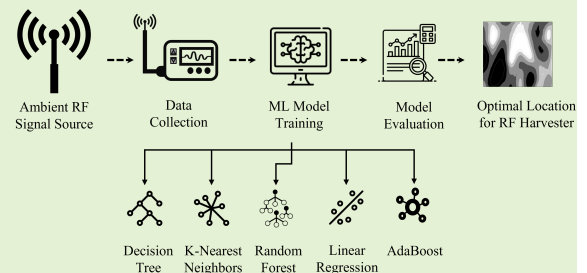


Machine Learning-based Predictive Modeling Analysis in Ambient RF Energy Harvesting for IoT Systems

Khatereh Nadali, *Graduate Student Member, IEEE*, Adnan Shahid, *Senior Member, IEEE*, Sam Lemey, *Member, IEEE*, Patrick Van Torre, *Member, IEEE*, and Max J. Ammann, *Fellow, IEEE*

Abstract—The Internet of Things (IoT) has already ingrained itself into our daily lives, with the number of connected devices that are growing rapidly. Particularly, low-power wireless sensing devices are anticipated to make significant contributions to this expansion. These compact devices are designed to operate for an extended duration, spanning years or even decades, but the growing demand for such devices poses challenges in terms of ensuring sustainable power supply. In order to sustainably power these devices, ambient radio frequency (RF) energy harvesting has emerged as a possible approach. However, placing the harvester in an optimal location is essential to maximize the reception of ambient RF energy and ensure reliable performance. In this paper, we investigate the **estimation** of the ideal location for RF energy harvesting by utilizing machine-learning (ML) techniques in real-world scenarios. The study involves a frequency-dependent analysis and a received signal intensity analysis. A comparison of **three different interpolation methods** with five supervised ML algorithms is conducted, and the effect of reduced measurement points on **estimation accuracy** is evaluated. The outcomes demonstrate how well ML **estimates** the optimal location for energy scavenging and offer insights into creating sustainable energy systems.

Index Terms—Internet of Things, Machine Learning, Supervised Learning, Prediction Modeling, Energy Harvesting.



I. INTRODUCTION

THE Internet of Things (IoT) is part of our everyday life, and we utilize digital tools increasingly to interact with the rapidly growing digitized environment. Projections for the growth of IoT devices vary widely. Intel **anticipated** 200 billion devices by 2020, while IHS Markit forecasted 75.4 billion by 2025. A revised forecast from Cisco estimated the number of connected devices to be between 25 billion and 125 billion by 2030 [1], [2]. The advancement of IoT technology has led to smaller, more cost-effective, and energy-efficient low-power wireless sensing devices that allow us to monitor and regulate the environment. However, as the number of these sensors increases, the IoT network faces challenges concerning the power supply of the devices, including limited battery capacity, complex charging processes, and high replacement costs. Ad-

This paper was submitted for review on 09/June/2023. This paper has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant numbers 18/CRT/6222 and 13/RC/2077-P2. Part of this research was funded by the Flemish FWO SBO S001521N IoBaLeT (Sustainable Internet of Battery-Less Things).

Khatereh Nadali (e-mail: khatereh.nadali@tudublin.ie) and Max J. Ammann is with the AHFR center, Technological University Dublin, Ireland (e-mail: max.ammann@tudublin.ie) are with the AHFR center, Technological University Dublin, Ireland.

Adnan Shahid, (e-mail: adnan.shahid@ugent.be), Sam Lemey (e-mail: sam.lemey@ugent.be), and Patrick Van Torre (e-mail: patrick.vantorre@ugent.be) are with IDLab, Department of Information Technology at Ghent University - imec, Ghent, Belgium .

ditionally, the growing number of low-power sensing devices in IoT raises concerns about ensuring their adequate and sustainable power supply as the IoT ecosystem expands [3]. To address this issue, ambient radio frequency (RF) energy harvesting has emerged as a promising technology that can power ultra-low-power IoT devices in a sustainable way.

Since IoT networks operate across diverse frequency bands, there is a potential to utilize different ambient RF sources as a power source for self-sustainable IoT devices. For instance, a temperature sensor connected to a smart home network can harvest energy from ambient Wi-Fi signals emitted by nearby devices, while a similar sensor connected to a smart agricultural network could harvest ambient cellular signals from nearby cellular towers as a source of energy. The amount of RF energy that can be harvested from the environment varies depending on the position, direction, and orientation of the receiving nodes, making it challenging to create a consistent and reliable IoT system that operates on harvested RF energy [4].

To effectively harvest RF energy from such sources, the use of multiband and wideband receiving antennas is becoming increasingly important. Besides, the positioning of the harvester device plays a crucial role in increasing the amount of captured energy. As the ambient RF energy that can be harvested from the surrounding environment is typically very low, **varying** from micro-watts to milli-watts [5], [6],

capturing every bit of energy is essential. The incident waves from ambient sources may encounter obstacles, resulting in reflections and diffraction that attenuate the received signal. Therefore, positioning the harvester in an optimal location reduces these effects to ensure maximum energy reception, leading to increased harvesting efficiency.

The best location for an RF energy harvester can be determined using a number of methods. One method, for instance, is to measure the signal strength at various points throughout a specific area using a spectrum analyzer or power meter. This method may not be reliable, though, as it just provides information on the signal strength at particular points rather than a comprehensive evaluation of signal propagation across the entire region. Another approach involves using mathematical models to investigate how radio waves reflect, diffract, and scatter in order to pinpoint the areas with the strongest signals. The complexity of signal propagation in indoor spaces, which includes man-made noise, objects, and people present in the space, causes the analysis to become more difficult and makes this method extremely challenging [7], [8]. The study can also be impacted by the frequency range of interest, i.e., higher operating frequencies result in larger propagation loss, multipath, and attenuation due to the absorption by surrounding objects.

The quality of communications in energy harvesting applications is a complex and evolving research area. In [9], the authors studied the problem of how to provide reliable communications for energy-harvesting wireless sensor networks and the analyses have been verified by numerical simulation. A comprehensive paper [10] provided an overview of the state-of-the-art in energy harvesting wireless networks, highlighting that the quality of communications in such networks depends on several critical factors, including channel conditions, energy arrival patterns, energy consumption models, network topology, and network objectives. However, our work focuses on a specific aspect of this broader field, which is demonstrating the potential of ML in finding the optimum location for an ambient RF energy harvesting device that has not been investigated in the literature, and we believe it contributes to the growing body of knowledge in energy-harvesting wireless communications.

In this paper, we analyze the potential of ML to estimate the optimal location of an RF energy harvester. With this method, the optimal location can be determined in an automated way on a limited set of observations, not only speeding up deployment but also ensuring the optimal position. Utilizing ML algorithms as a tool facilitates optimizing the performance of the energy harvester and maximizing energy harvesting efficiency, thus contributing to a more sustainable and energy-efficient environment.

The main contributions of the work are:

- Investigation of received signal intensity considering the small movement of people in the test environment,
- Frequency-dependent analysis of the received signal intensity in a controlled environment,
- Providing a comparison of three interpolation methods along with five supervised ML algorithms to identify the most effective one for estimating the optimal energy harvesting location in a real-life deployment scenario,

- Assessment of the impact of reducing the number of measurement points on the accuracy of estimating the optimal location for energy harvesting.

The remainder of this paper is structured in the following manner: Section II introduces the related research that has been conducted. Section III examines the system for harvesting ambient RF energy and provides a detailed description of the measurement setup. Section IV explores the utilization of various machine learning algorithms. Section V presents the numerical findings, along with an interpretation and analysis of the results. Section VI presents a specific application example, whereas Section VII concludes this work.

II. RELATED WORKS

Cost-effective energy prediction techniques are frequently implemented to accommodate restricted computing in low-power IoT systems. The exponentially weighted moving average (EWMA) prediction technique is one of the widely used techniques. EWMA computes the harvested energy at a given time by taking a weighted average of the energy collected at that time over the previous days. In [11], the authors created efficient distributed methods to utilize harvested energy and understand the changing nature of energy sources. They utilized an EWMA filter prediction model that accounts for solar energy patterns and seasonal variations. The authors in [12] introduced a weather-conditioned moving average (WCMA) prediction algorithm based on the EWMA algorithm for solar-based energy harvesting. Unlike EWMA, WCMA takes into account seasonal fluctuations such as changes in sunrise and sunset times and variances in solar intensity. Despite the works that have been done on modeling and predicting solar, wind, and other natural sources of energy harvesting, these sources have limitations due to their dependency on time and weather conditions. Ambient RF energy harvesting systems are distinct from other types of energy harvesting techniques, as they do not depend on nature and its variable elements, such as sunlight, wind, or temperature. Over the past years, RF energy harvesting has increasingly drawn attention, as the number of wireless devices and resulting RF signal emitters have been gradually increasing [13] and subsequently, RF energy is more accessible in indoor environments. Ambient RF energy harvesting can be utilized in low-power wireless devices due to the ease of their integration into small devices. However, the amount of harvested RF energy is determined by various factors such as the distance between the source and harvester, the frequency of the RF signal, interference from other signals, and environmental and operational conditions [14].

In recent years, ambient RF energy harvesting has been studied in the literature. In [15], the authors highlighted the importance of reducing power consumption for IoT devices and used a technique to achieve self-sustained sensors by significantly lowering power requirements. In [16], researchers investigated multi-input, multi-output (MIMO) wireless power transfer (WPT) to manage the spatial and temporal distribution of wireless power. They explored the idea of a wireless power system akin to data networks, capable of selectively delivering power to enhance energy harvesting for mobile

devices in optimized regions. Prediction/estimation of the RF energy pattern is essential to ensure the efficient operation of self-sustainable devices. Specifically, in [17] the low power density of ambient RF signals as a challenge associated with harvesting and efficiently storing RF energy was reviewed. The study was carried out in [18], to examine the possibility of harvesting ambient RF energy in both urban and semi-urban environments by surveying 270 underground stations in London. The findings showed that the amount of RF energy available in underground stations was significantly influenced by the station's position and proximity to signal sources such as Wi-Fi access points (APs) and cellular base stations. The application of ML in modeling RF energy has been explored in a very small number of research works. Authors in [19] compared two ML techniques, linear regression (LR) and decision trees (DT), to determine their effectiveness as predictive models for estimating the amount of RF energy that can be harvested from an RF power source, using measurements of different frequency bands for the period of several days. The work in [18] focused on utilizing ML algorithms to model the prediction of time-series RF energy data for efficient energy-harvesting communication systems. In [20], four ML algorithms were tested for time-series RF energy data estimation, with LR having the highest accuracy and stable performance, followed by the support vector machine (SVM) and random forest algorithm. Among them, the DT was the worst model as DT displays a greater error range, which indicates a significant fluctuation in its predictions across various data subsets and less consistency and stability in its performance. In [21], a new statistical model was developed for predicting the power density of mobile service channels using kernel density estimation and moving average techniques. The model features an accuracy of over 80% and adjusts the sampling frequency based on the channel power prediction.

According to the best of our knowledge, this is the first paper that presents an ML solution for estimating the optimal location for energy harvesting in a real environment. This is a significant contribution because traditional methods for determining the ideal position for an energy harvester, such as analytical models or simulations, may not be accurate in real-world circumstances due to the complexity and diversity of the environment. By using ML techniques, we are addressing a gap in the literature and also presenting a potential solution for the optimized positioning of an energy harvester device in real-world scenarios, which has practical implications for the development of sustainable energy systems.

III. SYSTEM DESCRIPTION

A. RF Energy Harvesters

A rectenna is required to transform RF signals into a usable form of energy to power the electronic parts of a harvester. A rectenna consists of an antenna and an RF-DC converter. Microwatts to milliwatts of power can be captured by a rectenna and be transformed into usable energy, depending on a number of variables such as the input power of the source, the distance from the source, and the efficiency of the power conversion [17]. Besides, a variety of factors must

be taken into account when designing the rectenna for use in low/ultra-low power RF energy harvesting applications, including the performance of the antenna and its resonance frequency, the effectiveness of the matching network and its circuitry components, and the distance between the harvester and the ambient RF source are some of these variables. Power conversion efficiency (PCE) is an important factor to consider when assessing a rectenna and it refers to the proportion of output power, P_{out} , to the power received by the antenna, P_{in} and can be calculated by

$$PCE = \frac{P_{out}}{P_{in}} = \frac{(V_{DC})^2}{R_{load}} \cdot \frac{1}{P_d A_{eff}}, \quad (1)$$

where V_{DC} represents the output DC voltage, R_{load} represents the load resistance, P_d is the power density distribution across the receiver aperture and A_{eff} is the effective area of the receiving antenna. In ambient RF energy harvesting, the transmitted power level is beyond our control and greatly impacts the available power density. However, the receiving antenna performance and efficiency of the rectifier are controlling factors that can be optimized to maximize the power reception for a certain operating condition. In this work, we opted for a circularly polarized antenna as the transmitting and receiving antenna with the aim of maximizing power reception. Circularly polarized antennas are commonly preferred in energy harvesting applications due to their reduced sensitivity to the polarization of the incident signal. The circular polarization characteristic of an antenna minimizes the effects of polarization mismatch and multipath interference, ultimately leading to higher and also more stable energy harvesting efficiency.

B. Propagation

The frequency of the signal, the distance between the transmitter and receiver, the presence of obstacles, and the atmospheric conditions are just a few of the factors that influence how far RF signals travel and define the signal range. In general, in the UHF and microwave bands, path loss (PL) and the presence of multipath propagation are the dominant factors contributing to signal attenuation. In an indoor setting, wireless signals may encounter objects and barriers, causing them to reflect, scatter, or diffract. As a result, the antenna receives signals through multiple paths, leading to multipath fading. Multipath can be time-varying, due to the movement of obstacles such as people. Path loss is the attenuation of signal strength caused by increasing distance, often measured in dB, and it can be calculated using

$$PL \text{ (dB)} = 10\alpha \log_{10}(d) + L_0, \quad (2)$$

where α represents the PL exponent, d denotes the distance between the transmitter and the receiver in meters, and L_0 represents a frequency-dependent constant that accounts for system losses [22], [23]. In indoor environments, the value of α varies from 0.8 to 1.8 for line-of-sight (LOS), and it can increase significantly up to 8.6 in non-line-of-sight (NLOS) scenarios [24].

There are mathematical models based on empirical data and physical laws to compute the received signal by taking path

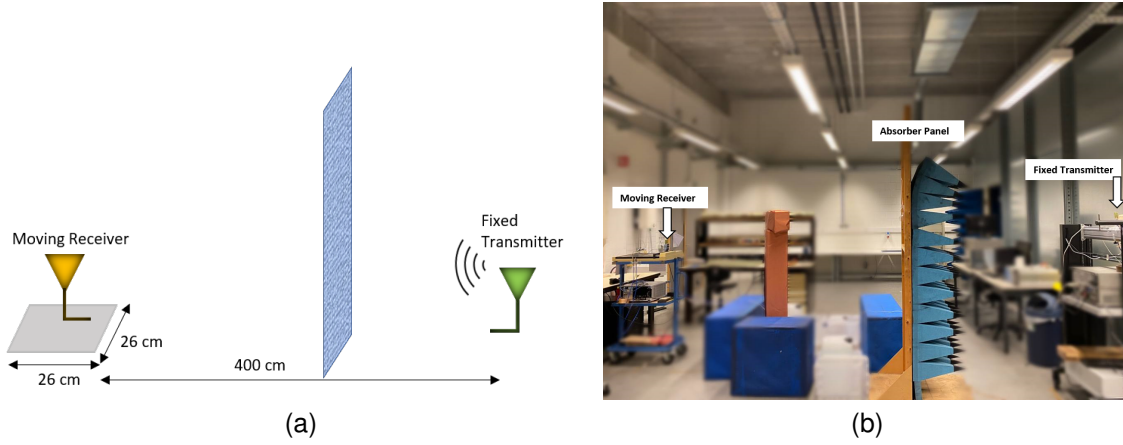


Fig. 1: Test setup, (a) measurement configuration, (b) test environment in Ghent University's laboratory.

loss, refraction, and diffraction into account. Friis transmission equation is one of the commonly used models for free space communication in LOS environments, in which the PL exponent is 2, i.e.,

$$P_R = (\text{PLF}) \cdot \frac{(P_T G_T G_R)}{(4\pi d \lambda)^2}, \quad (3)$$

where P_R is the power level at the receiver with antenna gain G_R , assuming the total power P_T transmitted from an antenna with gain G_T in free space, with a wavelength λ corresponding to the frequency of operation, while the distance between the transmitter and receiver is d . To take into account the difference in polarization between the transmit and receive antennas, it is necessary to multiply the received power by the polarization loss factor (PLF). Accurately determining received power in real-world scenarios is challenging due to the lack of information on transmitter characteristics, as well as factors such as interference, obstacles, and multipath propagation.

C. Measurement Setup and Data Collection

Fig. 1 shows the measurement setup configuration and the test environment. In our experiments conducted at Ghent University's laboratory [25], we deliberately chose a controlled indoor environment to ensure replicability and enable systematic comparisons of various ML algorithms and interpolation techniques. While the controlled environment has been selected for its manageable conditions, we introduce real-world complexities. This includes indoor propagation challenges with man-made noise, obstacles, and movement of a person. The laboratory setup features obstacles, and the receiver moves within the area of interest, simulating dynamic network scenarios. These additions aimed to mimic real-world challenges, enhancing the credibility and transparency of our work. In all the measurements a network analyzer generates RF signals at discrete frequencies. In this research, we use the antenna proposed in [26] due to its wide bandwidth, circularly polarized characteristics, and ease of fabrication design, which make it an appropriate candidate for our work. The transmit antenna is a circularly polarized printed slot antenna with

a 3dB-axial-ratio bandwidth (from 1.7 GHz to 2.69 GHz) and a maximum measured gain of 3.7 dBic at 2.2 GHz. The geometry of the antenna and its performance details are described in [26]. To receive signals, an identical antenna is positioned at a distance of 4.0 meters from the transmitter. The receiver is mounted on a plotter moving from one point to the next with a step-size of 2.57 mm, covering an area of $26 \times 26 \text{ cm}^2$ on the plotter. In this study, measurements are conducted at various frequencies in a frequency span from 2.35 to 2.50 GHz. The receive antenna is connected to a spectrum analyzer to measure and record the received power level. $P_f^{(i,j)}$ represents the measured power level at the frequency f while the receive antenna is located at point (i, j) , where $i = 0, 1, \dots, 100$ and $j = 0, 1, \dots, 100$. Fig. 2 shows the receive antenna mounted on the Roland DXY-1200 plotter. The yellow line represents its movement path as it collects data at its stopping points. The spectrum analyzer's sweep time, which measures the received power level at each point, is configured to be 800 milliseconds. Additionally, there is a communication delay of 500 milliseconds, and a sleep time of 200 milliseconds is introduced to ensure smooth antenna movement between measurement points. As a result, the total measurement time at each point is 1.5 seconds, and the duration of each measurement run amounts to 4 hours and 15 minutes for all the 10,201 (as total points in a 101×101 grid) measurement points.

IV. PROPOSED METHODOLOGY

In practical situations, various factors such as interference, obstacles, movement of people, and multipath propagation lead to a complicated distribution of power levels, making it difficult to determine the received power level accurately. Besides, there is no knowledge of the total value of power that is delivered to the ambient RF source, nor any information about the gain of the transmitter and the distance from the source. ML algorithms can determine the optimal location for an RF energy harvester in a room using extensive data, enhancing performance and maximizing energy harvesting efficiency.

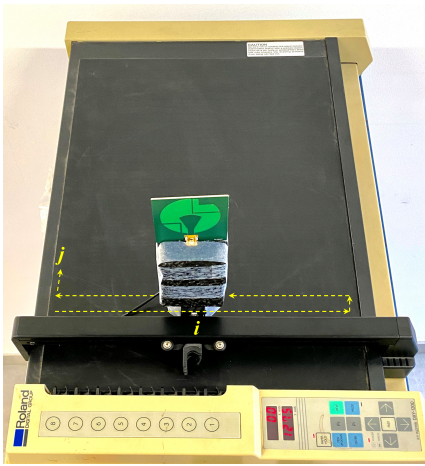


Fig. 2: The receiving antenna mounted on the plotter for taking measurements at a grid of 101×101 points.

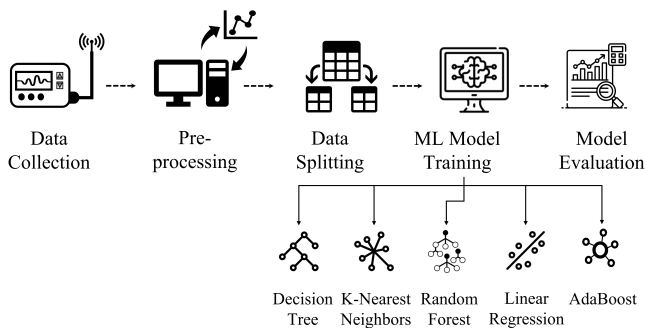


Fig. 3: Estimation module's development methodology.

Algorithm 1: Algorithm to compute the MAE

Input: Measured harvested power level in coordinates (i, j) at Frequency f

Output: Optimal location for the energy harvester

- 1: Initialize and pre-process the dataset.
 - 1-1: Filter the harvested power level values to reduce the presence of noise
 - 1-2: Normalize the numerical features
 - 2: **for** the target amount of data from 100% to 5% **do**
 - 3: Split dataset to training, validation, and test set
 - 4: **for** Random State in 1 .. 100 **do**
 - 5: Train the model
 - 5-1: Train a decision tree regressor
 - 5-2: Train a random forest regressor
 - 5-3: Train an Adaboost regressor
 - 5-4: Train a linear model regressor
 - 5-5: Train a K-nearest neighbors regressor
 - 6: Compute MAE for each model
 - 7: Evaluate the performance of each model
-

Machine learning (ML) is a technique that allows computers to learn and improve performance on a given task by leveraging example data and prior knowledge without the need for explicit programming. ML is a widely used method for data analysis that employs statistical algorithms to learn from data and improve task performance. In this study, we utilize a supervised learning framework to train and evaluate an ML model on a dataset consisting of relevant variables. The model is fitted on a training set, while a distinct test set is used to assess the model's effectiveness.

The pipeline of the **estimation** module for the RF power level in a particular area was conducted in a structured and sequential manner, as illustrated visually in Fig. 3.

1) **Data Collection:** The RF power level data in this work were captured inside a research laboratory at Ghent University. In each measurement run, the level of received RF signal at 10,201 points was measured and the measurements were repeated for 8 different frequencies **spanning** from 2.39 GHz to 2.5 GHz.

2) **Data Pre-processing:** This step in the implementation aims to prepare a dataset suitable for training ML algorithms. Since some ML algorithms are sensitive to outliers and noise in the dataset, in this step, all values below -80 dBm are identified as noise. To improve the accuracy of the **estimations** and refine the dataset, all values lower than -80 dBm are substituted with -80 dBm. In addition, to prevent any single feature from overpowering the learning process due to its scale, all the numerical features are normalized to a similar scale, using the Scikit-learn library (`sklearn`).

3) **Data Splitting:** ML algorithms commonly split datasets randomly into training/validation and test sets. This division allows for evaluating the ML model and assessing its accuracy. The model is trained entirely on training data and then tested on a separate test set to determine its validity. **Our intention in this research is to ensure that the training, validation, and test points are evenly distributed across the entire region of interest. This approach was chosen to prevent the ML algorithms from overlooking any specific region and to provide a comprehensive estimation of signal strength across the area of concern. To achieve this, a stratified random sampling technique is employed to divide the data into training, validation, and test sets, ensuring a balanced distribution of variables in the sets. In this respect, we split the dataset into randomized 70 : 15 : 15 training, test, and validation sets using Python, where 30% of the data is used for the test/validation set.**

4) **ML Model Training:** The module for ML model training utilizes ML algorithms. To determine the most effective model, five different algorithms are tested. This task is considered a regression problem since the aim is to **estimate** the RF power level and determine the optimal position for an RF energy harvester, involving the estimation of a continuous variable. In order to address this regression problem, well-known ML algorithms including linear regression, K-nearest neighbors, random forest, decision tree, and AdaBoost are employed to establish relationships between variables and generate **estimations** on the test set. As hyperparameters control the behavior of the algorithm, we can optimize the performance of the model and enhance its ability to make accurate **estimations** on unseen data, by tuning hyperparameters.

The Big O notation can be used to assess the complexity and scalability of the aforementioned algorithms. Among the methods listed, KNN has the lowest training complexity, with a constant time complexity of $\mathcal{O}(1)$ during training. This means that KNN does not require significant computing resources during training. It should be noted, however, that the prediction/**estimation** complexity of KNN is considerably higher, being $\mathcal{O}(n \times k)$, where n represents the number of training samples and k denotes the number of neighbors

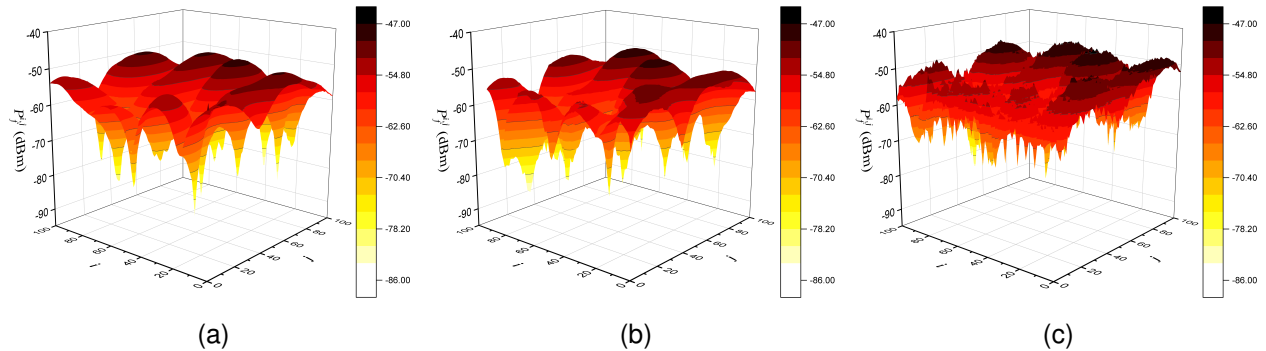


Fig. 4: Received power level at the frequency of 2.35 GHz, (a) no one in the test room, (b) a person stays still in the room, (c) a person moving constantly in the test room.

considered. The training complexities of LR and DT are the same. Both algorithms have an $\mathcal{O}(n \times m^2)$ training complexity, where m indicates the number of features. However, both LR and DT can be considered to have a similar prediction complexity of $\mathcal{O}(m)$, indicating that the computational resources required for prediction scale linearly with the number of features. AdaBoost and random forest are the most complex algorithms in terms of training and prediction complexity. Overall, the random forest has a higher training complexity of $\mathcal{O}(T \times n \times m^2)$, where T is the number of trees in the forest. The prediction complexity for the random forest is $\mathcal{O}(T \times m)$, indicating that the computational resources required for prediction scale linearly with the number of features and the number of trees in the forest. It is important to consider other factors such as the interpretability of the model, and the desired level of accuracy when determining the most suitable approach. This paper provides further analysis aimed at identifying the optimal ML algorithm for this particular study.

The Python programming language was utilized to implement the models, using the Scikit-learn library (`sklearn`) for tasks such as model construction, training, and evaluation. All of the experiments and evaluations were conducted on a PC equipped with an Intel(R) Core(TM) i7-9700 CPU.

5) *Model Evaluation*: After training the ML models on the training set and generating estimations on the test set, the next step is to evaluate the performance of the models. In this study, the evaluation is based on the mean absolute error (MAE) metric which is a commonly used metric for regression problems. MAE is a statistical measure used to evaluate the differences between paired observations that express the same phenomenon [27]. We use MAE to evaluate the accuracy of the estimated value of the received power level, using each model. The MAE is given by

$$\text{MAE} = \frac{1}{n'} \sum_{q=1}^{n'} \left| P_f^{(q)} - \hat{P}_f^{(q)} \right|, \quad (4)$$

where n' represents the number of test points, $P_f^{(q)}$ represents the measured power level at frequency f at a selected testset point q , and $\hat{P}_f^{(q)}$ represents the corresponding estimated value. In this study, the objective is to find the optimum algorithm to

minimize the MAE value in order to find the optimum location for the energy harvester device.

V. NUMERICAL RESULTS AND DISCUSSION

A. Data Collection at a Single Frequency

In this study, we initially conducted measurements of the received signal power using a signal generator and a fixed CP antenna with a realized gain of 3.4 dBic at a frequency of 2.35 GHz [26] on one side and a moving receive antenna identical to the transmit antenna with a spectrum analyzer for power measurement on the other side. An absorber stand was placed between the two antennas to block LoS signals and force the receiver to receive signals from indirect paths. We conducted the measurements under controlled conditions, including three different scenarios: a) no one in the test room, b) a person in the room with minimal movement, and c) a person moving constantly around the room, particularly between the two antennas. Fig. 4 depicts the heatmap of the received power for each scenario. It should be noted that we repeated the measurements at various time intervals throughout the day and discovered no significant changes in the results. As demonstrated in Fig. 4, there were no discernible variations in the received signal intensity when no one was present or when a person was there but remained still. The radio waves that are transmitted interact with objects in the environment, resulting in a complex pattern of reflection, diffraction, and scattering that alters signal intensity at the harvester. The placement of the receiver and transmitter, as well as the surrounding objects, all influence signal strength. If the person in the test setting moves as little as possible, it can be assumed that the items and structures in the room are generally fixed and that the effects of reflection, diffraction, and scattering are relatively constant. The heatmap of the received signal validates our assumption. However, when a person is constantly moving around the room, particularly between the two antennas under test, the signal strength varies due to the changing effects of the surrounding objects and structures on signal propagation. Despite this, the areas with the minimum and maximum chance of reception of the ambient RF signal remained the same, indicating that certain regions within the room consistently exhibited the weakest and strongest reception of the signal, regardless of the presence or movement of individuals.

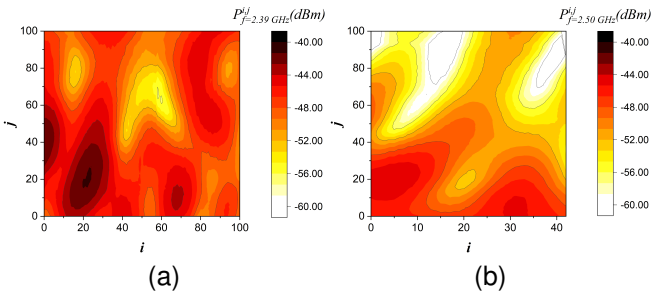


Fig. 5: Heatmap of the received signal at two different frequencies, (a) $f = 2.39$ GHz, (b) $f = 2.50$ GHz.

B. Data Collection at Various Frequencies

In the next step, we maintained a static environment with no one present in the test environment and conducted measurements covering a spectrum of frequencies from 2.39 to 2.50 GHz. This comprehensive study provided valuable insights into the behavior of ambient RF signals within this frequency range. As expected, the intensity of the received signal was found to be dependent on the frequency of the transmitted signal. The changes in the signal power that we measure depend on how objects in the environment affect electromagnetic waves. This phenomenon is intricately linked to the physical properties of the objects that the waves encounter and the characteristics of the electromagnetic waves themselves. These interactions result in different patterns that can either amplify or attenuate the signal, making it crucial to understand the frequency-dependent characteristics of the ambient RF signals. Fig. 5 illustrates the heatmap of the measured received signal over the entire area at two specific frequencies, 2.39 GHz and 2.50 GHz. It is evident from the figure that different frequencies yield distinct patterns of signal distribution, highlighting the dynamic nature of RF signal propagation within the environment. These findings emphasize the importance of frequency-aware analysis when optimizing the placement of energy harvesters. Understanding how signal strength varies across different frequencies enables us to make informed decisions regarding the selection of frequency bands for energy harvesting. Moreover, it highlights the necessity for a robust predictive model that can adapt to frequency-dependent variations in signal strength.

C. Estimation of the Received Signal

1) Single frequency:

a) *Machine learning techniques:* The process of estimating the harvested power at a specific frequency, such as 2.39 GHz, can be accomplished using various ML algorithms. However, in order to determine the most effective algorithm, a comparison was conducted using MAE as an evaluation factor. LR model has a significantly higher MAE of 1.980 in this scenario. This is due to the presence of non-linear relationships between the input variables and output that a linear equation is unable to effectively express. The nonlinearity of the received signal power, as depicted in Fig. 10(a), contributes to this behavior. LR model relies on the assumption that the input variables

and the target variable are linearly related, which makes it unable to accurately estimate the harvested power values for individual points within the area. The limitations of a linear hyperplane become apparent in this context. In comparison, the random forest algorithm outperforms other methods in estimating the harvested power accurately, as it has an MAE of 0.071 for estimating the harvested power at 2.39 GHz. It is a powerful and flexible algorithm that can handle non-linear relationships between the input features and output. Random forest is an ensemble learning method that constructs a multitude of decision trees and combines their estimations to get a more accurate and stable estimation. The AdaBoost and KNN algorithms have comparable performances, with MAEs of 0.120 and 0.105, respectively. However, they are still not as accurate as the random forest algorithm. The DT algorithm is prone to overfitting and may not generalize well to the unseen data. The DT algorithm has an MAE of 0.138, which is higher than that of AdaBoost, KNN, and random forest. This suggests that DT may not be the most appropriate algorithm for this estimation task. The results mentioned above are visually compared in Fig. 6. The values presented in this figure represent the average of all values obtained from a hundred distinct random seeds of training and validation splits. The

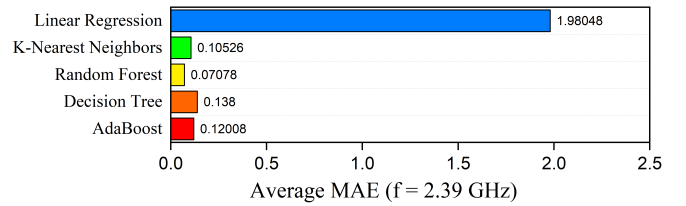


Fig. 6: MAE using different algorithms at the frequency 2.39 GHz.

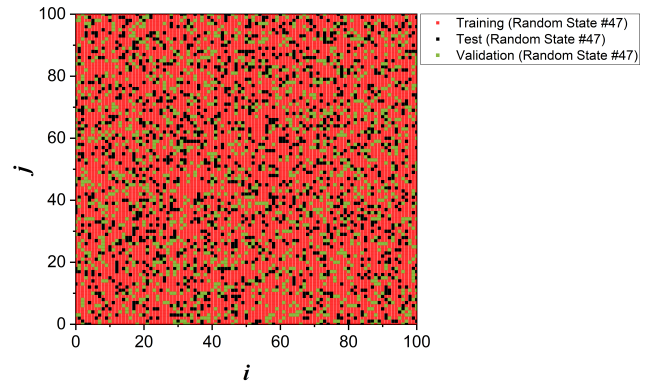


Fig. 7: Distribution of the training, test, and validation points in the measurement region corresponding to the random state number 47.

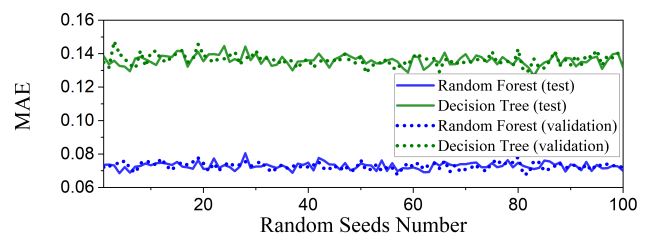


Fig. 8: MAE results for different random seeds.

spatial distribution of training, validation, and test points can have a significant impact on the performance of the prediction model. To mitigate the influence of point distribution on our findings and ensure the reliability of the results, we conducted our experiments using 100 different random distributions of the training and test sets. Fig. 7 provides an illustration of the distribution of the training, test, and validation points in the measurement region corresponding to the random state number 47, and Fig. 8 depicts the MAE achieved through a hundred executions of ML validations for the random forest and DT algorithms. The objective of utilizing multiple random seeds is to account for both the data unpredictability and the possibility of overfitting specific training sets. By comparing the results, it can be observed that the model consistently performs well across all the trials, with similar trends and patterns emerging in each run.

b) Interpolation techniques: Interpolation is a numerical technique to estimate values between data points, helping to provide a continuous representation of data in between discrete data points, where its different methods offer unique trade-offs. Nearest-neighbor interpolation, a straightforward and computationally efficient technique, assigns a value to a target point based on the nearest data point, yet it tends to generate discontinuous results, introducing a stair-step effect and a lack of smoothness. In contrast, linear interpolation constructs straight lines between adjacent data points, making it computationally efficient and widely applicable, but it may struggle to capture complex data variations with curvature between the points. Cubic interpolation, a more advanced approach, utilizes cubic polynomials to create smooth curves through a 2×2 grid of data points, offering superior flexibility and accuracy. However, this sophistication comes at the cost of higher computational complexity, which can impact performance in certain applications.

To evaluate interpolation methods, we applied nearest-neighbor, linear, and cubic interpolation to estimate received ambient RF signal intensity at a frequency of 2.39 GHz. The superiority of random forest over other ML techniques is evident in Fig. 6, hence, Fig. 9 visually compares interpolation methods only with random forest. Despite initial expectations that cubic interpolation might yield superior results compared to other interpolation methods, our experiments revealed that linear and cubic interpolation performed similarly. This observation may be attributed to the susceptibility of cubic interpolation to overfitting when data contains noise. Linear interpolation, being simpler, is less prone to overfitting. Regardless of the interpolation method used, random forest

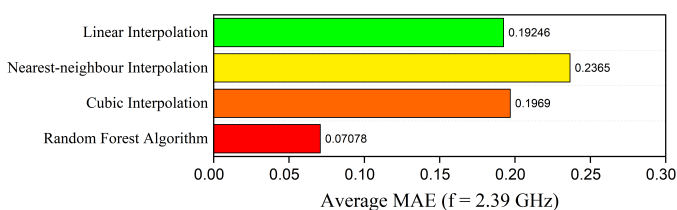


Fig. 9: Comparing random forest accuracy with different interpolation methods.

consistently outperforms them.

The performance of interpolation methods depends on dataset characteristics and underlying function behavior, making it crucial to conduct a thorough data assessment and experiment with various interpolation methods. Machine learning, like random forests, avoids the need for predefined interpolation choices and adapts to data. In contrast, traditional interpolation methods require selecting the right method, making machine learning advantageous in capturing complex signal fading and propagation behavior. Therefore, we focus on analyzing machine learning techniques for estimating ambient RF signal intensity in the following sections.

Fig. 10 and Fig. 11 provide a visual comparison of the actual measured data and the estimated values using different ML and interpolation techniques.

2) Multiple frequencies: In another study, we consider three variables (X, Y, f) , where the first two correspond to the (i, j) positions of the antenna, and f corresponds to the considered frequency. The results for two different approaches are presented in Fig. 12. The illustration provided in Fig. 12(a) compares the average MAE of a hundred random seeds for different ML algorithms when combining the measurement results of all frequencies into a single input variable. On the other hand, Fig. 12(b) illustrates the comparison of the MAE using different algorithms for each frequency individually, followed by the computation of the average MAE across all frequencies. Notably, in both approaches, the pattern of the MAE results across all frequencies closely mirrors the pattern of ML estimation observed at a single frequency of 2.39 GHz. The random forest algorithm appears to have the lowest MAE across all frequencies, and the LR algorithm shows higher MAE values across all frequencies compared to the other algorithms. It is worth noting that the number and distribution of data points collected at each frequency are precisely the same for all algorithms and all frequencies. Our results indicate that the MAE values for each frequency, using each algorithm, differ from one another. This difference can be attributed to the inherent differences in the signal characteristics at each frequency. Signal intensity, noise level, and other signal quality parameters may change at different frequencies, resulting in significant variations in the reliability of ML algorithms across different frequencies.

The MAE values obtained for each ML algorithm vary when computing the MAE using different approaches. Because each approach takes various aspects of the data into account, the outcomes from the two approaches differ. The first approach offers a comprehensive perspective of the algorithms' performance by combining results across all frequencies. In contrast, the second approach evaluates the performance of algorithms at each frequency individually before computing an average based on the processing of all individual frequencies. These differences can arise due to variations in the characteristics and patterns present in the data at different frequencies. Some algorithms may perform better or worse at certain frequencies, and assessing frequencies individually allows us to capture such variations. The average MAE across all frequencies provides a balanced evaluation, considering both the individual performance at each frequency and the overall performance

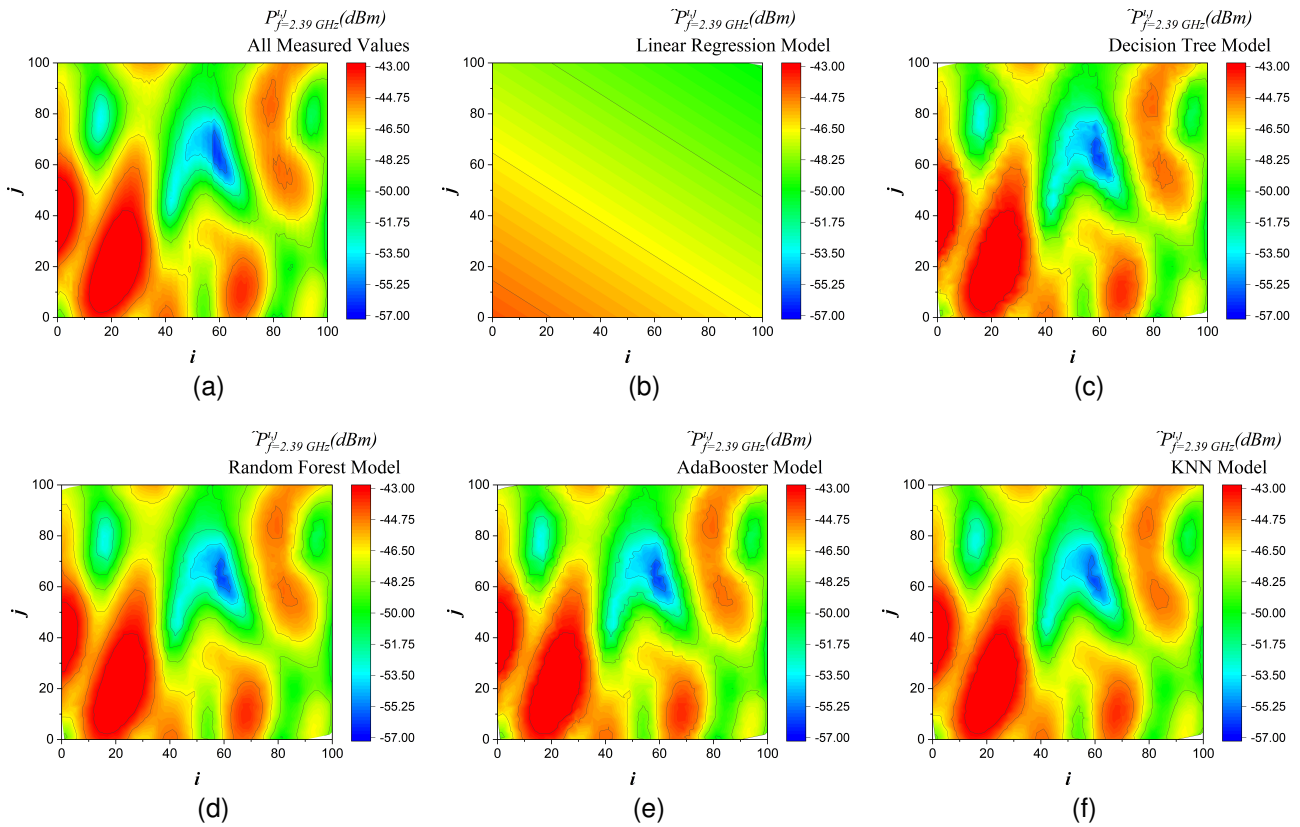


Fig. 10: Graphical comparison of the measured and estimated values when dataset is split into randomized 70 : 15 : 15 training, test, and validation sets, (a) all measured values, estimated values using ML (b) LR algorithm, (c) DT algorithm, (d) random forest algorithm, (e) AdaBooster algorithm, (f) KNN algorithm.

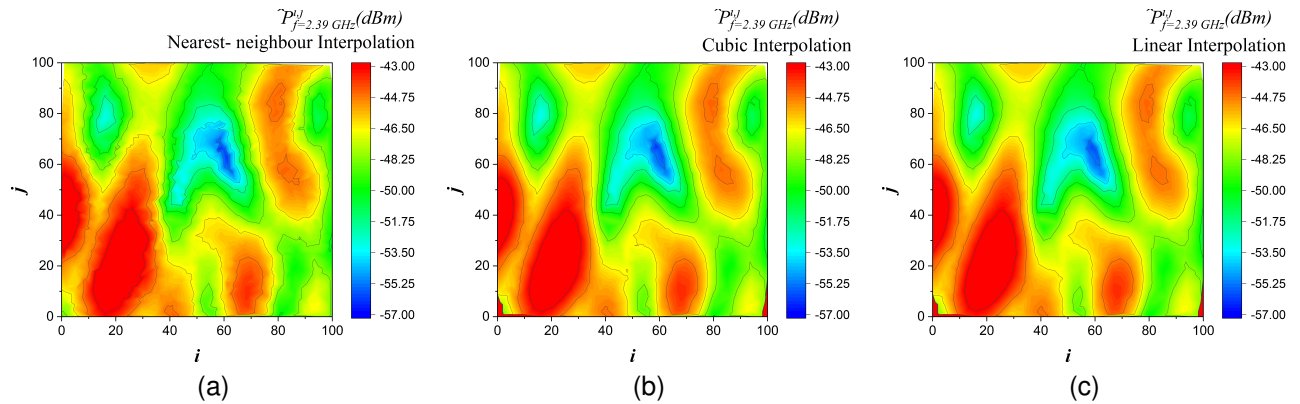


Fig. 11: Graphical comparison of the estimated values using interpolation models when the dataset is split into randomized 70 : 15 : 15 training, test, and validation sets, (a) nearest-neighbour interpolation, (b) cubic interpolation, (c) linear interpolation.

across all frequencies. Another potential factor contributing to the differences is that the first approach employs an input dataset eight times larger than the one used for training and evaluation at a single frequency, which increases the risk of overfitting. Given that the second approach yields more favorable outcomes, specifically lower MAE values, we will focus on utilizing the second approach for further analysis in this section.

3) *Reduction of sampling points* : Scanning an area with a high granularity between measurement points to find the

best spot for an energy harvester may not be feasible for many applications. As a result, this study considers the effects of reducing measurement points. The performance of ML algorithms for *estimating* the received power level is evaluated while the number of measurement points is reduced to 5.0% of the original data. When referring to 100% of the data, it means that the total number of measurement points used for training and testing in a single-frequency evaluation is 10,201. However, in the case of using 50% of the data, the number of measurement points considered for training and

TABLE I: MAE RESULTS FOR DATA REDUCTION ON SCANNING ROWS, CORRESPONDING TO THE SECOND APPROACH

	Data Reduction								
	100%	50.0%	33.6%	25.7%	20.6%	17.0%	12.0%	9.0%	5.0%
AdaBoost	0.15356	0.24873	0.26856	0.29196	0.31495	0.33281	0.33976	0.34158	0.37622
Decision Tree	0.17462	0.28037	0.30379	0.32979	0.35627	0.37717	0.38458	0.389	0.43277
Random Forest	0.10511	0.17154	0.18881	0.21103	0.23435	0.25451	0.26805	0.30284	0.37612
K-Nearest Neighbors	0.17377	0.2573	0.28763	0.31944	0.34421	0.36238	0.41872	0.45309	0.49347
Linear Regression	2.91863	2.94604	2.93666	2.92876	2.92952	2.92818	2.9923	3.00944	3.01065

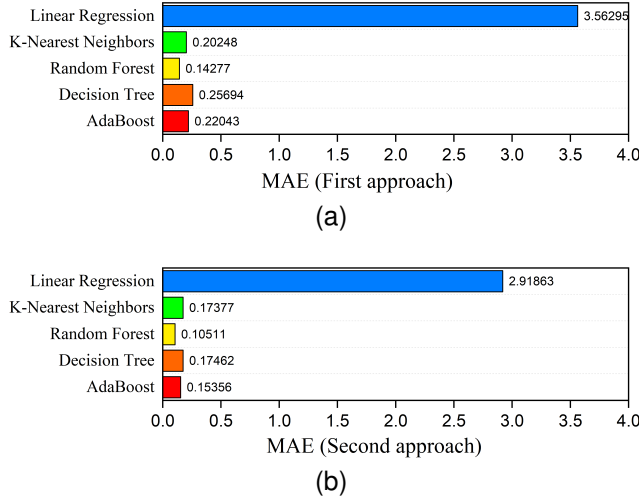


Fig. 12: MAE using different algorithms considering all frequencies as the third variable. (a) all frequencies are combined together in a single file, (b) each frequency is evaluated individually and then an average of MAE of all frequencies is computed.

testing is 5,151 and they are stored in a new file. Out of these 5,151 points, 70% (3,606) are used for training, and the remaining 30% (1,545) are used for testing. In order to address the challenge of scanning an area with high granularity to identify the optimal location for energy harvester placement, this strategy is employed to reduce the number of required measurement points. This is accomplished by selectively reducing the number of scanning lines. Initially, the number of scanning lines is halved by eliminating all the odd rows from the collected data. In this context, the rows represent different values of j . For example, in the case of reducing the data to 50%, all data values for (i,j) : $i=0,1,\dots,100, j=0,2,\dots,100$ are preserved, while other data are removed. Subsequently, a reduction process is implemented by selecting only the first line from a series of consecutive lines while discarding the rest. This reduction continues to effectively decrease the number of measured data points to only 5.0% of the original dataset. Despite the significant reduction, the retained data still contains the necessary information for analysis using ML algorithms. Table I and Fig. 13 provide an overview of the performance of different ML algorithms in *estimating* the received power level at a specific area in the presence of data reduction.

Further analysis of the results in Table I reveals that the LR algorithm performs the worst in all scenarios, with the

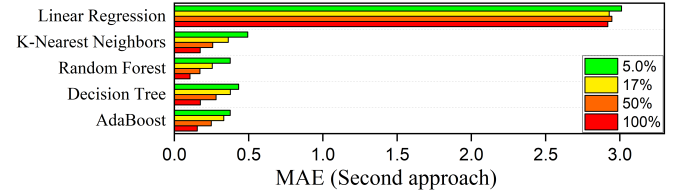


Fig. 13: MAE using different algorithms considering all frequencies while reducing the amount of data.

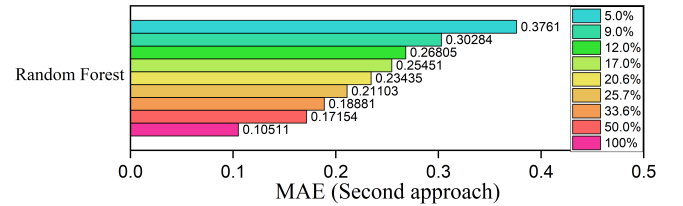


Fig. 14: MAE using random forest algorithm considering all frequencies while reducing the amount of data.

highest MAE observed for all levels of data reduction. The results indicate that even with a reduction in the amount of data, the random forest algorithm consistently outperforms the other algorithms across all data reduction scenarios. Fig. 14 presents an enhanced graphical representation of how the random forest algorithm performs. This suggests that the random forest algorithm is better able to generalize and make accurate *estimations* with limited data compared to the other algorithms. On the other hand, reducing the amount of data by reducing the number of scanning lines results in an increase in MAE for all algorithms.

In addition, in Fig. 15 it is evident that the random forest model can *estimate* the area with the highest ambient RF signal reception even when the data is reduced to only 5.0% of the initially collected data. These findings highlight the potential of ML techniques, especially the random forest algorithm, for energy harvesting *estimation* in scenarios where the amount of data available for training is limited. Moreover, by decreasing the number of measurement points to only 5.0% of the original quantity, the time required for running the measurement at each frequency will be significantly reduced from several hours to a matter of minutes. However, it should be noted that reducing the number of measurement points leads to a substantial decrease in accuracy. Thus, it is essential to find a balance between reducing measurement points and maintaining an acceptable level of accuracy.

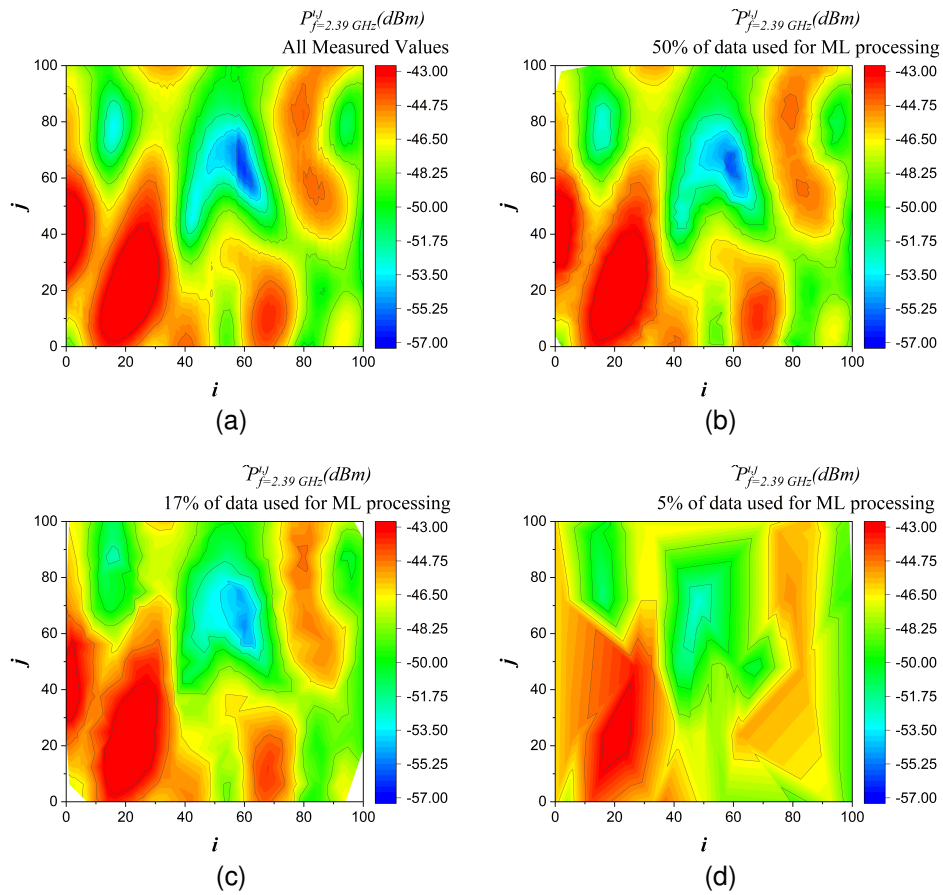


Fig. 15: Graphic comparison of the measured and **estimated** values using random forest algorithm at frequency 2.39 GHz. (a) All measured values, and **estimated** values while (b) reducing the amount of data to 50.0%, (c) reducing the amount of data to 17.0%, (d) reducing the amount of data to 5.0%, of the collected data.

VI. APPLICATION

There are potential applications in the fields of smart agriculture, smart home, and smart industrial monitoring and maintenance that can benefit from the investigations conducted in this work. Developments in ambient RF energy harvesting technology provide an opportunity to sustainably power ultra-low-power IoT devices. A moving or mobile device in an IoT network deployed in a smart home or smart farm structure could minimize its reliance on batteries by exploiting ambient RF sources, such as Wi-Fi signals or cellular signals, as a power source.

One specific potential application could be a robotic device used for environmental monitoring in smart agriculture fields. These robotic devices play a vital role in collecting and transferring environmental conditions to a control center for monitoring purposes. However, a significant challenge they face is the frequent need for recharging to activate the sensors required for wireless communication. To minimize power supply requirements, the power source of the robot is separate from the monitoring sensors. By integrating ambient RF energy harvesting technology into the robotic device, it could become more self-sustaining and reduce its dependency on battery recharging. **The sensing circuit or condition monitoring device, which can be mounted on the**

robot, typically consumes minimal energy and can efficiently rely on harvested RF energy. It can have the capability to store and manage this harvested energy, utilizing techniques to reduce power consumption for energy-efficient consumption [28], [29]. The device could harvest RF energy from ambient sources within the agricultural environment. Implementing multiband or wideband receiving antennas, along with optimal positioning based on the research findings, would ensure maximum energy reception and improve the device's energy harvesting efficiency. Furthermore, by utilizing ML algorithms, the device can **estimate** the optimal locations for energy harvesting based on real-time data, considering factors such as signal strength, obstacles in the field, and movement patterns of the robot. **The robot can periodically update its training data, allowing it to efficiently balance its primary task with ongoing optimization, minimizing the need for continuous data collection.** Implementing these advancements in the robotic device would enhance its performance, reduce its reliance on frequent recharging, and contribute to a more sustainable and energy-efficient smart agriculture environment. It would enable the robotic device to operate for longer durations without interruptions, ensuring continuous and accurate environmental monitoring. Additionally, by reducing the need for frequent battery recharging, maintenance efforts would be minimized,

allowing the device to focus more on its core monitoring tasks in the agriculture field. However, this is merely a suggested potential application that requires further exploration in future research endeavors.

VII. CONCLUSION

With the growing IoT adoption, sustainable and efficient power supply solutions are crucial. Harvesting ambient radio frequency (RF) energy offers promise for powering low-power IoT devices. Yet, capturing and utilizing RF energy is challenging due to signal propagation, interference, and environmental factors. Accurately estimating received signal intensity is vital for optimizing the placement of RF energy harvesters in IoT applications. In this paper, we investigated the collection and estimation of received signal data for energy harvesting applications. We conducted measurements at a single frequency under controlled conditions and observed no significant variations in signal intensity. We also conducted measurements at different frequencies and found that the received signal intensity was frequency-dependent due to the interaction of electromagnetic waves with objects in the environment.

Additionally, this research investigated the application of interpolation techniques and ML methods to estimate the optimal location for placing an RF energy harvester in a given area, considering real-world scenarios. By utilizing extensive data collected from controlled indoor measurements, we trained and evaluated three interpolation techniques and five different supervised ML algorithms. The performance of the models was assessed based on the MAE metric. Among the ML algorithms tested, random forest demonstrated the best performance at a single frequency, while LR performed poorly. When considering all frequencies together, the random forest still outperformed other algorithms, indicating its effectiveness in handling non-linear relationships. We also examined the impact of reducing the number of measurement points and found that even with a substantial reduction to only 5.0% of the original data, the ML algorithms were still capable of making estimations, albeit with a trade-off in accuracy. These findings contribute to the understanding of signal behavior in energy harvesting systems and emphasize the effectiveness of ML algorithms in estimating received signal power for optimal energy harvesting.

We consider exploring intricate real-world scenarios to be a significant path for future research, as well as delving into practical aspects as another potential direction for extending this work.

REFERENCES

- [1] D. Wilson, "IoT is creating massive growth opportunities," Oct. 2020, [online]. Available: <https://blogs.cisco.com/internet-of-things/iot-is-creating-massive-growth-opportunities>.
- [2] B. Posey, "Ultimate IoT implementation guide for businesses," Mar. 2022, [online]. Available: <https://www.techtarget.com/iotagenda/definition/IoT-device>.
- [3] I. Zhou *et al.*, "Internet of things 2.0: Concepts, applications, and future directions," *IEEE Access*, vol. 9, pp. 70961–71012, 2021.
- [4] A. Cammarano, C. Petrioli, and D. Spensa, "Pro-energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks," in *9th Int. Conf. Mobile Ad-Hoc Sensor Syst.* IEEE, Oct. 2012, pp. 75–83.
- [5] F. Yildiz, "Potential ambient energy-harvesting sources and techniques," *J. Technol. Stud.*, vol. 35, no. 1, pp. 40–48, Fall 2009.
- [6] M. Cansiz, D. Altinel, and G. K. Kurt, "Efficiency in RF energy harvesting systems: A comprehensive review," *Energy*, vol. 174, pp. 292–309, May 2019.
- [7] L.-G. Tran, H.-K. Cha, and W.-T. Park, "RF power harvesting: A review on designing methodologies and applications," *Micro Nano Syst. Lett.*, vol. 5, pp. 1–16, Feb. 2017.
- [8] H. H. R. Sherazi, D. Zorbas, and B. O'Flynn, "A comprehensive survey on RF energy harvesting: Applications and performance determinants," *Sensors*, vol. 22, no. 8, p. 2990, Apr. 2022.
- [9] V. N. Vo, H. Tran, E. Uhlemann, Q. X. Truong, C. So-In, and A. Balador, "Reliable communication performance for energy harvesting wireless sensor networks," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–6.
- [10] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, March, 2015.
- [11] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 6, no. 4, pp. 32–es, Sep. 2007.
- [12] J. R. Piorno, C. Bergonzini, D. Atienza, and T. S. Rosing, "Prediction and management in energy harvested wireless sensor nodes," in *1st Int. Conf. Wireless Commun., Veh. Technol., Inf. Theory Aerosp. & Electron. Syst. Technol.* IEEE, May 2009, pp. 6–10.
- [13] Y. Chen, *Energy Harvesting Communications: Principles and Theories*. Hoboken, NJ, USA: Wiley, 2019.
- [14] H. J. Visser and R. J. Vullers, "RF energy harvesting and transport for wireless sensor network applications: Principles and requirements," *Proc. IEEE*, vol. 101, no. 6, pp. 1410–1423, Jun. 2013.
- [15] K. S. Raghav and D. Bansal, "Power controlled system for self-sustained RF energy harvesting sensors," *Analog Integrated Circuits and Signal Processing*, vol. 113, no. 1, pp. 73–79, 2022.
- [16] D. Arnitz and M. S. Reynolds, "MIMO wireless power transfer for mobile devices," *IEEE Pervasive Computing*, vol. 15, no. 4, pp. 36–44, 2016.
- [17] S. Kim *et al.*, "Ambient RF energy-harvesting technologies for self-sustainable standalone wireless sensor platforms," *Proc. IEEE*, vol. 102, no. 11, pp. 1649–1666, Nov. 2014.
- [18] M. Piñuela, P. D. Mitcheson, and S. Lucyszyn, "Ambient RF energy harvesting in urban and semi-urban environments," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 7, pp. 2715–2726, Jul. 2013.
- [19] F. Azmat, Y. Chen, and N. Stocks, "Predictive modelling of RF energy for wireless powered communications," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 173–176, Jan. 2016.
- [20] Y. Ye, F. Azmat, I. Adenopo, Y. Chen, and R. Shi, "RF energy modelling using machine learning for energy harvesting communications systems," *Int. J. Commun. Syst.*, vol. 34, no. 3, p. e4688, Feb. 2021.
- [21] Z. Liang and J. Yuan, "Modelling and prediction of mobile service channel power density for RF energy harvesting," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 741–744, May 2020.
- [22] A. R. Mishra, *Advanced Cellular Network Planning and Optimisation: 2G/2.5 G/3G... Evolution to 4G*. Hoboken, NJ, USA: Wiley, 2007.
- [23] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.
- [24] D. Molkdar, "Review on radio propagation into and within buildings," in *IEE Proc. H-Microw., Antennas Propag.*, vol. 138, no. 1. IET, Feb. 1991, pp. 61–73.
- [25] Ghent University, "Electromagnetism," [online]. Available: <https://www.ugent.be/ea/idlab/en/research/research-infrastructure/electromagnetism.htm>.
- [26] K. Nadali, P. McEvoy, and M. J. Ammann, "A broadband circularly polarised slot antenna for ambient RF energy harvesting applications," in *Int. Workshop Antenna Technol.* IEEE, May 2022, pp. 153–156.
- [27] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, Dec. 2005.
- [28] H. Bello, Z. Xiaoping, R. Nordin, and J. Xin, "Advances and opportunities in passive wake-up radios with wireless energy harvesting for the internet of things applications," *Sensors*, vol. 19, no. 14, p. 3078, 2019.
- [29] G. Verma and V. Sharma, "A novel RF energy harvester for event-based environmental monitoring in wireless sensor networks," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3189–3203, 2021.