EasyChair Preprint

№ 8478

Deep Learning for Human Activity/Action Recognition Based Sensor and Smartphone

Youssef Errafik, Adil Kenzi and Younes Dhassi

July 16, 2022

Deep Learning for Human Activity/ Action Recognition based Sensor and Smartphone

Youssef Errafik[1*], Adil Kenzi [1], Younes Dhassi [1]

[1] Sidi Mohamed Ben Abdellah University,
Fes, Morocco
youssef.errafik@usmba.ac.ma

Abstract. In recent years, the field of sensor-based Human Action Recognition (HAR) has become one of the most developed research areas, benefiting from the evolution and availability of electronic devices in our daily lives, and from the exponential evolution of artificial intelligence (AI) as well. In this context, its powerful advances are constantly exploited by researchers to develop useful methods for obtaining the best performance needed to solve existing challenges, thus, putting its contributions in favor of medical applications, taking the example of the intelligent monitoring field which is characterized by automatic, continuous, remote, and real-time monitoring of the actions of the elderly, pregnant women at home, and even hospitalized patients suffering from all kinds of mental and behavioral disorders caused by neurodegenerative diseases such as Alzheimer's, Parkinson's disease and various types of addiction. The opportunities, as well as the advantages offered by HAR, are then largely exploited in medicine and other fields such as robotics (human-computer interaction), sports discipline, and many others. Thanks to the different architectures of recurrent neural networks, deep learning (DL) has proven its effectiveness and robustness in the fields of AI and computer vision to the extent of being equal to or even exceeding human skills in terms of particular tasks such as speech recognition, language translation, pattern recognition, and in particular HAR that are going to be covered in this article. Our goal is to examine and compare the performance of some pivotal approaches in the field of human activity recognition based on smartphone sensors. We analyzed both LSTM and GRU models. In order to ensure equality of treatment and to obtain a reliable comparison, the implementation of these architectures is carried out on several sets of data. The experimental results obtained relate to the regular indicative measurements of the HAR domain.

Keywords: Human activity/action Recognition (HAR); Deep Learning (DL); Convolutional Neural Network (CNN); Long Short-Term Memory (LSTM); Accelerometer and gyroscope sensor; time-series; Recurrent Neural Networks (RNN); Gated Recurrent Unit (GRU); Smartphone; Sensors

1 Introduction

The main objective of sensor-based human activity recognition is to identify human activities performed automatically from the data sequences sent by electronic sensors such as cameras, wearable devices, smart watches, smartphones and others.

According to the history (and literature) of the development of human action recognition, activity recognition has emerged as a sub-branch of human motion analysis that has received increasing attention from the computer vision scientific community.

Researchers began by using devices attached to the actors' bodies to capture signs and record them as behavioral data of the actors.

There are two main branches in this line of research:

(1) Video-based action/activity recognition

- The first one is an image-based approach that respects the computer vision approach, which uses only cameras and video devices for the detection and the classification of human activities in automated recognition systems. This approach uses RVG image sequences, depth maps and skeleton sequences.

(2) Sensor-based action/activity recognition

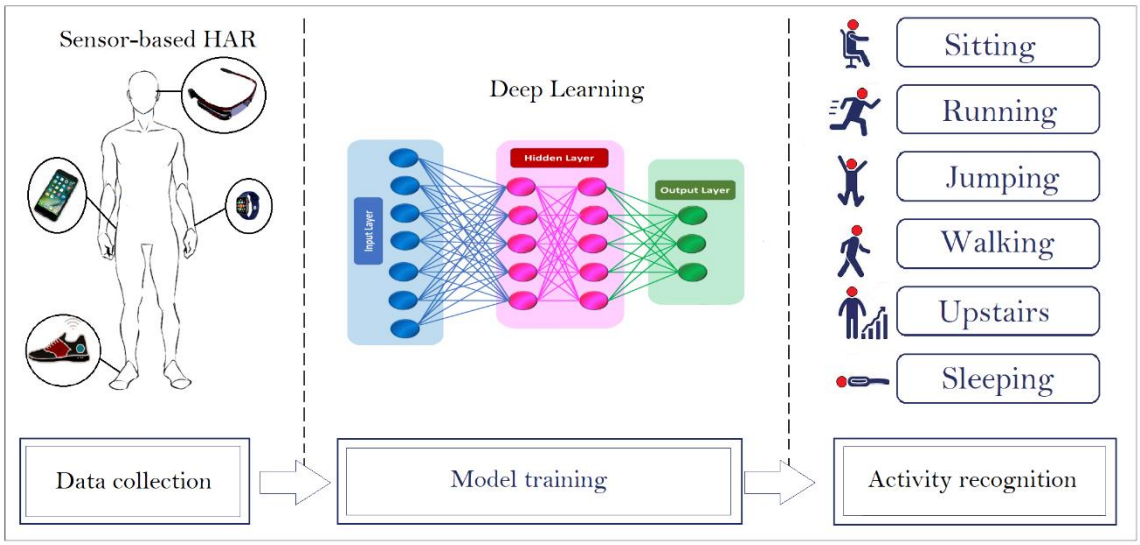


Fig. 1 The Human Activity Recognition (HAR) Sensor Framework consists of three main steps: (a) data collection for HAR using sensors; (b) formation of the established deep learning model, and (c) recognition (classification) of activities.

- As shown in Fig. 1. The second one named "Sensor-based human activity recognition" which exploits other sensors than the image such as Accelerometer, Gyroscope, Bluetooth, etc...

To determine human activity, systems with the ability to detect and classify human activities using one of three categories of sensors:

- Wearable sensor-based activity recognition,
- Activity recognition that is based on environmental/ambient sensors, and
- Activity recognition based on smartphone sensors

The work of researchers in the field of HAR always consists of addressing the persistent challenges that are detected for each type of device and sensor and then proposing new methods or improving the existing ones, and through implementation and experimentation, the advances are approved its performance in the laboratory, then in the field, they are integrated into different systems to exploit its high performance possible and another time the researchers resume the cycle by determining the new problems and challenges .

Although classical human action recognition approaches using machine learning methods have previously worked in a successful manner, these methods still rely heavily on human intervention. Therefore, both the design and the extraction of features manually are manually limited by human domain awareness, hence the performance of its approaches is limited in terms of accuracy and efficiency.

Hence, the appearance and realization of deep learning (DL) represents a technological revolution in artificial intelligence thanks to its design based on the artificial neural network (ANN) that resembles the human brain. The structure of the rest of this paper is as follows. In Section 2, we present related work on human activity recognition using sensors. In section 3, we describe the methodology followed to solve our work. In section 4, we describe the datasets and architectures (models) used in our work. In Section 5, is dedicated to the detailed discussion of the results of our experiments. Finally, Section 6 concludes the paper.

2 Related work

Mobile sensor-based HAR exploits the opportunities of machine learning algorithms in healthcare to determine the activities of elderly people with Parkinson's disease (PwP) through the use of three-dimensional accelerometer and gyroscope signals extracted by wearable sensors [1]. According to the survey [2] that analyzes the various systems based on HAR smartphones, it is found that the accuracy alone is insufficient to confirm the effectiveness of human activity recognition system tested on a specific data set pre-processed in the laboratory, that is, the number and the type of activities are well chosen, as well as the size of the population and the characteristics extracted by the researchers.

Hence the application of the so-called effective methods in the laboratory in real situations can be frustrating and catastrophic because, on the one hand, of changing experimental conditions such as orientation and location

of sensors on the body, and on the other hand the most accurate systems are usually very expensive in computation and memory and can take a long time to return the result of activity recognition, which is not always acceptable in real condition.

After the huge success of classical convolutional neural networks in various computer vision tasks, where the input is mainly a 2D image or a sequence of images. In [14]. Duffner, Berlemont, Lefebvre and Garcia (2014) attempted to introduce this type of convolutional neural networks (convnets) in human activity recognition using raw accelerometer and gyroscope signals, exploiting the effect and the ability of CONV networks to extract automatically features without human intervention.

Inspiration from researchers that are specialized in this branch and who concluded that convolutional neural networks could surpass several other methods of recognizing human gestures. And After these early experiments, Ronao and Cho[15] succeeded in building a powerful convolutional neural network (convnet) that operates both ways; as an automatic feature extractor and as a classifier from the raw sensory data of smartphones to recognize the activity (HAR) performed by the wearer of these devices.

Despite the multiple experiments to solve the problem of classification of human activities through convolutional neural networks CNN, researchers are convinced that this kind of network has reached its limits and that it does not have the necessary ability to extract the dependence on long-term time series, making it difficult even impossible to optimize its performance.

The RNN architecture is a great technological achievement that solves time series problems since it ensures that the system is able to learn the temporal context of the Input sequences in order to make highly accuracy predictions. After its great success in Natural language processing (NLP) [3] and Speech recognition [4-5], the Recurrent Neural Network (RNN) is approved for its performance in exploiting the temporal correlations between neurons in human activity recognition.

In [6], they proposed a structure of the recurrent neural network (RNN) named Residual-RNN, which allows predicting activities of residents from a sequence of data collected by sensors in the domestic setting.

Using the Massachusetts Institute of Technology (MIT) real-world dataset [7] and thanks to its attention mechanism, this model outperformed the two famous models: the short-term memory model (LSTM) and closed recurrent unit (GRU) in terms of the accuracy of classifying human activities in the smart home.

Bengio. et al [8-9] have shown through theoretical and experimental evidence that gradient descent causes great difficulty in RNN model training for tasks characterized by long-term dependencies. His work focuses mainly on the two problems: gradient leakage and gradient explosion.

After the considerable evolution of RNN structures, the emergence of LSTM model [10] to solve the various problems intractable by RNN mainly those concerning complex artificial tasks a with long lag. This combination of RNN and LSTM architectures has proven to be successful and cost-effective in cursive handwriting recognition and speech recognition [11].

In the work [12], to improve the efficiency and sensitivity of their HAR system based on LSTM-RNN model, they combined the image-based localization data with those of several sensors in the form of 3D accelerometer and 3D gyroscope signals. Thanks to this combination, disturbing activities will be eliminated which positively influences the recognition accuracy.

In the work [23], they presented a Long-Short Term Memory (LSTM) deep recurrent neural network model to classify activities of daily living from tri-dimensional accelerometer and gyroscope data. They managed to improve the average accuracy by 92% by using a batch normalization technique that dramatically reduces the number of training epochs.

Milenkoski et al. [13] proposed a new lightweight real-time system that classifies human activities based on data captured from a Smartphone's accelerometer.

After the success of the "Long-Short-Term Memory (LSTM)" model extension, named bidirectional LSTM (Bi-LSTM) especially in text classification [16] and speech processing tasks [17] and in generally in the field of natural language processing (NLP). It seems logical that the integration of this architecture would be cost-effective and efficient in the field of human activity recognition for the simple reason that both of its fields used time-series data. This result is well demonstrated (confirmed) in several works on its performance in the extraction of the long-term dependence in the series of temporal data collected by sensors: smartphone [18] and Wearable Sensors [19].

In [20], Alawneh et al. presented a detailed comparison between one-way LSTM and two-way LSTM models on two different datasets of human activity recognition; they found that the priority and efficiency of people activity classification are slightly in favor of the Bi-LSTM model versus unidirectional LSTM.

By remaining in the same RNN residual network category. Since the innovation of Cho et al. in 2014 [21], Gated Recurrent Units (GRU) recurrent neural networks and their extensions make it possible to manage long-

term dependencies. The GRU model resembles that of LSTM since they consist of memory cells capable of storing information from a time series. However, GRUs are characterized by a simpler structure than LSTMs. They contain two gates (the reset gate and the update gate) which control successive data instead of three in TM the LS model forget, input and output),This structure - which uses simple functions such as sigmoid - is able to remember long sets of data without losing useful information.

In the work of [22], researchers compared two model variants with Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) layers, using a data augmentation approach to enhance the robustness of the models RNN in case of the absence of one or more sensors. They concluded that the architecture using the GRU layers is more resistant to the absence of data than that using the LSTM layers.

3 Methodology

Sensor-based human activity recognition is a classification problem that relies on the analysis of time series data collected by wearable sensors (accelerometers, gyroscopes, etc.), the captured time series data will be segmented under shapes sampled at regular intervals and labeled for use in training and testing a deep neural network model. In this way, the model will be able to recognize the activities carried out by a person from the data collected by one or more sensors worn by this person. The most important feature of Deep Learning models over Machine Learning models is that they automatically extract and learn features without the intervention of humans (technicians or experts) from the raw time series data.

For the experimentation of the proposed models, we used a set of datasets WISDM [24], which is considered useful, and well known in the field of sensor-based HAR.

3.1 Data segmentation

After collecting and recording sensor data, the first phase performed in the human activity recognition process is segmenting the time series data and labeling it.

Before performing the segmentation of the data, we must clean the database which consists of the lines of a two-dimensional table, from which we delete the lines that have one or more missing values. Thanks to the sliding window technique, which is well defined in advance and which runs through the series of time series data, the segments are extracted in samples (Frames) of a fixed size of 128 time-tamps per sample and 3 values of the acceleration (corresponding to the 3 dimensions X, Y and Z) associated with each time-tamp. The overlap between two successive samples is limited to 50% as shown in Fig. 2.

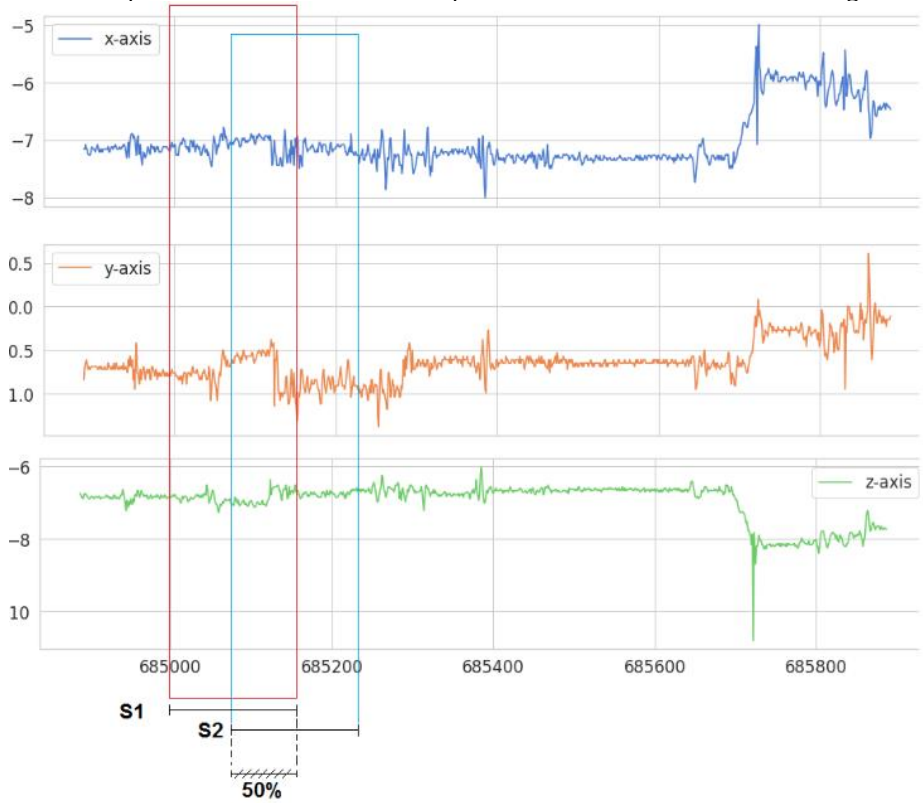


Fig. 2 The segmentation of samples

Like all RNN deep learning models and unlike machine learning models, feature extraction is performed automatically in both types of models that are considered on segments of raw accelerometer data Acc. x, Acc. y and Acc. z, these data used as input do not undergo any form of normalization or transformation.

In our work, we will test two RNN models: LSTM and GRU. In Fig. 2, the block diagram shows the architecture of the proposed LSTM and GRU classifiers for human activity recognition by sensors.

3.2 Feature extraction

The two types of models studied in this work use the characteristics and properties of RNN neural networks. This category of RNN neural network has one of the competitive advantages that differentiate it from other AI methods; extracting features is done automatically unlike ML models. In addition, this RNN architecture category can treat time-series type data effectively. The succession of the aforementioned data has enormous importance and through learning the neural network model that it becomes clear how to classify the activity performed efficiently from a series of data collected by sensors. Thus, this is how most approaches deal with phenomena characterized by the temporal aspect of sequence data. In relation to that, LSTM and GRU models are credible due to their performance and robustness in capturing long-term dependencies on raw speech signals and polyphonic musical data as it is shown in fig3, the internal architecture of LSTM and GRU cells.

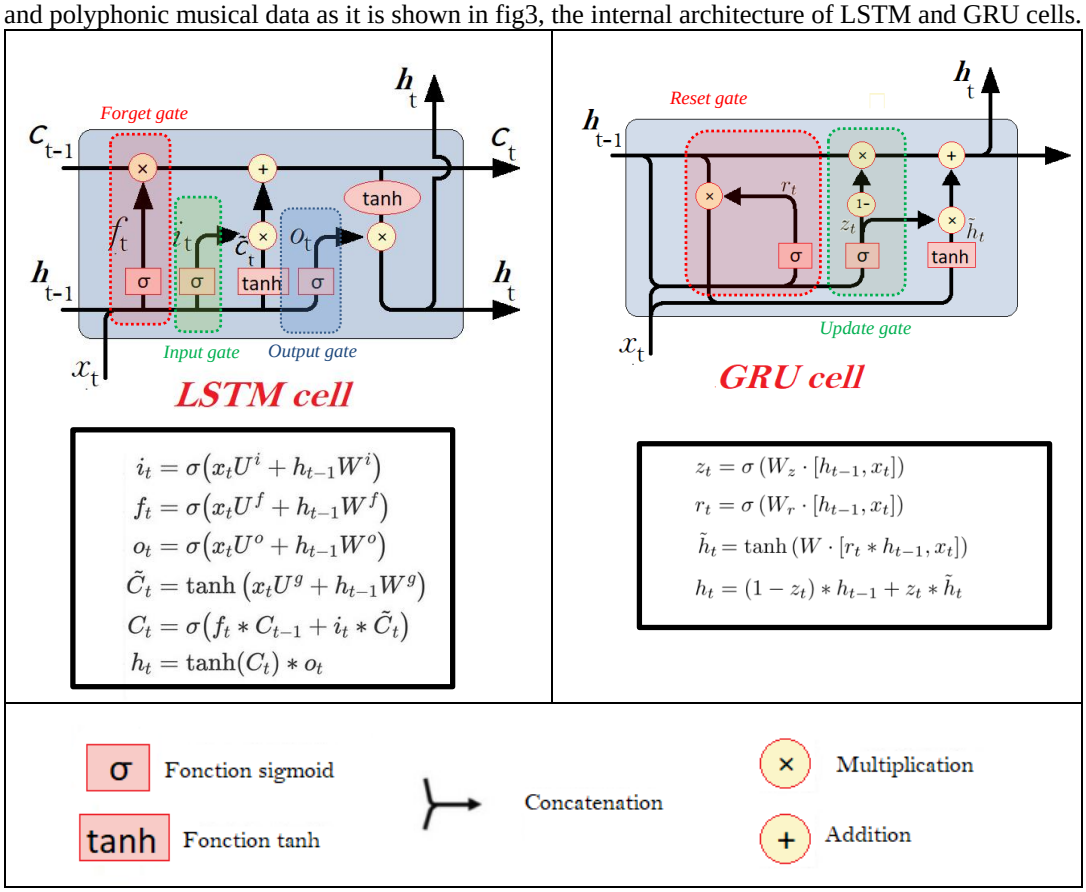


Fig. 3 the Architecture of basic LSTM and GRU.

Since its appearance, LSTM or "long short-term memory" is the recurrent unit that became very popular in the processing of time-series data because of its performance deemed superior and its ability to resist the problem of the leakage gradient.

The basic unit of LSTM is composed of the following three essential gates:

- Input gate
- Output gate
- Then, and Forget gate.

The GRU framework is an evolved extension of the LSTM framework; they have the same function but they do not have the same performance; in the structure, the structure of GRU is surely superior to that of LSTM in terms of classification accuracy and speed.

As shown in Fig. 3 this famous GRU structure is composed of two essential gates: an update gate (r) and a reset gate (z).

3.3 Model architecture

The functional diagram of the studied series of LSTM and GRU architectures is illustrated in the **Table 1** and **Table 2**.

List 1	Layer	Optimizer & Learning rate (LR) &Learning loss (LL)		Batch size & maximum epochs
LSTM 1	Input layer 2LSTM layers (64 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : LR set to 0.0025 LL set to 0.0015	Categorical Cross entropy	Batch size :128 Max epochs :50
LSTM 2	Input layer 2LSTM layers (32 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : LR set to 0.0025	Categorical Cross entropy	Batch size :128 Max epochs :50
LSTM 3	Input layer 2LSTM layers (16 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : LR set to 0.0025	Categorical Cross entropy	Batch size :128 Max epochs :50

Table 1 Series of studied LSTM models and their parameters.

List 2	Layer	Optimizer & Learning rate (LR)		Batch size & maximum epochs
GRU 1	Input layer 2GRU layers (64 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : Learning rate set to 0.0025	Categorical Cross entropy	Batch size :128 Max epochs :50
GRU 2	Input layer 2GRU layers (32 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : Learning rate set to 0.0025	Categorical Cross entropy	Batch size :128 Max epochs :50
GRU 3	Input layer 2GRU layers (16 UNITS) Dropout layer 3 FC layers Output layer	**Adam** : Learning rate set to 0.0025	Categorical Cross entropy	Batch size :128 Max epochs :50

Table 2 Series of studied GRU models and their parameters.

4 Experiments and results

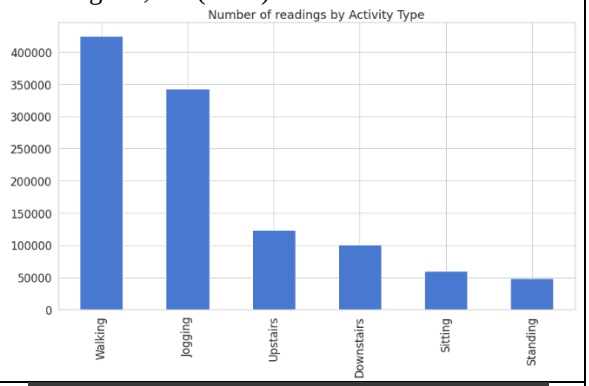
Our experiments were performed on the WISDM dataset .To facilitate the task and save the needed time, the focus has been on using the "**Keras**" and "**Tensorflow**" libraries on the google platform to execute validation processing of python codes in the field of artificial intelligence, machine learning, and deep learning; **Google Colaboratory** which manages notebook documents « **Jupyter** » with its powerful GPUs is accessible remotely, while **Google Colab** is considered the best platform to experiment and analyze AI models. Our models trained with the cross-entropy loss function. Table 1 summarizes the different hyper-parameters used in our models and the other hyper-parameters that are not mentioned take the default values. In the rest of this section, we present more details about the data sets used, the evaluation parameters checked and the results

obtained.

4.1 Dataset used

- ***WISDM dataset [24]:***

Fordham University's Wireless Sensor Data Mining Lab is working on a project called "Wireless Sensor Data Mining" which involves creating an activity recognition dataset of 36 riders performing nine daily activities (i.e. Walking, Sitting, Jogging, Downstairs, Upstairs, and standing); they relied on the sensors built into smartphones to collect 3-dimensional accelerometer type data, with a frequency of 20 Hz. This database is to be divided into two unequal parts: the first one, is a data part of 29 participants that are being used for the training, while the second one, and is a data part of 7 other participants, which is reserved for the realization of the test.
https://www.cis.fordham.edu/wisdm/dataset.php

<table>
<tr><td>Statistics</td><td colspan="2">Raw Time Series Data</td></tr>
<tr><td rowspan="4"></td><td>Number of examples :</td><td>1,098,207</td></tr>
<tr><td>Number of attributes :</td><td>6</td></tr>
<tr><td>Class Distribution :</td><td>Walking: 424,400 (38.6%)
Jogging: 342,177 (31.2%)
Upstairs: 122,869 (11.2%)
Downstairs: 100,427 (9.1%)
Sitting: 59,939 (5.5%)
Standing: 48,395 (4.4%)
 
</td></tr>
<tr><td colspan="2">

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1098203 entries, 0 to 1098203
Data columns (total 6 columns):
 #   Column     Non-Null Count    Dtype
---  ------     --------------    -----
 0   user       1098203 non-null  int64
 1   activity   1098203 non-null  object
 2   timestamp  1098203 non-null  int64
 3   x-axis     1098203 non-null  float64
 4   y-axis     1098203 non-null  float64
 5   z-axis     1098203 non-null  float64
dtypes: float64(3), int64(2), object(1)
memory usage: 58.7+ MB
```

```
print(df)
```

```
         user activity       timestamp    x-axis     y-axis    z-axis
0          33  Jogging   49105962326000 -0.694638  12.680544  0.503953
1          33  Jogging   49106062271000  5.012288  11.264028  0.953424
2          33  Jogging   49106112167000  4.903325  10.882658 -0.081722
3          33  Jogging   49106222305000 -0.612916  18.496431  3.023717
4          33  Jogging   49106332290000 -1.184970  12.108489  7.205164
...       ...      ...             ...       ...        ...       ...
1098199    19  Sitting  131623331483000  9.000000  -1.570000  1.690000
1098200    19  Sitting  131623371431000  9.040000  -1.460000  1.730000
1098201    19  Sitting  131623411592000  9.080000  -1.380000  1.690000
1098202    19  Sitting  131623491487000  9.000000  -1.460000  1.730000
1098203    19  Sitting  131623531465000  8.880000  -1.330000  1.610000

[1098203 rows x 6 columns]
```

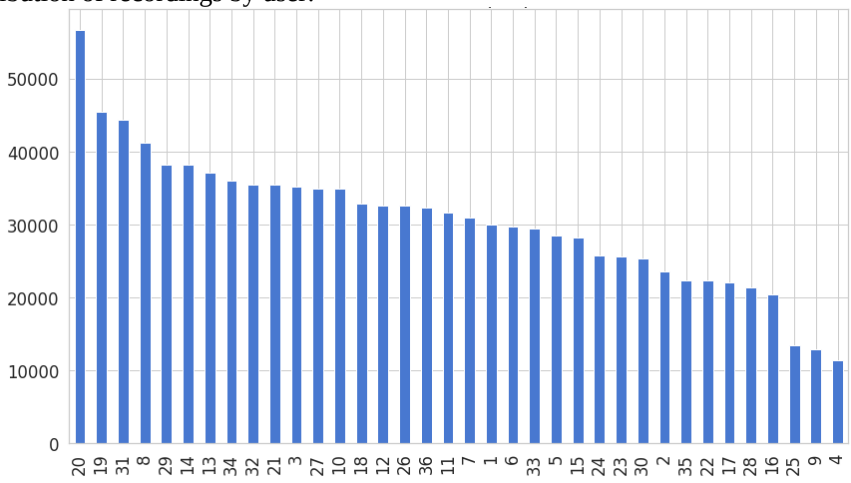
</td></tr>
<tr><td colspan="2">Distribution of recordings by user:
 
</td></tr>
</table>

Fig. 4 the following figure represents an example of sampling for each activity

4.2 Performance measures

To compare various models that are studied in this article, we have used the following performance measures: precision, F1 score, recall, accuracy and confusion matrix (CM).

To ensure the **accuracy**, we use the ratio between the number of samples correctly classified and the overall number of samples tested.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Here is the formula used:

Precision of a model is the ratio of the number of correct positive predictions to the overall number of positively classified samples.

Here is the formula used:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Recall is calculated by the ratio of the number of positives correctly predicted to the actual number of samples judged to be positive.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

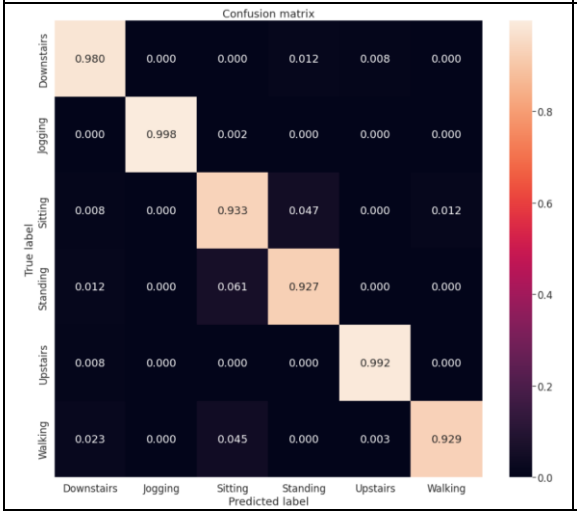
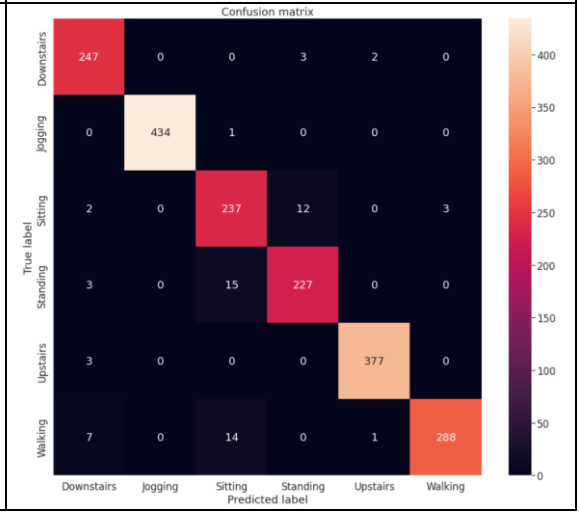
When the data set is unbalanced, we use the F1 score to calculate this value; we use the following mathematical formula:

$$\text{F1} - \text{score} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall})$$

Where:

TP is the number of samples tested and correctly classified as true positives. **TN** is the number of samples tested and correctly classified as true negatives. **FN** is the number of samples tested and misclassified as false negatives. **FP** is the number of samples tested and misclassified as false positives.

The Confusion matrix (CM) plots the true labels horizontally and the predicted labels vertically.
The normalized and not normalized confusion matrix gathers respectively the rate and the number of classification by class, it allows visualize the classification performances of a studied model. In a simplified way, the good predictions are classified by activity on the diagonal of this matrix and the others are false predictions.

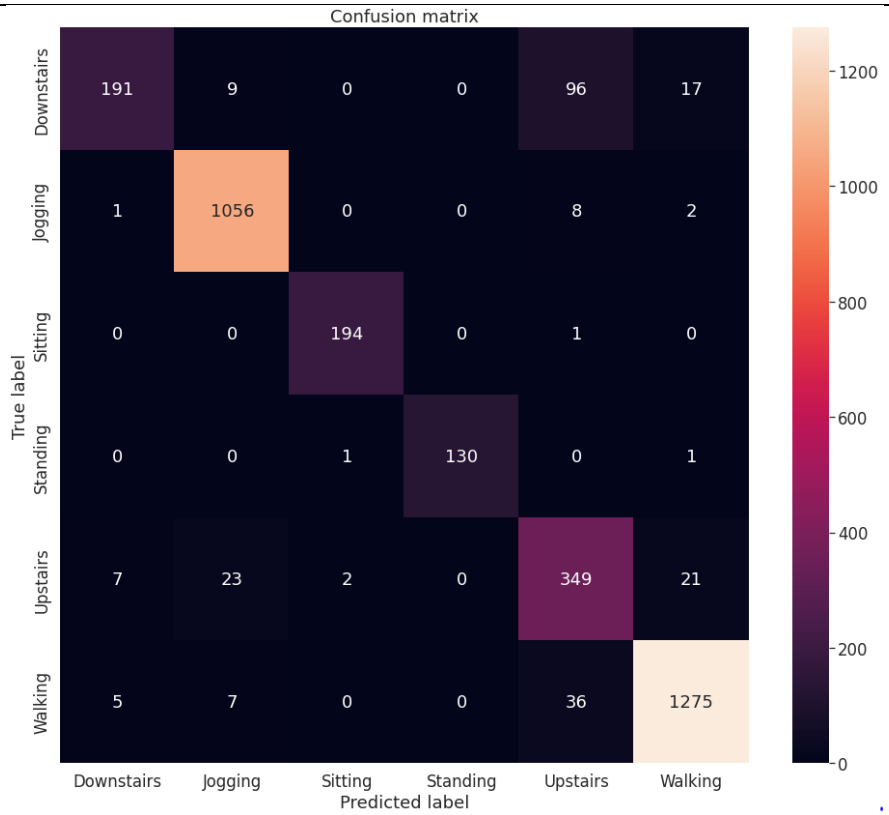
Type of confusion matrix	
Normalized confusion matrix	**Not-normalized confusion matrix**
	

Like most of the databases found in the net, the database used in our article is unbalanced, hence the Not-normalized confusion matrix is the only one used and which effectively presents the results obtained, for example, it gives positively the number of predictions classified in a total number of the same class is a more meaningful thing than a raw rate.

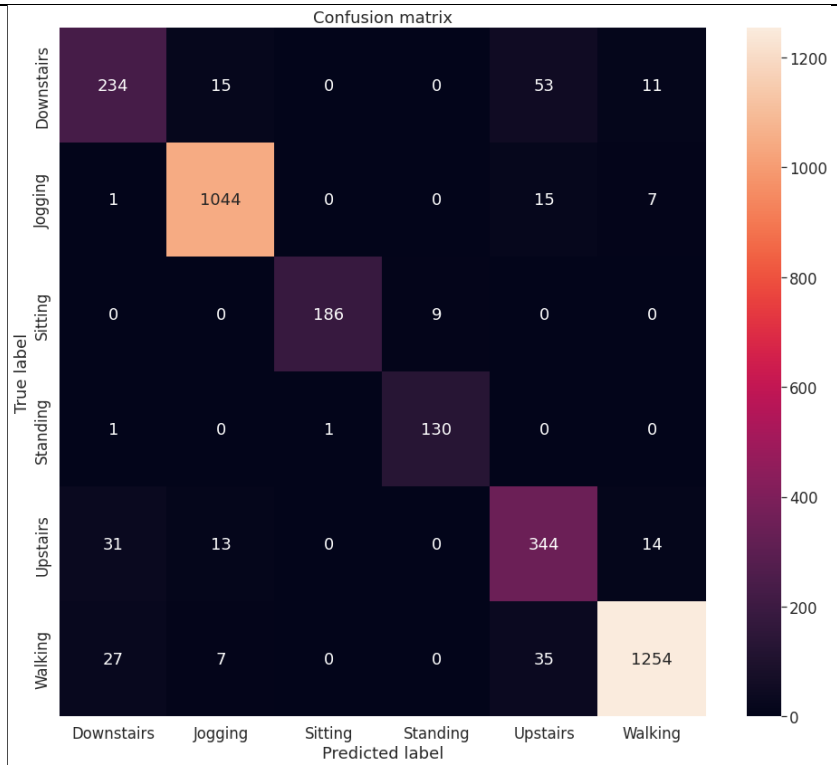
4.3 Results

The dataset used is randomized to 80% sampling to train the model and 20% remaining to perform tests on that model. The WISDM dataset is the most popular in the field of human activity recognition by sensors, they are used in our comparative study.

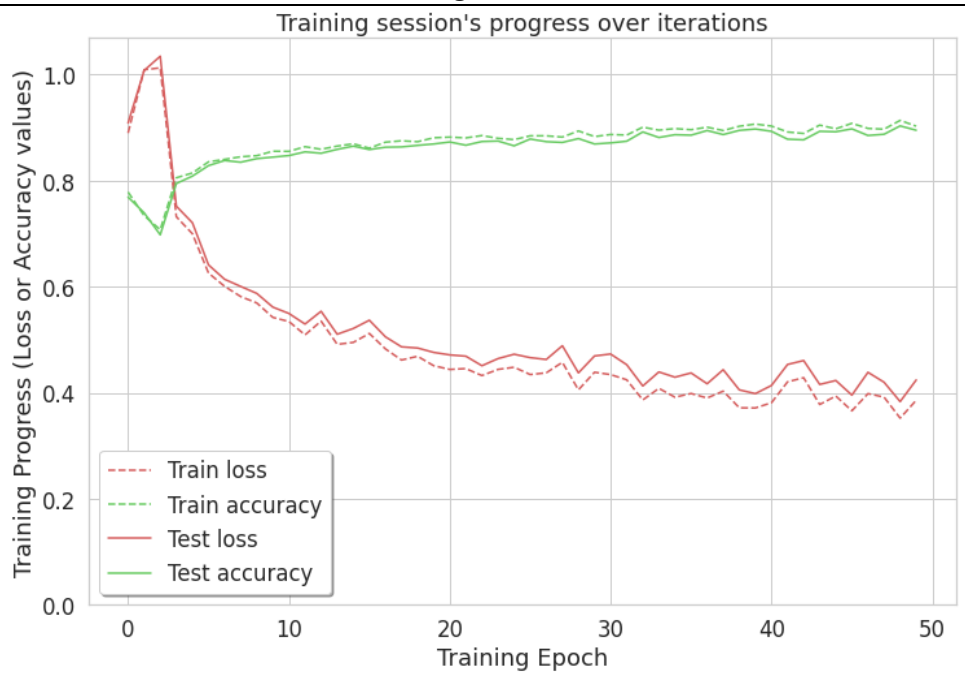
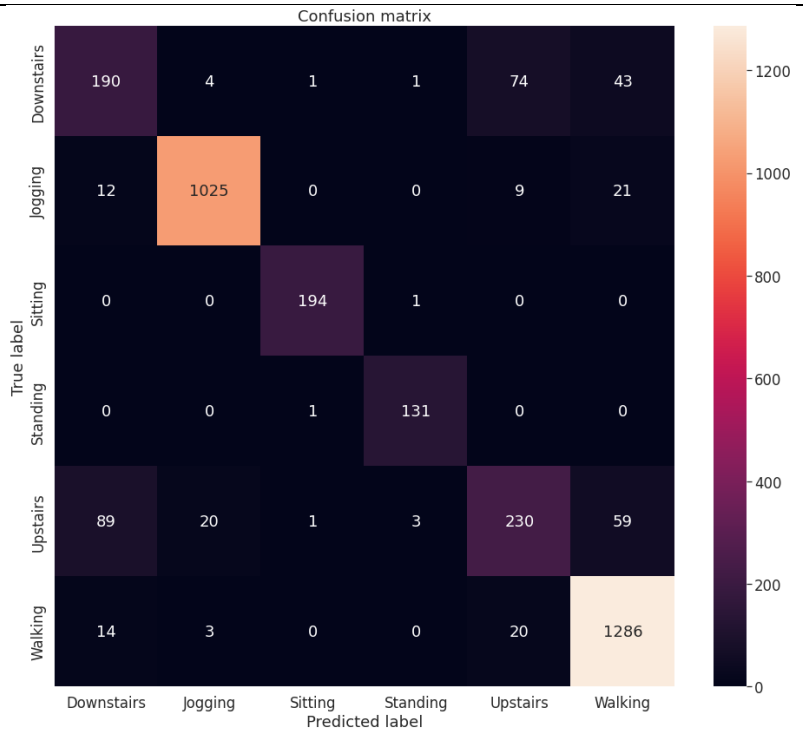
The following table lists illustrate respectively the results obtained for the WISDM data from the list of LSTM and GRU models.

	LSTM 1 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

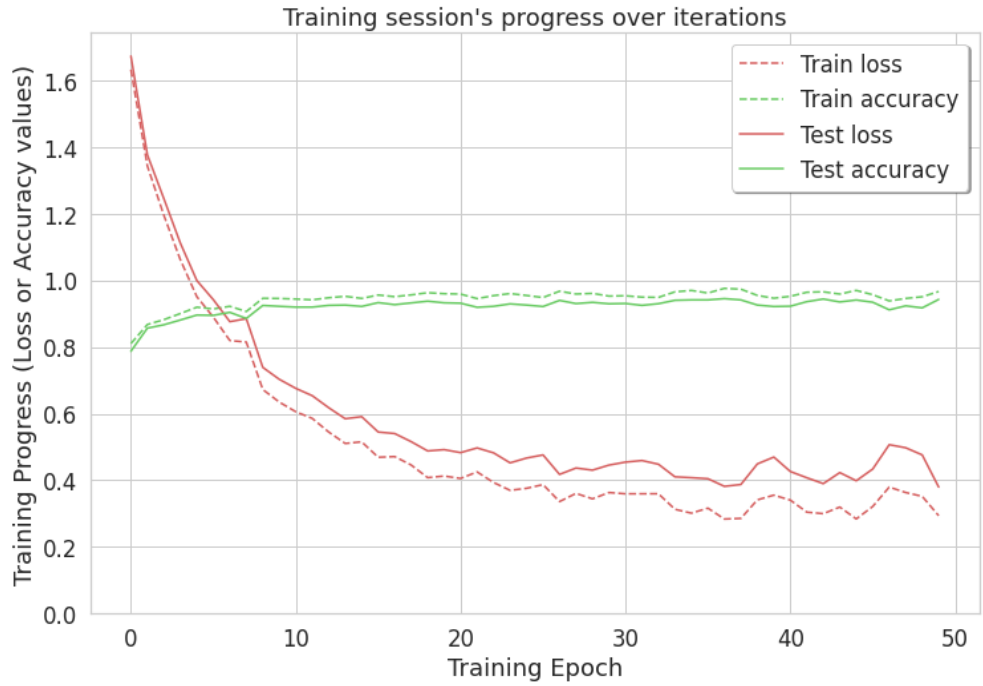
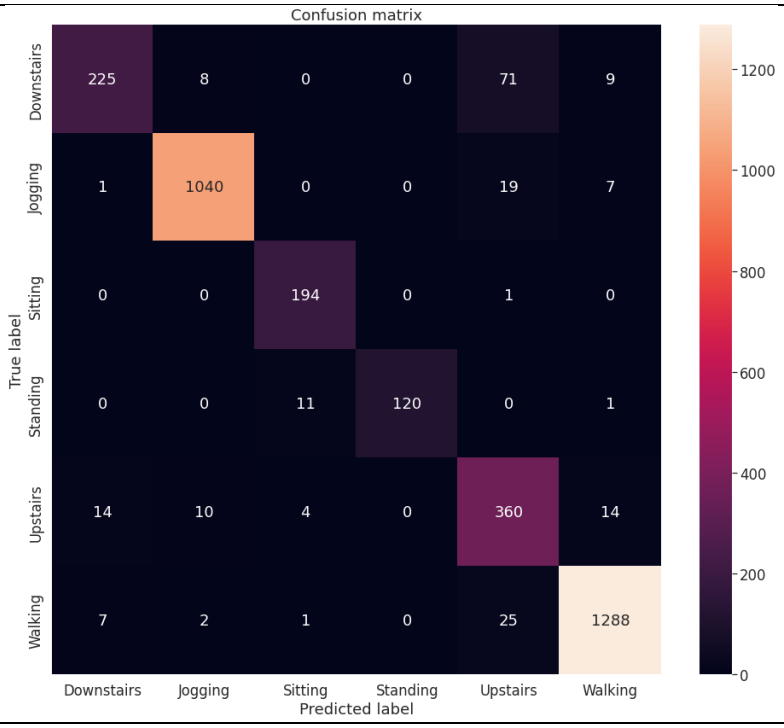
	precision	recall	f1-score	support
Downstairs	0.94	0.61	0.74	313
Jogging	0.96	0.99	0.98	1067
Sitting	0.98	0.99	0.99	195
Standing	1.00	0.98	0.99	132
Upstairs	0.71	0.87	0.78	402
Walking	0.97	0.96	0.97	1323
accuracy			0.93	3432
macro avg	0.93	0.90	0.91	3432
weighted avg	0.94	0.93	0.93	3432

	LSTM 2 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

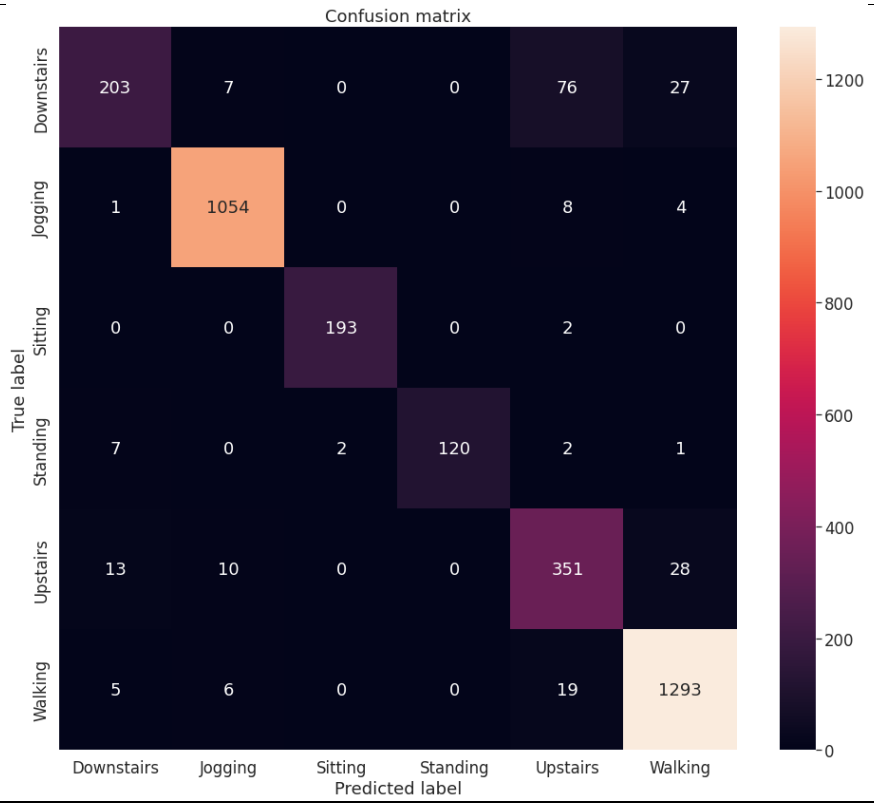
	precision	recall	f1-score	support
Downstairs	0.80	0.75	0.77	313
Jogging	0.97	0.98	0.97	1067
Sitting	0.99	0.95	0.97	195
Standing	0.94	0.98	0.96	132
Upstairs	0.77	0.86	0.81	402
Walking	0.98	0.95	0.96	1323
accuracy			0.93	3432
macro avg	0.91	0.91	0.91	3432
weighted avg	0.93	0.93	0.93	3432

	LSTM 3 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

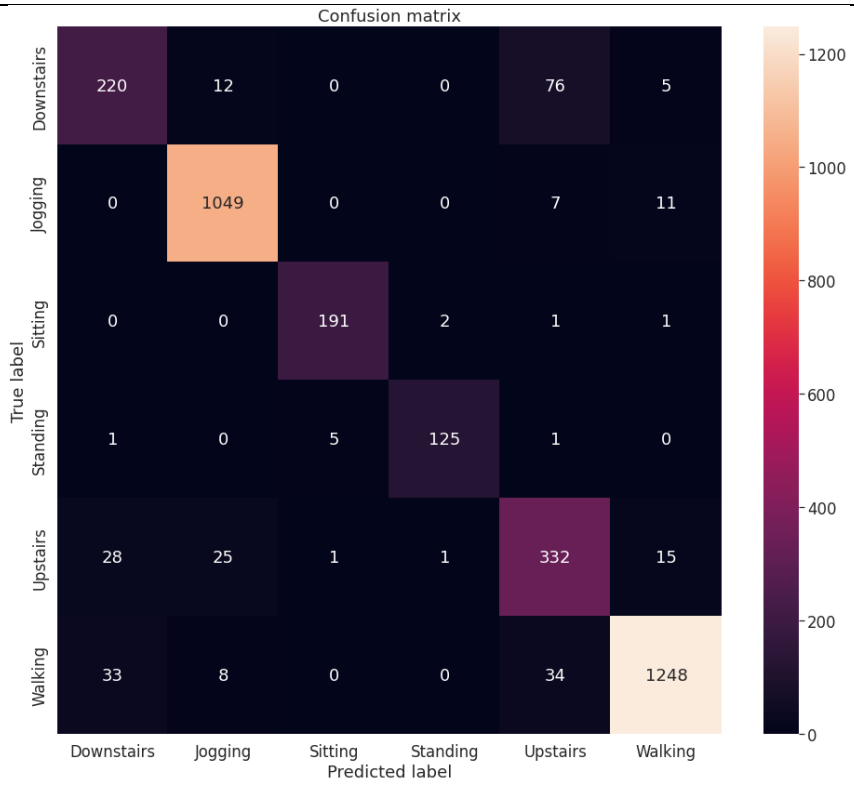
	precision	recall	f1-score	support
Downstairs	0.62	0.61	0.61	313
Jogging	0.97	0.96	0.97	1067
Sitting	0.98	0.99	0.99	195
Standing	0.96	0.99	0.98	132
Upstairs	0.69	0.57	0.63	402
Walking	0.91	0.97	0.94	1323
accuracy			0.89	3432
macro avg	0.86	0.85	0.85	3432
weighted avg	0.89	0.89	0.89	3432

	GRU 1 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

	precision	recall	f1-score	support
Downstairs	0.91	0.72	0.80	313
Jogging	0.98	0.97	0.98	1067
Sitting	0.92	0.99	0.96	195
Standing	1.00	0.91	0.95	132
Upstairs	0.76	0.90	0.82	402
Walking	0.98	0.97	0.98	1323
accuracy			0.94	3432
macro avg	0.92	0.91	0.91	3432
weighted avg	0.94	0.94	0.94	3432

	GRU 2 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

	precision	recall	f1-score	support
Downstairs	0.89	0.65	0.75	313
Jogging	0.98	0.99	0.98	1067
Sitting	0.99	0.99	0.99	195
Standing	1.00	0.91	0.95	132
Upstairs	0.77	0.87	0.82	402
Walking	0.96	0.98	0.97	1323
accuracy			0.94	3432
macro avg	0.93	0.90	0.91	3432
weighted avg	0.94	0.94	0.93	3432

	GRU 3 using Dataset WISDM
Train and Test Accuracy plot	
Confusion matrix	
Classification report	

	precision	recall	f1-score	support
Downstairs	0.78	0.70	0.74	313
Jogging	0.96	0.98	0.97	1067
Sitting	0.97	0.98	0.97	195
Standing	0.98	0.95	0.96	132
Upstairs	0.74	0.83	0.78	402
Walking	0.97	0.94	0.96	1323
accuracy			0.92	3432
macro avg	0.90	0.90	0.90	3432
weighted avg	0.92	0.92	0.92	3432

Architecture	Avg. Accuracy	Avg. Recall	Avg. F1 Score
LSTM 1	0.927	0.900	0.908
LSTM 2	0.908	**0.912**	0.907
LSTM 3	0.855	0.878	0.853
GRU 1	0.925	0.910	**0.915**
GRU 2	**0.932**	0.898	0.910
GRU 3	0.900	0.897	0.897

Table 3 General comparison table between the two series of models LSTM and GRU.

Activity	Models	Precision	Recall	F1 Score
Downstairs	*LSTM 1 / 2 / 3*	**0.94** / 0.80 / 0.62	0.61 / **0.75** / 0.61	0.74 / 0.77 / 0.61
	GRU 1 / 2 / 3	0.91 / 0.89 / 0.78	0.72 / 0.65 / 0.70	**0.80** / 0.75 / 0.74
Jogging	*LSTM 1 / 2 / 3*	0.96 / 0.97 / 0.97	**0.99** / 0.98 / 0.96	**0.98** / 0.97 / 0.97
	GRU 1 / 2 / 3	**0.98** / **0.98** / 0.96	0.97 / **0.99** / 0.98	**0.98** / **0.98** / 0.97
Sitting	*LSTM 1 / 2 / 3*	0.98 / **0.99** / 0.98	**0.99** / 0.95 / **0.99**	**0.99** / 0.97 / 0.99
	GRU 1 / 2 / 3	0.92 / **0.99** /0.97	**0.99** / **0.99** / 0.98	0.96 / **0.99** / 0.97
Standing	*LSTM 1 / 2 / 3*	**1.00** /0.94 / 0.96	0.98 / 0.98 / **0.99**	**0.99** / 0.96 / 0.98
	GRU 1 / 2 / 3	**1.00** / **1.00** /0.98	0.91 / 0.91 / 0.95	0.95 / 0.95 / 0.96
Upstairs	*LSTM 1 / 2 / 3*	0.71 / **0.77** / 0.69	0.87 / 0.86 / 0.57	**0.87** / 0.81 / 0.63
	GRU 1 / 2 / 3	0.76 / **0.77** / 0.74	**0.90** / 0.87 / 0.83	0.82 / 0.82 / 0.78
Walking	*LSTM 1 / 2 / 3*	0.97 / **0.98** / 0.91	0.96 / 0.95 / **0.97**	0.97 / 0.96 / 0.94
	GRU 1 / 2 / 3	**0.98** / 0.96 / 0.97	**0.97** / **0.97** / 0.94	**0.98** / 0.97 / 0.96

Table 4 Comparative table by activity between the two series of LSTM and GRU models.

The following two tables summarize the results obtained; the first shows the mean of precision, recall, and F1 score. The second demonstrates the same performance but by activity.
Following the list of models tested in the previous section, it is interesting to note that the GRU architecture is slightly better at classifying human activities.
The difference in precision between the two models LSTM and GRU varied between 0.5% and 7.7%. The difference in Recall between the two models LSTM and GRU varied between 0.2% and 3.3%. The difference in F1 score between the two models LSTM and GRU varied between 0.7% and 6.2%.

5 Conclusion

This scientific paper, we studied two types of recurrent neural networks for the recognition of human activities from time series of data captured by sensors, by focusing on the problem of forecasting. We have examined with interest two different architecture series LSTM and GRU by explaining their internal structures, hence discussing their properties and training procedures.
We carried out a comparative analysis of the prediction performances obtained by its various networks on a well-known database in the “sensor-based HAR” field. Thanks to the structures of the LSTM and GRU cells and their dynamic mechanisms which are constituted to keep the information and to perfectly model the nonlinear statistical dependencies extracted from the time series of data in order to effectively predict the class of human activity carried out.
According to the obtained results, the information provided makes it generally clear that the models based on RGU cells are a bit more efficient as compared to those found on LSTM cells. Besides, the structures which contain more numbers of LSTM cells or GRU are more effective at predicting human activities.

References

[1] Tahavori F, Stack E, Agarwal V, Burnett M, Ashburn A, Hoseini tabatabaei SA, Harwin W. Physical activity recognition of elderly people and people with parkinson's (PwP) during standard mobility tests using wearable sensors. In: 2017 international smart cities conference (ISC2). Wuxi, ChinaSept, 14–17 September 2017; 2012. p. 403–407.

[2] Physical activity recognition by smartphones, a survey; Morales J., Akopian D. (2017) Biocybernetics and Biomedical Engineering, 37 (3), pp. 388-400.

[3] J. Chai and A. Li, "Deep Learning in Natural Language Processing: A State-of-the-Art Survey," *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2019, pp. 1-6, doi: 10.1109/ICMLC48188.2019.8949185.

[4] A survey on the application of recurrent neural networks to statistical language modeling De Mulder W., Bethard S., Moens M.-F. (2015) Computer Speech and Language, 30 (1) , pp. 61-98.

[5] A. Kumar, S. Verma and H. Mangla, "A Survey of Deep Learning Techniques in Speech Recognition," *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2018, pp. 179-185, doi: 10.1109/ICACCCN.2018.8748399.

[6] J. Park, K. Jang, S.-B. Yang, Deep neural networks for activity recognition with multi-sensor data in a smart home, in 2018 4th IEEE World Forum on Internet of Things (IEEE, 2018), pp. 155–160.

[7] Tapia, E. M., Intille, S. S., & Larson, K. Activity recognition in the home using simple and ubiquitous sensors. In International Conference on Pervasive Computing, pp. 158-175, April 2004.

[8] R Pascanu, T Mikolov, Y Bengio -On the difficulty of training recurrent neural networks. International conference on machine learning, 2013.

[9] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166.

[10] Sepp Hochreiter, S Hochreiter, Jürgen Schmidhuber, and J Schmidhuber. 1997. Long short-term memory. Neural computation 9, 8 (1997), 1735–80. DOI:https: //doi.org/10.1162/neco.1997.9.8.1735 arXiv:1206.2944

[11] Alex Graves, Abdel-Rahman Mohamed, and Georey Hinton. 2013. Speech recognition with deep recurrent neural networks. In International Conference on Acoustics, Speech and Signal Processing (ICASSP). 6645–6649. DOI:https://doi. org/10.1109/ICASSP.2013.6638947 arXiv:arXiv:1303.5778v1

[12] W.-H. Chen, C.A.B. Baca, C.-H. Tou, LSTM-RNNs combined with scene information for human activity recognition, in 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (2017), pp. 1–6.

[13] M. Milenkoski, K. Trivodaliev, S. Kalajdziski, M. Jovanov, B.R. Stojkoska, Real time human activity recognition on smartphones using LSTM Networks, in Proceedings of 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (IEEE, 2018), pp. 1126–1131.

[14] Duffner, S. , Berlemont, S. , Lefebvre, G. , & Garcia, C. (2014). 3D gesture classification with convolutional neural networks. In Proceedings of international conference on acoustic,speech, and signal processing (ICASSP) (pp. 5432–5436) .

[15] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," Expert Systems with Applications, vol. 59, pp. 235–244, 2016.

[16] G. Liu and J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing, vol. 337, p. 325-338, 2019.

[17] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional lstm and other neural network architectures, Neural Networks 18 (5-6) (2005) 602–610.

[18] Zhao, Y.; Yang, R.; Chevalier, G.; Xu, X.; Zhang, Z. Deep Residual Bidir-LSTM for Human Activity Recognition UsingWearable Sensors. Math. Probl. Eng. 2018, 2018, 1–13.

[19] T. Su, H. Sun, C. Ma, L. Jiang and T. Xu, "HDL: Hierarchical Deep Learning Model based Human Activity Recognition using Smartphone Sensors," 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8851889.

[20] Alawneh, L.; Mohsen, B.; Al-Zinati, M.; Shatnawi, A.; Al-Ayyoub, M. A Comparison of Unidirectional and Bidirectional LSTM Networks for Human Activity Recognition. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerComWorkshops), Austin, TX, USA, 23–27 March 2020; pp. 1–6.

[21] Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural ¨ machine translation: Encoder–Decoder approaches. In Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation.

[22] J. Okai, S. Paraschiakos, M. Beekman, A. Knobbe and C. R. de Sá, "Building robust models for Human Activity Recognition from raw accelerometers data using Gated Recurrent Units and Long Short Term Memory Neural Networks," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2486-2491, doi: 10.1109/EMBC.2019.8857288.

[23] T. Zebin, M. Sperrin, N. Peek and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1-4, doi: 10.1109/EMBC.2018.8513115.

[24] Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SIGKDD Explor Newsl 12(2):74–82

[25] Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling.