

Exploiting Local Temporal Characteristics via Multinomial Decomposition Algorithm for Real-Time Activity Recognition

Manuel Abbas, Régine Le Bouquin Jeannes

▶ **To cite this version:**

Manuel Abbas, Régine Le Bouquin Jeannes. Exploiting Local Temporal Characteristics via Multinomial Decomposition Algorithm for Real-Time Activity Recognition. IEEE Transactions on Instrumentation and Measurement, 2021, 70, pp.2514011. 10.1109/TIM.2021.3108235 . hal-03412612

HAL Id: hal-03412612

https://univ-rennes.hal.science/hal-03412612

Submitted on 3 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

Exploiting Local Temporal Characteristics via Multinomial Decomposition Algorithm for Real-time Activity Recognition

Manuel Abbas and Régine Le Bouquin Jeannès, *Member, IEEE*

***Abstract*—With the current expansion of Internet of Things (IoT), tracking human physical activity through wearable devices is becoming increasingly frequent. A human activity recognition (HAR) module is needed to localize and classify the patterns of acquired signals. Although this topic has been a talking point in the past decades, the trade-off between efficiency, reliability, and computational complexity is still an open research challenge. In this work, we propose a novel recognition process, based on online feature vector computation followed by a multinomial decomposition algorithm (MDA). Specifically, the temporal characteristics of performed activities are encoded over local segments. Afterwards, a low-cost algorithm divides the multinomial classification into different stages, where the encoded patterns feed a neural network at each stage to classify the corresponding segment. The proposed approach has been evaluated on five public datasets and compared with a set of state-of-the-art methods. Experimental results show that with a few hundreds of floating point operations per second (flops), our solution can achieve competitive performance in terms of accuracy. The effects of the sliding window length, the sampling frequency, the size of the training set, and overfitting have been studied. A prototype was developed to display the behavior of this solution in real world conditions.**

***Index Terms*—Human activity recognition, wearable sensors, local temporal characteristics, online computation, multinomial decomposition, computational complexity.**

I. Introduction

HUMAN activity recognition (HAR) is a wide field of research, and is currently one of the most challenging topics. The goal is to reveal and predict the successive movements of a subject, using a 4-step process, namely (a) data acquisition, (b) data segmentation, (c) feature extraction, and (d) classification. The identification of human activities serves a broad range of applications in different domains. For instance, HAR-based techniques are an instrument for assisted living approaches and smart homes [1]. Moreover, monitoring certain activities of daily living (ADLs) is the cornerstone of healthcare applications [2]. Besides, motion detection is also used in security and surveillance systems [3], in order to assist human operators and improve their work.

HAR has been a trending research area in the past decades, and continues to be an active topic [4], [5]. It can be divided into two branches, namely (i) wearable sensor-based systems [6] and (ii) ambient sensor-based systems, using cameras for data acquisition [7]–[9]. The former are light-weight devices consisting of miniaturized sensors like accelerometers, gyroscopes, magnetometers, and barometers. Furthermore, soft wearable sensors are frequently employed in the context of gesture recognition [10]. They are a solution for gait analysis [11], ensuring long-term monitoring, as well as movement tracing and rehabilitation [12]. In terms of privacy, wearable trackers do not reveal the subject's identity, thus have an advantage over the second category. Moreover, since they are body-worn, they supply data anywhere continuously, unlike ambient sensors which suffer from visibility issues and are shortened to a certain environment. Besides, camera-based HAR has its limitations like diversity in camera views, variety in environmental conditions, and variety of human subjects [13]. Ideally, the activity is identified regardless the environment and/or the performing person. Hence, wearable sensors are considered in our study.

Now, inertial measurement units (IMUs) and magnetic angular rate and gravity (MARG) sensors are usually employed for inertial human motion tracking [14]. One important factor influencing the development of a recognition process is the power consumption of the device. It is worth noting that the current consumed by a gyroscope is huge compared to other sensors [15]. Furthermore, the reliability of a magnetometer in indoor environments could be questionable, especially with the existence of artificial fields provided by different sources like smartphones and electrical materials. Another important factor is the sampling frequency (F_s) of the sensors. While a lower sampling frequency deteriorates the shape of the windowed signal, a higher one increases the computational load of the system and could not provide low latency for real-time prediction. A third factor is the length L of the window over which data are processed to predict the performed activity. Generally, HAR systems should automatically localize the targeted ADLs under unsupervised conditions to satisfy the aforementioned applications. Therefore, while larger windows may inaccurately delimit the targeted activity by containing unrelated movements, smaller windows may exclude informative segments for the recognition.

Considering all the aforementioned elements, we propose a low-cost and highly accurate acceleration-based HAR approach. Although the number of degrees of freedom (DOF) is reduced after excluding the gyroscope and the magnetometer from the acquisition unit, data acquired with

This work was supported by the French National Research Agency (ANR) in the context of the ACCORDS project (grant n° ANR-17-CE19-0024-01).

M. Abbas and R. Le Bouquin Jeannès are with INSERM U1099, Rennes, F-35000, France and with Université de Rennes 1, LTSI, Rennes, F-35000, France. (e-mail: {manuel.abbas, regine.le-bouquin-jeannes}@univ-rennes1.fr).

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

micro-electromechanical system (MEMS) accelerometers can be exploited adequately to classify the different patterns representing the ADLs. For an accurate localization, the algorithm makes a decision each second. Thanks to an efficient online computation, the feature vectors are constructed at a low cost. These vectors encode the time characteristics of acceleration signals over local segments. Afterwards, they feed a novel classification unit, called multinomial decomposition algorithm (MDA), which divides the multinomial classification into subdivisions, creating a tree structure architecture. Specifically, the windowed data are labeled at each stage using a Neural Network (NN) until they reach a leaf. We will demonstrate the effectiveness of this low-cost approach, which requires few floating point operations per second (flops). In this paper, we address some challenges and overcome the potential drawbacks of HAR, including the effect of F_s and L on the accuracy and the computational complexity of the system. The main contributions of this work are fourfold:

1) Proposal of an online feature vector computation, in order to encode the temporal characteristics of ADLs over local segments
2) Proposal of a novel low-cost classification process, based on multinomial decomposition
3) Analysis of sampling frequency and sliding window length effects on the accuracy
4) Development of a prototype which operates in real-time under real world conditions.

The remainder of the paper is organized as follows. Section II briefly reviews the related studies and describes the problem formulation, by explaining the logic behind the feature extraction and the classification strategy. The full process of the HAR mechanism is introduced in detail in section III, including the implementation of an efficient online feature computation, as well as the novel classification model, namely MDA. Then, we present an exhaustive performance evaluation in section IV before describing the developed prototype in section V and concluding the paper in section VI.

II. Towards a Robust HAR system

A. Related Work

Wearable sensor-based HAR has been a talking point in the past decades, where many studies have dealt with this research topic. This section reviews and discusses its different aspects.

Output classes: the output of a HAR module depends on the application. For instance, some studies considered a relatively low number of classes [18], [19], like level walking, going upstairs/downstairs, and staying still. Others have targeted a large number of ADLs (up to 16 classes) [20], [21], with activities like brushing teeth and cutting food. Nonetheless, human beings might perform a wide range of activities or movements in their daily routine, which are not considered by the respective models developed in these studies. Having an additional class to filter out/exclude ADLs that are not of interest is needed, but has received little to no attention in the literature [18]–[21], [25], [29]. Even though state-of-the-art methods have achieved impressive accuracy in clinical trials, their performance will drop in free-living conditions.

Wearable sensors placement: several body placements have been considered in the literature, such as the waist, the wrist, the ankle, and so on. Moreover, multi-positional body sensing was also considered [22]. The body trunk, particularly the waist or the neck, is the best position while employing one sensing device when it comes to reliability [18], [23], [37]. Additionally, the trunk acceleration patterns are meaningful for post-processing units (like postural transitions for example), unlike the wrist or the ankle which generate complex and random movements.

Feature extraction and classification: the majority of HAR modules are based on machine learning (ML) techniques [24]. Most of these methods rely on the extraction of handcrafted features from acquired data, which feed a classifier to make a decision. In the literature, time-domain and frequency-domain features were proposed such as mean, median absolute deviation, interquartile range, entropy [25]–[27], as well as the Fast Fourier Transform (FFT) coefficients and spectral power [28], [29] to name a few. These features feed a classifier to predict the activity over a window. Some low-cost classifiers, like decision trees (DT) and support vector machines (SVM) with linear/quadratic kernel, may be preferred to memory-based ones such as *k*-nearest neighbors (KNN), which raise the computational complexity of the system. These features, which are computed globally over windowed signals, encode the statistical patterns of these signals but not their temporal structures, which may lead to undesirable confusion between different classes. On the other hand, deep learning techniques were also implemented to develop HAR systems, where feature extraction is done automatically from raw signals to learn and make intelligent decisions. For instance, convolutional neural networks (CNN) [30], [33] and recurrent neural networks (RNN) [31] were proposed. The long short-term memory (LSTM) network, which is a type of recurrent networks, was also developed for this purpose [32]. LSTM addresses the vanishing/exploding gradient problem that can happen with RNNs. However, its reliability is affected by the size of the training set. Therefore, a third architecture of recurrent networks has been proposed, namely gated recurrent unit (GRU) network, which (i) is faster than LSTM, and (ii) can outperform it [34]. More recently, Chen *et al.* proposed a feature fusion technique [35]. It consists of two separate modules: (a) handcrafted features followed by fully connected layers, and (b) automatic feature learning (deep learning). Both outputs are then concatenated to predict the activity.

Sampling frequency F_s and window length L: these parameters received little attention in previous studies. These two quantities, which determine the number of points in time-series (acceleration signals), affect the feature computation (handcrafted features or those learned automatically), thus the accuracy of HAR systems.

Computational complexity and embeddability: computational complexity is a key-point to consider. According to the limited power resources of a micro-controller, the feature computation and the selected classifier should require a reasonable number of flops. The computation of certain mathematical operations like exponential needs a software library instead of a hardware implementation like the addition or the multiplication. This

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

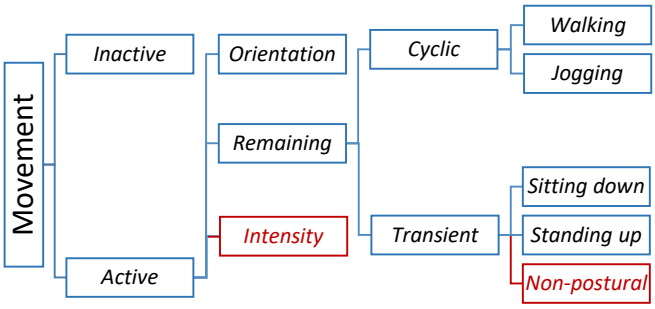


Fig. 1. The steps of the proposed HAR approach, starting from data acquisition to predict the activity (blue leaf) or exclude it (red leaf).

might increase the latency of the system. The trade-off between efficiency and computational complexity remains an open research challenge.

B. Problem Formulation

The proposed approach aims at analyzing (i) the postural transitions, (ii) cyclic/recurrent movements, and (iii) inactivity periods. Hence, our method considers basic ADLs, often encountered in daily routine of people of diverse profiles, namely (1) sitting down (walk-to-sit and stand-to-sit), (2) standing up (sit-to-stand and sit-to-walk), (3) lying down (sit-to-lie and stand-to-lie), (4) rising up (lie-to-sit and lie-to-stand), (5) walking (same level, climbing up, climbing down), (6) jogging, and (7) rest (standing, sitting, lying). For real-time recognition, a sliding window cuts the signals to make a decision, which may result in some windowed data representing random movements occurring in free-living conditions but not targeted by the system (section II-A), like the beginning/end of a cyclic activity (*i.e.* walk-to-stand, stand-to-walk, start/stop jogging). These movements constitute an eighth class called (8) others to be filtered out.

The goal behind the proposed model is to regroup the targeted activities into different clusters at each level, creating a tree structure classification process. This strategy aims at increasing the discrimination power of the system, and thus reduces the confusion between different output classes. Fig. 1 displays the proposed subdivision to label the movement. Firstly, the ADLs, which constitute the output of the proposed approach, are subject to a first-level classification into (i) inactive (*i.e.* rest) and (ii) active. If active, the second-level category is either (ii-a) change of trunk orientation (lying-down/rising-up), (ii-b) sudden change of movements intensity (rapid change in acceleration level), and (ii-c) the remaining movements. The remaining ADLs can be separated into cyclic (*i.e.* recurrent) and transient movements. ADLs reaching the cyclic node are either walking or jogging. The weightlessness state [36], occurring when the human body leaves the floor (for a short time) while jogging, and the variability of acceleration signals, are two useful quantities to discriminate between these two ADLs at this stage. Lastly, ADLs reaching the transient node are either sitting down, standing up, or some random non-postural transition. The up-down movements are symbolized by a sine wave in acceleration magnitude $||a||$. This phenomenon is scientifically interpreted in Appendix A. Moreover, the acceleration component directed towards the trunk fluctuates when the subject sits/stands. This is not the case for other similar non-postural transitions. Fig. 2 illustrates the acceleration patterns of sitting-down and standing-up.

III. Materials and Methods

This section presents the full process of the proposed HAR mechanism. It first begins by the description of local features to be extracted from acceleration signals. Next, an online feature vector computation approach is proposed to reduce the complexity of the system. Finally, the novel multinomial decomposition technique is detailed, which proves to be a low-cost and highly accurate solution. Throughout the rest of the paper, the acceleration magnitude $||a||$ is defined as follows:

$$||a|| = \sqrt{a_X^2 + a_Y^2 + a_Z^2} \tag{1}$$

where a_X is the vertical component (when the subject is standing), a_Y is the lateral component, and a_Z is pointing towards the human's trunk.

A. Encoding Temporal Structures

To employ the time characteristics of an activity, the first and second order moments of its corresponding time-series are considered, namely the mean value (μ) and the standard deviation (SD) σ:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} V_i \tag{2}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (V_i - \mu)^2} \tag{3}$$

where V is an acceleration time-series of length n, and V_i its i-th element. Since the standard deviation is proportional to the power of the signal, it quantifies the movements intensity.

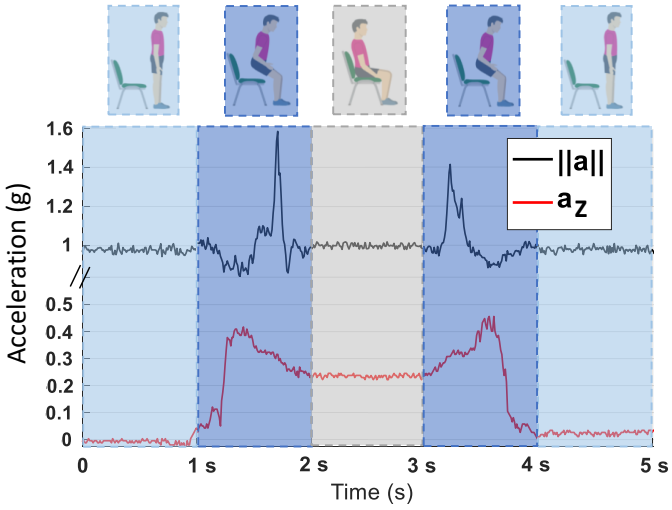


Fig. 2. The stand-to-sit and sit-to-stand (dark blue) are represented by two opposite sine waves in acceleration magnitude $||a||$. The a_Z component, directed towards the human trunk, fluctuates during these two transitions.

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

These two variables, when computed globally over the window as a whole (from end to end), determine the statistical patterns of a signal. Now, to encode the temporal structures of the windowed data, we computed these two features over local segments. Specifically, a_X and a_Z are divided into equal non-overlapping sub-segments of 1-second length to calculate their mean values. Meanwhile, $||a||$ is divided into 500-ms segments to calculate its local mean values. Successive local mean values reveal, in some way, the shape/pattern of the signal. Hence, μ is computed over smaller fragments for acceleration magnitude (500 ms for $||a||$ vs 1 s for a_X and a_Z), since it contains faster transitions, as shown in Fig. 2 (fast sine wave transition in $||a||$ vs slower deviation in a_Z). As for the intensity of movements, local SD values are calculated over 1-second segments of $||a||$, seeing that σ is sensitive to random fluctuations over very small sub-segments (like 500 ms), which might raise the confusion between the output classes.

Consequently, the succession of these mean values reveals the trunk orientation, the up-down movements, and locates the weightlessness states and the deviation in a signal. The succession of SD values detects the sudden change in movements intensity and estimates the variability of time-series. As discussed in sub-section II-B, these parameters are important to increase the discrimination power of the system. The ambiguity between different activity classes is expected to decrease considerably using the proposed local features.

B. Online Computation

The goal of this sub-section is to describe the online feature extraction technique, *i.e.* the calculation of local mean and SD values in parallel with data acquisition. Fig. 3 illustrates the process of online feature computation for a given acceleration component. The idea is to store data in m buffers $\zeta^{[i]}$ of length n, which represent the local segments over which the features are calculated, and to update the elements of a vector χ, which will be used to create the feature vectors when making a decision. With each new acquisition, *i.e.* when the accelerometer provides a new data-point, a new value ν is introduced by shifting $\zeta_j^{[i]}$ as indicated in Fig. 3, and the values χ_i are updated: $\chi_i \leftarrow \chi_i - \zeta_1^{[i]} + \zeta_n^{[i]}$. Now, let us detail the aforementioned operations.

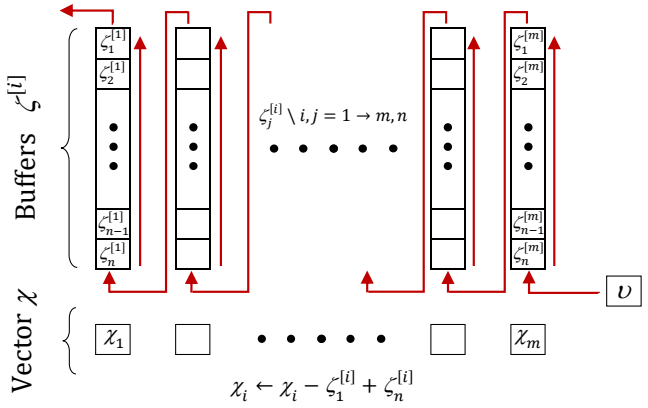


Fig. 3. The process consists of m buffers $\zeta^{[i]}$ of length n. The red arrows indicate the propagation of acquired data-points. For each new read value ν, the $\zeta_j^{[i]}$ values are shifted by one sample, $\zeta_1^{[1]}$ is excluded, and the elements χ_i are updated as indicated in the figure.

Propagation \zeta and \chi

Algorithm 1 Online Feature Computation

Define: L, F_s, $counter = 0$
Initialize: $\Lambda^{[i]}$, $\Upsilon^{[i]}$, λ, ρ are set to 0
while true **do**
  Read acquired acceleration value V_i
  $\nu \leftarrow V_i$
  Shift values $\Lambda_j^{[i]}$; $\Lambda_n^{[m]} \leftarrow \nu$
  $\lambda_i \leftarrow \lambda_i - \Lambda_1^{[i]} + \Lambda_n^{[i]}$
  $\nu \leftarrow V_i^2$
  Shift values $\Upsilon_j^{[i]}$; $\Upsilon_n^{[m]} \leftarrow \nu$
  $\rho_i \leftarrow \rho_i - \Upsilon_1^{[i]} + \Upsilon_n^{[i]}$
  $counter \leftarrow counter + 1$
  if $counter = Fs$ **then**
    Calculate μ_i and σ_i according to Eqs. (4) and (5)
    $counter \leftarrow 0$

Mean values: we consider a sliding window of length L and a sampling frequency equal to F_s. Moreover, buffers $\Lambda_j^{[i]}$ are considered to store acquired acceleration data-points V_i. $\lambda_i = \sum_{j=1}^{n} \Lambda_j^{[i]}$ is updated with each new acquisition. In this case, we get:

$$\nu \leftrightarrow V_i \quad ; \quad \zeta_j^{[i]} \leftrightarrow \Lambda_j^{[i]} \quad ; \quad \chi_i \leftrightarrow \lambda_i$$

For x-component and z-component, $m = L$ and $n = F_s$, whereas for $||a||$, $m = 2L$ and $n = \frac{F_s}{2}$. Consequently, at each decision, the mean value of the i-th segment is computed based on λ_i and $c = \frac{1}{n}$ (Eq. (2)):

$$\mu_i = c \times \lambda_i \tag{4}$$

SD values: in addition to $\Lambda_j^{[i]}$, new buffers $\Upsilon_j^{[i]}$ are considered to store V_i^2 values. Therefore, $\rho_i = \sum_{j=1}^{n} \Upsilon_j^{[i]}$ is updated with each new acquisition. Hence, we get:

$$\nu \leftrightarrow V_i^2 \quad ; \quad \zeta_j^{[i]} \leftrightarrow \Upsilon_j^{[i]} \quad ; \quad \chi_i \leftrightarrow \rho_i$$

Here, $||a||$ is involved, with $m = L$ and $n = F_s$. Consequently, at each decision, the SD of the i-th segment (i-th second) is calculated based on ρ_i, λ_i, and c (Eq. (3)):

$$\sigma_i = \sqrt{c(\rho_i - c\lambda_i^2)} \tag{5}$$

TABLE I
A SUMMARY OF THE MATHEMATICAL SYMBOLS (ONLINE COMPUTATION)

Symbol	Description
V_i	Raw acceleration value (acquired data)
$\Lambda_j^{[i]}$	A buffer to store raw acceleration values V_i
$\lambda_i = \sum_{j=1}^{n} \Lambda_j^{[i]}$	Sum of acceleration values (i^{th} second)
$\Upsilon_j^{[i]}$	A buffer to store squared acceleration values V_i^2
$\rho_i = \sum_{j=1}^{n} \Upsilon_j^{[i]}$	Sum of squared acceleration values (i^{th} second)
μ_i, σ_i	Mean and SD value of the i^{th} second respectively

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

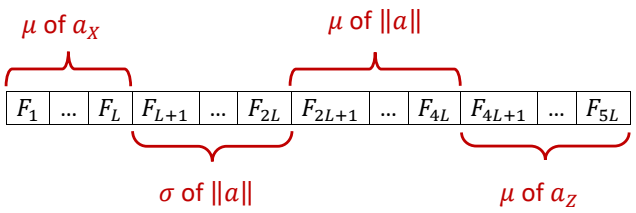


Fig. 4. The 5L extracted features using online computation.

TABLE I summarizes the mathematical symbols, and **Algorithm 1** provides the pseudo-code of this process.

C. Multinomial Decomposition Algorithm

With online extraction, $5L$ local features are calculated automatically in parallel with data acquisition, then stored in an array F. For instance, with a 5-second window ($L = 5$), 25 features are computed. Suppose that F_i is the i^{th} element of F, the extracted features are as follows (see Fig. 4):

- μ values of a_X over each second in the L-second sliding window (denoted by $F_{1\rightarrow L}$, *i.e.* first L elements of F)
- σ values of $||a||$ over each second in the L-second sliding window (denoted by $F_{L+1\rightarrow 2L}$)
- μ values of $||a||$ over each 500 ms in the L-second sliding window (denoted by $F_{2L+1\rightarrow 4L}$)
- μ values of a_Z over each second in the L-second sliding window (denoted by $F_{4L+1\rightarrow 5L}$)

We now introduce the MDA, which exploits the aforementioned features to predict ADLs.

ML Classifier: an artificial Neural Network (NN) is called at each level of the tree-structure classification process to make a prediction, until it reaches a leaf. This NN consists of one hidden layer of 10 neurons, with a customized transfer function $h'(x)$ detailed in Appendix B. It is derived from the hyperbolic tangent sigmoid $h(x)$, after introducing an expansion of exponential to reduce its computational complexity. Fig. 5 illustrates both functions, with the corresponding gap being displayed using double head arrows. Although the exponential increases faster as x increases, both curves have similar shapes and thus produce similar results (in terms of classification). It was experimentally shown (using parallel loops) that the logarithm operation is almost 25 times more complex than a multiplication. Finally, the output layer consists of a linear function.

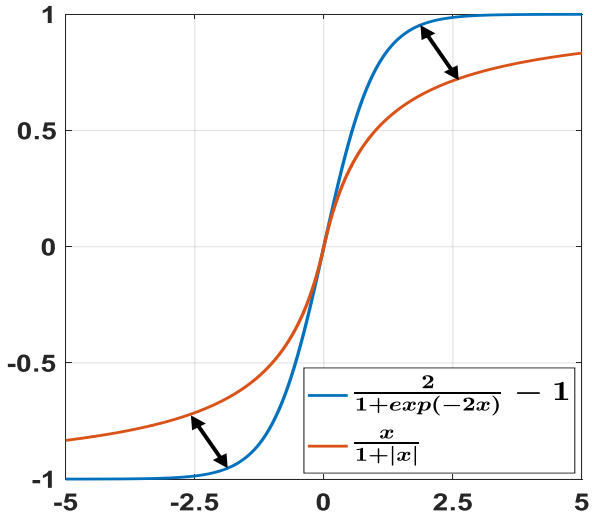


Fig. 5. Hyperbolic tangent sigmoid (blue) vs customized function (orange).

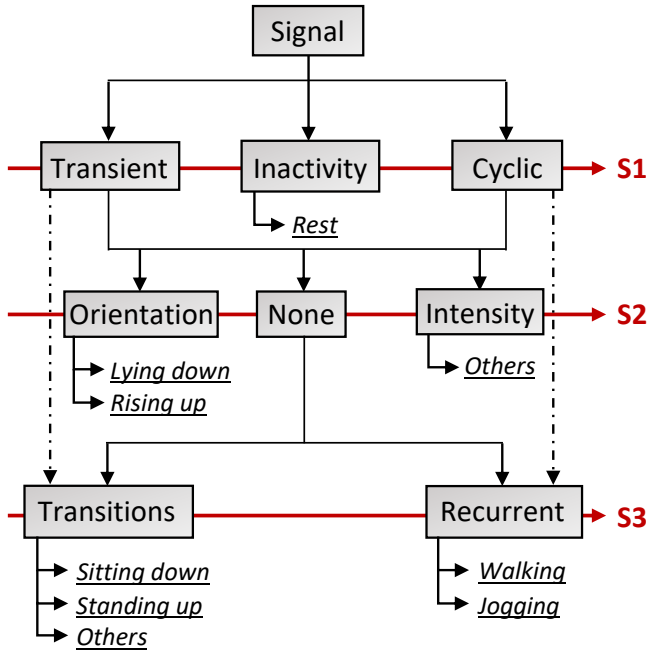


Fig. 6. The proposed multinomial decomposition algorithm, consisting of three stages S1 → S3, which predicts the performed ADL at a low-cost.

Classification by stages: MDA follows the logic of section II-B, and is illustrated, in its optimized form, in Fig. 6. The SD values of $||a||$ ($F_{L+1\rightarrow 2L}$) feed the predefined NN to assign a first-level label to the windowed data (S1), namely (I) Inactivity, (II) Cyclic, and (III) Transient. (I) is a leaf, and its corresponding output is "rest". Next, mean values of a_X and SD values of $||a||$ ($F_{1\rightarrow 2L}$) feed the NN to assign a second-level label (S2), namely (a) Orientation, (b) Intensity, and (c) None. The ADL reaching (a) is either "lying down" or "rising up", depending on the direction of the transition, whereas that reaching (b) is filtered out. Finally, the third-level stage (S3) depends on the previous attributes. For ADLs belonging to (II-c), a third NN, whose inputs are the mean and SD values of $||a||$ ($F_{L+1\rightarrow 4L}$) is applied, to see whether the subject is "walking" or "jogging". As for (III-c), the inputs of the third NN are the SD values of $||a||$ and the mean values of a_Z ($F_{2L+1\rightarrow 5L}$), to see whether the subject is "sitting down" or "standing up", or if the activity is to be excluded (non-postural transition).

IV. Performance Evaluation

To evaluate the performance of the proposed HAR approach, we resorted to five public datasets. Our method is compared with state-of-the-art techniques, particularly (a) a broad range of widely used handcrafted features, and (b) shallow learning (SL) and deep learning (DL) techniques to assess the efficiency

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

TABLE II
THE CHARACTERISTICS OF FIVE PUBLIC DATASETS IN TERMS OF EMPLOYED ACCELEROMETERS, SAMPLING FREQUENCY F_s, AND PLACEMENT OF THE DEVICE

Dataset	Accelerometer	F_s	Placement
FallAllD [37]	LSM9DS1	238 Hz	Neck, Waist
SisFall [38]	ADXL345	200 Hz	Waist
mHealth [39], [40]	Shimmer2	50 Hz	Chest
HAPT [41]	Galaxy S2	50 Hz	Waist
UMAFall [42]	MPU-9250	20 Hz	Chest, Waist

of MDA as well as that of the local temporal features. The experiments are detailed in the next sub-sections.

A. Experimental Frame

Five public datasets (see TABLE II) were used in the experiments. They employ different accelerometers with different sampling rates F_s. We resorted to these datasets to run extensive performance evaluations in order to study different aspects of a HAR system. Note that, for the five datasets, the components are adjusted in such a way that a_X represents the vertical component (oscillating around 1 g when the subject is walking) and a_Z is pointing towards the body. The corresponding data were segmented as follows.

FallAllD: the seven targeted classes (from (1) to (7)) are selected from this dataset. Moreover, some movements like falls, jumping, start/stop walking/jogging were selected randomly to constitute the eighth class "Others". These movements could occur in free-living conditions, but are to be filtered out by the algorithm. For transient ADLs, acceleration signals were segmented using 7 windows: the one centered on the transition moment, and six others shifted by $\pm 125/\pm 250/\pm 500$ ms. Afterwards, 5 out of 7 windows were chosen randomly. As for cyclic ADLs, 5 random windows were chosen to cut raw data. These operations resulted in 9100 samples consisting of eight classes: 1080 as "sitting down", 1080 as "standing up", 720 as "lying down", 720 as "rising up", 1130 as "walking", 520 as "jogging", 1095 as "rest" (inactivity), and 2755 non-targeted activities (others). The classifiers were first evaluated on this dataset using 10-fold cross validation, then trained on this dataset and tested on the remaining datasets. This choice was made since (i) FallAllD covers a wide-range of scenarios for the same ADL (like stand-to-sit and walk-to-sit for "sitting down" for example), (ii) the employed accelerometer is configured with the highest F_s (so data can be easily down-sampled), and (iii) this dataset is relatively larger to train most of DL models.

The remaining datasets: these datasets were used for validation. Here, only the seven targeted classes were considered. 3 windows were selected for each ADL. For transient activities, the centers of the 3 selected windows were randomly set between the transition moment ± 500 ms. For cyclic activities, the windows were selected randomly over the period of these movements.

Now, the proposed solution was first tested to quantify its performance (section IV-B). Next, we tested the following well-known handcrafted features from time and frequency domains, which have been shown to be effective [25], [28], [29], [40]:

time-domain: mean, SD, median absolute deviation, maximum, minimum, energy, interquartile range, and entropy for each of the three components; the first four autoregressive coefficients of $||a||$; correlation coefficients between each pair of two components; signal magnitude area.

frequency-domain: largest frequency component, weighted average, skewness, kurtosis, and energy for each component; angle between each pair of two components.

Afterwards, the classification was done using:

- classical classification strategy with one NN using the proposed feature vectors (section IV-C1).
- 9 shallow learning classifiers, namely multilayer perceptron (MLP), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), gradient boosting machine (GBM), decision tree (DT), naive Bayes (NB), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA).
- CNN using raw acceleration components.
- simple RNN, LSTM, and GRU using raw data.
- stacked autoencoder (SAE) of dense layers.
- feature fusion (FF) as in [35].

A comparative study (section IV-C) between the proposed HAR solution and the aforementioned techniques is conducted in terms of accuracy, computational load, overfitting, and training on relatively small datasets. It is worth noting that the tests were done using a set of different sampling frequencies F_s and window lengths L to understand their effect on the discrimination power of each classifier and ensemble of features. Finally, the embeddability of MDA is discussed (section IV-D). For reproducibility purposes, the architectures of DL models and the hyper-parameters of SL classifiers are provided in a pdf file attached to this manuscript.

B. MDA Performance Assessment

In this sub-section, data of FallAllD were used to validate the proposed approach. As indicated before, the achieved accuracy is reported using:
- a set of $F_s \in [20, 40, 60, 100, 238]$ Hz
- a set of $L \in [2, 3, 4, 5, 6]$ s

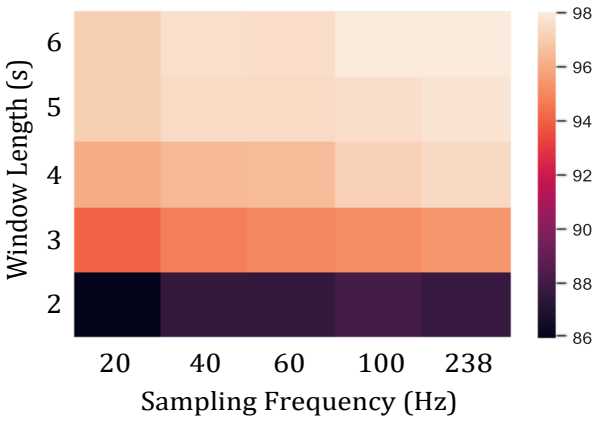


Fig. 7. The heat map consisting of the achieved accuracy (%) for each (F_s, L) pair after applying MDA.

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

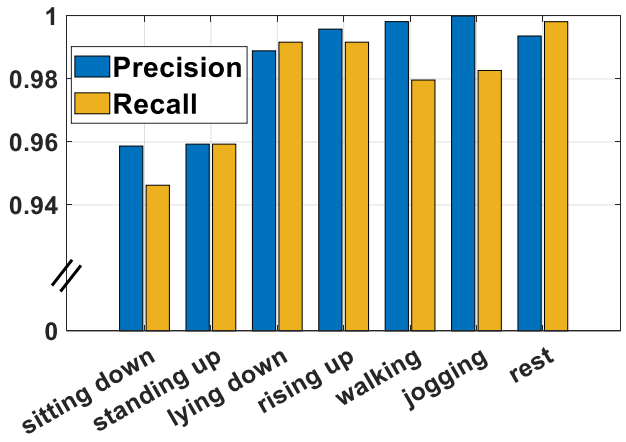


Fig. 8. The performance of MDA in terms of precision and recall for each targeted ADL using 1-vs-all strategy.

Hence, data were down-sampled, segmented using different window lengths, then cross-validated into 10 folds. The accuracy (%) for each (F_s, L) pair is illustrated in the heat map (see Fig. 7). It is clear that larger windows achieve higher accuracies (the color becomes lighter), since longer temporal structures are being encoded, thus more informative local features are available in the feature vectors. Moreover, the accuracy increases by enlarging F_s, but becomes somewhat saturated for very high frequencies ($\geq$ 100 Hz). MDA attains high accuracies (around 97-98%) which are quite satisfactory for a broad range of applications, proving its efficiency in terms of reliability. To assess the influence of the output class on the performance of MDA, precision and recall were calculated as displayed in Fig. 8. These two metrics have the lowest values (between 0.94 and 0.96) when it comes to sitting down and standing up. Now, MDA is highly sensitive to inactivity periods (highest recall) and is highly precise when it comes to cyclic ADLs.

C. Comparison with State-of-the-Art Methods

1) One stage of 8-class classification: the proposed feature vectors, *i.e.* local mean and SD values, are concatenated to form one vector of length $5L$, which constitutes the input of a NN with one hidden layer of 10 neurons. This is the case of traditional ML methods, where extracted features feed one single classifier. In this case, the accuracy drops from 98.03% to 95%. This proves the added-value of the proposed architecture of MDA. Encoding the temporal characteristics is valuable, since the achieved accuracy of 95% is still relatively high, but the way of interpreting them plays a huge role in improving the performance.

2) Deep learning vs Shallow learning: the performance of DL models, especially recurrent ones, depends on data size. Therefore, the considered DL classifiers were trained and tested on the same 9100 samples, *i.e.* training and test sets (TrSet and TeSet) were the same. Table III illustrates the accuracy of each model. It is obvious that the accuracy of RNN is poor. The same goes for LSTM, even though its performance is somewhat better. This was expected as

TABLE III
THE ACHIEVED RESULTS OF DL METHODS WHEN TRAINED AND TESTED ON THE SAME DATA

Model	CNN	RNN	LSTM	GRU	SAE
Results	96.89%	45.21%	83.31%	97.37%	97.82%

explained in section II-A. Simple RNN may suffer from vanishing/exploding gradient problem, while LSTM needs larger datasets to learn properly. However, GRU avoids both problems. CNN and SAE were also able to train, displaying quite high performance. Therefore, in a second step, only CNN, GRU and SAE were kept and a 10-fold cross validation (10-CV) was applied on FallAllD. In this step, the aforementioned nine SL techniques were also tested for comparison, considering the fifty handcrafted features extracted from windowed data (section IV-A). Additionally, the FF technique was also considered by combining GRU with handcrafted features. All these classifiers were trained then tested using all (F_s, L) pairs. Table IV illustrates the highest accuracy achieved by each model (10-CV on FallAllD), where values exceeding 98% are underlined. The proposed MDA is among the best performers, by providing competitive results. SAE is the worst DL model in terms of performance. As expected, FF outperforms GRU since it adds another dimension to the recognition process. As for SL models, KNN and GBM are the best performers with excellent accuracies.

Fig. 9 illustrates the variation in accuracy while fixing F_s and increasing L (left graph), and while fixing L and increasing F_s (right graph) for MDA, DL models and the two best SL performers (GBM and KNN). It is clear that very tight windows are not suitable for HAR. Now, for CNN, GBM, and KNN, larger windows increase the accuracy. However, for reccurent DL networks (GRU) and SAE, the ideal L is somewhere in the middle, since relatively larger windows make the accuracy decrease. Furthermore, in general, the accuracy increases while increasing F_s, except for SAE, since its accuracy is saturated around 89%. This is logical since higher F_s provides more data-points in time-series. Hence, the

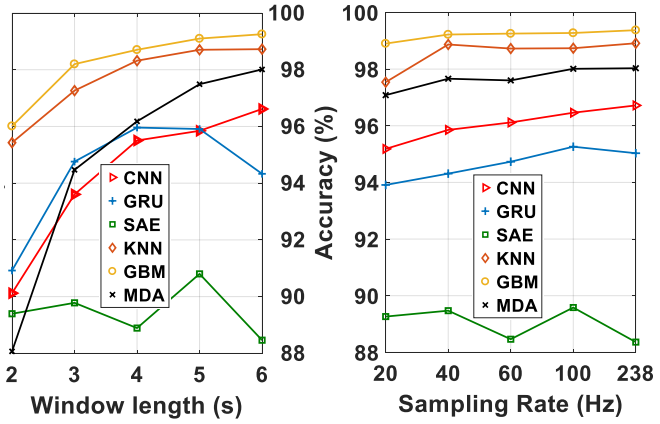


Fig. 9. The achieved accuracies (%) when (i) F_s is fixed and L is variable (left graph) and (ii) L is fixed and F_s is variable (right graph).

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

TABLE IV
THE ACHIEVED ACCURACY OF (I) 10-FOLD CROSS VALIDATION ON FALLALLD AND (II) VALIDATION ON THE REMAINING DATASETS

TrSet	*FallAllD*				
TeSet	10-CV	SisFall	mHealth	HAPT	UMAFall
MDA	98.03%	**97.75%**	**96.95%**	**94.12%**	**92%**
CNN	96.72%	70.56%	**97.26%**	**90.61%**	84.64%
SAE	90.8%	71.75%	**91.41%**	82.78%	85.79%
GRU	95.96%	83.54%	**95.21%**	**94.15%**	**90.07%**
MLP	91.26%	40.28%	72.71%	34.18%	68.7%
SVM	94.44%	38.03%	70.43%	46.46%	63.74%
KNN	99%	53.5%	81.84%	46.7%	58.77%
RF	95.82%	54.77%	87.71%	44.89%	61.43%
GBM	99.27%	51.75%	77.39%	44.56%	61.08%
DT	91.01%	42.84%	66.74%	44.06%	38.56%
NB	75%	11.12%	65.97%	36.4%	11.08%
LDA	83.4%	34.63%	83.26%	69.02%	44.22%
QDA	89.51%	23.68%	11.08%	21.25%	11.43%
FF	96.84%	58.84%	**91.19%**	55.22%	76.11%

pattern of signals delivers better knowledge to the machine. Note that, for SL models, the accuracy is somewhat the same for $F_s \geq 40$ Hz. It is worth mentioning that it has been already shown [37] that deeper CNNs are required for higher sampling frequencies, since the features extracted by the first layer is more local and thus a deeper architecture is needed to abstract them. This examination is important to choose the best configuration for a HAR system.

3) Overfitting: the overfitting phenomenon occurs frequently in ML. If the learning corresponds too closely to a particular dataset, it may fail to fit additional data or predict future observations accurately. Thus, the choice of extracted features and the classifier is delicate. Since 10-fold cross validation may not reveal overfitting, seeing that training and testing are done on data from the same sensing unit, same subjects, and under the same conditions (same clinical trial), we suggest another experiment. The aforementioned models were trained on FallAllD, then tested on the remaining datasets to assess their performance. Table IV illustrates the accuracy for each model on each dataset, where values exceeding 90% are in bold. In general, SL methods were exposed to overfitting when tested on completely unseen data. This is obvious since the handcrafted features (listed in section IV-A) were extracted globally from windowed data instead of exploiting temporal locality of signals. Hence, they reflect the statistical patterns of a signal without taking into consideration its temporal structure. Moreover, large feature vectors may mislead the machine. When it comes to DL, recurrent networks are more robust. This is also expected since they learn and remember over time sequences, contrarily to CNN for example, which exploits the spatial correlation of data. Finally, MDA avoids overfitting since it encodes local temporal characteristics, and thus provides much better results than most of the other approaches. It is worth mentioning that results on mHealth are relatively higher. The detection of recurrent movements and inactivity periods is easier than that of transitions. Besides, MDA and GRU achieved the lowest results on UMAFall since F_s is somewhat low (20 Hz). In conclusion, high accuracies may not always reflect the behavior of the system in real world conditions.

4) Computational Cost: ultimately, we measured the computational time of the proposed algorithm (MDA), DL models, and the top two SL performers. We ran a Python 3.6 implementation of the algorithms on an Intel(R) Xeon(R) Gold 5118 2.3-GHz processor with 64 GB RAM. Table V illustrates the elapsed time during the training phase and feature extraction + prediction phase on 9100 samples. KNN is the fastest when it comes to training, since it only places the labeled vectors in some metric space. However, classifiers are trained offline. Hence, this phase is not related to the latency of the system. On the other hand, it is clear that MDA provides the fastest response when it comes to predicting the ADL.This response is negligible compared to most of the other techniques. With the efficient online computation technique and the tree-structure classification process, the computational cost decreases drastically compared to the other techniques. Meanwhile, the computational load of the handcrafted feature extraction and that of recurrent network are huge. Even though GRU seems suitable for HAR and may achieve satisfactory results (see previous sections), its complexity and high latency are major drawbacks. It is worth mentioning that the resident set size (RSS), the virtual memory size (VMS), and the number of page faults (pFault) vary between [89.91, 89.96] MB, [99.95, 100.04] MB, and [29206, 29234] respectively, when MDA makes a prediction, depending on the number of accessed stages (S1, S2, and S3). With NN models being loaded, the memory usage before launching MDA is defined by RSS = 85.93 MB, VMS = 91.95 MB, and pFault = 26638.

D. Embeddability

The elapsed time to extract features and make a prediction is useful to evaluate the relative computational cost of the approaches. However, it does not reveal the ability of being embedded in wearable devices, seeing the limited power resources of a micro-controller. Consequently, Appendix C shows in detail the number of operations required at each stage of MDA. Since $\alpha = \beta = \gamma = 1$ flop (micro-controllers featuring FPU single precision), the computational complexity is equal to $\Xi = 12d + 11k + 30$ flops, where d and k are the length of feature vector and the number of output classes respectively. Suppose that $L = 5$, the complexity of each stage (leaf) is:

$$\{\text{'S1': 123, 'S2': 183, 'II-c': 232, 'III-c': 243}\}$$
$$\longrightarrow 123 \text{ flops (rest)} \leq \Xi \leq 549 \text{ flops (transitions)}$$

With only few hundreds of flops, MDA is able to provide a

TABLE V
THE TIME NEEDED (IN S) FOR TRAINING, FEATURE EXTRACTION, AND PREDICTION.

Method	**Training**	**Feature Extraction + Prediction**
MDA	15.12 s	0.173 s
CNN	34.16 s	2.54 s
SAE	10.46 s	0.55 s
GBM	107.13 s	23.018 s
KNN	0.55 s	24.235 s
GRU	4909.2 s	26.12 s
LSTM	4991.3 s	29.17 s
FF	4965.4 s	26.22 s

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

prediction. Note that these three types of operations are done in one cycle using a hardware implementation (which is almost negligible), without any need for a software library (the case of exponential and logarithm for example).

V. Prototype

A prototype was developed to illustrate the behavior of the proposed HAR system in real world conditions. The corresponding device was developed by RF-Track society in Cesson-Sévigné, France. LIS3DH from STMicroelectronics is the MEMS 3D accelerometer. It has dynamically user selectable full scales of $\pm$2g/$\pm$4g/$\pm$8g/$\pm$16g and it is capable of measuring accelerations with output data rates ranging from 1 Hz to 5.3 kHz. It provides a 16-bit output interface. Additionally, it provides a data rate configuration of $\{1, 10, 25, 50, 100, 200, 400\}$ Hz and $\{1.60, 1.344, 5.376\}$ kHz. Raw data are pre-processed by the ultra-low-power micro-controller STM32L431 from STMicroelectronics (with FPU) and saved on a micro SD card for further analysis. Each component of the 3D accelerometer output is multiplied by $\frac{4}{16384}$ to convert it to g units.

As previously mentioned, a decision is taken every second by the micro-controller. Once connected to an electronic device like mobile phone or tablet, and thanks to a developed mobile application, the user can display the performed ADLs instantaneously. The device sends immediately the label of the predicted activity to the electronic device. Now, if a decision is made each second, a total of 86400 predictions are done under 24 hours. With 98% accuracy (which is quite high), 1728 windows are misclassified if we want to consider the ADL of each sliding window. However, the goal of a HAR system in real world conditions is the accurate localization of ADLs (postural transitions and cyclic movements) for an ideal post-processing, and not blind recognition. Therefore, to ensure a robust output, the application applies a voting technique on the last three outputs after each prediction. The result of this operation is displayed on the screen. For example, if {(5), (6), (5), (5), (1), (1)} is the list of decisions:

$$\begin{cases} \{(5),(6),(5)\} \rightarrow display\,(5) : Walking \\ \{(6),(5),(5)\} \rightarrow display\,(5) : Walking \\ \{(5),(5),(1)\} \rightarrow display\,(5) : Walking \\ \{(5),(1),(1)\} \rightarrow display\,(1) : Sitting\,down \end{cases}$$

Note that, if the input of this voting technique is three different labels, or if the windowed data is filtered out, the screen shows a dash (—). A subject has worn the device for 1 minute and 3 seconds. A tool has been developed in MATLAB in order to show the acquired signals using Time Scope, to apply the proposed algorithm, and to illustrate the label of the activity as previously described. A video displaying this process is attached to this manuscript. With a 4-second sliding window and F_s=50 Hz, 59 predictions were made by the proposed algorithm. 48 windowed data were correctly classified, 8 were misclassified, and the remaining 3 windows were wrongly filtered out (classified as "Others"). Nevertheless, the effect of the 8 misclassifications is not visible in the visualization process, since the interface displays the correct ADL instantaneously, which shows the efficiency and the added-value of the proposed voting technique. The result of each prediction (at each second) is meaningless in real world conditions. What matters most is the summary of successive ADLs over a certain period of time and their localization based on acquired data.

VI. Conclusion

This paper addressed the challenges of a HAR system. By encoding local temporal characteristics of acquired acceleration signals, ADLs are accurately localized and recognized. The system presents low computational cost thanks to an efficient online feature extraction technique and a novel classification architecture. Extensive performance evaluations on five public datasets have proved the effectiveness of the proposed solution named MDA, which outperforms start-of-the-art handcrafted features and ML techniques in terms of accuracy, computational load, and capacity of avoiding overfitting. Furthermore, the sampling frequency and the window length were discussed to understand their effect on the accuracy of the system. The feasibility of MDA in real world conditions has been also investigated, by developing a prototype which accurately predicts the performed activity in real-time.

In a future work, barometric signals will be investigated as a post-processing tool. These data give another dimension to the HAR system, namely the altitude of the human body, which is an important feature for several applications.

Appendix A
Up-Down Acceleration Pattern

Let us assume that MEMS accelerometers are equivalent to an ideal case of single spring and single mass system without any force opposing the motion of the components. When we stand up, the spring is released from the equilibrium point with an upward velocity $-V_0$. Hence, two forces are acting upon the system, namely (a) the restoration force of the spring $F_s = kX$, and (b) the weight due to gravity $W = m_0 g$. Note that $X = x + s$ thus $F_s = kx + ks$: ks the restoration force of the spring opposing the pulling of gravity and kx the restoration force of the spring trying to get back to the equilibrium point. By applying Newton's second law:

$$\sum F = -kx + m_0 g - ks = m_0 a \tag{6}$$

Now, by setting the equilibrium as the reference point, *i.e.* $-ks + m_0 g = 0$, Eq. (6) could be rewritten:

$$\ddot{x} + \left(\sqrt{\frac{k}{m_0}}\right)^2 x = 0 \tag{7}$$

Assuming that $\sqrt{\frac{k}{m_0}} = w_0$, the general solution is:

$$x(t) = C_1 \cos(w_0 t) + C_2 \sin(w_0 t) \tag{8}$$

The mass was released from the equilibrium point:

$$x(t) = x_0 \sin(w_0 t)\,,\ x_0 < 0 \tag{9}$$

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

By replacing $x(t)$ in (7) by its value in (9), we get:

$$a(t) \sim a_0 \sin(w_0 t),\ a_0 > 0 \tag{10}$$

Note that for "sitting down", we get the same acceleration with an opposite sinusoidal function.

APPENDIX B
CUSTOMIZED TRANSFER FUNCTION

The Taylor series for exponential $\exp(u)$ (where $u > 0$) is:

$$\exp(u) = \sum_{i=0}^{+\infty} \frac{u^i}{i!} = 1 + u + ... \tag{11}$$

The hyperbolic tangent sigmoid transfer function $h(x)$ is:

$$h(x) = \frac{2}{1 + \exp(-2x)} - 1 \tag{12}$$

By replacing $\exp(u)$ by $1+u$ (order 1) in $h(x)$ we obtain the customized transfer function $h'(x)$:

$$x > 0 : h(x) = \frac{2}{1 + \frac{1}{\exp(2x)}} - 1 = \frac{2}{1 + \frac{1}{1+2x}} - 1 = \frac{x}{1+x}$$

$$x < 0 : h(x) = \frac{2}{1 + \exp(-2x)} - 1 = \frac{2}{1 + 1 - 2x} - 1 = \frac{x}{1-x}$$

$$\longrightarrow h'(x) = \frac{x}{1 + |x|} \quad \forall x \in \mathbb{R} \tag{13}$$

APPENDIX C
MDA COMPUTATIONAL COMPLEXITY

The addition/subtraction α, multiplication/division β, and multiply-and-accumulate γ are considered. The "sign" operation is free at the micro-controller level.

The NN begins by scaling the feature vector $\mathcal{F}$ of length d using the min-max normalization:

$$\hat{\mathcal{F}} = \frac{\mathcal{F} - min(\mathcal{F})}{max(\mathcal{F}) - min(\mathcal{F})} \sim a \times \mathcal{F} + b \tag{14}$$

This step requires $\xi_1 = d\alpha + d\beta$ operations.

Afterwards, $\hat{\mathcal{F}}$ constitutes the input of the hidden layer (fully-connected) of 10 neurons. Hence, it is the subject to the following transformation:

$$\hat{\mathcal{F}}' = \mathbf{M} \times \hat{\mathcal{F}}^{\intercal} + B_h \tag{15}$$

where $\mathbf{M}$ and B_h are of size $10 \times d$ and 10×1 respectively. This step results in $\xi_2 = 10d\gamma + 10\alpha$ operations.

Subsequently, the customized transfer function $h'(x)$ is applied on $\hat{\mathcal{F}}'$, resulting in $\widetilde{\mathcal{F}}$ as output while involving $\xi_2 = 10\alpha + 10\beta$ operations.

Finally, $\widetilde{\mathcal{F}}$ feeds the output layer to finalize the k-class classification process:

$$Out = \mathbf{H} \times \widetilde{\mathcal{F}} + B_o \tag{16}$$

where $\mathbf{H}$ and B_o are of size $k \times 10$ and $k \times 1$ respectively. Here, $\xi_4 = 10k\gamma + k\alpha$ operations are required. Consequently, each stage of MDA requires $\Xi = \sum_i \xi_i$ operations:

$$\Xi = \alpha(d + 20 + k) + \beta(10 + d) + 10\gamma(d + k)$$

REFERENCES

[1] G. Okeyo, L. Chen, and H. Wang, "Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes," in *Future Generation Computer Systems,* vol. 39, pp. 29-43, 2014.

[2] X. Zhou, W. Liang, K. I. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-Learning-Enhanced Human Activity Recognition for Internet of Healthcare Things," in *IEEE Internet of Things Journal,* vol. 7, no. 7, pp. 6429-6438, Jul. 2020.

[3] S. Huang, "An Advanced Motion Detection Algorithm With Video Quality Analysis for Video Surveillance Systems," in *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 21, no. 1, pp. 1-14, Jan. 2011.

[4] G. Panahandeh, N. Mohammadiha, A. Leijon, and P. Händel, "Continuous Hidden Markov Model for Pedestrian Activity Classification and Gait Analysis," in *IEEE Transactions on Instrumentation and Measurement,* vol. 62, no. 5, pp. 1073-1083, 2013.

[5] Y. Zhang, G. Tian, S. Zhang, and C. Li, "A Knowledge-Based Approach for Multiagent Collaboration in Smart Home: From Activity Recognition to Guidance Service," in *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 2, pp. 317-329, 2020.

[6] S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," *IEEE Sensors Journal,* vol. 15, no. 3, pp. 1321-1330, March 2015.

[7] K. Li and Y. Fu, "Prediction of Human Activity by Discovering Temporal Sequence Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, no. 8, pp. 1644-1657, Aug. 2014.

[8] X. Yang and Y. Tian, "Super Normal Vector for Human Activity Recognition with Depth Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 5, pp. 1028-1039, 2017.

[9] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang and Q. Tian, "Symbiotic Graph Neural Networks for 3D Skeleton-based Human Action Recognition and Motion Prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* doi: 10.1109/TPAMI.2021.3053765.

[10] J. Kwon, J. Park, S. Ku, Y. Jeong, N. Paik, and Y. Park, "A Soft Wearable Robotic Ankle-Foot-Orthosis for Post-Stroke Patients," in *IEEE Robotics and Automation Letters,* vol. 4, no. 3, pp. 2547-2552, Jul. 2019.

[11] M. Gholami, A. Rezaei, T.J. Cuthbert, C. Napier, and C. Menon, "Lower Body Kinematics Monitoring in Running Using Fabric-Based Wearable Sensors and Deep Convolutional Neural Networks," in *Sensors,* 19(23):5325, Dec. 2019.

[12] S. Bao, S. Yin, H. Chen, and W. Chen, "A wearable multimode system with soft sensors for lower limb activity evaluation and rehabilitation," in *Proc. IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* pp. 1-6, 2018.

[13] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.Y. Duan and A.C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 10, pp. 2684-2701, Oct. 2020.

[14] S.O.H. Madgwick, A.J.L. Harrison and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," *IEEE International Conference on Rehabilitation Robotics,* pp. 1-7, 2011.

[15] Q. Liu *et al.*, "Gazelle: Energy-efficient wearable analysis for running," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2531–2544, Sep. 2017.

[16] R.W Kressig *et al.*, "Temporal and spatial features of gait in older adults transitioning to frailty," *Gait & Posture,* Vol. 20, no. 1, pp. 30-35, 2004.

[17] S. Parvaneh, J. Mohler, N. Toosizadeh, G.S. Grewal, and B. Najafi, "Postural Transitions during Activities of Daily Living Could Identify Frailty Status: Application of Wearable Technology to Identify Frailty during Unsupervised Condition," *Gerontology,* , vol. 63, no. 5, pp. 479-487, 2017.

[18] Y. Ohtaki *et al.*, "Automatic classification of ambulatory movements and evaluation of energy consumptions utilizing accelerometers and a barometer," *Microsyst Technol,* vol. 11, pp. 1034–1040, Aug. 2005.

[19] A. B. Mesanza, S. Lucas, A. Zubizarreta, I. Cabanes, E. Portillo and A. Rodriguez-Larrad, "A Machine Learning Approach to Perform Physical Activity Classification Using a Sensorized Crutch Tip," in *IEEE Access,* vol. 8, pp. 210023-210034, 2020.

[20] A. Moncada-Torres, K. Leuenberger, R. Gonzenbach, A. Luft, and R. Gassert, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," *Physiol Meas.,* vol. 35, no. 7, pp. 1245-1263, Jul. 2014.

[21] A.G. Bonomi, A.H. Goris, B. Yin, K.R. Westerterp, "Detection of type, duration, and intensity of physical activity using an accelerometer," *Med Sci Sports Exerc.,* vol. 41, no. 9, pp. 1770-1777, Sep. 2009.

ACCEPTED MANUSCRIPT - CLEAN COPY

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2021.3108235, IEEE Transactions on Instrumentation and Measurement

[22] P. Bharti, D. De, S. Chellappan and S.K. Das, "HuMAn: Complex Activity Recognition with Multi-Modal Multi-Positional Body Sensing," *IEEE Transactions on Mobile Computing,* vol. 18, no. 4, pp. 857-870, Apr. 2019.

[23] I. Cleland *et al.*, "Optimal Placement of Accelerometers for the Detection of Everyday Activities," *Sensors*, vol. 13, no. 7, pp. 9183-9200, Jul. 2013.

[24] Z. Wang, Z. Yang, T. Dong, "A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital Signs in Real Time," *Sensors (Basel)*, vol. 17, no. 2, Feb. 2017.

[25] 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.

[26] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomput.*, vol. 171, pp. 754–767, 2016.

[27] L. Bao and S.S. Intille, "Activity recognition from user-annotated acceleration data," *Lect. Notes Comput. Sci.,* vol.3001, pp.1-17, 2004.

[28] A.G. Bonomi, A.H. Goris, B. Yin, and K.R. Westerterp, "Detection of type, duration, and intensity of physical activity using an accelerometer," *Medicine and Science in Sports and Exercise,* vol. 41, no. 9, pp. 1770-1777, Sep. 2009.

[29] L. Xie, J. Tian, G. Ding and Q. Zhao, "Human activity recognition method based on inertial sensor and barometer," in *Proc. IEEE International Symposium on Inertial Sensors and Systems (INERTIAL),* pp. 1-4, 2018.

[30] M. Zeng *et al.*, "Convolutional Neural Networks for human activity recognition using mobile sensors," *6th International Conference on Mobile Computing, Applications and Services,* pp. 197-205, 2014.

[31] S. Li, C. Li, W. Li, Y. Hou, and C. Cook, "Smartphone-sensors Based Activity Recognition Using IndRNN," In *Proc. ACM International Joint Conference and International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18),* Association for Computing Machinery, New York, NY, USA, pp. 1541–1547, Oct. 2018.

[32] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications,* pp. 50-53, 2016.

[33] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *Proc. International Joint Conference on Neural Networks (IJCNN),* 2016, pp. 381-388.

[34] S. Yang, X. Yu and Y. Zhou, "LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example," in *Proc. International Workshop on Electronic Communication and Artificial Intelligence (IWECAI),* pp. 98-101, 2020.

[35] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu and X. Li, "Smartphone Sensor-Based Human Activity Recognition Using Feature Fusion and Maximum Full a Posteriori," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 7, pp. 3992–4001, July 2020.

[36] Z. He, Z. Liu, L. Jin, L. Zhen, and J. Huang, "Weightlessness feature — a novel feature for single tri-axial accelerometer based activity recognition," in *Proc. 19th International Conference on Pattern Recognition,* pp. 1–4, 2008.

[37] M. Saleh, M. Abbas and R. Le Bouquin Jeannès, "FallAllD: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning Applications," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1849-1858, Jan. 2021.

[38] A. Sucerquia, J. López, and J. Vargas-Bonilla, "SisFall: A fall and movement dataset," *Sensors*, vol. 17, no. 12, p. 198, Jan. 2017.

[39] O. Banos *et al.*, "mHealthDroid: A Novel Framework for Agile Development of Mobile Health Applications," in *Proc. of IWAAL*, Belfast, Northern Ireland, Dec. 2014.

[40] O. Banos *et al.*, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *BioMedical Engineering OnLine*, vol. 14, pp. 1-20, 2015.

[41] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. ESANN*, 2013, pp. 437–442.

[42] E. Casilari, J.A. Santoyo-Ramón, J.M. Cano-García, "UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection," *Procedia Computer Science*, vol. 110, pp. 32-39, 2017.



Manuel Abbas received the Engineering degree in electronics and telecommunications from the Faculty of Engineering, Lebanese University, Branch II, Roumieh, Lebanon, in 2018, and the master's degree in signal and image processing from Ecole Centrale de Nantes, France, in 2018. He is currently pursuing the Ph.D. degree with LTSI (Laboratoire Traitement du Signal et de l'Image), Inserm U1099, Université de Rennes 1, France. His research activities mainly focus on signal processing and machine learning in the field of activity and health monitoring.



Régine Le Bouquin Jeannès (Member, IEEE) received her Ph.D. degree in Signal Processing and Telecommunications from Rennes University, France, in 1991. She is currently Professor in LTSI (Laboratoire Traitement du Signal et de l'Image, Inserm U1099) at Rennes University, France, and belongs to CRIBs (Centre de Recherche en Information Biomédicale sino-français), a French-Chinese laboratory associated with Inserm (Institut national de la santé et de la recherche médicale). Her research activities mainly focus on biomedical signals processing and modeling in the field of epilepsy and brain connectivity as well as on health monitoring and telecare.

ACCEPTED MANUSCRIPT - CLEAN COPY