Information Fusion 123 (2025) 103307

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Fuzzy processing applied to improve multimodal sensor data fusion to discover frequent behavioral patterns for smart healthcare

Carlos Fernandez-Basso [a,b,*], David Díaz-Jimenez [b], Jose L. López [b], Macarena Espinilla [b]

[a] Department of Computer Science, University of Granada, C/ Periodista Daniel Saucedo, Granada, Spain
[b] Department of Computer Science, University of Jaén, Campus Las Lagunillas s/n, Jaén, Spain

ARTICLE INFO

Keywords:
Data fusion
Sensor data
Sensor fuzzification
Smart healthcare

ABSTRACT

The extraction and utilization of latent information from sensor data is gaining increasing prominence due to its potential for transforming decision-making processes across various sectors. Data mining techniques provide robust tools for analyzing large-scale data generated by advanced network management systems, offering actionable insights that drive operational efficiency and strategic improvements. However, the sheer volume of sensor data, combined with challenges related to real-world sensor deployment and user interaction, necessitates the development of advanced data fusion and processing frameworks. This paper presents an innovative automatic fusion and fuzzification methodology designed to integrate multi-source sensor data into coherent, high-quality intelligent outputs. By applying fuzzy logic, the proposed system enhances the interpretability and interoperability of complex sensor datasets. The approach has been validated in a real-world scenario within sensorized homes of Type II diabetic patients in Cabra (Córdoba, Spain), where it aids healthcare professionals in monitoring and optimizing patient routines. Experimental results demonstrate the system's effectiveness in identifying and analyzing behavioral patterns, highlighting its potential to improve patient care through advanced sensor data fusion techniques.

1. Introduction

Nowadays, advanced sensor technologies and healthcare infrastructures in hospitals [1], nursing homes [2] or individual homes [3] for dependent adults can generate thousands of readings every minute from a wide range of sensors. Health sector entities are progressively recognizing the immense potential [1] that can be harnessed through the analysis and utilization of this information. Consequently, the prevailing trend is to store and process such data with the aim of extracting valuable insights regarding patterns of activity, behavioral tendencies, and identification of individuals' actions and behaviors [4,5].

In order to collect data in the real context, it is necessary to develop and deploy systems based on Internet of Things (IoT) technology. This type of systems [1,3,6] are characterized by sensorised devices that are capable of recording the activity of the users, communicating this information wirelessly [7] to the different elements of the system and storing all the data in the cloud. Different types of sensorised devices can be found in this field [8]: device free, tagged object and wearables. These wearable devices [9] are highly specialized in the healthcare field (Internet of Medical Things) and are characterized by being attached to the user's body and measuring different types of physiological signals, such as heart rate, temperature, breathing rate, among others.

Due to the large amount of information generated, the main challenge [10,11] in processing sensor-generated data is to find efficient methods and techniques to improve the quality of the measurement data obtained. One of the key steps is data fusion during the pre-processing phase because, in most cases, data from different sources must be imported. In addition, these data are usually of considerable size, as they are generated at low frequencies by numerous sensors, detailing an inherent context and correlation between them. Therefore, pre-processing and data fusion become integral elements to obtain a database that encapsulates the relationship and knowledge associated with the different sensors.

Sensor measurement data is commonly produced using numerical measurements that encapsulate values within a continuous range. This characteristic often complicates large-scale analysis due to its fine granularity. A fundamental approach comprises designating divisions of the

* Corresponding author.
E-mail address: cjferba@ugr.es (C. Fernandez-Basso).

https://doi.org/10.1016/j.inffus.2025.103307
Received 17 October 2024; Received in revised form 20 March 2025; Accepted 5 May 2025
Available online 20 May 2025

possible values into intervals, aiding algorithms in data processing. This partitioning, however, could be subjected to significant inconveniences since the results could vary drastically based on the applied division, and this division might not be intuitive for subsequent result analysis.

Fuzzy sets have proven to effectively represent data with soft edges, thereby enhancing the interpretability of results by associating meaningful linguistic labels with the resultant fuzzy sets [12–14]. Alternatively, interval programming methods can be utilized to address the imprecise characteristics contained in the data.

Although the literature presents various Artificial Intelligence (IA) and data fusion methods applied to healthcare, fuzzy logic [15] continues to be widely used due to its simplicity. This feature facilitates its use in environments with limited computational resources and real-time applications, and enhances the interpretability and trustworthiness [16] of the implemented AI model. Nevertheless, fuzzy logic-based systems still rely strongly on expert knowledge to define their rules and membership functions.

Due to these advantages, fuzzy logic is still widely used in various fields of application, such as healthcare [17], cybersecurity [18] and sentiment analysis [19]. In addition, it is frequently integrated into more complex hybrid models [20] that combine its characteristics with other methodologies, such as fuzzy machine learning systems, thus taking advantage of the complementary strengths of different approaches.

In this article, we propose a fusion that enriches sensor data to improve and create a database containing the knowledge and relationships between sensors. In addition, a fuzzification algorithm, which does not require expert knowledge, is used to preprocess the data adequately and, in a subsequent step, apply fuzzy data mining techniques to discover potentially helpful information that may be hidden in the data.

In particular, we have applied association rule discovery [12,21], an unsupervised technique capable of finding existing relationships between variables and their values. Furthermore, these results allow us to find frequent behavior in the behavior of people who live in sensorized houses. Additionally, sensors linked with activities and person-locating functionalities allow us to comprehend routines-based frequent behaviors or activities [1,4,6,22].

The entire system has been successfully applied in a sensorized house located in Cabra (Spain), obtaining a set of patterns that describe the behavior and activity patterns of the tenant of the house. However, the presented approach could also be applied in other types of houses, residences with more tenants or different sensors.

The patterns obtained describe the behavior and daily routines of the tenant, but could also help to discover routines that are not beneficial or unhealthy for people. Behavioral extraction is crucial in the healthcare field, for example in patients with type II diabetes mellitus. In this case, it is essential for healthcare professionals to know whether patients' routines are healthy in order to modify their treatment [3].

Therefore, this work proposes a fuzzy logic-based system for the discovery of behavioral patterns in older people through the analysis of data collected from multimodal sensors. These sensors have been deployed in a typical real home environment using a big data architecture, and present a novel process for modeling the uncertainty of the data using fuzzy techniques.

The primary contributions of this work are enumerated below:

1. Development of an automatic fusion and fuzzification methodology that integrates multi-source sensor data into a coherent and high-quality dataset while leveraging fuzzy logic to enhance the interpretability and interoperability of the data.
2. Application of fuzzy association rule mining to discover frequent behavioral patterns in smart healthcare environments, validated through a real-world deployment in a sensorized home with monitored Type II diabetic patients, demonstrating its effectiveness in optimizing daily routines.

The work is structured as follows. The next section reviews previous related research and introduces the necessary background of the related concepts. Section 3 describes the design of our system focusing on the developed fuzzification algorithm. The Section 4 presents our results for the use case of house located in Cabra (Spain). Finally, in Section 5 we summarize the conclusions and present possible future research lines.

2. Related works

Activity detection, particularly in human behavior and daily living, is a critical area of research with significant implications for healthcare, elder care, and pervasive computing environments. The primary goal of activity detection applications is to accurately recognize and monitor human activities using various sensing technologies and computational methods. This capability is crucial for developing intelligent systems that can provide timely assistance, enhance safety, and improve the quality of life for individuals, especially the elderly or those with special needs.

The state of the art in activity detection has evolved to include a variety of approaches and technologies. One standard method involves the use of wearable sensors, such as accelerometers, to capture motion data that can be analyzed to identify specific activities [23]. These sensors are advantageous due to their ability to gather data unobtrusively, providing continuous monitoring without significantly impacting the user's daily routine.

A range of studies have explored the use of fuzzy processing and fuzzy association rules in activity recognition. In [24,25] both propose methods for recognizing human activities, with the former using a Neuro-Fuzzy Finite State Machine and the latter using a fuzzy association rule mining algorithm. [26] focus on improving the performance of existing methods, with Borges adapting the k-Nearest Neighbors classifier to incorporate fuzzy rules and Chiang developing a fuzzy computing model for activity recognition in ubiquitous healthcare. [12,27] both address the challenge of recognizing activities in smart homes, with Gayathri proposing a fuzzy ontology-based approach and Mohmed introducing a clustering-based fuzzy finite state machine. In [28], fuzzy processing is extended to complex activity recognition. Sakr employs a combination of emerging patterns and fuzzy sets, while Nguyen proposes a method that integrates Fuzzy Markov Logic Networks with Fuzzy Qualitative Spatio-Temporal Activity Graphs.

Fuzzy logic has emerged as a powerful tool in activity detection, offering a way to handle the uncertainty and imprecision inherent in human activities. Fuzzy logic systems are capable of modeling the vagueness and variability of human behavior, making them well-suited for applications where traditional binary logic systems might fail [29]. These systems use fuzzy rules and inference mechanisms to process sensor data and infer activities, allowing for more flexible and adaptive activity recognition.

Neuro-fuzzy classifiers combine neural networks and fuzzy logic to enhance the classification and recognition processes. These classifiers are particularly effective in scenarios where the training data includes noise and variability, common in human activity data [13]. The neuro-fuzzy systems are trained using algorithms that adjust membership functions and rule parameters to optimize performance.

The integration of multiple sensors requires the use of data fusion techniques. This methodology represents a growing trend in this field, as it enhances the robustness and accuracy of artificial intelligence models and is commonly applied at different levels [15,30]: data, features, and decisions. In the classical literature [30], several techniques for data fusion can be found, such as Bayesian probabilistic methods, filters (e.g., Kalman filter), fuzzy logic and rule-based systems, and machine learning approaches. In recent years, driven by the proliferation of artificial intelligence, new data fusion methods have emerged, including deep learning [31] through deep neural networks and Dempster-Shafer evidence theory [32,33], an extension of traditional probabilistic methods. However, these recent approaches involve higher computational complexity, are more challenging to interpret and explain, and complicate the appropriate parameter tuning process. Conversely, fuzzy logic provides good results naturally and straightforwardly in environments characterized by ambiguity or imprecision, reduces computational cost, and is easier to interpret and parameterize.

In the specific case of activity recognition, combining data from multiple sources, such as video cameras and sensors, can provide a more comprehensive understanding of the user's environment and behavior [34].

In summary, the field of activity detection is characterized by a diverse array of methods and technologies, each offering unique advantages and suited to different applications. The ongoing development of these systems holds promise for creating more intelligent and responsive environments, ultimately enhancing safety and quality of life for users.

2.1. Data mining system

Data mining techniques have been extensively utilized across various scientific fields. One such field is energy, where notable studies [21,35,36] demonstrate the application of traditional data mining techniques to extract information related to construction. Similarly, in social science, numerous studies [37–39] employ text pre-processing to analyze data. Other sciences, such as Physics and Astronomy, apply pre-processing techniques to images [40–43]. Data mining techniques are also widely applied in the medical domain, as observed in studies like [10,44,45]. Medical research utilizing data mining generally falls into two categories: studies focusing on imaging data [46–49] and those analyzing human activity [2,50].

A specific study in the health sector, such as the analysis of brain signals [45], encounters challenges like small sample sizes and noisy signals. The authors propose a viable method for detecting and distinguishing directions from Electroencephalography (EEG) signals by utilizing feature extraction techniques for brain signal processing.

In another study [46], researchers developed an automated system for the extraction and classification of tumors from magnetic resonance images. This system consists of five main steps: tumor contrast, tumor extraction, multi-model feature extraction, feature extraction, and classification. Other studies, such as [51–54], focus on unsupervised algorithms.

Different data types often require distinct pre-processing techniques to enhance quality and outcomes. Structured data generally benefit from classical pre-processing techniques like data cleaning, data integration, data transformation, and data reduction [55–57]. Big data technology also finds numerous applications in the healthcare sector [11,44,58,59], including predictive modeling, clinical decision support, disease surveillance, and research. Big data analytics often leverages data mining methods such as classification, clustering, and regression.

In our study, we propose a series of techniques that differ from previous studies by uniquely applying and unifying these methods. This approach ensures that, by the end of the procedure, we have a refined set of data ready for knowledge extraction and interpretation.

We have enriched our dataset by incorporating various diagnoses occurring during the patient's medical action protocol to address the proposed diagnosis using external sources. Basic pre-processing techniques were employed, such as detecting and eliminating missing values within the medical data center and removing outliers by selecting non-relevant fields.

The next step in our procedure involves transforming our data using fuzzy techniques, ultimately creating association rules to detect relationships between different fields analyzed in our study.

2.2. Fuzzy-based system

Most of the challenges facing sensor data are related to structuring the information within databases and representing it for correct interpretation by end-users. It is essential to apply fuzzy techniques to transform imprecise data, including sensor data, into accurate and interpretable information for end users. In ¨citecarlos energia, the authors propose new measures of accuracy and usability for extracting fuzzy association rules from energy sensor databases. Their approach allows for a significant reduction in the number of rules without losing important information.

An intriguing example of the state of the art is the study [14], which focuses on applying fuzzy techniques in healthcare. It demonstrates how sensors can be used to develop a system that recommends specific prescriptions for patients with diabetes.

In addition, there are theoretical studies [34,60] that explore the use of fuzzy programming algorithms to solve linear problems, providing results in a way that eliminates uncertainty.

These systems can be fully scalable by employing distributive algorithms [61]. This study presents biomedical data stored in the cloud, showing how such algorithms are ideal for solving large-scale problems.

Our study focuses on the fuzzification of particular home sensors, analyzing how proximity and sensor data plus some wearables can yield results that categorize patients' activities as recurrent to understand their behavior and daily routine. This output and information will be obtained by processing the data and using fuzzy logic to add the proximity of individuals to places in the house, as well as giving more understandable and interpretable linguistic fuzzy labels to the continuous data.

2.3. Activity recognition

Another significant aspect addressed in this study is the analysis of activity recognition. Activity recognition involves identifying and categorizing individual actions using data from various sensors [1,4,6,22]. While activity data are often reported statistically, mainly in the context of academic research to inform health systems and public health agencies, few studies focus on dynamic activity recognition for initial diagnosis or continuous monitoring.

One study [62] focuses on recognizing activities to categorize the degrees of autism in children based on the different activity patterns they exhibit. In another study [63] researchers aimed to detect unnecessary medical procedures by analyzing patients' activity data. For instance, they identified patients with upper gastrointestinal bleeding and patients with unspecified gastrointestinal tract bleeding to examine the correlation between their activities and laboratory results, such as calcium levels and hemocytes in the blood. Their experiments involved labeling promising activity patterns and grouping patients according to their activity data.

A compelling case study is presented on diabetic retinopathy, an activity-induced condition generated by diabetes [5,9]. The study demonstrates how applying classification techniques, fuzzy logic, and data balancing methods can determine which patients are most at risk of developing diabetic retinopathy based on their activity patterns.

Our study extends the scope by focusing on activity recognition in medical records. By analyzing activity data, we can categorize patients as simple, standard, or complex using fuzzy logic, which helps us better interpret their medical history and condition based on their recognized activities.

3. A fuzzy-based system for pattern mining in big data

This work proposes a fuzzy logic-based system for the discovery of behavioral patterns in older people through the analysis of data collected from multimodal sensors. These sensors have been deployed in a typical real home environment using a big data architecture, and

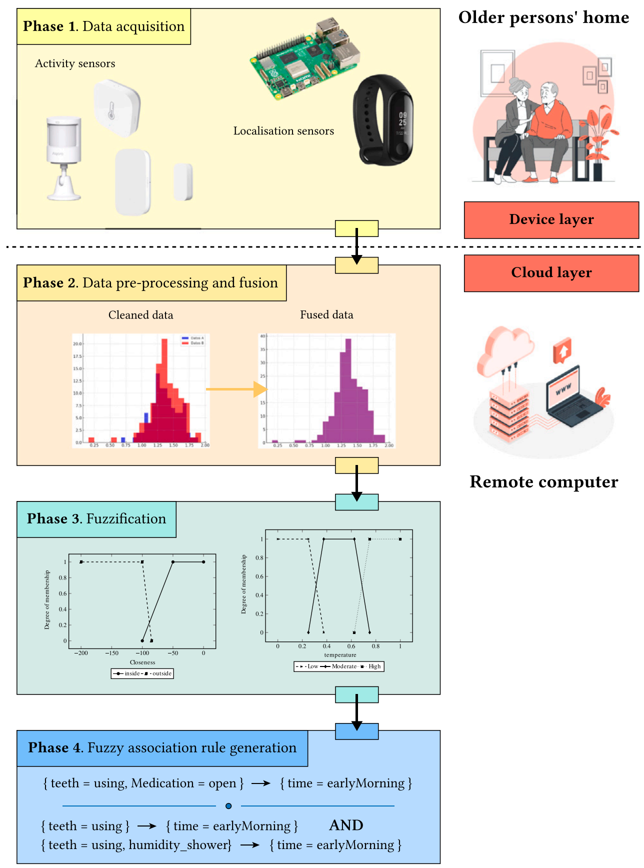

Fig. 1. Summary of the system workflow which has been divided into four main phases.

present a novel process for modeling the uncertainty of the data using fuzzy techniques.

In this research work, the presented process is divided into four main blocks: data acquisition, data pre-processing and fusion, fuzzification and generation of fuzzy association rules for automatic behavior extraction. First, data is collected from multimodal sensors using an architecture based on the IoT paradigm. Next, the collected data is processed and fused, leading to the fuzzification process. Finally, automatic behavior extraction is performed. The whole workflow is summarized in Fig. 1.

Each stage of the process is described in detail below.

3.1. Data acquisition

Data acquisition is the first stage of the proposed process. This stage is fundamental to provide input data for the following phases and thus to discover common behavioral patterns of older people. In this work it has been chosen to use multimodal sensors that record all the activities of the user in his habitual home.

This type of sensor has a number of advantages over other types of proposal. Firstly, these devices provide complementary information by capturing different types of data simultaneously, such as movement, location, temperature or interaction with the environment. In addition, combining data from different sources reduces potential errors and provides greater accuracy. Finally, devices can monitor the user in real time by processing and sending data instantly.

In the literature, different types of sensors can be found in smart environments. Considering the proposed data acquisition in older people's homes, systems based on device-free sensors are the most convenient because the user does not need to interact directly with the devices. However, in real multi-occupancy environments, some identification of the occupant is required to recognize who is performing the detected activity. Therefore, to achieve activity data collection, a hybrid device-free data acquisition system combined with wearable devices is used, based on the proposal of David et al. [3].

The system deploys a total of five distinct sensorised devices, which are as follows: an infrared-based motion sensor, an opening and closing sensor, an ambient sensor, a fixed location device and an activity wristband. All pertinent information concerning these devices is provided in meticulous detail within Table 1.

Specific devices are employed to monitor the daily behaviors of older adults in their domestic environments. The system records information pertinent to the most crucial health-related behaviors, including physical activity, sleep, hygiene, mealtime routine and medication adherence. Additionally, the system utilizes an indoor location system to ascertain the individual responsible for activating the sensors. These types of systems are widely employed within the scientific community and typically utilize radio frequency technology. In this scenario, the system is based on Bluetooth Low Energy (BLE) technology and uses the Received Signal Strength Indicator (RSSI) as a feature.

The following provides an overview of the types of events related to the different behaviors detected by the devices described in Table 1:

- **Motion sensor**: sleep, personal hygiene, mealtime routine.
- **Open and close sensor**: physical activity, medication.
- **Ambient sensor**: personal hygiene.
- **Fixed tracking devices**: location indoors.
- **Activity wristband**: physical activity, sleep, location indoors.

The system is based on an edge-fog-cloud architecture for the acquisition of data. In this environment, a variety of communication protocols are employed for this purpose. These include the MQTT protocol for communication between edge-fog layer devices, the ZigBee protocol for receiving data from commercial devices, such as Aqara devices, and the HTTPS protocol for communication between fog and cloud layer devices. Further clarification on all of these points can be found in [3].

Considering the sensitive nature of the data, several techniques are applied to ensure the protection of information and the integrity of communications at different levels: communication and storage. At the communication level, a range of mechanisms are employed, including:

- **MQTT protocol**: message encryption using TLS/SSL, authentication and authorization.
- **Web service**: message encryption using HTTPS and use of Json Web Tokens for authentication.
- **Database**: message encryption between this component and any other application, authentication and authorization.

Finally, at the database storage level, all information is encrypted using the AES-256 algorithm.

3.2. Pre-processing

In the domain of data pre-processing, an array of diverse methods is commonly employed. Notably among them is the technique of data granularity normalization. Sensor data, being collected at varying time intervals, necessitates alignment to standard time points. This particular transformation is illustrated in curated examples like Table 2, wherein time intervals have been collated and grouped per every 5 min of recorded data. Yet, crucial to note here is the flexibility of adjustment according to user specifications or analytic requirements.

The methodologies utilized for this kind of grouping activity can differ significantly in alignment with the nature of the data under scrutiny. Taking continuous data as a case in point, one approach could be to use average value computations, demonstrated in Table 2. Alternatively,

Table 1
A comprehensive description of the specific sensors employed in the system in the data acquisition phase.

Device	Data	Type	Range	Activation	Other features
Motion sensor[1]	Motion	Binary	–	Upon the occurrence of motion	The device offers 170° coverage and a range of up to seven meters. Furthermore, the sensitivity and frequency of activation can be modified according to the user's preferences. The device has a dimensions of 55 *x* 37 *x* 70 mm and is equipped with a battery that is capable of powering the device for a period of two years
Open and close sensor[2]	Mechanism state	Binary	–	When users interact with door and window mechanisms or other similar elements	The device has a dimensions of 41 *x* 22 *x* 11 mm and is equipped with a battery that is capable of powering the device for a period of two years
Ambient sensor[3]	Temperature	Integer	[−20, 50] °C	When significant temperature variation occurs	The device has a dimensions of 36 *x* 36 *x* 9 mm and is equipped with a battery that is capable of powering the device for a period of two years
	Humidity	Integer	[0, 100] % RH	When a significant variation in humidity occurs	
Fixed tracking devices[4]	RSSI	Integer	[0, −200]	Send data every three seconds if the wristband is found	General computing unit used to implement the anchor element of the indoor localization system. The device measures 85 *x* 49 mm
Activity wristband[5]	Steps	Integer	[0, ∞)	Sends data every minute	The device is equipped with an integrated accelerometer and gyroscope, enabling it to calculate steps. Its dimensions are 7.9 *x* 46.9 *x* 12 mm

1 https://www.aqara.com/en/product/motion-sensor-p1/.
2 https://www.aqara.com/en/product/door-and-window-sensor/.
3 https://www.aqara.com/en/product/temperature-humidity-sensor/.
4 https://www.raspberrypi.com/products/raspberry-pi-4-model-b/.
5 https://www.mi.com/cl/mi-band-3/.

Table 2
Example of the temporal granularity processing of some temperature data.

Time	Temperature
1/1/2020 15:14:30	16
1/1/2020 15:16:45	17
↓	
Time	Temperature
1/1/2020 15:15:00	16.5

one could opt for the use of the most recently documented sensor data value.

Crucially, this transformative exercise engenders conditions conducive towards the formation of a transactional data framework — an operative form particularly well adapted for data mining techniques. Extending our perspective on this, such normalization of data granularity not only makes the dataset systematically rigid but also highly conducive for temporal data interpretation and lends a sequential sense to otherwise scattered intervals.

Moreover, different formulas could be adopted to handle diverse data types. For binary or categorical data, the mode, simply put, the most appeared category within an interval, can be deployed. Standardization methods might follow the normalization process to further streamline the data to align with the machine algorithm feed requirements.

The gathered data is partially sourced from counters, which operate as accumulators. This data category is exemplified in Fig. 2, showing a consistently escalating value indicative of a person's step count throughout the day. This variable type has undergone a transformation function, translating into a representation of an individual's activity level for a specific duration.

Consequently, each data instance signifies the activity level during that respective timeframe. After processing the data and obtaining the desired granularity, transactions with lost values and transactions containing outlier measurements made by sensors are eliminated. The *sklearn* python library [64] was used to determine these outliers, specifically using the elliptic envelope fitting function [65].

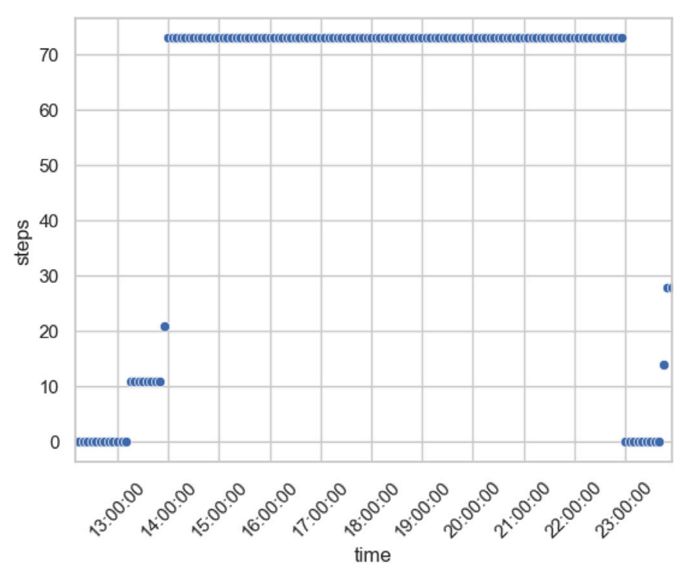

Fig. 2. Sensor steps.

3.3. Fusion data

The data collected by the sensors is often embedded with certain contextual significance, calling for an enrichment of the sensor's relationship with the user and the household environment. This renders the concept of data fusion relevant — integrating multiple data sources to produce more consistent, accurate, and useful information. The context encapsulates the interactions between the user and the varying sensor locations within the house, highlighting the significance of 'anchors'. These 'anchors' are utilized as guiding reference points to map a user's positional proximity within a zone of the house. Values approaching zero harness interpretation of an increase in closeness to a particular position, as exemplified in Fig. 3.

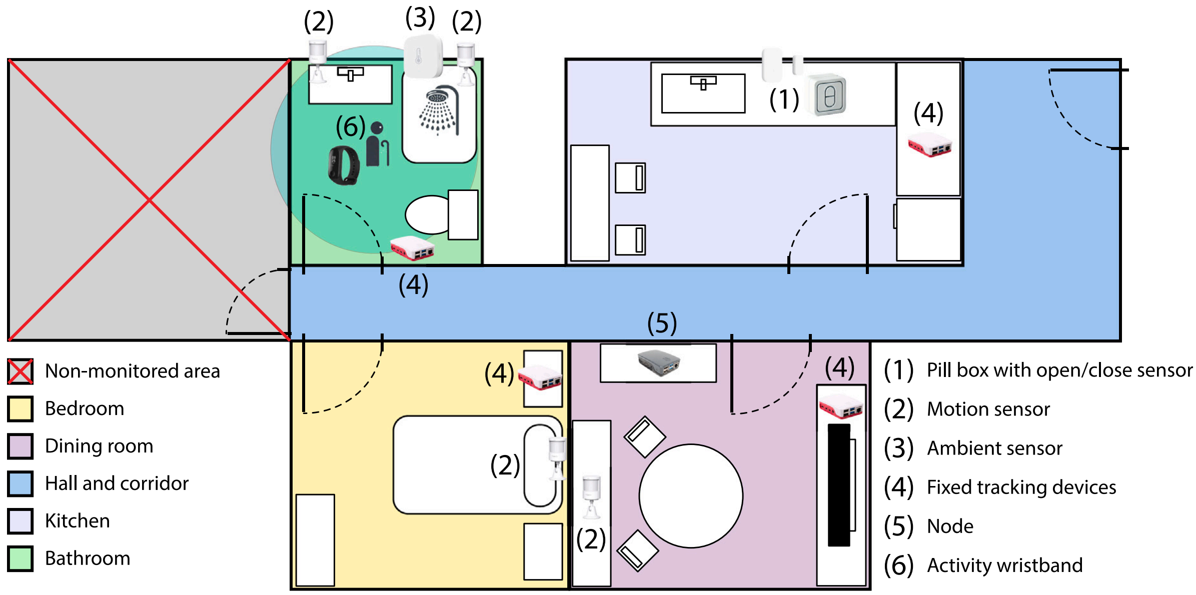

Fig. 3. Distribution of the dwelling used in the case study, location of the devices used and example of the activation range of the monitored elderly person obtained based on the proximity of these devices.

Data fusion plays a pivotal role in this context, adroitly synthesizing information from disparate sources, in this case, the user and the sensor environment. It enhances the understanding of the user's positioning and movement, augments data quality and results in a more comprehensive simulation of the household environment. This smart fusion of information from different contexts thus facilitates the next phases with more coherent and in-depth results.

In order to relate this context with the values of the inhabitants of the household, a process of enrichment and transformation of the database will be carried out to obtain a database with this added knowledge. To this end, the numerical value of the proximity will be transformed into two membership values that will mean that the person is in the room or not in the room. These membership values will help us count the transactions the sensors interact with the user and those that do not.

We can see in Fig. 4 how the values will be fuzzified using 2 labels for proximity. With them, the transactions that are important will be managed, that is, what sensor values have to be in the transaction because the person is in that area of the house. We can see in Fig. 4 how the values will be fuzzified using 2 labels for proximity. With them, the important transaction will be managed, that is, what sensor values have to be in the transaction because the person is in that area of the house. To exemplify this, we have Fig. 4 in which we see how on the left are all the sensors and areas close to the house and in the figure on the left what data would be the data that would be used to add to the transaction.

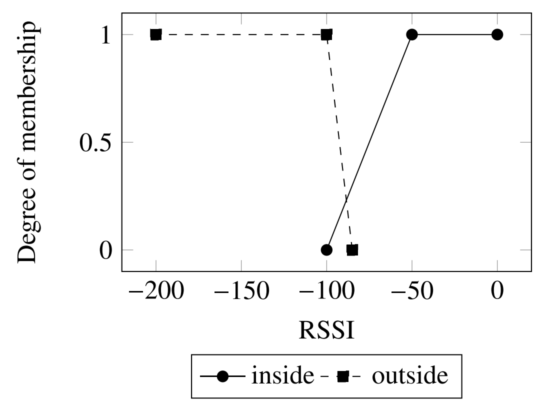

Fig. 4. Fuzzy membership functions to determine closeness as a function of RSSI value.

3.3.1. Fuzzification

Collecting data from various sources, such as sensors, counters, or building occupancy metrics, often challenges understanding the results obtained by end users due to their continuous nature and often complex measurements. Introducing a fuzzification process, a method that transforms these continuous values into fuzzy linguistic equations for better representation and interpretation can be very beneficial.

Fuzzification addresses the inherent complexity of these multiple data points by simplifying their representation into intuitive and easy-to-use formats, such as linguistic equations. This transformation increases their interpretability and, therefore, facilitates explainable results. Ultimately, by wrapping this sophisticated nature of the data in a simplified interpretable wrapper, fuzzification could significantly improve the accuracy of the results retrieved from mining algorithms and make them more acceptable to end users. It promotes a better understanding of the data and can lead to more confident and informed decision-making.

We propose a fuzzification algorithm that allows automatic processing defined by the machine using the data values according to their distribution. The general process is described in Algorithm 1. The algorithm has a dataset, a Python dictionary (hash), and an integer as input. The Python dictionary serves as is mechanism for expert-determined variable intervals and corresponding labels. In contrast, the algorithm computes automatic labels based on distribution for non-user-defined variables using a default label count. Along the advanced steps of the algorithm (as visible in line 6), a global variable is employed.

The initial procedure begins in line 8 of the algorithm, which is detailed in distinct aspects. Initially, the algorithm asserts whether the variable name finds a place in the hash list $Intervals$. On finding a match within the Python dictionary, new fuzzy variables are successfully created, bearing label names as stipulated by $Intervals$ and accessing relevant settings (e.g., computation of membership degrees) corresponding to the pre-set interval, as can be seen in lines 10–18 of Algorithm 1. Conversely, if the variable cannot be traced within the dictionary, the algorithm engages its automatic function to segment variable values into a series of intervals, as defined under $DefaultIntervals$. This division aligns proportionately with the

Algorithm 1 Fuzzification function

Input: *Data:* A transaction: $t_k = \{item_1, \dots, item_m\}$
Global distributed variable: *Intervals:* Hash-list of intervals for each variable : $\{Variable_1 : [\{Intervals\}, \{Labels\}], \dots, Variable_p : [\{Intervals\}, \{Labels\}]\}$
Input: *DefaultIntervals:* number of intervals automatically generated by the algorithm
Output: Fuzzy transaction

Start Algorithm

```
Features = Dataset.NameFeatures()
GlobalVariable(Features)
i=0
do
    # Check if the variable exists in the hash list
    if Feature[i] ∈ Intervals then
        Interval=Intervals[Feature[i][0]]
        Labels=Intervals[Feature[i][1]]
    else
        Interval = GenerateIntervals(DefaultIntervals,Data[Feature[i]])
        Labels = GenerateLabels(DefaultIntervals)
    end if
    for j = 0; j < |Labels| ; j++ do
        FuzzyData[Label]=FuzzyDivision(Interval[j], Interval[j+1], type)
        # type ="linear", "exponential", "logarithmic"...
        i++
    end for
while |Feature| > i
return FuzzyData
```

variable percentiles. To ensure scalability and efficient data processing, the code is implemented with Apache Spark's mapPartition functions. This functionality enables distributed computing across multiple nodes, thereby making the algorithm capable of handling large datasets and complex computations seamlessly. By leveraging Spark's distributed processing capabilities, the system efficiently scales to meet the demands of extensive healthcare data analysis, ensuring quick and accurate fuzzy variable creation even in big data environments.

Exemplified in Fig. 5 is a scenario where the variable $DefaultIntervals = 3$. Here, the y-axis symbolizes the degree of membership, while the x-axis represents the variable's percentile. Notably, the percentiles strategically selected are 25 and 37.5, outlining the first label's trapezoidal shape and the left portion of the second label. Simultaneously, percentiles 62.5 and 75 aid in defining the right part of the second label in conjunction with the third label. This indicates applying the $GenerateIntervals$ function that tactically segments the set into k evenly distributed fuzzy sets, using their relevant percentiles. For instance, if we consider $k = 4$, the subsequent percentiles are computed in the following manner:

$$\{\frac{100}{k+1}, \frac{100}{k+1} + \frac{100}{(k+1)(k-1)}, \frac{2\cdot 100}{k+1} + \frac{100}{(k+1)(k-1)},$$
$$\frac{2\cdot 100}{k+1} + \frac{2\cdot 100}{(k+1)(k-1)}, \frac{3\cdot 100}{k+1} + \frac{2\cdot 100}{(k+1)(k-1)},$$
$$\frac{3\cdot 100}{k+1} + \frac{3\cdot 100}{(k+1)(k-1)}\}$$

which results in

$$\{p_{20}, p_{26.6}, p_{46.6}, p_{53.3}, p_{73.3}, p_{80}\}$$

On the contrary, the $FuzzyDivision$ function uses the defined intervals contained in the global variable $Intervals$.

3.3.2. Scalability testing and performance analysis

To assess the scalability of the proposed system, tests were conducted using an augmented dataset. Due to the limited availability of real data, the original dataset was duplicated several times to create a significantly larger dataset. With the purpose of measuring the efficiency of our proposal, we have analyzed the *speed up* and the *efficiency* [66] according to the number of cores. For that, we have computed the well-known measure of speed up defined as [66]

$$S_n = T_1/T_n \quad (1)$$

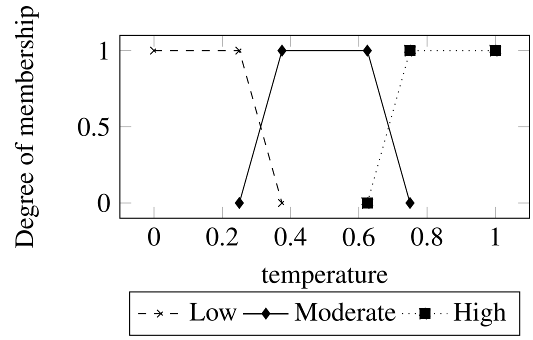

Fig. 5. Example of automatic execution with 3 default intervals.

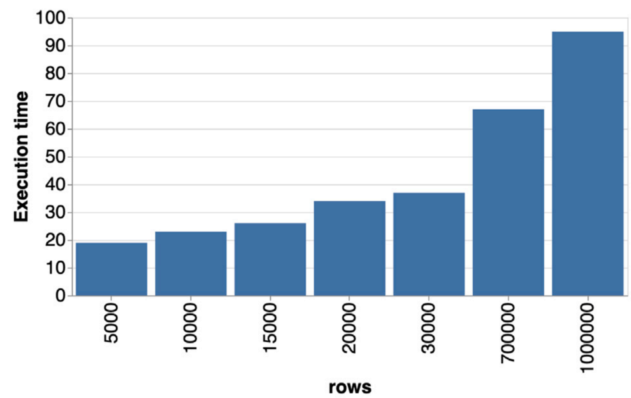

Fig. 6. Execution time of algorithm.

where T_1 is the time of the sequential algorithm and T_n is the execution time of the distributed algorithm using several cores. This approach was intended to emulate larger-scale scenarios and demonstrate the capability of the fusion process under extended conditions. The results, depicted in Fig. 11, illustrate how the algorithm scales with system resources, therefore it can process up to 100,000 records. Thus we can see how it demonstrates its ability to effectively handle increasingly complex and voluminous data entries. In addition to scalability testing, a speed-up analysis was conducted to assess the performance improvements achievable through resource optimization. Fig. 7 presents the results of this analysis, which reveal the speed-up of algorithm execution with increased computational resource allocation (between 10 and 30 percent). This exercise highlights the potential improvements in processing time, affirming that the system not only scales efficiently, but also optimizes resource usage to achieve better performance. By performing these analyzes, we demonstrate the robustness and efficiency of the system in large-scale healthcare data processing environments, reinforcing its applicability and scalability for real-world applications (see Fig. 6).

3.4. Data mining: Fuzzy association rules in big data

After data pre-processing, fuzzification and Fusion data, data mining techniques were applied to the processed data. In particular, an algorithm for association rule mining was applied in Big Data (BDFARE Apriori-TID Big Data Fuzzy Association Rules Extraction [67,68]). The algorithm was also developed using the MapReduce paradigm within the Spark Framework. This implementation allows for the efficient processing of large collections of fuzzy transactions to discover frequent itemsets and fuzzy association rules that surpass predefined support and confidence thresholds based on a given set of α-cuts. The algorithm operates in two main phases: firstly, extracting frequent items that

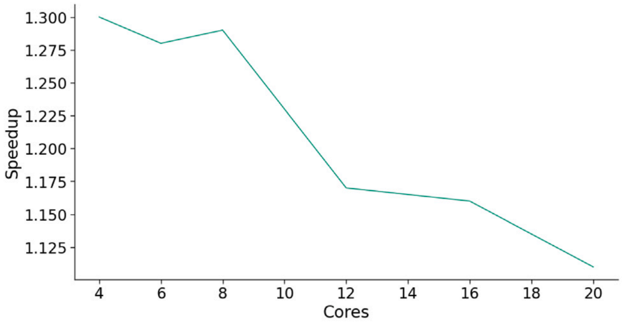

Fig. 7. Speed up of algorithm.

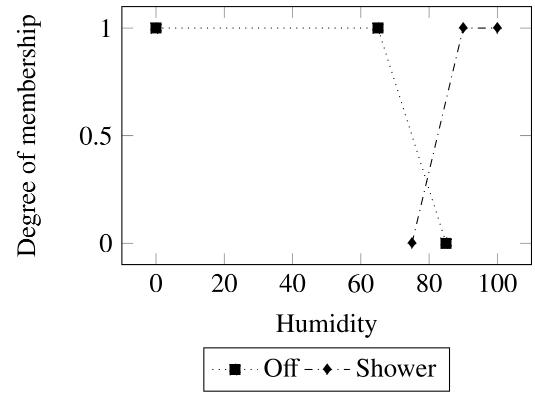

Fig. 8. Example of fuzzy-labels of the shower sensor.

satisfy a distributed form of support threshold, and secondly, extracting association rules that meet the minimum confidence threshold.

4. Use case: Smart home

In this section, a case study to validate the proposal is presented. The case study focuses on monitoring the home of an elderly person living alone. In order to accomplish this, context and devices deployed are illustrated, the fuzzification obtained and finally, the results are analyzed.

4.1. Sensor distribution

Activity recognition, especially in multi-occupant environments where several people may cohabit, requires the use of sensors in various locations where the desired activities are performed. Moreover, being a multi-occupancy environment, it is necessary to integrate an indoor location system, to be able to discern the person performing the activities.

The monitored dwelling is illustrated in Fig. 3, which provides a detailed visualization of the deployment of various devices across the dwelling.

In the kitchen, an open and close sensor is installed within the pill box the medication intake alongside with a fixed tracking device, to obtain the location.

For the bathroom, several devices are employed: a motion sensor monitors tooth brushing. In contrast, another motion sensor and an ambient sensor are placed in the shower area to track showering activity. Finally a fixed tracking device is installed in the wall.

In the bedroom, a Motion Sensor is located near the headboard to detect when the user rests, complemented by an fixed tracking device on the bedside table.

The dining area features a motion sensor to monitor meal timese a fixed tracking device and a central node that aggregates data from all sensors. Additionally, the user wears an activity wristband to monitor the location.

4.2. Data transformation and enrichment

Having all the raw data in our architecture, we will explain what the knowledge extraction process would be like. This will be done using the above data and analyzing the relationships obtained from the data set after pre-processing, cleaning and enriching the data. Our objective is to study the frequent behaviors and routines of the tenants of the houses so that this information can be used by end users such as doctors. To do this, it was necessary to carry out the steps described in Section 3.

In the course of our transformation procedure, one-minute intervals were respectively appointed for the grouping of our time variables. This approach is instrumental in capturing user interactions across the various components corresponding to household activities. Concurrently, sensor fusion was implemented by leveraging the process elucidated in Section 2. With this procedure in effect, variables are significantly reduced as proximity metrics are merged with the respective proximity, humidity, and temperature sensors.

The fusion yields dual benefits, ensuring computational efficiency by eliminating sequences of inactive sensors in the regions devoid of user presence in every transaction and augmenting these sensors with valuable user context and location information - a critical determinant for identifying commonly followed routines based on their conducted activities. This transformation method and sensor fusion not only streamline data but also enhance the understanding of user patterns within a household.

Ultimately, every variable underwent fuzzification, utilizing the method delineated in Section 2 and as demonstrated in Algorithm 1. This process fosters an improved interpretation for end-users, such as medical practitioners, who do not specialize in data mining. Creating varied linguistic labels contributes to identifying intriguing daytime patterns noteworthy for physicians, setting humidity boundaries indicative of shower usage (as shown in Fig. 8), and the like. In essence, the fuzzification step aids in translating large and complex data into more understandable and actionable insights for various end users.

Data preprocessing, particularly data fusion, plays a considerable role in minimizing the volume of data under analysis. This reduction occurs because, through these methods, our focus is narrowed solely to the sensors within the user's sphere of activity.

Any sensor data not associated with the user's primary location context is pruned. In other words, we keep sensors that the user interacts with and trim away those within inactive or irrelevant zones. This ensures that only interactionally pertinent sensor data is retained, reducing unnecessary clutter and enhancing the efficiency of our data analysis procedure.

Thus, data fusion, by integrating multiple data sources and facilitating a centralized overview of sensors working within the user's activity space, directly contributes to simplifying the analysis process and reducing computational burden. This optimization leads to enhanced processing and analytical speeds, potentially unlocking swift insights into user behavior patterns.

4.3. Results and discussion

In this section, our discussion focuses on the dual benefits of using data fusion techniques and fuzzy labels when applying association rules. In particular, all the processing discussed above has been applied, and in this section, we will check and analyze the results obtained by applying fuzzy association rules to the processed data. In particular, we want to study two features:

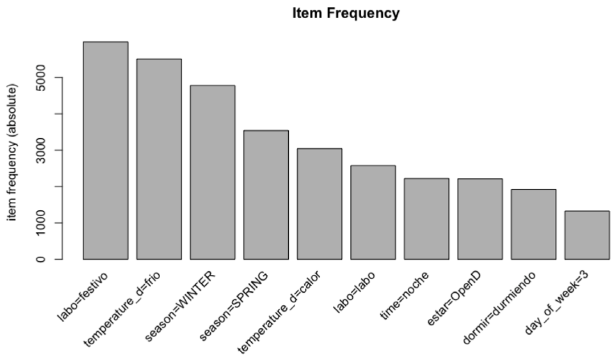

Fig. 9. Demonstration of most frequent elements with the standard fusion, illustrating a high incidence of unused or inactive sensors.

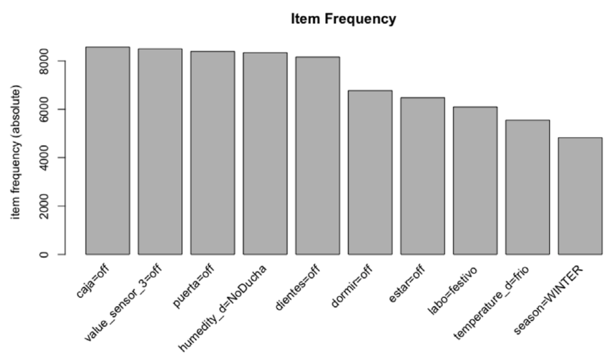

Fig. 10. Result of data fusion as presented in this work, predominantly featuring actively-used sensors, affirming the usage context person.

- The improved algorithm performance arises from the decreased data volume and results in fewer but more relevant rules for user perception.
- The association rules were revealed through our results, emphasizing interesting relationships and their interpretive improvement through fuzzy linguistic labels.

Our initial experimentation stage engaged the complete dataset, implementing standard aggregation methodologies for data fusion. Through this approach, we produced a consolidated set comprising 9000 transactions and encompassing 67 different features.

While evaluating the frequent elements with standard fusion (as depicted in Fig. 9), a preponderance of inactive or unused sensors (demonstrated as the 'off' status or where linguistic labels suggest inactivity) emerges in the dataset's frequency spectrum. This outcome can be attributed to the fusion technique where all the sensor data are considered during the rule application to the transactions, regardless of the user's presence in the respective area.

In contrast, Fig. 10 illustrates a more focused demographic, comprised mainly of sensors either 'on' or in active use. This significant distinction in the results owes itself to utilizing the data fusion method proposed in this work, which incorporates the individual's context. As a result, the items presented predominantly consist of sensors or elements actively engaged by the user, thereby leading to a more utilitarian and functional dataset. These findings prove the efficacy of our context-bearing data fusion in filtering sensor data to only reflect user-relevant, active interactions.

On the other hand, as illustrated in Fig. 11, we observed a pronounced reduction in algorithm execution time. This improvement stems from an effectively downsized dataset achieved via data fusion, substantially curtailing the required computations, resulting in a speedier process.

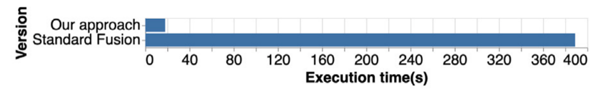

Fig. 11. Describe the execution speed of the algorithm using standard fusion and fusion with our approach.

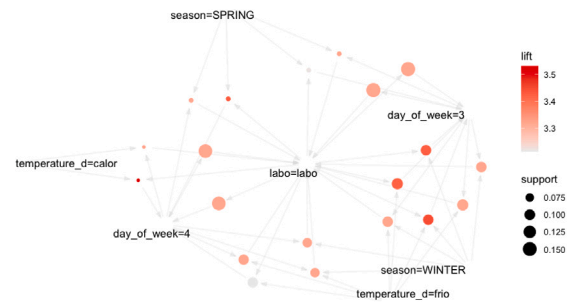

Fig. 12. Graphical representation of the rules obtained in the form of a graph.

4.3.1. Interpretation of the results

The set of rules obtained has allowed us to discover hidden patterns in the relationship between the different sensors and characteristics that appear in the data and relate them to the activities and routines carried out. This type of relationship allows us to study users' routines and behaviors in the system's data.

Having a look at the discovered patterns, we can highlight different rules. For example:

$$\{teeth = using, Medication = open\} \rightarrow \{time = earlyMorning\} \tag{2}$$

This rule describes the relationship between taking the medication (the sensor of the box that stores the pills) in the early morning and washing the diets. This demonstrates a clear routine that can be useful to end users, such as doctors knowing that it is performed frequently, indicating that the person is consistent in their healthy routines. On the other hand, some rules have been selected, where we can see different behaviors. For example, these rules were obtained:

$$\{teeth = using\} \rightarrow \{time = earlyMorning\} \tag{3}$$

and

$$\{teeth = using, humidity\_shower\} \rightarrow \{time = earlyMorning\} \tag{4}$$

These rules provide important information about the relationship between showering and brushing teeth. We can see how a clear routine is to brush your teeth in the early morning, and with the other rule, we can understand that you also shower before brushing them. In conclusion, a graphical presentation of the observed results is provided (Fig. 13). Notably, the weekdays, especially Wednesday and Thursday, emerge as the most active days. This pattern can be explained by the demographic profile of the household, consisting of older, active individuals. Moreover, some intriguing correlations are observed, such as temperature sensor readings showing patterns reflective of the changing seasons — cold in winter or heat in spring, demonstrating that the mining results reveal reliable and coherent associations (see Fig. 12).

4.3.2. Comparative results using the two fusion methods

In this section, we present a comparative analysis of the results obtained using the traditional fusion method and the proposed approach. One of the main limitations of conventional fusion techniques is the generation of a large number of association rules, many of which are produced by correlations between sensor data that do not provide meaningful insights into the user's activities. For instance, in traditional methods, frequent rules may include:

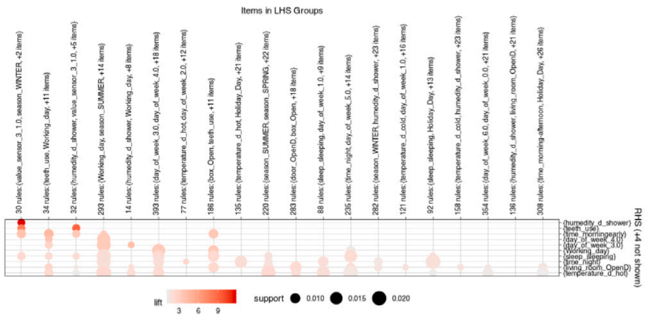

Fig. 13. Graphical representation of the rules obtained in the form of a group-table.

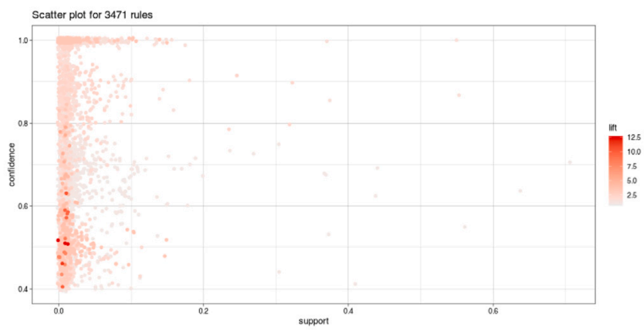

Fig. 14. Association rules generated by the traditional fusion method.

- High humidity levels (above 70%) in winter are associated with lower room temperatures.
- Temperatures exceeding a certain threshold are more frequent in summer.

While these correlations reflect environmental conditions, they do not offer actionable information about the user's behavior, routines, or interactions with their living environment. Such rules may increase the computational burden and reduce the interpretability of the model, making it challenging to derive useful insights for improving healthy routines. Our proposed methodology addresses these limitations by integrating user information, activity anchors, and linguistic variables into the fusion process. By incorporating semantic reasoning and context-awareness, the new approach significantly reduces the number of generated rules while enhancing their relevance. Some key advantages include:

- Reduction in Redundant Rules: By filtering out rules driven solely by environmental correlations, our method focuses on user-centered patterns.
- More Intuitive and Actionable Insights: Instead of generic environmental correlations, our approach generates rules that are directly linked to user behavior.
- Enhanced Contextual Relevance: The use of linguistic variables ensures that the extracted rules align with human reasoning, making them more interpretable for caregivers or automated recommendation systems.

To quantify the improvements, Figs. 14 and 15 compares the number of rules generated and their interpretability scores between the two methods. The results demonstrate that our approach not only reduces the overall rule set but also increases the proportion of contextually meaningful rules, improving decision-making in IoT-based health monitoring systems.

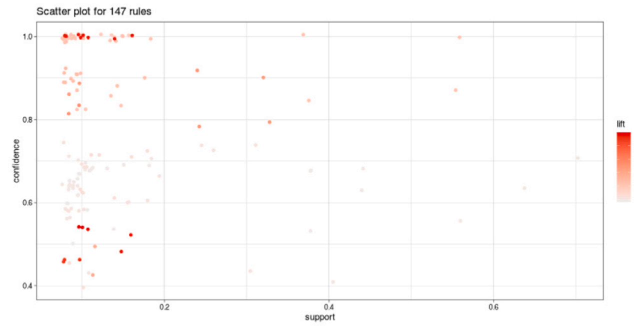

Fig. 15. Association rules generated by the fuzzy fusion method.

5. Conclusions and future work

The discovery and utilization of information from home sensors have gained significant attention over the past decade, primarily due to their economic and health-related benefits. Big Data and IoT technologies provide an ideal framework for implementing advanced analytical techniques that can effectively manage the vast amounts of data generated, particularly in healthcare systems. Additionally, the use of fuzzy logic enhances the interpretability of the collected data, offering more meaningful insights to end users.

This study aims to extract hidden knowledge from the sensorization of homes and their inhabitants, focusing on analyzing and interpreting this data. We have developed and implemented a data mining system, which was applied to a real-world dataset collected from a home in Cabra, Spain. In particular, we enriched certain features and applied a fuzzification algorithm to improve the performance of data mining techniques such as association rule mining. The results of this implementation highlight several key improvements:

- **Performance Improvement:** The application of data fusion techniques significantly reduced the overall dataset size by focusing on sensors within the user's activity domain. This optimization led to a reduction in algorithmic processing requirements, thereby improving efficiency, conserving computational resources, and reducing processing time. The introduction of fuzzified labels further minimized interpretive challenges, yielding more concise and focused rules that directly improved the end-user experience by providing actionable insights.
- **Improved interpretation of association rules:** The integration of fuzzy linguistic labels enhanced the interpretability of association rules, uncovering meaningful and informative connections within the data. This approach facilitated the generation of more comprehensible summaries of complex data patterns. The contextualization provided by fuzzy logic allowed end users to better understand and utilize these rules, transforming complex relationships into practical, user-friendly insights.
- **Scalability and generalization of the system**: Another key aspect of the system is its ability to adapt to different environments and populations. Although it has been implemented in a home environment in this study, its applicability can be extended to other healthcare scenarios, such as hospitals and nursing homes. Furthermore, experiments with extended datasets have shown that the system maintains optimal performance even with a larger volume of transactions and sensors. In terms of generalization, although this work has focused on discovering patterns in the older persons, the approach can be applied to other demographic groups or to people with chronic conditions such as cardiovascular disease or cognitive disorders. This provides valuable information about activity patterns and adherence to therapeutic routines, facilitating its use in different clinical contexts.

Our experimentation with real-world data demonstrates the viability of our proposed system. The system successfully identified valuable patterns, such as routines related to medication adherence and healthy hygiene habits (e.g., tooth brushing or showering). These insights can help end users predict and prevent health issues related to poor routine management and behavioral patterns.

This research serves as a foundation for further studies and opens new avenues for future work. The next step involves scaling the system to include a larger number of sensors and integrating Big Data tools to extract a wider array of association rules. This advancement would significantly enhance the knowledge extraction process. Additionally, future efforts should focus on improving the visualization of results, making them more informative and accessible for end users.

Furthermore, it is essential to reinforce the system's security mechanisms, given the sensitivity of the data handled, through approaches such as federated learning, which allows distributed analysis without compromising privacy. In this context, fuzzy logic is an exceptionally suitable approach, as it enables the detection of patterns in low-computing devices, maintaining processing efficiency and ensuring the protection of information.

Finally, it is imperative to conduct a qualitative and quantitative study in future research to evaluate healthcare professionals' experiences of utilizing the system and its impact on patients' adherence to treatment routines. A validation phase will be conducted in clinical and home environments, involving doctors, nurses and caregivers. These professionals will provide feedback on the system's usefulness, the clarity and interpretability of the generated information, and its ease of use in daily practice. Furthermore, the monitoring of key therapeutic adherence indicators will facilitate the determination of whether the identification of behavior patterns contributes to the detection of possible deviations in patients' routines, thereby enabling early interventions that improve treatment compliance.

CRediT authorship contribution statement

Carlos Fernandez-Basso: Writing – original draft, Investigation, Formal analysis, Conceptualization. **David Díaz-Jimenez:** Writing – original draft, Methodology, Investigation. **Jose L. López:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Macarena Espinilla:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Carlos Fernandez Basso reports financial support was provided by University of Jaen. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This result has been partially supported by grant PID2021-123960OB-I00, PDC2023-145863-I00 and grant PID2021-127275OB-I00 funded by MICIU/AEI, Spain/10.13039/501100011033 and by "ERDF A way of making Europe", grant PDC2023-145863-I00 funded by MICIU/AEI, Spain/ 10.13039/501100011033 and by "European Union NextGenerationEU/PRTR", and grant M.2 PDC_000756 funded by Consejería de Universidad, Investigación e Innovación, Spain and by ERDF Andalusia Program 2021–2027. Finally, the research reported in this paper is also funded by the European Union (BAG-INTEL project, grant agreement no. 101121309). Funding for open access charge: Universidad de Granada / CBUA

Data availability

Data will be made available on request.

References

[1] Alfredo J Perez, Farhan Siddiqui, Sherali Zeadally, Derek Lane, A review of IoT systems to enable independence for the elderly and disabled individuals, Internet Things 21 (2023) 100653.

[2] J. M. Quero, Claire Orr, Shuai Zhang, C. Nugent, A. Salguero, M. Espinilla, Real-time recognition of interleaved activities based on ensemble classifier of long short-term memory with fuzzy temporal windows, in: International Conference on Ubiquitous Computing and Ambient Intelligence, 2018.

[3] David Díaz-Jiménez, José L. López, Jesús González-Lama, Macarena Espinilla, A new horizon in healthcare: An innovative methodology for sensor-based adherence platforms in home monitoring of key treatment indicators, IEEE Internet Things J. (2024) 1, http://dx.doi.org/10.1109/jiot.2024.3424905.

[4] Majdi Rawashdeh, Mohammed GH Al Zamil, Samer Samarah, M Shamim Hossain, Ghulam Muhammad, A knowledge-driven approach for activity recognition in smart homes based on activity profiling, Future Gener. Comput. Syst. 107 (2020) 924–941.

[5] Anindita Saha, Sajan Rajak, Jayita Saha, Chandreyee Chowdhury, A survey of machine learning and meta-heuristics approaches for sensor-based human activity recognition systems, J. Ambient. Intell. Humaniz. Comput. 15 (1) (2024) 29–56.

[6] Ashwin Geet D'Sa, B.G. Prasad, An IoT based framework for activity recognition using deep learning technique, 2019, arXiv preprint arXiv:1906.07247.

[7] Muhammad Ali Jamshed, Kamran Ali, Qammer H. Abbasi, Muhammad Ali Imran, Masood Ur-Rehman, Challenges, applications, and future of wireless sensors in Internet of Things: A review, IEEE Sens. J. 22 (6) (2022) 5482–5494, http://dx.doi.org/10.1109/jsen.2022.3148128.

[8] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, Lisha Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognit. Lett. 119 (2019) 3–11, http://dx.doi.org/10.1016/j.patrec.2018.02.010.

[9] Alfredo J. Perez, Sherali Zeadally, Recent advances in wearable sensing technologies, Sensors 21 (20) (2021) 6828, http://dx.doi.org/10.3390/s21206828.

[10] M. Sabzevari, E. Imani, Separation of movement direction concepts based on independent component analysis algorithm, linear discriminant analysis, deep belief network, artificial and fuzzy neural networks, Biomed. Signal Process. Control. 62 (2020).

[11] A. Ponmalar, V. Dhanakoti, An intrusion detection approach using ensemble Support Vector Machine based Chaos Game Optimization algorithm in big data platform, Appl. Soft Comput. 116 (2022).

[12] K. Gayathri, K.S. Easwarakumar, Susan Elias, Fuzzy ontology based activity recognition for assistive health care using smart home, Int. J. Intell. Inf. Technol. (2020).

[13] Shirin Kordnoori, Arash Sharifi, Hamed Shah-Hosseini, Human fall detection using neuro-fuzzy models based on ensemble learning, Prog. Artif. Intell. 11 (3) (2022) 219–232.

[14] Juan F Gaitán-Guerrero, José L López, Macarena Espinilla, Carmen Martínez-Cruz, Towards an interpretable continuous glucose monitoring data modeling, IEEE Internet Things J. (2024).

[15] Xieling Chen, Haoran Xie, Zongxi Li, Gary Cheng, Mingming Leng, Fu Lee Wang, Information fusion and artificial intelligence for smart healthcare: a bibliometric study, Inf. Process. Manag. 60 (1) (2023) 103113, http://dx.doi.org/10.1016/j.ipm.2022.103113.

[16] Olusegun Agbabiaka, Adegboyega Ojo, Niall Connolly, Requirements for trustworthy AI-enabled automated decision-making in the public sector: A systematic review, Vol. 215, 2025, pp. 124076 http://dx.doi.org/10.1016/j.techfore.2025.124076.

[17] Abdulkareem Y. Abdalla, Turki Y. Abdalla, Adala M. Chyaid, Internet of Things-based fuzzy systems for medical applications: A review, IEEE Access 12 (2024) 163883–163902, http://dx.doi.org/10.1109/access.2024.3487812.

[18] Vipul Bansal, Ankit Bhardwaj, Jitendra Singh, Devvret Verma, Mohit Tiwari, Someshwar Siddi, Using artificial intelligence to integrate machine learning, fuzzy logic, and The IOT as A Cybersecurity System, in: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, 2023, pp. 762–769, http://dx.doi.org/10.1109/icacite57410.2023.10182967.

[19] Srishti Vashishtha, Vedika Gupta, Mamta Mittal, Sentiment analysis using fuzzy logic: A comprehensive literature review, WIREs Data Min. Knowl. Discov. 13 (5) (2023) http://dx.doi.org/10.1002/widm.1509.

[20] Jie Lu, Guangzhi Ma, Guangquan Zhang, Fuzzy machine learning: A comprehensive framework and systematic review, IEEE Trans. Fuzzy Syst. 32 (7) (2024) 3861–3878.

[21] Carlos Fernandez-Basso, Karel Gutiérrez-Batista, Roberto Morcillo-Jiménez, Maria-Amparo Vila, Maria J Martin-Bautista, A fuzzy-based medical system for pattern mining in a distributed environment: Application to diagnostic and co-morbidity, Appl. Soft Comput. 122 (2022) 108870.

[22] Pau Climent-Pérez, Francisco Flórez-Revuelta, Improved action recognition with separable spatio-temporal attention using Alternative Skeletal and Video Pre-Processing, Sensors (Basel, Switzerland) 21 (2021) http://dx.doi.org/10.3390/s21031005.

[23] Solaiman Ahmed, Taiki Kishi, M. Nii, K. Higuchi, S. Kobashi, Estimation of human activities by fuzzy classification systems for understanding subject Persons' Situation, Int. Conf. Mach. Learn. Comput. (2019).

[24] Gadelhag Mohmed, Ahmad Lotfi, Amir Pourabdollah, Long short-term memory fuzzy finite state machine for human activity modelling, Petra (2019).

[25] Gadelhag Mohmed, Ahmad Lotfi, Caroline S. Langensiepen, Amir Pourabdollah, Clustering-based fuzzy finite state machine for human activity recognition, UKCI (2018).

[26] Shu-Yin Chiang, Yao-Chiang Kan, Yun-Shan Chen, Ying-Ching Tu, Hsueh-Chun Lin, Fuzzy computing model of activity recognition on WSN movement data for Ubiquitous Healthcare Measurement, Ital. Natl. Conf. Sens. (2016).

[27] Gadelhag Mohmed, Ahmad Lotfi, Amir Pourabdollah, Human activities recognition based on Neuro-Fuzzy Finite State Machine, Technologies (2018).

[28] Nehal A. Sakr, Mervat Abu-Elkheir, A. Atwan, Hassan H. Soliman, A multilabel classification approach for complex human activities using a combination of emerging patterns and fuzzy sets, Int. J. Electr. Comput. Eng. (IJECE) (2019).

[29] Van Nguyen, Liam Mellor, Fuzzy MLNs and QSTAGs for activity recognition and modelling with rush, in: 2020 IEEE 23rd International Conference on Information Fusion, FUSION, IEEE, 2020, pp. 1–8.

[30] Federico Castanedo, A review of data fusion techniques, in: Y. Takama, D. Ursino (Eds.), Sci. World J. 2013 (1) (2013) http://dx.doi.org/10.1155/2013/704504.

[31] Mazhar Hussain, Mattias O'Nils, Jan Lundgren, Seyed Jalaleddin Mousavirad, A comprehensive review on deep learning-based data fusion, IEEE Access 12 (2024) 180093–180124, http://dx.doi.org/10.1109/access.2024.3508271.

[32] Huaping He, Fuyuan Xiao, A novel quantum Dempster's rule of combination for pattern classification, Inf. Sci. 671 (2024) 120617, http://dx.doi.org/10.1016/j.ins.2024.120617.

[33] Fuyuan Xiao, Junhao Wen, Witold Pedrycz, Masayoshi Aritsugi, Complex evidence theory for multisource data fusion, Chin. J. Inf. Fusion 1 (2) (2024) 134–159, http://dx.doi.org/10.62762/cjif.2024.999646.

[34] Carlos Fernandez-Basso, M Dolores Ruiz, Maria J Martin-Bautista, A fuzzy mining approach for energy efficiency in a Big Data framework, IEEE Trans. Fuzzy Syst. 28 (11) (2020) 2747–2758.

[35] Miguel Molina-Solana, María Ros, M Dolores Ruiz, Juan Gómez-Romero, María J Martín-Bautista, Data science for building energy management: A review, Renew. Sustain. Energy Rev. 70 (2017) 598–609.

[36] Cheng Fan, Fu Xiao, Mining gradual patterns in big building operational data for building energy efficiency enhancement, Energy Procedia 143 (2017) 119–124.

[37] J.J. Davis, A.J. Clark, Data preprocessing for anomaly based network intrusion detection: A review, Comput. Secur. 30 (6–7) (2011) 353–375.

[38] A. Chandra Pandey, D. Singh Rajpoot, M. Saraswat, Twitter sentiment analysis using hybrid cuckoo search method, Inf. Process. Manag. 53 (4) (2017) 764–779.

[39] J. Jeon, C. Lee, Y. Park, How to use patent information to search potential technology partners in open innovation, J. Intellect. Prop. Rights 16 (5) (2011) 385–393.

[40] H.A. Amirkolaee, H. Arefi, Height estimation from single aerial images using a deep convolutional encoder-decoder network, ISPRS J. Photogramm. Remote Sens. 149 (2019) 50–66.

[41] S. Zhang, Q. Shen, C. Nie, Y. Huang, J. Wang, Q. Hu, X. Ding, Y. Zhou, Y. Chen, Hyperspectral inversion of heavy metal content in reclaimed soil from a mining wasteland based on different spectral transformation and modeling methods, Spectrochim. Acta - A: Mol. Biomol. Spectrosc. 211 (2019) 393–400.

[42] L. Sun, X. Zhang, J. Xu, S. Zhang, An attribute reduction method using neighborhood entropy measures in neighborhood rough sets, Entropy 21 (2) (2019).

[43] L. Zhang, H. Sun, Z. Rao, H. Ji, Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds, Spectrochim. Acta - A: Mol. Biomol. Spectrosc. 229 (2020).

[44] A.L. Beam, I.S. Kohane, Big data and machine learning in health care, JAMA - J. Am. Med. Assoc. 319 (13) (2018) 1317–1318.

[45] C.H. Lee, H.-J. Yoon, Medical big data: Promise and challenges, Kidney Res. Clin. Pr. 36 (1) (2017) 3–11.

[46] M.A. Khan, I.U. Lali, A. Rehman, M. Ishaq, M. Sharif, T. Saba, S. Zahoor, T. Akram, Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection, Microsc. Res. Tech. 82 (6) (2019) 909–922.

[47] S. Vajda, A. Karargyris, S. Jaeger, K.C. Santosh, S. Candemir, Z. Xue, S. Antani, G. Thoma, Feature selection for automatic tuberculosis screening in frontal chest radiographs, J. Med. Syst. 42 (8) (2018).

[48] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. J. 74 (2019) 255–263.

[49] J. Chen, B. Hu, P. Moore, X. Zhang, X. Ma, Electroencephalogram-based emotion assessment system using ontology and data mining techniques, Appl. Soft Comput. 30 (2015) 663–674.

[50] Naseem Rao, Anil Kumar, T. Abbasi, Mobile health monitoring system using fuzzy logic, Bioinformatics (2016).

[51] A. Nikfarjam, A. Sarker, k. O'Connor, R. Ginn, G. Gonzalez, Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features, J. Am. Med. Inform. Assoc. 22 (3) (2015) 671–681.

[52] R. Bauder, T.M. Khoshgoftaar, N. Seliya, A survey on the state of healthcare upcoding fraud analysis and detection, Heal. Serv. Outcomes Res. Methodol. 17 (1) (2017) 31–55.

[53] L.F. Silva, A.A.S.M.D. Santos, R.S. Bravo, A.C. Silva, D.C. Muchaluat-Saade, A. Conci, Hybrid analysis for indicating patients with breast cancer using temperature time series, Comput. Methods Programs Biomed. 130 (2016) 142–153.

[54] S. Chakraborty, K. Mali, Fuzzy electromagnetism optimization (FEMO) and its application in biomedical image segmentation, Appl. Soft Comput. J. 97 (2020).

[55] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, G. Wang, Data processing and text mining technologies on electronic medical records: A review, J. Heal. Eng. 2018 (2018).

[56] F. Gargiulo, S. Silvestri, M. Ciampi, G. De Pietro, Deep neural network for hierarchical extreme multi-label text classification, Appl. Soft Comput. J. 79 (2019) 125–138.

[57] R. Catelli, F. Gargiulo, V. Casola, G. De Pietro, H. Fujita, M. Esposito, Crosslingual named entity recognition for clinical de-identification applied to a COVID-19 Italian data set, Appl. Soft Comput. J. 97 (2020).

[58] W. Ai, K. Li, K. Li, An effective hot topic detection method for microblog on spark, Appl. Soft Comput. J. 70 (2018) 1010–1023.

[59] S. Malla, A. P.J.A., COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets, Appl. Soft Comput. 107 (2021).

[60] F. Goodarzian, H. Hosseini-Nasab, J. Muñuzuri, M.-B. Fakhrzad, A multi-objective pharmaceutical supply chain network based on a robust fuzzy model: A comparison of meta-heuristics, Appl. Soft Comput. J. 92 (2020).

[61] B. Malysiak-Mrozek, M. Stabla, D. Mrozek, Soft and declarative fishing of information in big data lake, IEEE Trans. Fuzzy Syst. 26 (5) (2018) 2732–2747.

[62] Subinoy Adhikari, Ahona Ghosh, Prasenjit Das, Sandip Roy, Rajesh Bose, An image based activity recognition, Int. J. 8 (5) (2020).

[63] Zameer Gulzar, A. Anny Leema, I. Malaserene, Human activity analysis using machine learning classification techniques, Int. J. Innov. Technol. Explor. Eng.(IJITEE) (2019).

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[65] Victoria Hodge, Jim Austin, A survey of outlier detection methodologies, Artif. Intell. Rev. 22 (2) (2004) 85–126.

[66] Cristóbal Barba-González, José García-Nieto, Antonio Benítez-Hidalgo, Antonio J. Nebro, José F. Aldana-Montes, Scalable inference of gene regulatory networks with the spark distributed computing platform, in: Javier Del Ser, Eneko Osaba, Miren Nekane Bilbao, Javier J. Sanchez-Medina, Massimo Vecchio, Xin-She Yang (Eds.), Intelligent Distributed Computing XII, Springer International Publishing, Cham, 2018, pp. 61–70.

[67] Carlos Fernandez-Bassso, M Dolores Ruiz, Maria J Martin-Bautista, Fuzzy association rules mining using spark, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2018, pp. 15–25.

[68] Carlos Fernandez-Basso, M Dolores Ruiz, Maria J Martin-Bautista, Spark solutions for discovering fuzzy association rules in big data, Internat. J. Approx. Reason. 137 (2021) 94–112.