ACHIEVING SUSTAINABLE DEVELOPMENT GOALS THROUGH BOOSTING MOBILE CONNECTIVITY USING MACHINE LEARNING AND BIG DATA PROVIDED BY NATIONAL TELECOM VOLUNTEERS

Himanshu, Sharma[1]; Atul, Joshi[2]

[1,2]Ministry of Communications, Government of India

ABSTRACT

After the 5G launch in October 2022, its use cases in India have advanced significantly. The advent of 5G technology holds significant promise for advancing Sustainable Development Goals (SDGs) in India. The *potential impacts of 5G deployment on various aspects of sustainable development include focusing on economic growth, social inclusion, environmental sustainability, and governance. Hence, it is imperative to measure the 5G coverage evenly across the country to make sure that no one is left behind to benefit from development i.e. ensuring equitable justice. However, there is currently no method or tool for precise real-time monitoring of telecom technology (2G/3G/4G/5G) and Quality of Service (QoS) for individuals on the ground, resulting in a large portion of the population without reliable telecom connectivity.*

This paper proposes a methodology and tools to measure ground-level QoS, which the Indian Government can use to take proactive steps to provide high-quality telecom coverage to all citizens, focus on strategic areas, and achieve the goals of the National Digital Communication Policy-2018. The created smartphone app collects data from volunteers using scalable server architecture. Along with signal data, topographical, meteorological, and mobile tower data will be used. Machine Learning will predict coverage of places for which data is missing. The Google map hotspot will monitor telecom coverage, and the dead/grey zone will be improved to increase telecom coverage, and quality of life, and achieve SDGs.

Keywords – telecommunication, quality of service, deep learning, received signal strength indicator, Sustainable Development.

1. INTRODUCTION

As per the GSMA report, the Absence of Quality Telecom coverage hinders the performance in achieving targets of Sustainable Development Goals (SDG)[1]. Further, TRAI data shows that rural teledensity in India is 58.24%, vs 133.7% for urban, showing a lack of equitable justice and equal opportunity [2].

The current study aims to address the issue of the unavailability of a proprietary tool with the Govt. of India for measuring telecom coverage at the ground level. In summary, the research will develop a mobile app for volunteer citizens to install. These volunteers will periodically report their technology (2G/3G/4G/5G), location, mobile phone signal strength, and other telecom parameters. This real-time and historical data will be incorporated into a machine-learning (ML) model to forecast signal quality in missing places where no data was obtained. The solutions which initially tested for the patch of Geographical area can be scaled to cover the entire country and will give the Telecom coverage scenario of the Nation. This solution which is scalable and economical will replace existing solutions like drive test or MDT (Minimization of Drive Test) [3]. Recent research works have used deep learning models to predict mobile coverage in data sparsity conditions [4]. The performance of numerous Machine Learning algorithms has been compared, and the best-suited algorithm, closest to the desired outcome, has been recommended and used for processing data.

As per result of the prediction, the telecom operators will receive guidance regarding the necessary infrastructure or configuration adjustments, to address the gaps in telecom coverage based on the collected information.

3GPP (Release 18) is beginning to embrace ML (Machine Learning) capabilities as part of advanced network planning for future 5G deployments [5]. With the availability of geography-wise mobile network statistics, the conducive policy to cover the uncovered areas will help to get access to mobile to a large population and serve the purpose of '*Connecting the Unconnected*'. As per the GSMA report [1] Access to mobile products and services can provide an important route to prosperity for individuals and well-being. Most low and middle-income countries (LMICs) residents access the internet via it. These groups can access education, healthcare, and financial services via mobile devices. Mobile phones drive innovation and economic creation, helping achieve the SDGs, which UN Member States approved in 2015 under the 2030 Agenda for Sustainable Development.

1.1 Current scenario for measuring mobile coverage.

Existing solutions like MDT can certainly reduce the need for drive tests (for measuring telecom coverage), but there are still certain situations where MDT cannot replace drive tests [6]. Moreover, drive tests are limited in providing large-scale measurements for cities or countries [6]. As is the case with MDT, the solution proposed in the current study can assist telecom operators in coverage, mobility & capacity optimization, parametrization for common channels and QoS verification [6] and hence improving the Quality of living of individuals and ultimately improving SDG score of country.

1.2 Mobile coverage and its impact on SDGs

Poor mobile coverage leads to hindrances in improving the Standard of life of society and some of such unfortunate incidents in Indian Society are cited below:

a) Incidence published in [7] a child from Dapana village, Morni (Oct 2020) was found sitting on a tree branch to catch mobile signals to help other children complete their homework.
b) In another incident in New Delhi [8], a biker was killed in the Pragati Maidan tunnel (New Delhi) as a poor signal delayed an emergency call for medical help (May 2023).

The absence of mobile and internet connectivity directly adversely affects the targets mentioned in SDGs, e.g. Incidence in point (a) above hit target 4.2 of SDG- '*By 2030,* *ensure that all girls and boys have access to quality early childhood development, care and pre-primary education*', further incidence mentioned in point (b) mentioned above affects the target 3.6 of SDGs – '*halve the number of global deaths and injuries from road traffic accidents*.' Access to mobile connectivity is linked to the majority of 17 SDGs directly or indirectly. specifically, directly linked to SDG 1: *No Poverty*, SDG-2: *Ending Hunger* SDG-3: *Good Health and Well-being* SDG 4: *Quality Education*, SDG 5: *Gender Equality and SDG* 8: *Decent Work and Economic Growth*, SDG 9: *Industry, Innovation and Infrastructure, SDG* 10: *Reducing Inequalities*.

Take an example related to coverage of mobile and SDG, in India around half of the population depends on agriculture and as per estimates 40% of farmers in India are still taking loans from the informal banking sector and depend on moneylenders [9]. Programs like M-Pesa (Mobile Money) provide an opportunity to get rid of the debt trap of moneylenders, which may provide impetus to farmers for quick money transfers. Kenyan mobile money system M-PESA increased per capita consumption levels and lifted 194,000 households, or 2% of Kenyan households, out of poverty till 2020 [10]. In line with this, the Indian government launched PM-KISAN as direct benefit transfers to farmers. However, a deep dive into the digital arena indicates farmers' challenges due to low cell connectivity or service quality. Issues faced by farmers in Odisha are explored in [11].

Quality mobile connectivity also facilitates education, certification, skills training, financial services like transfers and payments, and M2M agricultural services like crop, soil, and weather monitoring. Thus, access to Quality mobile/internet services on the one hand will improve the livelihood of farmers on the other hand improve agricultural productivity focusing on SDG 2 - *end hunger, achieve food security and improved nutrition, and promote sustainable agriculture*.

1.3 Proposed Solutions and Benefits

The objective of this paper is to have a system owned by the Department of Telecommunications (DoT), Government of India which can tell the ground reality of the telecom services in any geographical area of interest. The system should be trustable by the governments and the citizens, and it should be free from any conflict of interest (as per the current scenario, the Telecom coverage data is being collected from telecom service providers (TSPs)). So, the objective will be fulfilled in the following ways:

i. A mobile application of DoT to be installed by users as volunteer citizens.
ii. Real-time telecom parameters will be captured by the application without any reliance on TSPs (Telecom Service Providers).
iii. Historic mobile coverage data will show how the signal is improving in an area and the technology of an area is getting upgraded e.g.,3G to 4G to 5G.
iv. Operators will receive guidance regarding the necessary infrastructure or configuration adjustments, such as height, tilt, and power, to address the gaps in telecom coverage based on the collected parameters [12].

The organization, Department of Telecommunications (DoT) under the Ministry of Communications, Government of India will benefit in the following ways:

a) FIELD MONITORING TOOL: The Department will get a mobile app for real-time field monitoring (Figure 1) and by assessing user experience, telecom coverage effectiveness may be assessed to make necessary improvements [13].

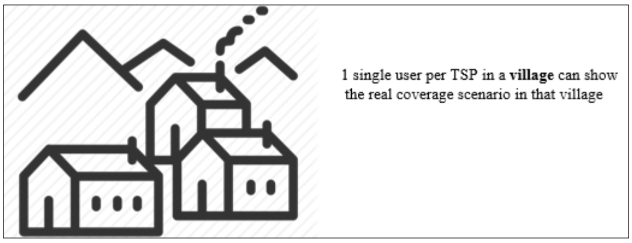

Figure 1 - A single telecom volunteer per village in India

b) PREDICTION MAPS: The dependency on TSP coverage forecast maps will be reduced significantly as an approach relying on automatic data collection from many user devices (volunteers) would likely provide better spectral maps (Figure 2) covering wide geographical areas than using data from a few accurate (but expensive) drive testing instruments [12].

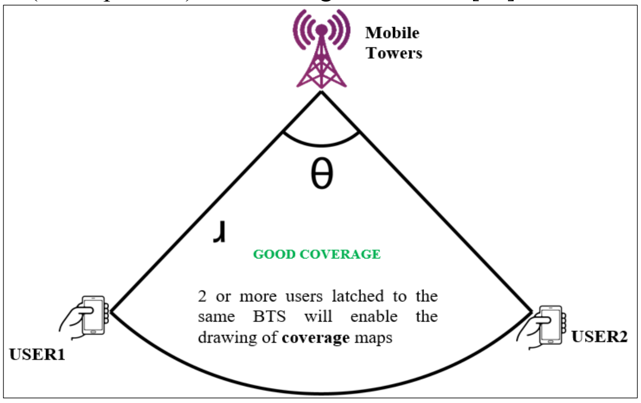

Figure 2 - Multiple telecom subscribers latched to the same BTS for coverage maps.

c) REDUCING EXPENDITURE: The dependence as of now is on RF (radio frequency) coverage prediction maps provided by the TSPs collected through driving a car/vehicle with a measuring kit along the main roads or in other accessible regions where vehicles can be driven [14]. Further, it is time-consuming, expensive, and not scalable [15], [16].

d) *FUTURE FOUNDATION*: The data collected will serve as a foundation for all future forms of data analysis (Figure 3) in the telecommunications industry. The designed application comprehensively captures all conceivable sets of attributes related to the telecommunications industry.

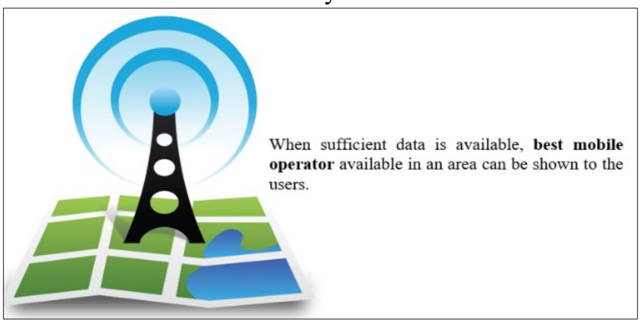

Figure 3 - shows the best mobile operator.

e) POLICY MAKING & STRATEGIC FUND ALLOCATION: DoT will possess the ability to identify regions across the country that exhibit the lowest signal quality, enabling them to develop relevant policies and allocate necessary finances to address this coverage issue.

2. OTHER RELATED WORKS

a) In a study done by FRACCARO, P. et al [15] regarding combining Geospatial Big Data and AI to predict Model Mobile Signal Strength, researchers gathered open-access geospatial data about several factors including weather conditions, tree coverage, land use patterns, altitude, and telecom infrastructure. This model was tested across the United Kingdom. The netBravo, crowdsourcing platform that was used didn't provide telecom operator information, so tower characteristics were averaged for locations with multiple operators using different transmitters. Current research employs a similar methodology by considering the average distance of the three nearest BTS as the measure of distance between the BTS and the telecom subscriber.

b) Alimpertis, E. [17] employed a machine learning framework and Android app data to anticipate missing values in mobile coverage maps. Location, time, BTS cell IDs, and device hardware were used. Current research produced an Android app. Similar to the current study, mobile crowdsourcing data from New York and Los Angeles metropolitan areas was employed. It developed a sophisticated Random Forests-based machine learning platform. The following study anonymised the dataset by assigning a random device ID to prevent monitoring of the original users. User data was uploaded to MongoDB and the current study used Cassandra. Both are NoSQL distributed databases.

c) In the study by M. F. Ahmad Fauzi et al. [5], out of various machine learning models, the Random Forest algorithm was selected and accolade as the predominant machine learning approach for developing a reliable RSRP (Reference Signal Receive Power) prediction model. In the current study as well, the Random Forest algorithm demonstrated the highest level of accuracy.

d) In the study by Wang, H et al. [18], the authors collected network measurement data from end-user's smartphones via crowd sensing and utilized machine learning techniques to create BSA (base station almanac) database.

e) In the study by F. Lyu et al. [4], the terrain was divided into 10 categories, including roads, buildings, manufactured items, tracks, vehicles, crops, trees, and rivers. It analyzed telecom coverage maps using machine learning. The current study used 18 terrain categories.

f) In the study by I. A. Saadi et al. [19] a drone was utilized in the study to forecast ground-level mobile signals. The artificial neural network predicted ground signal intensity from high-altitude data. It graded signal quality as excellent, good, fair, and bad, which matches the current methodology.

g) In the study by FRACCARO, P. et al [15], digital elevation models from The Shuttle Radar Topography Mission of NASA were used. It provides elevation data at a 30-meter resolution in latitude and longitude coordinates. In the current research, we used terrain data from the Space Applications Centre, ISRO (Indian Space Research Organization) having a spatial resolution of 56 meters.

h) In the study by D. Madariaga et al. [20] author showed that crowdsourced data from mobile devices can predict mobile signal strength while adding meteorological data enhancing regional accuracy.

i) In the study by D. F. S. Fernandes et al. [21], ANN (artificial neural networks) was used to predict path loss in mobile telecom networks.

3. GAPS OR INCONSISTENCIES IN EXISTING RESEARCH AND NEED FOR CURRENT STUDY

a) Minimization of Driving Tests is proposed as a 3GPP standard option for gathering measurement data from real users and assessing coverage [22]. MDT uses end-user devices to crowdsource measurements. MDT implementation in present networks is difficult because of imprecise positioning, limited data availability, and poor indoor reporting.

b) In another study by F. Lyu et al. [22] authors examined data from 31 cities, while the current analysis covers 711 districts out of the total 766 districts in India, representing coverage of 93% of the total districts of the country.

c) Dense urban areas with lots of buildings are prone to errors in positioning as noted by F. Lyu et al. [4] in the case of MDT. Precise location tracking in the mobile application developed in current research minimizes position errors. Signal loss makes indoor measurements impractical. Communication is crucial in larger buildings like airports. Thus, indoor positioning requires more accuracy than outdoors [6]. The current approach uses the telecom subscriber's precise location for extremely high location accuracy.

d) Further MDT seems to be more useful for coverage testing of a single telecom operator whereas the solution proposed in the current study considers all the telecom operators in India.

e) To protect user privacy, a study by FRACCARO, P. et al. [16] consolidated data into monthly releases at different resolutions (100 m and 1 km), resulting in less accurate results than current research. In the study by D. F. S. Fernandes et al. [22] a total of 12,194 mobile signal strength measures were used. However, the current study used a much larger dataset of 1.4 million mobile signal strength measurements.

f) M. F. Ahmad Fauzi et al. [5] simulated datasets of 12,011,833 samples, whereas in current research actual field-level measurement data was used.

g) According to the study by Alimpertis E [17], previous research often focused on evaluating the raw signal intensity and reducing mean square error, which may not align with telecom operators' priorities. Telecom subscribers' main issue is signal quality, good vs. bad coverage, hence current research also focuses on signal quality rather than signal strength.

4. NEED FOR CURRENT DEVELOPMENT STUDY

The need for current development as mentioned in this paper arises due to the following reasons.

i. Presently, there is no dedicated machine-learning model for signal strength interpolation in mobile networks in India. However, the signal strength value and tower position as features in ML can be used to calculate signal strength at new sites [23]. The novel findings of the current study will lay the groundwork for further study in this area which can be used by other countries as well.

ii. Most coverage prediction maps are unreliable. The Federal Communications Commission (FCC) distributes annual Broadband Deployment reports and shapefiles for telecom providers indicating their geographic coverage areas. Scholars caution this data may be false. M. Nekrasov et al. [13] found that Pennsylvania's broadband connection is considerably overestimated using Measurement Lab speed tests. Network providers sometimes exaggerate coverage with proprietary propagation models. Public systems for evaluating coverage regions and signal strength, especially in poor areas, are needed to assess mobile broadband accessibility.

iii. The DoT, responsible for telecom policy formulation in India, lacks an independent framework for signal quality assessment. Hence developing this system was imperative.

5. PROPOSED SOLUTIONS

5.1 Development of the mobile application

5.1.1 National Telecom Volunteer - mobile application

A smartphone application [24] was developed (Figure 4) to serve as a crowdsourcing platform, as DoT lacked a preexisting mobile application for this. The app was designed for installation by volunteer citizens who are prepared to donate telecom-related data, despite the potential impact on their mobile device's battery life and data consumption.

This tool collects statistics on voice quality, data network quality, mobile base stations, and mobile phone hardware characteristics. This massive data gathering is intended to cover all future research applications. Such crowdsourced measures are more cost-effective and efficient for predicting post-deployment coverage gaps caused by highway construction or consumer residential preferences etc. [25]. Adjusting antenna tilt, height, and power or adding more BTS addresses these coverage gaps, notably in the 5G [26]. The NTV (National Telecom Volunteer) app was developed and released on Google Play on August 4, 2023.

Figure 4 - NTV mobile application launched on Google Play

5.1.2 Feature of the App

The application displays to the user the mobile signals available on his/her mobile phone (Figure 5). The application shows the telecom operator's name. It also shows the telecom technology to which the system is latched whether 2G, 3G, 4G or 5G. Further, it displays the signal strength in dBm. Signal strength affects mobile network performance and QoS and mobile networks often struggle to meet 5G application requirements due to poor signal quality [27]. Telecommunications companies need accurate mobile signal strength estimates to optimize cell tower placement and improve consumer connectivity [15]. For non-telecom experts, the statistics indicate the signal's quality as GREAT, GOOD, MODERATE and POOR.

Figure 5 - NTV application 'Signals tab'.

5.1.3 Data Captured by the App

The application uses the Work Manager component of the Android API to periodically collect the specified data attributes (Table 1) in the background at 15-minute intervals. The application provides users with the option to pause and resume background data collection at their convenience. To motivate users, the application displays the total count of data samples collected, showing the user's contribution.

Table 1 – Data captured by the NTV mobile application.

Signal	Data Network	Hardware	Subscriber
Operator Technology Signal strength	Downstream Bandwidth	Manufacturer	Latitude
Channel No., Base Station Code, Tracking Area Code	Upstream Bandwidth	Brand	Longitude
Cell ID, Location Area Code	Interface name	Model	Timestamp
Mobile Country/Network code	Transport Type (Mobile/Wi-Fi)	Product	Other Network information

5.2 Server Architecture

A highly reliable architecture (Figure 6) has been used as a backend server for the mobile application, which serves as the recipient of data pushed by the mobile application. Amazon Elastic Compute Cloud (EC2) was used as the cloud computing platform to host the databases. Ubuntu Linux distribution v. 22.04.2 LTS is used as the operating system on the server. The live database of the mobile application utilizes Apache Cassandra version 4.0.10.

Apache Hive version 3.1.2 database was installed as a data warehouse to store historic old data. Apache Spark version 3.4.1 was installed for real-time batch processing of data to be fed to the Geospatial dashboard.

A geospatial dashboard was developed to show the coverage status on the political map of India. A machine learning model was used to predict the mobile coverage at places for which no data have been captured but have similar characteristics to the locations whose telecom data is present.

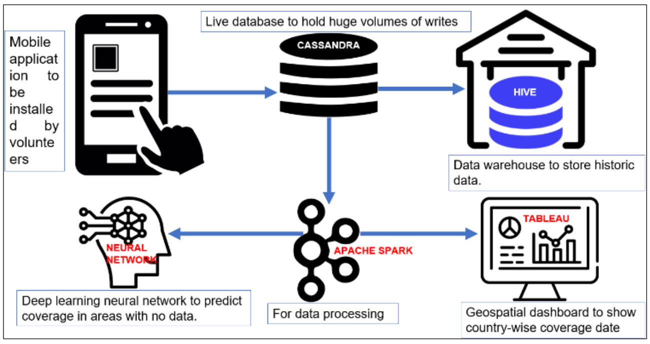

Figure 6 - NTV application back-end server architecture

5.3 Data Description and Collection

The application was built in a pilot mode during the research phase; hence the data was insufficient to train the machine learning model. The paper explains how supplemental data sources (Figure 7) fed the machine-learning model. Since Received Signal Strength is affected by trees, buildings, density, humidity, and latched BTS proximity [27], the authors used signal strength data and three other types of data.

a) MOBILE COVERAGE DATA (KUSHAL SANCHAR): This is a mobile coverage data dump of the Kushal Sanchar mobile application, which was released by DoT on March 12, 2022. There were 204,207 records available. Given the author's affiliation with the DoT as an employee, got privileged access to otherwise non-publicly available data.
b) MOBILE COVERAGE DATA (TRAI MYSPEED): Another mobile coverage data dump used was from the TRAI MySpeed mobile application, which TRAI launched in August 2016. A total of 13,32,850 records were accessible.
c) BTS DATA: The dataset contains 2.7 million Base Transceiver Stations from all TSPs in the country. Due to the author's DoT affiliation, we obtained this data from the National EMF site. Again, this highly sensitive data is not publicly available.
d) TERRAIN DATA: ISRO's Space Applications Centre (SAC) provided the earth terrain data. Geographical and landform data was added to telecom coverage data to improve mobile coverage estimation [22]. Signal strength forecast depends on location and terrain [16], [20]. Signals fade differently in different mobile radio channels due to the type of signal and channel characteristics like hills, trees, and buildings between the transmitter and receiver [28]. Terrain information was provided for each location of the sample telecom subscriber. Topography is divided into 18 classes. Public access to this data is restricted.
e) WEATHER DATA: The weather data comes from Open-Meteo, an open-source API. Open-Meteo's free Historic Weather API offered weather data for the telecom subscriber's location at the time of sampling. Raindrops and fog absorb radio wave power, affecting signal strength. According to the study [22], heat loss or scattering dissipates this received power. The air refractive index changes constantly, refracting radio waves. Hence, the received signal strength fluctuates with atmospheric conditions [19].

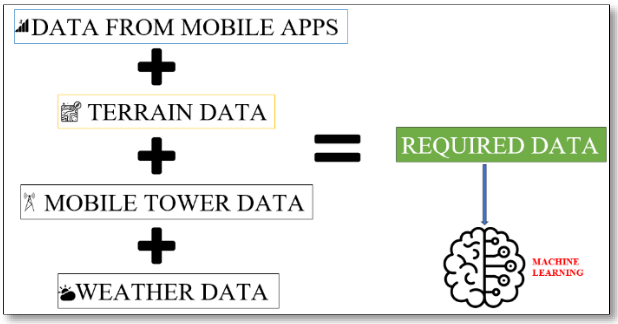

Figure 7 – Data utilization in analysis

5.4 Feature Engineering

Aggregating the datasets above prepared data (Table 2) for exploratory data analysis and machine learning. The telecom subscriber's latching BTS was unavailable, thus the telecom subscribers' average distance from the three nearest BTSs in the database, was calculated using the Haversine formula. This average distance was the dataset attribute. Another study [15] employed the three nearest BTS because the mobile subscriber's locked BTS was unavailable. The telecom subscriber will connect to the next nearest BTS if a mountain or structure blocks the nearest BTS.

Table 2 –Unique Data Frames Attributes.

S. N.	ATTRIBUTE	COUNT	UNIQUE VALUES(REMARKS)
1	Telecom Operator	4	'RJIL', 'AIRTEL', 'BSNL' & 'VIL'
2	Telecom Technology	2	'4G' & '3G' (Telecom operators are gradually phasing out 2G from India and 5G was just recently launched on 1st Oct 2023)
3	State	37	All the states and Union territories of India
4	District	711	711 districts are covered out of 766 districts in India (93% of districts are covered.)
5	Technology Type	7	***4G***: *'TDD LTE', 'FDD LTE-1800', 'FDD LTE-850', 'FDD LTE' & 'FDD LTE-2100'* ***3G***: *'UMTS-2100' & 'UMTS' (All possible 3G & 4G telecom technologies variants are covered)*
6	Terrain	18	*'Built-up', 'Kharif Crop', 'Double/Triple Crop', 'Rabi Crop', 'Deciduous Forest', 'Current Fallow', 'Plantation', 'Waterbodies max', 'Grassland', 'Degraded/Scrub Forest', 'Wasteland', 'Evergreen Forest', 'Waterbodies min', 'Littoral Swamp', 'Rann', 'Snow Cover', 'Zaid Crop', 'Shifting Cultivation'*
7.	Signal Quality	4	*'SIGNAL_STRENGTH_GREAT', SIGNAL_STRENGTH_GOOD', 'SIGNAL_STRENGTH_POOR' & 'SIGNAL_STRENGTH_MODERATE'*

5.4.1 Signal Strengths to Signal Quality

Telecom subscribers' QoS depends on signal strength. Different signal strengths indicate different signal quality. Within a range, telecom subscribers receive the same service quality. The operator prioritizes the prediction of good and bad coverage over signal strength [17]. Thus, these signal quality classifications (Table 3) were used.

Table 3 – Signal Quality ranges

S. N.	Signal Quality	Range of Signal Strength (dBm)
1	GREAT	>= -85
2	GOOD	>= -95 & <-85
3	MODERATE	>= -105 & < -95
4	POOR	>= -115 & <-105

5.4.2 Duplicate Samples Removal

A single telecom subscriber in the same territory could provide many data samples. Thus, data was sanitized when the telecom operator, technology, latitude, and longitude matched. Removing duplicate data samples from 14,28,399 entries yielded the pandas dataframe with 8,81,228 unique records.

5.4.3 Feature Selection

This involves choosing the most important and relevant features from collected data. This enhances the machine learning model by decreasing input data while maintaining crucial information for accurate predictions. Machine learning requires feature selection to simplify models, prevent overfitting, and increase generalisation to new, unobserved data. The following dataset features (Table 4) were fed to machine learning and deep learning algorithms:

Table 4 – Features selected.

S.N.	FEATURE	REASON FOR SELECTION
1	State	Telecom operators have separate network teams for each state and district configuring BTS in that area.
2	District	
3	Telecom Operator	Each telecom operator has a separate strategy for telecom coverage.
4	Telecom Technology	Every telecom technology and type have a separate science of telecom coverage
5	Technology Type	
6	Log Distance from mobile tower	The greater the distance from the tower, the lesser is mobile coverage.
7	Log Mobile Tower Height	The height of a tower is of utmost importance in establishing and sustaining a direct LoS between the tower and mobile devices.
8	Terrain	Signal strength forecast depends on location and terrain
9	Weather_Temperature	Raindrops and fog absorb radio waves' power affecting signal strength. As per study [15] heat loss or scattering dissipates the received power. The refractive index of air changes constantly refracting radio waves.[12]
10	Weather_Relative_Humidity	
11	Weather_Pressure	
12	Weather_Precipitation	
13	Weather_Rain	
14	Weather_Snowfall	
15	Weather_Cloud_Cover	
16	Weather_Windspeed	
17	Signal Quality	The target class

5.5 Machine Learning

Researchers anticipate signal strengths using Machine Learning. This prediction is a continuous classification problem. Linear Regression (LR), Locally Weighted Regression, Random Forest (RF), Artificial Neural Networks (ANNs), and others are popular [14] algorithms used.

Before inputting the data into the selected machine learning and deep learning models, a series of preprocessing steps were undertaken label encoding, random oversampling, standardization and one-hot encoding of output vector.

6. RESULTS WITH PROPOSED ML MODEL

After analyzing different machine learning and deep learning neural network algorithms, we have determined that the **Random Forest algorithm** shows the most potential as a machine learning model for predicting telecom signal quality in the topography of India. The algorithm exhibits an accuracy rate of 84% (Table 5), a figure that is approximately twice as high as the accuracy rates observed in the other examined algorithms tabulated below. Even in the study [5], among a range of machine learning models the Random Forest algorithm was chosen and strongly endorsed as the machine learning model for constructing a robust RSRP prediction model. It is also important to note that none of the FNN (Feedforward Neural Network) algorithm versions we tested learned the pattern to predict the link between features and target class with even limited accuracy. All FNN models had a substantial loss, indicating very poor predictions. Future studies are needed to find additional hyper-parameters to train this model accurately.

Table 5 – Results of various machine learning algorithms

S. N.	Algorithm	Accuracy
1	LOGISTICREGRESSION	38%
2	**RANDOMFORESTCLASSIFIER**	**84%**
3	GRADIENT BOOSTING CLASSIFIER	45%
4	AUTOSKLEARNCLASSIFIER	44%
5	FNN WITH ADAM OPTIMIZER	48.67%
6	FNN WITH SGD OPTIMIZER	48.44%
7	FNN WITH RMSPROP OPTIMIZER	43.1%
8	FNN WITH L1 REGULARIZATION	24.96%
9	FNN WITH L2 REGULARIZATION	36.77%
10	FNN WITH DROPOUT	48.38%
11	FNN WITH EARLY STOPPING	48.6%
12	TABNETCLASSIFIER	47.26%

7. CONCLUSION, LIMITATION AND FUTURE SCOPE

7.1. CONCLUSION

Current research has developed a very adaptable model that may be used in different locations of the country without prior mobile signal intensity knowledge. This makes it better than geographic interpolation. The current study uses machine learning to estimate signal intensity in data-poor areas, enabling the hypothetical planning of new telecom infrastructure for an entire region.

The strategy may also use the USOF (Universal Service Obligation Fund) to provide connectivity, as the government did in Left Wing Extremism-hit areas [31]. The current study is crucial for fair justice. SDG 8 promotes *decent work and economic growth*, while SDG 9 emphasizes *industry infrastructure and innovation*. Being an IT hub, India has proved most IT companies can function successfully from home and boost economic growth. India, a startup powerhouse, has a strong IT sector built on ICT, industry, and infrastructure. Equitable digital connectivity will provide Delhi and Dimapur employees an equal chance [31].

7.2. LIMITATIONS OF THE PROPOSED MODEL

1) The machine learning model of this study lacks information on the BTS to which the mobile user was connected, also highlighted in the study by Alimpertis, E. [17].

2) Public crowdsourcing systems to collect signal data primarily focus on prominent transit routes, hence excluding communities situated beyond these regions [32], [33].

3) In addition to antenna design, device materials can affect radio signal absorption in mobile phones. Receiver circuitry noise, nonlinearity, and frequency band support affect performance. M.-R. Fida and others [12] and T. K. Sarkar [28] found that user mobility, device orientation, humidity, and temperature affect signal variation.

4) External factors that could affect mobile internet QoS must be carefully examined like data from big crowds in limited locations like sports events, concerts, and street protests is useful, can be added to machine learning algorithms to improve signal strength forecasts [20].

5) For network-based services, QoS depends on throughput, delay, packet loss, and error rate rather than just RSRP. Numerous other network factors affect service quality. High interference (Reference Signal Received Quality) can degrade performance even with a robust RSRP [17].

6) Current research uses GPS (global positioning system) for precise position measurement. Note that not all gadgets have GPS [17].

7) Smartphones' own algorithms calculate signal strength differently. Device noise also affects wireless receiver sensitivity [17].

8) As was the case in the study by M. F. Ahmad Fauzi et al. [5] , data in the current study did not include 5G samples, since 5G was introduced in October 2022.

7.3. FUTURE SCOPE OF WORK

1) Using data from mobile app NTV developed in current research, telecom companies can identify landforms in inaccessible or unneeded areas where signal coverage is not needed [22]. This action could lower telecom provider costs.

2) Still a challenge is developing novel deep learning models for tabular data. The enhanced TabNet (iTabNet) described in the study [34] is the solution proposed. Once its Python library implementation is available, NTV data can be used to verify its efficacy.

3) Drive test and MDT data can be used to train models and improve signal quality predictions [14].

4) The two-dimensional and three-dimensional distance between the transmitter and receiver, height difference, transmitter tilt and azimuth angles, transmitting power, clutter and building heights information, etc. can be used in the ML algorithm to improve accuracy [5].

5) This app has been integrated into the existing Umang [35] mobile application having a user base of around 50 million.

6) e-delivery services utilizing Gig economy services like Zomato, Uber, and Amazon may include NTV as an auxiliary code to their apps to improve their Corporate Social Responsibility (CSR) efforts. This will greatly increase the amount and variety of data samples for our model.

7) NTV applications can be installed in vehicles, especially public service vehicles that run on long routes from one state to another, like buses, allowing them to act as participants in carrying out mobile crowd-sensing tasks [36].

8) It is crucial to implement security measures and regularly monitor the data collected to ensure the authenticity of the system is not compromised [36] through fake sensing attacks.

9) After government clearance, certification, and testing, Android and iOS platforms may grant unique rights to the mobile app NTV like accessing the phone's state without user authorization, enabling automatic data collection.

8. ACKNOWLEDGEMENT

We would like to express gratitude to our employer Department of Telecommunications for approving our research work for developing the tool and collecting the data from volunteers. Further, our heartfelt thanks to all the volunteers who have installed our NTV app displaying full trust. We also thank ISRO for providing us with terrain data of millions of locations. We express profound appreciation to our families for their steadfast support and appreciation of our aspirations and inclinations. Above all, thanks to the almighty for providing us with all privileges, and for becoming what we are today.

REFERENCES

[1] Telecom Regulatory Authority of India. (9th February 2024) Press Release on Indian Telecom Services Performance Indicator Report for the Quarter ending September 2023. [Online] [Accessed on 29th March 2024] https://www.trai.gov.in/sites/default/files/PR_No.05of2024.pdf

[2] GSMA. (16th September 2023) Improvements in Mobile Connectivity Correlate to Greater Impact Towards Sustainable Development Goals, According To Gsma Report. [Online] [Accessed on 29th March 2024] https://www.gsma.com/newsroom/press-release/improvements-in-mobile-connectivity-correlate-to-greater-impact-towards-sustainable-development-goals-according-to-gsma-report

[3] J. Johansson, W. A. Hapsari, S. Kelley and G. Bodog, "Minimization of drive tests in 3GPP release 11," in IEEE Communications Magazine, vol. 50, no. 11, pp. 36-43, November 2012, doi: 10.1109/MCOM.2012.6353680

[4] F. Lyu et al., "Coverage Estimation of Wireless Network Using Attention U-Net," 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Wuhan, China, 2022, pp. 1255-1260, doi: 10.1109/TrustCom56396.2022.00174

[5] M. Ahmad Fauzi et al., "Mobile Network Coverage Prediction Based on Supervised Machine Learning Algorithms," in IEEE Access, vol. 10, pp. 55782-55793, 2022, doi: 10.1109/ACCESS.2022.3176619

[6] A Daniel Baumann Supervisor: Tsvetko Tsvetkov Seminar Future Internet WS2013/2014 Chair for Network Architectures and Services Department of Computer Science, Technische Universität München Email: daniel.baumann@in.tum.de https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2014-03-1/NET-2014-03-1_02.pdf

[7] Education Today. (5th October 2020) Sonu Sood installs mobile tower in Haryana village after students struggle for

online access. [Online] [Accessed on 29th November 2023] https://educationtoday.co/blogs/?p=2939.
[8] Siddhanta Mishra & Abhay Singh. (24th May 2023) Delhi: Biker killed in Pragati Maidan tunnel; poor signal delays call. [Online] [Accessed on 29th November 2023] https://timesofindia.indiatimes.com/city/delhi/biker-killed-in-tunnel-poor-signal-delays-call/articleshow/100459130.cms?from=mdr
[9] NABARD. (25th Jan 2023) Trends and Patterns in Agriculture Credit in India: A District Level Analysis of Uttar Pradesh. [Online] [Accessed on 29th March 2024] https://www.nabard.org/auth/writereaddata/tender/2501235626trends-and-patterns-in-agriculture-credit-in-india.pdf
[10] Suri, T., & Jack, W. (2016). The long-run poverty and gender impacts of mobile money. Science. https://doi.org/aah5309
[11] The WIRE. (30th Nov 2023) Navigating PM-KISAN: A Deep Dive into Digital Challenges Faced by Farmers [Online] [Accessed on 29th March 2024] thewire.in/government/navigating-pm-kisan-a-deep-dive-into-digital-challenges-faced-by-farmers
[12] M. -R. Fida et al., "Impact of Device Diversity on Crowdsourced Mobile Coverage Maps," 2018 14th International Conference on Network and Service Management (CNSM), Rome, Italy, 2018, pp. 348-352
[13] M. Nekrasov et al., "Evaluating LTE Coverage and Quality from an Unmanned Aircraft System," 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Monterey, CA, USA, 2019, pp. 171-179, doi: 10.1109/MASS.2019.00029.
[14] Y. Liu et al., "3.5 GHz Outdoor Radio Signal Strength Prediction with Machine Learning Based on Low-Cost Geographic Features," in IEEE Transactions on Antennas and Propagation, vol. 70, no. 6, pp. 4155-4170, June 2022, doi: 10.1109/TAP.2022.3175214
[15] FRACCARO, P. et al. Modelling mobile signal strength by combining geospatial big data and artificial intelligence. ACM International Conference Proceeding Series, [s. l.], 2020. DOI 10.1145/3448823.3448879. Disponível em: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=edselc&AN=edselc.2-52.0-85102954047&site=eds-live&scope=site.%20Acesso%20em:%205%20set.%202023
[16] O. O. Erunkulu et al., "Cellular Communications Coverage Prediction Techniques: A Survey and Comparison," in IEEE Access, vol. 8, pp. 113052-113077, 2020, doi: 10.1109/ACCESS.2020.3003247
[17] Alimpertis, E. (2020). Mobile Coverage Maps Prediction. UC Irvine. ProQuest ID: Alimpertis_uci_0030D_16368. Merritt ID: ark:/13030/m5dc3933. Retrieved from https://escholarship.org/uc/item/1wr2b02f
[18] Wang, H et al. Big Data-Driven Cellular Information Detection and Coverage Identification. Sensors 2019, 19, 937. https://doi.org/10.3390/s19040937
[19] I. A. Saadi et al., "Ground Level Mobile Signal Prediction Using Higher Altitude UAV Measurements and ANN," 2022 32nd Conference of Open Innovations Association (FRUCT), Tampere, Finland, 2022, pp. 15-21, doi: 10.23919/FRUCT56874.2022.9953813
[20] D. Madariaga et al., "I'm Only Unhappy when it Rains: Forecasting Mobile QoS with Weather Conditions," 2018 Network Traffic Measurement and Analysis Conference (TMA), Vienna, Austria, 2018, pp. 1-6, doi: 10.23919/TMA.2018.8506509
[21] D. F. S. Fernandes et al., "Comparison of Artificial Intelligence and Semi-Empirical Methodologies for Estimation of Coverage in Mobile Networks," in IEEE Access, vol. 8, pp. 139803-139812, 2020, doi: 10.1109/ACCESS.2020.3013036
[22] F. Lyu et al., "Coverage Estimation of Mobile Network Using Supervised Learning Model on Artificial Estimation Dataset," 2021 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Hangzhou, China, 2021, pp. 214-220, doi: 10.1109/ICT-DM52643.2021.9664185
[23] M. Zaboub et al., "5G Mobile Network Signal Strength Interpolation Using Machine Learning GAN Algorithm," 2023 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Montreal, QC, Canada, 2023, pp. 362-368, doi: 10.1109/WiMob58348.2023.10187812
[24] Himanshu Sharma, "National Telecom Volunteer" Version 1, Android, 1st Aug 2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.dot.nbm&hl=en
[25] Molinari, M et al., M.K. C2B(I)D 2015 - Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data, Part of SIGCOMM 2015, 17th August 2015, :33-38 Language: English. Association for Computing Machinery, Inc DOI: 10.1145/2787394.2787395
[26] Luo, J., & Xu, S. (2023). NCART: Neural Classification and Regression Tree for Tabular Data. ArXiv. /abs/2307.12198
[27] M. Zaboub, M. Corici and T. Magedanz, "5G Mobile Network Signal Strength Interpolation Using Machine Learning GAN Algorithm," 2023 19th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Montreal, QC, Canada, 2023, pp. 362-368, doi: 10.1109/WiMob58348.2023.10187812
[28] T. K. Sarkar et al., "A survey of various propagation models for mobile communication," in IEEE Antennas and Propagation Magazine, vol. 45, no. 3, pp. 51-82, June 2003, doi: 10.1109/MAP.2003.1232163
[29] Chen, K et al. (2023). Trompt: Towards a Better Deep Neural Network for Tabular Data. ArXiv. /abs/2305.18446
[30] The Hindu. (27th April 2022) Cabinet approves ₹2,400-crore project to upgrade to 4G services in Left Wing Extremism-hit areas. [Online] [Accessed on 29th March 2024] https://www.thehindu.com/news/national/cabinet-approves-2400-crore-project-to-upgrade-to-4g-services-in-left-wing-extremism-hit-areas/article65360614.ece
[31] Jordan Pioth. (6th Nov 2023) The Problems with Bandwidth Usage While Working from Home. [Online] [Accessed on 29th March 2024] https://www.coeosolutions.com/news/problems-bandwidth-working-from-home
[32] X. Li et al., "Maximizing Mobile Coverage via Optimal Deployment of Base Stations and Relays," in IEEE Transactions on Vehicular Technology, vol. 65, no. 7, pp. 5060-5072, July 20 16, doi: 10.1109/TVT.2015.2458015
[33] Coverage bias: the impact of eligibility constraints on mobile phone-based sampling and data collection Helen SuichORCID Icon,Mandy YapORCID Icon &Trang Pham, https://doi.org/10.1080/13645579.2021.1957310, International Journal of Social Research Methodology
[34] Liu, J., Tian, T., Liu, Y. et al. iTabNet: an improved neural network for tabular data and its application to predict socioeconomic and environmental attributes. Neural Comput & Applic 35, 11389–11402 (2023). https://doi.org/10.1007/s00521-023-08304-7
[35] MeitY, Government of India, "UMANG" Version 2.0.103, Android, 3rd Jul 2024. [Online]. Available: https://play.google.com/store/apps/details?id=in.gov.umang.negd.g2c&hl=en_IN
[36] H. Braham et al., "Fixed Rank Kriging for Cellular Coverage Analysis," in IEEE Transactions on Vehicular Technology, vol. 66, no. 5, pp. 4212-4222, May 2017, doi: 10.1109/TVT.2016.259