Sensor-based Activity Recognition using Deep Learning: A Comparative Study

Imen Trabelsi
imen.trabelsi@universite-paris-saclay.fr
Université Paris-Saclay, CNRS, LISN
91400, Orsay, France

Jules Françoise
jules.francoise@lisn.upsaclay.fr
Université Paris-Saclay, CNRS, LISN
91400, Orsay, France

Yacine Bellik
Yacine.Bellik@universite-paris-saclay.fr
Université Paris-Saclay, CNRS, LISN
91400, Orsay, France

ABSTRACT

With the wide availability of inertial sensors in smartphones and connected objects, interest in sensor-based activity recognition has risen. Yet, recognizing human actions from inertial data remains a challenging task because of the complexity of human movements and of inter-individual differences in movement execution. Recently, approaches based on deep neural networks have shown success on standardized activity recognition datasets, yet few works investigate systematically how these models generalize to other protocols for data collection. We present a study that evaluates the performance of various deep learning architectures for activity recognition from a single inertial measurement unit, on a recognition task combining data from six publicly available datasets. We found that the best performance on this combined dataset is obtained with an approach combining the continuous wavelet transform and 2D convolutional neural networks.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Human Activity recognition, Inertial Sensors, Movement Computing, Continuous Wavelet Transform, Deep Learning, Convolutional Neural Networks.

ACM Reference Format:
Imen Trabelsi, Jules Françoise, and Yacine Bellik. 2022. Sensor-based Activity Recognition using Deep Learning: A Comparative Study. In *8th International Conference on Movement and Computing (MOCO'22), June 22–24, 2022, Chicago, IL, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3537972.3537996

1 INTRODUCTION

Computational movement modeling is a central topic in the Movement and Computing (MOCO) community, as it is the foundation for the design of interactive systems involving gesture and whole-body motion. The MOCO community has witnessed an increasing number of contributions to new conceptual frameworks for movement analysis[12], signal processing methods [25, 28, 44] and applications of Machine Learning (ML) to movement analysis and recognition [15, 21]. Yet, using ML for analyzing movement data from wearable sensors remains challenging, especially in applications domains where collecting and annotating data remains time-consuming or impractical, for instance in music and performing arts. Approaches such as interactive machine learning have shown success for designing gesture recognition systems from few examples [6, 19, 20, 22]. Yet, these approaches often rely on hand-crafted features and standard ML models, which can limit their use for modeling higher-level movement characteristics and concepts such as movement qualities [11, 17, 18].

In this project, we are interested in drawing upon Deep Neural Networks (DNNs) to learn representations of movement from inertial sensor data. While our goal is to apply such techniques to a broad spectrum of movement and computing applications, we consider in this paper the task of sensor-based Human Activity Recognition (HAR), where deep learning has shown success in recent years [14, 26, 32, 38, 40].

HAR based on wearable sensors finds applications in a wide range of areas such as health an well being, monitoring physical activity or surveillance. Sensor-based HAR entails recognising a person's activity (e.g. walking, running, etc.) from their physical movements by analysing data generated from on-body inertial sensors. These wearable sensors usually embed accelerometers, gyroscopes and magnetometers. Traditional approaches to HAR apply classic ML models – including Random Forests (RF), Bayesian Network (BN) or Support Vector Machines (SVM), – to movement features that are engineered by hand. In recent years, DNN-based methods have achieved tremendous success in sensor-based HAR, and several architectures have been proposed, including convolutional and recurrent neural networks [14, 26, 35, 40]. Yet, to date few approaches have investigated the potential of time-frequency representations using the wavelet transform in conjunction with convolutional neural networks (CNNs) which are extremely successful in computer vision. Considering that human activities captured with wearable sensors often involve characteristic temporal patterns, we hypothesize that the Continuous Wavelet Transform (CWT), that provides an accurate time-frequency representation, would result in a valuable representation for recognizing activities. Furthermore, algorithms are usually evaluated on standard datasets that are publicly available, that are built with a fixed protocol and a limited number of activities.

MOCO'22, June 22–24, 2022, Chicago, IL, USA

ACM ISBN 978-1-4503-8716-3/22/06…$15.00
https://doi.org/10.1145/3537972.3537996

Our contributions in this paper are twofold. First, rather than considering only individual datasets, we propose a more challenging task by combining activities from different datasets. This process introduces more variability in data collection protocols and sensor placement, and challenges the generalization power of HAR methods. Second, we propose a new method where accelerometer and gyroscope signals are first analyzed using the Continuous Wavelet Transform (CWT) to generate a 2D image representation, and then classified using a CNN 2D. We demonstrate that wavelet CNNs achieve competitive accuracies and outperformed the traditional statistical and other deep learning methods when combining activities from different datasets.

The remainder of this paper is structured as follows: Section 2 summarizes the related work. Section 3 presents the proposed methods and metrics. Section 4 describes the experimental setup and presents the results, Section 5 contains a discussion of the obtained results, and Section 6 concludes this paper and proposes future improvements.

2 RELATED WORK

HAR methods consists in monitoring and analyzing a person's activity behaviors to enable computing systems to proactively assist users in various tasks. Significant progress has been made by conventional ML methods. However, with the rapid development of DL, there is an emerging interest and effort in applying it. Figure 1 presents a typical flowchart of HAR from inertial measurement unit (IMU) sensor data using approaches based on feature engineering or approaches based on deep neural networks.

2.1 Approaches based on Feature Engineering

Conventional approaches to HAR often draw upon classic learning algorithms applied to handcrafted features. De Leonardis et al. [16] explored five classifiers (SVM, decision tree, K-Nearest Neighbor (KNN), Feedforward Neural Network (FNN) and Naïve Bayes (NB) for classifying eight activities performed by fifteen subjects from their own dataset. The results showed that all classifiers were able to correctly recognize more than 90% of activities, and that best performances were obtained by KNN. Catal et al. [13] conducted several experiments on the *Wisdm* dataset with decision trees, logistic regression, and multi-layer perceptron algorithms, and combines these classifiers with the average of probabilities combination rule. Gao et al. [23] evaluated the effect of multiple sensors on HAR. They compared four classifiers, and concluded that decision tree classifier was the best algorithm for this problem. They explain that traditiona artificial neural network (ANN), KNN, and SVM classifiers are computational expensive although they provide good performance. The proposed approach achieved 96.4% accuracy when recognizing on their own dataset. Batool et al. [5] explored the Particle Swarm Optimization (PSO) together with SVM algorithm. The results of their model shows that pre-classifier as PSO and SVM along with feature extraction module have shown accuracy of 87.50% over *MotionSense* dataset.

While conventional approaches have made progress in HAR, they remain limited in several aspects. First, traditional methods require handcrafted features, which heavily relies on human expertise. This lengthy process is task-dependent and for more general tasks, this will show performance degradation [27]. Second, handcrafted features often refer to some statistical information. They can only be used to recognize basic physical activities such as walking or standing, and can hardly infer high-level or context-aware activities [29].

2.2 Approaches based on Deep Neural Networks

Unlike conventional techniques, approaches based on DNNs can directly process raw data and have outperformed conventional approaches especially for the case of heterogeneous, large and unfamiliar datasets [24, 39]. There are important reasons behind deep learning's success: the availability of big data, advances in learning algorithms and hardware, and increase in computing power. Chen et al. [14] provide a recent review of sensor-based human activity recognition using deep learning. Muralidharan et al. [35] compared classic approaches based on feature engineering with standard ML algorithms, as introduced in the previous section, with 1D Convolutional Neural Network (CNNs) on UCI HAR dataset. They found that the SVM and the proposed 1D convolution neural network were the best-performing models. Qi et al. [40] proposed a fast and robust CNN to perform complex activity recognition from smartphone sensors and the accuracy reached 94.18% in the experiment. Ignatov [26] proposed using a 1D CNN for local feature extraction together with simple statistical features. The proposed CNN architecture consists of a convolutional layer, a max-pooling layer, the output of the max-pooling layer is then flattened and stacked together with additional statistical features, followed by a fully-connected layer. The authors obtained 97.62% accuracy on the UCI HAR dataset with the proposed method.

In [32, 48], the authors presented an LSTM network for the classification of six activities from UCI HAR dataset. The first one achieves 92% average accuracy, while the second achieves 91.61% of accuracy. To improve accuracy of predictions, researchers in [10, 32, 38] merged functionalities of CNN and LSTM architectures to create a new one, the so-called CNN-LSTM. They used the outputs of CNN as input data for LSTM and obtained good results.

Alemayoh et al. [2] proposed a CNN with a double-channeled time-domain input to identify activities. In the proposed architecture, data were shaped into double-channels to extract deep temporal and spatial features of the signals. In [45], the authors split the sensor data into small segments of 128 data points. A frequency representation is taken from these windows by applying the Discrete Fourier transform (DFT). A pre-trained deep neural network (AlexNet) extracts features from the produced images. For the classification task, a SVM is choosen with linear and nonlinear kernels. An overall accuracy of 94% was reached on the UCI HAR dataset. While a number of neural network architectures have been proposed for sensors-based HAR, few works have investigated the potential of time-frequency representations such as the continuous wavelet transform [37]. Moreover, most algorithms are evaluated on published dataset in isolation. In this paper, we propose a new method for sensor-based HAR that applies 2D CNN to scalograms computed from IMU data. We evaluate the proposed method against state of the art algorithms, both on isolated datasets and on a combined dataset that aggregates data from various sources.

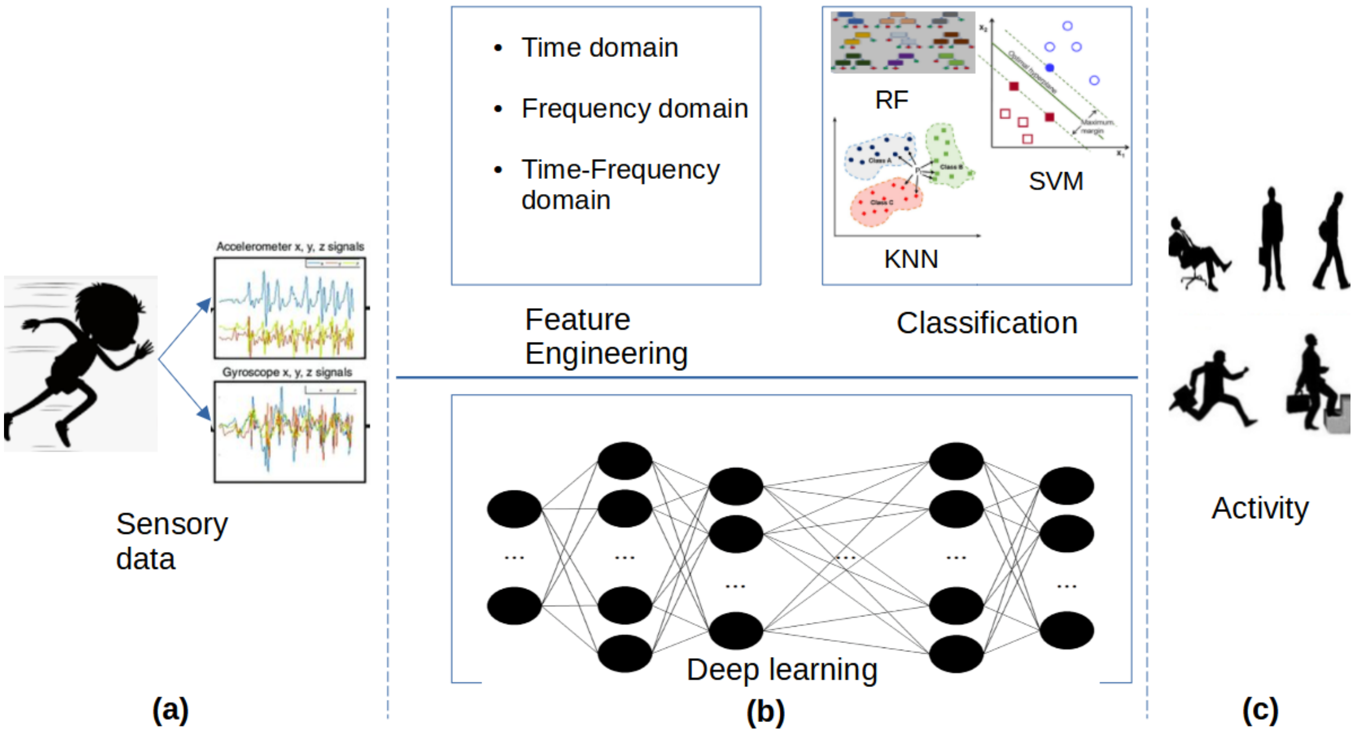

Figure 1: HAR from IMU sensor data comprises three main parts, (a) data capture and pre-processing (accelerometer and gyroscope data); (b) modeling and training, using either feature engineering with standard classifiers or deep learning extract representations from raw data, and (c) activity recognition.

3 METHODS

This Section describes the adopted four baseline approaches, as well as our method combining CWT and CNNs.

3.1 Baseline Approaches

3.1.1 Random Forests on Statistical Features. We first consider a conventional baseline that applies a standard classifier on a large set of both time-domain and frequency-domain features. Frequency domain signals were obtained by applying a Fast Fourier Transform (FFT). Following Anguita et al. [3], we used 561 features to train a classifier. We experimented with several popular algorithms (KNN, Adaboost, NB and RF) and found that Random Forests (RF) gave higher accuracy in this circumstance. RF classifiers are efficient on huge data and can manage hundreds of input variable without deletion [8]. It is also preferred because of its interpretability and low computational cost.

3.1.2 Convolutional Neural Network (CNN 1D). CNNs have been actively used and achieved impressive performance in many research area that include activity recognition [34]. CNNs consist of combinations of convolutional, pooling, normalization, and fully connected layers that extract progressive yet semantically rich features from the input data. CNNs were developed for image classification problems, where the model learns an internal representation of the input image, and produces a batch of 2D feature, in a process referred to as feature learning. This same process can be harnessed on one-dimensional sequences of data, such as in the case of acceleration and gyroscopic sensors. The model learns to extract features from sequences of observations and how to map the internal features to different activity types. CNN are able to capture the spatial information and connections of multimodal sensory data.

In this paper, we adopt the architecture proposed by Ignatov [26], that consists of a convolutional layer, followed by one Max pooling layer, followed by a flattening convolutional layer, then one fully-connected layer and finally a softmax classifier. It is denoted **CNN1D** in the following. The choice of this method is motivated by the fact that this algorithm is computationally efficient, and the obtained results significantly outperforms baseline approaches [26].

3.1.3 Long-Short Term Memory Networks (LSTM). Long Short-Term Memory (LSTM) are a special case of Recurrent Neutral Networks (RNN), having the capability to extract dependencies in time-series data and incrementally learn information through time intervals. Hence, they are appropriate for streaming sensory data in human activity recognition. The internal architecture of LSTM includes several gates where each gate processes the input from the previous gate and forward it to the next gate thereby controlling the flow of information towards the final output.

We compare two LSTM architectures. The **LSTM** architecture consists of one LSTM layer, followed by one fully-connected layer and finally a dense layer with a softmax activation function [10].

The **CNN-LSTM** architecture involves CNN layers for feature extraction on input data combined with LSTMs to support sequence prediction. The architecture seeks to leverage the combined power of both networks [36, 47]. The architecture consists of two sets of convolutional layers, followed by one Max pooling layer, followed by a flattening convolutional layer. These features are then passed through one LSTM layer, followed by one fully-connected layer and finally a dense layer with a softmax activation function for classification [9].

3.2 Proposed Method Combining CWT and 2D CNNs

We propose a new approach that draws upon a time-frequency representation of IMU data, combined with 2D convolutional networks, as illustrated in Figure 2. A similar approach has previously been proposed using the windowed Fourier transform [45]. This approach is motivated by the fact that many daily activities such as walking, running, involve specific temporal patterns that could be analyzed in the spectral domain. Time-frequency representations can then be considered images, and we can take advantage of methods developed in computer vision to analyze spectrograms as images. Rather than using the Short-Term Fourier Transform (STFT), we propose to consider the Continuous Wavelet Transform (CWT), which addresses limitations of the STFT [1, 31, 43]. Wavelets could be considered as viable options when considering feature extraction in activity recognition, as evidenced in [42]. Rather than imposing a fixed window length for analysis, the CWT translates and dilates a wavelet function with short-term influence, raising better resolution in both time and frequency [43]. The CWT also gives flexibility in the choice of the Wavelet basis used for analysis, and in the choice of the frequency range and resolution to use for analysis.

In this paper, we use the Complex Morlet wavelet, which is defined as a plane wave modulated by a Gaussian window. Using the complex version of the Morlet wavelet enables us to compute the scalogram (analog to the spectrogram with the Fourier transform), by taking the power of the transform. We used 32 frequency bands distributed in log scale, with a fixed number of 4 bands per octave. Applied to signals sampled at 50 Hz, the corresponding frequency range is $[0.11, 25]$ Hz. Since each signal has six components, each signal will also have six scalograms. The scalograms are placed on top of each other to create one single image with six channels, of shape $(128, 32, 6)$. For computational efficiency and reduced memory, the scalograms are downsampled by a factor 4 along the time axis, raising images of shape $(32, 32, 6)$. We can therefore use the images resulting from the mapping from the input sensors into a 2-dimensional space via CWT as input of a 2D CNN.

Given that images have small dimension and low complexity (compared to pictures), we used a simple 2D CNN architecture, summarized in Figure 2- (c), with 2 convolutional layers with 64 filters of size 3x3, interleaved with max pooling layers with kernel size 2 and stride 2, a flattening layer is used then to convert the output of the pooling layer, followed by two dense layers containing 1024, 128 units separately with ReLU activation, dropouts are applied after each dense layer. And the last dense layer, uses softmax as the activation function.

4 RESULTS

4.1 Experimental Protocol

We assessed the performance of the methods presented in Section 3.1 to predict human activities on several datasets. In order to ensure the reproducibility of our results, we chose to train and evaluate our methods on six publicly available datasets, that are summarized in Table 1. To further assess the generalization power of the various methods, we also created a new "combined" dataset aggregating data from these six datasets across five common activities (Walking, Upstairs, Downstairs, Sitting, Standing).

Each dataset was divided into a training set and for test set with a 70/30% split across users, guaranteeing that different users are used for training and evaluation. For the combined dataset, we kept the train and test split from each dataset to constitute the train and test sets.

For consistency, we used similar preprocessing for all of methods. We only used accelerometer and gyroscope data from a single IMU. Signals were resampled at a fixed frame rate of 50Hz, and segmented into 2.56s windows (128 samples) with 50% overlap. We used Scikit-Learn[1] for RF and all deep learning algorithms were implemented using Keras.[2] Models were trained using the Adam optimizer with a learning rate of 0.001, with a batch size of 128, for a maximum of 100 epochs with an additional early stopping mechanism.

4.2 Activity Recognition on Individual Datasets

In this Section, we examine a comparison between the different methods on individual datasets, including all activities from each dataset. We report the accuracy on the test set of each of the methods in Table 2. Overall, the classifiers performed better with the UCI HAR dataset followed by the MotionSense dataset.

Random Forest performs relatively better than the other methods, obtaining the best scores on 4 datasets: MotionSense, PAMAP2, w-HAR and Wisdm Watch. The proposed CWT-CNN2D yielded an accuracy of 92.46% for UCI HAR and 48.4% for Wisdm Phone. The results are poor on the Wisdm dataset for all methods. PAMAP2 showed also poor performances, with an accuracy inferior to 63%.

4.3 Activity Recognition on a Combined Dataset

Table 3 reports the recognition rates obtained by the different methods, when reducing the number of activities to the ones that are common across all six datasets: laying, sitting, walking, walking upstairs and walking downstairs. We observe an increase in recognition rates for all methods. The best results were given by our proposed method CWT-CNN2D on UCI HAR, w-HAR, Wisdm Phone, and Wisdm Watch datasets, followed by RF applied on statistical features. We also note that the MotionSense dataset has the highest recognition rate, with 93.9%. CWT-CNN2D gives the best recognition rate on the Combined Dataset. In particular, this result demonstrated an increase of approximately 9% over the RF method.

[1] https://scikit-learn.org/stable/
[2] https://keras.io/

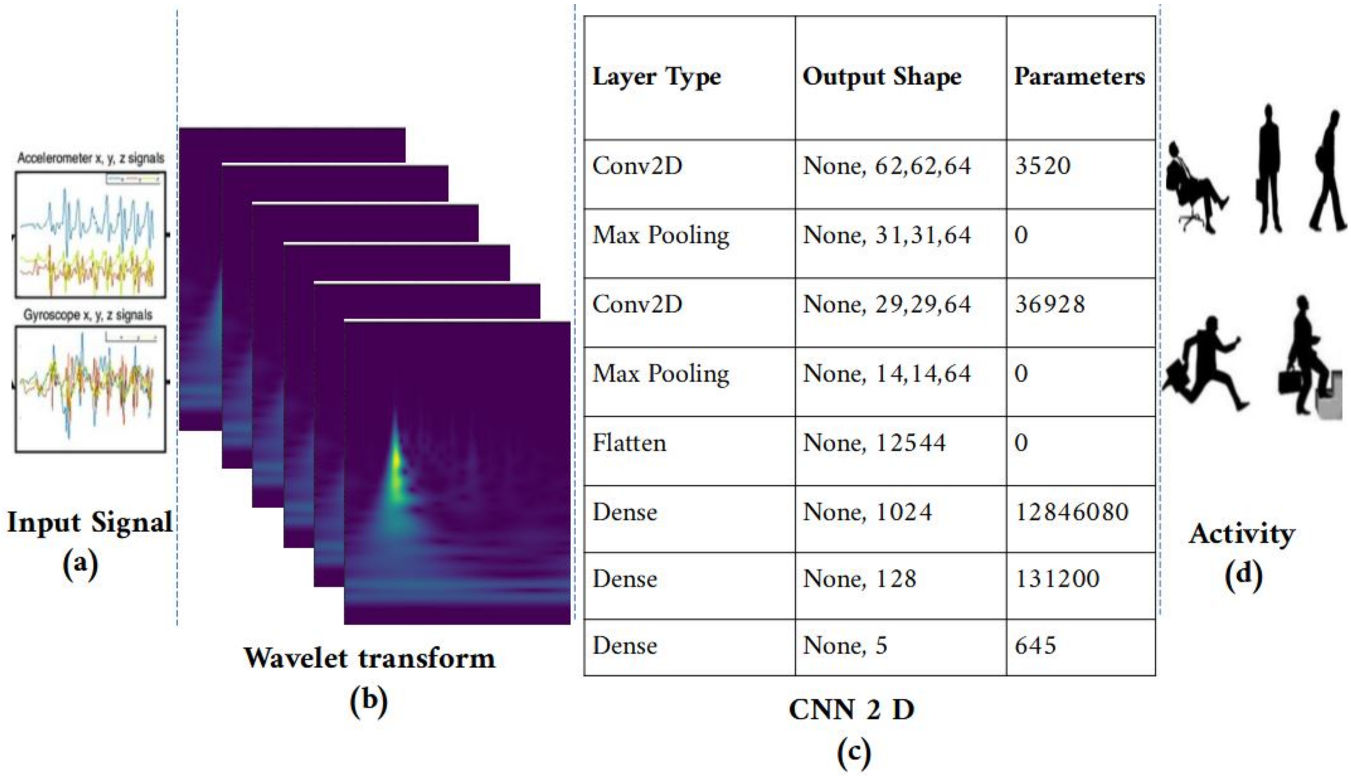

Figure 2: CWT-CNN2D framework comprises three main parts, (a) data pre-processing for sensor-based HAR (accelerometer and gyroscope); (b) continuous wavelet transform applied on data; (c) training phase, which utilizes CNN2D, and (d) activities recognition.

Table 1: HAR Datasets used for evaluation. The combined dataset was created using common activities from all six datasets.

Dataset	**Activities number**	**Sensor placement**	**Frequency**	**Users**
w-Har [7]	7	Ankle	250	22
UCI HAR[4]	6	Pocket	50	30
PAMAP2 [41]	18	Wrist	100	9
MotionSense [30]	6	Pocket	50	24
Wisdm (smartphone) [46]	18	Pocket	20	51
Wisdm (watch)[46]	18	Wrist	20	5
Combined dataset	5	multiple	100	136

Table 2: Classification results on different HAR datasets. Bold font indicates best results.

	CNN1D	**LSTM**	**CNN-LSTM**	**CWT-CNN2D**	**RF**
UCI HAR	92.2	87.5	89.8	**92.46**	86
MotionSense	82.6	80.08	86.1	88.9	**91.52**
PAMAP2	52.4	51.8	53.9	54.4	**62.17**
w-HAR	64.2	59.9	64.7	71.3	**74.33**
Wisdm Phone	29.8	26.9	40.75	**48.4**	38.52
Wisdm watch	54.2	49.7	47.4	49.4	**57.52**

Table 3: Classification results on HAR datasets, for the common activities. The last row represents the classfication results for the Combined Dataset. Bold font indicates best results.

	CNN1D	LSTM	CNN-LSTM	CWT-CNN2D	RF
UCI HAR	93.8	89.8	93	**93.9**	87.79
MotionSense	87.8	87.08	86.3	89.1	**93.9**
PAMAP2	77.1	76.1	78.1	74.3	**92.20**
w-HAR	75.2	69.8	80.8	**84.2**	78.68
Wisdm Phone	89.1	87.7	86.3	**91.8**	83.43
Wisdm Watch	64.2	63.18	69.3	**70.04**	63.17
Combined Dataset	63.8	65.2	71.4	**82.5**	74

5 DISCUSSION

We assessed the performance of the several deep learning and conventional methods to predict human activities from IMU data on separate datasets. Random Forests performs relatively better than the other methods followed by our proposed method CWT-CNN2D. The performance of the handcrafted method confirms the years of research experience in HAR community performed by standard HAR procedure: manually extracted features, feature selection and feature classification. Nevertheless, for Wisdm Phone dataset, Wisdm Watch dataset and PAMAP2, the results were relatively poor with all the methods. In fact, the Wisdm dataset have some specific characteristics. It is composed of complex activities. This complexity increases the difficulty of the recognition because of multiple transitions between different motions. Moreover, for this dataset, previous works usually considered user-dependent solutions in which, the models are trained and evaluated for a user using just her/his own data, and trained their models on a subset of activities [33]. For the PAMAP2 dataset, the poor accuracies are related to the high number of activities but also, because it includes basic movements, household activities and more unusual activities (e.g. rope jumping).

The proposed method combining the CWT with CNNs reached higher accuracy on 4 out of 6 datasets when selecting only common activities across different datasets. Interestingly, this method outperformed all other approaches on the combined dataset. This dataset offers a much larger training and testing database and embeds significant variability, because protocols for data capture differ dramatically across the original datasets. In particular, the combined dataset aggregates IMU data from different types of sensors (e.g. smartphones and smartwatches), placed on different body parts. This makes the recognition task more difficult than with isolated datasets, and requires models to further generalize. The CWT-CNN2D approach is particularly efficient with this data, and might be more robust in real-world settings where users' devices might be heterogeneous in type, quality and placement. We hypothesize that the CWT could present a better ability to capture temporal patterns while remaining invariant to positioning and direction. Yet, its performance decreases with a higher number of activities, which means that it might generalize well to new data when the number of activities is limited, but might fail to capture nuances in movement execution when working with a large number of classes. Finally, the proposed method combining CWT with 2D convolutional networks has higher computational complexity than other approaches based on 1D CNNs. Training remains efficient once scalograms are pre-computed for the entire database, and in our experiments the method is still efficient enough for real-time activity recognition, using 1s windows with 50% overlap.

6 CONCLUSION

In this work, we compared different deep learning techniques for sensor-based human activity recognition under three different scenarios: activities from individual datasets, five common activities from individual datasets and five common activities from a combined dataset. We used data from UCI HAR dataset, MotionSense dataset, PAMAP2 dataset, w-HAR dataset, Wisdm Phone dataset and Wisdm Watch dataset.

We proposed a new method combining the CWT, that provides an accurate time-frequency representation of the original signal, with 2D CNNs that have been extensively studied in image processing. The main findings from the direct comparison of our proposed method CWT-CNN2D against the baseline models is that: (i) CWT-CNN2D reaches a higher recognition rate, especially when the number of activities is small; (ii) it offers better generalization capacity illustrated by a higher recognition rate; (iii) conventional methods based on the combination between statistical feature analysis and Random Forest remain powerful even compared with modern deep learning architectures.

Technically, there is no model which achieves the best results in all tasks, so it is recommended to choose models based on the scenarios. The proposed CWT-CNN2D method appears more robust to variations in sensors placement than other techniques, yet it might remain limited when a large number of activities need to be recognized. We believe that this architecture could form the base of other movement analysis methods which require analyzing temporal dynamics independently from trajectories and directions. In future research, the proposed method will be extended to include a transfer learning approach in order to facilitate the design of movement analysis and recognition systems able to learn from few examples.

ACKNOWLEDGMENTS

This research was supported by the ELEMENT project (ANR-18-CE33-0002) from the French National Research Agency. We want to acknowledge all members of the project for fruitful discussions.

REFERENCES

[1] Paul S Addison. 2005. Wavelet transforms and the ECG: a review. *Physiological Measurement* 26, 5 (2005), R155. https://doi.org/10.1088/0967-3334/26/5/R01

[2] Tsige Tadesse Alemayoh, Jae Hoon Lee, and Shingo Okamoto. 2021. New sensor data structuring for deeper feature extraction in human activity recognition. *Sensors* 21, 8 (2021), 2814. https://doi.org/10.3390/s21082814

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez, and Jorge Luis Reyes Ortiz. 2013. 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*. 437–442. https://doi.org/10.3390/s20082200

[4] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez, and Jorge Luis Reyes Ortiz. 2013. 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*. 437–442. https://doi.org/10.3390/s20082200

[5] Mouazma Batool, Ahmad Jalal, and Kibum Kim. 2019. Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm. In *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*. IEEE, 145–150. https://doi.org/10.1109/ICAEM.2019.8853770

[6] Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, Fabrice Guédy, and Nicolas Rasamimanana. 2010. Continuous Realtime Gesture Following and Recognition. In *Gesture in Embodied Communication and Human-Computer Interaction*, Stefan Kopp and Ipke Wachsmuth (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 73–84.

[7] Ganapati Bhat, Nicholas Tran, Holly Shill, and Umit Y Ogras. 2020. w-HAR: An activity recognition dataset and framework using low-power wearable devices. *Sensors* 20, 18 (2020), 5356. https://doi.org/10.3390/s20185356

[8] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32. https://doi.org/10.1023/A:1010933404324

[9] Jason Brownlee. 2017. *Long short-term memory networks with python: develop sequence prediction models with deep learning*. Machine Learning Mastery.

[10] Jason Brownlee. 2018. *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

[11] Antonio Camurri. 2017. Computational Models of Expressive Movement Qualities in Dance. In *Music-Dance*. Routledge, 249–258.

[12] Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, and Corrado Canepa. 2016. The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. In *Proceedings of the 3rd International Symposium on Movement and Computing* (Thessaloniki, GA, Greece) *(MOCO '16)*. Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. https://doi.org/10.1145/2948910.2948927

[13] Cagatay Catal, Selin Tufekci, Elif Pirmit, and Guner Kocabag. 2015. On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing* 37 (2015), 1018–1022. https://doi.org/10.1016/j.asoc.2015.01.025

[14] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2021. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–40. https://doi.org/10.1145/3447744

[15] Eva Coupeté, Fabien Moutarde, and Sotiris Manitsaris. 2016. A User-Adaptive Gesture Recognition System Applied to Human-Robot Collaboration in Factories. In *Proceedings of the 3rd International Symposium on Movement and Computing* (Thessaloniki, GA, Greece) *(MOCO '16)*. Association for Computing Machinery, New York, NY, USA, Article 12, 7 pages. https://doi.org/10.1145/2948910.2948933

[16] G De Leonardis, Samanta Rosati, Gabriella Balestra, Valentina Agostini, Elisa Panero, Laura Gastaldi, and Marco Knaflitz. 2018. Human Activity Recognition by Wearable Sensors: Comparison of different classifiers for real-time applications. In *2018 ieee international symposium on medical measurements and applications (memea)*. IEEE, 1–6. https://doi.org/10.1109/MeMeA.2018.8438750

[17] Sarah Fdili Alaoui, Baptiste Caramiaux, M. Serrano, and Frédéric Bevilacqua. 2012. Movement Qualities as Interaction Modality. In *ACM Designing Interactive Systems (DIS'12)*. Newcastle, UK. https://doi.org/10.1145/2317956.2318071

[18] Sarah Fdili Alaoui, Jules Françoise, Thecla Schiphorst, Karen Studd, and Frederic Bevilacqua. 2017. Seeing, Sensing and Recognizing Laban Movement Qualities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'17)*. ACM, Denver, CO, USA. https://doi.org/10.1145/3025453.3025530

[19] Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. Human Model Evaluation in Interactive Supervised Learning. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 147. https://doi.org/10.1145/1978942.1978965

[20] Jules Françoise and Frédéric Bevilacqua. 2018. Motion-Sound Mapping through Interaction: An Approach to User-Centered Design of Auditory Feedback Using Machine Learning. *ACM Transactions on Interactive Intelligent Systems* 8, 2 (June 2018), 1–30. https://doi.org/10.1145/3211826

[21] Jules Françoise, Agnès Roby-Brami, Natasha Riboud, and Frédéric Bevilacqua. 2015. Movement Sequence Analysis Using Hidden Markov Models. In *Proceedings of the 2nd International Workshop on Movement and Computing (MOCO'15)*. ACM Press, Vancouver, BC, Canada, 29–36. https://doi.org/10.1145/2790994.2791006

[22] Jules Françoise, Norbert Schnell, Riccardo Borghesi, and Frédéric Bevilacqua. 2014. Probabilistic Models for Designing Motion and Sound Relationships. In *Proceedings of the 2014 International Conference on New Interfaces for Musical Expression (NIME'14)*. London, UK.

[23] Lei Gao, AK Bourke, and John Nelson. 2014. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical engineering & physics* 36, 6 (2014), 779–785. https://doi.org/10.1016/j.medengphy.2014.02.012

[24] Theodoros Georgiou, Yu Liu, Wei Chen, and Michael Lew. 2020. A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. *International Journal of Multimedia Information Retrieval* 9, 3 (2020), 135–170. https://doi.org/10.1007/s13735-019-00183-w

[25] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. 2011. Toward a minimal representation of affective gestures. *Affective Computing, IEEE Transactions on* 2, 2 (2011), 106–118. https://doi.org/10.1109/T-AFFC.2011.7

[26] Andrey Ignatov. 2018. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing* 62 (2018), 915–922. https://doi.org/10.1016/j.asoc.2017.09.027

[27] Muhammad Attique Khan, Muhammad Sharif, Tallha Akram, Mudassar Raza, Tanzila Saba, and Amjad Rehman. 2020. Hand-crafted and deep convolutional neural network features fusion and selection strategy: an application to intelligent human action recognition. *Applied Soft Computing* 87 (2020), 105986. https://doi.org/10.1016/j.asoc.2019.105986

[28] Caroline Larboulette and Sylvie Gibet. 2015. A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International Workshop on Movement and Computing*. ACM, 21–28. https://doi.org/10.1145/2790994.2790998

[29] Yong Li and Luping Wang. 2022. Human Activity Recognition Based on Residual Network and BiLSTM. *Sensors* 22, 2 (2022), 635. https://doi.org/10.3390/s22020635

[30] Mohammad Malekzadeh, Richard G Clegg, Andrea Cavallaro, and Hamed Haddadi. 2018. Protecting sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*. 1–6. https://doi.org/10.1145/3195258.3195260

[31] Stéphane Mallat. 2009. (third edition ed.). Academic Press, Boston. https://doi.org/10.1016/B978-0-12-374370-1.X0001-8

[32] Sakorn Mekruksavanich and Anuchit Jitpattanakul. 2021. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors* 21, 5 (2021), 1636. https://doi.org/10.3390/s21051636

[33] Sakorn Mekruksavanich, Anuchit Jitpattanakul, Phichai Youplao, and Preecha Yupapin. 2020. Enhanced hand-oriented activity recognition based on smartwatch sensor data using lstms. *Symmetry* 12, 9 (2020), 1570. https://doi.org/10.3390/sym12091570

[34] Sebastian Münzner, Philip Schmidt, Attila Reiss, Michael Hanselmann, Rainer Stiefelhagen, and Robert Dürichen. 2017. CNN-based sensor fusion techniques for multimodal human activity recognition. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 158–165. https://doi.org/10.1145/3123021.3123046

[35] K Muralidharan, Anirudh Ramesh, G Rithvik, Saket Prem, AA Reghunaath, and MP Gopinath. 2021. 1D Convolution approach to human activity recognition using sensor data and comparison with machine learning algorithms. *International Journal of Cognitive Computing in Engineering* 2 (2021), 130–143. https://doi.org/10.1016/j.ijcce.2021.09.001

[36] Ronald Mutegeki and Dong Seog Han. 2020. A CNN-LSTM approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 362–366. https://doi.org/10.1109/ICAIIC48513.2020.9065078

[37] Anna Nedorubova, Alena Kadyrova, and Aleksey Khlyupin. 2021. Human Activity Recognition using Continuous Wavelet Transform and Convolutional Neural Networks. *arXiv e-prints* (2021), arXiv–2106. https://doi.org/10.48550/arXiv.2106.12666

[38] Francisco Ordóñez and Daniel Roggen. 2016. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115. https://doi.org/10.3390/s16010115

[39] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. 2019. Deep learning vs. traditional computer vision. In *Science and information conference*. Springer, 128–144. https://doi.org/10.1007/978-3-030-17795-9

[40] Wen Qi, Hang Su, Chenguang Yang, Giancarlo Ferrigno, Elena De Momi, and Andrea Aliverti. 2019. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sensors* 19, 17 (2019), 3731. https://doi.org/10.3390/s19173731

[41] Attila Reiss and Didier Stricker. 2012. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*. IEEE, 108–109. https://doi.org/10.1109/ISWC.2012.13

[42] Abbas Shah Syed, Daniel Sierra-Sosa, Anup Kumar, and Adel Elmaghraby. 2021. A hierarchical approach to activity recognition and fall detection using wavelets and adaptive pooling. *Sensors* 21, 19 (2021), 6653. https://doi.org/10.3390/s21196653

[43] Christopher Torrence and Gilbert P Compo. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society* 79, 1 (1998), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

[44] Gualtiero Volpe. 2003. Computational models of expressive gesture in multimedia systems. *InfoMus Lab, DIST–University of Genova* 12 (2003). ftp://ftp.infomus.org/pub/Publications/2003/VolpeDissertation.pdf

[45] Daniel Wagner, Kathrin Kalischewski, Jorg Velten, and Anton Kummert. 2017. Activity recognition using inertial sensors and a 2-D convolutional neural network. In *2017 10th International Workshop on Multidimensional (nD) Systems (nDS)*. IEEE, 1–6. https://doi.org/10.1109/NDS.2017.8070615

[46] Gary M Weiss, Kenichi Yoneda, and Thaier Hayajneh. 2019. Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access* 7 (2019), 133190–133202. https://doi.org/10.1109/ACCESS.2019.2940729

[47] Kun Xia, Jianguang Huang, and Hanyu Wang. 2020. LSTM-CNN architecture for human activity recognition. *IEEE Access* 8 (2020), 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

[48] Tahmina Zebin, Matthew Sperrin, Niels Peek, and Alexander J Casson. 2018. Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 1–4. https://doi.org/10.1109/EMBC.2018.8513115