Comparative Analysis of the Features of a 5G Network Production Dataset: The Machine Learning Approach

Chinedu R. Okpara, Victor E. Idigo, and Chukwunenye S. Okafor

***Abstract* — 5G networks deployment is much data driven, leading to more energy consumption. The need to efficiently manage this energy consumption is a major drive in the comparative analysis of the features of a 5G production dataset. The features of the 5G production dataset generated with G-Net track pro were analyzed using Python programming language. From the correlation coefficient results obtained, the highest correlation value of 0.78 exists between the reference signal power and the received signal reference power of the neighbouring cells. Using the significant indicator, we observed that the signal to noise ratio is the most important of all the features. Using heat map and scatter plots, we further observed that there were good relationships between the key features selected from the significant indicator. These features will play a big role in improving the energy efficiency of a 5G network.**

***Keywords* — 5G, Artificial Intelligence, Dataset, Machine Learning.**

I. Introduction

5G networks deployment supports large number of devices and requires high data rates, needing too much energy consumption. The need to efficiently manage this energy is a major drive in determining which features of a production dataset plays key roles in improving the energy efficiency of a 5G network. According to [1] cellular technologies have seen gradual evolution from the first to the fifth evolution (5G) for meeting the demands in terms of bandwidth, throughput, latency and jitter. Panwar, Sharma and Singh [2] stated that each generation gave rise to energy consumption due to the addition of hardware to support applications and requirements. According to [3], in 2025 the amount of user data will increase four times compared to today's network. As a result, energy efficiency will be a significant factor in 5G as compared to earlier generations. In [4], it is explained that a fully operative and efficient 5G network cannot be complete without artificial intelligence (AI) and by integrating machine learning (ML) into 5G technology, intelligent base stations will be able to make decisions for themselves, and mobile devices will be able to create dynamically adaptable clusters based on learned data.

II. Reviewed Literature

Fonseca *et al.* [9] discussed that 5G networks are being designed to provide pervasive networking, high data rates, coverage, reliability and low latency and that meeting such diverse requirements has also resulted in increased ICT energy consumption; by 2025, the ICT industry itself could be responsible for 30% of power consumption globally. Johnson [5] tried to solve the problem of 5G energy consumption by the deployment of small cells. This makes the networks denser, leading to more energy consumption. Bjornson *et al.* [10] explained that to improve the cellular energy efficiency, without sacrificing Quality of Service (QoS) at the users, the network topology must be densified to enable higher spatial re-use. Their goal was to minimize the total power consumption while satisfying QoS constraints at the users and power constraints at the BS and small cell access points (SCAs). Rajoria *et al.* [6] employed massive MIMO, which increases power consumption due to the more hardware components required. The authors in [11] tried to answer the fundamental questions; what the optimal number of antennas are, active users and transmit power, for a multi-user MIMO system to be designed from scratch to uniformly cover a given area with maximal energy efficiency. Haidine *et al.* [7] states that AI and ML will unlock the power of software and algorithms that will allow for efficient deployment of assets and resources. In this work, ML approach was used in determining the key features of a 5G production dataset for improving the energy efficiency of a 5G network.

III. Materials and Method

5G production dataset of [8] generated from two mobility patterns (static and car), and across two application patterns (video streaming and file download); composed of client-side cellular key performance indicators (KPIs), comprised of channel related metrics, context-related metrics, cell-related metrics and throughput information was used. These metrics were generated from a well-known non-rooted android network monitoring application, G-Net Track Pro. Python programming language was used in the coding and analysis in determining the best features of the production dataset for improving the energy efficiency of a 5G network. The steps we took in the whole process include first, the 35 different

Submitted on February 06, 2023.
Published on April 02, 2023.
C. R. Okpara, Telecommunication Engineering, Federal University of Technology, Owerri, Imo State, Nigeria.
(e-mail: chinedu.okpara@futo.edu.ng)
V. E. Idigo, Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
(e-mail: ve.idigo@unizik.edu.ng)
C. S. Okafor, Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
(e-mail: nechuko@gmail.com)

datasets of [8] were merged as one and cleaned. It was then stored, and comparatively analyzed using significance indicator, correlation matrix and density heat maps. All data were converted to numeric data. The relationship between the features were explored and normalized using the log function in numpy library. Lastly, the best features for optimal performance for improved energy efficiency were determined.

IV. Results

From the results we obtained; using Pearson correlation coefficient to investigate how correlated the features are, it was found that the highest correlation exists between the reference signal power and the received signal reference power of the neighbouring cells (NRxRSRP) with a correlation value of 0.78 as shown in Fig. 1.

In determining the key important features of the production dataset of [8] for improving the energy efficiency of a 5G network, the signal to noise ratio was found to be the most important, given the significance indicator as shown in Fig. 2.

We used more visual plots in determining the features of the production dataset that needed to be normalized, since machine learning works better with normalized data. Fig. 3(a-j) shows the results in numerical columns with respect to determining the data transformation and normalization approach to take for each feature. From the results obtained, we observed from Fig 3a that the speed data distribution is skewed heavily to the left, thus needs normalization.

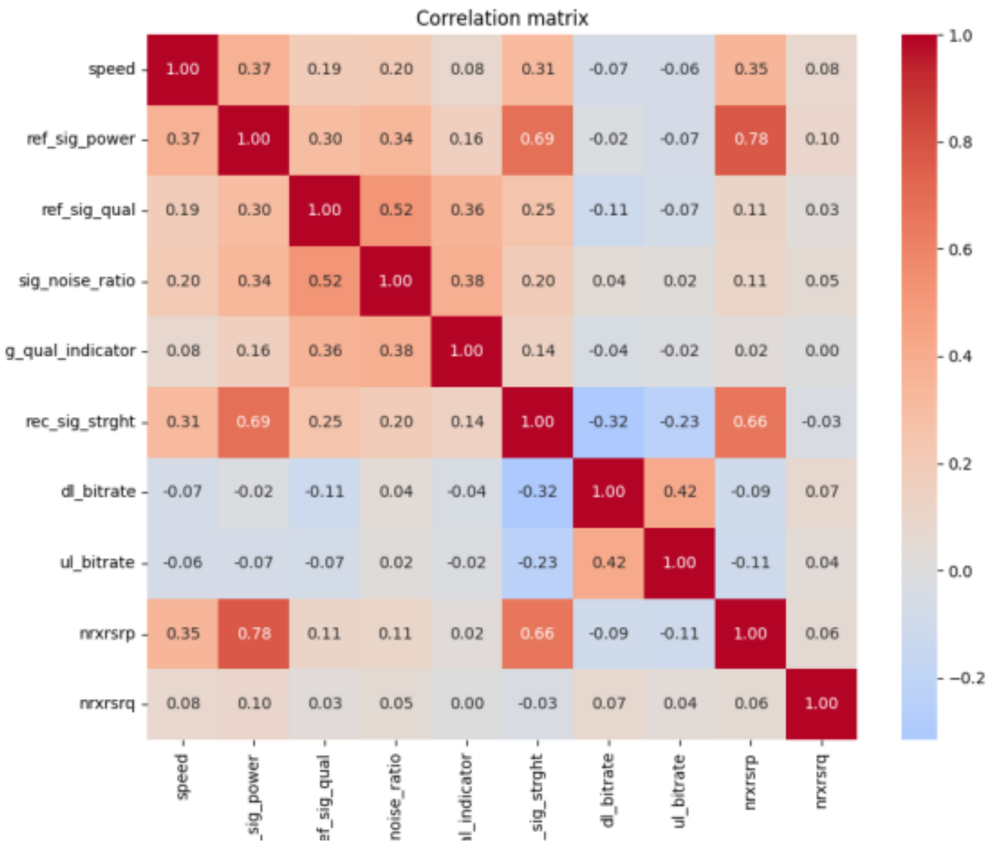


Fig. 1. Correlation investigation results of the production dataset features.

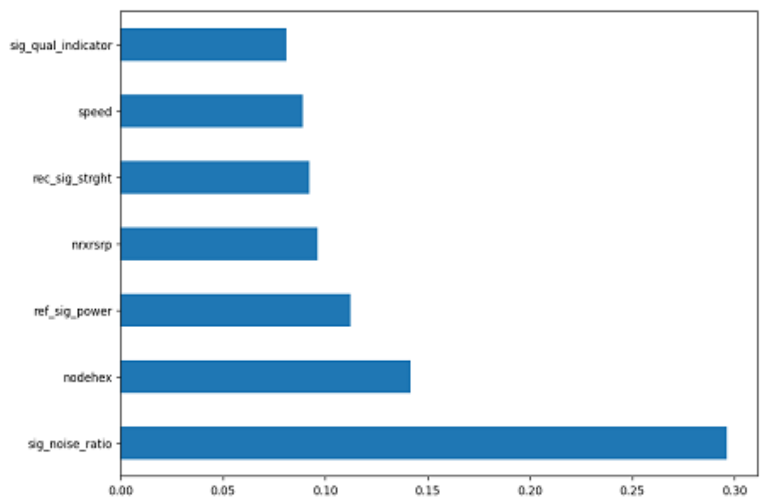


Fig. 2. Key features of the production dataset for improved energy efficiency.

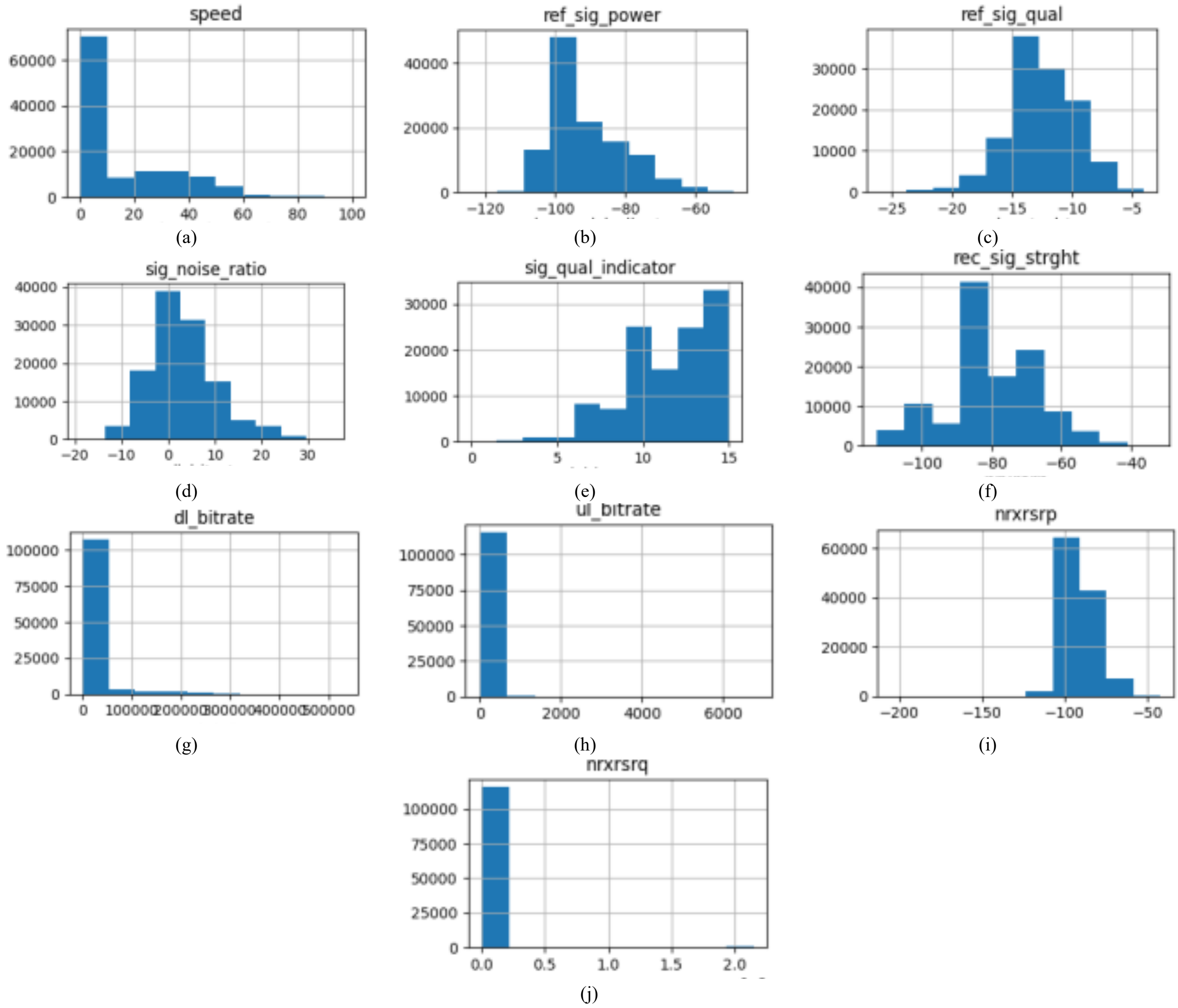


Fig. 3. Column distribution for data normalization approach; a) Speed data distribution chart; b) Reference signal power data distribution chart; c) Reference signal quality data distribution chart; d) Signal to noise ratio data distribution chart; e) Signal quality indicator data distribution chart; f) Received signal strength data distribution chart; g) Download bit-rate data distribution chart; h) Upload bit-rate data distribution chart; i) Received signal reference power of the neighbouring cells data distribution chart; j) Received signal reference quality of the neighbouring cells data distribution chart.

Fig. 3b shows that reference signal power data distribution is almost a Gaussian distribution, needing little or no normalization. Fig. 3c shows that the reference signal quality data distribution is a Gaussian distribution. Fig. 3d shows that the signal to noise ratio data distribution is a Gaussian distribution. Fig. 3e shows that the signal quality indicator data distribution is almost a Gaussian distribution, needing a little normalization. Fig. 3f shows that the received signal strength data distribution is a Gaussian distribution. Fig. 3g shows that the download bitrate data distribution needs further investigation of its values, since its values are centered between 0 and 1. Fig. 3h shows that the up-link bitrate data distribution needs further investigation, due to its values ranging from 0 to 1,000. Fig. 3i shows that the nRxRSRP data distribution is slightly skewed to the right, needing normalization. Fig. 3j shows that the nRxRSRQ data distribution needs further investigation since its values centered between 0 and 0.3.

Further, the density heat map which uses colour gradient to show how concentrated a given data is, at a particular area, and the scatter plot which shows how the data is distributed were used to show the relationships between the features.

The density heat map and scatter plot of Fig. 4 shows that the linear relationship between the received signal strength and the reference signal quality is good.

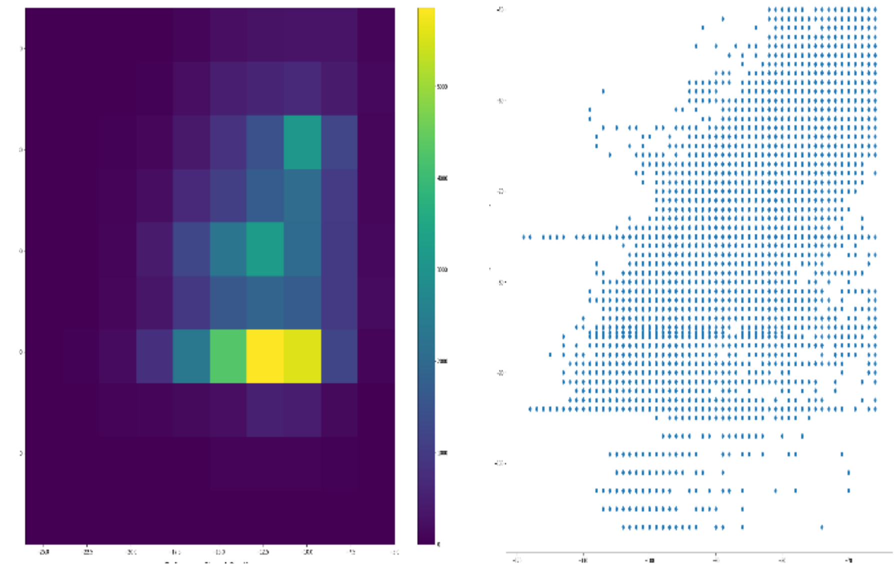
The density heat map and scatter plot of Fig. 5 shows that the linear relationship between the reference signal quality and the reference signal power is good.

The density heat map and scatter plot of Fig. 6 shows that very good relationship exists between the reference signal quality and the signal to noise ratio.

The density heat map and scatter plot of Fig. 7 shows that good relationship exists between the download bitrate and the upload bitrate of the data capture.

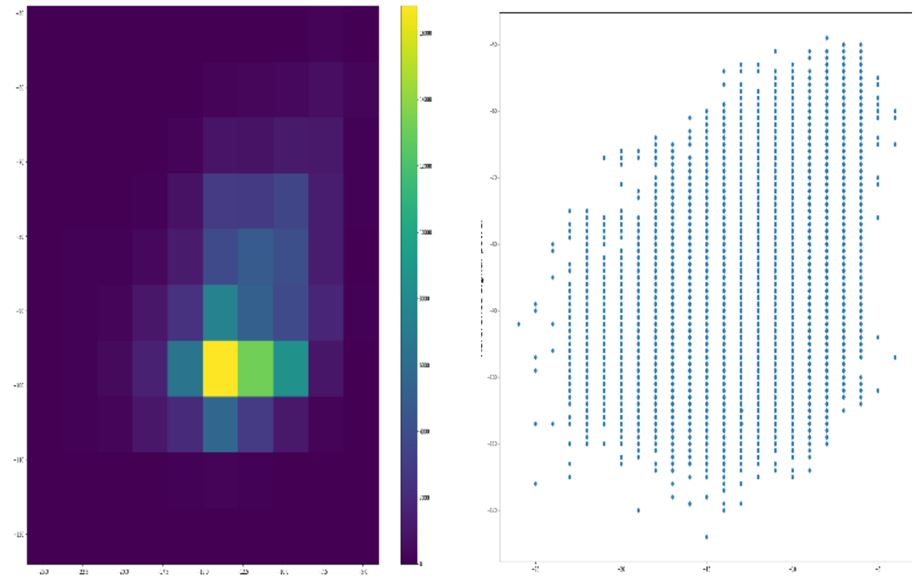
The density heat map and scatter plot of Fig. 8 shows that there is good relationship between the reference signal received power of the neighbouring cells (nRxRSRP) and the received signal power.

The density heat map and scatter plot of Fig. 9 shows that there is great relationship between the received signal power and the signal to noise ratio.



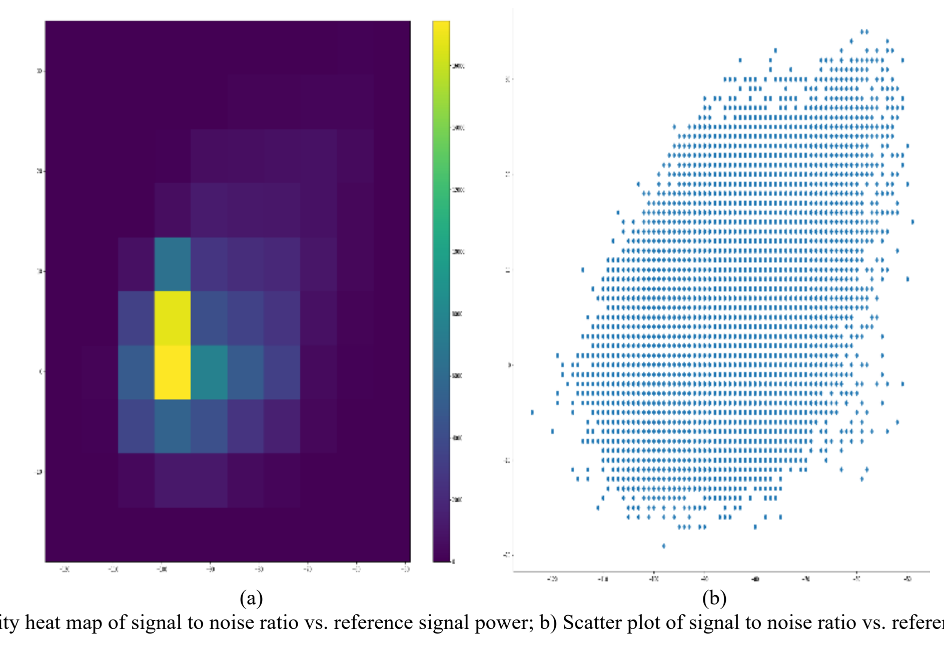
(a) (b)

Fig. 4.a) Density heat map of received signal strength vs. reference signal quality; b) Scatter plot of received signal strength vs. reference signal quality.



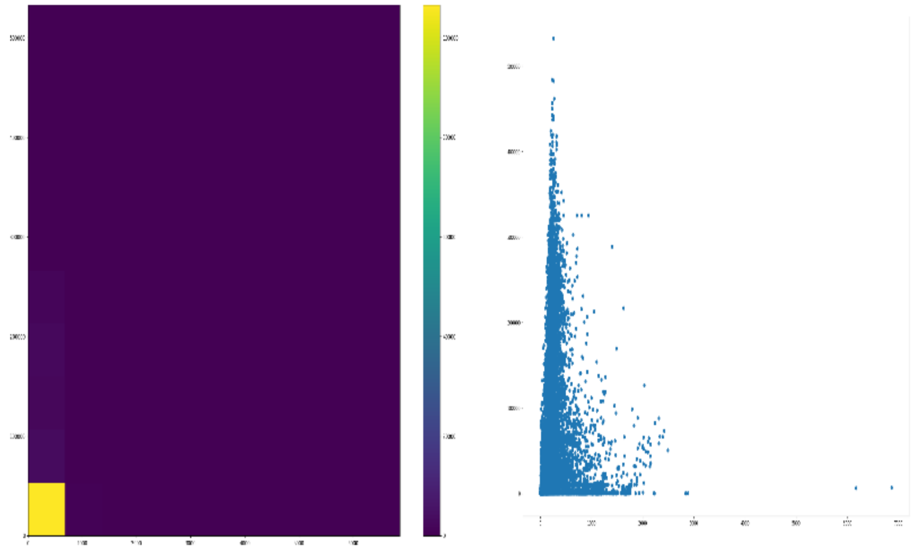
(a) (b)

Fig. 5.a) Density heat map of reference signal power vs. reference signal quality; b) Scatter plot of reference signal power vs. reference signal quality.



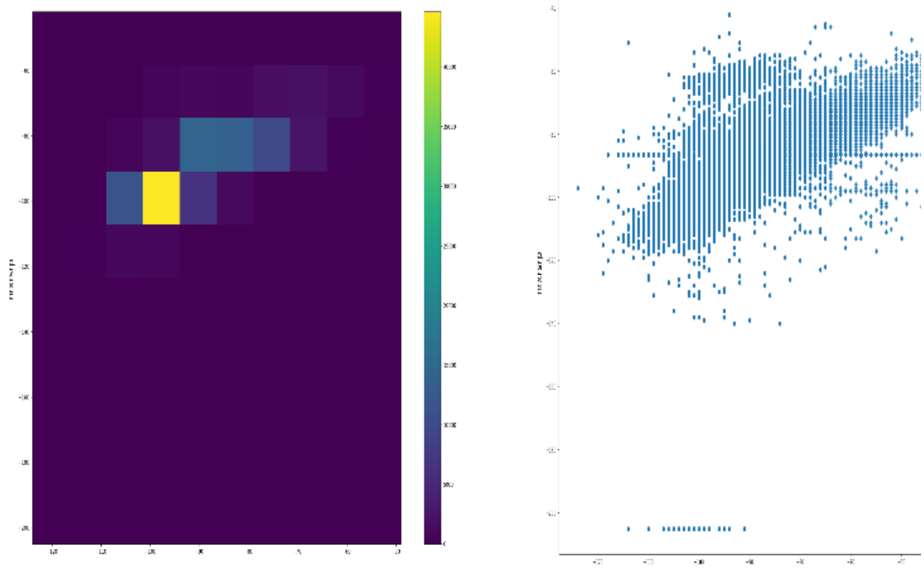
(a) (b)

Fig. 6. a) Density heat map of signal to noise ratio vs. reference signal power; b) Scatter plot of signal to noise ratio vs. reference signal power.



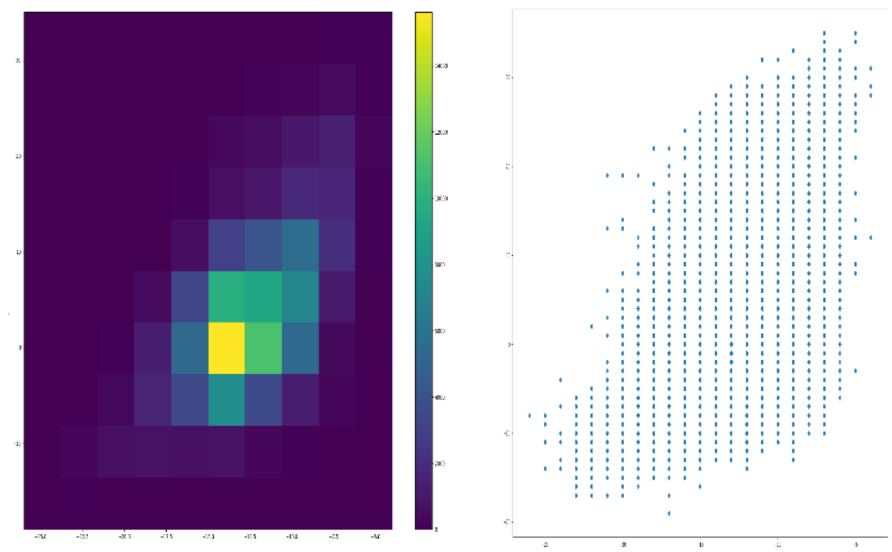
(a) (b)

Fig. 7. a) Density heat map of downlink bit-rate vs. uplink bit-rate of data capture; b) Scatter plot of downlink bit-rate vs. uplink bit-rate of data capture.



(a) (b)

Fig. 8. a) Density heat map of nRxRSRP vs. received signal power; b) Scatter plot of nRxRSRP vs. received signal power.



(a) (b)

Fig. 9. a) Density heat map of signal to noise ratio vs. reference signal quality; b) Scatter plot of signal to noise ratio vs. reference signal quality.

V. Conclusion

This research work on comparative analysis of the features of a 5G production dataset is carried out to identify which features in the dataset could make a significant impact in improving the energy efficiency of a 5G network.

From the foregoing results obtained in this work, we observed using the significance indicator that the signal to noise ratio is the most significant feature of the dataset, and we further observed that the reference signal quality, the signal to noise ratio, the reference signal power and the reference signal received power of the neighbouring cells all have key roles in improving the energy efficiency of a 5G network.

These features can also be used in developing models for determining the signal strength losses in mm-wave technology.

Acknowledgment

The authors thanks Jason Quinlan, Darijo Raca, Dylan Leahy and Cormac J. Sreenan for allowing the use of their conference paper on 'Beyond throughput, the next Generation: a 5G dataset with channel and context metrics' (MMSys '20) at Instanbul, Turkey for this research.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References

[1] Yadav R. Challenges and evolution of next generations wireless communication. *In Proc. Int. MultiConf. Eng. Comput. Scientists,* 2017; 2: 619-623.

[2] Panwar N, Sharma S, Singh AK. A survey on 5G: The next generation of mobile communication. *Phys. Commun.* 2016; 18: 64-84.

[3] Ericsson mobility report. [Internet] Retrieved from: https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf.

[4] DEEPSIG. *Howartificialintelligenceimproves5G wireless capabilities*. [Internet] Retrieve from: https://www.deepsig.ai/how-artificial-intelligence-improves-5g-wireless-capabilities.

[5] Johnson DD. *The 5G dilemma: more base stations, more antennas-less energy?*[Internet] Retrieve from: https://spectrum.ieee.org/energywise/telecom/wireless/will-increased-energy-consumption-be-the-achilles-heals-of-5g-networks 2018.

[6] Rajoria S, Trivedi A, Godfrey WW. A comprehensive survey: small cells meet massive MIMO. *Phys. Commun*. 2018; 26: 40-49.

[7] Haidine A, Salman FZ, Aqqal A, Dahbi A. Artificial Intelligence and Machine Learning in 5G and beyond: A Survey and Perspectives. *Moving broadband mobile communication forward intelligent technologies for 5G and beyond*. 2021. DOI: 10.5772/intechopen.98517.

[8] Raca D, Leahy D, Sreenan CJ, Quinlan JJ. Beyond throughput, the next generation: A 5G dataset with channel and context metrics. ACM Multimedia Systems Conference (MMSys), 2020, Instanbul, Turkey.

[9] Fonseca A, Kazman R, Lago P. A Manifesto for Energy-Aware Software. *IEEE Software*, 2019; 36(6): 79–82. https://doi.org/10.1109/ms.2019.2924498.

[10] Bjornson E, Kountouris M, Debbah M. Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination. *ICT 2013*, 2013. https://doi.org/10.1109/ictel.2013.6632074.

[11] Bjornson E, Sanguinetti L, Hoydis J, Debbah M. Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the answer? *IEEE Transactions on Wireless Communications*,2015; 14(6): 3059-3075. https://arXiv.org/pdf/1403.6150v2.

Chinedu R. Okpara obtained his B.Eng in Electrical/Electronic Engineering and M.Eng in Electrical/Electronic Engineering (Communication option) from the Federal University of Technology, Owerri, Imo State, Nigeria in 2008 and 2012 respectively and is presently working on his PhD dissertation in Communication Engineering at Nnamdi Azikiwe, University Awka, Anambra State, Nigeria.

He is a lecturer in Electrical/Electronic Engineering department and by extension Telecommunication Engineering at the Federal University of Technology, Owerri, Imo State, Nigeria.

Engr. Okpara is a registered member of the Nigerian Society of Engineers (NSE) and the Council for the Regulation of Engineering in Nigeria (COREN).

Victor E. Idigo is a Professor of Electronic and Computer Engineering at Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

He was the former Faculty of Engineering of the Institution.

He is a member of the following bodies: IEEE, IAENG and IACSIT.

Chukwunenye S. Okafor holds a Bachelor of Engineering (B.Eng.) Degree in Electrical/Electronic Engineering at Nnamdi Azikiwe University, Awka and a Master of Engineering (M.Eng.) degree in Electrical/Electronic Engineering (Electronic/Communication option) of the University of Benin, Benin City, and doctorate (PhD) degree in Communication Engineering of Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

He is a lecturer in Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka.

Dr. Okafor is a registered member of COREN, NSE and NIEEE.