Supervised Machine Learning for RSSI based Indoor Localization in IoT Applications

M.W.P. Maduranga
IIC University of Technology
Phnom Penh 121206
Kingdom of Cambodia

Ruvan Abeysekara
IIC University of Technology
Phnom Penh 121206
Kingdom of Cambodia

ABSTRACT

The Internet of Things (IoT) technology has revolutionized every aspect of everyday life by making everything smarter. IoT became more popular in recent years due to its vast applications in many fields such as smart cities, agriculture, healthcare, ambient assisted living, animal tracking, etc. Localization of a sensor node refers to knowing a sensor node's geographical location in the IoT network. In this research, we propose a device free indoor localization mechanism based on the Received Signal Strength Indicator (RSSI), a measure of the receiving signal from the sensor nodes, and supervised Machine Learning (ML) algorithms. An experimental test-bed was implanted in a controlled environment to collect RSSI values from the sensor nodes. The RSSI levels were collected by using multiple and published to a remote MQTT server over the Internet. In this research, RSSI values were used to train supervised ML algorithms, Linear Regression (LR), Polynomial Regression (PR), Decision Tree Regression (DTR), Support Vector Regression (SVR), and Random Forest Regressor (RFR) to estimate the accurate positioning of IoT related localization applications. The error between the actual measured values of the position and the estimated values are compared to validate the system model presented.

Keywords

IoT, Indoor Localization, Supervised Machine Learning, RSSI

1. INTRODUCTION

It is essential to apply localization technique to know the geographical location of a sensor node in many indoor IoT applications. For examples knowing a location of a moving animal, vehicle or a human using wireless sensor data. In these localization techniques, its not necessary to use additional hardware to detect the location, using the broadcasting signals form the sensor node itself can estimate its location. Moreover, it is significant as a sensor node's position affects the performance and accuracy of the information. On the other hand, the localization method could help estimate a sensor node's accurate position based on other neighbouring nodes' values. The IoT systems use many wireless technologies to communicate within the sensor nodes in their networks, such as Bluetooth, infrared, LoRaWAN, Zigbee, Wi-Fi, GPRS, and 3G. With these technologies, it can get the geographical location information of a node in different ways. Location-based services are the primary service of the IoT. Therefore, localization accuracy is an important issue. Many localization algorithms were developed for Wireless Sensor Networks (WSNs) applications and IoT. Most of the algorithms proposed in the literature for indoor localization are statistical-based algorithms [1-4]. Different hardware devices are utilized by most of the existing statistical localization solutions, which increase the cost and significantly limit the location-based applications.

Moreover, algorithms were developed for sensor node localization may inefficient and having difficulties deploying them on real IoT devices. Though significant related work exists on ML-based indoor localization for IoT systems, often proposed methods are evaluated using only a single ML algorithm, or no experimental test-bed implemented and evaluated. In this study it has designed and deploy an IoT testbed in the indoor environment and collect RSSI data. In the second phase, data were pre-processed and supervised ML algorithms investigated on estimating the accurate location of a sensor node and finally evaluate the performances of each proposed algorithm.

2. RELATED WORKS

The position estimation is a compulsory component of many IoT applications [5], especially in the tracking and monitoring applications such as [6-8]: animal behaviours monitoring, air quality monitoring, and smart cities. Besides, the location information enables various emerging applications such as inventory management, intrusion detection, road traffic tracking, health monitoring, etc. [9], [10]. Methods for determining position are usually based on geometric calculations such as trilateration (by measuring the angle to a fixed point or node with a known position) or trilateration (by measuring the distance between nodes) [11]. To determine the distance between two nodes in an IoT network, several techniques can be employed. For example, synchronization, RSSI, and the physical characteristics of the carrying wave [12]. The localization techniques in the IoT systems can be free of a previous position determination in the network, relying on a few specific sensors' position information and their inter-measurements in the network such as time difference of arrival, distance, angle of arrival, and connectivity [13]. For wireless-based indoor localization systems, the relationship between RSSI and distance is essential. The most common method is the location of fixed nodes based on triangulation [15]. According to [16], [17], [18], and [19], RSSI-based indoor positioning uses a variety of algorithms to identify mobile users in an indoor environment. These positioning algorithms are primarily trilateral measurements, arrival angle (AOA), arrival time (TOA), and arrival time difference (TDOA) [20] is based. Due to its simplicity and wide range of applications, the above algorithms' most popular is the trilateration algorithm.

Significant research works carried out on investigating ML for indoor localization problem in WSN and IoT [21], [22], [23], [24], [25]. Though significant related work exists on localization for WSN, most of the related work on Machine Learning-based indoor localization is for WSNs, not for IoT systems. Existing works on ML-based localization for IoT are limited to only one type of specific ML algorithm. Moreover, a lack of performance evaluation on different algorithms has been conducted. The majority of studies on deterministic localization algorithms factor in noise and various conditions; by using an ML-based approach, these factors and noise can be considered a part of the general environment in which the localization must be carried out.

3. EXPERIMENTAL TESTBED DESIGN

We designed the test-bed using two nodes named reference node and mobile node using an ESP-8266 chip. Figure 1 shows the test-bed arrangement, where 34 known location is marked with their x and y coordinates. ESP-8266 incorporates the IEEE 802.11 standard used in all indoor environments. Get compatible with Internet Protocol Version 4 (IPv4), Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Hypertext Transfer Protocol (HTTP) at low cost by using microchips. Beacon nodes connect to mobile nodes via Wi-Fi, the IEEE 802.11 standard in the 2.4 GHz frequency band, and the +20 dBm transmit power. Mobile node RSSI measurements for the reference node are sent to the remote server over Wi-Fi. The hardware design investigated includes mobile nodes, reference nodes, and IoT cloud architectures for sending RSSI data to remote servers, as shown in figure 2. The remote server uses supervised learning models to process RSSI data for indoor positioning. And figure 3 shows the methodology of the experiment.

The mobile node is implemented by an ESP-12E chip that records RSSI data via Wi-Fi via a reference node. The ESP-01 is used as a reference node with an independent 3.3V DC power supply via an ADP7158 linear regulator, as shown in figure 4 a and b. ESP-12E uses a lithium polymer secondary battery for storage. Experimental tests were conducted in an 8.02 square meter area that spans an open area surrounded by walls. The reference nodes were permanently placed at three corners of the above test area, and the mobile nodes can be moved to any known location. RSSI data is collected from 34 known sample locations by keeping the mobile node at a known location for about 60 seconds. Initially, data RSSI values received from the mobile node were collected using three reference nodes and in the initial step and number of reference nodes increased up to 4 and 5 by adding additional reference nodes in to the test bed.

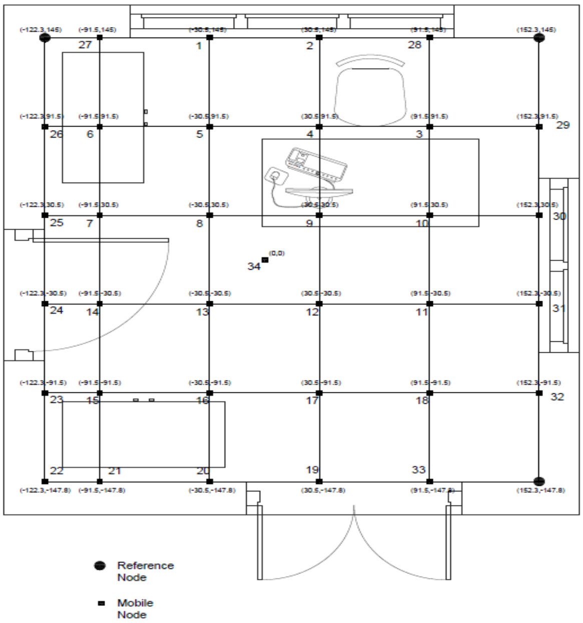

Figure. 1 Sensor node arrangement in the test-bed

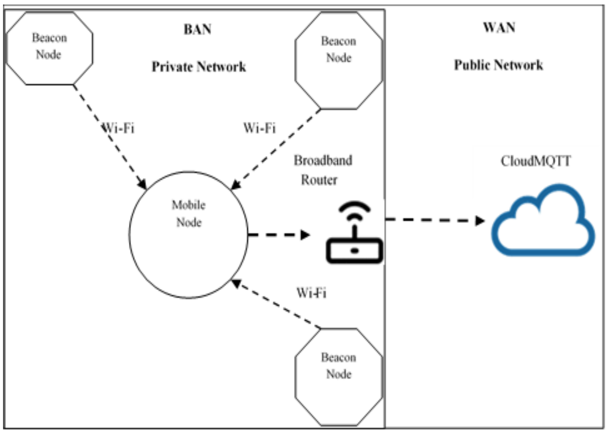

Figure. 2 System overviews of the test-bed

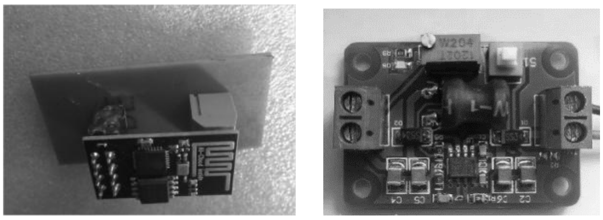
(a) (b)

Figure. 4 (a) reference sensor node and (b) 3.3V DC

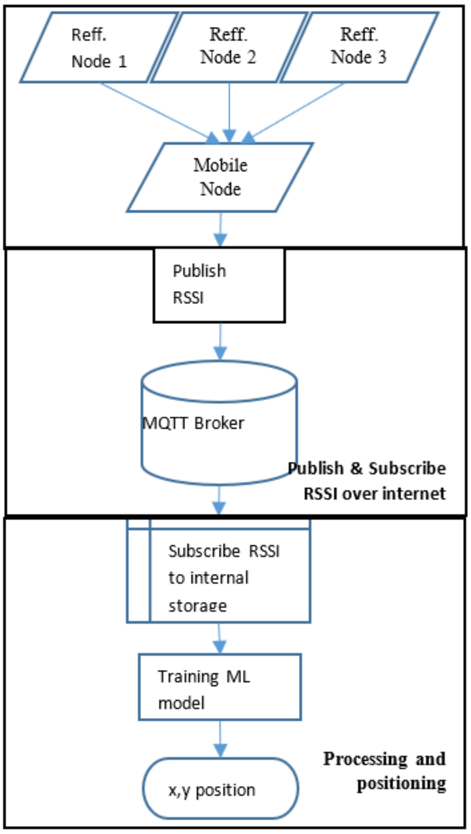

Figure. 3 Process of location estimating

4. DATA COLLECTION AND PRE-PROCESSING

In this design, CloudMQTT will be used as an IoT cloud platform, a managed Mosquitto server in the cloud. The MQTT broker provides a lightweight way to execute messages. This device is a low-power Wi-Fi module and remote computer. The mobile node publishes three RSSI data for the three beacon nodes to the MQTT broker over the Internet, subscribes to the RSSI data to a remote storage server, and further processes it with indoor localization.

5. PREDICTION MODELS DEVELOPMENT

The relationship between RSSI value and the distance can be expressed as The RSSI measure the received power level of an access point on a mobile client device. The RSSI units are in dB (decibel) or the similar 'dBm' (dB per milliwatt). The relationship between RSSI vs distance can be computed as follows;

$$\text{RSSI} = -(10\text{n}log_{10}d + A) \qquad (Eq.1)$$

where;

d - distance from the blind node to the reference node
n - Signal propagation constant
A - Received Signal strength at 1m distance

Supervised ML algorithms were investigated to estimate the accurate location of the sensor node. These algorithms are typically implemented in two phases. In the first phase, called the training phase, data is gathered and provided to the algorithm to learn patterns and create a model to classify data or predict data properties. In the second phase, called the testing phase, new data is tested against the model built during the training phase, and the effectiveness of the model is revealed. Such two-phase learning algorithms are called supervised learning algorithms. In this work, we investigate Linear Regression (LR), Polynomial Regression (PR), Support Vector Machine(SVM), Decision Tree Regression(DTR), and Random Forest Regression(RFR).All the algorithms were coded using Python 3 on Jupiter Notebook.

6. RESULT ANALYSIS

It has collected 4520 RSSI values from 32 known locations in the test-bed in one test. Then, the data set was trained using supervised ML algorithms Linear Regression (LR), Polynomial Regression (PR), Support Vector Machine (SVM), Decision Tree Regression (DTR), and Random Forest Regression (RFR). The Root Mean Squared Error (RMSE) and coefficient of determination, R2 were calculated can be compared. Further error histograms of each algorithm were plotted. Initially, the experiment started with three references nodes and step by step, the number of references nodes increased up to five. For each step, Root Mean Squared Error (RMSE) and Coefficient of Determination(R^2) values were calculated to compare the performances. Table 1 shows the results when the number of reference nodes is equal to 3. The RMSE value is decreasing in DTR as number of trees increases and R^2 is increasing. Also, error is decreasing in RFT as number of forest increases. Table 2 and 3 shows the RMSE value and the R^2 values when number of reference nodes increases up to 4 and 5 respectively. The RMSE and R^2 is improving as number of references nodes increases in the experimental test-bed. Figures 5(a) to 5(j) shows the error distribution of x coordinate and y coordinates for each algorithm. It’s interesting to see that error distribution is optimal at DTR supervised ML algorithm.

Table 1. Statistical comparison of each algorithm (Number of reference nodes=3)

ML Algorithm		Root Mean Squared Error (RMSE)		R^2	
		x	y	x	y
Linear Regression		77.73	71.84	0.2664	0.4183
Polynomial Regression		69.16	57.54	0.4192	0.6268
Support Vector Regression		73.37	66.84	0.34642	0.49634
Decision Tree Regression	No. of Trees=15	29.33	28.52	0.8955	0.9083
	No. of Trees=25	28.12	27.84	0.9040	0.9126
	No. of	28.1	27.8	0.9037	0.9124

	Trees=35	6	7		
Random Forest Regression	No. of Forests=40	31.14	29.09	0.8822	0.9046
	No. of Forests=70	30.84	28.80	0.8844	0.9064
	No. of Forests=100	30.90	28.70	0.8840	0.90714

Table 2. Statistical comparison of each algorithm (Number of reference nodes=4)

ML Algorithm		Root Mean Squared Error (RMSE)		R^2	
		x	y	x	y
Linear Regression		77.98	71.22	0.2773	0.4222
Polynomial Regression		67.33	54.33	0.4293	0.6322
Support Vector Regression		73.11	65.39	0.3692	0.4828
Decision Tree Regression	No. of Trees=15	28.33	28.22	0.8935	0.9183
	No. of Trees=25	27.13	26.99	0.9110	0.9189
	No. of Trees=35	25.12	24.17	0.9231	0.9321
Random Forest Regression	No. of Forests=40	29.10	29.11	0.8303	0.9189
	No. of Forests=70	28.20	28.80	0.8910	0.9292
	No. of Forests=100	28.92	28.21	0.8940	0.9174

Table 3. Statistical comparison of each algorithm (Number of reference nodes=5)

ML Algorithm		Root Mean Squared Error (RMSE)		R^2	
		x	y	x	y
Linear Regression		77.55	71.10	0.2910	0.4292
Polynomial Regression		65.93	52.14	0.4393	0.6502
Support Vector Regression		71.11	65.22	0.3600	0.4802
Decision Tree Regression	No. of Trees=15	27.31	27.10	0.8959	0.9191
	No. of Trees=25	27.03	25.93	0.9139	0.9123
	No. of Trees=35	24.12	23.17	0.9598	0.9478
Random Forest Regression	No. of Forests=40	28.92	29.11	0.8403	0.9191
	No. of Forests=70	28.01	28.31	0.8923	0.9298
	No. of Forests=100	28.11	28.02	0.8945	0.9187

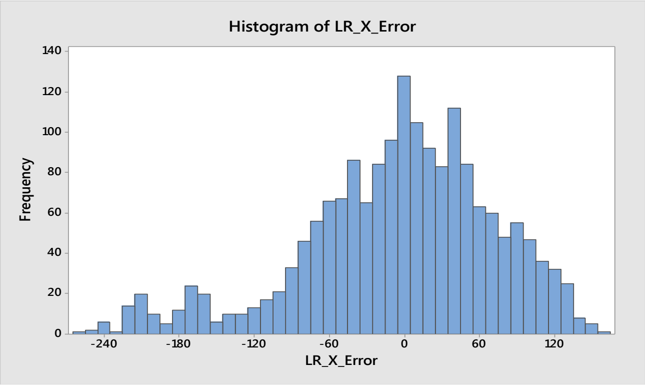

Figure 5(a): Error distribution of x coordinate in LR

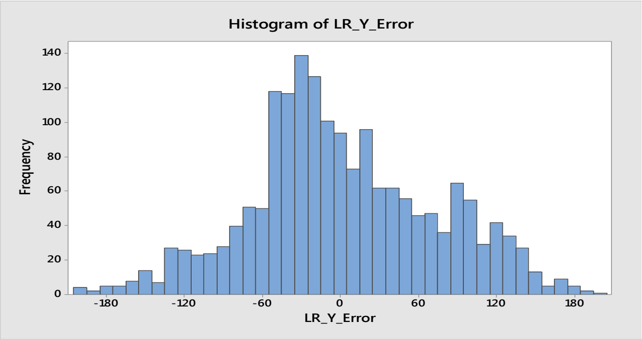

Figure 5(b): Error distribution of y coordinate in LR

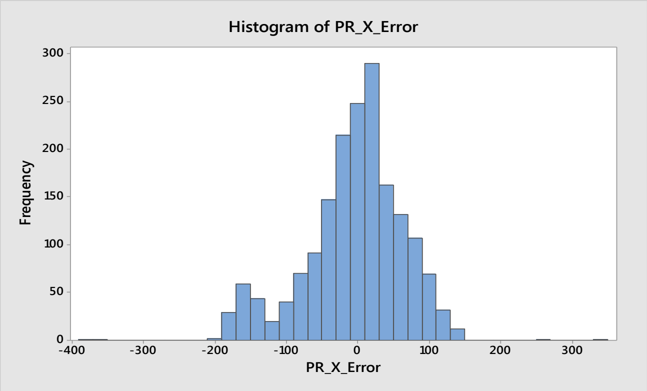

Figure 5(c): Error distribution of x coordinate in PR

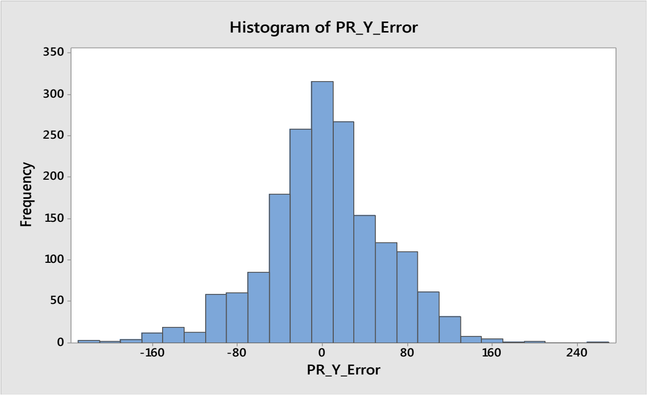

Figure 5(d): Error distribution of y coordinate in PR

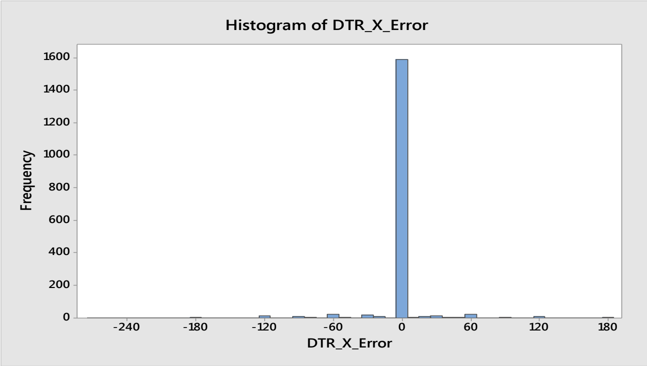

Figure 5(e): Error distribution of x coordinate in DTR

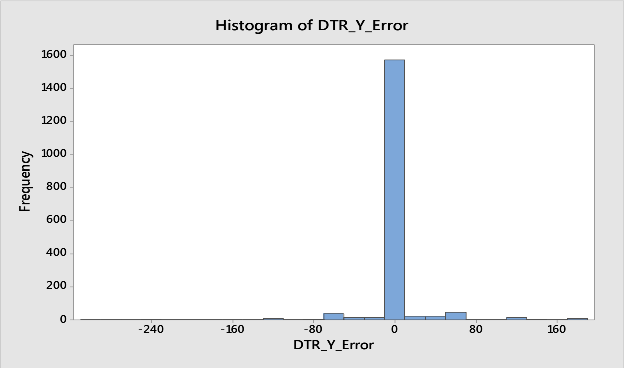

Figure 5(f): Error distribution of y coordinate in DTR

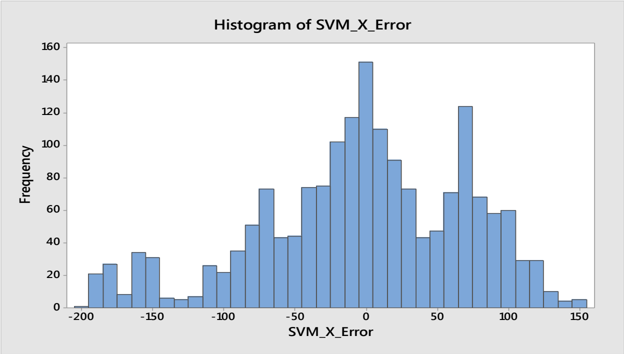

Figure 5(g): Error distribution of x coordinate in SVR

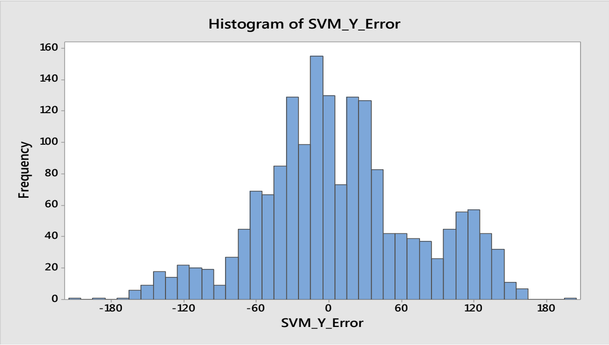

Figure 5(h): Error distribution of y coordinate in SVR

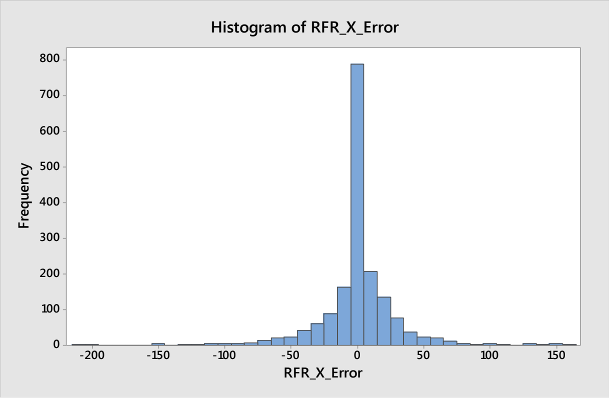

Figure 5(i): Error distribution of x coordinate in RFR

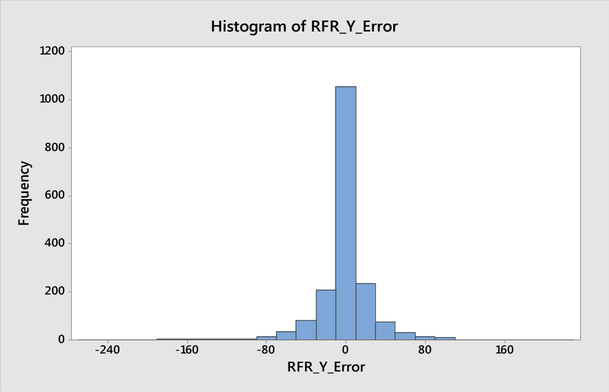

Figure 5(j): Error distribution of y coordinate in RFR

7. CONCLUSIONS

An experimental test-bed was designed to collect RSSI data, which emits by a mobile sensor node, and multiple fixed sensor nodes receive those RSSI values in the test-bed. The mobile node kept at 32 different known geographical location, and RSSI data were collected through an MQTT server. Collected data were filtered and pre-processed in order to train supervised ML algorithms. After the experiment with different supervised algorithms under different conditions, it becomes clear that Decision Tree Regressor (DTR) was the best-outperformed algorithm compare to the rest of the algorithms tested. The number of forests in DTR is matter to improve the location estimation accuracy and its significance in reducing the error. It observed that once increase the number of reference nodes in the test-bed, the accuracy and the error was significantly improved. Based on our experiments, we foresee using supervised ML algorithms to give improved results rather than deterministic localization.

8. REFERENCES

[1] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," In: *Proc. of IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, third quarter 2019

[2] P. Fonseka and K. Sandrasegaran, "Indoor localization for IoT applications using fingerprinting," In: *Proc. of 2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, Singapore, 2018, pp. 736-741

[3] S. S. Mohar, S. Goyal and R. Kaur, "A Survey of Localization in Wireless Sensor Network Using Optimization Techniques," In: *Proc. of 2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2018, pp. 1-6.

[4] M. N. Rahman, M. T. I. A. T. Hanuranto and S. T. M. T. R. Mayasari, "Trilateration and iterative multilateration algorithm for localization schemes on Wireless Sensor Network," In: *Proc. of 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*, Yogyakarta, 2017, pp. 88-92.

[5] P. Fonseka and K. Sandrasegaran, "Indoor localization for IoT applications using fingerprinting," In: *Proc. of* 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018, pp. 736-741.

[6] M. Asikainen, K. Haataja and P. Toivanen, "Wireless indoor tracking of livestock for behavioral analysis," In: *Proc. of* 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), Sardinia, Italy, 2013, pp. 1833-1838.

[7] M. Gor et al., "GATA: GPS-Arduino based Tracking and Alarm system for protection of wildlife animals," In: *Proc. of* 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, 2017, pp. 166-170.

[8] L.W. Turner, M.C. Udal, B. T. Larson, and S.A. Shearer. Monitoring cattle behavior and pasture use with GPS and GIS. Canadian Journal of Animal Science. 80(3):

[9] T. D. McAllister, S. El-Tawab and M. H. Heydari, "Localization of Health Center Assets Through an IoT Environment (LoCATE)," In: *Proc. of* 2017 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, 2017, pp. 132-137.

[10] X. Ma *et al*., "A Survey on Deep Learning Empowered IoT Applications," in *IEEE Access*, vol. 7, pp. 181721-181732, 2019, doi: 10.1109/ACCESS.2019.2958962.

[11] M. E. Rusli, M. Ali, N. Jamil and M. M. Din, "An Improved Indoor Positioning Algorithm Based on RSSI-Trilateration Technique for Internet of Things (IOT)," In: *Proc. of* 2016 International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, 2016, pp. 72-77.

[12] S. Sadowski and P. Spachos, "RSSI-Based Indoor Localization With the Internet of Things," in IEEE Access, vol. 6, pp. 30149-30161, 2018,

[13] Y. Zhang, W. Wu and Y. Chen, "A range-based localization algorithm for wireless sensor networks," in Journal of Communications and Networks, vol. 7, no. 4, pp. 429-437, Dec. 2005.

[14] Y. Wang, Q. Jin and J. Ma, "Integration of Range-Based and Range-Free Localization Algorithms in Wireless Sensor Networks for Mobile Clouds," In: *Proc. of* 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, Beijing, China, 2013, pp. 957-961.

[15] I. V. Korogodin, V. V. Dneprov and O. K. Mikhaylova, "Triangulation Positioning by Means of Wi-Fi Signals in Indoor Conditions," In: *Proc. of* 2019 PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring), Rome, Italy, 2019, pp. 2339-2345.

[16] B. Yang, L. Guo, R. Guo, M. Zhao and T. Zhao, "A Novel Trilateration Algorithm for RSSI-Based Indoor Localization," in *IEEE Sensors Journal*, vol. 20, no. 14, pp. 8164-8172, 15 July15, 2020, doi: 10.1109/JSEN.2020.2980966.

[17] Z. Jianyong, L. Haiyong, C. Zili and L. Zhaohui, "RSSI based Bluetooth low energy indoor positioning," In: *Proc. of 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Busan, Korea (South), 2014, pp. 526-533.

[18] W. Chen, K. Kao, Y. Chang and C. Chang, "An RSSI-based distributed real-time indoor positioning framework," In: *Proc. of* 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, Japan, 2018, pp. 1288-1291.

[19] E. Goldoni, A. Savioli, M. Risi and P. Gamba, "Experimental analysis of RSSI-based indoor localization with IEEE 802.15.4," In: *Proc. of* 2010 European Wireless Conference (EW), Lucca, Italy, 2010, pp. 71-77.

[20] U. Nazir, N. Shahid, M. A. Arshad and S. H. Raza, "Classification of localization algorithms for wireless sensor network: A survey," In: *Proc. of* 2012 International Conference on Open Source Systems and Technologies, Lahore, Pakistan, 2012.

[21] Y. Cheng, H. Chou and R. Y. Chang, "Machine-Learning Indoor Localization with Access Point Selection and Signal Strength Reconstruction", In*: Proc. of 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China,* 2016, pp. 1-5

[22] A. H. Salamah, M. Tamazin, M. A. Sharkas and M. Khedr, "An enhanced Wi-Fi indoor localization system based on machine learning," In: *Proc. of* 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcala de Henares, Spain, 2016, pp. 1-8, doi: 10.1109/IPIN.2016.7743586.

[23] M. Elbes, E. Almaita, T. Alrawashdeh, T. Kanan, S. AlZu'bi and B. Hawashin, "An Indoor Localization Approach Based on Deep Learning for Indoor Location-Based Services," In: *Proc. of* 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 2019, pp. 437-441.

[24] D. Milioris, "Efficient Indoor Localization via Reinforcement Learning," In: *Proc. of* ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 8350-8354.

[25] M. Dziubany *et al.*, "Machine Learning Based Indoor Localization Using a Representative k-Nearest-Neighbor Classifier on a Low-Cost IoT-Hardware," In: *Proc. of 2018 26th European Signal Processing Conference (EUSIPCO)*, Rome, Italy, 2018, pp. 2050-2054.