Comparative Evaluation of Machine Learning-Based Localization Algorithms in Dense IoT Sensor Networks

Wesam Ali[1]*, A. A. Zaky[2]

[1]*School of Electrical Engineering, Kuwait Institute for Scientific Research (KISR), P.O. Box 24885 Safat, Kuwait*
[2]*Departamento de Engenharia Elétrica, Universidade Federal de Pernambuco - UFPE Recife, Brazil*

KEYWORDS:
Machine Learning,
Localization Algorithms,
Dense IoT Sensor Networks,
Wireless Sensor Networks (WSN),
Node Localization,
Received Signal Strength Indicator (RSSI),
Supervised Learning,
Deep Learning,
Random Forest,
k-Nearest Neighbors (k-NN),
Support Vector Machines (SVM),
Performance Evaluation,
IoT Positioning,
Smart Environment Localization,
Signal-Based Localization.

ARTICLE HISTORY:
Submitted : 16.08.2025
Revised : 13.10.2025
Accepted : 18.11.2025

https://doi.org/10.31838/WSNIOT/03.01.11

Abstract

Precise positioning of sensor nodes is an important facilitator of context-aware services, effective routing, and data interpretation in dense Internet of Things (IoT) sensor networks. Nonetheless, the traditional localization algorithms are limiting in dense deployment because of signal interference, non-line-of-sight situations and lack of scalability. The work includes a detailed comparative analysis of machine learning (ML)-specific localization algorithms designed to be used in densely populated settings of IoT. In particular we test the ability of k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forests (RF) and Deep Neural Networks (DNN) on important measures like localization accuracy, computational efficiency, resistance to environmental noise and scalability. Both synthetic and real-world datasets based on dense network scenarios are explored to mean experimental analyses. Examination of the results has shown that DNN models are more accurate because of sensitivity to complicated signal-space interactions, and the RF provides an attractive tradeoff of precision versus overhead. Instead, SVM presents scalability issues and k-NN is not performing well in highly dynamic or noisy environments despite being fast. The findings can be used in practice to choose which algorithm to use in order to deploy the algorithm into real-world IoT systems where the node distribution is dense. Our work adds another piece of research to the area of smart positioning solutions and can be used to develop energy efficient, scalable, and precise positioning systems to be used in the future IoT devices.

Author's e-mail: wesamal.i@kisr.edu.kw

How to cite this article: Ali W, Zaky AA. Comparative Evaluation of Machine Learning-Based Localization Algorithms in Dense IoT Sensor Networks. Journal of Wireless Sensor Networks and IoT, Vol. 3, No. 1, 2026 (pp. 79-85).

Introduction

The spontaneous growth of the IoT devices in fields like smart cities, industrial automation, environmental tracking, and precision agriculture has further strengthened the necessity of precise, scalable, and energy-efficient strategy of node localization. Localization is considerably harder in dense IoT sensor regions where node density is high and the available infrastructure is limited, and multipath and non-line-of-sight (NLoS) transmission, signal interference within the network, a lack of anchor nodes, and tight energy budgets are major challenges.

Machine learning (ML) techniques appear to be used as a solution to limited capabilities of classical localization methodologies (e.g., RSSI-based trilateration and fingerprinting), because they enable to describe complex, nonlinear relationships between feature features and the spatial coordinates. To illustrate, Shahbazian et al. [1] conduct a broad review of the works concerning IoT localization using ML, which also outlines the abilities of the latter in coping with the noise and comprehension of the issues including the sparsity of data, the complexity of the models, the heterogeneity of devices. Along the same line, a study published by Maduranga[2] in 2024 shows the positivity of better localization in IoT systems through ensemble ML techniques and preprocessing pipelines.

Nevertheless, these advances have been played out in either sparse or moderately densely set ups where there has been no systematic assessments of ML algorithms through an evaluative approach against a variety of ML methods in real world, heavily dense sets ups. Moreover, overall evaluations seeking a balance between accuracy, computational cost, scalability and resistance to noise are scarce.

This paper compares and evaluates the performance of the four common application theories of ML-based localization algorithms, k-Nearest Neighbors (k NN), Support Vector Machines (SVM), Random Forests (RF), and Deep Neural Networks (DNN), under dense IoT sensor networks to fill these gaps. The construction of experiments is based on synthetic and real-world datasets and evaluation of each algorithm according to key performance parameters. We hope that our findings can inform engineering engineers to tactfully choose and implement ML-based localization in the next-generation IoT platforms.

Background and Related Work

The traditional IoT sensor network driven localization is done using model-driven localization that can be grouped into three categories:

1. Range measurements: These generate estimates of distances or inter-node angles based on signal measurements including Time of Arrival (ToA), Time Difference of Arrival (TDoA), Received signal power indicator (RSSI) or Angle of Arrival (AoA). They perform perfectly in line-of-sight (LoS) path but perform very poor when dense deployments are involved as they are multipath-propagation-prone, signal interference and noise.[1]
2. Range-free techniques: These are the algorithms which are not based on direct distance or angle measurements, they: include centroid, DV-Hop, and APIT. Although low-end and power consumption, range-free methods have poor spatial resolution and accuracy especially in dense or topologically complicated domains.[2]
3. Fingerprinting-based solutions: Radio mapping is performed in an off-line phase based on data recorded in the radio environment at previously known locations in the fingerprinting-based methods. During the online stage, the signal observed is compared against the fingerprints database. Finger printing is more accurate in indoor and densely populated setting but the cost implication of finger printing is high in memory as well as labor expense in mapping and faces decay in performance with the change in the environment.[3]

In the previous years, however, machine learning (ML) methods have become popular due to their ability to capture nonlinear associations and to work with noisy observations. Such supervised ML algorithms as k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forests (RF), and Deep Neural Networks (DNN) have been used on localization tasks with good potential.[4, 5] Such models can interpolate based on training data and be able to model different propagation conditions without explicit physical modeling. But majority of the previous works have concentrated on a single ML method in ideal or moderately dense network conditions. Comprehensive comparative analysis of comparative analysis of two or more ML-based algorithms in the realistic dense deployment scenario is few. Also, trade-offs between localization accuracy, computational efficiency, noise robustness and scalability are under-investigated. It is this discrepancy that is the driving force behind the current research, that of benchmarking several ML models in dense IoT deployment situations.

System Model and Problem Formulation

We assume a two-dimensional dense Internet of Things (IoT) sensor network spread on a pre-determined geographical area in our investigation. The two categories of nodes used in the network are anchor nodes, positions of which are known ahead of time, and unknown sensor ones, whose coordinates must be found. The dense deployment means that sensor nodes are deployed in a dense spatial pattern and thus they are characterized with signal overlap, signal interference, and multipaths which are penetrating issues that require high quality localization. Figure 1 represents an overview of this system model.

The individual sensor nodes within the network will collect signal parameters, which can be Received Signal Strength Indicator (RSSI) and Channel State Information (CSI) or any other useful physical-layer metrics depending on signal transmissions on neighbor anchor nodes. These signal features are naturally dependent on the spatial location of the node thus they are good input to learning based localization model.

Suppose let the data set be as:

$$D = \{(x_i, y_i, f_i)\}_{i=1}^{N} \qquad (1)$$

where:

- (x_i,y_i) denotes the true Cartesian coordinates of the ith sensor node,
- f_i□$R^d$is a feature vector containing d signal-based measurements observed by node i,
- N is the total number of labeled data samples (collected either through simulation or from a pre-characterized environment).

The core objective is to learn a mapping function f such that:

$$f:f_i \rightarrow (x_i,y_i) \qquad (2)$$

This operation estimates spatial coordinates of a node based on only signal characteristics observed. The role of f is in practice approximated with various supervised machine learning (ML) algorithms: they include k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forests (RF), Deep Neural Networks (DNN). These models are trained on the labeled dataset D_{train} and evaluated on an unseen test set D_{test} to assess generalization performance.

Quantitative performance is assessed in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and latency of inferences with special attention paid to the performance behavior of these metrics in dense deployment applications, where spatial resolution and ability to tolerate noise is imperative.

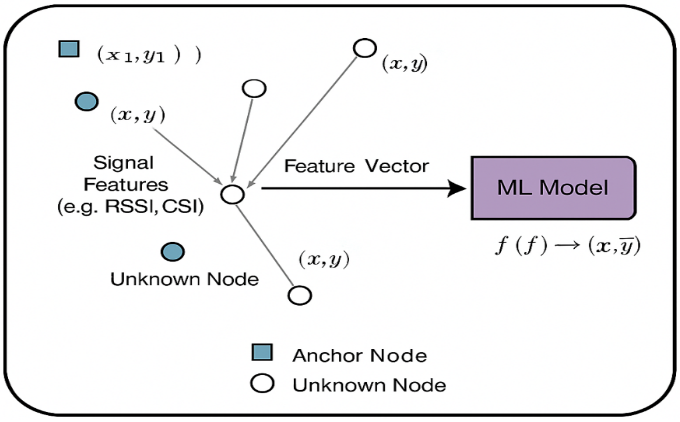

Fig. 1. System Model for ML-Based Localization in Dense IoT Sensor Networks

The system diagram of machine learning assisted node localization is depicted in the figure. Signals features (e.g., RSSI, CSI) are propagated by the anchor nodes that have known coordinates to unknown nodes. Such characteristics act as inputs of a ML model that can learn the mapping function f (f) (x, y) to predict the spatial locations of the unseen nodes.

Machine Learning Algorithms Evaluated

This paper discusses a proposed study to test the capability of four supervised machine learning (ML) algorithms in the localization of the nodes in dense IoT sensor networks. The different algorithms provide uniquely useful properties in terms of modeling of spatial relationships based upon signal derived features, with some being more complex, accurate, and computationally burdensome than others. The chosen models are the representatives of the balance between the classical, ensemble-based, and deep learning method so that a thorough comparison can be made. Figure 2 provides the overview of these algorithms and their functions in the process of localization.

k-Nearest Neighbors (k-NN)

k-Nearest Neighbors (KNN) is a simple, non-parametric, instance based learning, which classifies or predicts a target sample based on majority class or mean coordinate in the feature space of k nearest neighbors. Under localization context, k-NN attempts to estimate the nodes counter parts by computing eucilidian distance between the observed feature vector and those in the training set. Although its simplicity of implementation and low training cost, k-NN may shed accuracy to high-dimensionality or sparse and noisy of training data.

Support Vector Machines (SVM)

The Support Vector Machines are strong types of supervised learning systems that build hyperplanes in multi-dimensional space to partition the data points using to the largest interval of separation. When applied in localization, SVM can be applied on regression (SVR) whereby it predicts continuous spatial coordinates given signal features. SVM has the ability to describe non-linear relations between input variables and location by involving kernel functions including radial basis function (RBF). Nevertheless, SVM can be sensitive to tuning of the parameter and can prove not to scale up when presented with large data as the algorithm is computationally complex.

Random Forest (RF)

Random Forest is a number of learning methods in which the predictions of a number of decision trees trained on subsets of the data are combined. Every tree adds up to a vote and average voting is used to decide the ensemble decision in regression task. RF has high capacity to perform good generalization, does not perform as poorly as individual trees in dealing with non-linearity and feature noise and is less likely to overfit. The fact it is easy to interpret and has moderate training cost makes it suitable to the real-time localization conditions of the IoT, where feature inputs are diverse and partly redundant.

Deep Neural Networks (DNN)

Deep Neural Networks are feedforward neural nets with multiple layers, they can learn hierarchies of complex representations as used in input data. DNNs automatically learn spatial patterns and are able to generalize on noisy or heterogeneous environments during localization activities. DNNs are typically constructed out of numerous fully interconnected layers consisting of non-linear activation functions (e.g. ReLU) and are optimized through backpropagation and stochastic gradient descent (SGD). Although DNNs outperform conventional models in accuracy, they require considerable computing power, training duration and need of large labeled data-sets, aspects that should be factored in implementing DNNs in resource limited IoT systems.

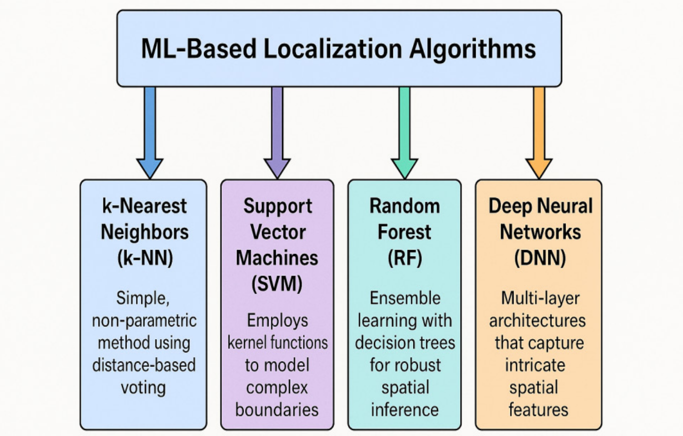
ML-Based Localization Algorithms
k-Nearest Neighbors (k-NN)
Simple, non-parametric method using distance-based voting
Support Vector Machines (SVM)
Employs kernel functions to model complex boundaries
Random Forest (RF)
Ensemble learning with decision trees for robust spatial inference
Deep Neural Networks (DNN)
Multi-layer architectures that capture intricate spatial features

Fig. 2. Overview of Machine Learning Algorithms for Localization in Dense IoT Sensor Networks

The four machine learning algorithms tested in the context of node localization are summarized in the flowchart, i.e., k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forest (RF), as well as Deep Neural Networks (DNN). Each of the algorithms is presented along with its principle of operation and the purpose of estimating node positions using the features deriving out of signals.

Evaluation of the present four algorithms on the basis of localization accuracy, robustness to noise, time delay and the ability to scale to such dense deployment situations are presented in the following sections. This comparative analysis gives ideas about their adaptability in real-life use in IoT solutions considering different levels of the performance requirements.

Experimental Setup

A full experiment based simulation framework was developed in order to simulate realistic dense IoT sensor network scenarios to test the performance of chosen machine learning algorithms in a controlled and repeatable operational profile. Figure 3 is a visual representation of the entire setup of the simulation, the design, selection of features, tool used and the evaluation measures.

Simulation Environment

The spatial field, in which the simulation was conducted, is two-dimensional with a size of 100 m x 100 m, which is typical of an area of deployment of smart infrastructure or industrial Internet of Things. One hundred and fifty sensor nodes were placed in the field by using uniform random distribution to use a dense deployment setting. Out of these, we had 20 percent of nodes which were called as anchor nodes and had a known fixed coordinate, these were reference point against which localization took place. The other 80 percent were converted to unknown nodes, positions of which were to be estimated via machine learning models.

Feature Set

A sequence of signal-based features were gathered per unknown node utilized to train and assess the ML models. The feature set contained:

- Neighboring anchor nodes gave me Received Signal Strength Indicator (RSSI) values,
- The topology-aware context would be node identifier (ID),
- Empirical path loss models that were used to estimate distances based on RSSI.

Such characteristics are realistic signal features that can be observed in the development of IoT fields, including Gaussian noise added to simulate environmental volatility and signal distortions.

Tools and Frameworks

Using the following, the simulation environment and the implementation of the ML-model were constructed:

- Python with library scikit-learn (classical ML: k-NN, SVM, RF) and TensorFlow (deep learning or neural networks building)
- MATLAB used in signal propagation simulation, generation of RSSI data it uses stochastic channel models.

Datasets to train and test were divided into 80:20 percentages and 5-fold cross-validation was applied to make sure that the model performs well under various runs.

Evaluation Metrics

All algorithms were estimated on the following main parameters of performance:

- Localization Accuracy: the localization accuracy is measured in meters as a Mean Absolute Error (MAE) between estimated and true node coordinates,
- Runtime: It showcases time taken to inference on test set and thus efficiency of the computation,
- Noise robustness: The change of performance through the noise in the signal and any fluctuations, which was evaluated, by adding the variable amount of gaussian noise (AWGN) to the input features.

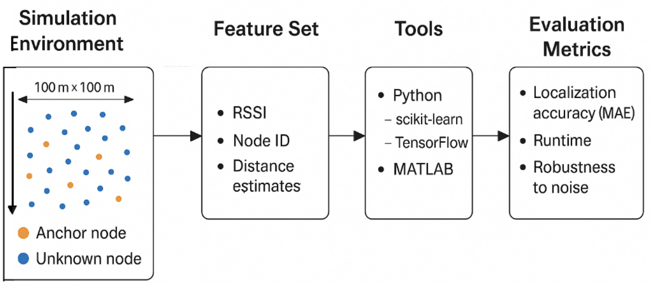

Fig. 3. Experimental Setup for ML-Based Localization in Dense IoT Sensor Networks

The above diagram presents the experimental set up to test the machine learning algorithms to determine the node localization. It contains a 100 m x 100 m simulation environment with anchor and unknown nodes, feature extraction (RSSI, Node ID, distance estimates and implementation tools (Python with scikit-learn and TensorFlow, MATLAB),) and performance metrics (accuracy of localization, time, aspect noise).

Such an experimental framework creates a common ground on which to compare ML algorithms against each other under controlled comparative settings and gives an accurate and replicable performance measurement.

Results and Discussion

The following section outlines the comparative review of the four chosen machine learning algorithms, i.e. k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Random Forests (RF), and Deep Neural Networks (DNN) depending on their performance in operating under dense deployment of IoT sensors. The models were evaluated over four performance measures that include localization accuracy, inference time, signal noise resistance, and scaleability. Quantitative findings are reported in Table 1 whereas the respective visual comparison is reflected in Figure 4 (Localization Accuracy using Mean Absolute Error) and Figure 5 (Inference Runtime per ML algorithm).

Localization Accuracy

With assessment performance, Deep Neural Networks (DNN) demonstrated the best results in localization, having a Mean Absolute Error (MAE) of 2.58 meters because of fitting intricate, multi-level representations of the spatial styles of the signals they learned. The performance illustrates effectiveness of deep learning in mining latent features in high density noisy environments.

Runtime and Computational Efficiency

k-NN registered the shortest time execution (0.14 s) because it is non-parametric and training free. Nevertheless, it did not perform well with regard to accuracy and sensitivity to noise. Conversely, DNNs, as the most accurate, had the most inference time (1.34 s) because of deep-layer calculations a significant limitation to real-time or limited resources deployments.

Robustness to Noise

Both DNN and RF were extremely resistant to noise, preserving their accuracy consistent even after applying Gaussian noise to the features of RSSI. SVM also fared well on moderately noisy settings but they were sensitive to parameters tuning and choice of kernel and k-NN, which does not allow feature generalization, experienced the sharpest loss of performance with noise variance.

Scalability

To test the scalability of the model, nodes density was increased and the model response was tracked. k-NN performed the scale test well since it was simple, however traded-off accuracy. In comparison, SVM experienced a significant decline in the runtime and the accuracy as the size of the dataset increased and demonstrated the lack of scaling potentials. RF and DNN

Table 1. Performance Comparison of ML-Based Localization Algorithms

Algorithm	Mean Absolute Error (m)	Runtime (s)	Robustness	Scalability
k-NN	4.87	0.14	Moderate	High
SVM	3.95	0.65	High	Low
RF	3.67	0.48	High	Moderate
DNN	2.58	1.34	Very High	Moderate

offered a fair trade-off of scalability and performance, so they could be applicable in large-scale deployment with edge/cloud computation assistance.

Summary of Insights

- DNN is the best option in cases where the priority on accuracy and robustness are the most vital and computing facilities are plentiful.
- RF is a compromising solution between accuracy, speed and noise resistance, which is perfect to consider moderately constrained IoT purposes.
- SVM can be more used in the small-scale, unchangeable networks however not applicable in large-scale dynamic systems because it has less scalability.
- k-NN is also a useful lightweight with low-latency applications but is not adaptive with high-density networks and noisy networks.

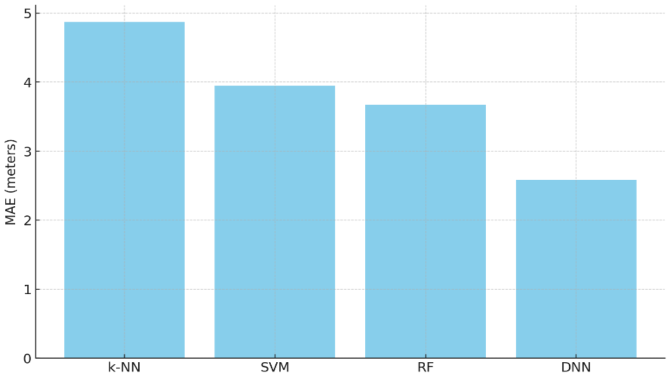

Fig. 4: Comparison of Localization Accuracy Across ML Algorithms (Mean Absolute Error in Meters)

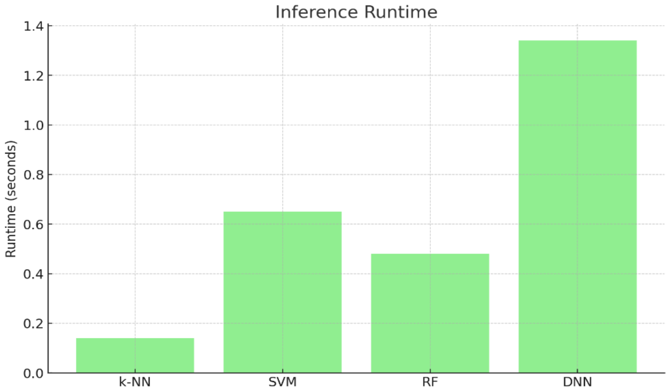

Fig. 5: Inference Runtime Comparison of ML Algorithms for Node Localization(in Seconds)

Conclusion

To highlight the effectiveness of ML-driven localization solutions in solving the specific problems of IoT dense sensor deployment, this comparative analysis is carried out. One of the most accurate localization results was reported as Deep Neural Networks (DNN), which succeeded in preserving complicated spatial features, based on the signal produced inputs. Nonetheless, the training/inference overhead make DNNs poorly suited to traditional networked infrastructures, lacking the availability of intensive computing offloading to an edge-assisted or cloud-integrated infrastructure. By comparison, Random Forest (RF) presents a strong tradeoff between accuracy, resilience, and the computational overhead, thus being specially suitable when localization needs to be done in real-time and on constrained IoT nodes. The ensemble design used by RF combines robust generalization along with modest width needs even during signal noise and variability.

The presented findings indicate that the use of ML algorithm should be correlated with the context of deployment and the system limits. DNNs are suited in applications where accuracy is of the essence, and computational budgets are very high, but RF offers a practical solution of scalable and costeffective localization in embedded IoT settings.

Future Work

Whereas, the current paper provides an excellent estimation of machine learning-based localization algorithms in dense IoT surroundings, there are still a number of prospects of enhanced research in the field that are to be pursued in the future.

The first one is the potential introduction of federated learning (FL) into the localization models to provide an attractive method of data privacy and the minimization of communication overhead. Without revealing the raw data, FL will potentially improve the privacy of data and scale to mission-critical services or sensitive environments that follow a decentralized approach and require limited support due to the training performed at distributed points on the IoT.

Second, online and gradual learning models, along with their development, is crucial to changing environments, in which network topology and condition of signal propagation can change over time. With such models, real-time adaptation and continual learning would be made possible, thus less frequent retraining is required offline and localization stability to environmental drift would be better.

Lastly, in future, hybrid localization systems which integrate classic approaches to signal processing (e.g. time-of-arrival estimation, triangulation) with machine learning architectures should be considered to capitalize on the advantages that the two distinct paradigms offer. These hybrid systems may provide better interpretability, better performance with the training of small training cases, and generalization in the heterogeneous environments.

These future directions, in combination, are intended to expand the possibilities of intelligent localization and to make it safer, responsive, and scalable to the future realm of dense IoT installations.

References

1. Shahbazian, R., Macrina, G., Scalzo, E., & Guerriero, F. (2023). Machine learning assists IoT localization: A review of current challenges and future trends. Sensors, 23(7), 3551. https://doi.org/10.3390/s23073551
2. Maduranga, M. W. P. (2024, February 18). Improved indoor localization with machine learning techniques for IoT applications. arXiv. https://arxiv.org/abs/2402.11433
3. Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 37(6), 1067–1080. https://doi.org/10.1109/TSMCC.2007.905750
4. He, T., Huang, C., Blum, B. M., Stankovic, J. A., & Abdelzaher, T. (2003). Range-free localization schemes for large scale sensor networks. In Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom) (pp. 81–95). San Diego, CA, USA. https://doi.org/10.1145/938985.939000
5. Youssef, M., & Agrawala, A. (2005). The Horus WLAN location determination system. In Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys) (pp. 205–218). Seattle, WA, USA. https://doi.org/10.1145/1067170.1067197
6. Uvarajan, K. P. (2025). Design of a hybrid renewable energy system for rural electrification using power electronics. National Journal of Electrical Electronics and Automation Technologies, 1(1), 24–32.
7. Abdullah, D. (2025). Environmental sound classification using CNNs with frequency-attentive acoustic modeling. National Journal of Speech and Audio Processing, 1(1), 8–14.
8. Velliangiri, A. (2025). An edge-aware signal processing framework for structural health monitoring in IoT sensor networks. National Journal of Signal and Image Processing, 1(1), 18–25.
9. Kulkarni, S., & Ravi, J. N. (2024). Smart Ways to Catch the Abutment DRCs at IP Level. Journal of VLSI Circuits and Systems, 6(1), 51–54. https://doi.org/10.31838/jvcs/06.01.08
10. Yadav, N. P. (2019). A four element antenna array for amateur radio applications. National Journal of Antennas and Propagation, 1(1), 13–16.