Predicting Energy Efficiency in 5G Advanced Network using Machine Learning

Ifeoma B. Asianuba
Department of Electrical/Electronic Engineering
University of Port Harcourt
Choba, Port Harcourt

Ugochukwu Nnamdi
Cybersecurity Analyst
IRCMS
Lagos, Nigeria

ABSTRACT
5G Advanced network is known to have substantial energy consumption challenges due to its massive machine-type communication, enhanced mobile broadband and ultra reliable low latency communications. Traditional optimization approach is limited by the high dimensional, dynamic and nonlinear nature of 5G networks. Machine learning is therefore deployed to provide an autonomous, adaptive and predictive energy optimization approach. This study therefore evaluates the performance of three regression models; Linear regression, decision tree regression, and Random Forest regression for predicting energy consumption based on key input features which includes; load, transmission power (TXpower), and energy-saving mode (ESMODE). Model performance was assessed using mean error/mean absolute error (MAE), R-squared values, and visual analysis through scatter plots comparing actual and predicted energy values. The Linear regression model achieved a mean error of 85.91 and an R-squared value of 0.55, indicating limited predictive capability and difficulty in capturing nonlinear relationships, particularly at higher energy levels where underprediction was observed. The Decision tree regression model showed improved performance with a mean absolute error of 41.82 and an R-squared value of 0.78, effectively modeling feature interactions but exhibiting underprediction at high energy values and discrete prediction patterns. The Random Forest Regression model delivered the best results, with a mean absolute error of 36.57 and an R-squared value of 0.81, demonstrating strong predictive accuracy and better generalization across energy ranges. Overall, the results indicate that ensemble-based methods, particularly Random Forest Regression, are more suitable for accurate energy prediction in this context, while simpler linear models are less effective due to their inability to model complex relationships in the data.

Keywords
5G, Decision Trees Machine Learning, Regression, Random Forest.

1. INTRODUCTION
The growth in the generation of cellular network which hinges on the growing demand for increased bandwidth, latency and throughput is also associated with a corresponding increased energy consumption. The evolutionary change in cellular network has transcended from simple wireless calls to high-speed internet applications. It is known that 5G consumes four times more energy than 4G due to the tendency to accommodate additional connected device alongside their corresponding applications [1]. It is designed to accommodate vast number of users without compromising the quality of service or introducing interferences.

Substantial amount of energy consumption is experienced at base station and access point levels. This is heightened by the deployment of small cells for network densification and massive MIMIO infrastructure. However, to fully harness the potentials of 5G, enabling technologies which include; cloud RAN (CRAN), software defined network, network function visualization, millimeter wave application, ultra dense network (UDN), mMTC, M2M, and D2D communication are applied at various cellular network operation. It is however expected that 6G will confront the challenges of energy consumption associated with earlier mobile generation by incorporating AI driven processes and optimization resources. 6G is also associated with cutting edge technology such as computational oriented communication (COM), machine type communication with the intension to connect everyone and everything and then cellular network communication which covers base station power and control activities. It is expected that 6G will utilize satellite communication for signal transmission rather than rely absolutely on the conventional base stations to achieve extensive mobile coverage and enhanced spectral efficiency. 6G will pose improved security features with advanced cryptographic algorithm to address the growing issue of data protection in mobile networks. It will support the development and deployment of smart objects, smart cities and IoT applications for improved integration and communication among devices. 6G will display high time as well as good phase synchronization accuracy better than what 5G can offer [2]. Although 6G is yet to be deployed as at the time of this research, there is need to investigate the energy challenges of the 5G Advanced network.

2. REVIEW OF RELATED WORKS
The advancement in technology which attracts the massive demand for data and the drive for the reduction of energy consumption towards the net zero agenda presents new dimension of challenge for network operation. However, this paper presents different machine models to predict energy consumption in 5G networks.

[3] proposed machine learning models to optimize power consumption in 5G base stations. The model leverages data obtained from a large-scale campaign. It is anticipated to serve as a foundational tool for optimizing network energy efficiency and gaining insights into the power consumption of 5G base stations (BSs). These studies highlight promising strategies for enhancing energy efficiency in 5G networks, particularly in base stations and radio interfaces [4] used different machine Learning techniques in developing and predicting models for energy consumption in 5G networks. XGBoost, CatBoost, Artificial Neural Network and ensemble method were used. Model performance matrices like mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the models. The ensemble method outperformed the other models achieving lowest MAPE (3.5620), RMSE (1.1524) and MAE (1.0245). In [5], an extensive review on current advancement in the

integration of ML and AI with wireless communication channel with emphasis on channel modelling, estimation and network management was discussed. The use of AI and ML to solve the unstructured and difficult problems which finds applications in 5G and Beyond (B5G) was proposed. It relied on the extracted wireless data features obtained from experimental findings thereby combining data driven insights with model-based approach to achieve accurate models in the next generation wireless communication networks.

[6] employed Gaussian mixture models (GMM) to improve energy efficiency. The GMM was trained with expectation-minimization for model development and decision making carried out with reinforcement learning. This model reduced energy per traffic unit by 14% and increased network throughput by 11% compared to baseline conditions

In [7], the research considered the state-of-art application of machine learning techniques in the 5G network to enable energy efficiency at the access, edge, and core network. The authors provided a taxonomy of machine learning applications in 5G networks to increase energy efficiency. A number of problems with energy efficiency in 5G networks that can be resolved with machine learning were discussed. Lastly, the authors further considered a number of issues that must be resolved in order to fully utilize machine learning's promise to increase 5G networks' energy efficiency.

[8] used machine learning techniques to improve the energy efficiency of 5G networks. The algorithms used includes random-forest algorithm, the lasso algorithm, the gradient boosting algorithm, the XGboost algorithm and the ridge stacking regression algorithm. From the results realised, the stacking algorithm outperformed the other individual algorithms while the XGboost was the best individual algorithm.

[9] offered a novel machine learning-based method that uses application-level data and passive network quality indicators to forecast the resulting uplink transmission power utilized for data transmissions. It is based on extensive field measurements of drive tests conducted in a public cellular network. With a mean average error of 3.166 dB, Random-Forest models clearly outperformed the other two machine learning techniques. The method worked well for long-term power estimates since the absolute total of errors converges to zero and, on average, drops below 1 dB after 28 predictions. The authors in [10] developed an optimized 5G base station energy consumption by improving on the traditional base station sleep strategy. The ECO-BS algorithm consists of application of interference to model real life systems, capacity limits of the user equipment and dynamic changes of UE to maximize the sleep number of BSs in real time. [11] developed a distributive intuitive online learning power allocation algorithm for multi-tier 5G heterogenous network which was used to reduce total power used while maintaining quality of service.

3. METHODOLOGY

Conventional optimization techniques such as convex optimization and heuristic algorithms adopts simplified models which require a complete knowledge of the system. This however makes them computationally expensive, poorly scalable and inefficient in real-time adaptive situations.

In the machine learning approach deployed in this study, the **5G-Energy consumption** [12] dataset was obtained from Kaggle and was provided by the International Telecommunication Union (ITU). It includes cell-level traffic statistics of 4G/5G sites collected on different days. A snippet of the data set is shown below in figure 1.

	Time	BS	Energy	load	ESMODE	TXpower
0	20230101 010000	B_0	64.275037	0.487936	0.0	7.101719
1	20230101 020000	B_0	55.904335	0.344468	0.0	7.101719
2	20230101 030000	B_0	57.698057	0.193766	0.0	7.101719
3	20230101 040000	B_0	55.156951	0.222383	0.0	7.101719
4	20230101 050000	B_0	56.053812	0.175436	0.0	7.101719

Fig 1: Snippet of the dataset.

The initial procedure required the dataset to be cleaned. The timestamp was dropped as the linear models and it does not require time variables. The BS (Base station) column was also dropped as it has too many unique variables. Linear regression, Decision tree regression and Random tree regression models were used to predict the 5G energy.

Linear regression, a supervised machine learning algorithm, finds a linear relationship to explain the correlation between the target and feature(s) variables. The general formula for linear regression is given as equation 1.

$$y = b_0 + b_1 x \qquad (1)$$

Where:

y = target or dependent variable

x = feature(s) or independent variable(s)

b_0 = constant

Decision tree regression uses tree-like structures for its model prediction. The model works by breaking the data into smaller parts based on simple rules taken from the input features thereby reducing errors in prediction. The general formula is given in equation 2.

$$f(x) = \sum_{m=1}^{M} c_m \cdot 1_{x\varepsilon R_m} \qquad (2)$$

M = number of leaf nodes (regions)

R_m = is the region (set of conditions or rules)correspondimg to the mth leaf

c_m = constant prediction value for region R_m

$1_{x\varepsilon R_m}$ = indicator function: 1 if x belongs to region R_m, 0 otherwise

The Random Forest Regressor uses an ensemble method to build multiple decision trees and outputs the average of their predictions. The random forest prediction is given by the formula in equation 3 below.

$$y(x) = \frac{1}{n}\sum_{i=1}^{n} T_i(x) \qquad (3)$$

Where:

$y(x)$ = final prediction for input x

$T_i(x)$ = predictions from each n individual decision trees

The different models are then validated using different r2 scores and mean square errors scores. The best model is then chosen. The general methodology is described in figure 2.

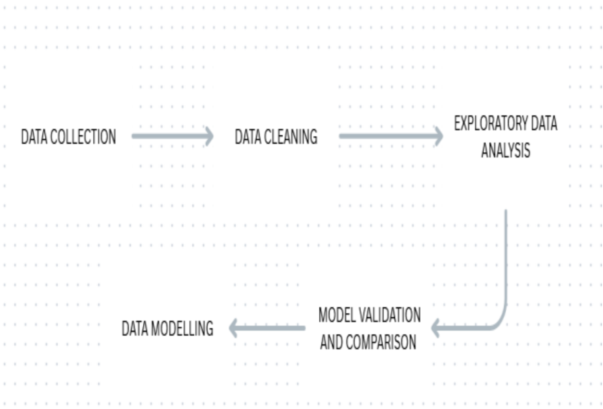

Fig 2: Methodology

R2 Scores: A statistical metric used in machine learning to assess a regression model's quality is called R-squared, or the coefficient of determination. By evaluating the percentage of variance in the dependent variable that the independent variables account for, it determines how well the model fits the data [13].

R2 can be calculated with the equation below.

$$R^2 = 1 - \frac{SSE}{SST}$$

Where SSE = Sum of Squares of Errors, SST = the Total Sum of Squares.

Mean Square error: The mean of the squared discrepancies between actual and anticipated outputs in a regression model is known as mean square error (MSE), and it is used to assess prediction accuracy. Because of the squaring of residual errors, it is also known as L2 loss and is less resilient to outliers [14].

4. RESULTS AND DISCUSSION

The heat map is shown below in figure 3. Energy is strongly affected the most by load and moderately by TXpower, while ESMODE has a small negative effect. This was very insightful for feature selection and modelling.

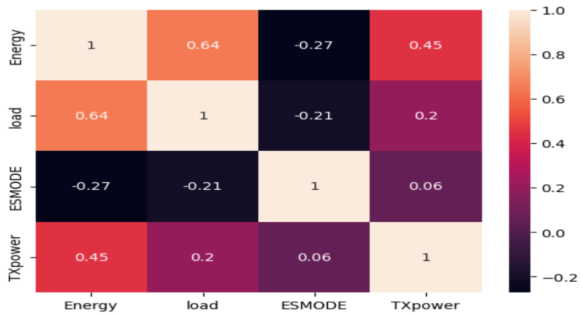

Fig 3: Heatmap showing correlation between Energy and other variables

The linear regression model had a mean error rate of 85.91 and a r-squared value of 0.55. Figure 4 shows a scatter plot comparing actual vs predicted energy values. The figure suggests that the model performs fairly well overall. However, the model struggled with high energy values, producing underpredictions. The model obtained from the Linear regression is given below:

$$\text{Energy} = -87.10 + (31.86 * \text{load}) + (-6.59 * \text{ESMODE}) + (15.96 * \text{TXpower})$$

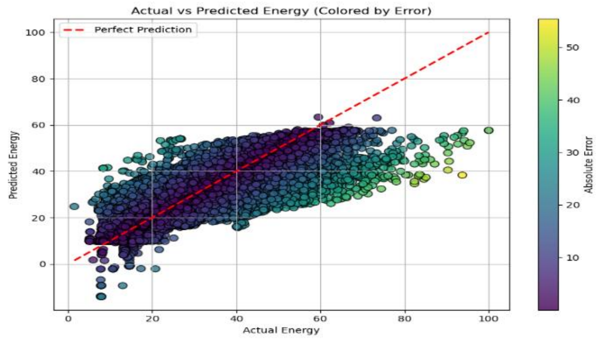

Fig 4: Scatter plot comparing actual vs. predicted Energy values for linear regression model

The decision tree regression model had a mean absolute error of 41.82 and r-squared value of 0.78. Figure 5 shows the comparison of the actual energy values (x-axis) to predicted energy values (y-axis), with each point colored based on the absolute error. The plot reveals that the model performs reasonably well at lower energy levels but tends to underpredict at higher actual values, as many points fall below the perfect prediction line. The presence of horizontal bands in the predicted values suggests the model often predicts a limited set of outputs, indicating possible bias or rounding. Overall, the model's accuracy declines as actual energy increases.

Figure 6 describes the visualization of a Decision Tree Regressor used to predict a numerical target based on features like TXpower, load, and ESMODE. Each node represents a decision rule that splits the data, with information about the number of samples, squared error (variance), and the predicted value at that point. The tree starts with a split on TXpower and continues dividing the data based on feature thresholds to minimize prediction error. Darker nodes indicate higher predicted values. Overall, the tree shows how combinations of input features guide the model in making predictions, with TXpower and load playing key roles.

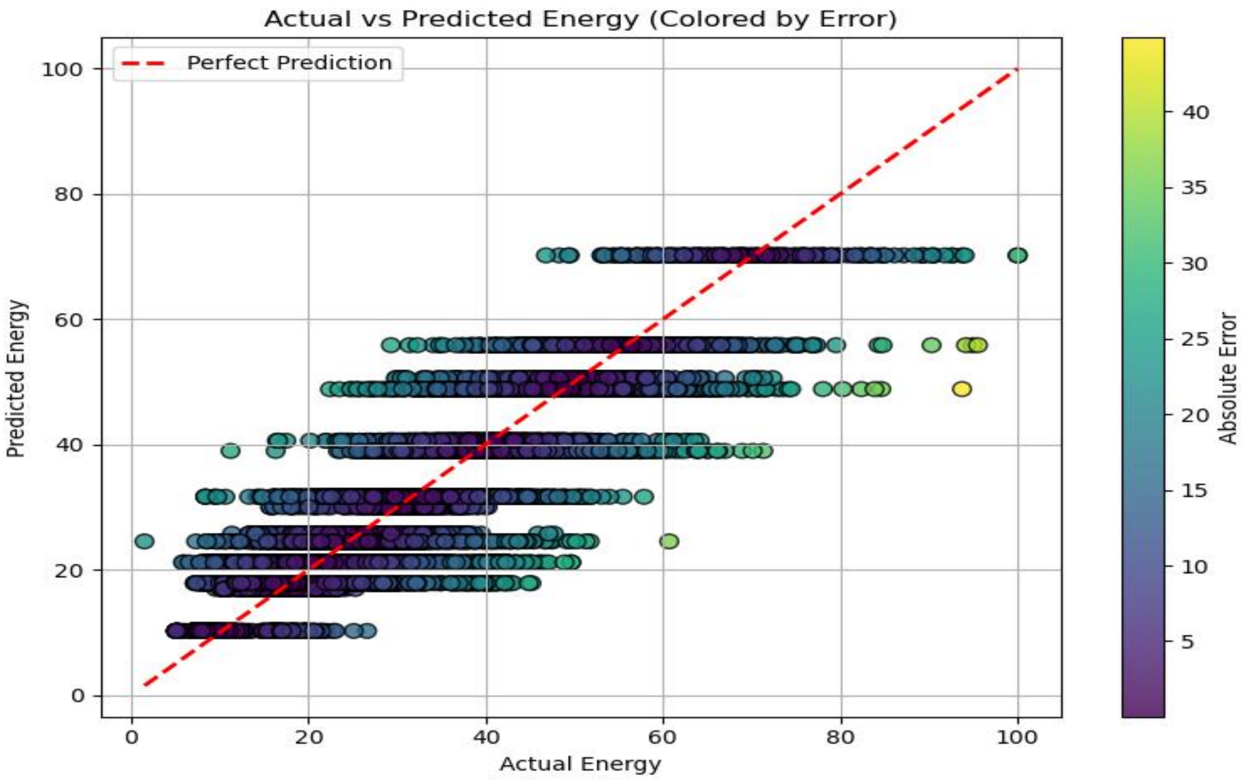

Figure 5: scatter plot comparing actual vs. predicted Energy values for Decision tree regression model

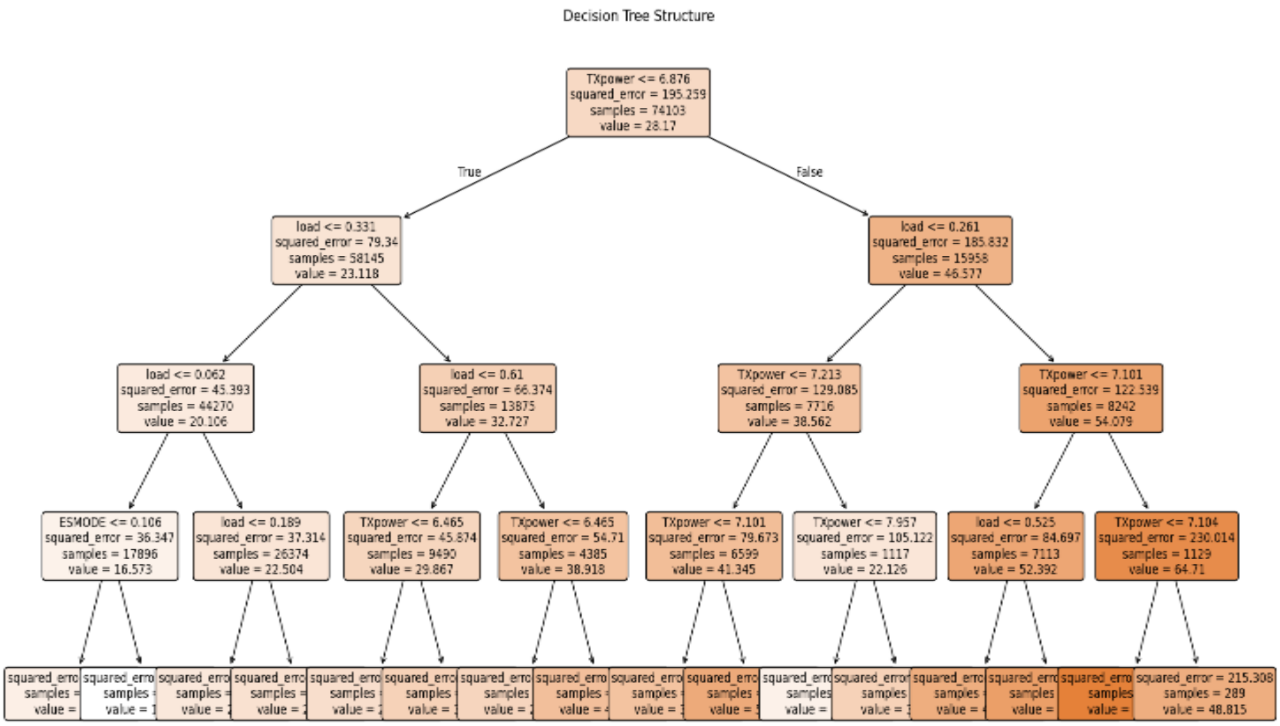

Figure 6: Decision tree structure

Finally, the random forest regression model had a mean absolute error of 36.57 and an r squared value of 0.81. This scatter plot in figure 7 shows the relationship between actual energy (x-axis) and predicted energy (y-axis). Most points are tightly clustered around the red line, suggesting the model performs well overall, especially for mid-range energy values. The colors are mostly dark (low error), but some green and yellow points (high error) appear, especially at higher actual energy levels, indicating the model occasionally struggles with extreme values.

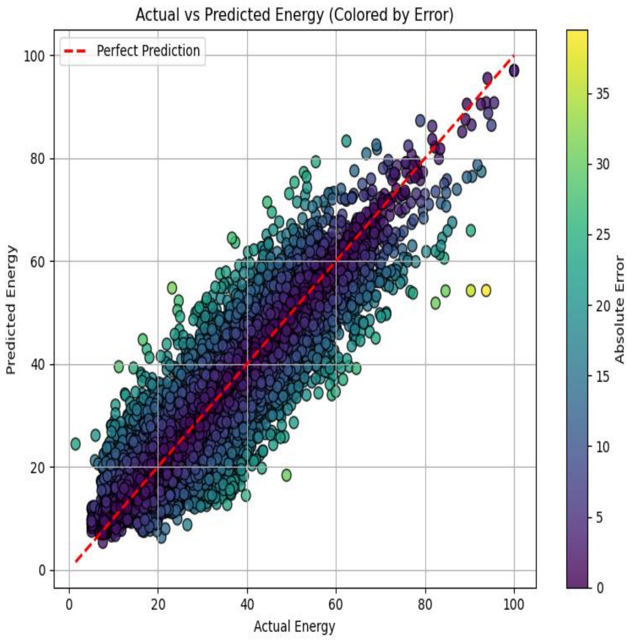

Figure 7: scatter plot comparing actual vs. predicted Energy values for random forest regression model

The MAE and r-squared values for the models are shown in table 1 below.

Table 1: Table showing MAE and r-squared value of each model

Model	MAE	R-SQUARED
Linear regression	85.91	0.55
Decision tree regression	41.87	0.78
Random forest regression	36.57	0.81

5. CONCLUSION

In 5G network, energy efficiency optimization is fundamental for a sustainable system. The inherent challenges of traditional optimization approach make it inefficient for 5G network which is known for its heterogenous, dynamic and complex nature.

Machine learning models were considered. Among the three models evaluated: Linear Regression, Decision Tree Regression, and Random Forest Regression. The Random Forest model demonstrated the best overall performance with the lowest mean absolute error (36.57) and the highest R-squared value (0.81), indicating a strong ability to explain the variability in energy consumption. Its scatter plot showed a tight clustering of points around the perfect prediction line, especially for mid-range energy values, although some underprediction still occurred at higher levels.

The Decision Tree Regression model also performed well, with a mean absolute error of 41.82 and an R-squared value of 0.78. While it captured relationships in the data more effectively than the linear model, it showed signs of prediction bias, particularly due to its tendency to produce discrete output values, as seen in the horizontal banding in the scatter plot. The decision tree structure visualization confirmed that features like TXpower and load were key influencers in the model's predictions. The Linear Regression model, with a mean error rate of 85.91 and an R-squared value of 0.55, performed the weakest. It captured the general trend but lacked precision, especially for high energy values where it significantly underpredicted. The linear nature of the model also limited its ability to capture complex patterns in the data. Overall, the Random Forest model offers the most reliable and accurate predictions for energy estimation, followed by the Decision Tree, while the Linear Regression model is best suited for simpler, more linear relationships

6. REFERENCES

[1] T. P. Fowdur and B. Doorgakant, "A review of machine learning techniques for enhanced energy efficient 5G and 6G communications," Eng. Appl. Artif. Intell., vol. 122, p. 106032, 2023, doi: https://doi.org/10.1016/j.engappai.2023.106032.

[2] Y.-C. Liang, D. Niyato, E. G. Larsson, and P. Popovski, "Guest editorial: 6G mobile networks: Emerging technologies and applications," China Commun., vol. 17, no. 9, pp. 90–91, 2020, doi: 10.23919/JCC.2020.9205979.

[3] N. Piovesan, D. Lopez-Perez, A. De Domenico, X. Geng, H. Bao, and M. Debbah, "Machine Learning and Analytical Power Consumption Models for 5G Base Stations," IEEE Commun. Mag., vol. 60, no. 10, pp. 56–62, Oct. 2022, doi: 10.1109/MCOM.001.2200023.

[4] I. Makroz, A. Nafidi, R. Gutiérrez-Sánchez, and S. Aourik, "A machine learning approach for energy consumption in 5G networks," ITU J. Futur. Evol. Technol., vol. 5, no. 4, pp. 447–457, 2024, doi: 10.52953/ijxz5881.

[5] C.-X. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial Intelligence Enabled Wireless Networking for 5G and Beyond: Recent Advances and Future Challenges," IEEE Wirel. Commun., vol. 27, no. 1, pp. 16–23, 2020, doi: 10.1109/MWC.001.1900292.

[6] Z. Alsalami, R.Kasthuri, S. S., P. Rani, D. K. Gupta, and I. V. G. K. Reddy, "Energy Efficiency in 5G Networks Using Gaussian Mixture Models," in 2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS), 2024, pp. 1–5. doi: 10.1109/IICCCS61609.2024.10763881.

[7] A. Mughees, M. Tahir, M. A. Sheikh, and A. Ahad, "Towards energy efficient 5G networks using machine learning: Taxonomy, research challenges, and future research directions," IEEE Access, vol. 8, pp. 187498–187522, 2020, doi: 10.1109/ACCESS.2020.3029903.

[8] C. R. Okpara, O. P. Ngwu, V. E. Idigo, and C. K. Agubor, "Improving The Energy Efficiency of A 5G Network: The Machine Learning Approach," IRE, vol. 2, no. 1, pp. 111–126, 2023.

[9] R. Falkenberg, B. Sliwa, N. Piatkowski, and C. Wietfeld, "Machine Learning Based Uplink Transmission Power Prediction for LTE and Upcoming 5G Networks Using Passive Downlink Indicators," IEEE Veh. Technol. Conf., vol. 2018-Augus, 2018, doi: 10.1109/VTCFall.2018.8690629.

[10] X. Ma, Y. Mu, Z. Liu, X. Jiang, J. Zhang, and Y. Gao, "Energy consumption optimization of 5G base stations considering variable threshold sleep mechanism," Energy Reports, vol. 9, pp. 34–42, 2023, doi: 10.1016/j.egyr.2023.04.026.

[11] I. AlQerm and B. Shihada, "Energy-Efficient Power Allocation in Multitier 5G Networks Using Enhanced Online Learning," IEEE Trans. Veh. Technol., vol. 66, no. 12, pp. 11086–11097, 2017, doi: 10.1109/TVT.2017.2731798.

[12] ITU, "Kaggle," 5G-Energy consumption Dataset. [Online]. Available: https://www.kaggle.com/datasets/nadiatriki/5g-energy-consumption-dataset

[13] Arize, "Arize," R Squared: Understanding the Coefficient of Determination. [Online]. Available: https://arize.com/blog-course/r-squared-understanding-the-coefficient-of-determination/./

[14] K. Tyagi, C. Rane, Harshvardhan, and M. Manry, "Chapter 4 - Regression analysis," in Artificial Intelligence and Machine Learning for EDGE Computing, R. Pandey, S. K. Khatri, N. kumar Singh, and P. B. T.-A. I. and M. L. for E. C. Verma, Eds., Academic Press, 2022, pp. 53–63. doi: https://doi.org/10.1016/B978-0-12-824054-0.00007-1.