SEMI-SUPERVISED LEARNING-BASED COVERAGE HOLE DETECTION IN CELLULAR NETWORKS

Shahriar Abdullah Al-Ahmed[1] and Muhammad Zeeshan Shakir[1]
[1]University of the West of Scotland, Paisley, United Kingdom

NOTE: Corresponding author: Shahriar Abdullah Al-Ahmed, shahriar.al-ahmed@uws.ac.uk

Abstract – *For any time-critical mobile network-dependent applications and services, coverage is one of the prominent factors for providing the best Quality of Service (QoS) and Quality of Experience (QoE). A simple Coverage Hole (CH) may degrade the performance and the reputation of any operator by reducing the Key Performance Indicators (KPIs). This is one of the important aspects which need to be planned from the phase of network deployment throughout the whole operational stage. Many factors can cause CH such as attenuation, obstacles and improper network planning. Traditionally, a Drive Test (DT) used to be carried out in order to assess the quality of the mobile network signal. With technological advancement, DT has been replaced by the Minimization of Drive Test (MDT) and included as a part of Self-Organizing Networkss (SONs). The MDT process is applicable to networks that operate in 3G, 4G and 5G technologies. With this method, operators are able to measure network performance with the help of end users' devices. Thus, the network can be managed more conveniently, performance is improved, quality is increased, and maintenance costs are reduced for the network. However, the processing of MDT at the operators' side remains time-consuming and complex especially for CH analysis and detection from mobile network data. Therefore, we present a method by utilising Semi-Supervised Learning (SSL) in this paper so that this task becomes uncomplicated with improved accuracy. Our results show that the proposed method achieves better accuracy than the usual classification algorithm.*

Keywords – 5G and beyond, coverage hole detection, machine learning, semi-supervised learning

1. INTRODUCTION

Technologies are improving expeditiously for the enhancement of our lives. One technology leads to the advancement of another for the same or many other purposes. Mobile communication technology is a good example of this. Mobile networks were once used solely for making phone calls but now the same technology serves many other purposes as well. As a result, the number of devices has also increased exponentially and created a new era of diverse mobile communication systems such as human-centric, machine-to-machine and human-to-machine. From one side, this communication system is helping mankind to achieve important goals. On the other side, network management is becoming complex due to the criteria of such a system. Among the criteria that can be considered are capacity, coverage and latency. The greatest importance of all lies in coverage since no devices can connect to the communication network if it has poor coverage or Coverage Holes (CHs). It is then possible to take into account latency and capacity if the wireless area network has sufficient coverage.

A critical element of network deployment is ensuring that no CHs are present from the outset. It is an area in the transmission footprint of a cell access point where the received signal level of the serving cell and its configured neighbour is below the threshold levels required to maintain the service at a minimum quality and healthy radio performance. Physical impediments (such as new structures and hills), inappropriate antenna parameters, hardware faults, inappropriate Radio Frequency (RF) planning, sleeping cells, and so on cause CHs.

Traditionally, a motor Drive Test (DT) is used to detect this with the use of specialised hardware and software to collect fixed radio measurements from cells by driving around the area of investigation [1]. As the infrastructure evolved over time, it was becoming more expensive and time-consuming to carry out a DT especially in dense areas with Non-Line-Of-Sight (NLoS) situations [2].

Although a DT provides more accurate measurements with the aid of scanning resources, some attributes including the limitations of a DT, automation of the network as well as restricting the Human-in-The-Loop (HiTL) model caused the 3rd Generation Partnership Project (3GPP) to announce the Minimization of Drive Test (MDT) in release 10 [3]. MDT has major advantages over DT in network planning, network testing, coverage estimation, User Equipment (UE) tracking, identifying abnormalities and above all reducing Capital Expenditure (CAPEX) and Operating Expenditure (OPEX). In MDT, the UEs sporadically provide the signalling of the network, geolocation and timestamp information automatically to the network operators. Therefore, the logged measurements help the network operators to implement coverage optimisation, capacity optimisation, Quality of Service (QoS) verification, tracking connections and radio channel characterisation [1].

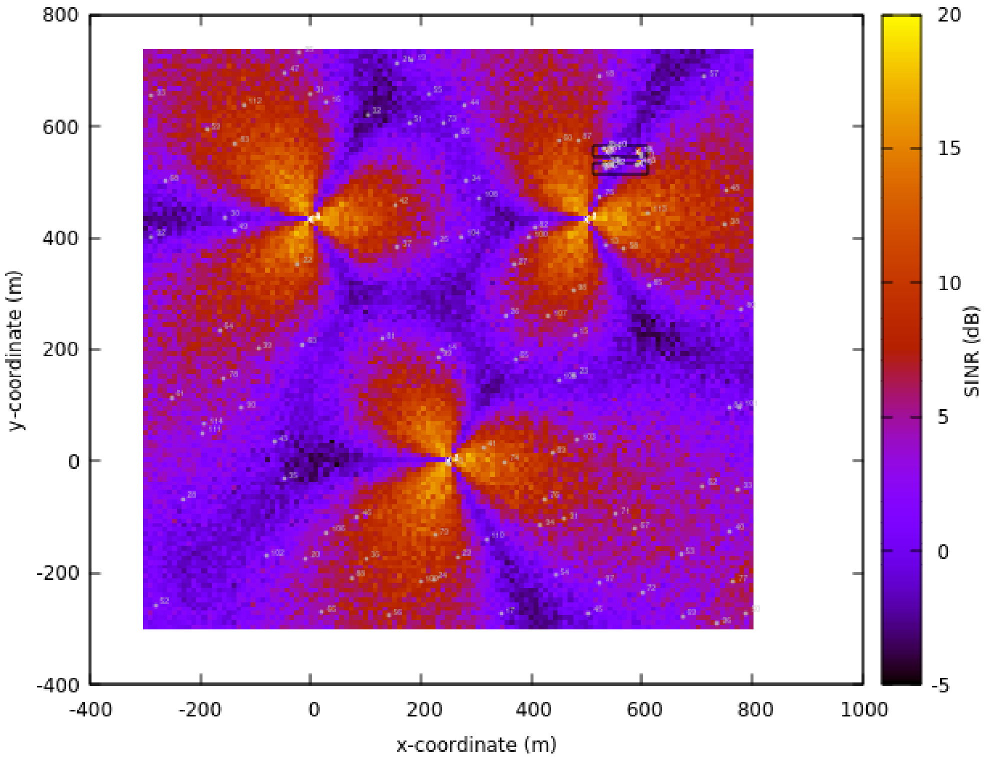


Fig. 1 – Radio Environment Map (REM) of LTE deployment in a real-world scenario.

During the MDT process, the UEs send information via the network to the operators for optimisation purposes. The UE with MDT features and GPS facilities (for location record) is used on the network. The reports are either reported at intervals or on request at the data centre of the operators. According to 3GPP Technical Specification (TS) 34.422, MDT has been defined as two types: area and subscription based on the network signalling [4]. The first one is the MDT measurement which is collected in an area and the second one is the MDT measurement from one UE only [5]. Furthermore, MDT can be classified into two types from a radio configuration perspective: immediate and logged. Immediate MDT occurs where information can be collected by a UE in connection mode in real time. On the other hand, the logged one occurs where information is collected in idle mode [1]. These definitions indicate the operators receive a huge amount of data at the data centre for processing purposes.

The MDT process can be widely applied to 5G and beyond networks, since it relies heavily on the end user's device. Furthermore, it can reduce the OPEX, increase network quality and performance and help to plan and expand the network. Small cells and femtocells are going to play a dominant role in 5G and beyond. Femtocells or Home eNodeBs (HeNBs) are mainly deployed indoors. So, more measurements are required from indoors to find the CH. In addition, the ultra-dense deployment of small cells or femtocells will lead to cell-less or cell-free architecture which will add higher interference with the Macro-Base Station (MBS) signals [6]. As a result, the CH hole will appear. MDT may enable the operators to mitigate the interference and make the deployed small cells or femtocells more intelligent [7].

Even though the MDT eliminates the disadvantages of DT, the subsequent handling of huge data from the MDT is largely manual on the operators' side [8]. For any operator, storing and analysing huge amounts of data is a complex task. In addition, QoS needs to be superior to earn revenues and compete. Accordingly, the recent advancements in technology and Machine Learning (ML) have led to a more straightforward way to accomplish these tasks. This paper proposes to use Semi-Supervised Learning (SSL) to examine the MDT data so as to improve operators' services by removing any CH or poor signals.

1.1 Related work

The competition among operators made the research of CH detection worthwhile. As a result, more revenue is generated by increasing QoS. Numerous concepts have been proposed by researchers since the beginning of the mobile network rolling out. The research became more intense with the announcement of the MDT from 3GPP. A remarkable research on CH detection can be found in [9] in which the authors considered inter-Radio Access Technology (RAT) handover information to identify CH. The information was mainly UE traced data consisting of geolocations and time stamps for investigation and troubleshooting with the help of IF-THEN conditional rules. A similar inter-RAT scenario has been considered by authors in [10]. Here, the researchers investigated the Base Station Subsystem Application Part (BSSAP) and radio resource management messages to identify the inter-technology handovers from 3G technology to 2G using the Hadoop platform. For Long-Term Evolution (LTE), the CH has an impact on the network due to the elements of the network [9]. Another study presented by authors in [11] which is a graph theory-based network insight analysis framework to detect CH. The authors used both network data and user behaviour data for their study but the accuracy of their study is limited to the lack of consideration of location. The authors in [12] proposed the use of spatial Bayesian geostatistics to build a Radio Environment Map (REM) to detect CH by considering UE data remotely. In this research, the size of the sample data was limited which raises the question of accuracy.

The same authors in [13] proposed a REM cognitive tool-based approach that provides REM where data was coming from location-aware devices or basically MDT. The approach seemed to function very well but the input data was from a planning tool and the obtained results were based only on models [14]. The authors of [14, 15] focused on the QoS evaluation using different Key Performance Indicators (KPIs) and correlated with location data to investigate how satisfied end users were. The authors in [16] also focused on the QoS and used ML algorithms such as k Nearest Neighbour (kNN) to characterise the satisfied and unsatisfied users. These studies were concerned with the QoS verification. Another classification approach has been observed in the study [17]. The authors proposed extended Radio Link Failure (RLF) reporting for mobility and coverage optimisation. On the basis of the results, the RLFs are classified into three groups: interference, downlink coverage and handover problems. The RLF is an event triggered method which will not provide the full picture of the network.

The studies in [18, 19, 20, 21, 22] investigated MDT to pinpoint the sleeping cells or network outage with the aid of Supervised Learning (SL) or Unsupervised Learning (UL). The anomaly detection was also investigated by authors in [23] utilising call detail records. Additionally, the authors applied SSL as their ML tool to evaluate their results. The sleeping cells or cell outage or anomaly is usually for large areas whereas the CH or poor reception is for a small area. It is possible to have poor reception without any RLF trigger.

In general, the above studies adopted effective data collection methods and ultimately applied tools to detect CH. Despite this, the techniques had some problems with simplicity and sufficiency for data processing and identification of issues. Due to these limitations, we have studied several ML tools to process data at the operators' side to remove complications. Thus, this is the first paper, based on the authors' knowledge, that processes mobile network data to detect CH at the operators' side using SSL.

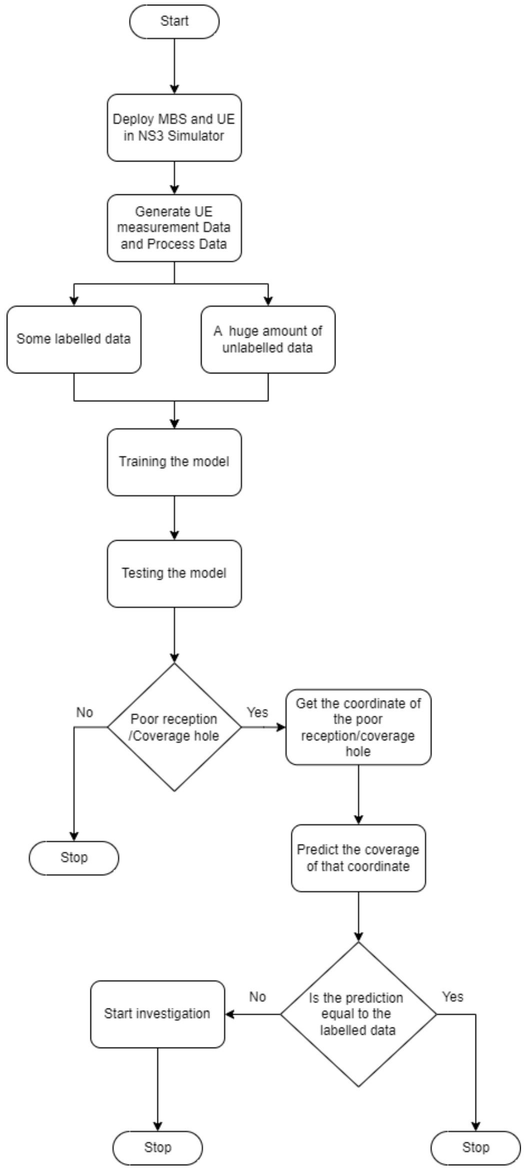


Fig. 2 – Flow chart of the proposed method.

1.2 Our contribution

The main contribution of this article is in two folds:

- We exploit the NS3 simulator to create an MDT database from UE measurements and apply semi- supervised learning to label the data to detect CH or poor reception.
- We predict the signal strength in certain coordinates based on the database using SL. It is conducted to compare the signals of certain coordinates and de- tect any abnormalities.

The rest of the article is organised as follows. In Section 2, we describe the methodology to obtain the data from the simulation. We discussed several types of ML in Section 3 Then, the results are discussed in Section 4. Finally, we conclude in Section 5.

Table 1 – Simulation parameters for the deployment of a cellular communication model

Parameter description	Value
Frequency band	1.8 Ghz
Macrocell sites	3 (each site has 3 cells);
Macrocell transmission power	46 dBm
Macrocell site distance	500m
Number of femtocell blocks	2-10
Femtocell transmission power	20 dBm
Femtocell deployment ratio	0.2
Femtocell activation ratio	0.5
UE uplink power	10 dBm
UE Mobility model	Constant Position
Macrocell Bandwidth	25 MHz
Area Margin Factor	0.5

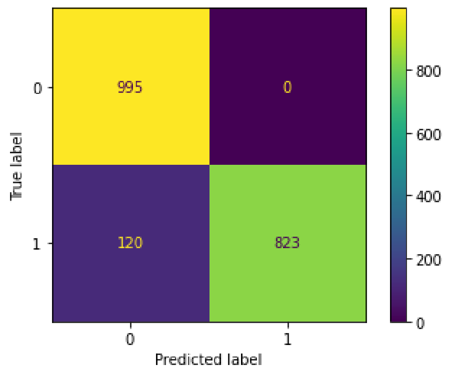


Fig. 3 – Confusion matrix to evaluate the performance of SSL.

2. OUR METHODOLOGY

To achieve the objective of this research, we needed to generate MDT data. Due to the limitations of getting the data from the operators, there were two options available in our hand. The first one is to take measurements at different geolocation sites using smartphones which is similar to a DT for the sake of research purposes. Second, to select an LTE simulator to reduce the amount of time required. In this case, the second option was chosen and NS-3 was selected out of many simulators due to its wider audience.

While deploying the LTE for generating the MDT data, we considered immediate data collection from a UE. The UE measurement was conducted in the simulator and generated a database file after plotting the LTE scenario. Three MBS were deployed where each site consisted of three cells. Some buildings were generated in simulation with femtocells placement. The femtocells deployment were based on 3GPP R4-092042 with variable density. More than one hundreds of UEs were deployed randomly with hybrid building propagation Path Loss (PL) while connecting to the nearest MBS. The architecture of LTE deployment was partially implemented from 3GPP R4-092042. Fig. 1 displays the radio environment map of the deployed LTE platform and Table 1 presents the LTE deployment parameters considered for this LTE deployment simulation.

In LTE, UE measurements are performed for MBS selection, re-selection and handover purposes. It also measures some parameters to generate an MDT database in logged mode. The parameters we considered from the NS-3 simulators to create an MDT database are discussed below:

- Reference Signal Received Power (RSRP): is the linear average of reference signal power (in Watts) measured over a particular bandwidth. This is one of the most significant measurements that a UE is required to carry out.
- Cell identity: is a unique number to identify each transceiver within a given area. Cell identities can be combined with many numbers based on antenna, location and mobile network operators.
- Latitude and longitude: are the Earth's coordinate system to locate the Earth's surface. In our simulation, we used coordinate systems such as the position of x and y.
- Timestamp: keeps track of the accurate timing of each recording. The simulator counts the number of seconds.
- Signal-to-Interference-Plus-Noise Ratio (SINR): is the strength of the desired signal compared to the interference and noise. We did not consider the measurement for our database because RSRP is one of the key elements for the UE measurement and MDT process [1, 3]. It is important to note that it was available from the NS-3 simulator.

Once the MDT database was generated, we applied our SSL algorithm to the database. The main objective here was to label all data first manually and take a small percentage of that data as labelled data. After that, the rest or the majority of the data is masked or unlabelled. With the aid of a small percentage of labelled data, we label the masked or unlabelled data. This method will help to measure the performance of the predicted labels for unlabelled data. The manually added label value is 0 or 1 where 0 means no CH and 1 means CH and the masked or unlabelled value is -1. The labelling was performed using the threshold value of RSRP -100 dBm. It needs to be noted the highest value of RSRP is -44 dBm and the lowest is -140 dBm. The cell edge RSRP is usually less than -100 dBm which was used for labelling the data [24]. By labelling the lowest RSRP, we can determine the CH. Fig. 2 is representing the flow chart of our methodology in this study to detect CHs or poor signals by labelling the data. To summarise, we needed to deploy the MBS and

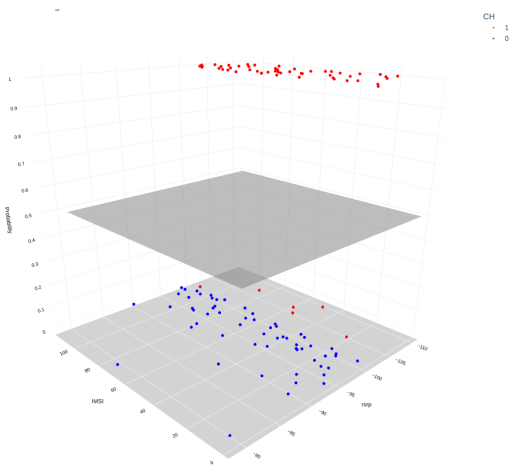


Fig. 4 – Graphical illustration of the SSL label propagation.

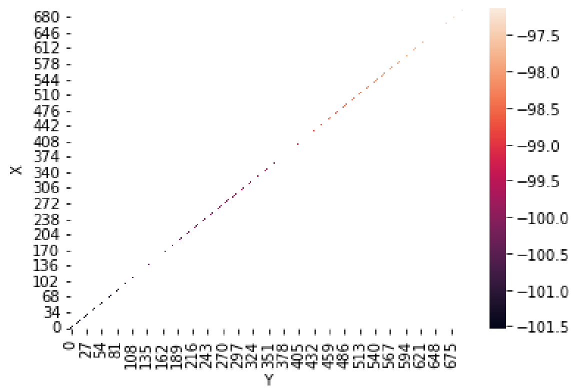


Fig. 5 – Predicting RSRP in coordinates by applying linear regression.

UE in the simulator to get measurements first. In the next step, the measurements are processed to apply SSL. Then a portion of labelled data is considered where the maximum data kept unlabelled. After this is done, the model is trained and tested. If there is any CH or poor signal then we get the coordinate of the labelled data. Following that, we predict the signal in that coordinate and compare it with the received data from the measurement in order to investigate.

3. ML FOR LABELLING THE DATA AND PREDICTION

ML is a branch of Artificial Intelligence (AI) which means the computers are able to self learn and adjust without programming explicitly [25, 26]. This self learning is based on algorithms which help to analyse statistical data, past experience, pattern recognition and computational theory [25]. ML can be categorised into four main types:

- SL: In this ML, the algorithm is trained based on labelled data to get specific output. The labelled data is taken as training data and then provided to the computing system to work as a supervisor that teaches the system to predict or classify the data. The main objective of the algorithm here is to map a function between the input variable (x) with the output variable (y) so that $y = f(x)$. From the definition, we can say that the SL can be categorised into two other groups such as classification and regression. Some real-world examples of SL can be given as weather predictions, fraud detection, email spam filtering, image classification, product or movie recommendations and so on [27].

- UL: As the name implies, this algorithm does not have a training data set to act as a supervisor which means we only have (x) and no output variables. Essentially, this algorithm helps the computing system learn from the data by analyzing its patterns, underlying structures and insights. UL is also divided into two groups: clustering and association. Some examples of the application of UL can be given as genetics, anomaly detection, data exploration, target marketing and so on [27].

- SSL: This type of ML is defined between the SL and and UL and bridges SL and UL. SSL contains large data (X) with some labelled data (Y) where the labelled data is repetitively applied to the unlabelled data. Labelling the unlabelled data is difficult, expensive and time-consuming in real-life scenarios. It is because of the lack of labelled data where unlabelled data is available [28]. SSL can be categorised in two settings: inductive and transduction learning based on the training function type [29]. Examples can be given as text document classification, speech recognition, web content classification, detecting human trafficking and so on.

- Reinforcement Learning (RL): This ML is about learning the optimal action or behaviour in an unknown environment based on a trial and error mechanism [30, 31]. The RL agent takes actions in the given environment where it receives positive and negative rewards [25]. These rewards are stored in the memory as experiences which help the agent to take optimal actions afterwards. From the discussion, we can say there are mainly two types of reinforcement learning: positive reinforcement and negative reinforcement. The applications of reinforcement learning include self-driving cars, playing games, marketing strategy, and industrial robotics, among others.

The common technique used for the SSL is label propagation. In this case, a classifier algorithm is trained with a small amount of labelled data. This data can be taken from the unlabelled data and then can be labelled according to the requirement. Generally, the best unlabelled data with their configured or predicted labels are added to the training set. Following that, the classifier is trained and the

Table 2 – Evaluation of the classification models

		Precision	Recall	F1-score	Support
Semi-supervised Learning	0	0.89	1.00	0.94	995
	1	1.00	0.87	0.93	943
	accuracy			0.94	1938
	macro avg	0.95	0.94	0.94	1938
	weighted avg	0.94	0.94	0.94	1938
Logistic Regression	0	0.86	0.82	0.84	1003
	1	0.82	0.85	0.84	935
	accuracy			0.84	1938
	macro avg	0.84	0.84	0.84	1938
	weighted avg	0.84	0.84	0.84	1938

procedure is repeated if required. Label propagation assumes that the data points with similar labels are closer together. Consequently, these class labels can be propagated through dense regions of unlabelled data. There are several steps in the algorithm that are iterative:

- Draw edges (links) between various nodes (data points) to form a connected graph.
- Weigh each edge so that stronger edges have larger weights (closer connection) and weaker edges have smaller weights (further distance). In general, larger edge weights increase the probability of labels propagating.
- Calculate a probability distribution for reaching a labelled point from each unlabelled point. In a random walk, all possible paths are explored until convergence is reached, at which point the probabilities remain the same.

The probabilities found by the above process are used to assign labels to unlabelled points [32]. Here, we pick SSL because labelled data is very expensive. The unlabelled data is vastly available in an unstructured way. Aside from that, SL will be unable to draw a decision boundary due to the lack of labelled data and UL will have reduced performance due to the absence of two defined clusters [32]. The data we receive from the MDT is mainly unlabelled data which can be used for CH detection and its location by applying SSL. Also, the SSL is superior to SL in many aspects including accuracy [33].

4. NUMERICAL RESULTS EVALUATION

In our methodology, first MDT data needs to be collected from the LTE deployment using the NS3 simulator. Then the data is processed after gathering from the simulator such as converting file formats to apply the ML algorithm. As previously discussed, the applied ML technique, SSL, can be useful for labelling the data. First, we needed to select a portion of labelled data based on the RSRP threshold value. A specific amount of data points (15%) was selected as labelled data. The rest of the unlabelled data was labelled to find the CH in a specific location by applying the above-mentioned method. As with any classification model, the *fit()* function can be used to fit the model and the *predict()* function to predict new data. The training data points provided to the *fit()* function were labelled as integer encoded (i.e., 0 or 1) whereas the unlabelled data points were marked as -1. Once the model has been fitted with kNN kernel for label propagation, a label will be assigned to the unlabelled examples. As soon as the model has been fitted, the estimated labels for the labelled and unlabelled data are available via the *"transduction_"* attribute on the *LabelPropagation* class [34]. From the simulation we conducted, we found that SSL had an accuracy rate of 94% and could be improved if the amount of labelled data was increased.

The logistic regression was also applied on the all data points where the accuracy was 84%. In both cases, the mobility model was considered as constant. Fig. 3 is the confusion matrix displaying SSL performance. In addition, Fig. 4 is displaying the illustration of the label propagation. Here, blue=0 is the true label of the no CH and red=1 is the true label of the CH. The points in the lower half of the probability axis indicate labels for predicted no CH and the points in the upper half of the probability axis indicate labels for predicted CH. That means the red points in the upper half and blue points in the lower half represent correctly recognised labels. On the other hand, the red points in the lower half represent incorrectly recognised labels due to the accuracy level of the algorithm. In the same figure, the International Mobile Subscriber Identity (IMSI) indicates the UE number at (x, y) coordinates. Furthermore, Table 2 is representing our results after applying SSL and logistic regression. In this study, one deployment scenario was sufficient to generate large quantities of data that could be used for labelling and to measure the performance of SSL if the simulation ran for a period of time. However, we also observed the performance of the SSL with different deployment scenarios by changing the UE deployment density. It appeared that the accuracy was getting reduced but stayed in an acceptable level.

In this research, we also employed linear regression of SL in order to predict the RSRP of any (x, y) coordinate.

The generated data was considered and fitted into the model. This may help to detect any abnormalities such as sleeping cells, cell outage, different CHs due to overshooting, pilot pollution or poor reception. By comparing the labelled data and predicted data, the operators may investigate the reason for CH. It needs to be noted that the CH will have a poor signal from a few UEs whereas cell outage will bring a number of UEs to send poor signal data from the same region. Fig. 5 shows the linear regression using the data for RSRP prediction in any location.

5. CONCLUSION

This paper presents an effective solution for CH detection driven from mobile network data by utilising ML algorithms such as SSL. The costly labelled data may help operators to reduce their CAPEX and OPEX. The evaluation of the proposed algorithm suggested that the SSL algorithm labelled the location of the poor signal area with higher accuracy than the classification algorithm. This paper also presents the prediction of signal strength based on mobile network data by employing a linear regression algorithm. This tool can be used to examine abnormalities by predicting the signal strength in a specific location. In future work, the prediction algorithm can be used for other research such as implementing a mobility-based solution for time critical applications or low/medium/high mobility traffic through the stronger signal areas. Also, XGBoost or better tools can be considered to identify the reason for abnormalities by predicting the signal coverage in a certain location.

ACKNOWLEDGEMENT

This publication was made possible by NPRP13S-0130-200200 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the author[s].

REFERENCES

[1] Daniel Baumann. "Minimization of drive tests (MDT) in mobile communication networks". In: *Proceeding zum Seminar Future Internet (FI) und Innovative Internet Technologien und Mobilkommunikation (IITM)*. Vol. 9. 2014, p. 7.

[2] Jakob Thrane, Matteo Artuso, Darko Zibar, and Henrik L. Christiansen. "Drive Test Minimization Using Deep Learning with Bayesian Approximation". In: *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. 2018, pp. 1–5. DOI: 10.1109/VTCFall.2018.8690911.

[3] Johan Johansson, Wuri A. Hapsari, Sean Kelley, and Gyula Bodog. "Minimization of drive tests in 3GPP release 11". In: *IEEE Communications Magazine* 50.11 (2012), pp. 36–43. DOI: 10.1109/MCOM.2012.6353680.

[4] Wuri A. Hapsari, Anil Umesh, Mikio Iwamura, Malgorzata Tomala, Bodog Gyula, and Benoist Sebire. "Minimization of drive tests solution in 3GPP". In: *IEEE Communications Magazine* 50.6 (2012), pp. 28–36. DOI: 10.1109/MCOM.2012.6211483.

[5] Seppo Hämäläinen, Henning Sanneck, and Cinzia Sartori. *LTE self-organising networks (SON): network management automation for operational efficiency*. John Wiley & Sons, 2012.

[6] Jiaying Zong, Yang Liu, Haitao Liu, Qingtian Wang, and Peng Chen. "6G Cell-Free Network Architecture". In: *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*. 2022, pp. 421–425. DOI: 10.1109/ICETCI55101.2022.9832308.

[7] Latif U. Khan, Ibrar Yaqoob, Muhammad Imran, Zhu Han, and Choong Seon Hong. "6G Wireless Systems: A Vision, Architectural Elements, and Future Directions". In: *IEEE Access* 8 (2020), pp. 147029–147044. DOI: 10.1109/ACCESS.2020.3015289.

[8] Ana Gómez-Andrades, Raquel Barco, Pablo Muñoz, and Inmaculada Serrano. "Data Analytics for Diagnosing the RF Condition in Self-Organizing Networks". In: *IEEE Transactions on Mobile Computing* 16.6 (2017), pp. 1587–1600. DOI: 10.1109/TMC.2016.2601919.

[9] Ana Gómez-Andrades, Raquel Barco, and Inmaculada Serrano. "A method of assessment of LTE coverage holes". In: *EURASIP Journal on Wireless Communications and Networking* 2016.1 (2016), pp. 1–12.

[10] Ömer Faruk Çelebi, Engin Zeydan, Ömer Faruk Kurt, Ömer Dedeoğlu, Ömer İieri, Burak AykutSungur, Ahmet Akan, and Salih Ergüt. "On use of big data for enhancing network coverage analysis". In: *ICT 2013*. 2013, pp. 1–5. DOI: 10.1109/ICTEL.2013.6632155.

[11] Murong Lin, Quanmin Ye, and Yinghua Ye. "Graph theory based mobile network insight analysis framework". In: *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE. 2016, pp. 1–7.

[12] A. Galindo-Serrano, B. Sayrac, S. Jemaa, J. Riihijärvi, and P. Mähönen. "Automated coverage hole detection for cellular networks using radio environment maps". In: *Proc. IEEE 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*. Tsukuba Science City, Japan, 13-17 May 2013.

[13] Ana Galindo-Serrano, Berna Sayrac, Sana Ben Jemaa, Janne Riihijärvi, and Petri Mähönen. "Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks". In: *8th International Conference on Cognitive Radio Oriented Wireless Networks*. 2013, pp. 37–42. DOI: 10.1109/CROWNCom.2013.6636791.

[14] Vassilis Dalakas. "Automate minimization of drive tests for qoe provisioning: The case of coverage mapping". In: *International Journal of Computer Applications* 975 (2015), p. 8887.

[15] Fedor Chernogorov and Timo Nihtila. "QoS Verification for Minimization of Drive Tests in LTE Networks". In: *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*. 2012, pp. 1–5. DOI: 10.1109/VETECS.2012.6240228.

[16] Fedor Chernogorov and Jani Puttonen. "User satisfaction classification for Minimization of Drive Tests QoS verification". In: *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 2013, pp. 2165–2169. DOI: 10.1109/PIMRC.2013.6666502.

[17] Jani Puttonen, Jussi Turkka, Olli Alanen, and Janne Kurjenniemi. "Coverage optimization for Minimization of Drive Tests in LTE with extended RLF reporting". In: *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. 2010, pp. 1764–1768. DOI: 10.1109/PIMRC.2010.5671907.

[18] Bilal Hussain, Qinghe Du, and Pinyi Ren. "Big data-driven anomaly detection in cellular networks". In: *2017 IEEE/CIC International Conference on Communications in China (ICCC)*. 2017, pp. 1–6. DOI: 10.1109/ICCChina.2017.8330468.

[19] Ahmed Zoha, Arsalan Saeed, Ali Imran, Muhammad Ali Imran, and Adnan Abu-Dayya. "A SON solution for sleeping cell detection using low-dimensional embedding of MDT measurements". In: *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. 2014, pp. 1626–1630. DOI: 10.1109/PIMRC.2014.7136428.

[20] Fedor Chernogorov, Jussi Turkka, Tapani Ristaniemi, and Amir Averbuch. "Detection of Sleeping Cells in LTE Networks Using Diffusion Maps". In: *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*. 2011, pp. 1–5. DOI: 10.1109/VETECS.2011.5956626.

[21] Jussi Turkka, Fedor Chernogorov, Kimmo Brigatti, Tapani Ristaniemi, and Jukka Lempiäinen. "An approach for network outage detection from drive-testing databases". In: *Journal of Computer Networks and Communications* 2012 (2012).

[22] Sergey Chernov, Michael Cochez, and Tapani Ristaniemi. "Anomaly Detection Algorithms for the Sleeping Cell Detection in LTE Networks". In: *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. 2015, pp. 1–5. DOI: 10.1109/VTCSpring.2015.7145707.

[23] Bilal Hussain, Qinghe Du, and Pinyi Ren. "Semi-supervised learning based big data-driven anomaly detection in mobile wireless networks". In: *China Communications* 15.4 (2018), pp. 41–57. DOI: 10.1109/CC.2018.8357700.

[24] CableFree Admin. *LTE RSSI, RSRP and RSRQ measurement*. Jan. 2018. URL: https://www.cablefree.net/wirelesstechnology/4glte/rsrp-rsrq-measurement-lte/.

[25] Shahriar Al-Ahmed, Muhammad Zeeshan Shakir, and Qasim Ahmed. "Deep Q-Network based coverage hole detection for future wireless networks". In: *AI for Emerging Verticals: Human-robot computing, sensing and networking*. IET, 2020, pp. 173–188.

[26] Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, and Amit P Sheth. "Machine learning for Internet of Things data analysis: A survey". In: *Digital Communications and Networks* 4.3 (2018), pp. 161–175.

[27] Wei-Lun Chao. "Machine learning tutorial". In: *Digital Image and Signal Processing* (2011).

[28] Georgios Kostopoulos, Stamatis Karlos, Sotiris Kotsiantis, and Omiros Ragos. "Semi-supervised regression: A recent review". In: *Journal of Intelligent & Fuzzy Systems* 35.2 (2018), pp. 1483–1500.

[29] Xiaojin Zhu and Andrew B Goldberg. "Introduction to semi-supervised learning". In: *Synthesis lectures on artificial intelligence and machine learning* 3.1 (2009), pp. 1–130.

[30] Shahriar Abdullah Al-Ahmed, Muhammad Zeeshan Shakir, and Syed Ali Raza Zaidi. "Optimal 3D UAV base station placement by considering autonomous coverage hole detection, wireless backhaul and user demand". In: *Journal of Communications and Networks* 22.6 (2020), pp. 467–475. DOI: 10.23919/JCN.2020.000034.

[31] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

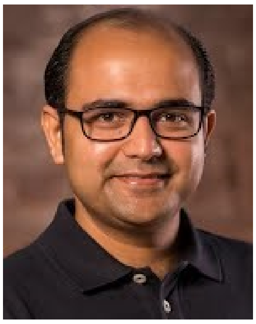
[32] Saul Dobilas. *Semi-supervised learning-how to assign labels with label Propagation Algorithm*. Feb. 2022. URL: https://towardsdatascience.com/semi-supervised-learning-how-to-assign-labels-with-label-propagation-algorithm-9f1683f4d0eb.

[33] Aarti Singh, Robert Nowak, and Jerry Zhu. "Unlabeled data: Now it helps, now it doesn't". In: *Advances in neural information processing systems* 21 (2008).

[34] Jason Brownlee. *Semi-supervised learning with label Propagation*. Dec. 2020. URL: https://machinelearningmastery.com/semi-supervised-learning-with-label-propagation/#:~:text=Semi%2Dsupervised%20learning%20refers%20to,learn%20from%20labeled%20training%20data..

AUTHORS

Shahriar Abdullah Al-Ahmed received a B.Eng. degree in computing and electronics from the University of East London, London, U.K., in 2008 and an M.Sc. degree in telecommunication (network pathway) from Queen Mary University of London, London, U.K. in 2012. Currently, he is working towards a Ph.D. degree in engineering and computing with the University of the West of Scotland, Paisley, Scotland, U.K. Before joining the Ph.D. degree, he was involved in teaching computer science in secondary educational institutes in the U.K. His research interests include artificial intelligence, Internet of Things, and wireless communication.



Muhammad Zeeshan Shakir is Professor of Wireless Communications at the University of the West of Scotland (UWS), UK and has received over £3m research funding from Innovate UK, ERASMUS, QNRF and Scottish Govt.

With over 15 years of research expertise in design and development of digital technologies he has published over 150 research articles and contributed to 10 books. Dr. Shakir is a recipient of numerous research awards/prizes including Scottish Association of Minority Ethnic Educators (SAMEE) STEM Inspiring Diversity and Inclusion award 2022 for successful delivery of CEMVO and Scottish Govt. funded project ”Stay Safe Scotland”, SICSA (The Scottish Informatics and Computer Science Alliance) Best Poster Award 2022, IEEE Communications Society Fred W. Ellersick Award 2021, IEEE Communications Society and China Institute of Communications Best Journal Article Award 2019 and UWS STARS Award 2020 & 2018 for outstanding research, teaching and enterprise performances.

Dr. Shakir is a fellow of Higher Education Academy, UK, senior member of IEEE, and an active member of IEEE Communications Society. He is a member of the Scottish Parliament Framework Agreement COVID-19 Pandemic to provide expert advice on the impact of COVID-19 on digital infrastructure, education, and healthcare. He is a member of the Royal Society of Edinburgh Young Academy of Scotland through a countrywide competitive selection process for building artificial intelligence capacity across Scotland.

Dr. Shakir has been serving as a chair and as part of the organising committee of several technical programs symposiums/workshops in IEEE flagship conferences, including Globecom, ICC and WCNC. He has been a frequent keynote speaker/tutorial speaker at IEEE flagship conferences such as IEEE Globecom and ICC and international events. He is a chair of IEEE Communications Society emerging technologies committee on backhaul/fronthaul, and Public Safety Technology Committee Informatics.