Journal of Reliable Intelligent Environments (2024) 10:45–54
https://doi.org/10.1007/s40860-022-00199-w

ORIGINAL ARTICLE

Two-stage RFID approach for localizing objects in smart homes based on gradient boosted decision trees with under- and over-sampling

Shadi Abudalfa[1] · Kevin Bouchard[2]

Received: 17 August 2022 / Accepted: 27 December 2022 / Published online: 12 January 2023

Abstract

Developing automated systems with a reasonable cost for long-term care for elders is a promising research direction. Such smart systems are based on realizing activities of daily living (ADLs) to enable aging in place while preserving the quality of life of all inhabitants in smart homes. One of the research directions is based on localizing items used by elders to monitor their activities with fine-grained details of the progress. In this paper, we shed the light on this issue by presenting an approach for localizing items in smart homes. The presented method is based on applying machine learning algorithms to Radio Frequency IDentification (RFID) tags readings. Our approach achieves the required task through two stages. The first stage detects in which room the selected object is located. Then, the second one determines the exact position of the selected object inside the detected room. Additionally, we present an efficient approach based on gradient boosted decision trees for detecting the location of the selected object in a real-world smart home. Moreover, we employ some techniques of over- and under-sampling with data clustering for improving the performance of the presented techniques. Many experiments are conducted in this work to evaluate the performance of the presented approach for localizing objects in a real smart home. The results of the experiments have shown that our approach provides remarkable performance.

Keywords Indoor positioning · RFID · Gradient boosting · Data clustering · Data balancing · Smart home

1 Introduction

There is a persistent demand to develop automated systems that provide suitable support to elders or patients who need long-term care [1]. The state-of-the-art has recently shifted toward implementing such systems in homes due to many circumstances such as dealing with the COVID-19 pandemic. The suggested systems should be implemented with a reasonable cost and preserve the quality of life for all inhabitants in smart homes [2].

Working principle of most suggested systems is based on recognizing human activities of daily living (ADLs) [3]. Recognizing ADLs mainly deals with monitoring exact human activity [4] (such as squat on chair and rotation of the wrist) or monitoring items used by inhabitants in smart homes [5]. Our work mimics the second research direction by detecting position of objects located in smart homes. Therefore, we present a convenient approach for achieving this task with a reasonable cost.

Our work is based on analyzing adequate data collected from Radio Frequency IDentification (RFID) antennas located in a smart home. Despite the progress that has been made so far, RFID-based localization is still imprecise and lacks robustness. Additionally, the RFID datasets available online for machine learning are scarce and too small. Wherefore, we add a contribution toward this direction by analyzing collected datasets with more machine learning methods based on gradient boosted decision trees. We also used 16 different techniques of over- and under-sampling to balance the collected data [6]. As a result of this work, the presented approach provides a competitive performance and can be easily implemented in any home.

To decrease the cost and facilitate the implementation of our approach, we investigated the exploitation of passive

✉ Shadi Abudalfa
sabudalfa@ucas.edu.ps

Kevin Bouchard
kevin_bouchard@uqac.ca

[1] IT Department, University College of Applied Sciences, Gaza, Palestine

[2] LIARA, Université du Québec À Chicoutimi, Saguenay, Canada

RFID tags [7], because they are inexpensive, small, and resistant in comparison with active RFID tags. The efficiency of using passive RFID tags has been evaluated by installing them on an object located in a realistic home environment. Then, the task was converted into a classification problem by labeling the collected readings.

The contributions of the present work toward solving the problem of localizing objects in smart homes are highlighted below:

- We have designed a hybrid technique by combining gradient boosted decision trees with over- and under-sampling. Based on our knowledge, our research is the first work that employs this hybrid technique for localizing objects in smart homes by using passive RFID tags.
- We present a technique that improves performance of object localization by combining over-sampling with data clustering.
- The presented technique can be employed for developing an integrated system that localizes objects in smart homes. This combination makes our approach unique.
- Extensive analysis has been presented to show characteristics of the collected data.
- We have evaluated the presented approach on a collected dataset and reported results through various perspectives.

The remainder of paper is organized as follows: Sect. 2 presents a review for some related studies. Section 3 explains the presented approach. Section 4 describes the experiment environment. Section 5 discusses the experiment results and provides due analysis. Finally, Sect. 6 concludes the paper and reveals some suggestions for future work.

2 Literature review

Our approach is grounded on the goal to make smart homes more informative while maintaining their invasiveness to the minimum. With passive RFID localization and tracking, the ambient system could not only have information that is more expressive about what happens, but also provide information that is more robust/reliable than simply depending on simpler sensors [8] (i.e. PIR, electromagnetic contacts, etc.). It is our belief that this will enable better services for decision making.

Localization in indoor environments [9] has gained popularity with the domain of wireless communication networks [10]. Various technologies in indoor environments have been explored with different applications such as IoT healthcare [11], security [12], and asset management [13]. Such technologies use RFID, Bluetooth and Wi-Fi to connect Internet of Things (IoT) devices in smart homes. Currently, RFID technology promises to revolutionize many fields due to its low-cost and low-power characteristics. However, many challenges still need to be addressed with this technology [14]. Our work sheds the light on tackling some challenges linked to this technology with localizing objects in smart homes.

There are mainly three categories of RFID indoor localization methods: Triangulations, Scene Analysis and Proximity [15]. The triangulation technique is based on estimating distances by using geometrical properties of triangles formed from the received signal strength indication (RSSI) of RFID tags [15]. The scene analysis method, which mimics our work, collects features from a specific scene and then estimates the location of the tagged object by using machine learning techniques. While the proximity method simply relies on the signal range of the antenna by considering the target object as a collocate object on its entire coverage [16].

Based on our exploration of the literature, there are limited works that achieved a working scene analysis approach. Previous related works apply directly classical data mining techniques for detecting the locations of targeted objects. Thereby, there is a limited performance provided by the previous related works. Some of pervious related works are described briefly in the sequel.

Zhang et al. [17] used a probabilistic approach by employing a Bayesian filter for estimating the location of a targeted object. While, Ma et al. [18] used statistical features to perform the classification task. They have evaluated the performance of applying classifiers using logistic regression (LR), support vector machine (SVM) and decision tree (DT) toward this research direction. Similarly, Xu et al. [19] employed a support vector regression algorithm to improve the positioning precision.

In the same context, Bergeron et al. [20] used decision trees to develop an Indoor Positioning System (IPS) for objects of daily life equipped with passive RFID tags. The same research group presented also a method [21] that uses the signal strength indication of RFID antennas with statistical features to perform relative positioning in a smart home.

There are hybrid techniques presented for improving the accuracy of indoor localization. For example, Xu et al. [22] used deep learning enhanced holography to create adaptive localization models for achieving high-positioning accuracy. Shen et al. [23] employed also deep learning for developing a relative localization method of RFID tags via phase and RSSI. On the other side, Shen et al. [24] improved the accuracy of indoor localization by using Spinning Antenna which is a rotating antenna that collects dynamic data for getting more information.

Some of the research works process data streams to support online learning. For example, Zhong and Liu [25] proposed an RFID localization algorithm based on an online learning system. Thus, their system can process data streams that continuously emerge in the environment.

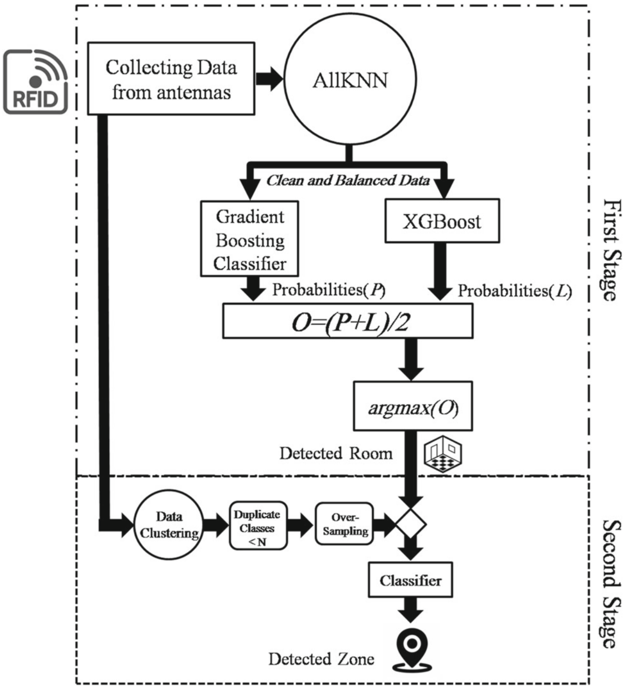


Fig. 1 The presented approach structure

Based on the previous discussion, we conclude that our research work presents additional knowledge for improving the performance of localizing objects in smart homes by using passive RFID technology through a scene analysis approach. Moreover, there are some unknown factors that were neither studied nor evaluated by previous related works. Thereby, our paper fills some gaps and enriches this research direction.

3 Solution approach

Our presented approach consists of two stages. The first stage detects in which room the selected object is located. While, the second one determines the exact position (zone) of the selected object inside the detected room. We present different techniques with each stage for improving performance of classifying the collected data. Figure 1 illustrates the structure of our presented approach which is described in the following paragraphs.

The presented approach is first based on collecting data from RFID antennas located in a smart home. Then, some noisy data are removed with balancing the data by using under-sampling algorithm entitled AllKNN [26]. We use this under-sampling algorithm since it provides competitive results with this phase in comparison with other under-sampling algorithms. The performance of applying AllKNN compared to other popular algorithms is shown in Sect. 5.2. Two classifiers based on gradient boosted decision trees are used to classify the restated data. Specifically, the used classifiers are two versions of implementing eXtreme Gradient Boosting (XGBoost) [27]. Ensemble method is used as well to combine the selected classifiers by averaging prediction probabilities. Our work employs gradient boosted decision trees based on our previous work which supports using this machine learning technique in comparison with others through this research direction.

The second stage of the presented approach is based on applying over-sampling techniques after clustering [28] the collected data. We applied over-sampling instead of using under-sampling since the number of samples collected per each zone is small for training the classifiers. In the same context, applying data clustering should precede over-sampling because the data is originally balanced as described previously. Thereby, applying data clustering will make the data imbalanced and distribute noisy samples into specific groups. Thus, applying over-sampling after data clustering will make the data balanced and increase the number of samples in each class for efficiently training the classifiers.

Our methodology is harmonious with mechanism of over-sampling which adds more samples to minority class to make it equal in size to majority class. It is worth also to mention here, that we experimentally checked effect of immediately applying over-sampling methods to the original data. The experiment results show that there is no improvement with this strategy and warning messages are appeared to clarify that the data is balanced and no need to apply over-sampling.

The data clustering step divides the data into similar groups. Whereas, each cluster contains different classes that include samples collected from some zones. All classes that contain small number of samples (N) are duplicated for increasing the possibility of training the classifier with noisy samples. Then, the over-sampling technique is used since the resulted data will be imbalanced. We use Mean-Shift algorithm [29] and Random Over Sampler [30] for applying data clustering and over-sampling, respectively. We selected Mean-Shift algorithm to avoid specifying the optimal number of clusters.

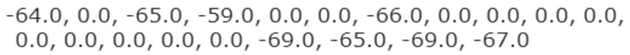
After that, we use decision trees for classifying the resulting data. We used classifiers that are related to decision tree since this technique provides competitive results as reported by Bergeron et al. [20]. The final output of this stage clarifies in which zone the target object is located.

4 Experimental setup

We conducted over 40 experiments to evaluate performance of the presented approach for localizing objects in smart homes. We used Google Colab [31] environment to show the performance of the presented approach. Some experiments are conducted by using GPU provided by Google Colab.

Table 1 Python modules

Tool	Purpose
Pandas	Importing and manipulating the used datasets
Numpy	Applying mathematical operation
Sklearn	Building and evaluating models
Imblearn	Applying over- and under-sampling to the used data
lightgbm	Building and learning LightGBM model
xgboost	Building and learning XGBoost model



```
-64.0, 0.0, -65.0, -59.0, 0.0, 0.0, -66.0, 0.0, 0.0, 0.0, 0.0,
 0.0, 0.0, 0.0, 0.0, 0.0, -69.0, -65.0, -69.0, -67.0
```

Fig. 2 One instance of the collected data

Table 1 shows the Python modules used for conducting our experiment work.

4.1 Datasets

Our experiment work is conducted by using datasets collected by Bergeron et al. [20] affiliated with the DOMUS' smart home at the Université de Sherbrooke. The smart home is divided into six different rooms: one bedroom, an entry hall, a kitchen, a dining room, a living room, and one bathroom. Twenty RFID antennas are used for covering the whole surface of the smart home. Each room in the smart home is divided into zones used to determine the exact location of the targeted object. The kitchen and dining room are divided into zones of 20 cm × 20 cm. While, the bedroom and bathroom (except the counter) are divided into zones of 60 cm × 60 cm. The counter in the bathroom has a higher precision of 30 cm to cover more complex activities such as brushing teeth and shaving. The remainder of the rooms are divided into zones of 75 cm × 75 cm.

The datasets are collected by gluing four RFID tags to an empty rigid plastic bottle of water and gathering readings from the RFID tags after placing the plastic bottle in the center of each zone. Figure 2 shows an instance of the collected data (on two lines for visibility). Whereas, Table 2 shows all datasets used with our experiment work. The datasets are all in the same normalized format and we do have a unified version including all the rooms. The results are presented and compared in Sect. 5.1. Number of samples here is referred to number of all RFID tags readings included in each dataset.

We can clearly notice that the first seven datasets are corresponding to the individual rooms in the smart room except the bathroom is divided into two datasets. Thus, the number of classes contained in these datasets is referred to number of zones. While the last dataset (number 8) contains all first seven datasets whereas the samples are labeled with the room name instead of precise zone. Thereby, this dataset includes seven classes (bathroom datasets are still separated).

4.2 Evaluation measures

Empirical results obtained from experiments provide a good way to evaluate performance of the presented techniques with object localization. Thereby, we use classification accuracy and F1-score [32] for evaluating the presented techniques. The accuracy is the ratio of all samples that are classified correctly. While, the F1-score is the harmonic mean of precision and recall, and its best value is 1 while the worst score is 0.

The Recall (which also known as sensitivity or true positive rate) is the ratio of samples which are classified correctly as positive to all positive samples. While the precision is the ratio of samples which are correctly classified as positive to all samples classified as positive. The F1-score is basically applied to binary classification and there are different types of averaging [33] used for applying multiclass classification such as macro, micro, and weighted. In this work, we use weighted average for studying how the approach performs across overall sets of data. This method calculates metrics for each label and finds their mean weighted by number of true samples for each label.

To evaluating the performance of the presented techniques and models, we used tenfold cross-validation [34] to report logical results. Additionally, we did not use randomize options with cross-validation. Therefore, reproducing this experiment work should provide close results. Specifically, we used cross-validation method entitled *RepeatedStratifiedKFold* to preserve the percentage of each class with each fold. We used same value of random state to make fair comparisons. We also run each experiment only once to decrease the duration of the training phase.

5 Results and analysis

We conducted over 50 experiments to evaluate performance of applying the presented techniques to the used datasets. This section reports our experiment results provided when applying the presented techniques through various perspectives.

5.1 Analyzing the collected data

We analyzed the collected data by evaluating performance of applying 16 methods[1] for balancing the data. Applying these methods show how much the data is noisy and imbalanced. We used LightGBM classifier with domus dataset (number 8

[1] https://imbalanced-learn.org/stable/references/index.html

Table 2 Used datasets

Class type	#	Dataset name (French)	Dataset (English)	Classes #	Samples #
Zones	1	Salle_manger_20cm	Dining_room_20cm	324	16,195
	2	Chambre_coucher_2tuiles	Bedroom_bed_2tiles	33	1650
	3	Salle_bain_60cm_comptoir	Bathroom_60cm_counter	27	1350
	4	Salle_bain_60cm_1	Bathroom_60cm_1	20	1000
	5	Salon_75cm	Living room_75cm	35	1750
	6	Hall_75cm	Lobby_75cm	16	800
	7	Cuisine_20cm	Kitchen_20cm	238	11,900
Rooms	8	Domus	House	7	34,645

Table 3 Performance of applying data cleaning and balancing to whole data with LightGBM classifier

Category	Method	Acc	F1
Raw data	Imbalanced data	95.399	95.405
Under-sampling	ClusterCentroids	81.286	81.265
	CondensedNearestNeighbour	94.909	94.503
	EditedNearestNeighbours	**99.902**	**99.902**
	AllKNN	99.804	99.803
	NearMiss	82.946	82.932
	OneSidedSelection	99.274	99.258
	RandomUnderSampler	84.304	84.241
	TomekLinks	95.411	95.413
Over-sampling	RandomOverSampler	90.714	89.949
	SMOTE	90.748	90.071
	ADASYN	86.845	86.724
	BorderlineSMOTE	88.688	88.671
	KMeansSMOTE	92.403	92.022
	SVMSMOTE	92.303	92.164
Combined	SMOTEENN	**99.384**	**99.386**
	SMOTETomek	93.374	93.052

Bold values indicate that EditedNearestNeighbour is the best with all techniques. While, SMOTEENN is best with combined Category

in Table 2) for showing performance of each selected method. We first applied sampling (over/under/combined) method to the whole data. Then, the resulting data have been classified by using the LightGBM model.

Table 3 shows all results reported by our experiment work in this direction. It is clear that the collected data includes outliers and noisy samples. Thus, cleaning the collected data improves the performance remarkably. The collected data include many noisy samples due to phenomenon of signal interference among antennas. The classification accuracy is improved as well after applying data balancing since the used dataset is imbalanced. These results encouraged us to employ data cleaning and balancing for developing our technique.

We can clearly note that using under-sampling with Edited Nearest Neighbors provides highest performance which cleans the data precisely. SMOTEENN provides a competitive performance as well due to combining SMOTE method for applying over-sampling and Edited Nearest Neighbors method for clearing the data. We applied more classifiers to resulting data to show that these remarkable results are due to specifications of unclean and imbalanced data and not to performance of LightGBM classifier. We used Support Vector Machine (SVM) and two models of gradient boosted decision trees (LightGBM, XGBoost) for this experiment. SVM is used, since it is a standard classifier that many pervious works employed with this research direction. Table 4 clearly shows that the accuracy is sharply increased after applying data cleaning and balancing with each selected classifier.

Another experiment is conducted for checking the need for using two stages with our approach. We evaluated performance of combining the first seven datasets illustrated in

Table 4 Performance of applying data cleaning and balancing with different classifiers

Balancing technique	Classifier	Acc	F1
Imbalanced and dirty raw data	SVM	94.530	94.419
	LightGBM	95.399	95.405
	XGBoost	96.776	96.746
SMOTEENN	SVM	98.544	98.547
	LightGBM	99.344	99.347
	XGBoost	98.398	98.402
EditedNearestNeighbours	SVM	98.887	98.878
	LightGBM	99.902	99.902
	XGBoost	99.688	99.684

Table 5 Performance of classifying a collection of all classes (zones) with data cleaning and balancing

	Classifier	Acc	F1
Raw data	SVM	71.560	72.826
	XGBoost	81.276	80.696
After applying SMOTEENN	SVM	93.406	93.606
	XGBoost	97.349	97.051

Table 6 Performance of applying different classifiers to the raw data

Classifier	Acc	F1
SVM	94.530	94.642
LightGBM	95.399	95.393
XGBoost	**96.761**	**96.790**
KNeighbors	95.235	95.268
Gaussian Naive Bayes	65.972	64.394
Random Forest	95.191	95.196
GradientBoosting	**96.389**	**96.395**
Histogram-based Gradient Boosting	95.708	95.708

Bold values shows the first and second highest values

Table 2. Thereby, the resulting dataset contains 34,645 samples distributed to 351 classes. We classified the resulting dataset by using SVM, LightGBM, and XGBoost. We classified the raw data and compared the results with performance of using SMOTEENN method. Table 5 shows the optimal performance which is in bold.

We can notice clearly that using SMOTEENN improved the performance of classification task. These results were expected since the dataset consists of imbalance classes (major class contains 300 samples while the minor 1 contains 50 samples). It is clear also that XGBoost classifier provides the best results while LightGBM provides the worst performance. Our analysis for these results is based on the characteristics of these classifiers. LightGBM takes a leaf-wise approach with building the tree. While, decision trees with XGBoost were built one level at a time. Thus, XGBoost will be consistent and build a tree with less depth when classifying huge number of classes.

When comparing Tables 4 and 5, we can notice that the performance is degraded with combining all zones in one dataset since the resulted data contains large number of classes (693 classes). This result encouraged us for developing our approach through two stages. Additionally, making two stages will simplify the building of classifiers based on gradient boosted decision trees. Moreover, the time consumed with training models will be decreased sharply.

5.2 Specific performance of the first stage

In this subsection, we show performance of classifying the domus dataset which contains seven classes for predicting in which room the object is located. As stated in previous subsection, removing noisy samples will improve classification accuracy dramatically. However, we should deal with noisy samples to mimic real scenarios with collecting data. In this subsection, we evaluate performance of classifying Domus dataset without applying data cleaning or balancing to testing data. Thus, we applied the presented techniques to only training data.

To select most suitable classifier for developing our technique, we applied many different classifiers to the raw data as shown in Table 6. We used default settings with all classifiers. As shown in the table, XGBoost provides the best performance. Thereby, we used it for developing our technique. We also used another version of Gradient boosted Decision trees (Gradient boosting) since it provides competitive results.

We also evaluated performance of applying under-, over, and combine sampling methods with XGBoost classifier as shown in Table 7. It is clear here that using over-sampling methods is not a good choice. While, using under-sampling and specifically AllKNN is the best strategy to improve performance of classifying the domus dataset. Under-sampling performs well in this direction due to applying data cleaning and removing noisy data.

Table 8 shows performance of the presented technique. The reported results show that the presented technique improves the performance for classifying the domus dataset. It is worth notice here that the reported results may be improved more after tuning all parameters used for building the technique. We did not tune the parameters, because our goal in this work is showing that the presented technique improves the performance in comparison with other methods. While, finding the best performance is our future research direction.

Table 7 Performance of applying data cleaning and balancing to training data with XGBoost classifier

Category	Method	Acc	F1	Category	Method	Acc	F1
Under-sampling	ClusterCentroids	93.030	92.433	Over-Sampling	RandomOverSampler	96.735	96.751
	CondensedNearestNeighbour	89.964	89.513		SMOTE	96.761	96.760
	EditedNearestNeighbours	96.479	96.724		ADASYN	96.181	96.162
	AllKNN	96.929	97.572		BorderlineSMOTE	96.259	96.236
	NearMiss	48.821	44.199		KMeansSMOTE	96.594	96.594
	OneSidedSelection	92.657	93.351		SVMSMOTE	96.608	96.594
	RandomUnderSampler	96.305	96.312	Combined	SMOTEENN	96.733	96.775
	TomekLinks	96.738	96.765		SMOTETomek	96.785	96.797
Raw data	Imbalanced data	96.761	96.790				

Table 8 Performance of the 1st sage

Method	Acc	F1
XGBoost	96.761	96.790
GradientBoosting	96.389	96.395
XGBoost + AllKNN	96.929	97.572
GradientBoosting + AllKNN	97.027	97.521
Ours	**97.137**	**97.718**

Bold values shows that our approach provides the best performance with the 1st stage

We also analyzed the results by showing which classes that perform worse accuracy. Figure 3 shows the average of all confusion matrixes resulted from the tenfold cross-validation. The figure shows clearly that the worst results are provided in the bathroom. Our explanation for this phenomenon is based on locating only two antennas in one corner inside the bathroom. This result sheds the light on dealing with this direction in the future.

5.3 Specific performance of the second stage

In this subsection, we show the performance of predicting the exact position (zone) of the selected object inside the detected room. Table 9 shows performance of the presented technique with the 2nd stage of our approach. We used XGBoostand Random Forest classifiers with this stage since they provide competitive results as stated previously. Additionally, we disabled the randomness with Random Forest classifier to make the comparisons fair enough. As shown in the table, the presented technique improves the performance in comparison with the used classifiers.

The table reports the values used for N parameter. We trained the presented technique by selecting some integer values from the range [3, 11] to set parameter N. Our findings show as well that XGBoost performs better with cuisine_20cm and salle_manger_20cm. While, RandomForest provides better results with the rest of datasets. This result clarifies that XGBoost works better with all dataset that consists large number of classes.

Additionally, it is clear that the under-sampling strategy is not a good choice with this stage as stated previously. Thereby, these results show that our idea is correct for employing over-sampling technique with this stage.

Moreover, it is clear that the overall performance has improved in comparison with previous related work. However, the accuracy of the detected zones in the kitchen and the dining room is still worst. These results are expected since the used datasets includes large number of classes where each class contains a small number of samples. Thereby, current number of samples is not enough to train the classifier in an optimal way. In the same context, using wider zones with these rooms will provide better results as shown in Fig. 4.

5.4 Threats to validity

It is not conclusive that applying the presented techniques to other datasets with different distributions would result in the same classification accuracy. We could not investigate this matter further because there were no other relevant public datasets available.

Additionally, our experiments revealed that results are sensitive to initial values that are used when setting the hyperparameters in models. We used default settings with all used models for making fair comparisons. Moreover, using other methods for applying cross-validation may provide different results.

6 Conclusion and future work

This work presents an approach for localizing objects located in smart homes. The performance of presented technique is improved by employing combination of gradient boosted

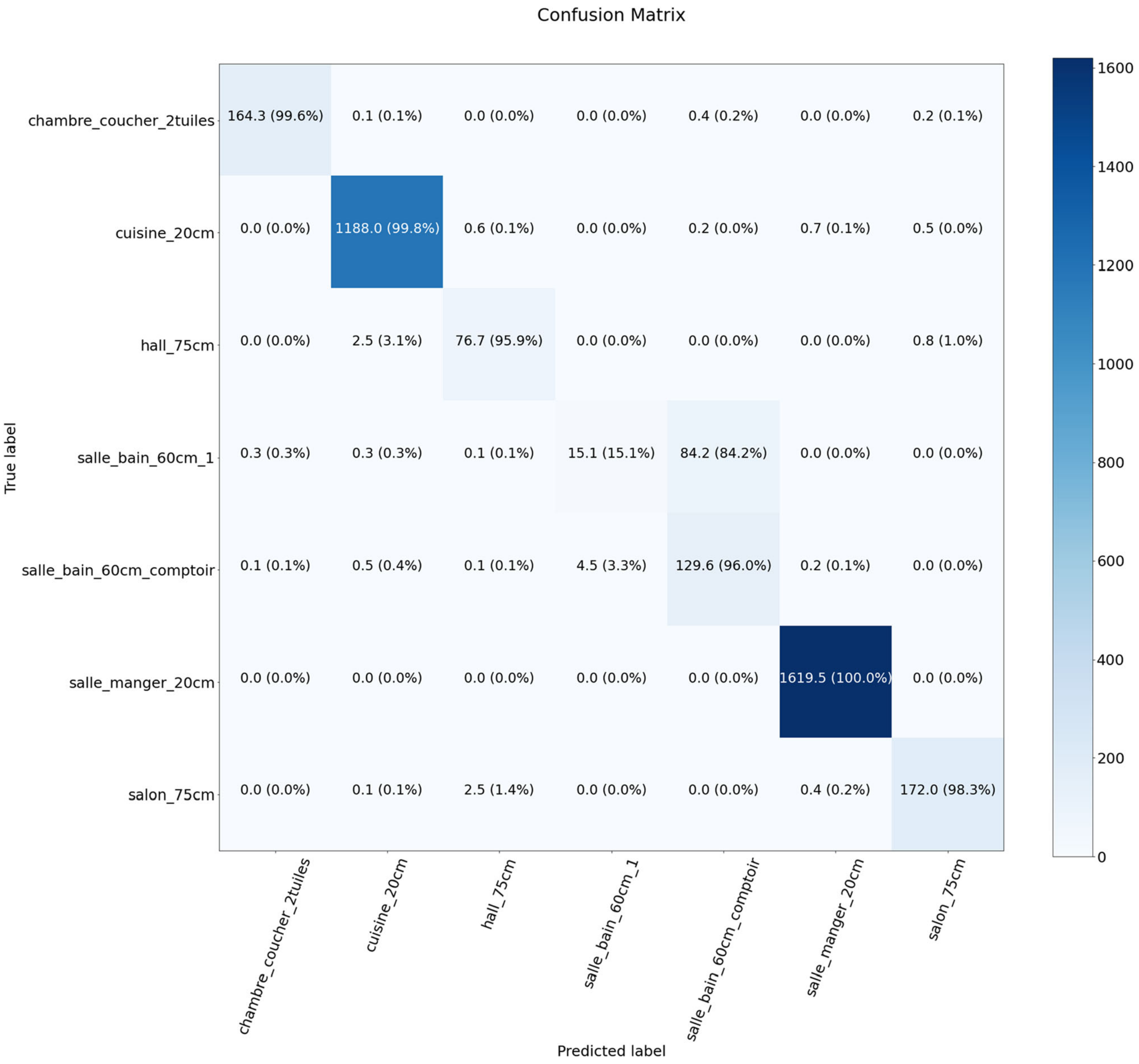


Fig. 3 Analyzing performance of the presented technique toward all classes

decision trees with over- and under-sampling. The data clustering method is used also to improve the classification accuracy. Our approach leads to a more accurate localization for objects located in homes.

Additionally, our work provides a possibility of using data balancing with RFID readings, and suggests that it is possible to perform object localization. It is worth noting here that, our work beats on the related work presented by Bergeron et al. [20] without using windowing and statistical features.

This research direction is still challenging when considering some perspectives such as tackling effect of the interactions appeared between many tagged objects in many rooms. Therefore, this work can be extended in different ways with promising research direction. First, it would be interesting to test performance of applying the presented approach to more datasets. Second, the presented approach could be evaluated by using more advanced unsupervised clustering techniques. Third, the future work maybe also extended by investigating the performance of applying feature engineering methods.

Moreover, it is interested to evaluate more random configurations when initializing the hyperparameters used with the Random Forest models. Using optimization methods for finding optimal values of hyperparameters is also a good direction that may improve the performance of applying the presented approach to object localization.

Table 9 Performance of the 2nd stage

	RandomForest		XGBoost		XGBoost + AllKNN		Ours			
Dataset	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Selected Model	*N*
Chambre_coucher_2tuiles	96.303	96.352	95.697	95.748	94.485	94.562	**96.364**	**96.406**	Random	5
Cuisine_20cm	88.588	88.841	89.714	89.922	85.966	86.301	**89.731**	**89.919**	XGBoost	4
Hall_75cm	97.500	97.572	97.375	97.445	95.750	95.708	**97.875**	**97.929**	Random	11
Salle_bain_60cm_1	94.900	94.914	94.200	94.266	91.600	91.766	**95.400**	**95.473**	Random	6
Salle_bain_60cm_comptoir	96.222	96.247	95.852	95.898	94.444	94.601	**96.741**	**96.776**	Random	7
Salle_manger_20cm	77.221	78.102	78.197	79.335	72.362	73.978	**78.469**	**79.537**	XGBoost	5
Salon_75cm	97.657	97.703	97.429	97.472	96.571	96.606	**98.057**	**98.089**	Random	6

Bold values shows that our approach provides the best performance with the 2nd stage

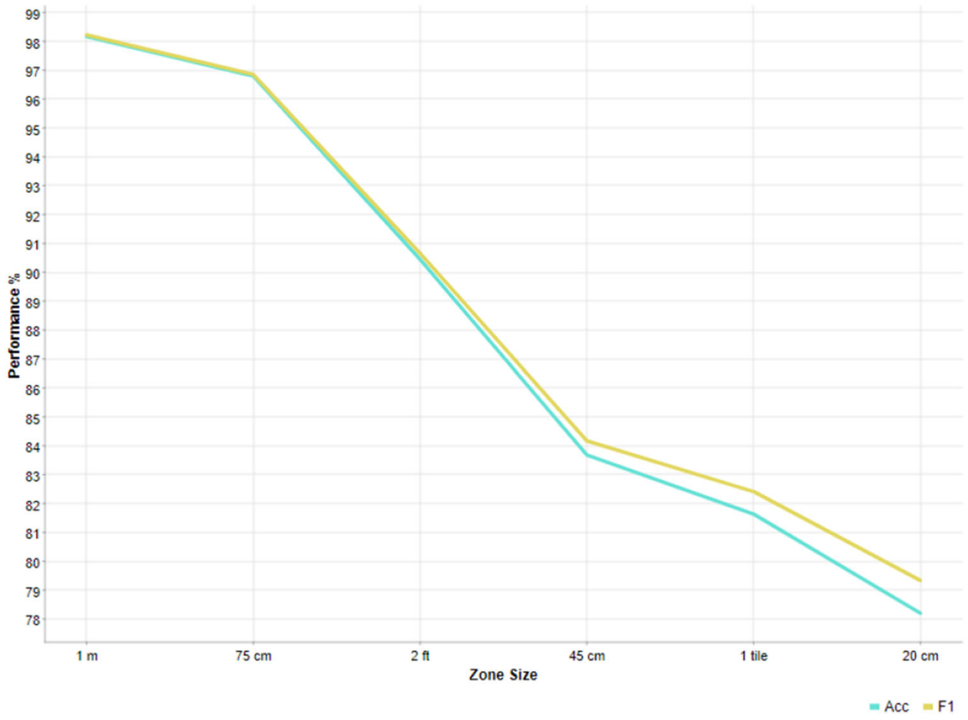


Fig. 4 Performance of applying XGBoost classifier by using zones with large size

Finally, collecting more RFID readings with other locations of antennas may also improve the classification accuracy specifically with semi-supervised learning. Our priority is imperiously based on employing all stated methods for enabling our approach mimics real-time localization.

Acknowledgements This work is conducted through The Palestinian Quebecer Science Bridge (PQSB), which promotes scientific collaboration in research between Quebec, Canada and Palestine through the Palestine Academy for Science and Technology and the Fonds de Recherche du Quebec (FRQ), Canada, and its three branches; the Fonds de recherche du Québec—Santé (FRQS), the Fonds de recherche du Québec—Nature et technologies (FRQNT), and the Fonds de recherche du Québec—Société et culture (FRQSC).

Author contributions SA: conducted the experiments and wrote the paper. KB: proposed the avenue of research and revised the paper.

Funding This work is conducted through The Palestinian Quebecer Science Bridge (PQSB), which promotes scientific collaboration in research between Quebec, Canada and Palestine through the Palestine Academy for Science and Technology and the Fonds de Recherche du Quebec (FRQ), Canada, and its three branches; the Fonds de recherche du Québec—Santé (FRQS), the Fonds de recherche du Québec—Nature et technologies (FRQNT), and the Fonds de recherche du Québec—-Société et culture (FRQSC).

Availability of data and materials Datasets are available on the Website of Kevin Bouchard.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The team used data made available online from previous research. All data were anonymous.

References

1. Klimova B, Valis M, Kuca K (2018) Exploring assistive technology as a potential beneficial intervention tool for people with Alzheimer's disease–a systematic review. Neuropsychiatr Dis Treat 14:3151
2. Naud D, Généreux M, Alauzet A, Bruneau JF, Cohen A, Levasseur M (2021) Social participation and barriers to community activities among middle-aged and older Canadians: differences and similarities according to gender and age. Geriatr Gerontol Int 21(1):77–84
3. Fahad LG, Tahir SF (2021) Activity recognition and anomaly detection in smart homes. Neurocomputing 423:362–372
4. Abudalfa S, Bouchard K (2020) Hybrid deep-readout echo state network and support vector machine with feature selection for human activity recognition. Big data technologies and applications Dec 11. Springer, Cham, pp 150–167
5. Kwon HB, Choi SH, Lee D, Son D, Yoon H, Lee MH, Lee YJ, Park KS (2021) Attention-based LSTM for non-contact sleep stage classification using IRUWB radar. IEEE J Biomed Health Inform 25(10):3844–3853
6. Ghorab AS, Ashour WM, Abudalfa SI (2022) An adaptive oversampling method for imbalanced datasets based on mean-shift and SMOTE. International conference on business and technology. Springer, Cham, pp 13–23
7. Ting SL, Kwok SK, Tsang AH, Ho GT (2011) The study on using passive RFID tags for indoor positioning. Int J Eng Bus Manag. 3:8
8. Polyzos GC, Fotiou N (2015) Building a reliable internet of things using information-centric networking. J Reliab Intell Environ 1:47–58
9. Facchinetti D, Psaila G, Scandurra P (2019) Mobile cloud computing for indoor emergency response: the IPSOS assistant case study. J Reliab Intell Environ 5:173–191
10. Farahsari PS, Farahzadi A, Rezazadeh J, Bagheri A (2022) A survey on indoor positioning systems for IoT-based applications. IEEE Internet Things J 9(10):7680–7699
11. Misra D, Das G, Das D (2020) An IoT based building health monitoring system supported by cloud. J Reliab Intell Environ 6:141–152
12. Panda PK, Chattopadhyay S (2020) A secure mutual authentication protocol for IoT environment. J Reliable Intell Environ 6:79–94
13. Tseng PY, Lin JJ, Chan YC, Chen AY (2022) Real-time indoor localization with visual SLAM for in-building emergency response. Autom Constr 140:104319
14. Oguntala G, Abd-Alhameed R, Jones S, Noras J, Patwary M, Rodriguez J (2018) Indoor location identification technologies for real-time IoT-based applications: an inclusive survey. Comput Sci Rev. 30:55–79
15. Wu C, Wang X, Chen M, Kim MJ (2019) Differential received signal strength based RFID positioning for construction equipment tracking. Adv Eng Inform 42:100960
16. Gu F, Hu X, Ramezani M, Acharya D, Khoshelham K, Valaee S, Shang J (2019) Indoor localization improved by spatial context—a survey. ACM Comput Surv (CSUR). 52(3):1–35
17. Zhang J, Lyu Y, Patton J, Periaswamy SC, Roppel T (2018) BFVP: a probabilistic UHF RFID tag localization algorithm using Bayesian filter and a variable power RFID model. IEEE Trans Ind Electron 65(10):8250–8259
18. Ma H, Wang Y, Wang K (2018) Automatic detection of false positive RFID readings using machine learning algorithms. Expert Syst with Appl. 91:442–451
19. Xu H, Wu M, Li P, Zhu F, Wang R (2018) An RFID indoor positioning algorithm based on support vector regression. Sensors 18(5):1504
20. Bergeron F, Bouchard K, Gaboury S, Giroux S, Bouchard B (2016) Indoor positioning system for smart homes based on decision trees and passive RFID. Pacific-Asia conference on knowledge discovery and data mining Apr 19. Springer, Cham, pp 42–53
21. Bergeron F, Bouchard K, Gaboury S, Giroux S (2021) RFID indoor localization using statistical features. Cybern Syst 52(8):625–641
22. Xu H, Wang D, Zhao R, Zhang Q (2019) AdaRF: adaptive RFID-based indoor localization using deep learning enhanced holography. Proc ACM Interact Mobile Wearable Ubiquitous Technol 3(3):1–22
23. Shen L, Zhang Q, Pang J, Xu H, Li P (2019) PRDL: relative localization method of RFID tags via phase and RSSI based on deep learning. IEEE Access. 7:20249–20261
24. Shen L, Zhang Q, Pang J, Xu H, Li P, Xue D (2019) ANTspin: efficient absolute localization method of RFID tags via spinning antenna. Sensors 19(9):2194
25. Zhong D, Liu F (2020) RF-OSFBLS: an RFID reader-fault-adaptive localization system based on online sequential fuzzy broad learning system. Neurocomputing 21(390):28–39
26. Tomek I (1976) An experiment with the edited nearest-neighbor rule. IEEE Trans Syst Man Cybern 6(6):448–452
27. Chen T, Guestrin C. Xgboost (2016) A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining Aug 13. pp 785–794
28. Abudalfa S, Mikki M (2013) A dynamic linkage clustering using KD-tree. Int Arab J Inf Technol 10(3):283–289
29. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
30. Menardi G, Torelli N (2014) Training and assessing classification rules with imbalanced data. Data Mining Knowl Discov 28(1):92–122
31. Bisong E (2019) Building machine learning and deep learning models on Google cloud platform. Apress, Berkeley, CA
32. Metz CE (1978) Basic principles of ROC analysis. Sem Nucl Med 8(4):283–298
33. Puthiya Parambath S, Usunier N, Grandvalet Y (2014) Optimizing F-measures by cost-sensitive classification. Adv Neural Inf Process Syst 27:2123–2131
34. Cawley GC, Talbot NL (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 11:2079–2107

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.