Semi-supervised Adaptive Method for Human Activities Recognition (HAR)

Fabio Mendoza Palechor[1(✉)], Enrico Vicario[2], Fulvio Patara[2], Alexis De la Hoz Manotas[1(✉)], and Diego Molina Estren[1(✉)]

[1] Department of Computer Science and Electronics, Universidad de la Costa, Barranquilla, Colombia
{fmendoza1,adelahoz6,dmolina}@cuc.edu.co
[2] Department of Systems Engineering, Università degli Studi di Firenze, Firenze, Italy

Abstract. Using sensors and mobile devices integrated with hardware and software tools for Human Recognition Activities (HAR), is a growing scientific field, the analysis based on this information have promising benefits to detect regular and irregular behaviors in individuals during their daily activities. In this study, the Van Kasteren dataset was used for the experimental stage, and it all data was processed using the data mining classification methods: Decision Trees (DT), Support Vector Machines (SVM) and Naïve Bayes (NB). These methods were applied during the training and validation processes with the proposed methodology, and the results obtained showed that all these three methods were successful to identify the cluster associated to the activities contained in the Van Kasteren dataset. The Support Vector Machines (SVM) method showed the best results with the evaluation metrics: True Positive Rate (TPR) 99.2%, False Positive Rate (FPR) 0.6%, precision (99.2%), coverage (99.2%) and F-Measure (98.8%).

Keywords: HAR · Data mining · Cluster · Evaluation metric · Dataset · Van Karesten

1 Introduction

Using sensors and mobile devices integrated with hardware and software tools for Human Recognition Activities (HAR), is a growing scientific field, the results have promising benefits for detecting regular and irregular behaviors in individuals during their daily activities. Many studies related to HAR can be found, studies like [1], where authors used several computational intelligence methods to identify actions like crouch, stand up, jump, run, walk and others, and there are applications in wide areas such as healthcare, security, domotics, sports and others [2–5]. Several data mining techniques (DM), machine learning (ML), and deep learning (DL) play a relevant role inside HAR, since they provide tools to perform tasks such as classification, segmentation and hybrid systems. These methods can be compared, based on their results, as you can see in [6–9]. The implementation of these methods can be found in several studies, such as K-Nearest Neighbor (KNN) [1, 8, 10, 11], Support Vector Machine (SVM) [1, 6, 10, 12, 13], Naive

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
K. Saeed and J. Dvorský (Eds.): CISIM 2022, LNCS 13293, pp. 3–17, 2022.
https://doi.org/10.1007/978-3-031-10539-5_1

Bayes (NB) [14, 15], Decision Trees (DT) [1, 16], and Neural Convolutional Networks (CNN) [17]. Datasets play an important role in the process of HAR, since scientists can develop new solutions for efficient identification of activities, in the literature you can see several studies where the authors use different datasets [1, 10, 12, 18, 19]. The most representative datasets related to HAR are Van Kasteren [20], CASAS Kyoto [21], CASAS Aruba [22], CASAS Multiresident [23], Opportunity [24–26], UCI HAR [27, 28] and mHealth [29, 30]. The first five datasets, contain data captured from individual interactions with sensors deployed in indoor environments and the last two datasets, contain data collected using body sensors (wearables) or smartphones. Ensuring the high quality of a HAR model is a complex task, due to the elevated number of evaluation metrics for validation, as you can see in [17, 31, 32]. The main objective of this study is to present a model based on data mining techniques for HAR, the proposed approach integrates the unsupervised cluster method K-Means with the classification techniques SVM, DT and NB. These were compared to achieve the best results in the validation metrics: precision, coverage, False Positive Rate FPR, True Positive Rate TPR and F-Measure. The dataset chosen by this study was Van Kasteren, due the high quality of its data and, the number of existing recommended studies of this kind. The study is structured as follows: in Sect. 2 you can find previous studies related to HAR, Sect. 3 presents the materials and methods used in the current study, Sect. 4 shows the methodology, Sect. 5 presents the results obtained, and finally Sect. 6, show the conclusions and future works.

2 Related Works

Several authors propose the implementation of techniques or methods for the development of tools that can correctly identify human activities, different hardware devices and software are used to collect data, for later analysis related to HAR. Next, you can find a brief description of some contributions in this field of study.

In [7], authors manifested difficulties in activity recognition based on sensor networks in SMART HOME, usually, probabilistic models such as Markov Hidden model (HMM) and Linear Discriminant Analysis (LDA), are used to classify activities. In that study, SVM based on minority sampling (SMOTE) overcome HMM, LDA and SVM, and this approach could lead to a significant increase in performance for the classification process. The Van Kasteren dataset was used during the experimentation process, and accuracy was the chosen evaluation metric. In [8], the performance of several classification algorithms was analyzed, based on a system for recognition of online activities under Android platforms. In this study, the classification algorithm KNN was improved with clustering algorithms. Grouped KNN eliminates the computational complexity, and the performance is evaluated in four test subjects for activities like walking, running, sitting and standing up. According to [1], the recognition of human activities using automated methods has surged recently as a key research topic, in this study, authors proposed an optical flow descriptor based on features derived from movement, human action is analyzed through a histogram containing local and global cinematic features, authors used the UCF101 and Weizman datasets, and they used the classification techniques KNN, DT, SVM and DL. The classification rates obtained were 98.76% using KNN for Weizman and 70% for UCF101. In [11], the study showed the percentage of the population that owns a smartphone has increased significantly recently, and this fact can be used to collect information about the context of the user. In this study, the authors proposed a HAR model, using smartphones to obtain data, several statistics from the sensor data were calculated to model each activity and, a learning supervised technique performed classification in real time. The methods KNN and a classifier based on K-means were compared, and the activities analyzed were still, walking, running and vehicle. According to [16], human movement detection is an important task to several areas such as healthcare, physical state and elder care. This study proposed a model for human movement detection using smartphone sensors and learning automated methods. The data was obtained from the accelerometer, gyroscope, step counter and cardiac frequency sensors. The algorithms used were KNN, DT, Random Forest, SVM with PCA and evaluation metrics were f-measure, accuracy and area under curve. In [17], the study shows that HAR is a field that has attracted much attention in the last few years, based on the demand in several areas. In this study, the authors proposed the implementation of a Deep Neural Convolutional Network (Convent) with the objective of making efficient and effective HAR based on smartphone sensors. The experimentation showed that Convent derive relevant and more complex features with each extra layer, even the complexity level diminishes from layer to layer. Neural networks overcome other data mining techniques in HAR, using a dataset collected from 30 volunteer subjects, achieving a general performance of 94.79% in the test dataset and 95.75% with extra information from temporal Fast Fourier Transform in the HAR dataset. According to [19], HAR represents a great contribution for health monitoring, in the study, the authors proposed an approach based on SMART HOME, data was collected through several sensors in different locations. The dataset used was Van Kasteren and CASAS. In their approach, authors first selected a class for each activity, and then the key features are used for activity recognition. The methods used were Multi-class SVM, Correct Incorrect Distance Learning and Ranking (CIDLR) and Direct Distance Minimization (DDM). The results obtained for HAR achieved 84.24% for Van Kasteren and 97.5% for CASAS. In [33], the study showed that human activity recognition must consider the semantics of the environment and the relationship between the elements on it. The authors called this "activity space", the location and the artifacts in it (identified by RFID tags), allowed to apply PMM to recognize the activity performed by the individual. Preliminary studies indicated that the proposed model was noise tolerant and had a high accuracy to detect the activity space of the individual, and the ability to handle big volume of data. The study in [34] showed a smart home with a ubiquitous intelligent monitoring system, with 4 sensors (ECG, Environment Temperature, Location Camera, Facial and Expression Sensor) and 7 contexts (ECG, pulse, body temperature, environment temperature, location, movement and facial expression). In this study, authors used LSVM (Support Vector Machine) and association rules to analyze captured patterns. The proposed model identified with a high level of accuracy (over 70%), the service needed by the human based on the analysis of the information provided by the sensors. Authors in [35], manifested that HAR provides valuable context information for wellness, healthcare and sports applications, recently many approaches have been proposed for automated learning, to identify activities from sensor data. Nevertheless, most methods are designed for off-line processing, and not on the sensor node, the proposed approach was implemented using an application for Android devices, 4 datasets were used: ActiveMiles, WISDM v1.1., Daphnet FoG and Skoda. Crossed validation was made with 10 subsets. In [36], the authors proposed the learning of composite features formed by 2D single corners in time and space, to recognize human activities. Each corner is codified with relation to their neighbors and the whole set. The extraction of composite features was made by data mining (association rules). The final classifiers can be used to perform activity localization and classification. According to [37], gesture and movement detection for hands is quite challenging, especially in real environments. This study performed the recording of a hand, and 5 different gestures, and the gestures are replaced later using the interaction with a computer mouse. The hand silhouette was extracted as a combination of different segmentation methods, to produce an invariant affine Fourier descriptor and then to be classified by SVM. The gestures are recognized was changes by a Finite State Machine FSM. According to [38], the authors proposed a hierarchical method to detect and recognize human activities in MPEG sequences. The algorithm presented 3 stages: 1. Using PCA analysis from the movement vectors in MPEG, grouping macroblocks per speed, distance and human proportions, 2. The DC DCT components of luminance and chrominance are compared with activity templates and human skin, 3. An exhaustive analysis of the regions without compression in the previous steps through segmentation and graphs. According to [39], authors presented an approach to group human activity analysis in a smart home environment based on auto adaptive neural network called Growing Self-Organizing Maps (GSOM). The dataset used was collected from several sensors in an apartment within a period of two weeks. The results indicated that GSOM showed attractive features for interpretation and analysis of useful patterns present in the data from daily activities.

3 Materials and Methods

3.1 Dataset Description

The dataset selected for the proposed method is Van Kasteren [20]. The data was collected by different sensors such as layer switches, mercury contacts, infrared passive sensors PIR and flotation sensors. The collection process was made during two weeks in an apartment occupied by two men. For the annotation process, they used Bluetooth headsets with voice recognition and an activity log handwritten. The data are binary type, and you can see the activation, on and off, from the sensors. The actions identified were grouped in the activities: brush teeth, shower, go to bathroom, take a bath, shave, breakfast, dinner, have snack, drink, load dishwasher and unload dishwasher and others. The dataset presents a structure formed by two plain text documents called KasterenSenseData and KasterenActData, these contain the information of the sensors and the set of activities studied respectively. To start with the analysis of the data, first you must perform a depuration process. First stage involves unifying the data in one single format, storing which sensors were used in each activity. Then, the data integration is determined by the number of times that each activity is found in the dataset, and this is shown in Fig. 1.

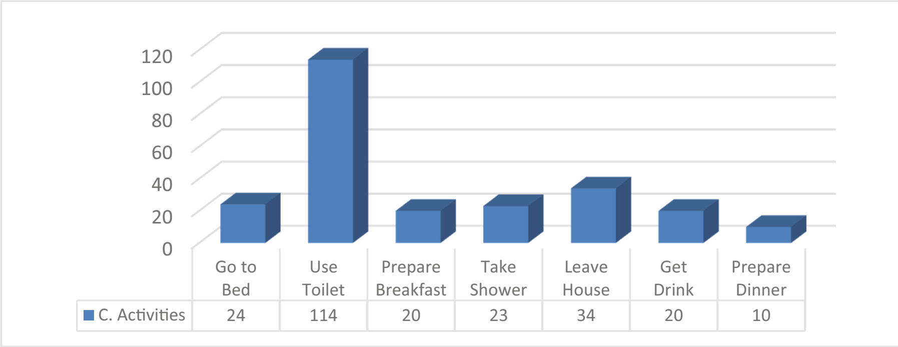

Fig. 1. Number of activities by category.

In Fig. 2, you can see the number of activities by capture date, based on the sensors installed inside the indoor environment of the study.

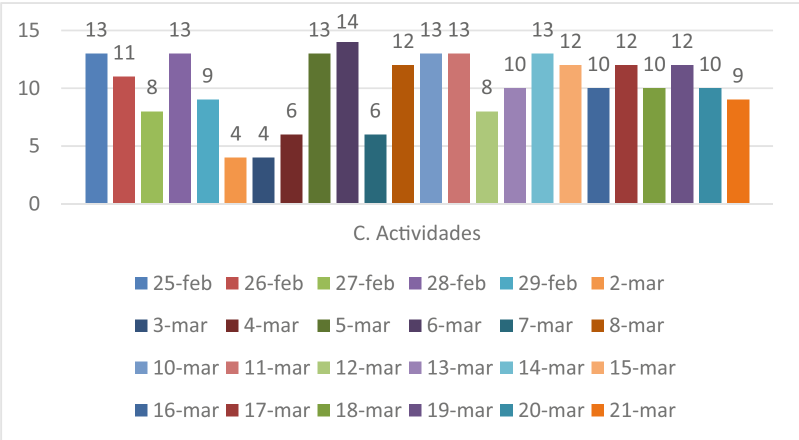

Fig. 2. Number of activities by date.

After the dataset was integrated, the process of separating and transforming the data was made, including an analysis of the distribution of the data with relation to the frequency of occurrence of each activity, considering that the dataset is unbalanced, also the authors proceeded to the segmentation of the data using the K-Means method, after obtaining the optimal selection of the number of clusters based on the elbow method. Finally, the integration of the segmentation algorithm and the classification algorithms SVM, DT and NB was introduced.

3.2 Simple K-Means

Clustering is a data mining task where analyzed data are divided in groups [40], according to [41], all data belonging to a single cluster are grouped, based on common features of the attributes. There are different methods such as partition, hierarchical, based on density. Simple K-Means is a popular clustering method, considered of hierarchical type, and it is used especially for grouping datasets of low dimension [41], highlighting its simplicity and efficiency, as mentioned by [42, 43]. In [43], authors indicated that the number of groups to create must be known previously.

3.3 Decision Trees

Decision Trees (DT) is a method used in classification processes to perform a partition of the data in subsets, usually the structure of a DT is formed by one single root node and several adjacent nodes called children, the combination of parent and children nodes allows to achieve an answer, found inside the terminal node as mentioned by [44, 45]. According to [46, 47], DT are used extensively for modelling of data in real time prediction systems inside real complex environments. In [48] you can see a DT is used to support decision making processes. In [49] DT are viewed as a supervised approach for classification processes, the most common algorithms are J48, C4.5 and Random Forest.

3.4 Naïve Bayes

According to [50], the Naïve Bayes (NB) method is based on the Bayes theorem with independence between the assumed predicates, a naïve Bayesian model is easy to build, without a complex estimation of the iterative parameters, which makes it useful for big datasets. In [51], the authors manifest that this method is based on classic mathematic theory, the model can be implemented for a wide range of activities, since it needs few estimated parameters, also the methods is insensitive to missing data. Bayesian networks are considered an alternative to classic expert systems oriented to decision making and prediction under uncertainty in probabilistic terms [52]. NB is a probabilistic classifier that calculates a probability set, counting the frequency and combination of the values given by the dataset [53]. The algorithm uses the Bayes theorem and assumes that all data are independent from the values of the class variable [54].

3.5 Support Vector Machines

Support Vector Machines (SVM) is considered a method of high level when the available data is limited, since it brings self-regulation to reduce the overfitting, it is used in many areas such as text and hypertext categorization, handwritten characters recognition, image classification and others [55–59]. Usually, SVM are used to solve classification problems efficiently, since it is an automatic learning system, based in the development of statistical learning systems [60].

4 Methodology

This study is based on the training of the methods DT, SVM and NB using the Van Kasteren dataset under the same training scenario. Previously, the dataset was integrated, as mentioned in Sect. 3.1, then all data was preprocessed and transformed to create the model under the proposed approach, where the best classification algorithm, Simple K-Means, is integrated. You can see each of the stages of the methodology below.

4.1 Data Cleaning and Preparation

In this stage, a meticulous analysis was performed on the data, after their integration, to identify atypical values, redundant and missing data. In this stage, it was noticed that the activities were unbalanced, being the most predominant the "use toilet" activity, which can create an issue of over training in that activity, and reduced accuracy capacity in the rest of the categories described in Sect. 3.1. You can see in Fig. 3, the distribution of the activities according to the dataset.

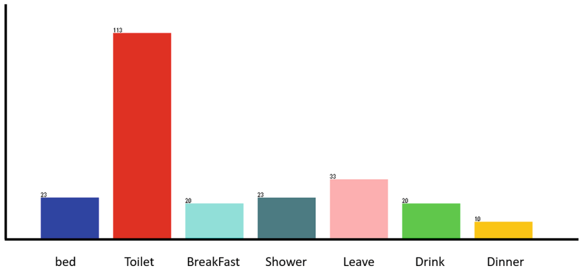

Fig. 3. Activity distribution before data balancing

With this situation, it was necessary to perform a balancing process, on the classes, to obtain best levels in the evaluation metrics, so it was decided to implement the SMOTE algorithm [61], to allow generation of synthetic data for minority classes, based on the number of near neighbors used and the percentage needed to increase the selected class and the random seed for the random sampling. In this stage, it is perfectly clear that all data must be prepared (remove missing data, atypical data, data normalization, etc.) before using SMOTE, since the selected neighbor chosen to generate synthetic data, could contain noise or disturbances, and the data produced would have low quality. Nevertheless, the SMOTE filter has a positive impact when the data are unbalanced, since the balancing processes diminish the probability of biased learning in favor of a majority class [62]. In Fig. 4, you can see the distribution of the activities after the SMOTE filter was applied.

4.2 Cluster Adjustment and Preparation

After the preparation and transformation stage was finished, the next step was to determine the optimal number of clusters to use to create the proposed model. For this task, it was used the elbow method. According to [63], the elbow algorithm allows to determine the real number of groups in a dataset, based on [64], it is a visual method, where k is set starting in 2, the right number of generated groups. Next, in Fig. 5, you can see the results shown by the elbow method.

Based on Fig. 5, you can see the elbow shows in $k = 4$, which means the optimal number of clusters is 4. Then, the generation of 4 groups of the Simple K-Means was implemented. Next you can see the features of data for each group created.

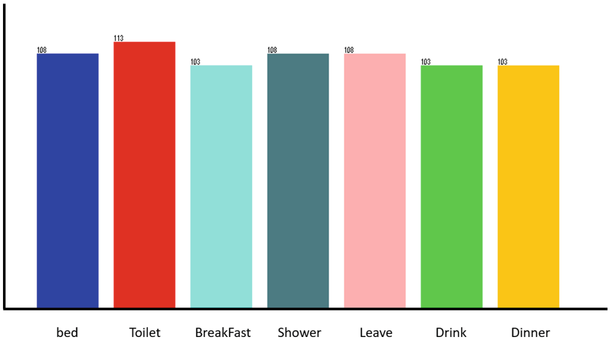

Fig. 4. Activity distribution after data balancing with SMOTE

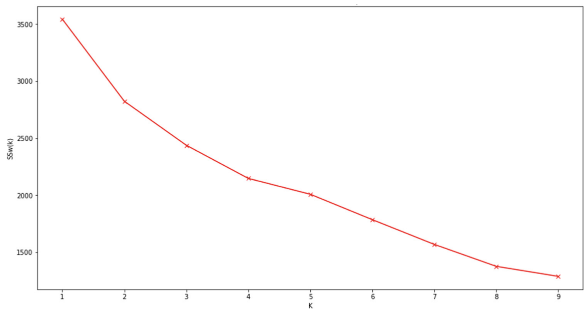

Fig. 5. Elbow method results

- Cluster No 1: Group of the activities associated with the activity “Drink”, where the activation of the sensors CupsCupboard, Fridge and Freezer is present. This group has 103 records, representing 14% of the data.
- Cluster No 2: Group of the activities associated with the activity “Shower”, where the activation of the sensors HallToiletDoor, HallBathroomDoor, ToiletFlush and HallBedRoomDoor is present. This group has 118 records, representing 16% of the data.
- Cluster No 3: Group of the activities associated with the activity “Bed”, where the activation of the sensors HallToiletDoor, HallBathroomDoor, Fridge, FrontDoor, ToiletFlush and HallBedRoomDoor is present. This group has 319 records, representing 43% of the data.
- Cluster No 4: Group of the activities associated with the activity “Breakfast”, where the activation of the sensors Microwave, CupsCupboard, Fridge, PlatesCupBoard, Freezer, GroceriesCupBoard and PansCupBoard is present. This group has 206 records, representing 28% of the data.

With these clusters, you can see the number of activities to be recognized is reduced from the data, but these groups can be related to other activities, as shown in Table 1.

Table 1. Relationship of clusters and activities

Cluster	Activity
1	Drink
2	Shower, Toilet
3	Bed, Leave
4	Breakfast, Dinner

4.3 Activity Classification

After all clusters were defined, the training and testing process was performed, using the classification methods selected: DT, SVM and NB. These are compared under the same test scenario, using crossed validation with 10 subsets to guarantee the dataset was divided randomly, one part for training and one part for test and evaluation. The evaluation metrics implemented were precision, coverage, FPR, TPR and F-Measure. Next, you can see the results obtained by the implemented methods. In Table 2, you can see the results obtained by DT for the activities classification associated with the clusters described in Sect. 4.2. The results were considered successful, with a precision rate of 98.9%, indicating a wrong classification in 8 records only.

Table 2. DT J48 results + crossed validation: 10 folds

Class	TPR	FPR	Precision	Coverage	F-measure
Cluster 0	96.1%	0.6%	96.1%	96.1%	96.1%
Cluster 1	1%	0%	1%	1%	1%
Cluster 2	99.1%	0.9%	98.8%	99.1%	98.9%
Cluster 3	99.5%	0%	1%	99.5%	99.8%
Average	98.9%	0.5%	98.9%	98.9%	98.9%

In Table 3, you can see the results obtained by SVM for the activities classification associated with the clusters described in Sect. 4.2. The results were considered successful, with a precision rate of 99.2%, indicating a wrong classification in 6 records only.

In Table 4, you can see the results obtained by NB for the activities classification associated with the clusters described in Sect. 4.2. The results were considered promising, with a precision rate of 97.6%, indicating a wrong classification in 20 records only.

With these results from the experimentation phase, the best algorithm was SVM with crossed validation 10 subsets, and it was integrated with the Simple K-means method finally. Next, you can see in Fig. 6 the comparison of the algorithms.

Table 3. SVM results + crossed validation: 10 folds

Class	TPR	FPR	Precision	Coverage	F-measure
Cluster 0	1%	0%	1%	1%	1%
Cluster 1	94.9%	0%	1%	94.9%	97.4%
Cluster 2	1%	0.14%	98.2%	1%	98.4%
Cluster 3	1%	0%	1%	1%	1%
Average	99.2%	0.6%	99.2%	99.2%	98.8%

Table 4. NB results + crossed validation: 10 folds

Class	TPR	FPR	Precision	Coverage	F-Measure
Cluster 0	99.0%	0%	1%	99.0%	99.5%
Cluster 1	99.2%	0.29%	86.7%	99.2%	92.5%
Cluster 2	94.4%	0.2%	99.7%	94.4%	96.9%
Cluster 3	1%	0.2%	99.5%	1%	99.8%
Average	97.3%	0.6%	97.6%	97.3%	97.4%

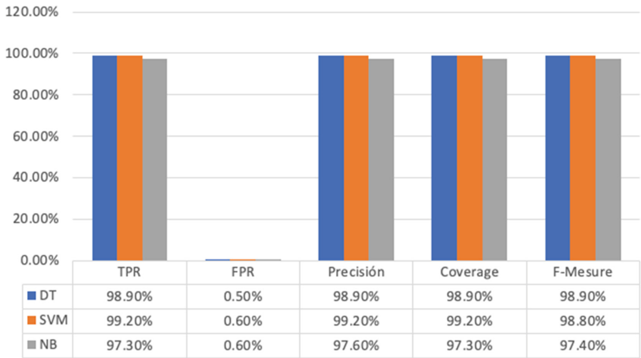

Fig. 6. Comparison results of DT, SVM and NB algorithms

5 Results and Discussion

HAR is a field of study in constant growing, and researchers are working more in the efficient identification of different activities, the technology devices have an important role in the process, since they provide the capture and compilation of the information correctly. The computational intelligence techniques framed inside the disciplines of DM, ML and DL have a substantial place inside HAR, since they all use data to discover patterns that become useful knowledge. In this study, the authors compared the classification methods DT, SVM and NB, during the processes of training and validation of the proposed approach, under the same test scenario. The results obtained showed that, all three methods are successful to recognize the cluster associated with each of the activities contained in the dataset Van Kasteren. The SVM method obtained the best result with TPR (99.2%), FPR (0.6%), Precision (99.2%), Coverage (99.2%) and F-Measure (98.8%), surpassing the results proposed by similar works as [31]. These results clearly show that it is possible to create hybrid systems where traditional classification methods can be integrated with unsupervised methods, corresponding to clustering this time, since the latter permit to segment the activities with certain degree of similarity, obtaining a reduction of the variables, which it is quite useful with datasets with a high number of attributes. Finally, the presented results showed that data mining hybrid methods represent an alternate solution inside HAR, and their findings can be used to support decision making processes.

6 Conclusions and Future Works

The recognition of human activities in daily life is research field with growing interest in recent years, different authors have decided to study monitoring several activities, using sensors connected through a network infrastructure, devices like smartphones and wearables, they also have built many datasets that permit the implementation of data mining methods with the goal of performing classification and detection processes. This study used the dataset Van Kasteren [20], in the same way as many other studies. To analyze the data, it was necessary to perform several procedures, highlighting integration and representation of the data to obtain a single format capable of being processed.

In this research, the classification methods DT, SVM and NB were compared under the same test scenario, during the training and validation processes, the results showed that all three methods are successful to identify the cluster associated to the activities contained in the dataset Van Kasteren. The SVM method obtained the best result with TPR (99.2%), FPR (0.6%), Precision (99.2%), Coverage (99.2%) and F-Measure (98.8%), surpassing the results proposed by similar works as [31].

Acknowledgment. This work was partially supported from the REMIND Project from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 734355.

References

1. Ladjailia, A., Bouchrika, I., Merouani, H.F., Harrati, N., Mahfouf, Z.: Human activity recognition via optical flow: decomposing activities into basic actions. Neural Comput. Appl. **32**(21), 16387–16400 (2019). https://doi.org/10.1007/s00521-018-3951-x
2. Banos, O., Damas, M., Pomares, H., Prieto, A., Rojas, I.: Daily living activity recognition based on statistical feature quality group selection. Expert Syst. Appl. **39**(9), 8013–8021 (2012)
3. Casale, P., Pujol, O., Radeva, P.: Human activity recognition from accelerometer data using a wearable device. In: Vitrià, J., Sanches, J.M., Hernández, M. (eds.) IbPRIA 2011. LNCS, vol. 6669, pp. 289–296. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21257-4_36

4. Calabria-Sarmiento, J.C., et al.: (2018). Software applications to health sector: a systematic review of literature (2018)
5. Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z.: Sensor-based activity recognition. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **42**(6), 790–808 (2012)
6. Arifoglu, D., Bouchachia, A.: Activity recognition and abnormal behaviour detection with recurrent neural networks. Proc. Comput. Sci. **110**, 86–93 (2017)
7. Fergani, B.: Comparing HMM, LDA, SVM and Smote-SVM algorithms in classifying human activities. In: El Oualkadi, A., Choubani, F., El Moussati, A. (eds.) Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015. Lecture Notes in Electrical Engineering, vol. 381, pp. 639–644. Springer, Cham (2016)
8. Paul, P., George, T.: An effective approach for human activity recognition on smartphone. In: 2015 IEEE International Conference on Engineering and Technology (Icetech), pp. 1–3. IEEE, March 2015
9. Dao, M.S., Nguyen-Gia, T.A., Mai, V.C.: Daily human activities recognition using heterogeneous sensors from smartphones. Proc. Comput. Sci. **111**, 323–328 (2017)
10. Liu, Y., Nie, L., Liu, L., Rosenblum, D.S.: From action to activity: sensor-based activity recognition. Neurocomputing **181**, 108–115 (2016)
11. Concone, F., Gaglio, S., Lo Re, G., Morana, M.: Smartphone data analysis for human activity recognition. In: Esposito, F., Basili, R., Ferilli, S., Lisi, F. (eds.) AI*IA 2017. LNCS, vol. 10640, pp. 58–71. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70169-1_5
12. Fahad, L.G., Khan, A., Rajarajan, M.: Activity recognition in smart homes with self verification of assignments. Neurocomputing **149**, 1286–1298 (2015)
13. Manzi, A., Dario, P., Cavallo, F.: A human activity recognition system based on dynamic clustering of skeleton data. Sensors **17**(5), 1100 (2017)
14. Tran, D., Sorokin, A.: Human activity recognition with metric learning. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 548–561. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_42
15. Robertson, N., Reid, I.: A general method for human activity recognition in video. Comput. Vis. Image Underst. **104**(2–3), 232–248 (2006). ISSN 1077-3142. https://doi.org/10.1016/j.cviu.2006.07.006
16. Balli, S., Sağbaş, E.A., Peker, M.: Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. Meas. Control **52**(1–2), 37–45 (2019)
17. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. **59**, 235–244 (2016)
18. Tan, Y.E., Lo, C.C., Shieh, C.S., Miu, D., Horng, M.F.: Adaptive confidence evaluation scheme for periodic activity recognition in smart home environments. In: 2019 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 1077–1081. IEEE, August 2019
19. Tahir, S.F., Fahad, L.G., Kifayat, K.: Key feature identification for recognition of activities performed by a smart-home resident. J. Ambient. Intell. Humaniz. Comput. **11**(5), 2105–2115 (2019). https://doi.org/10.1007/s12652-019-01236-y
20. van Kasteren, T.L.M., Englebienne, G., Kröse, B.J.A.: Activity recognition using semi-Markov models on real world smart home datasets. J. Ambient Intell. Smart Environ. **2**(3), 311–325 (2010). https://doi.org/10.3233/AIS-2010-0070
21. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: Casas: a smart home in a box. Computer **46**(7), 62–69 (2013). https://doi.org/10.1109/MC.2012.328
22. Cook, D.J.: Learning setting-generalized activity mdoels for smart spaces. IEEE Intell. Syst. **99**(1) (2011). https://doi.org/10.1109/MIS.2010.112

23. Singla, G., Cook, D.J., Schmitter-Edgecombe, M.: Recognizing independent and joint activities among multiple residents in smart environments. J. Ambient. Intell. Humaniz. Comput. **1**(1), 57–63 (2010). https://doi.org/10.1007/s12652-009-0007-1
24. Chavarriaga, R., et al.: The Opportunity challenge: a benchmark database for on-body sensor-based activity recognition. Pattern Recogn. Lett. **34**(15), 2033–2042 (2013). https://doi.org/10.1016/j.patrec.2012.12.014
25. Roggen, D., et al.: Collecting complex activity data sets in highly rich networked sensor environments. In: Seventh International Conference on Networked Sensing Systems (2010). https://doi.org/10.1109/INSS.2010.5573462
26. Lukowicz, P., et al.: Recording a complex, multi modal activity data set for context recognition. In: 23th International Conference on Architecture of Computing Systems (2010). http://www.opportunity-project.eu/challengeDataset. Accessed 3 July 2018
27. Anguita, D., et al.: A public domain dataset for Human Activity Recognition using smartphones. In: 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. ESANN 2013, Bruges, Belgium, pp. 437–442, April 2013
28. Ronao, C.A., Cho, S.: Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models. In: 2014 10th International Conference on Natural Computation, pp. 681–686 (2014). https://doi.org/10.1109/ICNC.2014.6975918
29. Banos, O., et al.: mHealthDroid: a novel framework for agile development of mobile health applications. In: Pecchia, L., Chen, L.L., Nugent, C., Bravo, J. (eds.) IWAAL 2014. LNCS, vol. 8868, pp. 91–98. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13105-4_14
30. Banos, O., et al.: Design, implementation and validation of a novel open framework for agile development of mobile health applications. BioMed. Eng. OnLine **14**(S2:S6), 1–20 (2015)
31. Chetty, G., White, M., Akther, F.: Smart phone based data mining for human activity recognition. Proc. Comput. Sci. **46**, 1181–1187 (2015)
32. Keyvanpour, M.R., Zolfaghari, S.: Augmented feature-state sensors in human activity recognition. In: 2017 9th International Conference on Information and Knowledge Technology (IKT), pp. 71–75). IEEE, October 2017
33. Yamada, N., Sakamoto, K., Kunito, G., Yamazaki, K., Tanaka, S.: Human activity recognition based on surrounding things. In: Tomoya Enokido, L., Yan, B.X., Kim, D., Dai, Y., Yang, L.T. (eds.) EUC 2005. LNCS, vol. 3823, pp. 1–10. Springer, Heidelberg (2005). https://doi.org/10.1007/11596042_1
34. Choi, J., Shin, D., Shin, D.: Ubiquitous Intelligent Sensing System for a Smart Home. In: Yeung, D.Y., Kwok, J.T., Fred, A., Roli, F., de Ridder, D. (eds.) SSPR/SPR 2006. LNCS, vol. 4109, pp. 322–330. Springer, Heidelberg (2006). https://doi.org/10.1007/11815921_35
35. Ravi, D., Wong, C., Lo, B., Yang, G.Z.: Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In: 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 71–76. IEEE, June 2016
36. Gilbert, A., Illingworth, J., Bowden, R.: Scale invariant action recognition using compound features mined from dense spatio-temporal corners. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 222–233. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_18
37. Okkonen, M.A., Kellokumpu, V., Pietikäinen, M., Heikkilä, J.: A visual system for hand gesture recognition in human-computer interaction. In: Ersbøll, B.K., Pedersen, K.S. (eds.) SCIA 2007. LNCS, vol. 4522, pp. 709–718. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73040-8_72
38. Ozer, B., Wolf, W., Akansu, A.N.: Human activity detection in MPEG sequences. In: Proceedings Workshop on Human Motion, Austin, Texas, USA, pp. 61–66 (2000). https://doi.org/10.1109/HUMO.2000.897372

39. Zheng, H., Wang, H., Black, N.: Human activity detection in smart home environment with self-adaptive neural networks. In: 2008 IEEE International Conference on Networking, Sensing and Control, Sanya, pp1505–1510 (2008). https://doi.org/10.1109/ICNSC.2008.4525459
40. Kaur, S.: Survey of different data clustering algorithms. Int. J. Comput. Sci. Mob. Comput. **5**(5), 584–588 (2016)
41. Virdi, G., Madan, N.: Review on various enhancements in K means clustering algorithm (2018)
42. Du, W , Lin, H., Sun, J., Yu, B., Yang, H.: A new projection-based K-means initialization algorithm. In: Proceedings of 2016 IEEE Chinese Guidance, Navigation and Control Conference, China (2016)
43. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **7**, 881–892 (2002)
44. Friedl, M.A., Brodley, C.E.: Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. **61**(3), 399–409 (1997)
45. Dai, W., Ji, W.: A mapreduce implementation of C4. 5 decision tree algorithm. Int. J. Database Theory Appl. **7**(1), 49–60 (2014)
46. Lausch, A., Schmidt, A., Tischendorf, L.: Data mining and linked open data—new perspectives for data analysis in environmental research. Ecol. Model. **295**, 5–17 (2015)
47. Daszykowski, M., Korzen, M., Krakowska, B., Fabianczyk, K.: Expert system for monitoring the tributyltin content in inland water samples. Chemom. Intell. Lab. Syst. **149**, 123–131 (2015)
48. Magerman, D.M.: Statistical decision-tree models for parsing. In: Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics, pp. 276–283. Association for Computational Linguistics, June 1995
49. Zhao, Y., Zhang, Y.: Comparison of decision tree methods for finding active objects. Adv. Space Res. **41**(12), 1955–1959 (2008)
50. Suresh, K., Dillibabu, R.: Designing a machine learning based software risk assessment model using Naïve Bayes algorithm. TAGA J. **14**, 3141–3147 (2018)
51. Naik, D.L., Kiran, R.: Naïve Bayes classifier, multivariate linear regression and experimental testing for classification and characterization of wheat straw based on mechanical properties. Ind. Crops Prod. **112**, 434–448 (2018)
52. Picard, R.W., et al.: Affective learning—a manifesto. BT Technol. J. **22**(4), 253–269 (2004)
53. Patil, T.R., Sherekar, S.S.: Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int. J. Comput. Sci. Appl. **6**(2), 256–261 (2013)
54. O'Reilly, K.M.A., Mclaughlin, A.M., Beckett, W.S., Sime, P.J.: Asbestos-related lung disease. Am. Family Phys. **75**(5), 683–688 (2007)
55. James, A., Abu-Mostafa, Y., Qiao, X.: Nowcasting recessions using the SVM machine learning algorithm. Available at SSRN 3316917 (2018)
56. Vapnik, V.: Statistical Learning Theory. Wiley, Hoboken (1998)
57. Papageorgiou, C., Oren, M., Poggio, T.: A general framework for object detection. In: Proceedings of the International Conference on Computer Vision (1998)
58. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 137–142. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026683
59. Kim, Y., Ling, H.: Human activity classification based on micro-Doppler signatures using a support vector machine. IEEE Trans. Geosci. Remote Sens. **47**(5), 1328–1337 (2009)
60. Da Silva, F., Niedermeyer, E.: Electroencephalography: Basic Principles. Clinical Applications, and Related Fields. William & Wikins, Baltimore (1993)

61. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
62. Palechor, F.M., de la Hoz Manotas, A.: Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in brief **25**, 104344 (2019)
63. Ng, A.: Clustering with the k-means algorithm. Mach. Learn. (2012)
64. Kodinariya, T.M., Makwana, P.R.: Review on determining number of cluster in K-means clustering. Int. J. **1**(6), 90–95 (2013)