the ANN model against the testing dataset.

Predicted → Reference ↓	downstairs	fast_walk	normal_walk	sit_to_stand	sitting	slow_walk	standing	upstairs
downstairs	284	0	0	1	0	0	0	0
fast_walk	1	260	1	0	0	0	0	0
normal_walk	0	5	279	0	0	0	0	0
sit_to_stand	0	0	0	293	0	0	0	0
sitting	0	0	0	0	278	0	0	0
slow_walk	0	0	2	0	0	273	0	1
standing	0	0	0	0	0	0	298	0
upstairs	1	0	0	0	0	0	0	267

Analyzing the performance of the system it seems that walking activities have a lower recognition rate. To learn more about what this result entails, a confusion matrix has been created. As shown in Table 3, the highest number of misclassifications is between the *normal_walk* and the *fast_walk*. This error is due to the nature of the dataset. The dataset was created by people at different times and places in a completely independent way and with few if almost no guidelines. So for instance one person can consider the *normal_walk* at the speed of 6 km/h, instead, another person can recognise the same speed as a *fast_walk*.

Overall, the produced solution outperforms the findings of the other studies examined; nevertheless, the number and kind of participants are insufficient to cover the entire population, which could lead to a loss of performance if the system is employed on a subject that differs a lot from those examined.

5. Conclusion

In this paper a Human Activity Recognition solution that may be non-invasive for the user has been discussed. A smart insole consisting of pressure sensors and inertial sensors is used as unique device able to capture activity information from the user. Moreover, with the aim of creating a replicable pipeline that is not bound to the subjects studied so as to be able to expand the number of subjects if necessary, the smart insole was supported by an Artificial Neural Network which allows the processing of raw data and auto extraction of meaningful information from them. The proposed system can resolve the human activity recognition with an overall accuracy of 99.47%. The results obtained exceed the existing solutions analysed, however it turned out that the lack of guidelines for collecting data has resulted in problems in the recognition of some activities such as normal and fast walking, since not having defined a standard speed, each user has decided their own.

Inasmuch as, not all types of subjects were treated, the number of people assessed narrows the prospective targets for everyday use. As a result, the number of subjects included in the study may be increased in the future, allowing the system to cover a larger number of users. In addition, to cover as many scenarios of daily life, it is possible to expand the number of activities involved.

Acknowledgments

Luigi D'Arco is supported by the Ulster-Beitto Collaboration Program, Graham McCalmont is supported by Department of Economy PhD Scholarship.

References

[1] H. Zheng, H. Wang, N. Black, Human activity detection in smart home environment with self-adaptive neural networks, in: 2008 IEEE International Conference on Networking, Sensing and Control, IEEE, 2008, pp. 1505–1510.

[2] W. Wolf, I. B. Ozer, A smart camera for real-time human activity recognition, in: 2001 IEEE Workshop on Signal Processing Systems. SiPS 2001. Design and Implementation (Cat. No. 01TH8578), IEEE, 2001, pp. 217–224.

[3] M. Leo, T. D'Orazio, P. Spagnolo, Human activity recognition for automatic visual surveillance of wide areas, in: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks, 2004, pp. 124–130.

[4] G. McCalmont, H. Zheng, H. Wang, S. McClean, M. Zallio, D. Berry, A lightweight classification algorithm for human activity recognition in outdoor spaces, in: Proceedings of the 32nd International BCS Human Computer Interaction Conference 32, 2018, pp. 1–5.

[5] S. Irene, N. Shwetha, P. Haribabu, R. Pitchiah, Development of zigbee triaxial accelerometer based human activity recognition system, in: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, IEEE, 2015, pp. 1460–1466.

[6] S. Ashry, R. Elbasiony, W. Gomaa, An lstm-based descriptor for human activities recognition using imu sensors, in: Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics, ICINCO, volume 1, 2018, pp. 494–501.

[7] W. Gomaa, Statistical and time series analysis of accelerometer signals for human activity recognition, in: 2019 14th International Conference on Computer Engineering and Systems (ICCES), IEEE, 2019, pp. 351–356.

[8] E. Kańtoch, Human activity recognition for physical rehabilitation using wearable sensors fusion and artificial neural networks, in: 2017 Computing in Cardiology (CinC), IEEE, 2017, pp. 1–4.

[9] F. Nurwanto, I. Ardiyanto, S. Wibirama, Light sport exercise detection based on smartwatch and smartphone using k-nearest neighbor and dynamic time warping algorithm, in: 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE, 2016, pp. 1–5.

[10] W. Ugulino, M. Ferreira, E. Velloso, H. Fuks, Virtual caregiver: a system for supporting collaboration in elderly monitoring, in: 2012 Brazilian Symposium on Collaborative Systems, IEEE, 2012, pp. 43–48.

[11] C. Moufawad el Achkar, C. Lenoble-Hoskovec, A. Paraschiv-Ionescu, K. Major, C. Büla, K. Aminian, Instrumented shoes for activity classification in the elderly, Gait & posture 44 (2016) 12–17.

[12] R. De Pinho André, P. H. F. Diniz, H. Fuks, Bottom-up investigation: Human activity recognition based on feet movement and posture information, in: Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction, 2017, pp. 1–6.

[13] C. Pham, N. N. Diep, T. M. Phuong, e-shoes: Smart shoes for unobtrusive human activity recognition, in: 2017 9th International Conference on Knowledge and Systems Engineering (KSE), IEEE, 2017, pp. 269–274.

[14] X. Zhang, J. Zhang, Subject independent human activity recognition with foot imu data, in: 2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), IEEE, 2019, pp. 240–246.

[15] A. J. Paydarfar, A. Prado, S. K. Agrawal, Human activity recognition using recurrent neural network classifiers on raw signals from insole piezoresistors, in: 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), IEEE, 2020, pp. 916–921.

[16] M. Kubat, R. C. Holte, S. Matwin, Machine learning for the detection of oil spills in satellite radar images, Machine learning 30 (1998) 195–215.

[17] P. Domingos, Metacost: A general method for making classifiers cost-sensitive, in: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 155–164.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.

[19] A. Bulling, U. Blanke, B. Schiele, A tutorial on human activity recognition using body-worn inertial sensors, ACM Computing Surveys (CSUR) 46 (2014) 1–33.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (2014) 1929–1958.