Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

A Clustering-Based Artificial Intelligence Approach for Minimizing of Ionizing Radiation Exposure in Uyo Metropolis, Nigeria

Imeh Umoren[1], Saviour Inyang[2], Ubong E. Etuk[3], Daniel Essien[4]

[1] *School of Computing and Information Technology, Ikot Abasi, Nigeria*
[2] *Department of Computing, Topfaith University, Mkpatak, Nigeria*
[3] *School of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK*
[4] *Independent Researcher, Halifax, Nova Scotia, Canada*

[1]*imehumoren@aksu.edu.ng*, [2]*saviour.inyang@topfaith.edu.ng*, [3]*u.etuk.1@research.gla.ac.uk*, [4]*kanalt2700@gmail.com*

ARTICLE INFORMATION

Article History:
Received: October 23, 2024
Last Revision: March 14, 2025
Published *Online*: March 30, 2025

KEYWORDS

Electromagnetic Radio (EMR),
Clustering Algorithms,
Radiation,
Ionizing,
Machine Learning

CORRESPONDENCE

Phone: +2348160521616
E-mail: saviour.inyang@topfaith.edu.ng

ABSTRACT

Electromagnetic Field (EMF) radio frequency exposure is a growing concern due to its impacts on public health and the environment. This study aims to develop a data-driven framework for clustering and analyzing long-term far-field EMF exposure in Uyo Metropolis, Nigeria, with a focus on identifying exposure patterns and assessing their implications. Data were measured at multiple locations using smart meter strategically deployed across three major roads in uyo metropolis to capture variations in exposure levels. The preprocessing steps involved data cleaning and normalization to enhance data quality and reliability for meaningful analysis. Four clustering algorithms, namely, K-Means, Hierarchical Clustering, DBSCAN, and Gaussian Mixture Model (GMM), were employed to analyze the distribution of radiation levels. The Silhouette score was used to evaluate the different clustering methods with respect to cohesion within clusters and separation from other clusters. The best results were obtained by Hierarchical Clustering and GMM, each achieving a mean Silhouette score of 0.81, indicating well-defined and highly contrasting clusters. K-Means performed moderately well, with an average Silhouette score of 0.73, while DBSCAN, due to its sensitivity to noise and parameter settings, achieved a lower score of 0.62. These findings highlight significant spatial variability in EMF exposure across different urban zones, emphasizing the need for targeted regulatory measures. The study underscores effectiveness of machine learning and offers a scalable approach for characterizing EMF exposure. Results reported offer scalable and data-driven framework for characterizing exposure patterns, with important implications for public health policies, urban planning strategies, and regulatory interventions.

1. INTRODUCTION

Wireless communication systems are becoming increasingly integral to our daily lives due to advancements in cellular technology. Central to these systems are base radio stations [1], through which all communication must pass. User demand for multimedia services, including internet access in addition to voice and text, grows throughout urban areas, the need for more base stations escalates, influenced by factors such as geographical terrain, user density, and bandwidth consumption. Each base station functions as an electromagnetic source, resulting in an increase in the electromagnetic field (EMF) exposure correlated with the number of base stations deployed. Additional sources of EMF include radio and television transmitters, high-voltage transmission lines, power transformers, and various wireless transmissions. The rising concentration of telecommunication equipment, including base stations and radio transmitters, contributes to a more intricate electromagnetic environment [2]. Electromagnetic radiation has emerged as the fourth most significant form of pollution affecting human life, following water, air, and noise pollution. This phenomenon, termed electromagnetic pollution, represents a novel type of environmental degradation and is considered one of the most serious risks to public health.

According to [3], the natural levels of electric and magnetic fields are set. In contrast, artificial electric and magnetic fields produced by electrical installations, communication systems, and electrical energy transport networks are characterized as electromagnetic pollution, defined as the presence of excessive amounts of electromagnetic spectrum radiation. Electromagnetic emissions have long been a concern, particularly since the rise of interconnected systems and technological applications reliant on electromagnetic radiation. Examples include mobile phones, television, media, and power transmission lines [3]. Wireless sensor networks (WSNs) are used in applications like military, health monitoring, and surveillance, with sensor nodes collecting data in challenging environments, often constrained by limited power [4] . In these scenarios, coordination and communication among specific nodes are essential.

A comparative analysis of clustering techniques applied to WSNs revealed that hybrid clustering approaches, integrating k-means and hierarchical clustering, achieved an accuracy improvement over traditional single-method clustering [5]. Furthermore, optimization techniques led to a 22% reduction in electromagnetic exposure levels within urban areas [6]. These findings highlight the practical advantages of combining clustering methods for EMF mitigation, particularly in local environments where network density and urban infrastructure impact exposure levels. However, because electric and magnetic fields are smaller in size and frequency compared to ionizing radiation, all these applications are classified as non-ionizing radiation. Ionizing radiation occurs when electric and magnetic fields can dislodge electrons from their atomic orbits, with frequencies oscillating. Therefore, electromagnetic fields below are categorized as non-ionizing radiation. Nonetheless, they can generate energy through heat. Concerns about exposure to electromagnetic radio waves are particularly pronounced potential health risks.

In May 2011, the International Agency for Research on Cancer (IARC) classified radio frequency EMF as probably carcinogenic to humans (Group 2B) [7]. This classification was based on evidence linking wireless phone use to a higher incidence of glioma, a malignant brain cancer. Over the past decade, numerous scientific advancements have led to variations in the levels of electromagnetic fields that the public encounters have hazards which can have negative impact on their wellbeing [8]. Electromagnetic fields consist of a series of waves that oscillate at specific frequencies, with a distinct distance between them known as wavelength. EMFs cover a wide bandwidth, ranging from low-frequency electrical supply lines with wavelengths of hundreds of meters to high-frequency medical X-rays.

2. RELATED WORK

To enhance understanding of urban electromagnetic exposure, various researchers have conducted studies on electromagnetic radiation for analysis and monitoring [9]delves into the measurement of radiation, its impact on biological systems, the potential hazards it poses to healthcare professionals, and practical guidelines tailored for different medical roles. Radiation is characterized as energy in motion. Also, studies have shown that radiation from mobile phone base stations poses risks to both humans and wildlife. Exposure to non-ionizing electromagnetic radiation has been linked to health issues such as fatigue, sleep disturbances, headaches, skin irritation, memory loss, and even infertility [10], [11] undertook a study to measure radiofrequency electromagnetic field (RF-EMF) levels, collecting data by walking through 51 distinct outdoor microenvironments in 20 cities in Switzerland.

Additionally, [12] assessed electromagnetic radiation levels at specific locations within Amman City and evaluated the findings against the global safety benchmarks set by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) [13] employed an unsupervised machine learning approach to assess the impact of EMF exposure in educational settings. Their methodology included the use of geographic information systems to collect, store, analyze, and present data on electromagnetic radiation. This approach allowed for geographical mapping and graphical representations of spatial relationships. However, they encountered challenges in identifying areas suspected of having high electromagnetic radiation levels, particularly distinguishing between urban and semi-urban areas due to differing population densities. Hierarchical clustering methods were utilized to compute the dissimilarity matrix, revealing that the mean exposure across the 205 schools studied remained below the recommended exposure values [11] conducted RF-EMF measurements from 2016 to 2018 to compare their results against the acceptable limits established by the International Commission on Non-Ionizing Radiation Protection (ICNIRP). They performed both short- and long-term measurements at 500 locations over a two-year span, concluding that RF-EMF levels did not exceed ICNIRP limits.

In dense urban areas of Beijing [14] assessed electromagnetic radiation exposure levels along major streets using ordinary kriging as an interpolation technique to analyze radiation exposure across large outdoor spaces via a car-mounted measuring device. The study allowed for the mapping of exposure levels to highlight hotspot regions. Furthermore, [15] evaluates the variation in Entrance Skin Dose (ESD) across different age groups in diagnostic areas like RGU, MCU, Fistulogram, and pelvic X-rays, focusing on identifying factors influencing ESD fluctuations for better radiation management [16] noted that wireless communication often exhibits unpredictable behavior, with quality influenced by environmental factors, the specific frequency spectrum in use, modulation schemes, and the devices involved [17]. An innovative approach for creating electromagnetic field (EMF) exposure maps in a designated urban region in France was conducted by [18].

The primary goal is to generate these maps using data gathered from spatially distributed sensors [18]. Recent advancements have leveraged AI and machine learning models to predict and classify EMF exposure levels in various settings, improving accuracy and enabling proactive management of exposure risks [19]. However, hierarchical clustering methods have since been refined using AI-based models that dynamically adjust cluster parameters based on exposure patterns, as seen in [20]. AI-powered clustering and predictive models have been recently proposed to optimize network planning while minimizing EMF exposure in urban environments. This AI-based clustering technique can inform network planners about optimal locations for infrastructure, thereby aiding in the minimization of EMF exposure in densely built environments [21]. Comparing these results with more recent deep-learning-based clustering techniques shows that AI-driven approaches provide better granularity in detecting high-exposure regions [22].

To the best of our knowledge, there is no study that adopts Clustering-Based AI Approach for Minimizing of Ionizing Radiation Exposure in Uyo Metropolis. This aims to fill this gap by implementing a clustered approach to minimize radiation exposure in urban environments, ultimately enhancing public health.

3. METHODOLOGY

To tackle the challenges identified in existing systems, this study employs the framework illustrated in Figure 1. The proposed framework is designed to assess and analyze long-term exposure to electromagnetic field radiation, using Uyo Metropolitan City, Nigeria, as a case study.

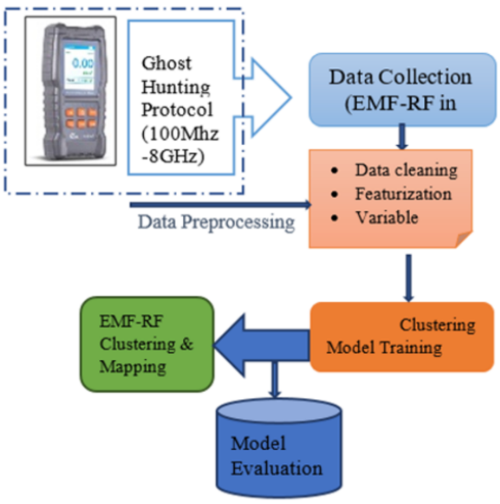


FIGURE 1. PROPOSED FRAMEWORK

The methods utilized in this research are outlined below:

3.1 Collection

In this research, data collection involves gathering both quantitative and qualitative information on specific variables to evaluate outcomes or gain actionable insights. It is a systematic process of collecting observations or measurements that can support the evaluation of hypotheses. Regardless of the research topic, data collection is typically the initial and most crucial step in the research process. Depending on the required information, various data collection methods are employed across different fields of study. The primary objective is to ensure that the data is rich and reliable, enabling statistical analysis that informs data-driven research decisions. In this study, the focus is on assessing and classifying far-field radio frequency electromagnetic field spatial exposure in the urban areas of Uyo Metropolis. Due to the need for accurate and continuous measurement of electromagnetic field (EMF) radiation levels, a smart meter was employed. Smart meters provide real-time readings, which are essential for precise environmental monitoring.

Data for this study was collected through observation. This method involves watching and documenting a subject's behavior, serving to gather valuable information through direct observation. It is often referred to as a participatory approach, as the researcher needs to establish a rapport with participants by immersing themselves in their environment. This allows for accurate recording and note-taking based on what is observed. In this research, data collection took place over a month at the University of Uyo metropolis, specifically along Oron Road, Nwaniba, and Ikot Ekpene Road. A smart meter was employed to detect and measure electromagnetic field (EMF) radiation emissions on these three streets within Uyo metropolis, facilitating effective clustering.

Data points were collected at consistent time intervals throughout the day to account for variations in EMF exposure due to environmental and operational factors, ensuring a comprehensive dataset.

3.2 Data Pre-Processing

Data preprocessing is a critical mining technique that transforms raw data into a usable and efficient format. This step is essential in preparing the data for model building. Raw EMF readings are often affected by noise and inconsistencies, which can impact the accuracy of analysis. Therefore, preprocessing techniques such as outlier detection, data normalization, and feature engineering were applied. In this study, the data was cleaned, and relevant variables were identified. Outlier detection was conducted using the Interquartile Range (IQR) method to filter extreme values caused by environmental interference. Missing values were addressed using interpolation techniques.

Featurization and normalization were performed to scale the data appropriately, aligning it with the continuous measurements obtained from the EMF smart meter. Various terms such as variables, features, fields, attributes, and dimensions are used to describe these elements. Table 1 illustrates how we categorize the statistical inputs from the data based on measurement units. To ensure uniform scaling, Min-Max normalization was applied to the EMF values. This technique preserves the relationships within the dataset while bringing all values into a standard range between 0 and 1.

TABLE 1. FIELD DATA ATTRIBUTES CATEGORIZATION

Attributes	Category / Units
Location	Text
Mode of EMF	(Magnetic, Radio Frequency, Electric)
EMF Value	Numeric
Distance in meters	Meters

3.3 Clustering

Clustering is fundamentally an unsupervised learning technique, which means it focuses on extracting patterns from datasets that contain input data without any labeled responses. This approach aims to identify significant structures, elucidate underlying processes, and create features and groups within a sample set. Essentially, clustering involves dividing a population or dataset into distinct groups, ensuring that data points within the same group are more like each other than to those in other groups. It revolves around categorizing objects based on their similarities and differences. When selecting a clustering algorithm, it is crucial to consider its scalability, especially since machine learning datasets can comprise millions of samples, and not all algorithms can handle large datasets effectively. Many clustering algorithms calculate the similarity between every pair of instances. Clustering can be approached in various ways, as detailed in a Comprehensive Survey of Clustering Algorithms [23]. Each method is tailored to specific data distributions. Nevertheless, this study adopts four types of clustering algorithms which are used for the clustering of radiation data to minimize the exposure rate within the urban location in uyo.

This study employs four clustering algorithms i.e K-Means, Hierarchical Clustering, DBSCAN, and Gaussian Mixture Models (GMM) based on their suitability for the dataset:

- K-Means is selected for its efficiency and ability to handle large datasets while optimizing intra-cluster variance [24].
- Hierarchical Clustering is used because of its interpretability and ability to create a dendrogram representation without pre-defining the number of clusters [25].
- DBSCAN is chosen due to its capability to detect noise and non-spherical clusters, making it useful for detecting areas of high EMF exposure with irregular distribution [26].
- GMM is included as a probabilistic model that assigns soft cluster memberships, which is particularly effective in scenarios where the dataset follows multiple overlapping Gaussian distributions [27].

3.3.1 K-Means Clustering

K-means clustering is a widely used machine learning algorithm used to separate n data into k clusters, where individual clusters are represented by their centroid [27]. It is widely used for segmentation and anomaly detection tasks. It minimizes the variance within individual clusters and efficiently groups data points that are similar [27], [28] which is shown in Eq. 1.

$$j = \sum_{i=1}^{k} \sum_{x \in c_i} \|x - \mu_i\|^2 \qquad (1)$$

Where:
K: number of clusters
c_i : set of points in cluster i
μ_i : Centroid of cluster i
$\|x - \mu_i\|^2$: Squared Euclidean distance between a data point x and μ_i

The K-means algorithm takes a series of iterative steps to separate data into k clusters. First, the k centroids are randomly selected methods. Once the centroid is selected, the algorithm assigns individual data points to the nearest centroid according to the Euclidean distance, forming clusters or groups. After all the points have been assigned, the centroids are monitored and updated by calculating the mean position of all the data points in the cluster which is presented in Eq. 2.

$$\frac{1}{C_i} \sum_{x \in c_i}^{x} \qquad (2)$$

Where:
μ_i : Centroid of cluster C_i
x : Cluster data point

3.3.2 Hierarchical Clustering

Hierarchical Clustering is an unsupervised machine learning algorithm that builds structures called dendrogram. The dendrogram was performed using two methods such as agglomeration (bottom-up) and divisive (top-down) methods [29] The main advantage of the clustering algorithm is that it does not require a specific number of clusters beforehand. However, the key limitation is that it requires high computational complexities, especially with large datasets [30]. This method is used in various domains such as gene expression analysis and social network analysis.

3.3.3 DBScan Clustering

Density Based Spatial Clustering of Application with Noise (DBSCAN) is a density-based clustering algorithm that is used in identifying clusters with unpredictable shapes or data points and handling noise in datasets [31]. The inner workings of the algorithm occur by marking as core points those that have minimum number of neighboring points within a specific distance, ε (epsilon) and separating the dense regions from noise. It relies on the radius for neighbors and minimum number of points represented by ε and MinPts respectively. The core point is defined by: Core Point: $|N(\varepsilon,\ p)| \geq MinPts$ DBSCAN has been generally used for clustering spatial data and big data with complex structures. Recent advancements in DBSCAN include parallelization techniques to improve computational efficiency [32], enhancements in DBSCAN algorithms to make them for efficient [33], [34]. These studies show the capability and wide range of use cases for the DBSCAN algorithm.

3.3.4 Gaussian Mixture Model

The Gaussian Mixture Model is a probabilistic model used for clustering; it assumes the data is generated by multiple gaussian distributions [35]. It is mostly used for clustering and data estimation in cases where the data displays complex patterns. The GMM model provides a more flexible approach by assigning each data point to a cluster based on probability rather than hard assignment like the k means clustering method. This approach helps it outperform methods like k-means clustering when handling more complex patterns. The clustering process occurs by maximizing the likelihood of the data using the assumption of gaussian components. It achieves this using the Expectation-Maximization (EM) Algorithm which consists of the E-step and M-step. The E-Step calculates the posterior probability that an individual data point belongs to each Gaussian component using the Bayes Theorem and the M-Step updates the parameters of the Gaussian Components to increase the likelihood. This can be represented in Eq.3.

$$\log L(\theta|X) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{k} \pi_k N(x_i \,|\mu_k, \Sigma_k\right) \qquad (3)$$

3.4 Model Evaluation

Traditional metrics for assessing the accuracy and effectiveness of supervised learning models in classification and regression are not applicable in this study. Clustering is primarily associated with Unsupervised Learning, while Classification and Regression pertain to Supervised Learning. The distinction between these two areas is based on the nature of the data used. In Supervised Learning, data are labeled with a category (Classification) or a numerical value (Regression), whereas Unsupervised Learning deals with unlabeled data, making analysis more challenging. To evaluate the performance of clustering algorithms, one must focus on model effectiveness and accuracy. This is crucial since clusters are often evaluated manually and subjectively to assess their relevance.

In this project, the Silhouette score and plot are utilized to measure model accuracy. The Silhouette Score quantifies the distance between clusters, providing insights into the proximity of data points within a cluster compared to those in neighboring clusters. This metric ranges from [-1, 1] and serves as a valuable tool for visually assessing the similarities and differences among clusters. The score is derived from the average intra-cluster distance and the average nearest-cluster distance for each data sample. The Silhouette coefficient for an individual sample is defined as:

$$\frac{n-i}{\max(i,n)} \qquad (4)$$

An object is considered to fit well within its assigned cluster and poorly with neighboring clusters if its score is close to 1. Conversely, a score near -1 indicates that the object may be incorrectly assigned to its cluster. The Silhouette Score helps to evaluate the effectiveness of clustering methods and determine the optimal number of clusters for a given dataset. Given a data sample, we can use Silhouette score to calculate the best cluster in a data set as in eqn5.

$$S(i) = \frac{b(i) - \mathrm{a}(i)}{\max(a(i), \mathrm{b}(i))} \qquad (5)$$

Where:

a. (i) is the average distance from i to other data points in the same cluster.
b. (i) is the smallest average distance from i to data points in a different cluster.

4. RESULT AND DISCUSSION

Result evaluation involves a thorough and unbiased assessment of an ongoing or completed project. The objective is to determine the extent to which project goals have been achieved, as well as to evaluate effectiveness, efficiency, impact, and sustainability. In this research, the K-means clustering technique, a form of unsupervised machine learning, is employed to categorize EMF radiation emitted by various devices in urban areas within the Uyo metropolis. To achieve this, R programming is utilized to preprocess our variables from the collected data, providing a summary of all parameters, including mean, median, maximum, and mode, as illustrated in Figure 2.

```
> summary(dat)
   Location          Mode_of_EMF      EMF_value        Distance_in_Meters
 Length:84          Min.   :1.000   Min.   :   0.00   Min.   : 0.020
 Class :character   1st Qu.:1.000   1st Qu.:   0.00   1st Qu.: 0.020
 Mode  :character   Median :2.000   Median :  10.00   Median : 1.000
                    Mean   :1.981   Mean   :  74.07   Mean   : 3.722
                    3rd Qu.:3.000   3rd Qu.:  34.90   3rd Qu.: 1.000
                    Max.   :3.000   Max.   :1021.00   Max.   :50.000
                    NA's   :31      NA's   :31
```

FIGURE 2. VARIABLES PROCESSING

4.1 Training and Testing

In machine learning, the training dataset consists of the actual data used to train the model for various tasks [36]. This real-world data is what the models learn from during their ongoing development, utilizing different APIs and algorithms to allow the machine to function independently. One of the core objectives in machine learning is to create algorithms that can learn from data and make predictions. These algorithms generate a mathematical model [37] based on input data, leading to data-driven predictions or conclusions. In this study, the training set was derived from remotely collected statistical field data in Uyo, Akwa Ibom State, Nigeria, and underwent preprocessing to eliminate noisy and redundant information. The dataset was split into a 70% training set and a 30% test set for the K-means clustering model. Figure 3 illustrates the division of the dataset into training and testing subsets for the K-means algorithm.

Location	Mode_of_EMF	EMF_value	Distance_in_Meters
Ikot ekpeneRoad	1	58.8	0.02
Ikot ekpeneRoad	1	0.1	1.00
OronRoad	2	17.0	0.02
OronRoad	2	14.0	1.00
Ikot ekpeneRoad	3	8.6	0.02
NwanibaRoad	3	4.6	1.00
NwanibaRoad	1	0.5	0.02
NwanibaRoad	1	0.0	1.00
NwanibaRoad	2	13.0	0.02
NwanibaRoad	2	10.0	1.00
NwanibaRoad	3	1.2	0.02
NwanibaRoad	3	0.9	1.00
NwanibaRoad	1	0.0	0.02
OronRoad	1	0.0	1.00
OronRoad	2	765.0	0.02
OronRoad	2	50.0	1.00
OronRoad	3	30.0	0.02
OronRoad	3	4.6	1.00
OronRoad	1	0.0	1.00
NwanibaRoad	1	0.0	20.00
NwanibaRoad	2	1021.0	1.00
NwanibaRoad	2	36.0	20.00
NwanibaRoad	3	54.7	1.00
NwanibaRoad	3	9.8	20.00
NwanibaRoad	1	0.0	0.02
NwanibaRoad	1	0.0	1.00
NwanibaRoad	2	406.0	0.02

1 to 27 of 84 entries, 4 total columns

FIGURE 3. DATASET

Furthermore, we carried out normalization which is essential and necessary for clustering. Figure 4 depicts the normalized data, or we can also say scaled data.

	Mode_of_EMF	EMF_value	Distance_in_Meters
[1,]	-1.19622319	-0.07661876	-0.4183322
[2,]	-1.19622319	-0.37119196	-0.3396924
[3,]	0.02300429	-0.28638298	-0.4183322
[4,]	0.02300429	-0.30143783	-0.3396924
[5,]	1.24223177	-0.32853656	-0.4183322
[6,]	1.24223177	-0.34860969	-0.3396924
[7,]	-1.19622319	-0.36918465	-0.4183322
[8,]	-1.19622319	-0.37169379	-0.3396924
[9,]	0.02300429	-0.30645611	-0.4183322
[10,]	0.02300429	-0.32151096	-0.3396924
[11,]	1.24223177	-0.36567185	-0.4183322
[12,]	1.24223177	-0.36717734	-0.3396924
[13,]	-1.19622319	-0.37169379	-0.4183322
[14,]	-1.19622319	-0.37169379	-0.3396924
[15,]	0.02300429	3.46729253	-0.4183322
[16,]	0.02300429	-0.12077965	-0.3396924
[17,]	1.24223177	-0.22114531	-0.4183322
[18,]	1.24223177	-0.34860969	-0.3396924
[19,]	-1.19622319	-0.37169379	-0.3396924
[20,]	-1.19622319	-0.37169379	1.1849567
[21,]	0.02300429	4.75197293	-0.3396924
[22,]	0.02300429	-0.19103561	1.1849567
[23,]	1.24223177	-0.09719372	-0.3396924
[24,]	1.24223177	-0.32251462	1.1849567
[25,]	-1.19622319	-0.37169379	-0.4183322
[26,]	-1.19622319	-0.37169379	-0.3396924
[27,]	0.02300429	1.66572902	-0.4183322
[28,]	0.02300429	-0.29641955	-0.3396924
[29,]	1.24223177	-0.28889212	-0.4183322
[30,]	1.24223177	-0.34108227	-0.3396924
[31,]	-1.19622319	-0.37169379	-0.4183322
[32,]	-1.19622319	-0.37169379	-0.3396924
[33,]	0.02300429	2.90022658	-0.4183322
[34,]	0.02300429	-0.19103561	-0.3396924
[35,]	1.24223177	-0.19655572	-0.4183322
[36,]	1.24223177	-0.34108227	-0.3396924
[37,]	-1.19622319	-0.37169379	-0.3396924
[38,]	-1.19622319	-0.37169379	3.5922973
[39,]	0.02300429	1.29437609	-0.3396924
[40,]	0.02300429	-0.22114531	3.5922973
[41,]	1.24223177	-0.20609046	-0.3396924
[42,]	1.24223177	-0.33506033	3.5922973
[43,]	-1.19622319	-0.37169379	-0.4183322
[44,]	-1.19622319	-0.37169379	-0.3396924
[45,]	0.02300429	-0.23318919	-0.4183322
[46,]	0.02300429	-0.30846343	-0.3396924

FIGURE 4. NORMALIZED DATASET

Furthermore, to determine the number of clusters in k-means approach, this research adopted an elbow method. The elbow technique is a heuristic for determining the number of clusters in a data set in cluster analysis [38]. Figure 5 contains two plots that assess the performance of clustering with Dindex values and their second-order differences to determine the optimal number of clusters: Left Plot (Red Line: Dindex Values vs. Number of Clusters): This plot shows how the distinctiveness of the clusters changes as the number of clusters increases. The Dindex values decrease with a steadily increasing number of clusters, indicating that the clusters become less well-separated as the number of clusters grows. This behavior suggests that fewer clusters (e.g., 3 to 5) provide a better separation of the data compared to a higher number of clusters. At the Right Plot (Blue Line: Second Differences of Dindex Values): This plot highlights the points where the rate of change in cluster separation slows down significantly. A noticeable drop in the second differences can be seen around 8 clusters, which may mark a possible solution, but this might also reflect instability in clustering quality. Before this point (3 to 5 clusters), the changes in Dindex values are more stable, implying better clustering quality.

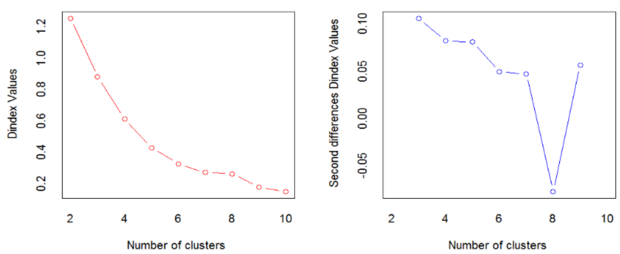


FIGURE 5. DETERMINATION OF NO CLUSTERS

After the determination of optimal clusters, four clustering algorithms was adopted for the clustering of the radiation data based on the location features of the data set. Figures 6 to 9 shows the clustering algorithms resulting clusters in K-means, Hierarchical clustering, DBSCAN clustering, and Gaussian Mixture Model respectively. Each method highlights different aspects of the dataset. DBSCAN is effective for identifying outliers, while K-Means and GMM provide clear, well-separated clusters for datasets with structured distributions. Hierarchical clustering is useful for understanding cluster relationships.

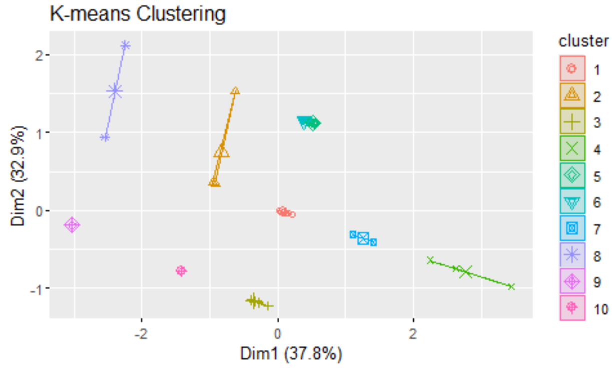


FIGURE 6. K-MEANS CLUSTERING

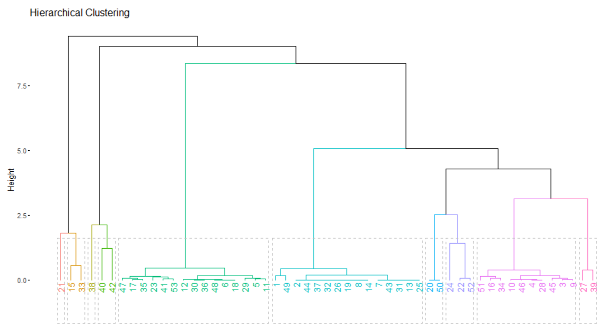


FIGURE 7. HIERARCHICAL CLUSTERING

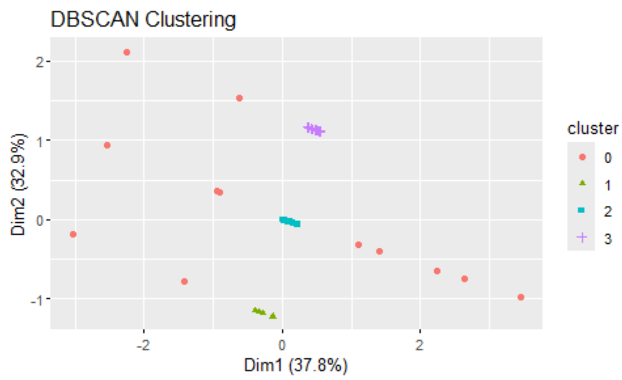


FIGURE 8. DDBSCAN CLUSTERING

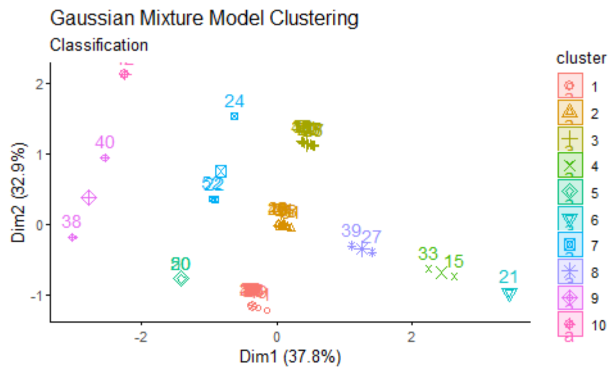


FIGURE 9. GAUSSIAN MIXTURE MODEL CLUSTERING

Using Silhouette Score as the performance measure of each clustering algorithm utilized in this study, the results of each individual models are presented in figure 10,11,12 and 13 respectively.

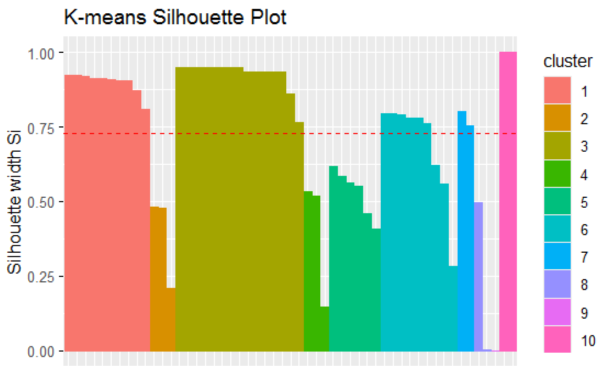


FIGURE 10. K-MEANS SILHOUETTE

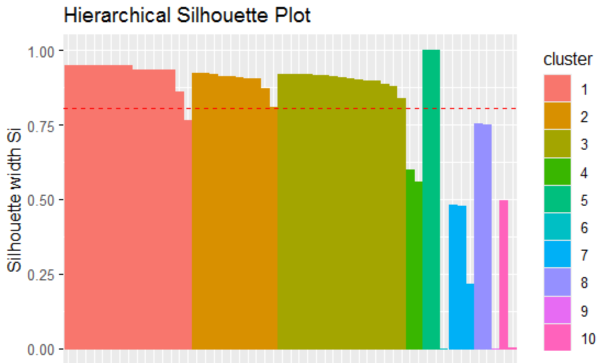


FIGURE 11. HIERARCHICAL SILHOUETTE

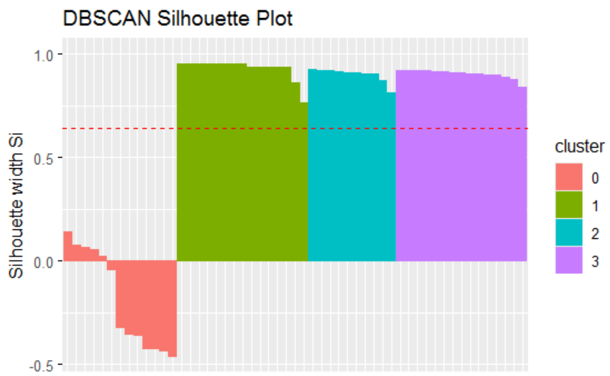


FIGURE 12. DBSCAN SILHOUETTE

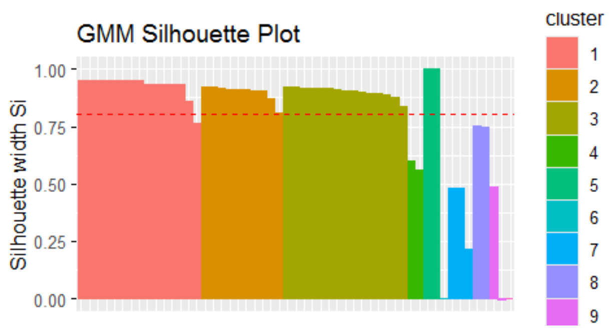


FIGURE 13. GMM SILHOUETTE

Furthermore, the average o Silhouette Score for each of the models were computed and the results are presented in Table 2.

TABLE 2. AVERAGE SILHOUETTE PERFORMANCE SCORES

Clustering Models	**Average Silhouette Width**
K-M means	0.73
Hierarchical	0.81
DBSCAN	0.64
GMM	0.81

Comparatively, the bar plot in figure 14 shows the performance comparison for the average Silhouette Width values for the four clustering algorithms: DBSCAN, GMM, Hierarchical, and K-means. Silhouette Width gives the quality of clustering, the higher the value, the better the cluster definition. In DBSCAN, this algorithm has the minimum average Silhouette Width compared to the other methods, as such it works less effectively in clustering this dataset as compared to the others. GMM is working well with a high average Silhouette Width, almost equal to that of the Hierarchical clustering method. This implies that GMM provides well-separated clusters in this dataset. Hierarchical method is the best-performing algorithm for this comparison, with the highest average Silhouette Width, slightly outperforming GMM.

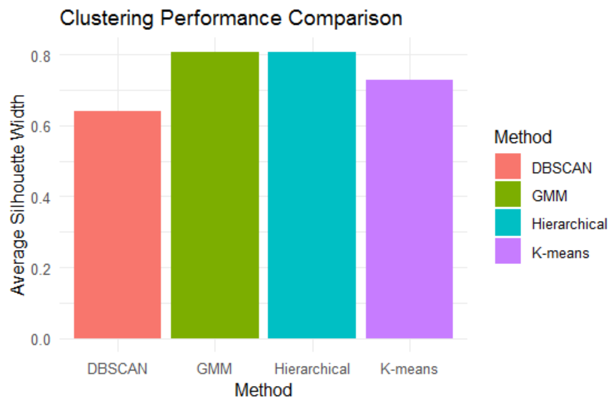
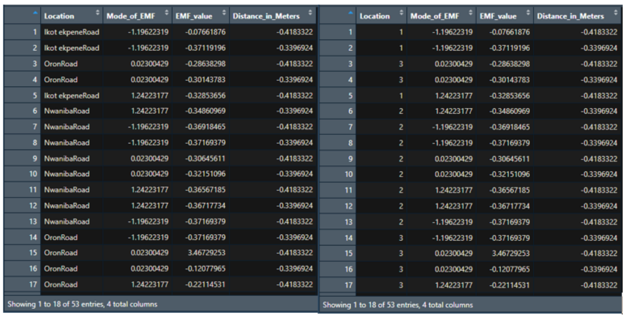


FIGURE 14. COMPARATIVE ANALYSIS OF CLUSTERING

Furthermore, in Figure 15, we scale the location column with other alongside other features in the data set so that we can then visualized from the clustering algorithm how each location varies based on the clusters.



	Location	Mode_of_EMF	EMF_value	Distance_in_Meters
1	Ikot ekpeneRoad	-1.19622319	-0.07661876	-0.4183322
2	Ikot ekpeneRoad	-1.19622319	-0.37119196	-0.3396924
3	OronRoad	0.02300429	-0.28638298	-0.4183322
4	OronRoad	0.02300429	-0.30143783	-0.3396924
5	Ikot ekpeneRoad	1.24223177	-0.32853656	-0.4183322
6	NwanibaRoad	1.24223177	-0.34860969	-0.3396924
7	NwanibaRoad	-1.19622319	-0.36918465	-0.4183322
8	NwanibaRoad	-1.19622319	-0.37169379	-0.3396924
9	NwanibaRoad	0.02300429	-0.30645611	-0.4183322
10	NwanibaRoad	0.02300429	-0.32151096	-0.3396924
11	NwanibaRoad	1.24223177	-0.36567185	-0.4183322
12	NwanibaRoad	1.24223177	-0.36717734	-0.3396924
13	NwanibaRoad	-1.19622319	-0.37169379	-0.4183322
14	OronRoad	-1.19622319	-0.37169379	-0.3396924
15	OronRoad	0.02300429	3.46729253	-0.4183322
16	OronRoad	0.02300429	-0.12077965	-0.3396924
17	OronRoad	1.24223177	-0.22114531	-0.4183322

Showing 1 to 18 of 53 entries, 4 total columns

	Location	Mode_of_EMF	EMF_value	Distance_in_Meters
1	1	-1.19622319	-0.07661876	-0.4183322
2	1	-1.19622319	-0.37119196	-0.3396924
3	3	0.02300429	-0.28638298	-0.4183322
4	3	0.02300429	-0.30143783	-0.3396924
5	1	1.24223177	-0.32853656	-0.4183322
6	2	1.24223177	-0.34860969	-0.3396924
7	2	-1.19622319	-0.36918465	-0.4183322
8	2	-1.19622319	-0.37169379	-0.3396924
9	2	0.02300429	-0.30645611	-0.4183322
10	2	0.02300429	-0.32151096	-0.3396924
11	2	1.24223177	-0.36567185	-0.4183322
12	2	1.24223177	-0.36717734	-0.3396924
13	2	-1.19622319	-0.37169379	-0.4183322
14	3	-1.19622319	-0.37169379	-0.3396924
15	3	0.02300429	3.46729253	-0.4183322
16	3	0.02300429	-0.12077965	-0.3396924
17	3	1.24223177	-0.22114531	-0.4183322

Showing 1 to 18 of 53 entries, 4 total columns

FIGURE 15. LOCATION SCALE DATA

From the location scaled, 1 represents Ikot Ekpene Road, 2- represent NwanibaRoad and 3 represent OronRoad respectively. The scaled locations alongside the other features can now be fed in to the defend clustering model built to determine the location that has a higher rate of ionizing radiation emission. Figure 16 shows the radiation emission based on the 3 different locations considered in this study, it is observed that location 1 forms the most clustered point for ionizing radiation followed by location 2 and 3 had a and insignificant clustered points.

This finding shows that exposure rate minimization of radiation in urban environment like in Uyo is highly required to curb the effects it might bring to long term residence in the area.

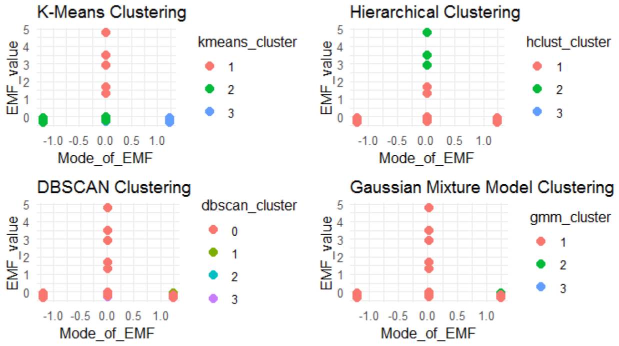


FIGURE 16. CLUSTERED LOCATION BASED ON RADIATION

5. CONCLUSIONS

This study demonstrates the applicability of machine learning clustering techniques in analyzing electromagnetic field radiation exposure in urban areas, focusing on Uyo Metropolis, Nigeria. Using data collected from three major roads which includes: Oron Road, Nwaniba Road, and Ikot Ekpene Road. the research provides critical insights into the spatial variability of EMF radiation levels. The clustering methods employed, such as K-Means, Hierarchical Clustering, DBSCAN, and Gaussian Mixture Model, have shown the efficiency of machine learning in identifying distinct radiation patterns. From the results, Hierarchical and GMM proved to be the best algorithms with the highest average Silhouette scores of 0.81, showing that the clusters were separated and clear. K-Means also performed moderately well with a Silhouette score of 0.73, but DBSCAN, which is supposed to handle noise as well handle noise as well, had the lowest score of 0.62 likely because it is sensitive toward parameters and due to the nature of the data. These findings prove the potential of clustering algorithms in environmental data analysis, yielding actionable insights for urban planners, environmentalists, and policymakers. This research extends the fast-growing area of EMF radiation analysis by offering a robust framework for classifying and interpreting spatial data. Future work may integrate more features such as population density, building materials, and health records, to further enrich the analysis. Furthermore, hybrid approaches that combines strengths of different clustering methods could be explored for further to improve the clustering accuracy. Ultimately, insights arising from this study can help develop policies while ensuring public health and sustainable urban development.

REFERENCES

[1] P. Ranjan, R. Rao, K. Kumar, and P. Sharma, *Wireless communication: Advancements and challenges*. 2022 [Online]. Available: https://www.taylorfrancis.com/books/mono/10.1201/9781003181699/wireless-communication-prashant-ranjan-ram-shringar-rao-krishna-kumar-pankaj-sharma

[2] B. B. Levitt, H. C. Lai, and A. M. Manville, "Effects of non-ionizing electromagnetic fields on flora and fauna, part 1. Rising ambient EMF levels in the environment," *Reviews on Environmental Health*, vol. 37, no. 1, pp. 81–122, Mar. 2022 [Online]. DOI: 10.1515/REVEH-2021-0026/MACHINEREADABLECITATION/RIS. [Accessed: 12-Apr-2025]

[3] M. E. Ekpenyong, D. E. Asuquo, and I. J. Umoren, "Evolutionary Optimisation of Energy-Efficient Communication in Wireless Sensor Networks," *International Journal of Wireless Information Networks*, vol. 26, no. 4, pp. 344–366, Dec. 2019 [Online]. DOI: 10.1007/S10776-019-00450-X/METRICS. [Accessed: 12-Apr-2025]

[4] M. F. ; Alomari, M. A. ; Mahmoud, R. Ramli, M. F. Alomari, M. A. Mahmoud, and R. Ramli, "A Systematic Review on the Energy Efficiency of Dynamic Clustering in a Heterogeneous Environment of Wireless Sensor Networks (WSNs)," *Electronics 2022, Vol. 11, Page 2837*, vol. 11, no. 18, p. 2837, Sep. 2022 [Online]. DOI: 10.3390/ELECTRONICS11182837. [Accessed: 12-Apr-2025]

[5] I. Shafi, M. Chaudhry, E. Montero, … E. A.-I., and undefined 2024, "A Review of Approaches for Rapid Data Clustering: Challenges, Opportunities and Future Directions," *ieeexplore.ieee.orgI Shafi, M Chaudhry, EC Montero, ES Alvarado, IDLT Diez, MA Samad, I AshrafIEEE Access, 2024•ieeexplore.ieee.org* [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10681061/

[6] I. Hussam, A. Elzanaty, Y. H. Al-Badarneh, and M.-S. Alouini, "EMF-aware cellular networks in RIS-assisted environments," *ieeexplore.ieee.orgH Ibraiwish, A Elzanaty, YH Al-Badarneh, MS AlouiniIEEE Communications Letters, 2021•ieeexplore.ieee.org* [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9576730/

[7] L. Hardell and M. Nilsson, "A Critical Analysis of the World Health Organization (WHO) Systematic Review 2024 on Radiofrequency Radiation Exposure and Cancer Risks," *Journal of Cancer Science and Clinical Therapeutics*, vol. 9, pp. 9–26, 2025. DOI: 10.26502/jcsct.5079261. [Accessed: 12-Apr-2025]

[8] D. Davis *et al.*, "Wireless technologies, non-ionizing electromagnetic fields and children: Identifying and reducing health risks," *Current Problems in Pediatric and Adolescent Health Care*, vol. 53, no. 2, p. 101374, Feb. 2023. DOI: 10.1016/J.CPPEDS.2023.101374. [Accessed: 12-Apr-2025]

[9] S. Akram and Y. S. Chowdhury, "Radiation Exposure Of Medical Imaging," *StatPearls*, Nov. 2022 [Online]. Available: http://europepmc.org/books/NBK565909

[10] B. Samaila and M. N. Yahaya, "Residential exposure to non-ionizing electromagnetic radiation from mobile base stations: a systematic review on biological effects assessment;" [Online]. DOI: 10.15406/mseij.2023.07.00203. [Accessed: 12-Apr-2025]

[11] C. Kurnaz and M. Mutlu, "Comprehensive radiofrequency electromagnetic field measurements and assessments: a city center example," *Environmental Monitoring and Assessment*, vol. 192, no. 6, pp. 1–14,

Jun. 2020 [Online]. DOI: 10.1007/S10661-020-08312-3/METRICS. [Accessed: 12-Apr-2025]

[12] A. Alshboul and M. Aqeel, "Enhancing Sustainable Urban Development: Towards Safe Electromagnetic Environment in Jordan, the Case of Amman," 2024 [Online]. DOI: 10.18280/ijsdp.191209. [Accessed: 12-Apr-2025]

[13] A. Aris, K. Yiannis, V. Softa, C. Kappas, C. Tyrakis, and K. Theodorou, "RF-EMF EXPOSURE LEVELS IN SENSITIVE LAND USE IN GREECE: EDUCATIONAL UNITS CENSUS IN THE MUNICIPALITY OF KORYDALLOS," *Radiation Protection Dosimetry*, vol. 190, no. 2, pp. 193–199, Aug. 2020 [Online]. DOI: 10.1093/RPD/NCAA090. [Accessed: 12-Apr-2025]

[14] J. Liu, M. Wei, H. Li, X. Wang, X. Wang, and S. Shi, "Measurement and mapping of the electromagnetic radiation in the urban environment," *Electromagnetic Biology and Medicine*, vol. 39, no. 1, pp. 38–43, Jan. 2020 [Online]. DOI: 10.1080/15368378.2019.1685540. [Accessed: 12-Apr-2025]

[15] M. Salma, M. Obayda, … I. N.-E. J. of, and undefined 2024, "Evaluating radiation exposure for patients during pelvic X-ray exams in Dhaka Metropolitan Area, Bangladesh," *ej-med.orgMU Salma, MA Obayda, IJ Nawreen, T ParvinEuropean Journal of Medical and Health Sciences, 2024•ej-med.org*, vol. 6, 2024 [Online]. DOI: 10.24018/ejmed.2024.6.1.2002. [Accessed: 12-Apr-2025]

[16] I. Umoren and S. Inyang, "Methodical performance modelling of mobile broadband networks with soft computing model," *researchgate.netIJ Umoren, SJ InyangInternational Journal of Computer Applications, 2021•researchgate.net*, vol. 174, no. 25, pp. 975–8887, 2021 [Online]. DOI: 10.5120/ijca2021921157. [Accessed: 12-Apr-2025]

[17] I. Umoren, D. Asuquo, O. Gilean, and M. Esang, "Performability of Retransmission of Loss Packets in Wireless Sensor Networks," *ComputerandInformation Science*, vol. 12, no. 2, 2019. DOI: 10.5539/cis.v12n2p71. [Accessed: 12-Apr-2025]

[18] M. T. H. Mallik, "Electromagnetic Field Exposure Reconstruction by Artificial Intelligence," Dec. 2023 [Online]. DOI: 10.34894/VQ1DJA. [Accessed: 12-Apr-2025]

[19] J. Zheng, Q. Lan, X. Zhang, W. Kainz, and J. Chen, "Prediction of MRI RF Exposure for Implantable Plate Devices Using Artificial Neural Network," *IEEE Transactions on Electromagnetic Compatibility*, vol. 62, no. 3, pp. 673–681, Jun. 2020. DOI: 10.1109/TEMC.2019.2916837. [Accessed: 12-Apr-2025]

[20] W. Dai, H. Cui, X. Kan, Y. Guo, S. Van Rooij, and C. Yang, "Transformer-Based Hierarchical Clustering for Brain Network Analysis," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2023-April, 2023. DOI: 10.1109/ISBI53787.2023.10230606. [Accessed: 12-Apr-2025]

[21] P. Zhou and Y. Chang, "Automated classification of building structures for urban built environment identification using machine learning," *Journal of Building Engineering*, vol. 43, Nov. 2021. DOI: 10.1016/J.JOBE.2021.103008. [Accessed: 12-Apr-2025]

[22] M. Mallik *et al.*, "Towards Outdoor Electromagnetic Field Exposure Mapping Generation Using Conditional GANs," *Sensors 2022, Vol. 22, Page 9643*, vol. 22, no. 24, p. 9643, Dec. 2022 [Online]. DOI: 10.3390/S22249643. [Accessed: 12-Apr-2025]

[23] D. Xu, Y. T.-A. of data science, and undefined 2015, "A comprehensive survey of clustering algorithms," *SpringerD Xu, Y TianAnnals of data science, 2015•Springer*, vol. 2, no. 2, pp. 165–193, Jun. 2015 [Online]. DOI: 10.1007/s40745-015-0040-1. [Accessed: 12-Apr-2025]

[24] G. Zhu, X. Li, S. Zhang, X. Xu, and B. Zhang, "An improved method for k-means clustering based on internal validity indexes and inter-cluster variance," *International Journal of Computational Science and Engineering*, vol. 25, no. 3, p. 253, 2022. DOI: 10.1504/IJCSE.2022.123112. [Accessed: 12-Apr-2025]

[25] T. Dinh *et al.*, "Data clustering: an essential technique in data science," Dec. 2024 [Online]. Available: http://arxiv.org/abs/2412.18760

[26] M. Ouabdou, F. Zohra Diar, Y. Benbouras, B. El Fahime, and M. Radouani, "Innovations in Defect Detection: Integrating AI and NDT Techniques in Composite Materials," *researchgate.netM Ouabdou, Y Benbouras, FZ Diar, B El Fahime, M RadouaniRevue des Composites et des Materiaux Avances, 2025•researchgate.net* [Online]. DOI: 10.18280/rcma.350119. [Accessed: 12-Apr-2025]

[27] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020. DOI: 10.1109/ACCESS.2020.2988796. [Accessed: 12-Apr-2025]

[28] R. Vankayalapati, K. Balaso Ghutugade, R. Vannapuram, and B. Pooja Sree Prasanna, "K-Means algorithm for clustering of learners performance levels using machine learning techniques.," *academia.eduR Vankayalapati, KB Ghutugade, R Vannapuram, BPS PrasannaRev. d'Intelligence Artif., 2021•academia.edu* [Online]. DOI: 10.18280/ria.350112. [Accessed: 12-Apr-2025]

[29] A. Sangaiah, A. Javadpour, … F. J.-I. transactions on, and undefined 2022, "Hierarchical clustering based on dendrogram in sustainable transportation systems," *ieeexplore.ieee.orgAK Sangaiah, A Javadpour, F Ja'fari, W Zhang, SM KhaniabadiIEEE transactions on intelligent transportation systems, 2022•ieeexplore.ieee.org* [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9971785/

[30] L. Yin, M. Li, H. Chen, and W. Deng, "An Improved Hierarchical Clustering Algorithm Based on the Idea of Population Reproduction and Fusion," *Electronics 2022, Vol. 11, Page 2735*, vol. 11, no. 17, p. 2735, Aug. 2022 [Online]. DOI: 10.3390/ELECTRONICS11172735. [Accessed: 12-Apr-2025]

[31] D. Deng, "DBSCAN Clustering Algorithm Based on Density," *Proceedings - 2020 7th International Forum on Electrical Engineering and Automation, IFEEA 2020*, pp. 949–953, Sep. 2020. DOI:

10.1109/IFEEA51475.2020.00199. [Accessed: 12-Apr-2025]

[32] M. A. Sattarov, “PARALLELIZATION METHODS OF DATA MINING ALGORITHMS: ENHANCING PERFORMANCE IN THE AGE OF BIG DATA,” *researchweb.uzMA*, vol. 3, no. 12, 2024 [Online]. DOI: 10.5281/zenodo.14567380. [Accessed: 12-Apr-2025]

[33] A. Sarma *et al.*, “μdBSCAN: An Exact Scalable DBSCAN Algorithm for Big Data Exploiting Spatial Locality,” *Proceedings - IEEE International Conference on Cluster Computing, ICCC*, vol. 2019-September, Sep. 2019. DOI: 10.1109/CLUSTER.2019.8891020. [Accessed: 12-Apr-2025]

[34] A. Awaad, H. H.-I. J. of C. and Digital, and undefined 2023, “Parallel Implementation of Statistical DBSCAN Algorithm for Spark-based Clustering on Google Cloud Platform,” *inass.orgAM Awaad, H HefnyInternational Journal of Computing and Digital Systems, 2023•inass.org*, vol. 16, no. 2, p. 2023 [Online]. DOI: 10.22266/ijies2023.0430.23. [Accessed: 12-Apr-2025]

[35] M. Löffler, A. Zhang, H. Z.-T. A. of Statistics, and undefined 2021, “Optimality of spectral clustering in the Gaussian mixture model,” *projecteuclid.orgM Löffler, AY Zhang, HH ZhouThe Annals of Statistics, 2021•projecteuclid.org*, vol. 49, no. 5, pp. 2506–2530, 2021 [Online]. DOI: 10.1214/20-AOS2044. [Accessed: 12-Apr-2025]

[36] A. Ekong, A. Silas, and S. Inyang, “A Machine Learning Approach for Prediction of Students’ Admissibility for Post-Secondary Education using Artificial Neural Network,” *researchgate.netA Ekong, A Silas, S InyangInt. J. Comput. Appl, 2022•researchgate.net*, vol. 184, no. 27, pp. 975–8887, 2022 [Online]. DOI: 10.5120/ijca2022922340. [Accessed: 12-Apr-2025]

[37] I. Umoren, S. Inyang, and A. Ekong, “A Fuzzy Knowledge-Based System for Modeling Handoff Prediction in Mobile Communication Networks” [Online]. Available: https://www.researchgate.net/publication/363914597

[38] S. Inyang and I. Umoren, “Semantic-Based Natural Language Processing for Classification of Infectious Diseases Based on Ecological Factors,” *International Journal of Innovative Research in Sciences and Engineering, 2023*, vol. 3, no. 7, 2023 [Online]. Available: https://www.researchgate.net/profile/Saviour-Inyang/publication/372983540_Semantic-Based_Natural_Language_Processing_for_Classification_of_Infectious_Diseases_Based_on_Ecological_Factors/links/64d25420806a9e4e5cf99bc9/Semantic-Based-Natural-Language-Processing-for-Classification-of-Infectious-Diseases-Based-on-Ecological-Factors.pdf

AUTHORS

Imeh J. Umoren

He holds B.Sc., M.Sc., and Ph.D. degrees in Computer Science. With over 12 years of experience in academia, research, and the global ICT industry, he currently serves as the Acting Provost of the School of Computing and Information Technology, FUTIA, Ikot Abasi, Nigeria. A recipient of the Total E&P Graduate Scholarship (2010–2012), he is a strategic thinker passionate about leveraging computational intelligence for development. Dr. Umoren is a member of IEEE, NCS, Polish Neural Networks Society, and the Internet Society, with 60+ publications. His research interests include Mobile Computing, Algorithms, Complexity Theory, ML, and NLP.

Saviour J. Inyang

He holds a National Diploma, B.Sc. (CGPA 4.40), and M.Sc. (CGPA 4.53) in Computer Science from Akwa Ibom State University, Nigeria. He lectures in the Department of Computing at Topfaith University, Mkpatak. His research interests span Natural Language Processing, Machine Learning, Soft Computing, Computational Intelligence, Artificial Intelligence, and Communication Network Modeling. He has authored and co-authored several papers in reputable local and international journals. Committed to excellence, he is dedicated to offering his expertise to organizations.

Ubong Etuk

He holds a BSc from the University of Calabar, an MSc from the University of Ibadan, and a PhD from the University of Glasgow. With 12 years of lecturing experience, he has contributed significantly to academia through both teaching and research. His research interests span software engineering, information security and privacy, distributed systems, and machine learning.

Daniel Essien

He is currently pursuing a Ph.D. in Computer Science at Dalhousie University, Canada with a research focus on multimodal data fusion for livestock welfare monitoring using advanced machine learning techniques. His areas of interest include machine learning for digital agriculture, IoT and cybersecurity. Currently, he serves as a research assistant at Dalhousie University, contributing to cutting-edge research in livestock behavior analysis using AI.