Impact of Radio Map Size on Indoor Localization Accuracy

Madikana S. Sediela[1(✉)], Moses L. Gadebe[2], and Okuthe P. Kogeda[3]

[1] Department of Computer Science, Faculty of ICT, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa
sedielams@tut.ac.za

[2] Department of Advanced Internet of Things, Council for Scientific and Industrial Research (CSIR), P.O. Box 395, Pretoria 0001, South Africa
MGadebe@csir.co.za

[3] Department of Computer Science and Informatics, Faculty of natural and agricultural sciences, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa
KogedaPO@ufs.ac.za

Abstract. Nowadays Indoor Positioning Systems (IPS) are attracting attention in literature because of Global Positioning System (GPS) challenge to track and navigate indoors. These IPSs intend to provide information about a wireless object's current position indoor. GPS-based localization is the most successful Location-Based Service (LBS) application deployed in an outdoor environment. However, GPS faces a challenge of the line of sight indoor. GPS is affected extensively by the multipath-effects. IPS technologies such as Wi-Fi are deployed for indoor localization, in an attempt to alleviate the GPS indoor challenges. Most IPS employs the Fingerprinting algorithm, whereby a radio-map is generated during the offline phase by collecting measurements of Received Signal Strength Indicator (RSSI) at known locations and, positioning of devices at an unknown location is performed during the online phase by utilizing Machine Learning (ML) Algorithms. The radio-map dataset plays a major role in the accuracy performance of the classifiers deployed during the online phase. RSSI fluctuates indoors because of fading, interferences, and shadowing, therefore, the correction of the radio-map RSSI measurements is mandatory to improve the classifiers performance. In this paper, we looked into the impact of the size of the calibration radio-map on the accuracy of the predictive model. We applied the Mean and Standard Deviation filter on three datasets of different sizes to reduce the RSSI instability at each required point and conducted comparative performance on how ML classification models perform on the three radio-map different in size. The radio-map was generated using our EMPsys application. The results of the simulations show that the accuracy of the Kernel Naïve Bayes significantly improved with filter as the radio-map size increased, from 64.4% with 453 observations in the first scenario to 95.4% with 1054 observations in the third scenario. We, therefore, conclude that the performance of the classifier to be used during the online phase of the fingerprinting algorithm relies on both the size of the radio-

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022, LNCS 13375, pp. 529–543, 2022.
https://doi.org/10.1007/978-3-031-10522-7_36

map and the filtering methods used to correct the RSSI measurements of the radio-map.

Keywords: Fingerprinting · Mean · Standard deviation · RSSI · GPS · Wi-Fi · IPS · LBS

1 Introduction

With a rapid development of a diverse wireless sensor networks, Location-Based Services (LBS) applications have attracted more attention in the research community. LBS provides useful information about a person's or object's current position [1]. LBS positioning technologies can be categorized into an outdoor and indoor system. The GPS technology is the most successful outdoor, however, its applicability is constrained in a closed indoor environment because of non-line-of-sight propagation and multipath effects [2–4]. With GPS challenge indoor, Indoor Positioning Systems (IPS) are explored to close the GPS gap Indoor. The main goal of the IPS is to provide a cost effective, high accuracy and real time solution [5]. IPS utilizes short-ranged technologies such as Wi-Fi [3,6] and Bluetooth [7,8].

These short-ranged technologies are usually deplored with the Fingerprinting Algorithm due to its low deployment cost and convenient signal acquisition [2]. The Fingerprinting Algorithm is divided into two phases, the offline phase for radio-map generation and the online phase for location estimation using the Machine Learning (ML) algorithms. The measurements of the Received signal strength indicator (RSSI) are collected together with the transmitter device's MAC Address and SSID to form a radio-map at known locations during the offline phase. In the online phase, the RSSI scanned at unknown location by the receiver device is matched with the data of the radio-map by utilizing a prediction algorithm.

The success of the model that utilizes the Fingerprinting algorithm is eminently dependent on the radio-map or the dataset collected during the offline phase. Nevertheless, the RSSI measurements are sensitive to environmental settings and without RSSI correction of the radio-map, the localization error of the prediction algorithms in an online phase is too high to achieve the needs of IPS. Continuous attention to correct the RSSI measurements in an effort to improve the performance of the ML model has been explored in literature, without considering the effect of the size of the radio-map on the performance of the ML algorithm. Radio-map generation is a very time consuming process [2], and its size may impact the outcome of models and thus also affect the classification performance [9,10].

In this study, we investigate the impact of the size of radio-map, and correct the RSSI measurements to simulate the effect of the outcomes of the performance of the prediction model. Our RSSI filtering approach presented in [1] uses the mean and standard deviation to identify and correct the RSSI measurements. We discuss our three incremental dataset scenarios resulting in one dataset.

We, therefore, provide comparative performance metrics of Machine learning classification algorithms and analyse the performance of the models as the size in radio-map increases. Our results show that the accuracy of the Kernel Naïve Bayes significantly improved with filter as the radio-map size increases, from 64.4% with 453 observation in the first scenario to 95.4% with 1054 observations in the third scenario. We therefore conclude that the performance of the classifier to be used during the online phase of the fingerprinting algorithm relies on both the size of the radio-map and the filtering method used to correct the RSSI measurements of the radio-map.

The remainder of this paper is structured as follows: In Sect. 2, we present related work. In Sect. 3, we present the methodology followed to improve indoor location estimation. In Sect. 4, we present testing and results. Finally, we provide conclusion and future work in Sect. 5.

2 Related Work

Several studies in literature have continuously attempted to address the issue of unstable RSSI measurements of the radio-map during the offline phase of the fingerprinting algorithm, to improve the performance of the ML algorithms deployed during the online phase. A little work was done to evaluate both the effect of the size of the radio map and corrected RSSI measurements on indoor localization model accuracy.

The authors in [11], examined the ability of the predictive model to induce based on different untrained dataset. The study utilized MATLAB to evaluate if the dataset size has an effect on the accuracy of the predictive model. The study discovered that the training dataset size should be good enough to represent the entire dataset. The author utilized three Neural Networks and the simulation results showed that the learning model produced the most accurate and stable results with the largest dataset.

Similarly, the authors in [9], examined the effect of medical domain dataset size on the performance of supervised Machine Learning algorithms. The authors used three datasets of different sizes to analyze the models performance with respect to accuracy, precision, recall, f-score, specificity and under the ROC curve (AUC). The study concluded that the success of a classifier depends on the dataset representation of the original observations rather than dataset size. Both studies in [9,11] evaluated the impact of the dataset size, however, the studies were not focusing on the hash indoor localization radio-map RSSI values. Our studies evaluate the impact of both radio-map size and the RSSI filtering mechanism.

In recent years, the Bluetooth and Wi-Fi localisation technologies gained a boost because of the evolution of smartphone devices with in-built sensors [12]. The authors in [13], developed a system for locating wireless nodes in a home environment requiring a single access point based on Wi-Fi. Wi-Fi requires that the area in which the devices are used have Access point. Received signal strength indicator was used as a sensor reading (RSSI). Bayesian filtering was used to compute location estimates based on sample sets derived by Monte Carlo sampling. The model proposed in [13] has similarities to the model proposed in [14]. The author in [14] exploited sensor fusion approach by combining the

landmarks, Wi-Fi and Pedestrian Dead Reckoning (PDR). Their system runs fully on the smartphone in an effort to cut costs, then used the Kalman Filter to denoise the RSSI measurements. The landmark approach was introduced to overcome the error of initial positioning, which in turn improves the accuracy when predicting the walking directions. Nevertheless, the RSSI training dataset of the Fingerprinting algorithm remains vulnerable because of changes in the environmental conditions.

More research are rather focusing on dealing with the RSSI challenges because of the environmental conditions. Up to a certain extent, acceptable results have been achieved. The authors in [15] proposed a Wi-Fi integrated Device-free localisation (Dfl) system for intrusion detection of the human body. The system uses RSSI patterns generated when a person in motion passes through the system. All the generated RSSI measurements are therefore filtered using the Alpha Trim Mean Filter, which considers extreme high and extreme low RSSI as noise or outliers. The study in [16], has a different view because according to the authors, extremely high RSSI values represent almost the real value of RSSI.

The above mentioned studies have one thing in common, they are all attempting to close the gap that comes with the GPS-based localisation indoor. In attempting to close the gap, more attention was given to correction of measurements by filtering the RSSI errors. In this study, we focus on both the impact of the size of radio-map and corrected the RSSI measurements of the radio-map.

3 Methodology

In this section, we investigate the impact of the size of radio-map generated during the offline phase of the fingerprinting algorithm on the classification performance. In order to achieve this goal, we developed a system called EMPsys that was implemented following the fingerprinting algorithm. The EMPsys was developed using Java programming language on Android platform using Android Studio Integrated Development Environment (IDE) and XML for interface design. The EMPsys was used during experimentation to generate the radio-map during the offline phase. Three radio-maps of different sizes were used to examine the impact of increasing the size of radio-map on the performance of Machine learning algorithms used for location determination. Each radio map was filtered using the mean and standard deviation to evaluate both the radio-map size and filter impact. The performance of the classifiers is evaluated with respect to accuracy, precision and recall. In the following sub-sections, we discuss the system architecture and classification models.

3.1 System Architecture

The design of EMPsys is based on the fingerprinting algorithm. It is divided into two phases, the offline phase and the online phase. During offline phase the radio-map is generated by collection RSSI measurements together with the transmitter device's SSID and MAC Address at known locations and store the data in a real-time Firebase database. The radio-map is then filtered for outliers at each required point, by using the mean and standard deviation filtering approach.

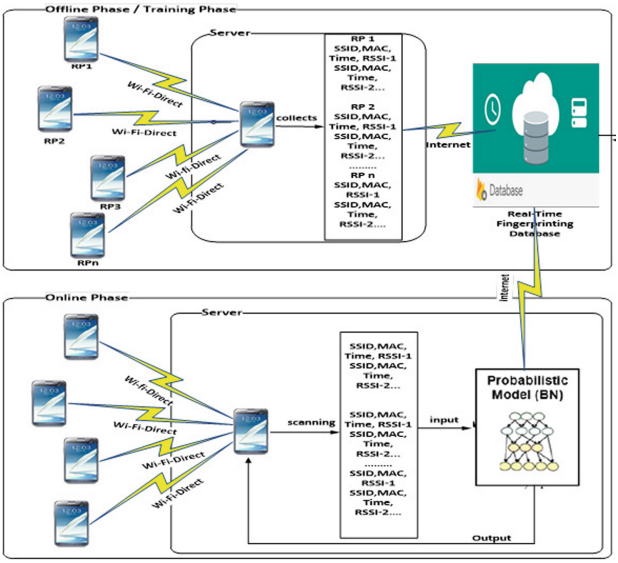

Fig. 1. System architecture [1]

All RSSI values found to be outliers are replaced with the mean value. These radio-map plays a major role during the online phase, because it is used for matching purposes by the prediction model. Machine learning algorithms such as the KNN, Naïve Bayes and decision trees are used during the online phase to localize transmitter devices at unknown locations. The system architecture is presented in Fig. 1.

3.2 Radio Map Generation

We therefore focus on the offline phase collection of the radio-map and correction of the RSSI measurements, and then evaluate the effect of its size on the classification algorithm to be utilized during the online phase as the radio-map size increases. In Fig. 2, Fig. 3, Fig. 4 and Fig. 5, we present the screenshots of the EMPsys application that is used to generate the radio-maps used in this study. We used the offline phase of the EMPsys as represented in the navigation bar of the application in Fig. 2. When the user clicks on the offline tap, the calibration interface is activated as presented in Fig. 3.

The user would therefore be required to click on Off button to start the calibration phase. During this process, the application checks if the receiver device Wi-Fi is enabled or not, if not, the application uses the WiFiManager library to switch the device Wi-Fi on. Then the calibration phase begins with the user prompted to specify the location of the transmitter device at known location to start gathering the radio-map dataset as illustrated in Fig. 4. The application therefore starts scanning for the RSSI, SSID and MAC Address of

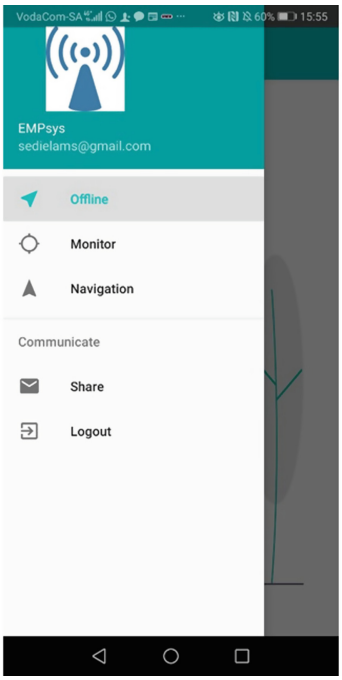

Fig. 2. Offline phase



Fig. 3. Calibration interface

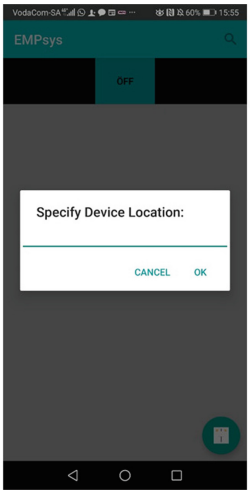

Fig. 4. Start radio-map generation



Fig. 5. Gathering radio-map dataset

the device that is broadcasting its signal at known location as presented in Fig. 4. The application keeps on scanning if it senses change in signal strength, otherwise, the user also has control to re-scan using the floating button.

3.3 Classification Algorithms

The radio-map is used by the location estimation algorithms such as the KNN [29], Decision Tree [30] and Naïve Bayes [31–33]. We looked into how these models localise and made comparisons.

The **K-Nearest Neighbourhood (KNN)** is used to perform the classification by determining the distance between nodes [17]. When using the KNN algorithm, the coordinates of a location from a measured transmitters are acquired by averaging the values of the Euclidean distance of K required points [18]. A given weight determined by the difference between the true measurements and positioning points. When the average value of the Euclidean distance is calculated, the fingerprinting K data are weighted.
The KNN classifier follows the following steps:

1. Initialise the value of K,
2. Compute the distance the measured location and the true location,
3. Sort the distances,
4. Choose the top K nearest neighbors,
5. Apply simple majority and,
6. Estimate location with more neighbors from measured location.
7. The location estimate is returned by Eq. (1):

$$L(x, y) = \frac{\sum_{m=1}^{k} d_m}{\sum_{n=1}^{k} d_n}(x_m, y_m) \tag{1}$$

where:

- *L(x,y)* is the coordinates of the measured location,
- (x_m, y_m) is the coordinates of the m-th required point and,
- d_m is the Euclidean distance between the RSSI readings at location *(x, y)* and (x_m, y_m).

The Euclidean distance d_m is given by Eq. (2):

$$d_m = \sum_{m=1}^{k} \sqrt{(RSSI_{(x,y)} - RSSI_{(x_m,y_m)})^2} \tag{2}$$

A **decision tree** is a non-parametric supervised learning method for classification and regression algorithms expressed as a recursive partition of the instance space [19]. It splits dataset into smaller subsets, while at the same time an associated decision tree is incrementally developed, and the results is a tree with decision notes [20]. For instance, in localisation, The RSSI and orientation are used in the internal node conditions, and the locations or required points

are used in the leaf nodes [21]. Classification of data collected is performed by navigating the tree from the root down to the leaf, based on the outcome of the conditions [22]. The main goal of the decision trees is to deduce the suitable decision tree by reducing the error.

The **Bayesian Network** uses a probabilistic approach to localisation. It utilises the conditional probability and Bayesian inference to estimate the location of a device at unknown location. The Bayesian algorithms such as the Naïve Bayes classifier models the statistical distribution of signal strengths collected at each required point during the offline or training phase of the fingerprinting algorithm as a histogram. Then use it as a prior in a Bayesian framework, and use it to determine the probability of having a specific histogram of signal strengths at a new location using a Naïve Bayes algorithm [23]. For location estimation, the Naïve Bayes is given by Eq. (3):

$$p(L_i|RSSI_i) = \frac{p(L_i).p(RSSI_i|L_i)}{\sum_{i=1}^{n} p(L_n).p(RSSI_i|L_n)} \tag{3}$$

where:

- L_i is the location to be estimated,
- $RSSI_i$ is the recent received RSSI at an unknown location i,
- $p(L_i)$ is the prior probability of location i,
- $p(RSSI_i \mid L_i)$ is the likelihood function and,
- $p(L_i \mid RSSI_i)$ is the posterior function or conditional probability.

Using Eq. (3), the conditional probability at L_i computed for fingerprints and the location corresponding to the maximum likelihood is chosen as the best matching location of the device at unknown location.

In Table 1, we provide a comparison of these location estimation models.

Table 1. Comparison of machine learning algorithms

Algorithm	Advantages	Limitations
KNN	No Training Period (It does nor learn anything in the training period).	Does not work well with large dataset and sensitive to noisy data, sing values and outliers [24]
Decision Tree	Requires less effort for data preparation and produce accurate result [25]. Missing values do not affect the process of building a decision tree [26]	A small change in the data can cause large change in the structure of the decision tree causing instability. Can be computationally expensive.
Naïve Bayes	It requires short computational time for training. It improves the classification time by removing the irrelevant features [27]	Requires very large number of records to obtain good results.

4 Experimentation and Results

The datasets or radio-maps used in this study are of different sizes. The datasets increase in each EMPsys data collection experimentation and, simulations are performed in MATLAB. The classification models are constructed based on the size of the radio-maps generated in each scenario generated during the offline phase of the Fingerprinting algorithm, with known target outputs. The data collection experiments were performed in an indoor environment with a dimension of 9 m × 10 m. The building has five rooms named Room 1, Room 2, Room 3, Room 4 and Room 5, which are our points of data collection during the offline phase and the target outputs during the online phase. The experimental environment is presented in Fig. 6.

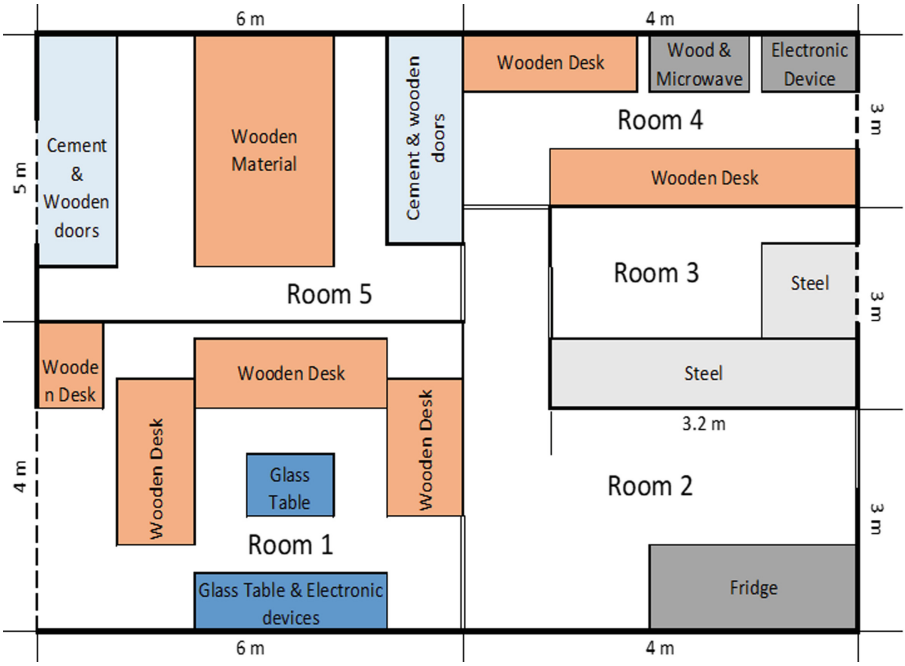

Fig. 6. The experimental environment

Scenario 1: In the first data collection experiment, 15 points were selected at each room and six measurements from each point were collected from the transmitter device used for data collection. A total of 90 measurements were taken from each room. The radio-map size for the first experiment was 453 observations. A sample of the measurements collected from room 1 is shown in Table 2.

Table 2. Comparison of machine learning algorithms

ID	SSID	MacAddress	RSSI	RequirePoint
71	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−56	Room 1
72	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−46	Room 1
73	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−45	Room 1
74	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−44	Room 1
75	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−44	Room 1
76	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−46	Room 1
77	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−46	Room 1
78	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−45	Room 1
79	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−44	Room 1
80	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−44	Room 1
81	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−50	Room 1
82	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−48	Room 1
83	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−48	Room 1
84	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−48	Room 1
85	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−48	Room 1
86	vodafoneMobileWifi-1C6633	33:37:8b:c1:1c:66	−35	Room 1

We therefore, performed the simulation in MATLAB to determine the impact of the radio-map size generated in the first scenario, and the effect of filter method (Table 3).

Table 3. Scenario 1 simulations results

Algorithm	Accuracy		
	Without filter	With Filter	% change
Gaussian Naïve Bayes	37.8 %	64.4 %	70.4%
Kernel Naïve Bayes	44 %	68.9 %	56.6%

The simulation result of the first scenario with dataset of size 453 observations shows that the accuracy result of the Gaussian and Kernel have both improved with filter feature. The Kernel Naïve Bayes classifier out-performed the Gaussian Naïve Bayes classifier with 68.9%. However, at this stage we observed that the percentage change was high for the Gaussian Naïve Bayes with 70.4%, which shows that the filter was more effective for the Gaussian Nave Bayes classifier.

Scenario 2: In the second data collection experiment, the number of points remained at 15 at each room, however, the measurements taken at each point has increased by 66.67% from 6 measurements at a point in the first scenario to

additional 4 measurement to make 10 measurements per point. A total of 150 measurements were taken from each room. Therefore, the size of the radio-map was 753 observations (Table 4).

Table 4. Scenario 2 simulations results

Algorithm	Accuracy		
	Without filter	With Filter	% change
Gaussian Naïve Bayes	50.7 %	70.7 %	39.4%
Kernel Naïve Bayes	53.3 %	76.0 %	42.6%

At the second scenario results with a radio-map of size 753 observations, we observed that both percentage change and the model accuracy favoured the Kernel Naïve Bayes Classifier with 42.6% percentage increase and 76.0% accuracy with filter. The model accuracy also improved in accuracy from 68.9% to 76.0% with the Kernel Naïve Bayes Classifier from the radio-map size 453 to 753 observations. At this stage, we observed that both the radio-map size and the filter have an impact on the predictive model accuracy.

Scenario 3: In the third data collection experiment, the number of points remained unchanged. The number of measurements collected from each point increased from 10 measurements in the second scenario to 14 measurements with a percentage increase of 40%. At this point of the experimentation, 211 measurements were taken from each room, to give a radio-map with 1054 observations size (Table 5).

Table 5. Scenario 3 simulations results

Algorithm	Accuracy		
	Without filter	With Filter	% change
Gaussian Naïve Bayse	45.7 %	79.0 %	72.9%
Kernel Naïve Bayse	49.5 %	95.2 %	92.3%

During the third and final simulation results with a radio-map of size 1054 observations, we observed that the Kernel Naïve Bayes outperforms the Gaussian Naïve Bayes with an accuracy of 95.2% to 72.9%. The accuracy of the Kernel Naïve Bayes with filter also increased from scenario 1, 2 and 3 from 68.9%, 76.0% and 95.2%, which shows that the predictive model depends on both the radio-map size and the filtering mechanism, because the results without a filter are too low in all three scenarios with different radio-maps to achieve the goal of IPS that its model should be accurate.

We further applied the radio-map generated in scenario 3 on different filtering methods and different classifiers in Fig. 7.

The classifiers used in Fig. 7, exhibited different reaction on radio-map of size 1054 observation based on different filters used. The result shows that the accuracy of a localisation model is affected by both the radio-map size and the filtering mechanism. Our proposed filter and the Median (+-3) filter mechanisms performed well with the radio-map of size 1054 observations. It is notable that our proposed method with the kernel Naïve Bayes and the Fine Decision Tree are most effective. Lastly, the moving average filter and exponential smoothing results were not effective in correction of RSSI measurements and thus resulted in poor performance on the predictive models with the radio-map of size 1054 observations.

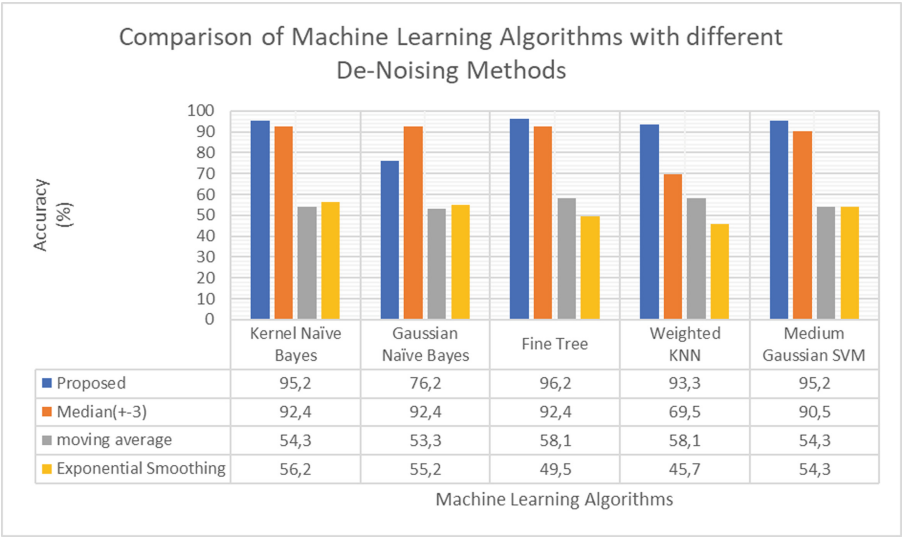

Fig. 7. Effect of different filters using radio-map size 1054 on ML accuracy

Finally, we made comparison with Dataset generated by another study in Table 6.

Table 6. Comparison with dataset generated by another study.

	Our study dataset	Dataset used in [28]
Radio-Map size	1054	56099
Kernel Naïve Bayes Accuracy with Filter	95.4%	94.5%
McConville et.al [28]	N/A	92%

Interestingly, we made a comparison between our study with a radio-map of 1054 observations and the radio-map generated by a study in [28], and we

observed a percentage decrease 0.9% in the Kernel Naïve Bayes Accuracy with Filter, which indicates that there is a very low difference in the model's accuracy between the two datasets. We also observe the Kernel Naïve Bayes Accuracy with Filter outperformed the model used in [28], with 94.5% to 92% respectively. We, therefore, conclude that up to a certain extent that both the size of the radio map and the filter method have an effect on the accuracy of the classification.

5 Conclusion

In recent years, indoor positioning systems have attracted attention, with an effort to close the GPS gap indoor and localize objects or human beings holding devices. The Fingerprinting algorithm is being used because of its cost-effective mechanism. Nevertheless, the Fingerprinting algorithms utilized ML algorithms during its online phase to localize and these ML models performance eminently depend on the radio-map. In this study, we investigated both the impact of the size of the radio-map and the effect of the RSSI filtering method on the performance on classification models. We developed a system called the EMPsys that we utilized to generate the radio-maps used to evaluate the effect on classification algorithms. The data scenarios were divided into three radio-maps of different sizes. We observed that the accuracy of the Kernel Naïve Bayes significantly improved with filter as the radio-map size increases, from 64.4% with 453 observation in the first scenario to 95.4% with 1054 observations in the third scenario. We, therefore, conclude that the performance of the classifier to be used during the online phase of the fingerprinting algorithm relies on both the size of the radio-map and the filtering method used to correct the RSSI measurements of the radio-map.

In future, we seek to investigate what constitute radio-map to be too small or too large in order for a classifier to maximize its performance.

Acknowledgements. The authors would like to thank Tshwane University of Technology and University of the Free State for financial support. The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Sediela, M.S., Gadebe, M.L., Kogeda, O.P.: Indoor localization with filtered and corrected calibration RSSI. In: Zitouni, R., Phokeer, A., Chavula, J., Elmokashfi, A., Gueye, A., Benamar, N. (eds.) AFRICOMM 2020. LNICST, vol. 361, pp. 59–73. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-70572-5_4
2. Qiao, W., Kang, X., Li, M.: An improved XGBoost indoor localization algorithm. DEStech Trans. Comput. Sci. Eng. cisnr (2020)
3. Owuor, D.L., Kogeda, O.P., Agbinya, J.I.: Three tier indoor localization system for digital forensics. Int. J. Electr. Comput. Energetic Electron. Commun. Eng. **11**(6), 602–610 (2017)

4. Khan, M., Kai, Y.D., Gul, H.U.: Indoor Wi-Fi positioning algorithm based on combination of Location Fingerprint and Unscented Kalman Filter. In: 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 693–698. IEEE (2017)
5. Li, S., Rashidzadeh, R.: Hybrid indoor location positioning system. IET Wirel. Sensor Syst. **9**(5), 257–264 (2019)
6. Zhou, Z., Yang, Z., Wu, C., Sun, W., Liu, Y.: LiFi: Line-of-sight identification with WiFi. In: 2014 Proceedings IEEE INFOCOM, pp. 2688–2696. IEEE (2014)
7. Oksar, I.: A Bluetooth signal strength based indoor localization method. In: 2014 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 251–254. IEEE (2014)
8. Huh, J.-H., Seo, K.: An indoor location-based control system using bluetooth beacons for IoT systems. Sensors **17**(12), 2917 (2017)
9. Althnian, A., et al.: Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Appl. Sci. **11**(2), 796 (2021)
10. Rácz, A., Bajusz, D., Héberger, K.: Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. Molecules **26**(4), 1111 (2021)
11. Ajiboye, A., Abdullah-Arshah, R., Hongwu, Q.: Evaluating the effect of dataset size on predictive model using supervised learning technique (2015)
12. Roy, P., Chowdhury, C.: A survey of machine learning techniques for indoor localization and navigation systems. J. Intell. Robot. Syst. **101**(3), 1–34 (2021). https://doi.org/10.1007/s10846-021-01327-z
13. Zàruba, G.V., Huber, M., Kamangar, F., Chlamtac, I.: Indoor location tracking using RSSI readings from a single Wi-Fi access point. Wireless Netw. **13**(2), 221–235 (2007)
14. Chen, Z., Zou, H., Jiang, H., Zhu, Q., Soh, Y.C., Xie, L.: Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization. Sensors **15**(1), 715–732 (2015)
15. Pirzada, N., Nayan, M.Y., Hassan, M.F., Subhan, F.: Multipath fading in device-free indoor localization system: measurements and interpretation. Mehran Univ. Res. J. Eng. Technol. **34**, no. S1 (2015)
16. Xue, W., Qiu, W., Hua, X., Yu, K.: Improved Wi-Fi RSSI measurement for indoor localization. IEEE Sens. J. **17**(7), 2224–2230 (2017)
17. Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064), vol. 2, pp. 775–784. IEEE (2000)
18. Zhu, X.: Indoor localization based on optimized KNN. Network Commun. Technol. **5**(2), 1–34 (2021)
19. Cleophas, T.J., Zwinderman, A.H.: Decision Trees. In: Machine Learning in Medicine, pp. 137–150. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-7869-6_14
20. Seçkin, A.Ç., Coçkun, A.: Hierarchical fusion of machine learning algorithms in indoor positioning and localization. Appl. Sci. **9**(18), 3665 (2019)
21. Sánchez-Rodríguez, D., Hernández-Morera, P., Quinteiro, J.M., Alonso-González, I.: A low complexity system based on multiple weighted decision trees for indoor localization. Sensors **15**(6), 14809–14829 (2015)
22. Maimon, O., Rokach, L.: Data mining and knowledge discovery handbook (2005)
23. Mirowski, P., Milioris, D., Whiting, P., Ho, T.K.: Probabilistic radio-frequency fingerprinting and localization on the run. Bell Labs Tech. J. **18**(4), 111–133 (2014)

24. Mohamed, A.E.: Comparative study of four supervised machine learning techniques for classification. Inf. J. Appl. Sci. Technol. **7**(2) (2017)
25. Archana, S., Elangovan, K.: Survey of classification techniques in data mining. Int. J. Comput. Sci. Mob. Appl. **2**(2), 65–71 (2014)
26. Dhiraj, K.: Top 5 advantages and disadvantages of Decision Tree Algorithm. ed (2020)
27. Jadhav, S.D., Channe, H.: Comparative study of K-NN, naive Bayes and decision tree classification techniques. Int. J. Sci. Res. (IJSR) **5**(1), 1842–1845 (2016)
28. McConville, R., Byrne, D., Craddock, I., Piechocki, R., Pope, J., Santos-Rodriguez, R.: Understanding the quality of calibrations for indoor localisation. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), pp. 676–681. IEEE (2018)
29. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K.: KNN model-based approach in classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds.) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. LNCS, vol. 2888. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-39964-3-62
30. Brijain, R., Patel, R., Kushik, M.R., Rana. K.: A survey on decision tree algorithm for classification (2014)
31. Huang, Y., Li, L.: Naive Bayes classification algorithm based on small sample set. In: IEEE International Conference on Cloud Computing and Intelligence Systems 2011, pp. 34–39 (2011). https://doi.org/10.1109/CCIS.2011.6045027
32. Ssengonzi, C., Kogeda, O.P., Olwal, T.O.: A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization. Elsevier: Array **14**, 100142 (2022). https://doi.org/10.1016/j.array.2022.100142. https://www.sciencedirect.com/science/article/pii/S2590005622000133. ISSN 2590-0056
33. Gadebe, M.L., Kogeda, O.P., Ojo, S.: A Smartphone Naïve Bayes Human Activity Recognition using Personalized dataset. J. Adv. Comput. Intell. Intell. Inf. **24**(5), 685–702 (2020). https://doi.org/10.20965/jaciii.2020.p0685. https://www.fujipress.jp/jaciii/jc/jacii002400050685/. ISBN: 1343-0130/1883-8014