# Enhanced Online Segmentation and Performance Evaluation Method for Real-Time Activity Recognition in Smart Homes

1st Abbas Ramadan
*Lab-STICC, UMR CNRS 6285*
*University of South Brittany*
Lorient, France
abbas.ramadan@univ-ubs.fr

2nd Farida Saïd
*LMBA, UMR CNRS 6205*
*University of South Brittany*
Lorient, France
farida.said@univ-ubs.fr

3rd Florent Frizon De Lamotte
*Lab-STICC, UMR CNRS 6285*
*University of South Brittany*
Lorient, France
florent.lamotte@univ-ubs.fr

***Abstract*—This paper presents a robust method for real-time recognition of Activities of Daily Living (ADLs) in smart home environments using IoT data. Our approach improves the segmentation of sensor data streams into distinct activities by leveraging IoT sensor spatiotemporal features and applies the Needleman-Wunsch method to align predicted and actual activities. Testing on the Aruba dataset achieved 83.2% accuracy, demonstrating superior performance in segmentation and activity recognition compared to existing dynamic methods. Future work will focus on developing a sensor installation simulator to further enhance accuracy and reliability.**

***Index Terms*—Activity of Daily Living, ADLs, IoT Data, Human Activity Recognition, Smart Home, Real-time Activity Recognition, Spatiotemporal Segmentation, Dynamic Segmentation Algorithms.**

## I. Introduction

The rapid development of ambient computing has led to the creation of various sensors and devices used in assistive environments for elderly or disabled individuals, allowing them to remain in their own homes. At the heart of this technology is Ambient Assisted Living (AAL), a rapidly growing field focused on creating integrated ecosystems of wireless networks, sensors, computers, mobile devices, and software applications. These systems support personal assistance, health monitoring, and telemedicine services [1].

The continuous use of these devices allows for collecting data that tracks an individual's daily activities, offering valuable insights into their behavior, habits, and routines. Analyzing this data helps, for instance, evaluate a person's health and well-being, facilitates the early identification of physical vulnerabilities, and helps prevent emergencies.

However, the variability in how individuals perform daily activities can lead to inconsistencies in the collected data, complicating sensor data segmentation into activities and their labeling. In this regard, considering spatiotemporal correlations between sensors can improve segmentation accuracy and provide more reliable activity recognition, especially during online recognition of activities of daily living (ADLs) [2], [3].

Furthermore, the integration of artificial intelligence algorithms and Internet of Things (IoT) technologies has significantly enhanced the automatic recognition of human activities [4]. Central to this advancement is the creation of digital twins within AAL environments. These digital twins model smart home data and use activity recognition algorithms to improve system functionality. Specifically, they enable the simulation of system installations, monitoring caregivers' activities, and generating predictive scenarios, thereby optimizing care and management in smart home environments [5]. In this context, our research questions are as follows: 1) How can we improve real-time dynamic segmentation of sensor data by more effectively calculating spatiotemporal correlation parameters? 2) How can we address the challenge of evaluating dynamic recognition results, especially when the actual activities and those predicted by algorithms differ in number and sequence of events? To answer these questions, we propose a new approach to calculate spatiotemporal correlation parameters to improve real-time dynamic segmentation of sensor data. Next, we introduce a method to address the challenge of evaluating dynamic recognition results by quantifying the discrepancies in number and sequencing between the actual activities and those predicted by algorithms.

Our paper is structured as follows. In section II, we outline the context of the study, followed by the methodology developed to answer our research questions in section III. We present our results in section IV and draw a conclusion and some perspectives in sectionV.

## II. Background

Activities of Daily Living (ADLs) are the activities necessary to live, maintain well-being, and participate in social life. These activities include eating, bathing (personal hygiene), toileting, continence, dressing, and transferring.

### A. ADLs Analysis

Analysis of ADLs is essential in evaluating and studying an individual's physical vulnerability. In the literature, some key points highlight the benefits of analyzing ADLs for individuals:

- Assessing independence: There are several tools for assessing a person's ability to perform a specific group of essential activities of daily living (ADLs), such as the Katz index [6] or AGGIR [7]. Automatic recognition of human activity can assess an individual's functional autonomy by studying their ability to perform the activities listed in the Katz index. This approach provides objective information about a person's degree of independence in their daily activities, which can be invaluable to healthcare professionals and caregivers in planning appropriate care and support.
- Detecting fragility: Fragility can increase susceptibility to developing health problems. Analyzing ADLs can help identify signs of frailty in individuals at an early stage [8]. By tracking changes in ADLs, healthcare professionals can intervene earlier to prevent health deterioration.
- Care planning: ADLs are an essential indicator for determining the level of care needed at home. Measuring ADLs helps plan the resources and personnel needed to respond to an individual's care needs [9].
- Improving quality of life: By analyzing individuals' behavior through ADLs, we can identify their difficulties and then deploy interventions or automate services to enhance their quality of life [10].

### *B. ADLs measurement*

With the development of AI technologies, it has become possible to automatically recognize Activities of Daily Living (ADLs) using data collected from connected devices in the home. Ambient Assisted Living (AAL) utilizes a variety of environmental sensors such as acoustic sensors, temperature, CO2, humidity sensors, electrical power sensors, passive infrared (PIR) sensors, and more advanced sensors like radar, lidar, and cameras [11]. To address variations in ADL recognition algorithms caused by different data sources, we plan to develop an adaptive method that can efficiently process any Boolean data. This method will involve the implementation of a transformation layer, which will convert sensor readings into Boolean values by applying predefined thresholds. This process will standardize the input data to ensure consistent algorithm performance.

### *C. Dataset*

We leveraged the widely-used Aruba dataset from the CASAS smart home project [12], known for its extensive coverage and detailed sensor logs, to train and evaluate our activity recognition model. The dataset captures the daily activities of an elderly woman living alone, with regular visits from her family, over a period of seven months, from November 4, 2010, to June 30, 2011. The smart home was equipped with 31 wireless binary motion sensors, 5 temperature sensors, and 3 door sensors, carefully positioned to monitor the resident's movements and interactions (Figure 1).
Each entry in the dataset includes a timestamp, sensor identifier, sensor status, and activity annotations, offering a rich data stream for robust activity recognition tasks. With 1,719,557 recorded events, this dataset provides one of the most comprehensive resources for evaluating models in this field. Due to its broad use in human activity recognition research, the Aruba dataset serves as an established benchmark for comparing various algorithms and methodologies.
To optimize real-time processing while maintaining high recognition accuracy, we selected a subset of 11 sensors (motion and door sensors) associated with the following activities: "Bed to toilet", "Eating", "Entering home", "Housework", "Leaving home", "Preparing meals", "Relaxing", "Sleeping", "Washing dishes", and "Working". Figure 2 illustrates the sequence of 'Sleeping' and 'Bed to Toilet' events recorded during the night of May 15, 2011, which aligns with typical nighttime behavior.

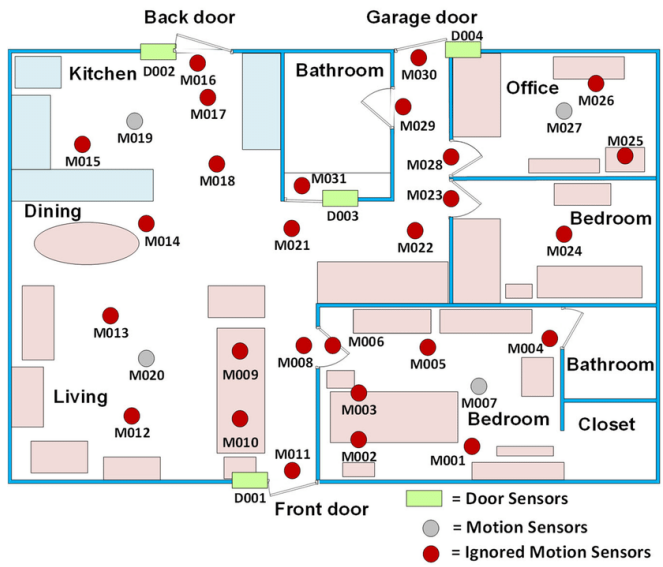


Fig. 1. Aruba layout.

| Date | Time | Sensor | Status | Annotation |
|---|---|---|---|---|
| 2011-05-15 | 04:22:43.082116 | M003 | ON | |
| 2011-05-15 | 04:22:44.624551 | M003 | OFF | Sleeping end |
| 2011-05-15 | 04:22:46.649038 | M005 | ON | |
| 2011-05-15 | 04:22:50.398987 | M004 | ON | Bed_to_Toilet begin |
| 2011-05-15 | 04:22:50.472761 | M005 | OFF | |
| 2011-05-15 | 04:22:57.782528 | T005 | 21 | |
| 2011-05-15 | 04:22:58.533178 | M004 | OFF | |
| 2011-05-15 | 04:22:58.623508 | M007 | OFF | |
| 2011-05-15 | 04:25:44.619463 | M004 | ON | |
| 2011-05-15 | 04:25:45.993194 | M007 | ON | |
| 2011-05-15 | 04:25:50.590135 | M004 | OFF | Bed_to_Toilet end |
| 2011-05-15 | 04:25:50.727788 | M005 | ON | |
| 2011-05-15 | 04:25:52.659384 | M005 | OFF | |
| 2011-05-15 | 04:25:55.446432 | M003 | ON | Sleeping begin |
| 2011-05-15 | 04:25:57.258257 | M007 | OFF | |

Fig. 2. Raw samples of sensor dataset.

### *D. Literature review*

A thorough literature review was conducted to provide an in-depth overview of previous research in human activity

recognition. This section delves into historical contributions and significant advancements within this specific area. The primary objective is to present a comprehensive survey of the methodologies and approaches employed thus far, shedding light on both the resolved and outstanding challenges. The criteria for selecting scientific articles on ADL recognition included:

- Real-time analysis: This is crucial for ADL detection as it enables the immediate identification and categorization of activities, facilitating prompt responses in emergencies and ongoing monitoring of a person's daily routines [13].
- Ambient sensors: These sensors are embedded within a person's living environment and include devices like motion detectors, temperature sensors, and door sensors, as well as controllers for lights and shutters. They are pivotal in unobtrusively gathering data about a person's daily activities and behaviors [14].
- Spatiotemporal aspects: These aspects consider daily activities' spatial ('where') and temporal ('when') dimensions. Indeed, activities can take place in different locations and follow specific temporal sequences. Recognizing and applying these dimensions are vital for accurately pinpointing activities and placing them within a broader contextual framework [15].
- Deep learning classifiers differ from older Activity of Daily Living (ADL) recognition methods because they can automatically extract features, eliminating the need for manual design of descriptors. These classifiers, especially deep neural networks, outperform traditional methods by capturing complex patterns and modeling non-linear relationships. Their adaptability and ability to generalize from diverse data are crucial for ADLs, where activities vary significantly between individuals. Besides, automatic parameter learning reduces the need for extensive manual intervention, while the ability to adapt to different dataset sizes offers significant scalability [16].

Table I synthesizes the outcomes of our literature review, highlighting six distinct approaches extracted from a pool of 3,222 publications that used the keywords 'Human Activity Recognition,' 'Activity of Daily Living,' 'Ambient,' and 'Real-time.' We selected two specific methods [2] and [3] for further in-depth study, testing, and comparison.

| References | Ambient sensors | Dynamic segmentation | Spatiotemporal | Classifier | Accuracy | Dataset |
|---|---|---|---|---|---|---|
| C. Chen et al. [17] | Yes | No | No | NB, MP, SMO, LB | - | CASAS |
| D. Liciotti et al. [18] | Yes | No | No | LSTM | 0.89 | CASAS |
| J. Wan et al. [19] | Yes | No | Yes | NB, BN, DT, NBT, HMM | 0.91 | CASAS |
| D. Bouchabou et al. [20] | Yes | No | No | LSTM | 0.84 | CASAS |
| H. Najeh et al. [2] | Yes | Yes | Yes | CNN2D | 0.81 | CASAS |
| Z. Xu et al. [3] | Yes | Yes | Yes | CNN-LSTM | 0.78 | CASAS |

TABLE I
COMPARATIVE ANALYSIS OF DIFFERENT APPROACHES

## III. METHODOLOGY

Although the two selected methods focus on human activity recognition in smart home environments using dynamic segmentation and emergent modeling techniques, they present different methodological innovations and specific approaches. In [3], The authors present a model that uses spatio-temporal correlations to manage the segmentation of streaming data. This approach ensures that data segmentation is sensitive to both the location of sensor activations and the timing between these events, which helps accurately define each segment's context. This method uses stigmergy to build activity features explicitly represented as a network, aiding in the contextual interpretation of sensor data without needing deep domain knowledge. In [2], the authors, inspired by the above work, emphasize a convolutional neural network (CNN) that is bootstrapped to dynamic segmentation combined with stigmergy-based encoding. Stigmergy is used here to create a feedback loop within the environment, enhancing the learning process by encoding and classifying segments of activities more efficiently. The approach strongly relies on the capabilities of CNNs to process and analyze the segmented data. The main difference between the two approaches is that Xu et al. employ stigmergy to construct activity features represented in a network model directly, whereas Najeh et al. use stigmergy to enhance the encoding process within a neural network framework. There is a difference in how stigmergy is applied—one is for building a self-updating model that adapts based on the activity context, and the other is for improving neural network training.

Both methods are based on the same principle but operate differently. They consist of two main phases: an offline phase for calculating and optimizing the parameters used to segment the real-time data stream, and an online phase for segmenting the real-time data stream, and then encoding and classifying the data segments to identify the specific activity being performed. The procedural steps for both phases are illustrated in Figure 3.

### A. Dataset overview

Our data is organized to facilitate precise tracking and analysis of activities, providing detailed insights into the sequence and nature of captured events. It consists of three sets: Events (E): This set includes individual events, each characterized by attributes such as date, time, sensor ID, and sensor value. E = $\{E_1, E_2, \ldots, E_n\}$ and each event $E_i$ is defined as a quadruplet (date, time, sensorID, sensorValue).

Activity Segments (W): This set comprises a series of activity segments, each segment $W_i$ encompassing a list of events

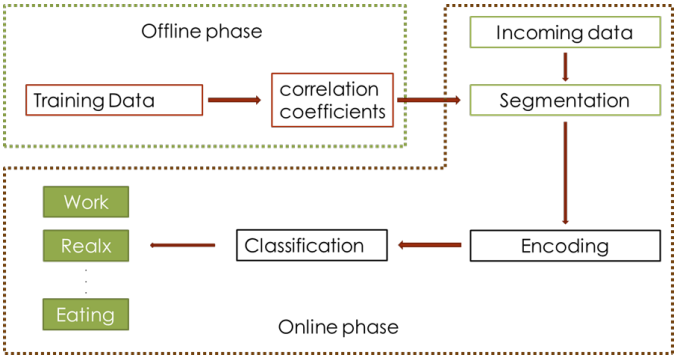


Fig. 3. Presentation of ADL recognition methodology.

$E_i$. For example, in Figure 3, the "Bed to Toilet" segment captures the sequence of activities between the "Begin" and "End" states.

Sensor Identifiers (S): This set contains the identifiers of the specific sensors used in data collection, ranging from $S_0$ to $S_{33}$. These identifiers represent the various devices deployed throughout the environment to monitor and record activity data.

### B. Offline phase

During the offline phase, we compute spatial and temporal correlation coefficients, which are essential for accurately segmenting sensor-derived events in real-time. These coefficients consider both the geographical context of the sensor locations and the time intervals between events.

### C. Spatial Correlation

This concept is widely recognized across various fields as an indicator of the inherent relationships between entities within a given space. In our context, spatial correlation is critical for discerning the interdependencies among activities captured through sensor data within the physical dimensions of the environment. We consider three distinct methods to measure this correlation.

1) Pearson Product Moment Correlation (PMC): This method [2] calculates the linear correlation between sensor events using the covariance and standard deviations of $E_i$, as depicted in Equation 1. The PMC provides a straightforward quantification of linear dependencies.
2) Mutual Information: This approach [3] evaluates the dependency between sensors by assessing the probability distributions of their events. A Sensor Correlation Matrix (SCM) is constructed based on offline data to map these dependencies, as outlined in Equation 2. Significant correlations are identified by applying thresholds and considering geographical proximity among sensors.
3) Sequential Correlation Evaluation: This method [19] enhances the SCM technique by providing a detailed analysis of sensor interrelationships over time. It achieves this by evaluating correlations based on consecutive occurrences in the sensor event sequence (see Equation 4 for details). Sequential Correlation Evaluation: We took inspiration

$$\rho_{\mathbf{XY}} = \frac{\mathbf{cov}(\mathbf{X}, \mathbf{Y})}{\sigma_{\mathbf{X}}\ \sigma_{\mathbf{Y}}} \tag{1}$$

$$\text{SCM}(s_i, s_j) = \frac{1}{|\mathbf{W}|} \sum_{k=1}^{|\mathbf{W}|} \theta(s_i, s_j) \tag{2}$$

$$\theta(s_i, s_j) = \begin{cases} 1, & \text{if } (s_i, s_j) \in \mathbf{W}_k \\ 0, & \text{else} \end{cases} \tag{3}$$

$$\text{SCM}(s_i, s_j) = \frac{1}{|\mathbf{W}|} \sum_{p=1}^{|\mathbf{W}|} \sum_{k=1}^{|\mathbf{W}_p|-1} \gamma(s_k, s_i)\gamma(s_{k+1}, s_j) \tag{4}$$

$$\gamma(s_i, s_j) = \begin{cases} 1, & \text{if } s_p = s_i \\ 0, & \text{if } s_p \neq s_i \end{cases} \tag{5}$$

Together, these approaches establish a comprehensive framework for comprehending sensor dependencies in dynamic environments. Building on that, we have developed a new approach that expands on the methods outlined in Equations 2 and 4, with an additional focus on the spatial separation of activity . In this updated method, when computing the SCM matrix, We reduce the values for the sensor pairs, from the first and last sensor of each activity, with those of the sensor pairs from the first and last sensor of all other activities. The following equation expresses the modified SCM calculation:

$$\begin{aligned} \text{SCM}(s_i, s_j) = & \frac{1}{N} \sum_{k=1}^{m} \frac{1}{L_k} \sum_{(s,t) \in W_k} \left(\delta_{s,i}\delta_{t,j} + \delta_{s,j}\delta_{t,i}\right) \\ & - \frac{1}{F} \sum_{k=1}^{m} \sum_{(s,t) \in W_k} \Bigg( \frac{1}{L_k} \sum_{u \in S \setminus W_k} \left(\delta_{s,i}\delta_{t,u}\right. \\ & \left. + \delta_{s,j}\delta_{t,u} + \delta_{u,i}\delta_{t,s} + \delta_{u,j}\delta_{t,s}\right) \Bigg) \end{aligned} \tag{6}$$

where:

- $\text{SCM}(s_i, s_j)$ represents the value of the SCM matrix for the pair $(s_i, s_j)$
- $N$ is the total number of activity segments
- $m$ is the number of distinct activities
- $L_k$ is the number of sensor pairs in the activity segment $W_k$
- $F$ is a normalization factor, for example $F = 4 \times m \times (n - L_k)$
- $\delta_{x,y}$ is the Kronecker delta function, defined as:

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

- $S$ is the set of sensor identifiers.

This formula calculates the correlations between sensors for various activity segments, considering the specific relationships between sensor pairs and adjusting for sensors excluded from the activity segments. The results of the sensor correlation matrix (SCM) are displayed in Figure 4.

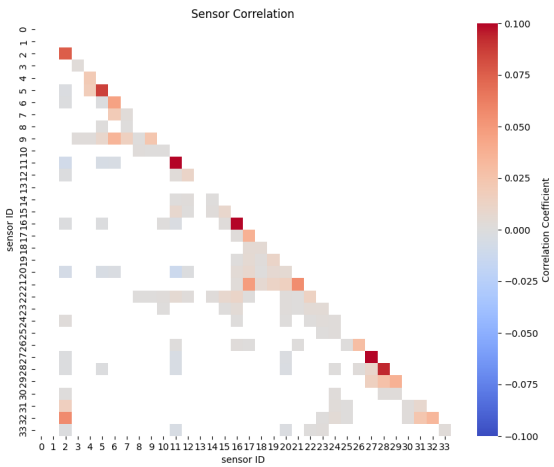


Fig. 4. Correlation value between pairs of sensors in the SCM matrix

*1) Time correlation:* The temporal dependency factor can be used to improve dynamic segmentation. Two thresholds are used in the literature to determine whether an event belongs to a segment. Maximum Time Interval (MTI) examines the maximum duration of each activity as determined by sensor identifiers. Maximum Time Span (MTS) represents the maximum duration between two events.

$$\mathrm{MTI}(s_i, s_j) = \max_{p=1}^{|\omega_k|-1} (T_{p+1} - T_p)\, \delta_{p,i}\delta_{p+1,j} \tag{7}$$

$$\mathrm{MTS}(s_i, s_j) = \max_{k=1}^{|\omega|}(T_{\text{last},j} - T_{\text{first},i}) \tag{8}$$

MTI is calculated using Equation 7, and the maximum time between sensor events is determined for each consecutive pair. MTS (Equation 8) represents the maximum time between each activity's first and last events.

### D. Online phase

The online phase aims to segment incoming sensor data using spatial and temporal correlation conditions. After segmentation, the next step involves encoding and classifying the data to identify activity.

*1) Segmentation:* Dynamic data segmentation seeks to group sensors with significant spatial and temporal correlation of events using the values of SCM, SCT, MTS, and MTI. Algorithm 1 illustrates the process.

*2) Encoding:* The encoder layer represents the input segment in a latent space as an adjacency matrix. To achieve this, we start by constructing the directed network; then, we use the Directed Weight Network (DWN) approach to

**Algorithm 1** Dynamic Sensor Data Segmentation Algorithm.

**Require:** Streaming sensor data: $E = (E_1, E_2, \ldots, E_n)$
**Require:** Initial segmentation window: $W_1 = \{\}$
**Ensure:** Segments: $W_1 = \{E_1, E_2, \ldots, E_k\}$
**Conditions:** SCM/SCT, MTI and MTS
**While** incoming sensor event **do**
  $add =$ False
  **if** $W_i$ is empty or $add =$ False **then**
    $W_1 = \{E_i\}$
    $E_{\text{first}} = E_i$
    $E_{\text{last}} = E_i$
  **else**
    $T_{\text{cor_last}} > MTS(S_i, S_{\text{last}})$
    $T_{\text{cor_first}} > MTS(S_i, S_{\text{last}})$
    $S_{\text{cor}} = SCM(S_i, S_j) \geq SCT$
    **if** $S_{\text{cor}}$ and $T_{\text{cor_first}}$ and $T_{\text{cor_last}}$ **then**
      $W_1 = W_i + \{E_i\}$
    **else**
      $add =$ True

calculate the weight of each sensor state change based on its duration [3]. These weights are then stored in a matrix, as illustrated in Figure 5. The complete process is as follows: part (a) shows the boxed events representing a segment. The event preceding this segment corresponds to the last sensor activated in the previous segment and constitutes the first sensor in the DWN, as illustrated in (b). This sensor is followed by the active sensors (SensorValue = ON) within the segment. The equation 9 used to calculate the weights is given by Equation 9:

$$I = \frac{(1-\rho)^{(T_e - t_e)} - (1-\rho)^{(T_e - t_s)}}{\rho} \tag{9}$$

where $T_e$ is the time of the last event in the segment, $t_e$ and $t_s$ are the activation and deactivation times of the sensors, respectively, and $\rho$ is set to 0.2 [2], as shown in part (c) of the figure. The final step corresponds to part (d) of the figure: the calculated weights feed into the adjacency matrix, whose dimension corresponds to the number of sensors.

*3) Classification:* Activity recognition classification assigns labels or categories to temporal sequences based on observed actions or behaviors. This task is mainly carried out using deep learning models that analyze patterns from sensor-captured data. By training on annotated datasets, the model can predict activities in real-time by associating specific labels with data sequences. For our classifier, which processes input data as two-dimensional matrices, we use Convolutional Neural Networks (CNNs) designed explicitly for 2D data, known as CNN2D. These networks are effective for this type of data because they can discern spatial relationships, share weights across the network, prioritize key features, and handle visual information efficiently [21].

Our CNN architecture includes specialized layers to enhance performance and accuracy. Batch Normalization is implemented to speed up learning convergence, stabilize gradi-

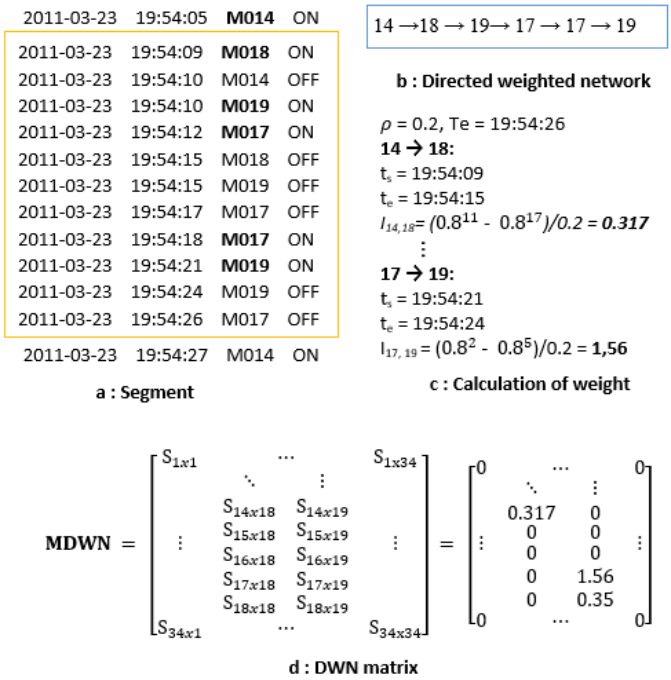


Fig. 5. Representation of segment encoding procedure.

ents, and mitigate overfitting. A Flatten layer transforms the two-dimensional matrix data into a one-dimensional vector, preparing it for processing in the subsequent fully connected Dense layer. This Dense layer is crucial as it generates predictions by computing a linear combination of inputs and applying an activation function. Regularization techniques such as Dropout and MaxPooling2D are used to refine the model further. Dropout helps reduce overfitting by randomly omitting subsets of features during training, while MaxPooling2D is used to decrease the spatial dimensions of the activation layer by summarizing the most significant features from the input data. Figure 6 illustrates the configuration and functionality of each layer and technique used in our model.

### *E. Performance Evaluation*

Our study involves two main tasks. The first is segmenting the incoming data stream using reference segments, followed by labeling these segments. Accurate performance measures are essential for both tasks to evaluate their effectiveness.

We assess our model's classification performance using key metrics such as accuracy, precision, recall, and the F1 score, which are essential for understanding how well the model performs under various conditions.

These metrics are based on indicators that compare the model's predictions with the actual labels in the dataset, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Accuracy measures the proportion of all predictions that the model identified correctly, both positives and negatives,

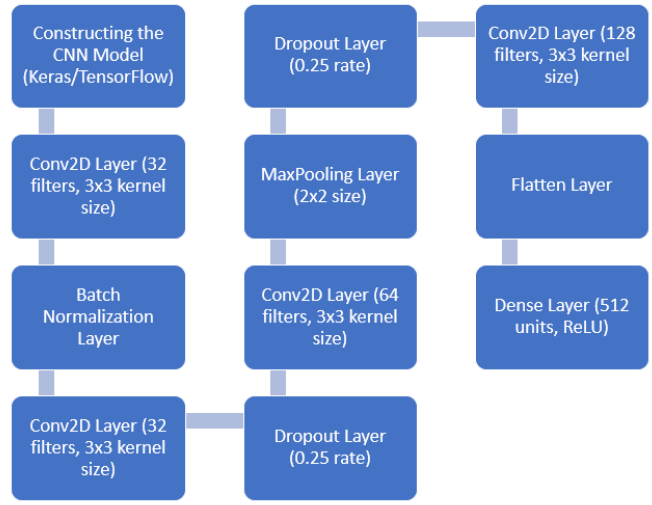


Fig. 6. Classification model

calculated as:

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

While accuracy provides a straightforward gauge of overall performance, it can be misleading in cases of class imbalance.

Precision (P) focuses specifically on the model's positive predictions, measuring the accuracy of these predictions. It is defined as the proportion of positive identifications that were actually correct:

$$P = \frac{TP}{TP + FP}$$

Recall (R), or sensitivity, measures the model's ability to identify all actual positives from the data correctly:

$$R = \frac{TP}{TP + FN}$$

The F1 score balances precision and recall. It is the harmonic mean of these two metrics:

$$F1 = 2 \times \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

A high F1 score only occurs when both precision and recall are high, making it a robust indicator of the model's accuracy. This robustness is especially crucial in situations where class imbalance might skew the perceived effectiveness of the model if only one metric were considered.

In online dynamic segmentation, evaluating the model's results can be challenging because the number of segments detected may sometimes differ from the actual number. To tackle this issue, we suggest calculating similarity scores to evaluate the resemblance between two sequences of events. These scores will consider various criteria, such as matches, substitutions, insertions/deletions (indels), and the order of events. Substitutions involve replacing one event in a sequence with another, while indels refer to adding or removing events from the sequence. These similarity scores

are essential in fields like sequence alignment, phylogeny, and genetic analysis, where precise measurement of sequence similarity is critical [22]. Sequence alignment is typically divided into two categories: global alignment and local alignment. Global alignment, performed by algorithms such as Needleman-Wunsch, aligns sequences end-to-end to provide a comprehensive comparison [23]. On the other hand, local alignment, often executed by algorithms like Smith-Waterman, focuses on identifying highly similar sub-sequences within larger sequences [24]. In this study, we chose the global alignment method as it aligns with our task of identifying exact sequences of events. However, recognizing activities through local alignment of events remains an interesting option. Below is an explanation of how the Needleman-Wunsch method operates.

A score matrix $S$ of size $(n+1) \times (m+1)$ is initialized, where $n$ is the length of the actual labels and $m$ is the length of the predicted labels. The first row and column are initialized with a gap penalty for an insertion or deletion, here defined as -1.

To fill the Score Matrix, each cell $S[i][j]$ is calculated as the maximum of:

$$S[i][j] = \max \begin{cases} S[i-1][j-1] + M(P[i], R[j]) \\ S[i-1][j] - d \\ S[i][j-1] - d \end{cases}$$

where $M(P[i], R[j])$ is the match score, typically $+1$ for an exact match and $-d$ for a mismatch.

Backtracking is then performed to obtain the optimal alignment. Starting from $S[n][m]$, the algorithm traces back to find the optimal alignment, counting the exact matches, insertions, and deletions.

Finally, accuracy percentage is calculated using the number of exact matches $C$:

$$\text{Accuracy} = \frac{C}{n} \times 100$$

## IV. Experiments and Results

To evaluate the accuracy of the recognition algorithm on its own, the system was first run without the segmentation step. The model was trained using 80% of the Aruba dataset and then tested on the remaining 20% annotated segments. The results demonstrated a high accuracy rate of 98.3% and an F1 score of 0.982, which indicates that the classification model is highly effective in terms of both precision and recall.

Figure 7 presents the classification results, including the confusion matrix and accuracy percentages for all activities.

The system appears to have difficulty distinguishing between 'Enter_Home' and 'Leave_Home' as well as 'Meal_Preparation' and 'Wash_Dishes' due to similarities in sensor activations. To clarify, here is what each of the activities "Enter_Home" and "Leave_Home" entail:

- 2010-11-04 11:43:53 D004 OPEN Leave_Home begin
- 2010-11-04 11:43:55 M030 ON

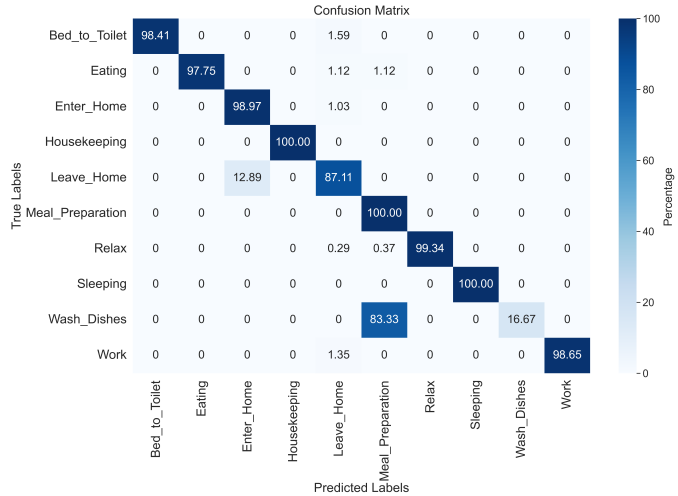


Fig. 7. Confusion matrix and accuracy percentages for all activities

- 2010-11-04 11:43:58 D004 CLOSE Leave_Home end
- 2010-11-04 11:47:48 D004 OPEN Enter_Home begin
- 2010-11-04 11:47:48 M030 ON
- 2010-11-04 11:47:50 M030 OFF
- 2010-11-04 11:47:52 D004 CLOSE Enter_Home end

In our evaluation of the comprehensive system, which includes dynamic segmentation and activity recognition, we used 80% of the Aruba dataset for training and reserved the remaining 20% for testing. The segmentation process generated 1629 segments, slightly more than the actual 1611 segments due to inherent variability in user task execution. It is important to note that with dynamic segmentation, the generated segments may not perfectly align with the actual segments, and the number of predicted activities may not perfectly match the number of actual activities. Therefore, using confusion matrices to evaluate the performance of the activity recognition phase may not be suitable. Instead, we used the Needleman-Wunsch method to measure consistent activities while considering variations between predicted and actual segments. Figure 8 illustrates the global alignment between predicted and actual activities as performed by the Needleman-Wunsch algorithm.

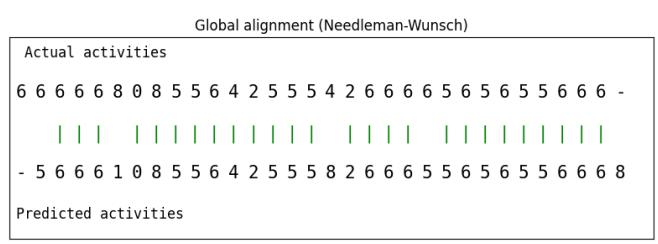


Fig. 8. Global alignment of actual and predicted activities using the Needleman-Wunsch method.

Of the 1629 predicted activities, 1340 corresponded to actual activities, resulting in an accuracy rate of 83.2%.

In Table II, we present a comparative analysis of our method's accuracy performance in contrast to similar methods that focus on spatiotemporal data and real-time segmentation. Our approach has demonstrated significant improvement compared to the performance of these studies.

| Approach | Ambient sensors | Dynamic segmentation | Spatiotemporal | Classifier | Accuracy | Dataset |
|---|---|---|---|---|---|---|
| Z. Xu et al. [3] | Yes | Yes | Yes | CNN-LSTM | 0.78 | CASAS |
| H. Najeh et al. [2] | Yes | Yes | Yes | CNN2D | 0.81 | CASAS |
| Our method | Yes | Yes | Yes | CNN2D | **0.83** | CASAS |

TABLE II
COMPARATIVE ANALYSIS OF DIFFERENT APPROACHES

## V. CONCLUSION

Our research introduces a robust method for real-time recognition of Activities of Daily Living (ADLs) using IoT data in smart home environments. Our results confirm the effectiveness and reliability of our approach, particularly when compared to state-of-the-art dynamic segmentation methods that leverage spatiotemporal data from IoT sensors. A key contribution of our work is an improved segmentation process that evaluates predicted activities by globally aligning sequences of events, enhancing the system's ability to detect activities more accurately.

In addition to segmentation, computational efficiency is crucial in real-time human activity recognition (HAR), particularly on embedded devices. Building on insights from Najeh et al. [24], who optimized HAR algorithms for resource-constrained hardware, we anticipate that our system, similar to theirs, can maintain high accuracy on low-power, low-cost devices. Our system differentiates itself from that of Najeh et al. by incorporating an additional layer dedicated to analyzing ADL progress using the Needleman-Wunsch algorithm. This layer is optimized to ensure compatibility with embedded systems that have limited resources. However, further testing across diverse hardware and software configurations is planned to fully validate our approach.

In our results, we occasionally encountered difficulties in distinguishing between very similar activities. This challenge is addressed in the literature through various methods, such as advanced feature extraction (e.g., temporal, frequency-based, and spatiotemporal features), deep learning approaches (e.g., CNNs, RNNs, and transformers), and methods incorporating context and multi-modal sensor fusion. A combination of these approaches is known to improve the system's ability to differentiate between similar activities. Our hybrid approach integrates spatiotemporal features, deep learning models, multi-modal sensor data fusion, and contextual segmentation, all tailored to individual user behavior. Looking ahead, we plan to explore the incorporation of additional sensors (e.g., entry/exit points) and the use of more contextual information, such as activity duration and sensor-activity signatures. However, adding these elements will incur additional costs in terms of data collection and integration, which must be carefully evaluated.

Furthermore, it is important to emphasize that the spatial arrangement of sensors during installation significantly affects performance. In the future, we aim to develop a sensor installation simulator to refine sensor placement, thereby improving the accuracy and reliability of ADL detection across various smart home configurations.

## ACKNOWLEDGMENT

This work was conducted within the framework of Chaire Maintien à Domicile (M@D), supported by the foundations of the University of Southern Brittany. It is also part of operation 4.2.7: Measurement and Analysis of Activities of Daily Living (ADL) as part of the Handicap Innovation Territoire (HIT) project supported by Lorient aglomeration. We would like to extend our thanks to Ndeye-Maguette Diagne, intern at the Chaire M@D, for her invaluable contribution to this project and to the elaboration of part of this work.

## REFERENCES

[1] M. Memon, S. R. Wagner, C. F. Pedersen, F. H. A. Beevi, and F. O. Hansen, 'Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes', Sensors, vol. 14, no. 3, pp. 4312–4341, Mar. 2014, doi: 10.3390/s140304312.

[2] H. Najeh, C. Lohr, and B. Leduc, 'Convolutional Neural Network Bootstrapped by Dynamic Segmentation and Stigmergy-Based Encoding for Real-Time Human Activity Recognition in Smart Homes', Sensors, vol. 23, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/s23041969.

[3] Z. Xu, G. Wang, and X. Guo, 'Online Activity Recognition Combining Dynamic Segmentation and Emergent Modeling', Sensors, vol. 22, no. 6, Art. no. 6, Jan. 2022, doi: 10.3390/s22062250.

[4] Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., and Selby, P. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 22(4), 714-724.

[5] Ramadan, A., De Lamotte, F. F., and Julien, N. (2023, September). Digital Twin Architecture for Ambient Assisted Living. In International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing (pp. 124-134). Cham: Springer Nature Switzerland.

[6] S. Katz, A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe, 'Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function', JAMA, vol. 185, no. 12, pp. 914–919, Sep. 1963, doi: 10.1001/jama.1963.03060120024016.

[7] Coutton, V. (2001). Évaluer la dépendance à l'aide de groupes iso-ressources (GIR): une tentative en France avec la grille aggir. Gérontologie et société, 2499(4), 111-129.

[8] L. M. Verbrugge and A. M. Jette, 'The disablement process', Soc. Sci. Med. 1982, vol. 38, no. 1, pp. 1–14, Jan. 1994, doi: 10.1016/0277-9536(94)90294-1.

[9] J. N. Morris, B. E. Fries, and S. A. Morris, 'Scaling ADLs within the MDS', J. Gerontol. A. Biol. Sci. Med. Sci., vol. 54, no. 11, pp. M546-553, Nov. 1999, doi: 10.1093/gerona/54.11.m546.

[10] L. N. Gitlin, M. Corcoran, L. Winter, A. Boyce, and W. W. Hauck, 'A randomized, controlled trial of a home environmental intervention: effect on efficacy and upset in caregivers and on daily function of persons with dementia', The Gerontologist, vol. 41, no. 1, pp. 4–14, Feb. 2001, doi: 10.1093/geront/41.1.4.

[11] G. Diraco, G. Rescio, P. Siciliano, and A. Leone, 'Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing', Sensors, vol. 23, no. 11, Art. no. 11, Jan. 2023, doi: 10.3390/s23115281.

[12] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, 'CASAS: A Smart Home in a Box', Computer, vol. 46, no. 7, Jul. 2013, doi: 10.1109/MC.2012.328. [8] P. Skocir, P. Krivic, M. Tomeljak, M. Kusek, and G. Jezic, 'Activity Detection in Smart Home Environment', Procedia Comput. Sci., vol. 96, pp. 672–681, Jan. 2016, doi: 10.1016/j.procs.2016.08.249.

[13] P. Skocir, P. Krivic, M. Tomeljak, M. Kusek, and G. Jezic, 'Activity Detection in Smart Home Environment', Procedia Comput. Sci., vol. 96, pp. 672–681, Jan. 2016, doi: 10.1016/j.procs.2016.08.249.

[14] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, 'A Survey on Ambient Intelligence in Healthcare', Proc. IEEE, vol. 101, no. 12, pp. 2470–2494, Dec. 2013, doi: 10.1109/JPROC.2013.2262913.

[15] S.-L. Chua, L. K. Foo, and H. W. Guesgen, 'Predicting Activities of Daily Living with Spatio-Temporal Information', Future Internet, vol. 12, no. 12, Art. no. 12, Dec. 2020, doi: 10.3390/fi12120214.

[16] H. Zhu, S. Samtani, and R. Brown, 'A Deep Learning Approach for Recognizing Activity of Daily Living (ADL) for Senior Care: Exploiting Interaction Dependency and Temporal Patterns', MIS Q., vol. 45, pp. 859–896, Jun. 2021, doi: 10.25300/MISQ/2021/15574.

[17] C. Chen, B. Das, and D. Cook, 'A Data Mining Framework for Activity Recognition in Smart Environments', presented at the Proceedings - 2010 6th International Conference on Intelligent Environments, IE 2010, Aug. 2010, pp. 80–83. doi: 10.1109/IE.2010.22.

[18] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, 'A sequential deep learning application for recognising human activities in smart homes', Neurocomputing, vol. 396, pp. 501–513, Jul. 2020, doi: 10.1016/j.neucom.2018.10.104.

[19] J. Wan, M. J. O'Grady, and G. M. P. O'Hare, 'Dynamic sensor event segmentation for real-time activity recognition in a smart home context', Pers. Ubiquitous Comput., vol. 19, no. 2, pp. 287–301, Feb. 2015, doi: 10.1007/s00779-014-0824-x.

[20] D. Bouchabou, S. M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, 'Using Language Model to Bootstrap Human Activity Recognition Ambient Sensors Based in Smart Homes', Electronics, vol. 10, no. 20, Art. no. 20, Jan. 2021, doi: 10.3390/electronics10202498.

[21] Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. European Conference on Computer Vision, 818-833.

[22] Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772–780.

[23] Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology, 48(3), 443-453.

[24] Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. Journal of Molecular Biology, 147(1), 195-197.

[25] Najeh, H., Lohr, C., and Leduc, B. (2024). Real-Time Human Activity Recognition on Embedded Equipment: A Comparative Study. Applied Sciences 14, no. 6: 2377.