Comparing Deep Learning and Human Crafted Features for Recognising Hand Activities of Daily Living from Wearables

Eleni Diamantidou*†, Dimitrios Giakoumis*, Konstantinos Votis*, Dimitrios Tzovaras* and Spiridon Likothanassis†
*Centre for Research and Technology Hellas, Information Technologies Institute, 6th km Charilaou-Thermi, 57001 Thermi, Greece
†University of Patras, Computer Engineering and Informatics Department, 26504 Patra, Greece

***Abstract*—This work presents a comparative analysis of human-crafted and automated feature extraction approaches for the discrimination of hand-based activities among eating, drinking and smoking. In this scheme, accelerometer and gyroscope sensors were utilised to capture activity signals. For this reason, wearable devices that embed the aforementioned sensors were employed to collect activity data from 12 office workers. The two approaches that were developed for feature mapping were evaluated equally on the collected dataset. Both the proposed schemes achieved to classify the hand-based activities. However, based on the experimental process, this study shows that the human-crafted features that extracted valuable information from the time and frequency domain of the raw signal measurements outperformed the automated feature mapping that utilised deep learning advances. The relevant results prove that the human-crafted features can recognise hand-based activities with 0.9109 and, on the other hand, automated features with a 0.907 F1 weighted score over the dataset.**

***Index Terms*—human activity recognition, activities of daily living, wearables, hand-based activities, feature extraction**

I. Introduction

In recent years, human activity recognition (HAR) has become one of the most popular research disciplines. Daily activity tracking is gaining popularity since it can provide useful information about a person's everyday life and well-being [1]. The study and understanding of human activities and behavioural patterns are associated with HAR. Humans have a variety of daily routines based on their needs, while a pile of them are associated with hand-based actions [2]. Hand gestures are commonly used in people's daily life for completing actions, such as drinking, eating, answering a phone, smoking and plenty of equivalent examples. The aforementioned activities are a longstanding objective of the HAR research since humans execute similar fine-grained body movements in order to establish them. The monitoring of these activities can allow both the individuals and the researchers to support several wellness goals. One of the most appealing challenges in the recognition of hand-based daily activities is the in-between identification and discrimination. For instance, the performance of daily routines concerning eating, drinking and smoking activities involve almost identical hand movements, as Figure I illustrates. The monitoring of these activities is of major importance, aiming to understand for example, whether humans skip their meals and their hydration and detect systematic smoking events.

Fig. 1. Representation of a typical human arm movement involved in all, complete eating, drinking and smoking activities.

The monitoring of hand-based activities can be achieved based on either obtrusive or unobtrusive methods. Many of the existing human activity detection systems utilise obtrusive techniques, such as vision schemes [3], [4] to capture in most cases non-free-living activity events. Appealing studies employing depth cameras achieved to recognise automatically essential activities of daily living such as cooking, eating, dishwashing and watching TV [5]. Moreover, audio-based approaches have been studied for the HAR challenge, investigating an effective feature mapping [6]. However, all the aforementioned approaches involve a high level of obtrusiveness in individual daily life [4]. To counteract for the obtrisiveness problem, the present work employs inertial sensors, like accelerometers and gyroscopes that are embedded in wearable devices such as smartwatches. Several existing works discuss and discover several HAR approaches based on wearable devices [7]. An accelerometer measures the accelerations of a specific body segment, in this case, the arm to which is attached. However, a gyroscope measures the rotation rate of the hand movements. Both accelerometers and gyroscopes record body movements signals in the form of time series which enables different approaches to the classification problem. Deep Learning (DL) based-methods have been increasingly employed for the HAR task in the last decade [8]. A major advantage of DL techniques for feature extraction is that it does not require domain-specific expertise [9] and so they can be employed for several applications such as HAR. However, DL methods aiming to fulfil an accurate feature extraction require data from a specific target domain depending on the problem. Therefore, recent studies introducing DL methods that use mobile and wearable devices

have pushed the boundaries in the HAR [7]. On the other hand, common traditional features are also employed for classification purposes. Many types of human-crafted features such as mean, standard deviation, and energies are commonly used in the literature for HAR [10]. Identical studies have been studied and analysed the effects of using human-crafted feature mapping toward the automated feature extraction for the HAR task [11].

The presented work focuses on implementing two approaches of feature learning for the task of classifying and distinguishing eating, drinking and smoking events. In particular, this work provides:

- a comprehensive analysis and processing of data regarding the eating, drinking and smoking events recorded from wearable sensors.
- a comparison between DL high-level features and human-crafted features for the HAR task.

The rest of the paper is organised as follows. Section II provides important information for the datasets and explains the data pre-processing. Moreover, section III describes the modelling of the HAR architectures that were investigated for detecting eating, drinking and smoking events, along with a comprehensive analysis of the experimental results. Finally, Section IV concludes the work with fundamental findings and remarks the important features of this study.

II. Data Analysis and Signal Processing

A set of free-living experiments were carried out to obtain the dataset regarding the hand-based data towards activity recognition. The four selected activities were eating, drinking, smoking, and idle status. A group of 12 office workers volunteers with age ranging from 28 to 38 years supported this study. For each one of them, almost one hour of data for each activity were captured, respectively. During the data collection, each subject performed naturally the activities of interest while wearing a Samsung Galaxy Watch 3 smartwatch which was placed by the user himself on the preferred (dominant) arm. No instructions were given to participants, concerning following some protocol of activities; thus each person was performing the typical activities that would be performing at home, within the monitoring period. The annotation of the recorded data was reported after each data collection manually by the user. The tasks were performed in an office environment where the volunteers were asked to perform freely the activities in order to create a realistic dataset.

The obtained dataset consists of sensor measurements recorded by the smartwatch embedded accelerometer and gyroscope, while these sensors capture triaxial linear acceleration and angular velocity information. Moreover, the sensor signals were recorded at a sampling rate of 50 Hz. An indicative example of the raw sensor measurements is presented in Figure II, where a volunteer performs a smoking event.

The raw signals were further processed by applying denoising filters to eliminate noise that may was captured during the recordings and caused by sensor failures. On the basis of corresponding signal pre-processing techniques for HAR data, a specific filtering methodology was applied [12]. Pursuant to that, a median filter and a 3rd order low-pass Butterworth filter with a 20 Hz cutoff frequency were applied at moth accelerometer and gyroscope raw measurements aiming to reduce noise in these signals.

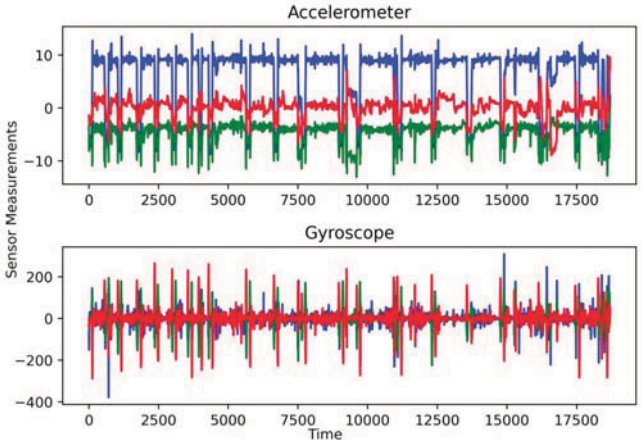

Fig. 2. Example of raw sensor measurements regarding a smoking event.

Thereafter, the filtered signals were divided into smaller segments of action. Based on a detailed analysis, an average person requires almost 3 seconds to complete a cycle of eating, drinking or smoking actions. As a result, the signals were sampled into a predefined window of 3-sec with a 50% overlap between them. The segmentation method was applied to the x, y, and z axes of both accelerometer and gyroscope measurements. Since the sensors capture body movements with a 50Hz sampling rate, each cycle window involves 150 signal measurements. From each segment of the recorded data, features were extracted by following either the human-crafted feature mapping or the automated feature mapping procedures, as further explained below.

1) Human-crafted feature mapping: The human-based feature mapping concerns the extraction of knowledge from sensor signals based on well-known functions that can be applied in the time or frequency domain of the raw signals. As a result, from each cycle window from the raw signal, a set of features was obtained. The feature mapping was calculated using common measurements well-known utilised in the HAR literature [13], among mean, correlation, min, max and signal magnitude area. Additional features extracted from the frequency domain, including energies, entropies and angles between axes vectors. These functions were applied on each sensor axes separately. Besides, magnitude values were calculated for the triaxial accelerometer and gyroscope measurements using the

$$sensor\_magnitude = \sqrt{x^2 + y^2 + z^2}$$

equation. Thus, human-based feature mapping was related directly to the sensor axes and to the overall magnitude. Table I contains the list of all the feature measures functions that are applied to the time and frequency domain signals.

TABLE I
LIST OF HUMAN-CRAFTED FEATURE VECTORS

Function	Description
mean	Mean value
std	Standard deviation value
mad	Median absolute value
max	Maximum values in sensor measurements
min	Minimum values in sensor measurements
sma	Signal magnitude area
percentile	Score below which a given percentage k
moment	Characteristic of signal distribution
energy	Average sum of signal squares in different frequency bands
iqr	Interquartile range
skewness	Frequency signal skewness
kurtosis	Frequency signal kurtosis
correlation	Correlation coefficients between measurements axes
spectralEntropy	Uniformity of signal energy distribution in the frequency-domain
valueEntropy	Signal entropy
meanFreq	Mean frequency coefficients

2) Automated feature mapping: One of the most important advantages of DL neural networks is their ability for automatic feature extraction since they can extract useful patterns that even not a human eye can obtain. In this study, a DL approach was also employed to create a high-level feature mapping [14]. A one-dimensional custom Convolutional Neural Network (CNN) was developed to extract features from the raw signal measurements. Accordingly, 1D convolutional layers, with 32 filters were utilised to extract valuable information at the spatial domain of the signal. In this approach, the 3-sec segments of raw signal activity were fed directly to the 1D CNN to recognise patterns for the final classification. The output of the CNN architecture was fed into a densely connected DL classifier.

III. EXPERIMENTS AND RESULTS

To solve the problem of classifying hand-based activities, a custom densely connected neural network was designed. The classifiers' architecture is presented in Figure III. To avoid increasing the complexity of the model, the proposed architecture was developed based on an experimental process that finalised the number neural network layers and nodes. The evaluation of the two feature mapping approaches was handled using the dataset collected from the wearables. This dataset was used to train firstly the human-crafted extracted features, and then the output of the CNN feature extractor. In the first case, the features were trained directly using the dense classifier. In the same notions, the feature extracted using the CNN were fed evenly to the same classifier. Further details regarding the training and evaluation processes are described below.

A. Training

Two main training experiments were conducted. In the first case, the classification was held out employing human-crafted features that extracted from time or frequency domain of the raw signals. In the second case, the classification process was based on automated feature extraction that generated features from a CNN architecture, as discussed earlier. A densely connected neural network was designed to train the features from each case of experiment. The dataset was randomly partitioned into two distinct sets to enable performance evaluation, with 70% of the data selected for training and the remaining 30% for testing. Adam optimiser [15] with default parameters [13] was employed. The softmax [16] activation was chosen to predict the probability distribution of each training class. Each experimental phase was trained for 100 epochs. The experimental process was implemented using Python frameworks, such as a Keras [17] implementation of CNN with TensorFlow [18] back-end.

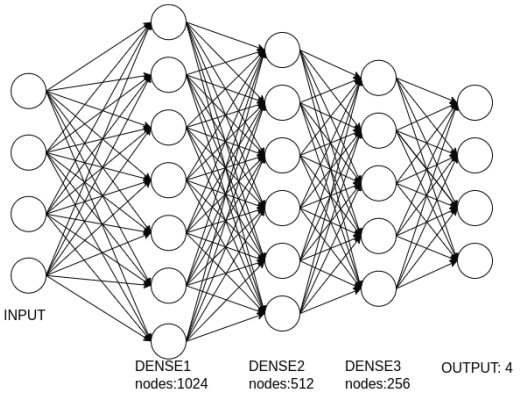

Fig. 3. Densely connected neural network architecture for detecting eating-drinking-smoking events

B. Experimental Results

Each feature generation approach for the acceleration and gyroscope data focused on the set of specific target activities defined as eating, drinking, smoking and idle.

The methods quality was measured in terms of precision, recall and F1-score. The classification performance for each class for each approach is shown in terms of the classification report. Table II presents the evaluation performance of activity recognition based on human-crafted features and Table III presents the corresponding performance based on high-level feature mapping utilised by 1D CNN modelling.

Observing the evaluation results, both the two feature mapping approaches obtained promising evaluation scores since they correctly classified the activities of eating, drinking, smoking and idle events. Fine recognition accuracy can be obtained utilising convolutional layers. However, a noticeable improvement in activity recognition was obtained by the utilisation of human-crafted feature mapping. According to the results at Table II and Table III, it should be considered that using feature mapping based on a CNN provided a more forthright way to extract information from sensor signals compared to the human-crafted case. However, the experimental process presented in this work shows that a CNN did not outperform conventional human-crafted features. The target human activities involve hand fine-grain movements that are

very comparable. A CNN architecture was not capable of identifying similarities or dissimilarities in this type of data. On the other hand, human-crafted feature mapping achieved to extract beneficial knowledge from the raw signals, employing characteristics such as signal energies and entropies. Based on the evaluation results, the human-crafted features represented more suitable the disparities of the training data, reaching higher F1-score.

TABLE II
CLASSIFICATION REPORT REGARDING THE HUMAN-CRAFTED FEATURE LEARNING EXPERIMENT

Activity	Precision	Recall	F1-score
drinking	0.9163	0.8223	0.8668
eating	0.9006	0.9129	0.9067
smoking	0.8874	0.9149	0.9009
idle	0.9395	0.9468	0.9431
Weighted average	**0.9114**	**0.9112**	**0.9109**

TABLE III
CLASSIFICATION REPORT REGARDING THE AUTOMATED FEATURE LEARNING EXPERIMENT

Activity	Precision	Recall	F1-score
drinking	0.7726	0.7245	0.7478
eating	0.7970	0.8154	0.8061
smoking	0.8055	0.8375	0.8212
idle	0.8196	0.8193	0.8192
Weighted average	**0.8712**	**0.902**	**0.907**

IV. CONCLUSIONS

In this study, a comparative comparison between human-based and automated feature extraction techniques was presented for the task of identifying eating, drinking and smoking events, which is very little analysed in the literature. Promising results were obtained during the experimentation. This research was based on a dataset collected for this specific task using wearable sensors such as smartwatches. The findings of each approach were equally presented.

The human-based feature mapping obtained improved results compared with the corresponding DL method. These outcomes were induced by the small complexity level of data that DL architectures require to learn and recognise patterns. Moreover, additional extensive experiments and data collection sessions will be handled to further analyse the outcomes in larger datasets and improve the effects of the method that has been studied in this work. Future work will have to investigate the necessary steps to enhance the discrimination of more hand-based activities.

ACKNOWLEDGMENT

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Name: ACTIVE, Project Number: HFRI-FM17-2271)

REFERENCES

[1] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *IEEE pervasive computing*, vol. 9, no. 1, pp. 48–53, 2009.

[2] S. Zhang, Y. Zhao, D. T. Nguyen, R. Xu, S. Sen, J. Hester, and N. Alshurafa, "Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 4, no. 2, pp. 1–26, 2020.

[3] Y. Bai, W. Jia, Z.-H. Mao, and M. Sun, "Automatic eating detection using a proximity sensor," in *2014 40th Annual Northeast Bioengineering Conference (NEBEC)*. IEEE, 2014, pp. 1–2.

[4] M. Vasileiadis, S. Malassiotis, D. Giakoumis, C.-S. Bouganis, and D. Tzovaras, "Robust human pose tracking for realistic service robot applications," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 1363–1372.

[5] D. Giakoumis, G. Stavropoulos, D. Kikidis, M. Vasileiadis, K. Votis, and D. Tzovaras, "Recognizing daily activities in realistic environments through depth-based user tracking and hidden conditional random fields for mci/ad support," in *European Conference on Computer Vision*. Springer, 2014, pp. 822–838.

[6] A. Vafeiadis, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, "Audio-based event recognition system for smart homes," in *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2017, pp. 1–8.

[7] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, p. 1476, 2022.

[8] Q. Zhu, Z. Chen, and Y. C. Soh, "A novel semisupervised deep learning method for human activity recognition," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3821–3830, 2018.

[9] Y. LeCun, Y. Bengio, G. Hinton *et al.*, "Deep learning. nature, 521 (7553), 436-444," *Google Scholar Google Scholar Cross Ref Cross Ref*, 2015.

[10] M. Dong, J. Han, Y. He, and X. Jing, "Har-net: Fusing deep representation and hand-crafted features for human activity recognition," in *International Conference On Signal And Information Processing, Networking And Computers*. Springer, 2018, pp. 32–40.

[11] F. Cruciani, A. Vafeiadis, C. Nugent, I. Cleland, P. McCullagh, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, "Comparing cnn and human crafted features for human activity recognition," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2019, pp. 960–967.

[12] D. Anguita, A. Ghio, L. Oneto, X. Parra Perez, and J. L. Reyes Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, 2013, pp. 437–442.

[13] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern recognition letters*, vol. 29, no. 16, pp. 2213–2220, 2008.

[14] M. Jogin, M. Madhulika, G. Divya, R. Meghana, S. Apoorva *et al.*, "Feature extraction using convolution neural networks (cnn) and deep learning," in *2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*. IEEE, 2018, pp. 2319–2323.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in *Proc. 8th Aust. Conf. on the Neural Networks, Melbourne*, vol. 181. Citeseer, 1997, p. 185.

[17] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[18] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.