Novelty Detection Algorithms to Help Identify Abnormal Activities in Elderly´s Daily Lives

A. M. R. Fernandes, V. R. Q. Leithardt, *Senior Member, IEEE,* and J. F. P. Santana

***Abstract*—In old age, several common health conditions, chronic illnesses, and disabilities affect the individual's physical and mental health and prevent him from carrying out Activities of Daily Living. In this context, this article presents a comparative study between some Machine Learning algorithms used to identify behavioral abnormalities based on ADL (Activities of Daily Living), through the Novelty Detection technique. ADL data from eHealth Monitoring Open Data Project database were used to create a model that defines the baseline behavior of an elderly person, and new observations, to verify significant changes in behavior, are classified as outliers or abnormal. The Local Outlier Factor, One-class Support Vector Machine, Robust Covariance, and Isolation Forest algorithms were analyzed, and the Local Outlier Factor obtained the best result, reaching a precision and F1-Score of 96%. As elderly people can have completely different routines, the data from the dataset used are not generalizable, but specific to everyone. In this work, the issue of model retraining is not evaluated, however, a variation is recommended in the period of weeks necessary for model retraining. Despite the good performance obtained, it is necessary to consider reproducing the experiments with data from other databases, to improve the generalization of the proposed solution, as well as to carry out a more refined validation. It is also necessary to carry out experiments to evaluate whether the variation in the types of activities carried out throughout a day by an elderly person, as well as the inclusion of new activities in the elderly person's routine, can impact the performance of the proposed model.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8373

***Index Terms*— Activities of Daily Living, Machine Learning Algorithms, Novelty Detection.**

I. Introduction

According to the World Health Organization (WHO), the world's population is aging at a rapid pace. It is estimated that from 2015 to 2050 the number of people over 60 will almost double, from 12% to 22% of the population, which represents 2.1 billion people [1]. In addition, the number of seniors over 80 is expected to triple to 526 million.

This situation poses a challenge to global society in terms of providing healthy and inclusive aging. In this sense, one of the biggest problems refers to how to provide solutions that can help in the monitoring of the elderly in chronic diseases that are quite common at this stage of life and affect around 80% of the elderly population [2, 3].

Chronic illnesses include heart disease, depression, Alzheimer's disease, and dementia. Many degenerative diseases prevent older people from being able to independently perform Activities of Daily Living (ADL) [4]. It is estimated that more than 56% of the elderly have some type of moderate to severe disability, around 256 million people [5]. In the domestic context, the difficulty of performing such tasks can be minimized through Ambient Assisted Living (AAL) technologies. These technologies aim to develop personal health and remote monitoring systems [6].

In this scenario, it is essential to monitor the ADL to identify abnormalities in the performance of these tasks. This is because significant changes in the normal routine of the elderly may represent declines in physical or mental health. This detection allows the elderly support network to be aware of these changes, and to identify any intervention needs [7]. The detection of ADL to detect anomalies and consequently provide better care for the elderly, has been the focus of research involving Wireless Sensor Networks, IoT and Machine Learning algorithms [8, 9, 10].

Recognition of ADLs is done by Human Activity Recognition (HAR) techniques [6]. The recognition of these activities for a certain period defines the baseline behavior of the individual, and from this pattern, abnormalities can be detected [11]. This pattern can be established using Machine Learning algorithms through Novelty Detection, ND. The use of this technique is related to the detection of rare events and anomalies. It is used when there are training data that characterize the target class well, however, data from other classes are scarce or not well characterized [12, 13, 14]. Within this context, this article presents a comparative study between the Local Outlier Factor (LOF), One-Class Support Vector Machine (OC-SVM), Robust Covariance (RCE), and Isolation Forest (iFortest) algorithms, to verify which one presents the best performance to create a baseline of elderly behavior, as well as establish new observations to verify significant variations that are classified as outliers or anomalous. In addition to presenting the algorithm with the best performance, this article also presents the anomaly detection model created from this algorithm. This model consists of two steps. The first is responsible for identifying daily anomalies, and the second is only executed when an anomaly is detected and aims to identify the activity or activities that caused the anomaly.

II. Selected Algorithms

To select the algorithms to be analyzed, a bibliographic

A. M. R. Fernandes is with University of Vale do Itajaí, Itajaí, SC 88302-901, Brazil (e-mail: anita.fernandes@Univali.br).

V. R. Q. Leithardt is with the Instituto Superior de Engenharia de Lisboa (ISEL), Lisboa, Portugal (e-mail: valderi@ieee.org).

J. F. P. Santana is with Expert Systems and Applications Laboratory (ESALAB), Faculty of Science, University of Salamanca, Salamanca, 37008, Spain (e-mail: fcofds@usal.es).

survey was carried out to verify the most cited ones for detecting novelties. Systematic reviews were analyzed on the application of Novelty Detection algorithms in different contexts, such as Business Financial Systems [15], IoT Environment [16], Prevention and Fighting against web attacks [17], and IoMT (Internet of Medical Things) [18]. In addition to these works, research correlated to this work was analyzed. Koutli et al [19], present a solution to detect abnormal behavior in the space-time context, combining classification and regression algorithms. For the classification task, the outlier detection algorithms RCE, OC-SVM, iForest, LOF, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) were evaluated. However, the LOF algorithm showed superior results. For the regression task, the algorithms Seasonal Auto-Regressive Integrated Moving Average (SARIMA), Random Forest, and a neural network Long Short-Term Memory Network were evaluated. The combined approach of LOF (classification) and SARIMA (regression) showed better results than those used individually, therefore, they were selected to compose the combined analysis.

Yahaya, Lofti, and Mahmud [20] present an approach based on an ensemble (set) of novelty detection algorithms named Consensus Novelty Detection Ensemble (CNDE) to detect abnormalities in Activities of Daily Living. The CNDE is based on the concepts of external and internal consensus. Internal consensus is defined by internal voting on each model and its child models (its variations). The external consensus is defined by the external voting of each model in the ensemble. The weight of each model is defined based on its performance and internal score, called the Normality Score. Data are classified as normal if they are within the normality score and classified as abnormal if they exceed the defined normality limit. A solution for hazard detection using common Wi-Fi devices and the OC-SVM outlier detection algorithm for monitoring activity in high-privacy environments such as bathrooms are presented by Zhang et al [21]. Hazardous conditions are characterized by data samples that deviate from the pattern learned by the model. These resources are used for training the OC-SVM, which is used to detect dangerous conditions for the elderly in the bathroom. Zekri et al. [22], present a framework to analyze the evolution of the behavior of the elderly for long periods, to detect anomalies in the performance of the ADL. Based on historical data, DBSCAN is used to model the normal behavior of the elderly, grouping each activity based on its start time and duration. Activities are evaluated individually according to statistical values extracted from clusters. Howedi, Lotfi e Pourabdollah [23] present the detection of anomalies in activities of daily living, based on entropy measures. For this, the OC-SVM algorithm was studied.

Table I presents a comparison between the works cited and the model proposed in this work, considering the following characteristics: algorithms and techniques used, dataset used, and types of sensors to be used to collect data.

Considering Table I, it appears that, of the five works analyzed, four used the OC – SVM algorithm. The LOF, RCE and iForest algorithms were used in two research each. And the DBSCAN algorithm was used in only one work. Therefore, in this work we chose to use the 4 most cited algorithms.

TABLE I
COMPARTIVE ANALYSIS

Characteristics	Paper 1 [19]	Paper 2 [20]	Paper 3 [21]	Paper 4 [22]	Paper 5 [23]	Proposed Model
Algorithms and Techniques						
iForest	x	x				x
OC-SVM	x	x	x		x	x
LOF	x	x				x
RCE	x	x				x
DBSCAN	x					
Type of Sensors						
Environmental	x	x	x		x	x
Medical Equipment Bluetooth	x			x		
Dataset						
Created a dataset			x	x	x	
Activities of Daily Living Recognition Using Binary Sensors Version 1.0		x				
CASAS H111		x				
eHealth Monitoring Open Data Project						x

Regarding the type of sensors used, 4 works used environmental sensors, and only one used Medical Equipment Bluetooth. Although this work does not include data collection in its scope, the data analyzed refer to environmental variables. As for the databases used, they did not specifically deal with elderly people, so it was decided to research a database whose data focused on the target audience of the work. In this way, the e-Health Monitoring open data Project base was chosen, which explains the daily lives of dependent person (such as an elderly person).

From the analysis of these works, it was verified that the LOF, OC-SVM, RCE, and Isolation Forest algorithms were considered the most promising for the type of problem that this work addresses, mainly due to the size of the database.

LOF [24] is an unsupervised approach for anomaly detection, based on the concept of local density [25]. Density is estimated by looking at the k nearest neighbors. This is done by comparing the distance of an object with the local densities of its neighbors. Objects that have substantially lower densities than their neighbors are considered outliers. OC-SVM works with the concept of decision boundary between classes, through the creation of hyperplanes and associated margins. However, in the problem of detecting anomalies or outliers, there are no labeled data, OC-SVM considers that all data belong to the normal class. In this way, the model learns the inherent pattern of the data and considers data samples that do not belong to this pattern as anomalies or outliers. This algorithm is known to be sensitive to outliers. Due to the use of the margin to avoid overfitting, this model can be considered a variant of linear and logistic regression models, as they use regularization for the same objective [26].

iForest is based on the idea that, when defining a set of data, it is possible, from the most central data of the cluster, to organize a binary tree based on the proximity between the clusters. Each node receives a score, quantifying its proximity to the center of the cluster. Lower scores represent behaviors more common to the data set, and the cluster's position in the tree will be towards the leaf nodes. The anomalies will have higher scores and the tendency will be to be closer to the root node of the tree [27]. RCE is a distance-based algorithm that uses geometric measurements to define the proximity between data samples in multidimensional space. Points further away than most are considered anomalous. The most used measure is the Mahalanobis Distance, used in multivariate spaces. What the algorithm uses to detect anomalies is how far from the covariance value of the variables is about the mean of the set, thus, values much above the mean are considered anomalous [28].

III. Methodology

For the detection of anomalies in ADL, historical data from the performance of these activities are used for training ND models. These models learn the normal behavior of the elderly in carrying out these activities, and based on this, evaluate whether new daily data present anomalies about the learned behavior. Data are grouped and summarized by day, extracting contextual and temporal attributes. In this work, the contextual ones boil down to the day of the week (from Monday to Sunday, respectively from 0 to 6), and the information if the day is a working day or part of the weekend (0 working days and 1 for the end of week). This information is used to distinguish routines on specific days or on weekends. Temporal attributes boil down to the frequency with which the activity was performed during the day (sum of the number of times the activity is performed), the duration of the activity during the day (sum of activity duration times), and the reason between duration and frequency. The use of these temporal attributes was inspired by the work of Bozdog et al [29].

The detection process works in two stages, the first is responsible for identifying daily anomalies. According to Fig. 1, **S** weeks are used for training the **M** model, and the data of the new day are submitted to the model for classification as normal or abnormal. Regarding the training weeks, the minimum number for an accuracy greater than 90% is evaluated, training and evaluating the models using 2 to 20 weeks of data. Generalization of the temporal attributes´ duration, frequency, and duration/frequency is also evaluated. For model M, LOF, iForest, OC-SVM, and RCE will be evaluated.

The final model configuration must include a minimum number of weeks of training data to reach accuracy and F1-Score greater than 90%, for both steps of the anomaly detection process; the first stage, will use predictive attributes such as the contextual and temporal attribute of greater generalization; each step will use the model that performs best. Among the databases identified in the literature [30, 31, 32, 33], the eHealth Monitoring Open Data Project database [30] was selected. This database consists of one year of simulated data, referring to the performance of ADL by a dependent person (such as an elderly person), on SMAF (Functional Autonomy Measurement System) [34] dependency profiles. This scale ranges from P1 (autonomous individual with some level of supervision and help needed) to P14 (completely dependent).

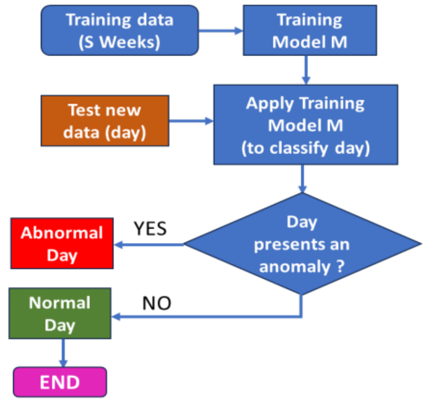


Fig. 1. Stage 1: Normal/Abnormal day classification.

This database was selected for the following reasons: (i) observations are presented at the level of activities and not sensors; (ii) presents a long period of labeled observations (one year); (iii) guarantees the non-existence of outliers within the same dependency profile; (iv) well-founded data simulation process [35]; (vi) no other dataset with the same characteristics was found. The data from this database are organized into two main sets, the first consisting of one-year observations on the P1 profile. The other consists of a year of observations, with changes in the dependence profile every 3 months, respectively in profiles P1, P3, P6, and P8/P9. For this work, the dataset named monitoringP1HRversion.txt was used. Fig. 3 presents a clipping of this dataset.

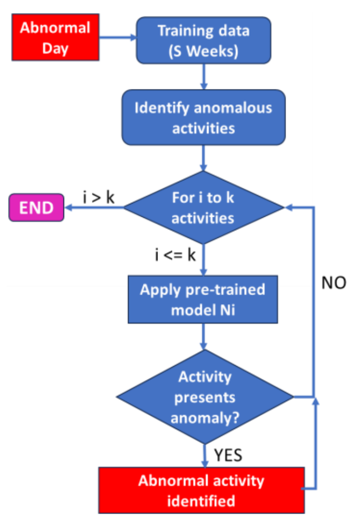


Fig. 2. Stage 2: Identification of anomalous activities on abnormal days.

Among the 21 activities observed in the selected dataset, 6 activities were used, selected arbitrarily, with a view to facilitating experiments and analysis. The selected activities were: Changing clothes, Eating, Making coffee, Making hot food, Toileting, and Washing hands/face.

The pre-processing of the monitorinP1HRversion dataset covered 5 steps. The first step consists of (i) creating a CSV file based on existing columns in the dataset; (ii) the columns day and name of the activity maintained; (iii) the column "duration" was added, which is the difference in seconds between the column end time and start time of the activity. The partial result of this step can be seen in Table II.

```
************************************************************
*   One-Year Scenario of Activities - Elderly in the SMAF Profile: P1
*   Human readable version - 28/03/2015
*   Format:
*   <starting time> <end time> <action/activity>
*
*   eHealth Monitoring Open Data Project
*   https://sourceforge.net/projects/ehealthmonitoringproject/
*   Members:
*       Tayeb Lemlouma (Tayeb.Lemlouma@irisa.fr)
*       Haider Mshali (hayder-hasan.mshali@u-bordeaux.fr)
*       Damien Magoni (magoni@labri.fr)
************************************************************
01 day - 08:03:32 01 day - 08:22:40 Washing (take shower)
01 day - 08:23:46 01 day - 08:26:53 Hair dray
01 day - 08:28:50 01 day - 08:38:39 Change clothes
01 day - 08:40:37 01 day - 08:50:24 Toileting
01 day - 08:52:12 01 day - 08:55:38 Washing hand/face
01 day - 08:57:36 01 day - 09:05:53 Make coffee
01 day - 09:07:38 01 day - 09:12:52 Washing hand/face
01 day - 09:13:57 01 day - 09:21:10 Make sandwich
```

Fig. 3. Dataset monitoringP1HRversion presenting some examples of activities per hour [30].

TABLE II
PRE-PROCESSING STEP 1

	Day	Duration	Activity
0	1	1148.0	tshower
1	1	187.0	hairdray
2	1	589.0	cclothes
3	1	587.0	toileting
4	1	206.0	whandface

The second step of pre-processing consists of (i) grouping the activities, counting the number of occurrences, originating the "frequency" column, as well as the sum of the activity durations, creating the "duration" column; (ii) creation of the "duration/Frequency" column, which is the ratio between the duration and frequency attributes; (iii) creation of the day_of_week column representing the days of the week from Monday to Sunday, with the respective values from 0 to 6; (iv) creation of the weekend column, with the values 0 and 1, 0 representing weekdays and 1 weekend. The partial result of this step is shown in Table III. Step 3 of the pre-processing consists of creating 3 datasets, each pivoting the duration, frequency, and duration/frequency attributes, using the activities as columns, and keeping the attributes day, day_of_week, and weekend.

TABLE III
PRE-PROCESSING STEP 2

	Day	Day of the Week	Week-end	Activi-ty	Dura-tion (D)	Fre-quen cy (F)	D/F
0	1	1	0	cclothes	1,319.0	3	439.7
1	1	1	0	celluse	4,986.0	5	997.2
2	1	1	0	eating	4,407.0	2	2203.5
3	1	1	0	goout	7,865.0	1	7865.0
4	1	1	0	hairdray	187.0	1	187.0

The columns referring to the following activities were maintained and renamed: Change clothes, Eating, Making coffee, Making hot food, Toileting and Washing hands/face renamed to respectively clotheses, eating, mcoffee, mhotfood, toileting and whandface. Table IV partially presents the 3 new datasets.

TABLE IV
PRE-PROCESSING STEP 3

a. dataset attribute duration

Day	Day of the Week	Weekend	cclothes	eating	mcoffee	mhotfoot	toileting	whandface
1	1	0	1319.0	4407.0	1600.0	2684.0	2840.0	2971.0
2	2	0	1242.0	7409.0	2122.0	4710.0	1937.0	2659.0
3	3	0	1513.0	3845.0	1157.0	6173.0	2800.0	2359.0
4	4	0	959.0	7010.0	1400.0	5820.0	2401.0	2862.0
5	5	1	947.0	8027.0	846.0	5823.0	1429.0	1066.0

b. dataset attribute frequency

Day	Day of the Week	Weekend	cclothes	eating	mcoffee	mhotfoot	toileting	whandface
1	1	0	3.0	2.0	4.0	1.0	6.0	10.0
2	2	0	3.0	3.0	5.0	2.0	4.0	10.0
3	3	0	3.0	2.0	3.0	2.0	7.0	8.0
4	4	0	2.0	3.0	3.0	2.0	5.0	11.0
5	5	1	2.0	3.0	2.0	2.0	3.0	4.0

c. dataset attribute duration/frequency

Day	Day of the Week	Weekend	cclothes	eating	mcoffee	mhotfoot	toileting	whandface
1	1	0	439.667	2203.50	400.000	2684.0	473.333	297.1000
2	2	0	414.000	2469.67	424.000	2355.0	484.250	265.9000
3	3	0	504.333	1922.50	385.667	3086.5	400.000	294.8750
4	4	0	479.500	2336.67	466.667	2910.0	480.200	260.1818
5	5	1	473.500	2675.67	423.000	2911.5	476.333	266.5000

The fourth stage of pre-processing involves the creation of a dataset for each evaluated activity. These datasets will be used in the second stage of the process, that of identifying the activity(ies) that originated the abnormal classification of the day. For each dataset, contextual and temporal attributes were used. The partial result can be seen in Table V, showing the dataset created for the Change clothes activity. The fifth stage of pre-processing consists of creating training, validation, and test datasets. For the training data, the first 20 weeks (140 days) of the pre-processed datasets are used. For validation data, data from weeks 21 to 33 (12 weeks - 84 days) are used. The test data is a copy of the validation data, with the addition of artificial anomalies to part of this data.

TABLE V
PRE-PROCESSING STEP 4

Day of the Week	Week end	Activity	Duration (D)	Frequency (F)	D/F
1	1	0	1,319.0	3	439.67
2	2	0	1,242.0	3	414.00
3	3	0	1,513.0	3	504.33
4	4	0	959.0	2	479.50
5	5	1	947.0	2	473.50

Artificial anomalies were added to the first 42 days (6 weeks) of test data. Randomly, for each day, 2 to 6 activities were selected, and randomly one of the following modifications was added: (i) added from 20 to 60% increment to the frequency attribute, and applied a decrease from 20 to 60 % in the duration attribute; (ii) added a 20 to 60% decrease to the frequency attribute, and applied a 20 to 60% increase to the duration attribute; (iii) for days where the observed frequency is less than 2, an increase of 20 to 60% was applied to the duration attribute, and a decrease of 20 to 60% was applied to the frequency attribute. The duration/frequency attribute has been recalculated to reflect changes made to test data across all test datasets. The first 42 days and their modified activities were marked as abnormal and the final 42 days and their respective activities were marked as normal. Fig. 4 presents the perturbations added to the data, in comparison to the same data before being modified, for the Change clothes activity.

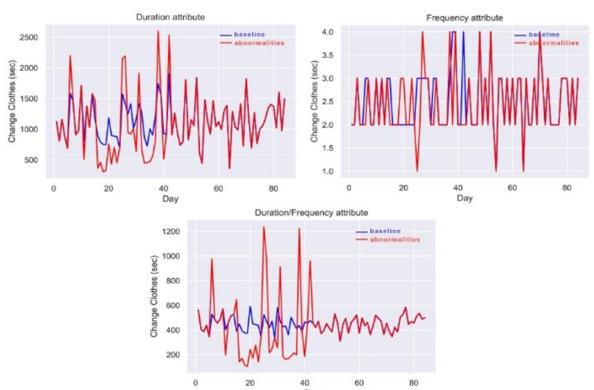


Fig 4. Abnormalities added to Change clothes activity.

IV. RESULTS

For the detection of anomalies in ADL, historical data from the performance of these activities are used for training ND models. These models learn the normal behavior of the elderly in carrying out these activities, and based on this, evaluate whether new daily data present anomalies about the learned behavior. This section presents the performance of the models for the steps of detecting abnormal days and identifying anomalous activities on abnormal days. For the first stage, the temporal attributes were evaluated; for both stages, the number of weeks of training and the overall performance of the models were evaluated. The steps were evaluated separately, to later compose the proposed solution.

Fig. 5 presents the generalization of the models according to the temporal attributes, considering the Accuracy [35]. It can be observed that, regardless of the model, the temporal attribute duration/frequency allowed a greater generalization in all tested algorithms. The Duration attribute allowed a greater generalization than the Frequency attribute when the LOF and OC-SVM algorithms were applied. The Frequency attribute had a better performance compared to the Duration attribute in the application of the iForest and RCE algorithms.

Fig. 6 shows the overall performance of the models using contextual attributes and the temporal attribute duration/frequency. The LOF and RCE models performed better than the others. Analyzing the performance of the models in terms of Accuracy, it appears that the LOF and RCE models obtained the best performance, and the OC-SVM model obtained the worst performance. Regarding Precision, again the LOF and RCE models had the best performance, and the iForest model had the worst performance.

Considering the Recall, the LOF, RCE, and iForest models had a very similar performance, however, the OC-SVM model performed well below the others. About F1-Score, again the LOF and RCE models performed better. The OC-SVM model had the worst performance. It is also verified that, for this metric, the LOF and RCE models perform better than the iForest model.

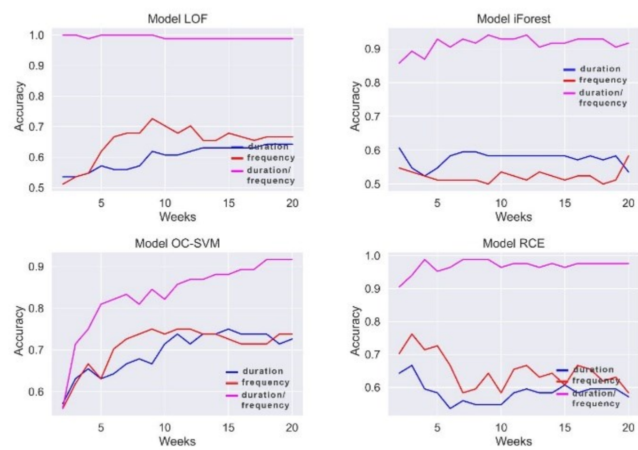


Fig 5. Generalization of predictive temporal attributes according to respective

The performance can also be seen in the bar chart shown in Fig. 7, restricted to 10 weeks of training. It can also be seen that the OC-SVM model presented the lowest performance among the models in this step. Observing the result of Accuracy, over the analyzed weeks, the LOF model presented the best performance, followed by the RCE model and the iForest model. It is noticed that the OC-SVM model performs well below the others. Regarding Precision, the LOF model obtained the best performance, according to the RCE model. In the second week of observations, the RCE and OC-SVM models had practically equal performances, while the iForest model achieved the worst performance. In weeks 3, 5, 6, and 7, the OC-SVM and iForest models performed similarly, but below the performance of the LOF and RCE models.

In the evaluation of the Recall and the F1-Score, the LOF model again obtained the best performance. It is noteworthy that in the evaluation of the Recall, the OC-SVM model had the worst performance in all the analyzed weeks. Based on the

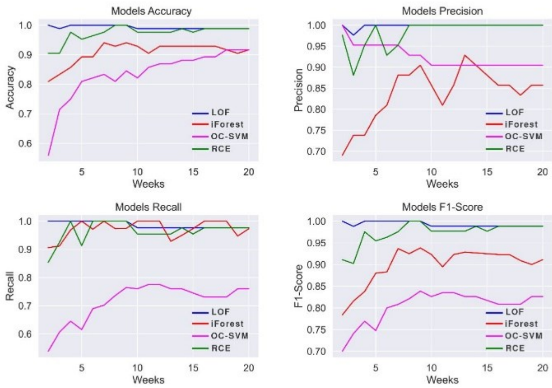


Fig 6. Performance of models with duration/frequency predictive attribute.

presented data, it is concluded that the LOF model with the use of the temporal attribute duration/Frequency, manages to reach accuracy and F1-Score superior to 90% with two weeks of training data. Table VI presents the metrics for the LOF with the duration/frequency attribute, using 2 to 6 weeks of training data.

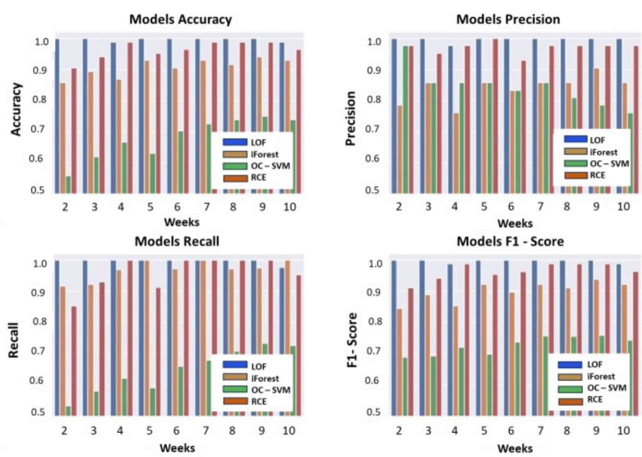


Fig 7. Performance of models with duration/frequency predictive attribute.

TABLE VI
LOF PERFORMANCE, 2 TO 6 WEEKS OF TRAINING WITH DURATION/FREQUENCY ATTRIBUTE

Weeks	TP	TN	FP	FN	Ac	Pc	Rc	F1
2	42	42	0	0	1	1	1	1
3	42	42	0	0	1	1	1	1
4	41	42	1	0	0,99	0,98	1	0,99
5	42	42	0	0	1	1	1	1
6	42	42	0	0	1	1	1	1

Abbreviations: TP - true positive, TN -true negative, FP - false positive, FN - false negative, Ac - Accuracy, Pc - Precision, Rc - Recall, and F1 - F1-Score.

A. Detection of Anomalous Activities on Abnormal Days

Fig. 8 shows the performance of the models for the activities Change Clothes, Eating, and Making Coffee. And Fig. 9 presents the generalization of predictive temporal attributes for the activities Making hot food, Toileting, and Washing hand/face.

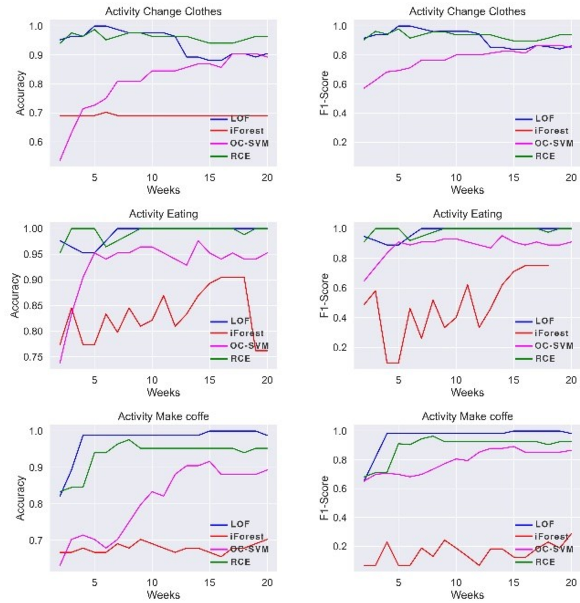


Fig. 8. Generalization of predictive temporal attributes for the activities of Change Clothes, Eating, and Making coffee.

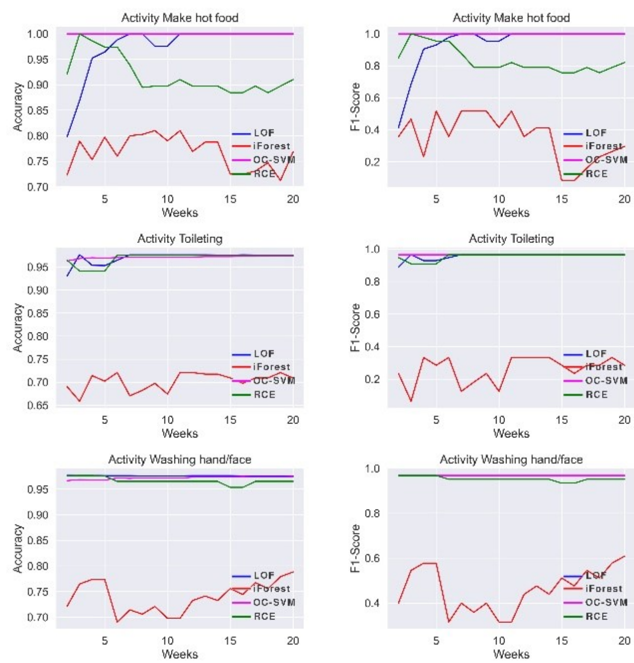


Fig. 9. Generalization of predictive temporal attributes for the activities Making hot food, Toileting, and Washing hand/face.

Table VII presents the accuracy values for the first 5 weeks for the LOF models. It can be observed that the LOF presents an accumulated average higher than the RCE, however, with a little significant difference. However, we chose to use the LOF, and both models can be used with accuracies greater than 90% for this set of activities. Based on the data presented, it is concluded that the LOF model can achieve accuracy and F1-Score greater than 90% in all activities with two weeks of training data.

TABLE VII
PERFORMANCE OF THE LOF MODEL FOR THE EVALUATED ACTIVITIES

Weeks	LOF- Accuracy						
	Cc	eat	mcf	mhf	toil	whf	av
2	0.95	0.98	0.82	0.80	0.93	0.98	0.91
3	0.96	0.96	0.89	0.87	0.98	0.98	0.94
4	0.96	0.95	0.99	0.95	0.95	0.98	0.96
5	1	0.95	0.99	0.96	0.95	0.98	0.97
6	1	0.98	0.99	0.99	0.96	0.99	0.99

Abbreviations: Cc - Change Clothes; eat - Eating; mcf - Making Coffee; mht - Making hot food; toil - Toileting; whf - Washing hand/face; and av - average.

B. *Suggested Model*

The final configuration for the proposed model includes the training of a LOF model using the predictive attribute duration/frequency for the detection of daily anomalies. The detection of anomalous activities on abnormal days, consists of training a LOF model, for each evaluated activity, using all contextual and temporal attributes. Regarding the number of weeks for training the model, it can be trained with and from two weeks of training data, reaching an overall average above 90%. Table VIII presents the metrics achieved by the model with two to six weeks of training data.

TABLE VIII
GENERAL PERFORMANCE OF THE PROPOSED MODEL

Weeks	Accuracy Stage 1	Accuracy Stage 2	Average
2	1	0.91	0.95
3	1	0.94	0.97
4	0.99	0.96	0.97
5	1	0.97	0.98
6	1	0.99	0.99

Based on the data presented, it is concluded that the proposed model can achieve accuracy and F1-Score greater than 90% with only 2 weeks of training data. Fig. 10 presents the configuration of the proposed model.

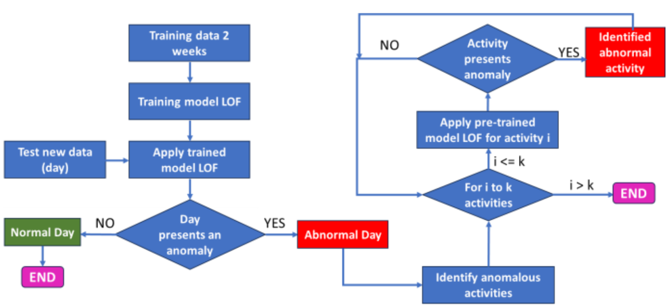


Fig. 10. Configuration of the proposed model.

Based on the proposed model, it is possible to create a system that identifies abnormalities in activities of daily living, based on a person's monitoring data. Fig. 11 presents the overview of the system.

As shown in Fig. 11, the system receives data regarding the day's activities. Such data can be collected via sensors or typed by a user who accompanies the person being monitored. The day's data are pre-processed by a routine created in Python and then sent to apply the algorithms in two stages. In the end, the system displays whether the day was normal or if there was an anomaly and what the anomaly was.

This system is in the implementation phase, for later testing in a real environment.

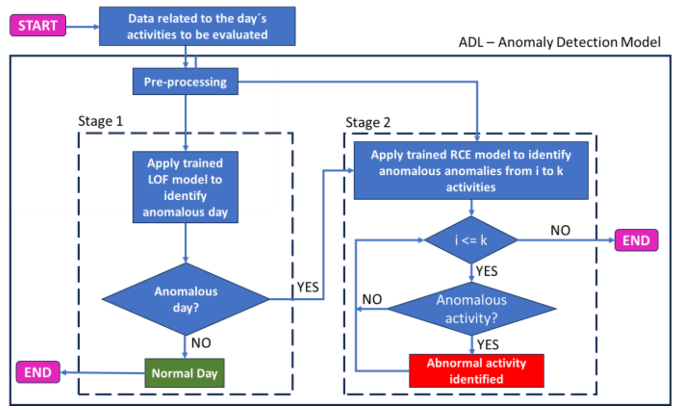


Fig. 11. System Overview.

V. CONCLUSIONS

This work aimed to create a solution to identify an abnormality in ADL of the elderly in the domestic context with the use of Novelty Detection. Its relevance lies in considering the number of weeks of follow-up of the individual and not just one week. The number of weeks of training data is relevant, as the implementation of monitoring systems and the initiation of monitoring of the elderly depend on this. Existing solutions do not define the minimum amount of training data for satisfactory model performance. In this way, the contribution of this work is to propose a model for the detection of anomalies in ADL reaching accuracy and F1-score superior to 90% with only two weeks of training data.

The database used was the eHealth Monitoring Open Data Project. This base was selected because it presents several desirable properties for the study. Such as the number of observations, the absence of anomalies in the data, and the number of activities observed. In the pre-processing stage, the data were organized into training, validation, and testing, artificial anomalies were added to the test data, to enable performance analysis.

The proposed solution consisted of two steps, the first was to identify abnormal days, and the second was to identify anomalous activities on abnormal days. Among the objectives of the work are, the identification of the amount of training data necessary for a greater than 90%, the evaluation of the temporal attribute of greater generalization for the first stage, the evaluation of the performance of the model's LOF, iForest, OC-SVM and RCE.

The LOF and RCE models performed better in both stages and the first stage the OC-SVM presented the worst results. However, in the second stage, OC-SVM presented good results, and iForest presented inferior results. Even so, the LOF model performed better than the others in both stages. Regarding the predictive temporal attribute, the duration/frequency attribute showed better generalization regardless of the model. Since, in the second stage, the contextual and temporal attributes were used for training, as it is the assessment of activities individually. Regarding the minimum training data, it was observed that two weeks of training were enough to reach the desired metrics.

The models tested in this work showed similar performance, considering contextual or temporal attributes individually. The iForest model obtained the worst performance in all situations, and the LOF and RCE models obtained the best performances.

Thus, for the proposed solution, the LOF model was selected for both the identification of abnormal days and the identification of anomalous activities on abnormal days. Making use of two weeks of training data achieving accuracy and F1-Score greater than 90%. However, the use of the RCE algorithms for the first stage, and the OC-SVM for the second, are not ruled out, as they present relevant performance. It is important to highlight that the model considered data relating to an interval of six weeks, that is, a month and a half. This range was considered the minimum satisfactory for a range of everyday situations.

Other aspects that must be considered refer to more complex scenarios. For situations with more than one person being monitored in the same environment, and people with diseases such as Alzheimer's, it is necessary to include new attributes in the model to consider these issues.

Consequently, new experiments must be carried out to adjust the model. It is also important to note that the application of this study in real-world environments still requires a series of studies. It is considered that there is a need to develop a platform based on AIoT (Artificial Intelligence of Things), which supports all necessary data collection, as well as the processing of this data and provision of insights and information to users. This platform must be developed considering the adaptability of the model.

Novelty detection techniques have the advantage of not requiring data referring to the abnormal class. However, this can represent a problem, because to evaluate the performance of the models it is necessary to create artificial anomalies. Since these anomalies may not represent the full complexity of the real scenario, where everyone has their behavior pattern and different health conditions. In this way, it is necessary to implement the proposed solution in a real scenario, for an assertive evaluation. As future works, it is considered to apply the experiments of this study in a real scenario, as well as to evaluate the performance of the proposed solution with ADL data collected directly from real environments.

REFERENCES

[1] WHO – World Health Organization, 2021. *Aging and Health*. Accessed on: November 16th, 2022 [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-andhealth.

[2] WHO – World Health Organization, 2017. *10 Facts on Ageing and Health*. Accessed on: January 10th, 2023. [Online] Available: https://www.who.int/news-room/factsheets/detail/10-facts-on-ageing-and-health.

[3] NCOA - National Council on Aging, 2021. *Chronic Conditions for Older Adults*. Accessed on January 3rd, 2023. [Online] Available: https://www.ncoa.org/article/the-top-10-most-common-chronic-conditions-in-older-adults.

[4] JCHS – Join Center for Housing Studies of Harvard University, 2016. Accessed on February 12th, 2023. [Online] Available: https://www.jchs.harvard.edu/research-areas/reports/projections-and-implications-housing-growing-populationolder-households.

[5] UN - United Nations, 2022. *Aging and disability*. Accessed on December 18th, 2022. [Online] Available: https://www.un.org/development/desa/disabilities/disability-andageing.

[6] R. M., Gomathi and M. Nithya. "Human activity recognition with smartphones using machine learning process". *Proc. Of International Conference on Computer Communication and Informatics* (ICCCI), Hammamet, Tunisia, 2022, pp. 1-10. DOI: 10.1109/ICCCI54379.2022.9740773.

[7] D. Arifoglu and A. Bouchachia, "A. Detection of abnormal behavior for dementia sufferers using Convolutional Neural Networks". *Artificial intelligence in medicine*, v. 94, p. 88-95, 2019. DOI: 10.1016/j.artmed.2019.01.005.

[8] H. Ghayvat, M. Awais, S. Pandya, H. Ren, S. Akbarzadeh, S. Chandra Mukhopadhyay, C. Chen, P. Gope, A. Chouhan, W. Chen. "Smart Aging System: Uncovering the Hidden Wellness Parameter for Well-Being Monitoring and Anomaly Detection". *Sensors*. 2019; 19(4):766. DOI: 10.3390/s19040766.

[9] C. Del-Valle-Soto, R. A. Briseño, L. J. Valdivia, R. Velázquez, J. A. Nolazco-Flores. "Non-Invasive Monitoring of Vital Signs for the Elderly Using Low-Cost Wireless Sensor Networks: Exploring the Impact on Sleep and Home Security". *Future Internet*. 2023; 15(9):287. DOI: 10.3390/fi15090287.

[10] L. G. Fahad, S. F. Tahir. "Activity recognition and anomaly detection in smart homes". *Neurocomputing*. 2021, vol 423, 362-372. ISSN 0925-2312. DOI: 10.1016/j.neucom.2020.10.102.

[11] S. W. Yahaya, A. Lotfi, A. and M. Mahmud, "Detecting anomaly and its sources in activities of daily living". *SN Computer Science*, v. 2, n. 1, p. 1-18, 2021. DOI: 10.1007/s42979-020-00418-2.

[12] B. Ouafae, L. Oumaima, R. Mariam and L. Abdelouahid, "Novelty Detection Review State of Art and Discussion of New Innovations in The Main Application Domains,"*2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 2020, pp. 1-7, DOI 10.1109/IRASET48871.2020.9092147.

[13] V. R. Leithardt, D. Nunes, A. G. Rossetto, C. O. Rolim, Geyer C. F., J. S. Silva, "Privacy Management Solution in Ubiquitous Environments Using Percontrol". *Journal of Ubiquitous Systems & Pervasive Networks*, v.5, n.2, p. 21-28, 2014. DOI: 10.5383/JUSPN.05.02.004.

[14] A. Howedi A., Lotfi and A. Pourabdollah, "An entropy-based approach for anomaly detection in activities of daily living in the presence of a visitor". *Entropy*, v. 22, n. 8, p. 845, 2020. DOI: 10.3390/e22080845.

[15] S. O. P Pinto and V. S. Sobreiro, "Literature review: Anomaly detection approaches on digital business financial systems". *Digital Business*, v. 2, Issue 2, 2022, 100038, ISSN 2666-9544. DOI: /10.1016/j.digbus.2022.100038.

[16] M. Fahim and A. Sillitti, "Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review," in *IEEE Access*, vol. 7, pp. 81664-81681, 2019, DOI: 10.1109/ACCESS.2019.2921912.

[17] T. Sureda Riera, J. R. Bermejo Higuera J. Bermejo Higuera, J. J. Martínez Herraiz, J. A. Sicilia Montalvo. "Prevention and Fighting against Web Attacks through Anomaly Detection Technology. A Systematic Review". *Sustainability*. 2020; 12 (12): 4945. DOI: 10.3390/su12124945.

[18] M. P. de Freitas, V. A. Piai, R. H. Farias, A. M. R. Fernandes, A. G. de Moraes Rossetto, V. R. Q. Leithardt, "Artificial Intelligence of Things Applied to Assistive Technology: A Systematic Literature Review". *Sensors*. 2022; 22(21):8531. DOI: 10.3390/s22218531.

[19] M. Koutli, N. Theologou, A. Tryferidis, D. Tzovaras, “Abnormal Behavior Detection for Elderly People Living Alone Leveraging IoT Sensors”. *Proc. of IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2019, pages 922-926DO 922-926. DOI: 10.1109/BIBE.2019.00173.

[20] S. Yahaya, A. Lotfi, M. Mahmud, “A Consensus Novelty Detection Ensemble Approach for Anomaly Detection in Activities of Daily Living”. *Applied Soft Computing.* Volume 83, 2019, 105613, ISSN 1568-4946. DOI: 10.1016/j.asoc.2019.105613.

[21] Z. Zhang, S. Ishida, S. Tagashira, A. Fukuda, “A. Danger-Pose Detection System Using Commodity Wi-Fi for Bathroom Monitoring”. *Sensors*. 2019; 19(4):884. DOI: 10.3390/s19040884.

[22] D. Zekri, T. Delot, M. Thilliez, S. Lecomte, M. Desertot, “A Framework for Detecting and Analyzing Behavior Changes of Elderly People over Time Using Learning Techniques”. *Sensors*. 2020; 20(24):7112. DOI: 10.3390/s20247112.

[23] A. Howedi, L. Ahmad and P. Amir, “An Entropy-Based Approach for Anomaly Detection in Activities of Daily Living in the Presence of a Visitor”. *Entropy*. 2020, 22, no. 8: 845. DOI: 10.3390/e22080845.

[24] A. Omar, A. Raed, S. Terence, M. Xiaogang, “A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams”. *Big Data Cogn. Comput*. 2021, 5(1), 1. DOI: 10.3390/bdcc5010001.

[25] C. Zhangyu, Z., Chengming, D., Jianwei. “Outlier detection using isolation forest and local outlier factor”. *Proc. of the Conference on Research in Adaptive and Convergent Systems (RACS ’19). Association for Computing Machinery*, 2019, New York, NY, USA, 161–168. DOI: 10.1145/3338840.3355641.

[26] C. C. Aggarwal. *Outlier Analysis*. Berlim, Germany: Springer; 2nd ed., 2017. ISBN: 978-3-319-47578-3.

[27] X. Tao, Y. Peng, F. Zhao, P. Zhao, Y. Wang, Y. “A parallel algorithm for network traffic anomaly detection based on Isolation Forest”. International *Journal of Distributed Sensor Networks*, v. 14, n. 11, 2018, DOI: 10.1177/1550147718814471.

[28] P. Skelsey, “Forecasting risk of crop disease with anomaly detection algorithms”. *Phytopathology R*, v. 111, n. 2, p. 321-332, 2021. DOI:10.1094/PHYTO-05-20-0185-R.

[29] I. A. Bozdog, T. Daniel-Nicusor, M. Antal, C. Antal, T. Cioara, I. Anghel, I. Salomie, "Human Behavior and Anomaly Detection using Machine Learning and Wearable Sensors,"*2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, Romania, 2021, pp. 383-390, DOI: 10.1109/ICCP53602.2021.9733684.

[30] eHealth Monitoring Open Data Project, 2018. Accessed on: November 22nd, 2022. [Online] Available: https://sourceforge.net/projects/ehealthmonitoringproject/.

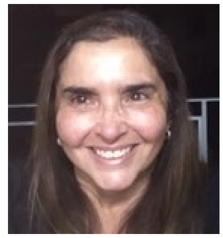
[31] CASAS: Center of Advanced Studies in Adaptive System. Accessed on December 13th, 2022. [Online] Available at: http://casas.wsu.edu/datasets/.

[32] DaLiAc – Daily Life Activities. Accessed on December 1st, 2022. [Online] Available at: https://www.mad.tf.fau.de/research/activitynet/daliac-daily-lifeactivities/.

[33] ADLs: Activities of Daily Living (ADLs) Recognition Using Binary Sensors Version 1.0. Accessed on January 10th, 2023. [Online] Available: https://archive.ics.uci.edu/ml/datasets/Activities+of+Daily+Living+(ADLs)+Recognition+Using+Binary+Sensors.

[34] H. Mshali, T. Lemlouma, and D. Magoni, "Analysis of dependency evaluation models for eHealth services," *2014 IEEE Global Communications Conference*, Austin, TX, USA, 2014, pp. 2429-2435, DOI: 10.1109/GLOCOM.2014.7037172.

[35] R.R. de Lima, A. M. R. Fernandes, J. R. Bombasar, B.A. da Silva, P. Crocker, V. R. Q. Leithardt, “An Empirical Comparison of Portuguese and Multilingual BERT Models for Auto-Classification of NCM Codes in International Trade”. *Big Data and Cognitive Computing*. 2022; 6(1):8 DOI:10.3390/bdcc6010008.



A. M. R. Fernandes received the Ph.D. degree from University of Santa Catarina in 2000. Her research interests are Artificial Intelligence Applied to Health and Data Science. She is currently a Professor with Universidade do Vale do Itajaí – Brasil.

R. Q. Leithardt is with the Instituto Superior de Engenharia de Lisboa (ISEL), Lisboa – Portugal, and Senior Member of IEEE. He received the Ph.D. degree in Computer Science from INF-UFRGS, Brazil, in 2015. He is also a Collaborating Researcher with the Expert Systems and Applications Laboratory, University of Salamanca, Spain.

J. F. P. Santana received the Ph.D. degree in Computer Science from the University of Salamanca, Spain, in 2003, 2005, 2007, and 2010, respectively. He is currently a Full Professor with the University of Salamanca and a Researcher with the Expert Systems and Applications Laboratory.