

Outlier Detection in Indoor Localization and Internet of Things (IoT) using Machine Learning

Mansoor Ahmed Bhatti, Rabia Riaz, Sanam Shahla Rizvi, Sana Shokat, Farina Riaz, and Se Jin Kwon

Abstract: In Internet of things (IoT) millions of devices are intelligently connected for providing smart services. Especially in indoor localization environment, that is one of the most concerning topic of smart cities, internet of things and wireless sensor networks. Many technologies are being used for localization purpose in indoor environment and Wi-Fi using received signal strengths (RSSs) is one of them. Wi-Fi RSSs are sensitive to reflection, refraction, interference and channel noise that cause irregularity in signal strengths. The irregular and anomalous RSS values, used in a Wi-Fi indoor localization environment, cannot define the location of any unknown node correctly. Therefore, this research has developed an outlier detection technique named as *iF_Ensemble* for Wi-Fi indoor localization environment by analyzing RSSs using the combination of supervised, unsupervised and ensemble machine learning methods. In this research isolation forest (iForest) is used as an unsupervised learning method. Supervised learning method includes support vector machine (SVM), K-nearest neighbor (KNN) and random forest (RF) classifiers with stacking that is an ensemble learning method. For the evaluation purpose accuracy, precision, recall, F-score and ROC-AUC curve are used. The evaluation of used machine learning method provides high accuracy of 97.8 percent with proposed outlier detection methods and almost 2 percent improvement in the accuracy of localization process in indoor environment after eliminating outliers.

Index Terms: Internet of things, localization, outliers, outliers detection.

I. INTRODUCTION

INDOOR localization is a process through which nodes of a network obtain their location in indoor environments such as smart homes, using smart devices. Localization, sometimes referred as positioning, self-organizing and geo location in the literature, is defined as the method of finding the specific location of unidentified nodes [1]. A GPS based method has been used

Manuscript received December 19, 2019; revised June 1, 2020; approved for publication by Linglong Dai, Guest Editor, June 3, 2020.

This work was supported by Basic Science Research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A3B04031440).

M. A. Bhatti, R. Riaz, and S. Shokat are with the University of Azad Jammu and Kashmir, Muzaffarabad, 13100, Pakistan, email: mansoor.ab6@gmail.com, rabia.riaz@ajku.edu.pk, and snagul@yahoo.com.

S. S. Rizvi is with Raptor Interactive (Pvt) Ltd, Eco Boulevard, Witch Hazel Ave, Centurion 0157, South Africa, email: sanam.shahla@raptorinteractive.com.

F. Riaz is an independent researcher, email: farina.qau@gmail.com.

S. J. Kwon is with Dept. of Computer Engineering, Kangwon National University, 346 Joongang-ro, Samcheok-si, Gangwon-do 25913, Korea, email: sjk-won@kangwon.ac.kr.

S. J. Kwon is the corresponding author.

Digital Object Identifier: 10.1109/JCN.2020.000018

for localization but they are not applicable in indoor environment because of weak satellite signals [2]. Strength of satellite signals decreases due to attenuation and scattering of signals by roof, walls and other objects. These satellite signals can only be used to attain precision in a definite range. Bluetooth, a short-range wireless mobile ad-hoc network, is another technology that is used for localization in indoor environment, but it is useful only for short range. In contrast to GPS and Bluetooth based methods, a better approach for localization in indoor environment is Wi-Fi using received signal strength indicator (RSSI). It is standardized by IEEE 802.11 and also its hardware components are pre-installed in most of the smart devices like smartphones, laptops etc.

The term outlier, also known as anomaly is originally taken from the field of statistics [3]. Outliers can be raised because of human error, machine error, mechanical faults and changes in the behavior of system or may be due to natural deviance in the environment. While outlier detection is the method of finding patterns from a given set of data that significantly differs or deviates drastically from the normal or average of the data set. Outlier detection is defined as finding patterns in data that behaves unexpectedly [4]. Objective of outlier detection is to find devices by their behavior that differs from the expected and previously observed in the field of IoT [5].

While in the case of Wi-Fi indoor localization environments, outliers are irregular and distinct RSS values caused by the following limitation of Wi-Fi indoor localization environment:

A. Accuracy of Localization

RSSs of a Wi-Fi network are actual signal strengths measured at receiver's device in decibel-milliWatts (dBm) or milliWatts (mW). RSS based techniques for indoor localization, such as Wi-Fi and bluetooth low energy (BLE) device are easy to deploy and cost efficient. Their signal strengths can be affected by the external forces such as reflection, refraction, diffraction, interference etc. due to the inherent nature of signals and from physical obstructions. The reference node in the wireless environment will receive different phase delayed and attenuated versions of the same signal. Almost all the techniques used for localization such as time difference of arrival (TDoA), time of flight (ToF), angle of arrival (AOA), including RSS depends upon these signals from the user's device or a reference node. A phase delayed and attenuated signal affected by reflection, refraction, diffraction, interference etc. causes irregularity in the RSS at reference node. This irregularity of RSS values affects the accuracy of localization and positioning in indoor environment.

B. Radio Environment

Wi-Fi based indoor networks use radio signals for communication. Radio signals are greatly affected by the physical obstacles such as walls, roof, doors and even from the human beings. Electric and magnetic field created by other electronic devices that are being used in indoor environment also effect radio signals. Radio signal strengths received at reference node also depend upon the location of node. These discussed factors about radio environment should essentially be taken into account for an accurate localization in indoor environment.

C. Bad Positioning

Bad positioning is actually incorrect position or location in indoor environment. For example, a smart device's user is trying to reach in the admin office of a department that is located on the second floor of a large building. A good indoor location system will show him exact position but a bad location system may show his location on first floor, third floor or may be an unexpected location. Hence, it's good to say that without positioning is better than bad positioning. Bad positioning in Wi-Fi indoor localization system can also be due to the irregular and anomalous RSSs caused by the external effects and physical obstructions of indoor environment.

Therefore, this research presents an outlier detection technique "iF_Ensemble" for Wi-Fi based indoor localization environment using the combination of supervised, unsupervised and ensemble leaning methods. To detect the outliers in the data of RSSs this research proposed to use unsupervised learning method such as isolation forest (iForest) to initially classify the data into two classes as normal and abnormal. After that machine learning classifiers support vector machine (SVM), K-nearest neighbor (KNN), random forest (RF) and Elliptic envelop has been applied separately on the data. However, their results were not satisfactory. Therefore, stacking an ensemble learning method has been proposed for the improved separation of outlier and normal RSSs.

The rest of the paper is organizing as follows. Section II addresses related work of outlier detection. Section III describes material and method used for outlier detection. Section IV describes the results and discussion. Finally, Section V concludes this paper.

II. RELATED WORK

Data analysts have always been concerned with the detection and correction of outliers of data because of security threats and unusual events. Outlier detection is an imperative area in studies and research of data mining and have various applications such as weather prediction, fraud discovery and interruption detection for a networks [6]. Detection of outlier has long been a field of study and research. An outlier detection technique called localization anomaly detection (LAD) was proposed in [7]. Their outlier detection scheme tries to detect the outliers and perform negotiation resilient localization without eliminating the anomalous node from the network. An anomaly detection technique to deal with the anomalous sensor's data has been presented in [8]. Authors discussed that low cost RSSs are mostly used for localization indoor environment but these RSSs are greatly affected

by the environment changes. They presented their technique for outlier detection purpose to deal with anomalous RSS data in indoor localization so that they can obtain reliable measures of RSS data for localization purpose.

An algorithm based on robust RSS for a secure localization has been presented in [9]. The proposed algorithm uses robust localization and time synchronization primitive, which correctly joined and allows attack's detection in the localization within an accurate attacker's model. A technique for outlier detection measurements that depends upon graph rigidity has been proposed in [10]. In the proposed approach authors examine that how much knowledge is required for the identification of outlier measurements. They propose the idea of confirmable edges by using the rigidity theory, and also develop the terms for an edge to be confirmable. A range based algorithm for outlier detection called localization density based outlier detection (LD-BOD) has been proposed that deals with the premise of density based anomaly detection strategies from the field of data mining [1]. To select a paramount candidate point, density based outlier detection uses the distance to K-nearest neighbor that is an average used for the estimation of location of unidentified node.

An Outlier detection approach using Bayesian belief networks (BBNs) for wireless sensor networks (WSNs) has been proposed in [11]. In their proposed approach authors discussed that in the user's perspectives; data consistency is an imperative problem in the context of WSNs steamed data. The consistency of data is affected by the severe condition of environment, interference in radio signals and bad quality sensor's usage. So, the data produced by the sensors can be incorrect and inconsistent that is main cause of missing values and outliers in the data. A clustering based outlier detection algorithm for WSNs called outlier detection based on clustering (ODC) has been presented in [12]. According to authors their proposed algorithm (ODC) will catches hidden outliers by the detection of labels of cluster in the slots of time.

An anomaly detection approach using improved iForest has been proposed in [13]. The proposed approach used simulated annealing method for the optimization of iForest. This simulated annealing method improved the speculation ability of iForest, eliminates terminated isolation tree (iTree) to minimize the computational issues. It improved speed of detection, stability and enhanced the performance of data outlier detection. An IoT architecture for the detection of outliers in the dataset of a forest environment has been proposed in [14] by applying four models of statistics such as gradient boosting machine (GBM), RF, classification and regression trees (CARD) and linear discriminant analysis (LDA).

An effective multi-dimensional process for user behavior anomaly detection (UBAD) has been developed using test of multi-dimensional statistics for anomaly detection [15]. They discussed that user behavior outlier detection is an imperative phase of user behavior analysis that can be used for the detection of events of an anomalous user from network traffic data and event. A methodology for the detection of abnormal actions of devices by examining their data of events has been proposed in [5]. Authors discussed that devices are associated with the internet and they are sending their log information to the server.

They mainly focus on the analysis of this data for automatically detecting infrequent patterns. They apply three steps i.e., feature or attribute generation, attribute aggregation and data analysis to process the log data.

According to best of our knowledge, in all above discussed previous studies of outlier detection in indoor localization and Internet of Things, machine learning methods have not been used along with ensemble learning as proposed in our research.

III. MATERIAL AND METHODS

A. The Data Set

The test dataset that has been used in the proposed iF_Ensemble technique for anomaly detection purpose has been taken from openly available database repository provided by the University of California, Irvine (UCI)¹. The dataset was collected by observing signal strengths of seven Wi-Fi routers using smartphone and saved in a text file to perform experimentation of indoor localization. The dataset contains 2000 entries (rows) with eight features (columns). Seven columns of the dataset contain the values of RSSs of seven Wi-Fi routers. The eighth column is class label as 1 to 4 that corresponds to different locations of an office such as conference room, kitchen or the indoor sports room etc.

B. Methods Used for Outlier Detection and Indoor User's Localization

Techniques used in the proposed approach include following machine learning methods.

1) Data Preprocessing

Raw data is mostly incomplete, inconsistent and may not be in a suitable format to perform an action. Data preprocessing is performed to clean, transform or normalize this data to bring it in the required and useful format.

In data preprocessing, data has been shifted from text file to .csv because .csv file provide well representation of data in the shape of rows and columns. In the .csv file headings for seven columns containing RSS values was given as rssi1 to rssi7 and for heading of class label 'y' has been used, so that it can be easily used for coming stages. After that .csv file was imported in a python program using python libraries such as pandas and required columns containing the values of RSS has been selected using columns function. At the end transformation of data into two dimensions (2d) has been made using decomposition library, principal component analysis (PCA) and its transform function in python.

2) Supervised Learning

Supervised learning is training of machine using some data that has already been labeled correctly. In this learning style corresponding proper outputs are provided along with the instances [16]. Some of the important supervised learning techniques that are proposed for outlier detection purpose in this research are:

- SVM with the parameter setup as $kernel='rbf'$ and $gamma = 0.0001$. $GridSearch()$ has also been used for the optimization of the parameters but the above given parameter setup is best for the underlying dataset.
- RF with parameter as $random_state = 0$, $n_estimators = 100$ and $maxdepth = 2$.
- K-nearest neighbors with $n_neighbors = 5$.

3) Unsupervised Learning

Unsupervised Learning is the training of a machine using information that is neither labeled nor classified. In this learning style corresponding proper outputs are not provided along with the instances [16]. As in unsupervised learning there is no need of human expert to manually label the data so it is extensively appropriate than supervised learning. Unsupervised learning method used in this research includes iForest with the parameter setup such as $n_estimators = 5$ and $contamination='auto'$. Local outlier factor (LOF) also has been used to compare with proposed iForest with $n_neighbors = 5$.

4) Ensemble Learning

Ensemble learning is a machine learning technique used to combine multiple weak learners into one but more effective model to get the better results [17]. It is a technique that uses multiple machine learning models instead of single model to solve a particular problem. The three terms used to describe the ensemble learning are Bagging, Boosting and Stacking. In this research we prefer to use stacking that is used for combining different classifiers to improve their predictive force. It is used to combine and learn numerous different weak learners by training a meta-model for the final predictions results. This meta-model based on the predictions results of weak learner models for the final output.

C. Evaluation Methods

For the evaluation and accuracy checking of outlier detection methods as well as classification methods of the proposed approach, following method has been used:

1) Confusion Matrix

Confusion matrix is used to measure the performance of classifier in machine learning on a test set of data that's true values are already known. The proposed approach of outlier detection has been implemented in python programming and python programming provide a library named metrics and the classes in it $confusion_matrix()$ and $classification_report()$. These python classes take the actual class labels of the underlying problem and the predicted result of classifier as an input and returns in a form of matrix. Results of confusion matrix are used to calculate other performance metrics such as accuracy, recall, precision, F-score and specificity.

2) Receiver Operating Characteristic (ROC) curve

ROC curve is a graph that is used for the evaluation of classifier's performance in machine learning. Its plots the true positive rate (TPR) on y-axis versus false positive rate (FPR) on x-axis at different classification thresholds. It is a probabilistic curve and

¹<https://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization>

specifies an area under curve (AUC) that represents the degree of separability. ROC (AUC) is mostly used when classes of the underlying problem are balanced.

3) Precision-Recall curve

A Precision-Recall curve is a graph used for the evaluation of classifier’s performance in machine learning using precision and recall scores. Precision is also known as Positive Predictive values while Recall is also known as sensitivity or TPR. In the plot of Precision-Recall, Recall or TPR is placed on x-axis while precision is placed on y-axis. It is mostly used when classes of underlying data are unbalanced.

IV. IMPLEMENTATIONS

The proposed approach of outlier detection iF_Ensemble has been implemented in Python programming language using tools *Jet Brains PyCharm* and *Jupyter notebook*. Python is most famous programming language that is used in the field of data science because it provides numerous libraries, built in classes and function for data processing and visualization. Following steps have been performed in iF_Ensemble using python program for detection of outliers from the RSS data of indoor localization environment.

- Input data is taken as a .csv file.
- To check the localization accuracy, user’s localization is performed using classification methods KNN, SVM and Naïve Bayes according to the class labels (1 to 4) used for localization.
- Data preprocessing is performed to select the required features for outlier detection. These features are seven columns containing RSSs values.
- iForest an unsupervised machine learning method is then applied on the selected features.
- Data is classified as normal with class label 1 and abnormal with class label –1 as predicted by the iForest.
- Stacking, an ensemble learning technique that is used to improve the performance of classifiers is then applied. SVM, KNN and RF have been used as base learner and RF as meta learner for stacking.
- Elimination of the data observation predicted by stacking as outliers is performed.
- To check the improvement in localization accuracy after applying the outlier detection technique, user’s localization is performed using classification methods KNN, SVM and Naïve Bayes.

Semantic diagram of proposed outlier detection scheme is shown in the Fig. 1.

V. RESULTS AND DISCUSSION

In this section calculation of results and evaluation of the proposed approach has been made using the test dataset that was collected in indoor environment. A python code was implemented for this purpose and it was executed ten (10) times/iterations. For each iteration; precision, recall, F1-score and accuracy was computed from the results of

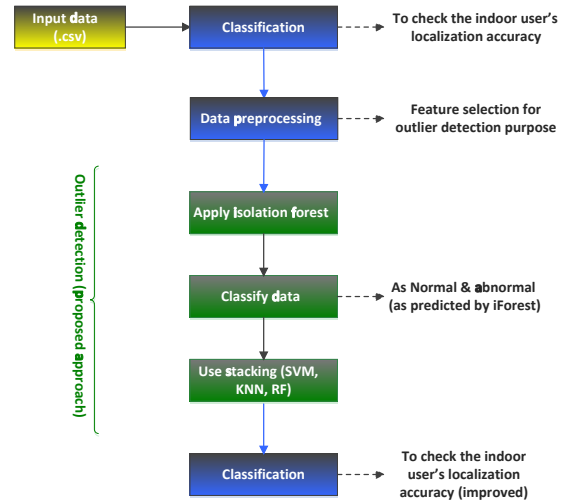


Fig. 1. Semantic diagram of proposed outlier detection approach iF_Ensemble.

Table 1. Results of user’s localization before applying the outlier detection approach.

Model	Precision	Recall	F1-score	Accuracy
SVM	96.41%	96.33%	96.25%	96.4%
KNN	96.45%	96.22%	96.22%	96.6%
Naïve Bayes	94.24%	94.16%	94.14%	94.2%

Table 2. Results of unsupervised learning methods.

Model	Accuracy
iForest	90.75%
Local outlier factor	88.45%

confusion_matrix() and *classification_report()*. Final results are computed using average of all ten values.

A. Indoor User’s Localization Accuracies before Applying the Proposed Outlier Detection Approach

Classification methods KNN [18], SVM [19] and Naïve Bayes [20] are performed on the indoor dataset to observe the localization accuracy in Wi-Fi connected indoor environment. The results of this step were later used to measure the difference in localization accuracy obtained after applying the proposed approach of outlier detection. Table 1 shows the accuracy results of user’s localization in Wi-Fi indoor localization environment before applying the proposed outlier detection approach.

B. Results of Proposed Outlier Detection Technique: iF_Ensemble

In iF_Ensemble, initially seven columns containing RSS values of Wi-Fi routers has been selected. As in case of raw data there was no clue that which portion of data is normal and which is abnormal (outliers). Therefore, iForest and LOF, two unsupervised learning methods, were applied on the selected features. As it is evident from Table 2 that accuracy of iForest is better than LOF thus we proposed to use iForest for separating normal data from abnormal in our proposed outlier detection technique.

In iForest result is predicted in the form of 1, –1 where 1 represents normal and –1 was used for abnormal data. Underlying dataset was assigned to a *dataFrame* in python program

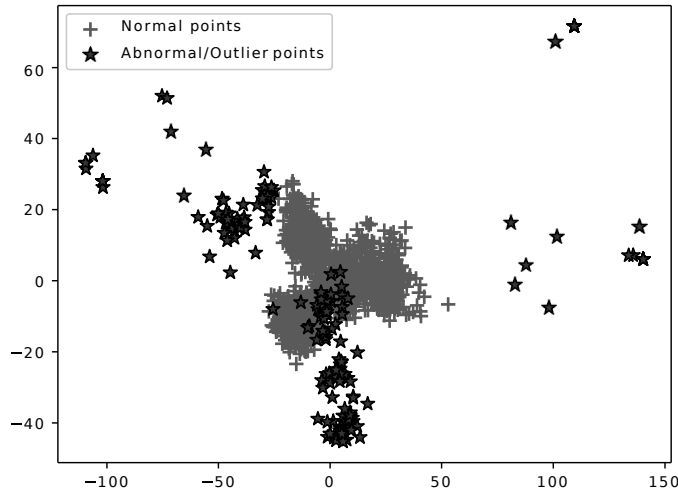


Fig. 2. Scatter plot of normal and abnormal classified data using iForest.

Table 3. Results of outlier detection.

Model	Precision	Recall	F1-score	Accuracy
SVM	96.18%	94.57%	95.64%	96%
Elliptic envelope	93%	91.93%	92.32%	92%
RF	94.79%	94.57%	93.73%	94.6%
KNN	96.89%	96.86%	96.85%	96.71%
Stacking	97.62%	97.67%	97.62%	97.8%

and labels 1 and -1 (predictions of iForest) were concatenated with the *dataFrame* against each entry (row) as eighth column. This eighth column is used as class label so that we can apply the classification methods for outlier detection. The results of iForest have been visualized in Fig. 2 using a two dimensional scatter plot of normal and outlier data.

To increase the accuracy level of proposed outlier detection method stacking is used. The main reason for using stacking is to enhance the accuracy of abnormal data identification with the help of classification methods.

Stacking uses base learner for initial predications and then a meta learner for final predication. Therefore, in this step of proposed approach KNN, SVM and RF has been used as base learners for stacking, while RF has been used as meta learner for the final predication to achieve the better results of outlier detection. These machine learning classifiers such as SVM, KNN, RF including elliptic envelope has also been applied separately (one by one) on the dataset to compare their results with proposed stacking.

All used models including stacking has been trained with 80 percent of the data observations and then whole dataset was tested using trained models instead of 20 percent of the remaining test data because our task was to differentiate outlier and normal data from the whole dataset. It is evident from Table 3 that stacking method showed highest accuracy of 97.8%.

Fig. 3 shows the accuracy of the algorithms used for outlier detection. Stacking algorithm achieves highest accuracy among all other algorithms.

We did not only evaluate the results of outlier detection using metrics such as precision, recall, F1-score and accuracy but also evaluates the results by plotting Receiver Operating Character-

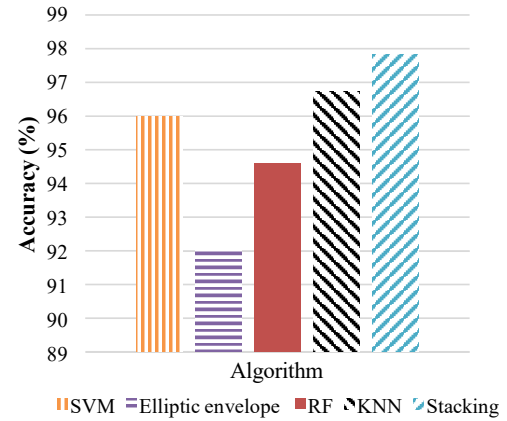


Fig. 3. Comparison of algorithms used for outlier detection.

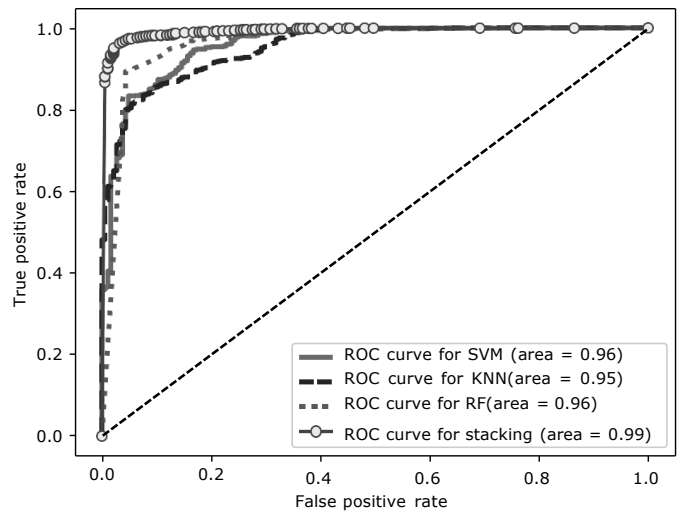


Fig. 4. ROC-AUC curve of outlier detection using SVM, ELPEN, RF and Stacking technique.

istic curve that is shown in the Fig. 4. The ROC graph shows the ROC curves and AUC of four different algorithms applied on the dataset that was classified as normal and abnormal by iForest. ROC curve shows the measurement of sensitivity considering ROC curve by plotting a function with (1-specificity) and the point on ROC expresses as sensitivity/specificity pairs those are comparing with a threshold [21]. The AUC and results will be better and high when a ROC curve passes from (0, 1) coordinate or ROC curve passes from near the upper left corner.

ROC graph plotted in the Fig. 4 shows 95 percent AUC for of elliptic envelope and 96 percent for both SVM and RF. It also shows ROC-AUC of proposed approach that is stacking with almost 99 percent AUC that is better than others.

The results of outlier detection have also been evaluated using precision-recall curve. Precision-recall curve is plotted in the Fig. 5 that includes the precision-recall curves of SVM, elliptic envelope and RF. It also shows precision-recall curve of proposed approach that is stacking. Precision-recall curve also

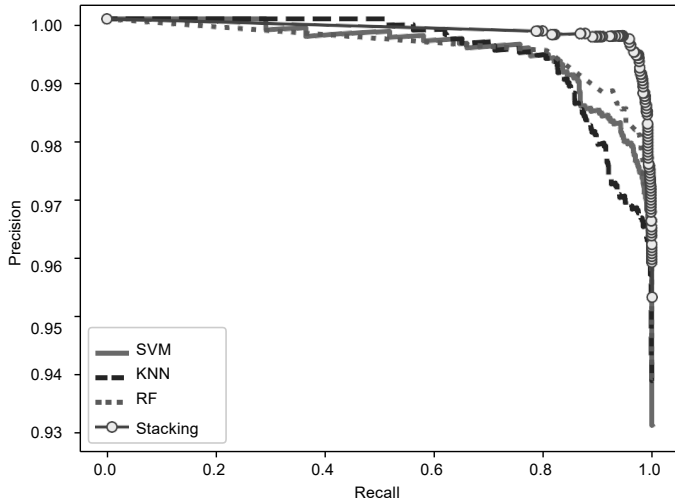


Fig. 5. Precision-recall curve of outlier detection using SVM, ELLEN, RF, and stacking technique.

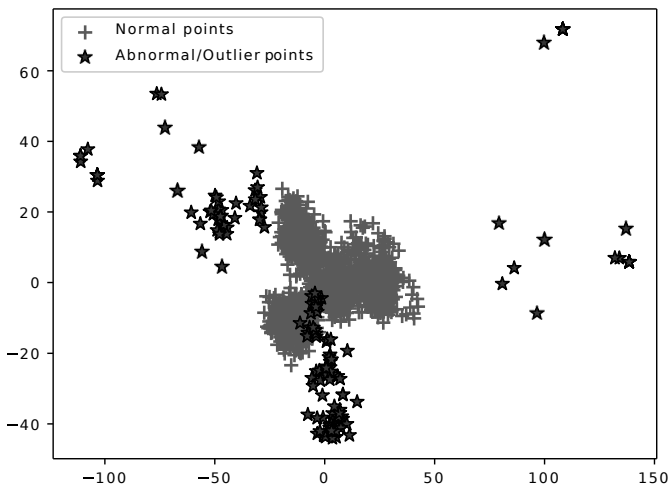


Fig. 6. Scatter plot of normal and abnormal classified data using iF_Ensemble, the proposed stacking based outlier detection approach.

indicates the better results of stacking.

After separating the normal and outlier data in the proposed approach outlier and normal data has been transformed into two dimension using PCA and transform related functions in python. Then these transformed values have been plotted in a scatter plot to visualize the result of proposed outlier detection approach as shown in Fig. 6. Figure shows that normal data points that were declared abnormal by iForest (Fig. 2) have been correctly identified as normal by applying stacking. Thus the proposed approach "iF_Ensemble" detects outliers more efficiently.

C. Indoor User's Localization Accuracies after Applying the Proposed Outlier Detection Approach iF_Ensemble

After applying the iF_Ensemble approach, outlier observation has been eliminated by skipping the observation containing class label -1 so there exist only the normal, accurate and anomaly free data for Wi-Fi indoor user's localization. Then

Table 4. Results of user's localization after applying the outlier detection approach.

Model	Precision	Recall	F1-score	Accuracy
SVM	98%	98%	98%	98.30%
KNN	98.25%	98.18%	98.20%	98.50%
Naïve Bayes	98%	98%	98%	98%

Table 5. Accuracy comparison of user's indoor localization.

Model	Before	After
SVM	96.4%	98.4%
KNN	96.6%	98.3%
Naïve Bayes	94.2%	98.2%

Table 6. Comparative results based on detection rate/recall.

Model	Detection rate/recall
SVM	93%
Neural network	92%
Naïve Bayes	88%
Decision tree	87%
iF_Ensemble(Proposed)	97.67%

classifiers that were previously used for user's localization in indoor environment, i.e., SVM, KNN, and Naïve Bayes have been used. Table 4 shows the results of user's indoor localization after applying outlier detection approach. Comparison of accuracies of indoor user's localization using classification methods such as SVM, KNN and Naïve Bayes before and after applying the proposed outlier detection approach is shown in the Table 5. It clearly indicates that accuracy of indoor localization improved after applying the proposed iF_Ensemble approach.

D. Comparative Analysis

To confirm the accuracy and performance of the iF_Ensemble it has been compared with the results of classification based methods for outlier detection in [22]. In which authors provides a complete and structured overview of classification based outlier detection techniques used in Wireless Sensor Networks. Comparative results based on detection rate (DR) or recall of the proposed approach and other classification based algorithms are shown in Table 6 where better results of our proposed approach can be observed.

VI. CONCLUSIONS

In indoor localization and IoT location information plays imperative role for accurate localization. But the RSSs value in Wi-Fi indoor environment are affected by the limitation of radio signals that causes irregularity in RSSs. These irregular and anomalous signals could not correctly identify a node in a local network. Therefore, this research has developed an outlier detection technique named as "iF_Ensemble" for a Wi-Fi indoor localization environment by analyzing RSSs using machine learning methods. The results evaluation of proposed approach using precision, recall, F-score and accuracy showed better results with 97.62 percent of precision, recall, and F-score and with 97.8 percent accuracy. The AUC of ROC and precision-recall curves plotted for the evaluation also confirm the better results of the proposed approach with almost 99 percent of AUC. In this research the accuracy results of user's indoor localization has also been calculated before and after out-

lier detection and elimination of outlier. It is observed that indoor user's localization accuracy improved by 2 percent. So, it is concluded that this can be useful and reliable approach for outlier detection in indoor localization and IoT. As general, there are little techniques for outlier detection in localization environments and wireless sensor networks. The applicability of artificial intelligence (AI) techniques for outlier detection in localization and WSNs can be investigated in the future.

REFERENCES

- [1] K. K. Almuzaini and A. Gulliver, "Range-based localization in wireless networks using density-based outlier detection," *Wireless Sensor Network*, vol. 2, no. 11, p. 807, Nov. 2010.
- [2] H. Aly, A. Basalamah, and M. Youssef, "Accurate and energy-efficient GPS-less outdoor localization," *ACM Trans. Spatial Algorithms Syst.*, vol. 3, no. 2, pp. 1–31, July 2017.
- [3] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys & Tuts.*, vol. 12, no. 2, pp. 159–170, 2010.
- [4] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Network Computer Applicat.*, vol. 60, pp. 19–31, 2016.
- [5] I. Cramer *et al.*, "Detecting anomalies in device event data in the IoT," in *Proc. IoTBDs*, 2018, pp. 52–62.
- [6] P. Yang and B. Huang, "A modified density based outlier mining algorithm for large dataset," in *Proc. FITME*, 2008.
- [7] J. F. Liu and Z. Ning, "Localization anomaly detection for wireless sensor networks," in *Proc. IEEE ICIS*, 2010.
- [8] Y. C. Chen and J. C. Juang, "Outlier-detection-based indoor localization system for wireless sensor networks," *Int. J. Navigation Observation*, vol. 2012, pp. 1687–5990, 2012.
- [9] S. Capkun, S. Ganerwal, F. Anjum, and M. Srivastava, "Secure RSS-based localization in sensor networks," Technical report/Swiss Federal Institute of Technology Zurich, Department of Computer Science, vol 529, 2011.
- [10] Z. Yang *et al.*, "Detecting outlier measurements based on graph rigidity for wireless sensor network localization," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 374–383, 2012.
- [11] D. Janakiram, V. A. Reddy, and A. P. Kumar, "Outlier detection in wireless sensor networks using Bayesian belief networks," in *Proc. COMSWARE*, 2006.
- [12] K. Niu, F. Zhao, and X. Qiao, "An outlier detection algorithm in wireless sensor network based on clustering," in *Proc. IEEE ICCT*, 2013.
- [13] D. Xu, Y. Wang, Y. Meng, and Z. Zhang, "An improved data anomaly detection method based on isolation forest," in *Proc. IEEE ISCID*, 2017.
- [14] N. Nesa, T. Ghosh, and I. Banerjee, "Outlier detection in sensed data using statistical learning models for IoT," in *Proc. IEEE WCNC*, 2018.
- [15] T. S. Prarthana and N. D. Gangadhar, "User behaviour anomaly detection in multidimensional data," in *Proc. IEEE CCEM*, 2017.
- [16] B. S. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging Artificial Intell. Applicat. Computer Eng.*, vol. 160, pp. 3–24, 2007.
- [17] C. Zhang and Y. Ma, eds. "Ensemble machine learning: Methods and applications," Springer Science & Business Media, 2012.
- [18] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers," arXiv preprint arXiv:2004.04523, 2020.
- [19] K. P. Soman, R. Loganathan and V. Ajay, "Machine learning with SVM and other kernel methods," PHI Learning Pvt. Ltd., 2009.
- [20] I. Rish, "An empirical study of the Naive Bayes classifier," in *Proc. IJCAI*, 2001.
- [21] S. A. Abbas, R. Riaz, Z. H. Kazmi, S. S. Rizvi and S. J. Kwon, "Cause analysis of Caesarian sections and application of machine learning methods for classification of birth data," *IEEE Access*, vol. 6, pp. 67555–67561, 2018.
- [22] A. Ayadi *et al.*, "Performance of outlier detection techniques based classification in wireless sensor networks," in *IEEE IWCMC*, 2017.



Mansoor Ahmed Bhatti was born in Haveli Kahuta, Azad Jammu & Kashmir, Pakistan in 1993. He received the MC.S. degrees from the University of Azad Jammu & Kashmir Muzaffarabad, Pakistan, in 2017 and currently enrolled for the degree of M.Phil in Department of Computer Science and Information Technology University of Azad Jammu & Kashmir Muzaffarabad, Pakistan, since 2017. He worked as IT Assistant in Information Technology Board of Azad Jammu & Kashmir. He is currently working as an IT Assistant at High Court of AJ&K Muzaffarabad, Pakistan.



Rabia Riaz did her Master's in Computer Sciences from Fatima Jinnah Women University Rawalpindi, in 2002. She received her Ph.D. in Electronics Engineering from AJOU University South Korea, with specialization in Network Security in 2009. She has more than ten years of Teaching and Research Experience in national and international organizations. Currently she is working as HOD/Assistant Professor at University of Azad Jammu and Kashmir in Department of Computer Sciences and Information Technology. She has also been Head of Department of Software Engineering and Department of Electrical Engineering. Her research interests include security of wireless and sensor networks, security and privacy of social networks, cryptography and network security.



Sanam Shahla Rizvi received her B.S.C. degree in Computer Science from Shah Abdul Latif University, Khairpur, Pakistan, in 2003, and M.C.S. degree in Computer Science from KASBIT University, Karachi, Pakistan, in 2004, M.S. degree in Software Engineering from Mohammad Ali Jinnah University, Karachi, Pakistan, in 2006, and Ph.D. degree in Information and Communication Engineering from AJOU University, Suwon, South Korea, in 2010. She has been working at MCS NUST and Preston University Pakistan as Assistant and Associate Professor in Department of Computer Sciences, respectively. Currently, she is working as Head of Research & Development at Raptor Interactive (Pty) Ltd, South Africa. Her research interests include flash memory storage, data management, database systems, CPU scheduling algorithms, and wireless sensor networks.



Sana Shokat is a Ph.D. scholar in Department of Computer Science and IT at University of Azad Jammu and Kashmir, Muzaffarabad. She has received her B.S. (Computer Science) degree from Pir Maher Ali Shah University of Arid Agriculture Rawalpindi, Pakistan in 2007 and her M.S. (Software Engineering) degree from Bahria University Islamabad, Pakistan in 2013. She has been working as a Lecturer from last 10 years in the Department of CS & IT at the University of Azad Jammu and Kashmir. Her research areas are human computer interaction, usability engineering and machine learning.



Farina Riaz received the BSCS degree in Computer Science from Quaid-e-Azam University, Islamabad Pakistan, in 2007, and the M.S. degree in Software Engineering from National University of Science and Technology, Islamabad Pakistan, in 2010. She has been working as Adjunct Lecturer in Manipal University, Dubai Campus for more than two years. Her research interests include wireless networks, sensor networks, data security, and encryption, authentication mechanism, and vehicle to vehicle communication



Se Jin Kwon received his M.S. and Ph.D. degrees in Computer Engineering from AJOU University, Korea, in 2008 and 2012, respectively. He worked as a Research Professor at the Department of Information & Computer Engineering, AJOU University, from 2013 to 2016. He worked as a Postdoctoral Researcher at University of California, Santa Cruz in 2016. He is currently an Assistant Professor at Department of Computer Engineering at Kangwon National University, Kangwon-do, 25806, Republic of Korea. His current interests include non-volatile memory systems, reliable storage systems, and large database systems.