International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Research Analysis of Signals using Machine Learning Techniques

Navpreet Kaur, Inderdeep Kaur Aulakh

Abstract: ***Enhancements in installed microchips, low-control simple and computerized hardware, and radio correspondences have empowered the improvement of little and low-estimated sensor hubs or nodes (SNs) that made remote sensor systems, WSNs one of the promising advances amid the previous decade. Over the most recent couple of years wireless sensor systems (WSNs) have drawn the consideration of the exploration network, driven by an abundance of hypothetical and viable difficulties. This dynamic research in WSNs investigated different new applications empowered by bigger scale systems of sensor hubs fit for detecting data from nature, process the detected information and transmits it to the remote area. [1][2][3]***

Keywords: WSN, SN, DT, SVM

I. INTRODUCTION

WSNs are generally utilized in, low data transfer capacity and postpone tolerant, applications going from common and military to natural and social insurance checking. WSNs for the most part comprise of at least one sink and maybe range of nodes are dispersed in a physical space [4]. With expanding capabilities of detecting and calculating, the sensor nodes can detect the rough data then process it and further forward it to the sink. The sink thus questions the sensor hubs for data. [5][6] [7][8].

Following are the basic features for each WSNs:

1. Restricted Energy Utilization.
2. Ability to renew the failed nodes.
3. Supports dynamic nature of nodes.
4. Nodes can different types of data.
5. Possible to deploy nodes on large scale.
6. Capacity to work in unsuitable environments.
7. Simple to develop and understand.

A. Types of WSN: [7]

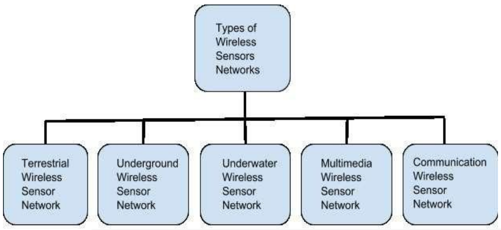


Fig.1. Types of Wireless Sensor Networks

Manuscript published on 30 September 2019

* Correspondence Author

Er. Navpreet Kaur*, Assistant Professor in the Department of Computer Science, University Institute of Engineering, Chandigarh University, Punjab, India.

Dr. Inderdeep Kaur Aulakh, Associate Professor in the Department of Information Technology, University Institute of Engineering and Technology, Panjab University, Chandigarh, India.

In Fig 1, these systems constitute of nodes that can be self-processed and work together with the real conditions. The dynamic nodes or hubs can sense and process. The WSNs are substantially more compliant than the static networks. The upper side of dynamic nature of network over the static network is improved inclusion, bettered energy proficiency, etc.

II. ML AND IT'S TECHNIQUES

The first step in making machine learn is with understanding, for example, immediate involvement, or suggestions, so as to look instances in observed details and select apt and superior preferences afterwards depending on the frameworks provided. The fundamental point is to make the machines come to terms naturally in the absence of external help and reorient events as required.

ML frameworks are oft-times separated as Supervised ML and Unsupervised ML.

- Supervised ML calculations can practice the earlier acquired data to newest findings by opting the marked guides for anticipating upcoming events. Initializing by inspecting the available dataset, the learned observations are derived in order to forecast the next events.
- In differentiate, Unsupervised ML is applied when the selected data is not structured. Unsupervised learning helps in developing ways that can convert the unstructured (not labelled) data into structured information.
- Reinforcement ML adds up with the existing scenarios by providing procedures and deduces rewards. To improve its execution, it allows machines and specialists to accordingly plan the best move so as to enhance its working

Distinction among Supervised ML and Unsupervised learning ML

Supervised Technique	**Unsupervised Technique**
Loaded information is Labeled	Input information is Unlabeled
Utilizations Preparing dataset	Uses simply the input dataset
For the most part utilized for prediction	Mainly utilized for analytics
Better Accuracy	Less Precise
Relapse and classification	Clusterin

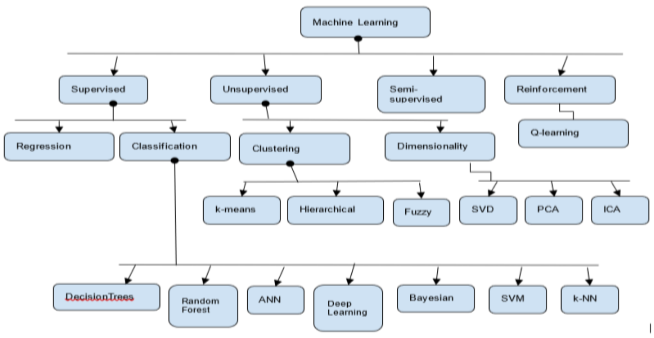


Fig 2. ML and its techniques

A. Decision Trees

In Decision Trees, the information is by default a section in accordance to a particular parameter. It consists of two sub-parts, the leaves as final outcomes and nodes as repository of information sections.

B. Support vector machine

SVMachine locates an ideal hyper-plane for sorting the information. SVMachine plays out the best characterization utilizing hyper-plane and arrange singular perception. The vast majority of preparation information is repetitive once a limit built up and a lot of focuses distinguishes the limit. Applying SVMachine for WSNs have tended to issues in WSNs, for example, restriction, availability issue, flaw discovery, steering, and blockage control.

C. Random Forests

This is group learning strategy to label and backtrack and various undertakings, by building DT in bulk amount at processing time and harvesting the character which is part of the characterization or backtrack path of the decision trees.

D. K-Means Clustering:

Technique used when there is unlabeled information, the pivotal use of this is to find clusters in the unclassified data. It works in loop to classify K clusters depending on the specific variables that are given. Information clusters are grouped depending on likeness.

III. PROJECT TECHNICALITIES

For target area converge issue, choosing an ideal no. of sensor hubs to cover the zone is effectively gotten by Machine Learning methods and can be utilized to isolate the defective sensor hubs from ordinary sensor hubs and improve the effectiveness of the system, Coverage and Connectivity are a standout amongst the most significant testing issues in WSNs. Coverage implies how proficiently each conveyed sensor screens the region of intrigue. The organization of hubs in system has haphazardly relied upon application. In a large portion of the WSNs application, arbitrary organization was doable when contrasted with deterministic arrangement. Coverage can be for the most part characterized into two classifications, for example,

1.Full Coverage
2.Partial Coverage

A. TECHNIQUES FOCUSED ON

- DECISION TREES
- MULTICLASS RANDOM FOREST
- SVM
- BAYESIAN LEARNING

Workflow While Implementing Machine Learning

- MULTICLASS CLASSIFICATION
- PROBLEM FORMULATION:

Classify the Signals received, based on their RSSI values from Source Node to Sink Node.

10 nodes are taken into consideration and each one is transmitting packets at a fixed interval to the sink node. Measurements are taken at the sink node.

- DATA ANALYSIS:

Since Manual collecting of data would require a large number of hardware capability which at the moment isn't at our disposal, We took the help of a MATLAB. Labels were added corresponding to the entries. The following points are noteworthy:

IV. NETWORK CONFIGURATION

The naming convention of each .csv file is defined as: node"i".csv, where "i" stands for id of the sensor node.

A. Each .csv file contains columns:

1. mean value of received signal strength over the path between node i and sink node which is multi hoping.
2. standard deviation of received signal strength over the path between node i and sink node which is multi hoping.
3. mean value of link quality indicator over the path between node i and sink node which is multi hoping.
4. standard deviation of link quality indicator over the path between node i and sink node which is multi hoping.

The dataset was unordered initially. It was cleaned and adjustments were made to the corresponding rows. Labels were added corresponding to the signal levels:

Signal Strength	Expected Quality
-30 dBm	Maximum signal strength, you are probably standing right next to the access point.
-50 dBm	Anything down to this level can be considered excellent signal strength.
-60 dBm	Good, reliable signal strength.
-67 dBm	Reliable signal strength.
-70 dBm	Not a strong signal.
-80 dBm	Unreliable signal strength, will not suffice for most services.
-90 dBm	The chances of even connecting are very low at this level.

CLASSIFICATION OF SIGNAL STRENGTHS AND LABELS:

Table I. Signal Parameters

SIGNAL STRENGTH	LABEL	REMARK
0 TO -30 dB	0	GOOD CONNECTION
-30 TO -67 dB	1	OKAY CONNECTION

-67 TO 89	2	POOR CONNECTION
-89 AND BELOW	3	LITTLE TO NO CONNECTIVITY

V. DATA PREPROCESSING:

We are going to import three libraries for our purpose:

- Pandas: One of the most popular libraries for data manipulation and storage. This is used to read/write the dataset and store it in a dataframe object. The library also provides various methods for dataframe transformation.
- Numpy: The library used for scientific computing. Here we are using the function vectorize for reversing the factorization of our classes to text.
- Sklearn: The library is used for a wide variety of tasks, i.e. dataset splitting into test and train, training the random forest, and creating the confusion matrix.

We are going to use 60% of the data for training and the remaining 40% as test data.We are not going to create cross validation datasets, as they are used when hyperparameter training is involved. Also, the reason for such high number of test case percentages is due to fewer numbers of rows for the model. Generally, 80/20 rule for train-test is used when data is sufficiently high. The below code uses the prebuilt function 'train_test_split' in a sklearn library for creating the train and test arrays for both independent and dependent variable. Also, random_state = 20 is assigned for random distribution of data.

We will perform the following steps:

1) Define a scaler by calling the function from sklearn library.
2) Transform train feature dataset (X_train) and fit the scaler on train feature dataset.
3) Use the scaler to transform test feature dataset (X_test).

VI. RESULTS:

Improvement Of Results:

The model was trained on 60% of data on a certain. We made similar execution on the other 9 nodes. The results are put side by side

Metrics Considered:

- PRECISION
- RECALL
- F1 SCORE

A. Requirement Gathering

RSSI AND LQI INDICATORS

There are two popular practical approaches that are used to define the energy of wireless received packets; the received-power and the estimated received link-quality indicator, RSSI and LQI respectively. The calculated power of a received broadcasting signal is defined as RSSI. RSSI is applied and extensively used in 802.11 standards while received power can be calculated from RSSI. On the other hand, the estimation of how simply the received signal can be modulated when considering noise existing in the communication channel is defined by LQI.

$$Pr = Pt * Gt * Gr(\lambda/4\Pi d)^{2}$$

Where:

Pt, Pr : The Power of the Transmission and Receive signals.

Gt, Gr : Transmitter and Receiver gains.

ƛ : Signal wave length.

d : Separation distance (sender and receiver).

$$RSSI = 10 * log_{10}(Pr/PRef)$$

Feasibility Study

This project is purely intended for research purpose. There is no public labeled dataset available for physical realization of the project. Also the manual cost setting up of hardware service is time consuming and expensive. We did our best to simulate and give a demonstration of a comparative analysis on a relatively small dataset.

VII. TEST RESULTS:

A. DATASET:

The dataset earlier comprised of 11 attributes namely RSSI, LQI, RSSI_MEAN, MEAN_PATH_LENGTH etc. After the feature extraction, only two features were extracted to serve our models.

The features extracted are: RSSI and LQI.

After this extraction, the new dataset was labelled by three target values 0,1,2. Target value 0 for RSSI_MEAN>-30, Target value 1 for RSSI_MEAN>-60 and Target value 2 for RSSI_MEAN > -90. These values were decided after lot of research and reading as no such labelled dataset exits publicly.

Table II. the Dataset


data - DataFrame

Index	RSSI_MEAN	RSSI_STDEV	LQI_MEAN	LQI_STDEV	Unnamed: 4	Unnamed: 5	Unnamed: 6	name:	MEAN_PATH_LENGTH	Unnamed: 9	Unnamed: 10	Unnamed: 11	U
0	-82.85	10.328	103.25	25.137	-93.525	2.6994	682.61	553…	2	0	0.43478	21.801	[illegible]
1	-80.706	3.8154	105.02	5.5094	-94.02	0.25452	692.54	558…	2	0	0.44797	21.769	[illegible]
2	-59.064	2.3576	96.702	3.1453	-96.298	13.869	685.5	554…	2	0	0.56165	21.76	[illegible]
3	-85.489	4.7388	103.22	8.5729	-93.756	2.1847	684.44	558…	2	0	0.51429	21.716	[illegible]
4	-59.229	1.9267	97.146	2.8329	-94.917	9.1597	703.28	573…	2	0	0.50153	21.672	[illegible]
5	-86.123	1.9034	102.96	5.8759	-94.123	0.10745	691.53	566…	2	0	0.60623	21.645	[illegible]
6	-86.292	1.9149	102.79	6.415	-92.604	11.906	700.67	571…	2	0	0.97612	21.645	[illegible]
7	-86.478	2.3365	102.41	3.9381	-93.391	5.2817	692.62	566…	2	0	0.93906	21.645	[illegible]
8	-87.733	2.24	100.09	26.703	-92.933	50.062	682.59	559…	2	0	0.918	21.632	[illegible]
9	-87.082	2.6464	100.84	33.892	-94.061	0.057476	669.71	542…	2	0	0.90423	21.604	[illegible]
10	-87.653	1.6143	101.55	6.5331	-93.347	5.1245	705.41	575…	2	0	0.94416	21.564	[illegible]
11	-87.292	1.6233	101.67	5.5972	-97.438	115.12	699.65	568…	2	0	0.96267	21.523	[illegible]
12	-87.385	1.1213	101.46	7.6331	-93.077	43.456	696.35	560…	2	0	0.98163	21.47	[illegible]
13	-88.068	1.7454	99.818	12.058	-93.454	3.5207	722.58	593…	2	0	0.99167	21.427	[illegible]
14	-87.298	3.4006	100.85	14.765	-94.043	0.040742	677.74	555…	2	0	0.91715	21.39	[illegible]
15	-88.34	0.73517	99.553	14.077	-93.617	4.0661	674.25	548…	2	0	0.92157	21.375	[illegible]
16	-88.688	1.9648	98.021	16.479	-93.604	4.7808	701.03	569…	2	0	0.92541	21.34	[illegible]
17	-88.143	1.6735	99.286	15.918	-92.531	49.27	688.44	552…	2	0	0.962	21.307	[illegible]
18	-88.5	1.625	98.563	23.246	-94.146	0.49957	700.69	565…	2	0	0.978	21.289	[illegible]
19	-88.234	1.6261	99.34	11.884	-94.192	0.15482	672.97	545…	2	0	0.958	21.282	[illegible]
20	-88.383	1.5129	99.808	11.9	-93.575	4.1593	688	554…	2	0	0.94	21.264	[illegible]
21	-88.449	1.5127	99.225	14.99	-92.571	49.51	689.11	554…	2	0	0.94023	21.238	[illegible]

Format Resize Background color Column min/max Save and Close Close

In the tables below, parameters such as ACCURACY, NODE PRECISION, RECALL and F1 SCORE have been calculated for 10 nodes individually and values of these describe the stability of the models.

The values in the Accuracy table indicate that the model is stable and does not fluctuate much and is able to correctly predict the validity of the model. Here Precision values depict the correctness of the classification. Moreover, Recall values depict the no of correctly observed classifications. F1 Score provides us with average of recall and precision.

A. Decision trees

Table III. Results while using DT

NODE	ACCURACY	NODE PRECISION	RECALL	F1 score
1	96.25	83	66	45
2	93.75	80	54	56
3	93.75	93	54	43
4	95	85	87	91
5	93.75	94	94	90
6	95	95	95	95
7	97.5	95	97	96
8	97	98	95	93
9	88.75	73	69	67
10	86.45	78	73	69

B. RANDOM FOREST CLASSIFIER

C.

Table IV. Results

NODES	ACCURACY	NODE PRECISION	F1 SCORE	RECALL
1	67.5	83	66	68
2	51.25	66	44	51
3	60	56	49	60
4	51.25	68	51	48
5	48.75	61	43	49
6	75	60	41	58
7	70	73	68	70
8	67.5	73	67	72
9	59.75	64	58	59
10	58.75	68	48	59

D. SVM - Sigmoid Kernel

Table V. Results

NODE	MODEL ACCURACY	NODE PRECISION	F1 SCORE	RECALL
1	95	93	96	89
2	93.75	95	97	92
3	78.75	82	79	78
4	85	80	85	82
5	83.75	75	84	79
6	87.75	81	88	84
7	83.75	78	84	80
8	90	86	90	88
9	87.5	85	88	86
10	77.5	83	78	76

E. Seaborn visualization of data

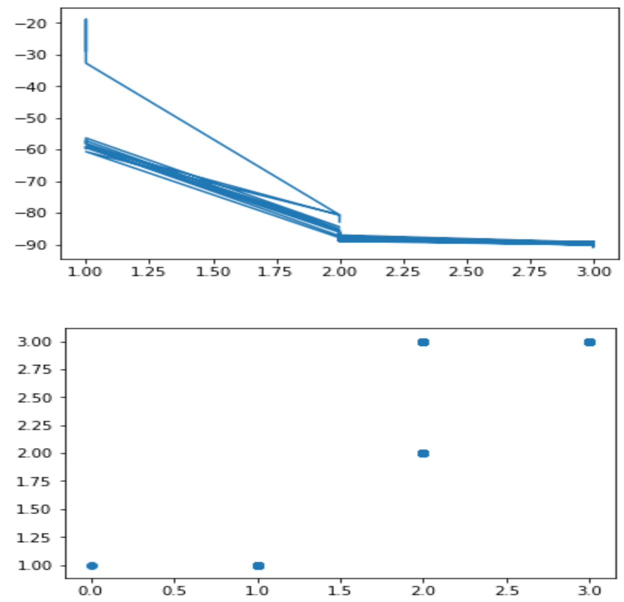
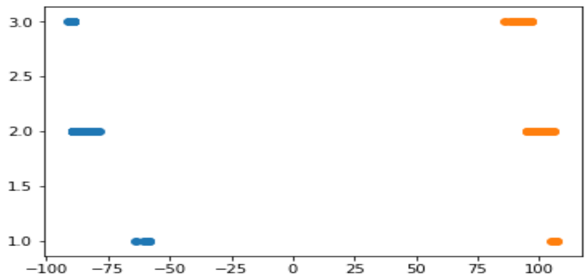
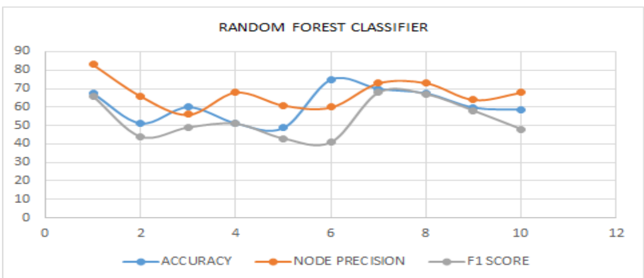


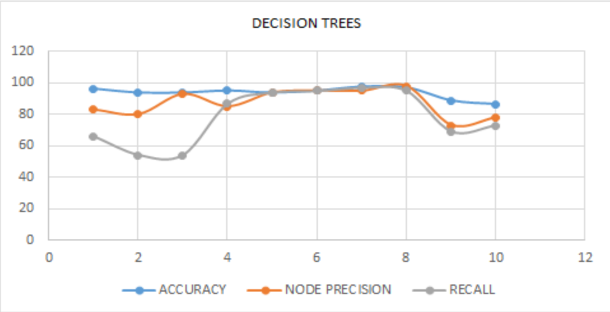
Fig 3. LQI+DTREE graph

This is a validation test: the X-axis represents the test labels for the dataset and the Y axis represents the prediction labels for the dataset.

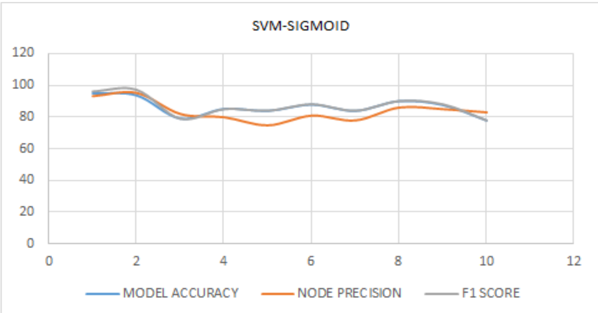
F. Scatterplot of test vs prediction samples

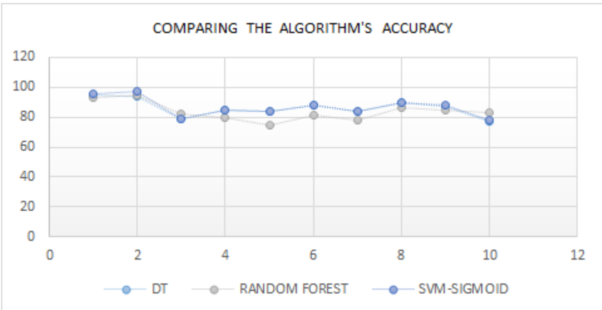






This Decision Tree output suggests that this decision tree output for a particular node suggests that the majority of signals are being classified as medium signals based on rssi values alone whereas based on LQI values, the signals majorly fall into medium and strong category. Since we are using 2 features, we can expect this anomaly





VIII. FUTURE SCOPE

- The outcome of this project is expected to be further utilized by various agencies as follows:

- Disaster Recovery Agencies: At the time of disaster communication, cables are somehow affected and further to carry out the communications, an efficient routing path must be made available. With the help of the RSSI value parameter, we can predict which machine learning algorithm is best.

- Research Agencies: Machine Learning has become a boom in current research trends, therefore integrating the ML with wireless sensor networks can be very helpful in providing a more efficient routing path and the outcome of this project would help us identify the best Machine Learning algorithm for the purpose.

- Defense Organization: During calamitous times Machine Learning based routing can be efficiently used to find the alternative best route to transmit the data in real time by finding the optimal path for data delivery.

REFERENCES

1. B. Zeng and Y. Dong, "An improved harmony search based energy-efficient routing algorithm for wireless sensor networks," Appl. Soft Comput. J., vol. 41, pp. 135–147,2016.
2. S. Ara, R. H. Filho, J. J. P. C. Rodrigues, R. D. A. L. Rabelo, N. D. C. Sousa, and J. C. C., L. S. Filho, "A Proposal for IoT Dynamic Routes Selection Based on Contextual Information,", Sensors,vol. 18,no. 2, pp. 1–16, 2018.
3. S. K. P. Assistant, T. Revathi, and K. Muneeswaran, "Multi-constraint multi-objective QoS aware routing heuristics for query driven sensor networks using fuzzy soft sets," Appl. Soft Comput. J., vol. 52,pp. 532-548,2016.
4. M. Z. Hasan, H. Al-rizzo, and F. Al-turjman, "A Survey on Multipath Routing Protocols for QoS Assurances in Real-Time Wireless Multimedia Sensor Networks," IEEE Communications Surveys & Tutorials,vol. 19,no. 3,pp. 1424-1456, 2017.
5. A. Sarkar and T. S. Murugan, "Routing protocols for wireless sensor networks : What the literature says ?," Alexandria Eng. J., vol. 55, no. 4, pp. 3173–3183, 2016.
6. W. Zhang, Y. Liu, G. Han, Y. Feng, and Y. Zhao, "An Energy Efficient and QoS Aware Routing Algorithm based on Data Classification for Industrial Wireless Sensor Networks," IEEE Access, vol. 6, pp. 46495-46504, 2018.
7. P. Prasad, "Recent trend in wireless sensor network and its applications : a survey," Sensor Review, Vol. 35 Issue: 2, pp.229-236, 2015.
8. D. Yuan, S. S. Kanhere, and M. Hollick, "Instrumenting Wireless Sensor Networks — A survey on the metrics that matter," Pervasive Mob. Comput., vol. 37, pp. 45–62, 2017.
9. Praveen Kumar D., *Tarachand Amgoth*, Chandra Sekhara Rao Annavarapu, Machine learning algorithms for wireless sensor networks: A survey, *Information Fusion* , vol. 49, pp. 1-25, 2019 (Elsevier).
10. K. Yau, K. P. Chow, S. M. Yiu and C. F. Chan, "Detecting anomalous behavior of PLC using semi-supervised machine learning," *2017 IEEE Conference on Communications and Network Security (CNS)*, Las Vegas, NV, 2017, pp. 580-585.
11. S. Ehsan and B. Hamdaoui, "A Survey on Energy-Efficient Routing Techniques with QoS Assurances for Wireless Multimedia Sensor Networks," in *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 265-278, Second Quarter 2012.
12. Jing, Xiangpeng and Dipankar Raychaudhuri. "Global Control Plane Architecture for Cognitive Radio Networks." *2007 IEEE International Conference on Communications* (2007): 6466-6470.

AUTHOR PROFILE

Er. Navpreet Kaur is Assistant Professor in the Department of Computer Science, University Institute of Engineering, Chandigarh University, Punjab, India. She is pursuing Ph.D. in Wireless Sensor Networks from Department of Information Technology, University Institute of Engineering and Technology, Panjab University, Chandigarh, India. She has 4.7 years of experience in teaching undergraduate and postgraduate students of Information Technology and research in the U.I.E.T Department of Panjab University. She has guided students in graduate thesis work and currently is supervising projects in the field of wireless sensor networks and IOT. Er. Navpreet has published 12 papers including international publications. She remains actively involved in recent developments of her field by attending and organizing seminars, workshops, conferences etc.

Dr. Inderdeep Kaur Aulakh is Associate Professor in the Department of Information Technology, University Institute of Engineering and Technology, Panjab University, Chandigarh, India. She has a Ph.D. in cognitive radio networks. She has 25 years of experience partly in Communication Equipment Industry and rest in teaching undergraduate and postgraduate students of Information Technology and Electronics Communication Engineering. She has guided 23 students in graduate thesis work and currently is supervising two PhD scholars in the field of wireless sensor networks and IOT. Dr. Aulakh has published 72 papers including 47 international publications. She remains actively involved in recent developments of her field by attending and organizing seminars, workshops, conferences etc.