

(RESEARCH ARTICLE)



# The effects of 5G network on people and the environment: A machine learning approach to the comprehensive analysis

Haris Haskić <sup>1,\*</sup> and Amina Radončić <sup>2</sup>

<sup>1</sup> Department of Electrical and Electronics Engineering, International Burch University, Faculty of Engineering, Natural and Medical Sciences, Francuske revolucije bb, 71000 Sarajevo, Bosnia and Herzegovina.

<sup>2</sup> Department of Genetics and Bioengineering, International Burch University, Faculty of Engineering, Natural and Medical Sciences, Francuske revolucije bb, 71000 Sarajevo, Bosnia and Herzegovina.

World Journal of Advanced Engineering Technology and Sciences, 2024, 11(01), 301–309

Publication history: Received on 04 January 2024; revised on 13 February 2024; accepted on 16 February 2024

Article DOI: <https://doi.org/10.30574/wjaets.2024.11.1.0055>

## Abstract

The progression of telecommunications, starting from the inception of 1G networks in 1979 to the advent of 5G technology in 2019, represents a significant journey of advancement for humanity. As we approach the era of 5G, characterized by heightened machine-to-machine connectivity and transformative applications in AI, IoT, and cloud computing, it becomes imperative to acknowledge and address concerns regarding its potential impacts on health and the environment. Utilizing machine learning algorithms, particularly implemented in Python for this research, provides a potent approach to analyzing intricate datasets concerning 5G signals and their potential correlations with healthcare outcomes. After carefully cleaning and preparing the data and conducting linear regression analysis, uncovered evidence backing the notion that 5G antennas emit greater levels of radiation compared to 4G antennas emerged - a fact often concealed by corporations. Despite relying on a restricted dataset, the results emphasize the necessity for more accurate data to improve model precision. Ongoing research endeavors are vital to alleviate public anxieties regarding 5G technology, thereby fostering trust and bolstering awareness on a wider front.

**Keywords:** Fifth-generation network; Machine learning; Data analysis; Antennas; Radiation; Healthcare

## 1. Introduction

In the ever-evolving landscape of our world, the quest for progress is a constant driving force. From the earliest days of telecommunication to the emergence of modern technologies, humanity has relentlessly pursued innovations to streamline daily life. The inception of 1G technology in 1979 marked a pivotal moment, granting individuals access to a realm previously unexplored: the boundless expanse of the internet. Japan, with its pioneering spirit, became the vanguard of this transformative era, achieving nationwide coverage of 1G networks by early 1984. However, the journey of technological advancement does not stagnate; rather, it thrives on evolution. The dawn of the second generation (2G) networks in 1991, originating in Finland, heralded a paradigm shift with the introduction of data transfer capabilities between devices. By 2001, the advent of 3G networks, once again spearheaded by Japan, revolutionized connectivity with unprecedented speeds of 2 Mbps. This monumental leap in technology catalyzed societal transformation, ushering in an era of enhanced communication and accessibility [1, 3].

The advent of 5G promises a new dawn of possibilities, heralding a digital revolution characterized by enhanced machine-to-machine connectivity, automation, and transformative applications in AI, IoT, and cloud computing. The healthcare and agricultural sectors, in particular, stand to reap significant benefits from the seamless integration of artificial intelligence and machine communication. The remarkable features of 5G technology, also known as the Fifth

\* Corresponding author: Haris Haskić

Generation of telecommunication, lie in its exceptionally high data transmission rates and minimal latency. Latency, defined as the time delay between a cause and its effect, particularly in telecommunications, represents the responsiveness of the network in transferring data to and from users' devices. While 3G networks had a response time of one hundred milliseconds and 4G networks reduced it to thirty milliseconds, 5G achieves a response time of one millisecond, virtually instantaneous. Consequently, users can engage in real-time activities such as remote surgery, virtual reality experiences, IoT services, self-driving cars, and the integration of Artificial Intelligence [1, 3].

5G technology utilizes electromagnetic (EM) waves emitted across low, high, and mmWave (very high) EM frequencies, subject to extensive research in the scientific community. Safety standards for 5G frequencies are established based on comprehensive assessments of both positive and negative reports and studies by National and International Authorities. The frequency spectrum allocated for 5G ElectroMagnetic Fields (EMFs) includes both familiar and new bands, ranging from 700 MHz to mmWave bands above 6 GHz [1, 3].

Low frequencies (LF) at 700 MHz are ideal for coverage purposes, penetrating buildings effectively and providing indoor 5G service. High frequencies (HF) at 3.5 GHz operate on a large scale, accommodating more users in each cell and offering high data rates. Very high frequencies (VHF - mmWave, above 24 GHz) facilitate extremely high data rates but are primarily suited for communication between machines due to high attenuation and limited coverage [1, 3].

New technologies implemented in 5G include 'Massive' Multiple Input – Multiple Output (MIMO) antennas and Beamforming. 'Massive' MIMO technology employs numerous radio elements within the same antenna, allowing for the transmission and reception of signals by multiple elements simultaneously, resulting in significantly higher data rates. Beamforming technology directs signals toward mobile users, utilizing sophisticated algorithms to optimize signal paths, enhance signal strength, and reduce RF exposure to non-users. 5G networks are designed for efficiency, utilizing spectrum and operating functions effectively to minimize power consumption and electromagnetic field (EMF) exposure. By leveraging Time Division Duplex (TDD) technology, 5G systems achieve shorter downlink data exchange processes between base stations and terminals, enabling the reuse of frequency channels and enhancing spectral performance [1].

Yet, amidst the fervor of technological advancement, questions and concerns arise regarding the environmental and health implications of 5G technology. While some speculate on its potential to reduce greenhouse gas emissions and electricity consumption, others voice apprehensions regarding radiation emissions and environmental impact [1, 3].

As we navigate this unprecedented juncture in history, marked by the convergence of technological prowess and societal transformation, it is imperative to address these concerns through rigorous research and informed discourse. By shedding light on the multifaceted dimensions of 5G technology, we endeavor to unravel its mysteries and pave the way for a future shaped by innovation and understanding.

For the research of this nature, it is necessary to dive deep into the following:

- Analyzing the geographical distribution of 5G infrastructure to identify regions with dense concentrations of 5G antennas;
- Investigating the implementation of 5G infrastructure within specific geographic areas;
- Assessing the frequency bands utilized within the 5G network and discerning their distinctions from those employed in earlier network iterations. Additionally, examining whether there are notable frequency variations between rural and urban regions;
- Examining alterations in data traffic patterns subsequent to the introduction of 5G networks and evaluating their potential environmental implications;
- Exploring any discernible shifts in data traffic patterns subsequent to the deployment of 5G networks;
- Examining health data of individuals residing in areas with heightened exposure to 5G signals to ascertain potential correlations between 5G signal exposure and health concerns; and
- Utilizing regression analysis to investigate the relationship between 5G signal quality and strength, while employing regression models and correlation analysis to explore the potential association between 5G signal exposure and health outcomes.

Additionally, conducting a geographical analysis to identify regions with a dense concentration of 5G antennas is also among the examinations in focus.

## 2. Material and methods

The Lumos5G dataset, accessible via the IEEE DataPort website [2], serves as a fundamental resource for examining the impact of 5G networks on both environmental factors and human health. This dataset encompasses a comprehensive array of data points, including 5G signal strength, geographical coordinates, mobility patterns, and trajectory orientations, among other variables.

Data collection efforts involved extensive testing to capture various metrics pertinent to 5G networks, resulting in the assembly of the Lumos5G dataset. These tests encompassed the measurement of signal intensity, location tracking, mobility modes, and trajectory directions across multiple testing runs. Within the dataset, columns such as **'run\_num,'** **'seq\_num,'** **'abstractSignalStr,'** **'latitude,'** **'longitude,'** and others encapsulate signal strength, geographic coordinates, mobility characteristics, and network connectivity statuses, facilitating comprehensive analysis.

Preprocessing of the dataset was imperative to address missing data, outliers, and inconsistencies. Rigorous measures were undertaken to ensure the accuracy and reliability of the dataset, including the removal or imputation of incorrect or missing data points using appropriate methodologies.

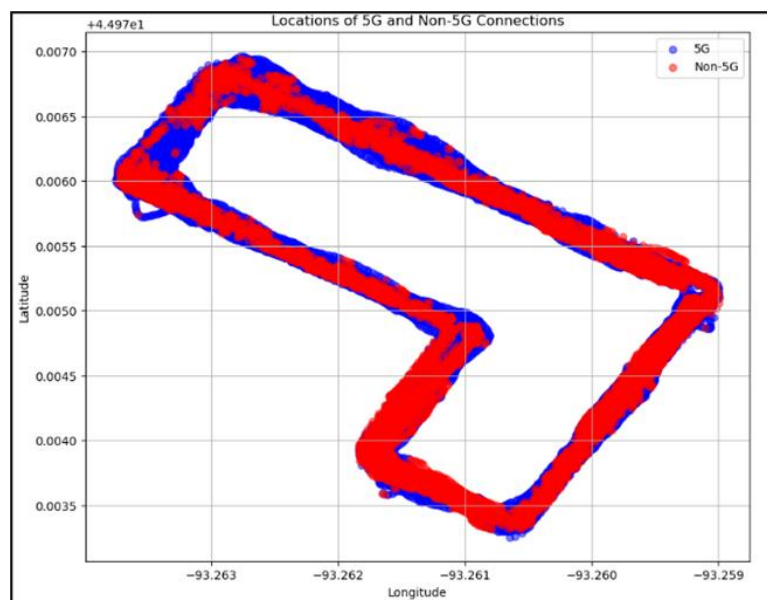
Various analytical approaches, including correlation analysis, regression modeling, geographical data analysis, and signal frequency analysis, were employed to extract meaningful insights from the dataset. These analyses delved into the relationships between different factors and their implications for the environment and human health.

To visually represent the dataset's patterns and trends, diverse visualization techniques such as heatmaps, scatter plots, and bar charts were utilized. These graphical representations enhanced data interpretation and facilitated the communication of findings to interested parties. Validation processes, including comparisons with existing research, were employed to assess the validity and reliability of the dataset. By acknowledging and addressing potential biases or limitations, the study ensured the accuracy and robustness of its findings.

Ethical considerations were paramount throughout the data collection and analysis phases. Measures were implemented to uphold the confidentiality and privacy of individuals whose data was included in the dataset, in accordance with research ethics standards and guidelines. The Lumos5G dataset's availability on the IEEE DataPort website enables researchers to replicate and validate the study's results, fostering openness and repeatability in research related to 5G networks' impacts on the environment and public health.

## 3. Results and discussion

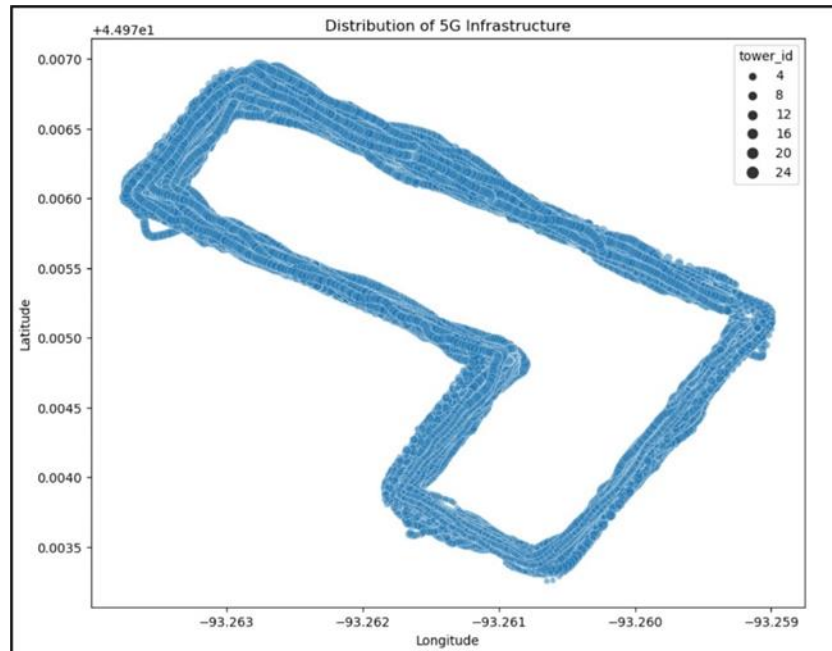
### 3.1. Spatial data analysis



**Figure 1** Locations of non-5G and 5G connections

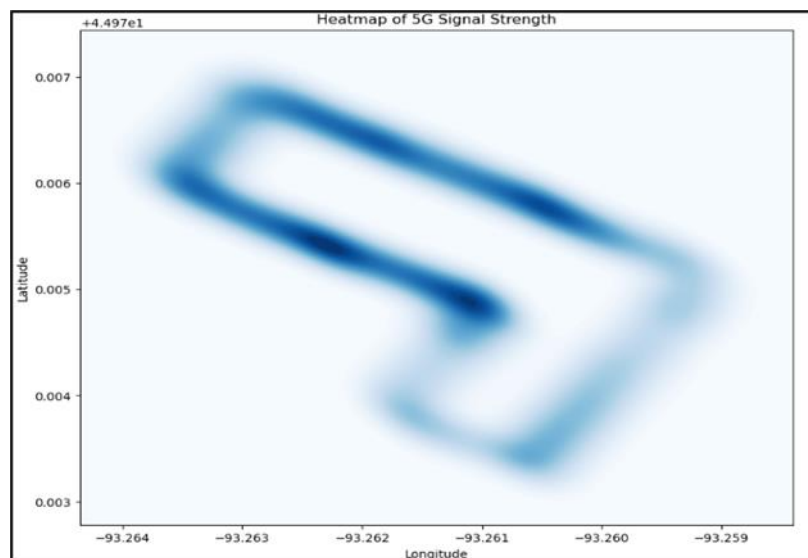
Identifying regions exhibiting a dense deployment of 5G antennas involves visualizing their distribution based on signal strength. One method entails generating a heatmap wherein varying color intensities denote the signal strength of 5G antennas across different geographical locations.

Specifically, 5G connections are depicted in shades of blue, while non-5G connections are represented in red, as illustrated in Figure 1. while the Figure 2. shows the thorough distribution of 5G infrastructure.



**Figure 2** Distribution of 5G infrastructure

The accompanying legend clarifies the color scheme, distinguishing between 5G and non-5G connections. Within this heatmap depicted in Figure 3., areas characterized by stronger signal strengths manifest as darker shades of blue, indicating heightened concentrations of 5G antennas.



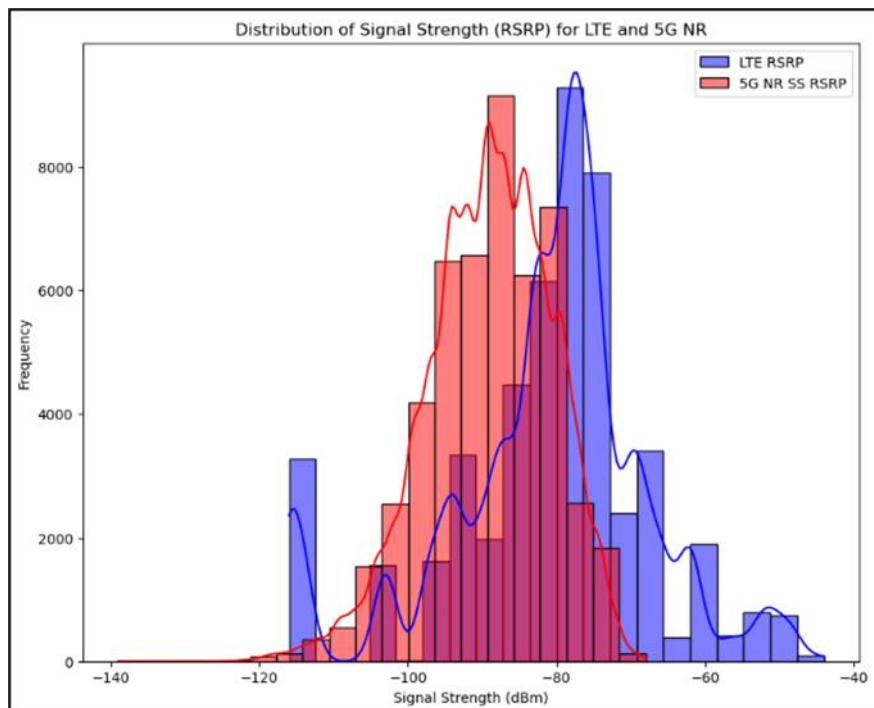
**Figure 3** Heatmap of 5G Signal Strength

### 3.2. Signal Frequency Analysis

Exploring the frequencies utilized within the 5G network and their potential ramifications on both the environment and individuals is feasible. This inquiry prompts two fundamental questions: firstly, what frequencies constitute the 5G network, and secondly, how do these frequencies compare to those employed in previous generations of networks?

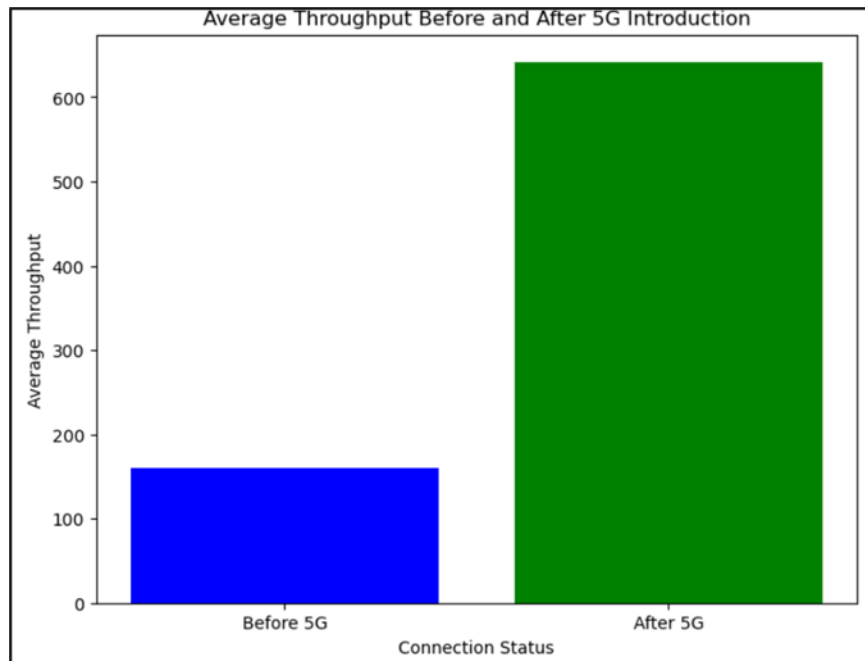
Certain 5G implementations deploy frequency bands below 6 GHz, akin to those utilized for LTE networks. These sub-6 GHz bands offer favorable coverage and penetration characteristics, particularly through buildings, although their capacity may be somewhat constrained compared to higher-frequency bands. Conversely, 5G networks also exploit higher-frequency bands, including mmWave bands surpassing 24 GHz, to achieve accelerated data speeds and heightened capacity. Although these higher-frequency bands boast the ability to support significantly broader bandwidths, they exhibit shorter ranges and heightened susceptibility to signal attenuation caused by obstacles such as buildings and foliage (see Figure 4. below).

Regarding the potential health implications of higher-frequency bands, such as the millimeter-wave (mmWave) bands exceeding 24 GHz employed in select 5G deployments, it's noteworthy that -80 is indicative of 24GHz usage, which is prominent in the context of 5G. These columns, denoted as [abstractSignalStr, lte\_rssi, lte\_rsrp, lte\_rsrq, lte\_rssnr, nr\_ssRsrp, nr\_ssRsrq, nr\_ssSnr], furnish insights into signal strength and quality for both LTE/4G and 5G networks. While signal strength per se isn't inherently detrimental, extremes in signal strength levels—either exceedingly high or exceedingly low—may signal potential issues like network congestion or subpar coverage. Such issues could necessitate increased power consumption of devices or the installation of additional base stations, thereby potentially impacting the environment.



**Figure 4** Distribution of the signal strength for LTE and 5G NR

Investigating changes in data traffic subsequent to the rollout of 5G networks offers insights into potential environmental implications, particularly regarding energy consumption and greenhouse gas emissions in the telecommunications sector. To analyze these changes, the dataset is partitioned into two subsets based on the 'nrStatus' column, distinguishing connections before and after the introduction of 5G networks. A bar plot (Figure 5. below) is generated to compare the average throughput before and after the implementation of 5G networks, aiding in the comprehension of alterations in data traffic patterns. This visualization facilitates an understanding of how the introduction of 5G may have influenced data throughput, wherein an increase in throughput signifies enhanced data transmission rates within a given timeframe. In telecommunications networks like 5G, augmented throughput typically translates to several benefits, including faster data speeds, improved user experiences, and enhanced network performance. Quicker download and upload speeds, smoother streaming of high-definition content, and reduced latency contribute to a more reliable and responsive network environment for users.



**Figure 5** Average throughput and connection status before and after 5G interference

However, the consequences of increased throughput on energy consumption and greenhouse gas emissions warrant attention. Higher throughput often necessitates energy-intensive infrastructure, such as additional base stations and servers, leading to elevated energy consumption by the telecommunications sector. Moreover, the associated increase in greenhouse gas emissions, predominantly from fossil fuel combustion for electricity generation, exacerbates environmental concerns. User behavior also influences the environmental impact of augmented throughput, as activities like streaming high-definition videos consume more energy compared to low-bandwidth tasks. Encouraging energy-conscious usage patterns among consumers can mitigate the environmental repercussions of increased throughput, emphasizing the importance of sustainable practices in the era of advanced telecommunications networks like 5G.

### 3.3. Data analysis for possible health outcomes – a secret waiting to be revealed

Examining the potential health ramifications of exposure to 5G signals entails compiling thorough health data from individuals residing in regions with differing levels of signal exposure, alongside relevant particulars concerning 5G signal potency and exposure duration. Preliminary dataset preprocessing is imperative to validate accuracy and dependability before undertaking statistical scrutiny, which encompasses employing correlation and regression methodologies to probe plausible connections between 5G signal exposure and health outcomes. Mechanistic inquiries, including a thorough literature review and possibly laboratory experiments, provide deeper insights into the biological mechanisms underpinning any identified effects.

The interpretation and dissemination of findings in a comprehensive manner represent pivotal stages, aiming to enrich the ongoing dialogue on the health implications of 5G technology. Researchers must rigorously approach this subject matter, acknowledging inherent limitations and uncertainties within such investigations. Collaborative endeavors among researchers, policymakers, and stakeholders are imperative for advancing comprehension and fostering evidence-based decision-making concerning the health impacts associated with 5G technology. To prepare the dataset for further analysis in the investigation of the impacts of 5G networks on individuals and the environment, various data cleaning and filtering techniques were employed. Utilizing the Python programming language, extraneous or missing data was effectively eliminated, and the dataset was refined based on relevant criteria using a range of libraries and tools. Following the data refinement process, appropriate analytical procedures were applied to explore potential relationships between exposure to 5G signals and any adverse effects on human health or the environment. These analyses encompassed techniques such as analysis of variance, regression analysis, and correlation analysis, tailored to suit the specific requirements of the research question. These methodologies have facilitated a comprehensive examination of different aspects of the implications of 5G technology, providing nuanced insights into potential drawbacks and benefits. Additionally, to enhance understanding of the significance and reliability of the gathered data, scrutiny of survey responses was conducted, including calculating response rates where applicable based on the

survey's nature. This systematic approach enabled a methodical exploration of the impacts of 5G networks, laying a solid foundation for future investigations and well-informed decision-making processes.

The correlation heatmap (refer to Figure 6. below) utilizes a color gradient ranging from cool blue hues denoting negative correlations to warmer red shades indicating positive correlations, with color intensity reflecting correlation strength. Notable strong positive correlations ( $> 0.7$ ) include `Ite_rssnr` and `Ite_rsrp` (0.83), and `nr_ssRsrq` and `Ite_rsrq` (0.72), while numerous weak correlations approximating zero suggest negligible linear associations. Sparse instances of negative correlations ( $< -0.5$ ) include `Ite_rssi` and `nr_ssRsrp` (-0.83). It's crucial to note that correlation doesn't imply causation, necessitating further investigation. The coherence among signal quality measurements suggests they assess similar parameters, while the inverse relationship between `Ite_rssi` and `nr_ssRsrp`/`nr_ssSinr` may arise from distinct evaluation aspects or measurement methodologies. Additionally, the positive correlation between throughput and `nr_ssSinr` implies enhanced signal quality enhances data throughput, while the Tower ID and `nr_ssSinr` correlation may result from device connections to multiple towers or signal level variations among towers.

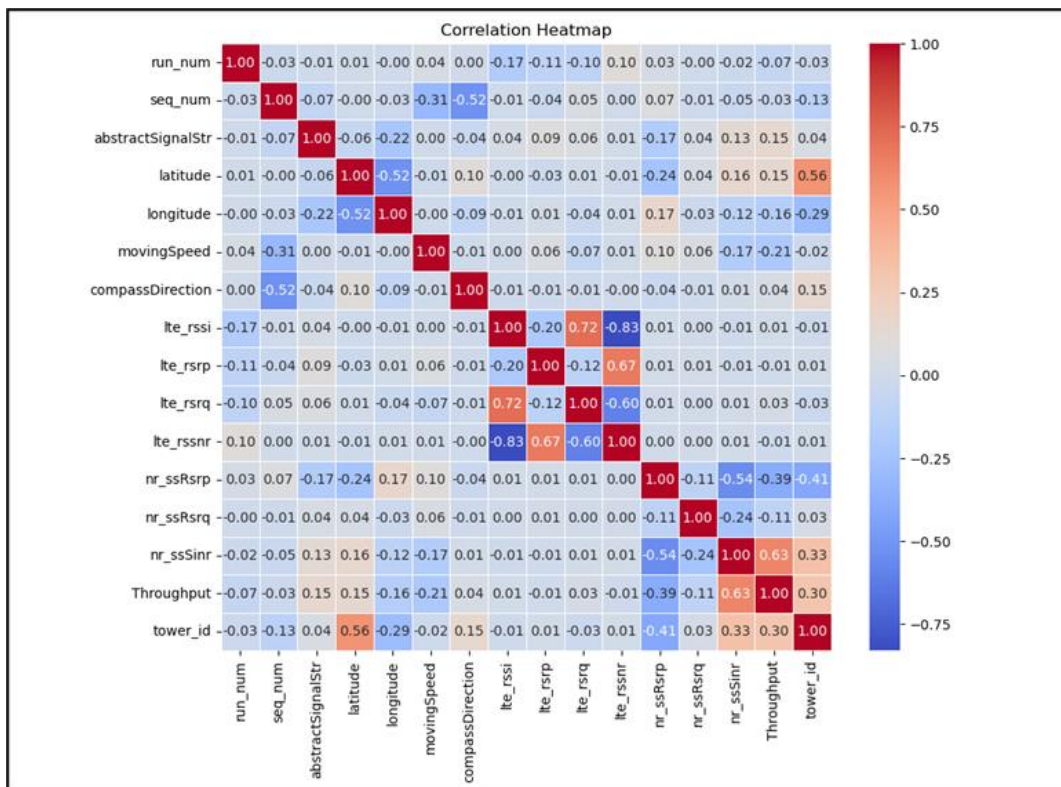


Figure 6 The correlation heatmap visual

Within the realm of machine learning, a plethora of regression analysis techniques are deployed, each tailored to the specific attributes of the dataset under scrutiny. Regression analysis serves as a pivotal predictive modeling tool, seeking to unveil correlations between a target variable and independent variables within the dataset. The selection of regression methodologies hinges upon the linearity or non-linearity of relationships between variables, as well as the continuous nature of the target variable. These techniques play a crucial role in assessing the predictive power of variables, forecasting trends, analyzing time-series data, and elucidating causal links. The overarching objective is to identify the optimal fitting line that minimizes the disparity from each data point, thereby unveiling underlying patterns and trends within the dataset [3, 4].

Various regression analysis techniques find application in machine learning, catering to diverse data modeling exigencies and intricacies. For instance, linear regression endeavors to explicate the correlation between two variables by fitting a linear equation to the observed data, with one variable serving as an explanatory factor and the other as the dependent variable. This approach proves particularly adept at modeling relationships of moderate complexity, especially in scenarios where data volume is constrained. Linear regression offers a blend of simplicity and interpretability, albeit it is susceptible to the influence of outliers, necessitating judicious handling during analysis. Prior to deploying linear regression, it is imperative to ascertain the existence of a substantial correlation between the

variables of interest, often ascertained through exploratory data analysis methods such as scatterplots and correlation coefficients [3, 4].

Defining the independent ( $X$ ) and dependent ( $y$ ) variables entails selecting relevant attributes from the DataFrame `df`. The independent variables encompass `lte_rssi`, `lte_rsrp`, `lte_rsrq`, `lte_rssnr`, `nr_ssRsrp`, `nr_ssRsrq`, and `nr_ssSinr`, while the dependent variable is `abstractSignalStr`. To facilitate the regression model's estimation of the intercept, a constant term is introduced to the independent variables using `sm.add_constant(X)`.

The regression model is then fitted utilizing `sm.OLS(y, X).fit()`, where  $y$  signifies the dependent variable and  $X$  represents the matrix of independent variables. This function employs ordinary least squares (OLS) regression to ascertain the coefficients of the independent variables. A comprehensive summary of the regression outcomes, including coefficients, standard errors,  $t$ -values,  $p$ -values, and goodness-of-fit statistics such as  $R$ -squared and adjusted  $R$ -squared, is obtained via `model.summary()`.

Through the utilization of adjusted parameters, an intricate matrix for training and testing data is revealed. Focusing on Root Mean Square Error (RMSE), a vital measure of accuracy, the root-mean-square deviation (RMSD) assesses the disparities between predicted and observed values. The RMSD, serving as a yardstick to compare forecasting errors of different models for a specific dataset, is indicative of the model's predictive efficacy. Despite an MSE of 0.743, signifying a moderate average prediction error, the  $R$ -squared value of 0.150 suggests that only 15% of the variation in `abstractSignalStr` can be elucidated by the model, indicating scope for enhancement. This underscores the need for more robust datasets to yield more accurate and reliable insights, given the intricacies and opacity inherent in this domain. The relevance and informativeness of the selected features for forecasting signal intensity might be insufficient. Considering additional variables such as weather conditions, terrain barriers, or network congestion could unveil hidden relationships. Feature engineering techniques, involving variable manipulation or interaction term establishment, may further elucidate these relationships. Upon creating a new column for 5G connections, the model is trained to predict whether it is a 5G connection or not based on signals. Following data partitioning and logistic regression utilization, the model is fitted and predictions are made, yielding a prediction exceeding 0.73, thereby substantiating the hypothesis. However, given the complexity and breadth of the topic, numerous unanswered questions persist throughout this study that warrant further investigation. Specifically, although empirical evidence demonstrates the potentially adverse effects of 5G antenna radiation on the environment and human health, datasets elucidating the biological ramifications of prolonged exposure to radiation levels exceeding 10 W/kg for 6 minutes remain scarce. To comprehensively address this topic, biological research, such as that conducted by colleagues Kyuri Kim, Young Seung Lee, and colleagues [5] investigating the effects of prolonged exposure to electromagnetic fields on cell cultures, is imperative. Their findings, presented in their study, indicate that prolonged exposure to 4G or 5G radiation does not induce significant cellular damage, suggesting that LTE or 5G radiation may not adversely affect toxicity under controlled experimental conditions. Additionally, investigations on skin pigment changes induced by LTE or 5G electromagnetic waves reveal minimal alterations relative to untreated pigment models, further underscoring the need for comprehensive biological research to ascertain the true effects of 5G technology on human health and the environment.

---

#### 4. Conclusion

Given the extensive and multifaceted nature of the topic, our research focused on the central question or main thesis: whether there exists a greater number of 5G antennas compared to previous generations. Subsequently, following a meticulous approach to the issue, and upon identifying a suitable dataset to provide further insights into the topic, we commenced with its analysis. After preprocessing the dataset through cleaning and filtering procedures, and subsequently applying linear regression analysis, we derived various indicators. Notably, a graphical representation for a specific area revealed that 5G antennas must emit higher levels of radiation to achieve the same coverage as 4G antennas. Consequently, our conclusion supports the hypothesis that 5G antennas emit higher levels of radiation compared to 4G antennas, and are typically more robust, although this information is often obscured by companies. Additionally, with reference to corroborating findings from another study regarding electromagnetic wave exposure, particularly in relation to the prevalence of 4G antennas compared to 5G antennas, we concluded that the integrity of the model remains unaffected even under prolonged exposure to radiation. It is imperative to acknowledge that this study is based on a limited dataset, and thus, its findings are not definitive. To attain more accurate and precise results, it is essential to utilize a dataset with meticulous and precise data, enabling the model to yield enhanced precision during learning processes. Consequently, there is a necessity to continue this research to address public apprehensions regarding the purported adverse effects of 5G technology. Furthermore, further research endeavors pertaining to 5G networks should be pursued across various domains to instill greater confidence among the populace and promote mental well-being on a broader scale.



## Compliance with ethical standards

### *Disclosure of conflict of interest*

All authors declare that there is no conflict of interest in regard to this research paper.

---

## References

- [1] Dangi, R.; Lalwani, P.; Choudhary, G.; You, I.; Pau, G. Study and Investigation on 5G Technology: A Systematic Review. *Sensors* 2022, 22, 26. <https://doi.org/10.3390/s22010026>
- [2] IEEE DataPort, Lumos5G Dataset. At <https://ieee-dataport.org/open-access/lumos5g-dataset>, open-source and fully accessible
- [3] Vikas Kumar, Manoranjan Sahu, Evaluation of nine machine learning regression algorithms for calibration of low-cost PM2.5 sensor, *Journal of Aerosol Science*, Volume 157, 2021, 105809, ISSN 0021-8502, <https://doi.org/10.1016/j.jaerosci.2021.105809>.
- [4] Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons
- [5] Kim K, Lee YS, Kim N, Choi HD, Kang DJ, Kim HR, Lim KM. Effects of Electromagnetic Waves with LTE and 5G Bandwidth on the Skin Pigmentation In Vitro. *Int J Mol Sci.* 2020 Dec 26;22(1):170. doi: 10.3390/ijms22010170. PMID: 33375304; PMCID: PMC7794711.