

# Machine Learning Based Analysis of Cellular Spectrum

Muhammad Yasir<sup>1</sup>, Zafi Sherhan Shah<sup>1</sup>, Sajjad Ali Memon<sup>1</sup>, Zahid Ali<sup>1,2</sup>

<sup>1</sup>IICT, Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan

<sup>2</sup>Chongqing University, China

E-mail address: yasirshaikh671@gmail.com, zafisherhan.shah@faculty.muett.edu.pk,  
sajjad.memon@faculty.muett.edu.pk, za4zahidali@gmail.com

Received: 15 January 2021; Revised: 02 February 2021; Accepted: 01 March 2021; Published: 08 April 2021

**Abstract:** One of the key issues of wireless communication networks is the spectrum crisis, and studies noted that static licensed bands are in the under-utilization stage. Recently Cognitive Radio Network facilitates a solution to minimize the spectrum crisis in which unlicensed users can utilize the licensed spectrum without transmission interference. To achieve this task we used Machine Learning techniques for analyzing spectrum occupancy which is an efficient method to analyze spectrum occupancy and provides high accuracy. Supervised machine learning algorithms namely Logistic regression, K nearest neighbor, and Naive Bayes are used to classify a given frequency band. In this paper we collect spectrum samples of GSM 900, 1800, and 2100 bands using Rohde & Schwarz FSH6 Handheld Spectrum Analyzer for developing a dataset, using that dataset we trained the classifiers and analyze their classification performance accuracy. Results have shown the best performance on the validation and testing partition for various Unweighted Average Recall (UAR) of each classifier. Here the Logistic Regression classifier learns the best representation from their feature vector. This research is helpful to measure the spectrum occupancy of different static allocated licensed bands for 24/7. This will give better ideas about spectrum utilization, future spectrum allocation and comfort to serve more users in the limited spectrum. The occupancy measurements of current allocated spectrums can not only provide a convincing basis for making future spectrum allocation policies, but also provide technical support for the development of new communication technologies.

**Index Terms:** Cognitive radio, Spectrum occupancy, Machine learning analysis, Cellular network, Spectrum analyzer.

## 1. Introduction

A cognitive radio network (CRN) consists of two groups of users, first the licensed users (PU's) and second unlicensed users (SU's). The main aim behind CR is to allow unlicensed users to access the licensed bands without interfering with licensed user interference. To achieve this task, a realistic understanding of the dynamic usage of the spectrum is needed. According to the study of the Federal Communications Commission (FCC) that spectrum wastage is from 15% to 85% for frequency spectrum band of 0-6 GHz [1]. In several countries, the ability of CR to maximize spectrum usage is -recognized and policies are being formulated to access the fallow spectrum bands. Several spectral surveys were performed in various parts of the world to survey different spectrum bands [2, 9]. The average power spectral density levels of received signals and bandwidth used over time in various spectrum bands are appeared by these spectral survey reports. In [10] measurement results show that most of the allocated spectrum is utilized with low efficiency, Improvement in the low spectrum occupancy band can be occurred by emerging dynamic access technology. In [11], analyze the usage of spectrum in a practical traffic scenario through various cellular technologies such as GSM, CDMA, UMTS, and WiMAX. 60% of the persecuted simulation research studies remain unused for 24 hours. In [12], GSM band percentages of spectrum occupancy were determined by RF Explorer 6G Combo spectrum Analyzer for 8 different thresholds from -40 dBm to -75 dBm. The spectrum occupancy models were tested by all the foregoing studies using traditional probabilistic or statistical methods and techniques, but these methods are usually limited due to assumptions needed to solve their theories.

In this work, we analyze spectrum occupancy in three transmission bands that are GSM 900, 1800, and 2100 in order to determine unused spectrum within the three bands in indoor and outdoor locations using the Rohde & Schwarz FSH6 spectrum analyzer. Using MATLAB, the measured results of all three bands were determined against the power spectral density versus frequency plots and show that there exists a lot amount of unused spectrum in a licensed band [13]. To find spectrum occupancy using MATLAB is not also an efficient solution because using MATLAB one

dedicated person required analyzing spectrum occupancy and 24 hours analysis is also very difficult. Manual analyzes of spectrum occupancy are time-consuming. Spectrum scarcity is a major concern to wireless communication.

### 1.1 Aims and Objectives

Aim of this research is to develop ML algorithms to analyze spectrum occupancy for 24 hours.

- Collect the samples of GSM 900, 1800 and 2100 frequency band.
- Perform the feature's augmentation of data.
- Labelled the data named as "used" and "unused" with respect to threshold level of power through power loss model.
- Trained logistic regression, K nearest neighbor and Nave Baye's classifiers by using given dataset to achieve desired results.

### 1.2 Problem Statement

As studied in literature [1, 13], researchers analyzed spectrum occupancy using conventional methods but current technology is moving for autonomous examination therefore conventional method is not competent solution to analyze occupancy of spectrum even cannot process continuously for 24 hours.

### 1.3 Limitation

In case study [1, 13], authors evaluated the spectrum occupancy model using traditional statistical or probabilistic tools but they have some limitations because these tools require assumption to drive their theories either value is a random variable or a randomly process.

## 2. Related Work

In [14,18], there are very few studies on the use of ML technique to find out spectrum occupancy, In this paper researchers work on the development of ML-based system to analyze the spectrum occupancy. In paper [18] studied different ML algorithm to analyze spectrum occupancy and shows that Machine learning (ML) is a very strong technique which has recently gained growing attention. As a result, they have greater precision than traditional probabilistic and statistical instruments in many stages. After studied paper [17,18] we conclude that Auto analyze spectrum occupancy for 24 hours is possible. Above ML literature studied different algorithms for classification accuracy of classifiers, but In this paper, we used Logistic Regression, KNN, and Naive Bayes classifiers for analyzing spectrum occupancy in a transmission band. In [19], Logistic regression (LR) keeps up with one of the most broadly used methods in data mining for binary data classification. This research focuses on providing an overview of the most relevant aspects of LR as used for data analysis, especially for algorithmic and machine learning point of view; also consider how LR can be applied to imbalanced and unusual data events.

This paper provides an overview of some of the algorithms and corrections that enable LR from a machine learning point of view to be both fast and accurate. In [20], KNN learning-based classification technique is implemented for cooperative spectrum sensing (CSS), KNN required a very small amount of time for training the classifiers. In [21], this paper tests cognitive radio network (CRN) spectrum occupancy based on the naive Bayesian classifier (NBC), The motive for this work is the problem of classification in spectrum sensing, where secondary users (SUs) must sense and use the free channel for their purpose of transmission/reception. NBC is considering all features independently and give a good model for classification in the future that's why it is considered.

## 3. Measurement System and Method

We developed ML algorithm to analyze spectrum occupancy efficiently because ML is power full tool no assumption required to derive their theories as needed in conventional method. ML based method provide higher accuracy than conventional method.

### 3.1 Measurement System

By the use of Rohde & Schwarz FSH6 handheld spectrum analyzer, the spectrum occupancy analysis has been a practice in the Qasimabad area of Hyderabad city which is located in Pakistan. We performed this activity in an indoor environment. This spectrum analyzer measure frequency in the range of 100 kHz to 6 GHz. GSM antenna is associate with a spectrum analyzer via fiber optical cable, the antenna specification is given in Table 1. The laptop is also associate with spectrum analyzer via USB optical cable as shown in Fig. 1, the spectrum band analyzing features is recorded in a Microsoft Excel sheet and developed a dataset, initially, this dataset has raw data and irrelevant features exists in the dataset, In the data cleaning and labeling step where we manually remove the irrelevant feature and labeled to the relevant features and keep meaningful features in the dataset. The accuracy of ML algorithms also depends on the number of features that's why we used the Okumura-Hata, Cost 321-Hata, and Cost 321-Walfisch Ikegami LOS power

loss models to increase features in the dataset. The training partition comprised 60% of the examples in the dataset whereas validation and test partitions each had 20% of the dataset. We experiment with our selected supervised ML algorithms using 60% of the dataset after train the dataset we analyze validation and test prediction of classifiers with 40% of the dataset and found desired results. The flow chart for research methodology is given below in Fig. 5.



Fig.1. Spectrum Occupancy Measurement System

Table 1. Description of Antenna [1].

The Electrical description	
BW in MHz	136
Frequency in MHz	824 – 960
Gain in dBi	3
Impedance	50
Voltage Standing Wave Ratio (VSWR)	2
Input Power (Max)	50
Polarization	Vertical
Input connector type	SMA male
Antenna model	AMXT-900-3
The Mechanical description	
Height of Antenna (mm)	210
temperature ( °C)	-40 to 60
Radome Color	Black

### 3.2 Cellular Service Providers in Pakistan

In Pakistan, four service providers are working such as Jazz, Ufone, Zong, and Telenor. These operators are using techniques such as LTE, WCDMA and GSM to provide a high quality of data and voice services to subscribers. Frequency allocation chart of GSM 900, 1800 and 2100 band which are operating in Pakistan is available on Pakistan Telecommunication Authority website we attached below in Fig. 2, 3 and 4.

#### 850 & 900 MHz

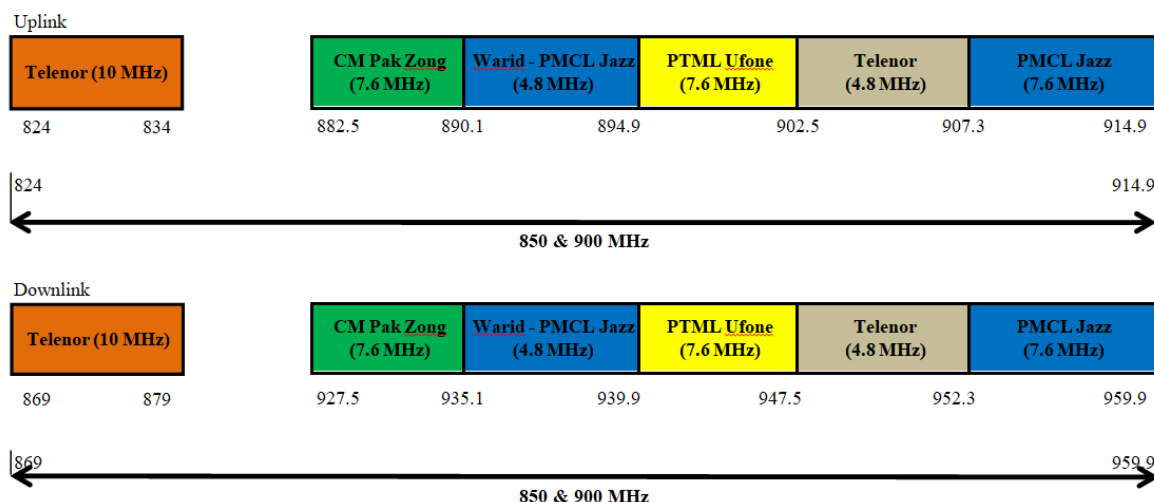


Fig.2. Frequency GSM allocation band in Pakistan

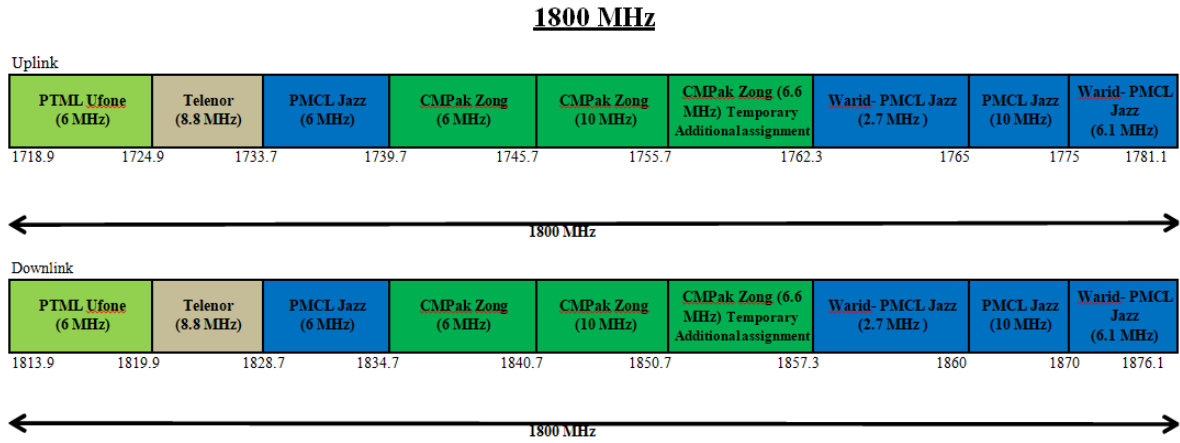


Fig.3. Frequency Data allocation band in Pakistan

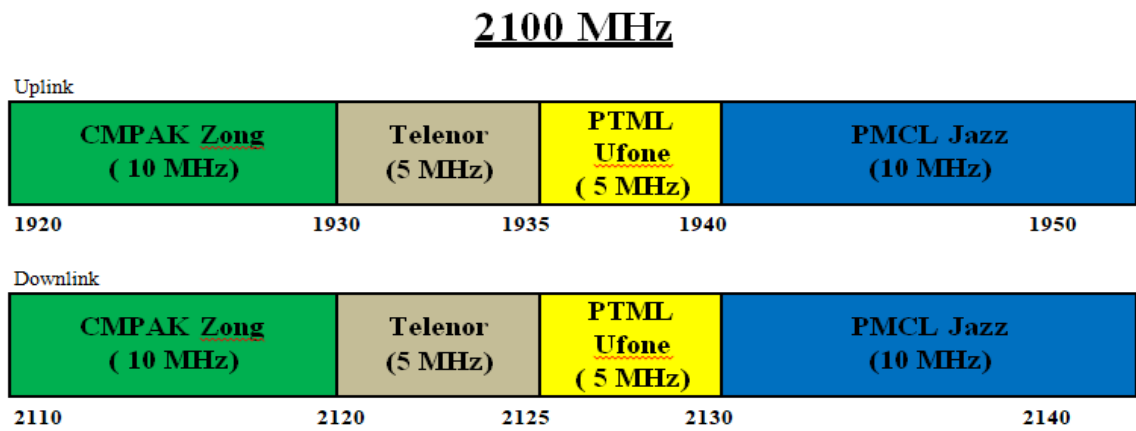


Fig.4. Frequency Data allocation band in Pakistan

Table 2. The SA settings for GSM 900, DCS 1800 and 2100 bands and GSM [1].

Pakistan Service Providers(SA Parameters)	GSM	Range of frequency	Central Freq in MHZ	Freq. Span	Resolution in kHz	Frequency measuring point
Ufone		Uplink 894.9-902.5 Downlink 939.9-947.5	Uplink 898.7 Downlink 943.7	Uplink 7.6 Downlink 7.6	200	301
Mobilink		Uplink 907.3- 914.9 Downlink 952.3- 959.9	Uplink 911.1 Downlink 956.1	Uplink 7.6 Downlink 7.6	200	301
Telenor		Uplink 902.5- 907.3 Downlink 947.5- 952.3	Uplink 904.9 Downlink 949.9	Uplink 4.8 Downlink 4.8	200	301
Zong		Uplink 882.5- 890.1 Downlink 927.5- 935.1	Uplink 886.3 Downlink 931.3	Uplink 7.6 Downlink 7.6	200	301
Warid		Uplink 890.1- 894.9 Downlink 935.1- 939.9	Uplink 892.5 Downlink 937.5	Uplink 4.8 Downlink 4.8	200	301
GSM 900		Uplink 890-915 Downlink 935- 960	Uplink 902.5 Downlink 947.5	Uplink 25 Downlink 25	200	301
DCS 1800		Uplink 1718.9-1781.1 Downlink 1813.9-1876.1	Uplink 1750 Downlink 1845	Uplink 62.2 Downlink 62.2	200	301
2100		Uplink 1920-1950 Downlink 2110-2140	Uplink 1935 Downlink 2125	Uplink 30 Downlink 30	200	301

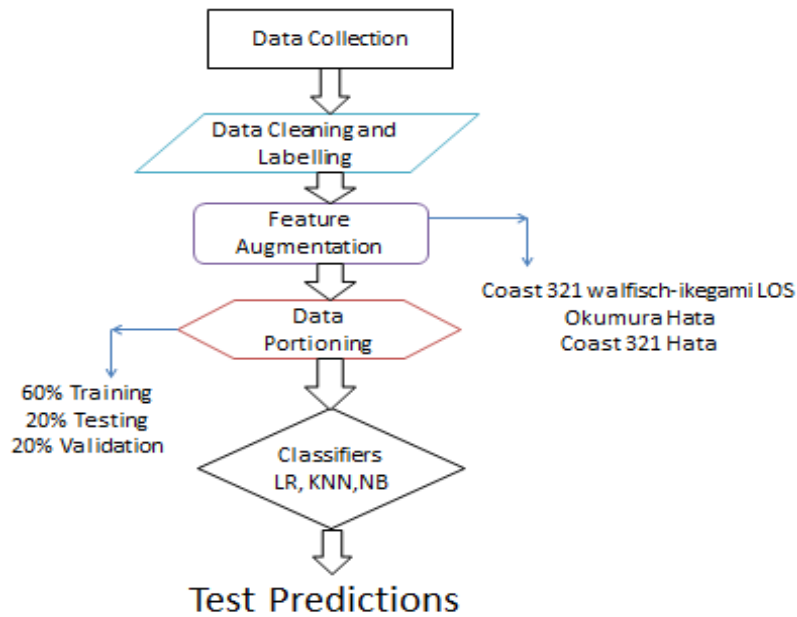


Fig.5. Flow chart of Research Methodology

### 3.3 Cleaning Data and Labelling

Using Rohde Schwarz FSH6 handheld spectrum analyzer we collect data in the indoor environment of Qasimabad, Pakistan. Initially, the collected data was raw data, we manually remove irrelevant features to make it meaningful data for ML to achieve nearest to desired results. We labeled the dataset for training the model to make the correct decision. In labeling, if power above the threshold level is labeled as used if power below the threshold it is unused.

### 3.4 Feature Augmentation

To train the ML algorithms we need more and more feature in a dataset for achieving accuracy, But the Rohde & Schwarz FSH6 spectrum analyzer just measures frequency, power, and amplitude, so using Okumura Hata, Cost 321 Hata, and Cost 321 Walfisch Ikegami LOS power loss model we calculate power loss of frequency spectrum GSM 900, 1800, and 2100 at the distance of 1m, 500m and 1km to increase more features in a dataset because the accuracy of the model depends on bulk amount of meaningful features. In Okumura Hata and Cost 321 Hata we put area value 4 in the formula for the city unless Cost 321 Walfisch Ikegami LOS does not depend on area value. In the power loss model, we analyze that power increase with an increase in distance.

Table 3. Power loss measurement using models

Okumura Hata	Cost 321 Hata	Cost 321 Walfisch Ikegami LOS
Power loss Higher than Cost 321 Hata and Cost 321 Walfisch Ikegami LOS	Power loss Higher than Cost 321 Walfisch Ikegami LOS but lower than Okumura Hata	Power loss Lower than Okumura Hata and Cost 321 Hata

In addition to the measurements collected from Rohde & Schwarz FSH6 spectrum analyzer, we used four types of propagation loss models to augment the feature vector. These models include the Okumura-Hata, Cost 321-Hata, and Cost 321-Walfisch Ikegami LOS power loss models. Here, we computed the power loss using these models for frequencies of 900 MHz, 1800 MHz, and 2100 MHz in the GSM/DCS band at distances of 1 m, 500 m, and 1000 m from the transmitter. We modeled propagation for rural, suburban, or urban environment depending on where spectrum measurements were collected. The resultant feature vector for each sample of measurements has a dimensionality of 15 and carries a binary label of either “used” or “unused”.

### 3.5 Data Portioning

In the data portioning step we kept 60% data for train the model, 20% for testing the model, and 20% to check the model validation. We trained to the Logistic regression, KNN, and Naive Bayes classifiers using that 60% data of dataset, after developing model 40% data used for testing and validation purpose and found pleasant results.

### 3.6 Machine Learning

In pursuit of automated identification on unused spectral bands, we experimented with three types of machine algorithms, namely logistic regression, k-nearest neighbors, and Naïve Bayes, to classify a given frequency band of “used” or “unused”. To train and evaluate the performance of these classifiers for the task at hand, we partitioned the dataset into training, validation, and test partitions. The training partition comprised 60% of the examples in the dataset whereas validation and test partitions each had 20% of the dataset. To ensure reproducibility of partitions across the three classifiers, we set the seed to a fixed predefined value. We note that the dataset was imbalanced with 80% of examples belonging to the “used” class whereas a minority 20% of examples belong to the “unused” class. Due to this reason, the performance of classifiers was measured using the Unweighted Average Recall (UAR), a parameter that is commonly used as a classification metric for imbalanced classification tasks.

## 4. Results and Discussion

In Table 4, we present a summary of results for the classification performance of logistic regression classification for various hyper-parameters. Here, we note that the best performance on the validation partition in terms of UAR is 94.93% which is achieved with the value of the regularization constant at  $10^5$ . This model achieves a UAR of 95.80% on the test partition.

Table 4. Summary of classification performance for Logistic Regression classifier of various hyper-parameters.

Regularization (C)	Unweighted Average Recall (%)	
	Val. Partition	Test Partition
$10^{-6}$	66.98	67.71
$10^{-5}$	87.47	88.18
$10^{-4}$	87.62	88.33
$10^{-3}$	89.80	89.84
$10^{-2}$	92.04	92.93
$10^{-1}$	92.27	94.02
$10^0$	92.67	94.36
$10^1$	92.76	94.44
$10^2$	93.51	93.96
$10^3$	93.38	94.27
$10^4$	94.00	94.73
$10^5$	<b>94.93</b>	<b>95.80</b>
$10^6$	94.64	96.04

A summary of the classification performance of the K-NN classifier with various hyper-parameters is provided in Table 5. Here, we note that the best performing model achieves a UAR of 78.11% on the validation partition which translates into a UAR of 82.00% for the test partition.

Table 5. Summary of classification performance for k-NN classifier for various hyper-parameters.

Num. Clusters	Unweighted Average Recall (%)	
	Val. Partition	Test Partition
2	72.02	79.20
4	77.49	79.71
8	77.02	81.44
16	<b>78.11</b>	<b>82.00</b>
32	75.36	79.24
64	73.82	75.31

Given that the Naïve Bayes’ classifier does not contain a hyper-parameter, we report that it achieved a UAR of 88.11% for the validation and 92.02% for the test partition. The overall summary of the best performing models is provided in Table 6. Here, we note that the Logistic Regression classifier learns the best representation from our feature vector which enables it to achieve an accuracy of 95.80% for the test partition. This is followed by the Naïve Bayes’ classifier with 92.02% whereas the K-NN classifier achieves a relatively poorer accuracy of 82.00% for the test partition.

Table 6. Summary of classification results.

Classifier	Unweighted Average Recall (%)	
	Val. Partition	Test Partition
Logistic Regression	94.93	95.80
K-NN	78.11	82
Naïve Bayes'	88.11	92.02

## 5. Conclusion

This study analyzed that when we train to the classifiers with collected samples of GSM 900, 1800, 2100 frequency band which have 80% used and 20% unused of frequency band they give the performance for validation and testing accuracy for various UAR. We analyzed that best performing model which is provided in Table 4, the Logistic Regression classifier learns the best representation from our feature vector which enables it to achieve an accuracy of 95.80% for the test partition. This is followed by the Naïve Bayes's classifier with 92.02% whereas the K-NN classifier achieves a relatively poorer accuracy of 82.00% for the test partition.

## 6. Future Work

As we saw that Logistic regression, K nearest neighbor and Nave Bays classifiers evaluate the classification performance accuracy for collected data of indoor environment and analyzed the desired results, In future we will conduct this type of activity for an outdoor environment and will analyze the classification performance accuracy of classifiers.

## References

- [1] S. Ali, Z. Chen, and F. Yin, "Spectrum Occupancy of Cellular Networks in Pakistan for Cognitive Radio — Measurements using Spectrum Analyzer," *Int. J. Inf. Electron. Eng.*, vol. 6, no. 1, pp. 26–31, 2015.
- [2] M. J. Marcus, "Spectrum policy for radio spectrum access," *Proc. IEEE*, vol. 100, no. SPL CONTENT, pp. 1685–1691, 2012.
- [3] V. Valenta, Z. Fedra, R. Maršálek, G. Baudoin, and M. Villegas, "Analysis of spectrum utilization in suburb environment - Evaluation of potentials for cognitive radio," *2009 Int. Conf. Ultra Mod. Telecommun. Work.*, no. 2008, pp. 0–5, 2009.
- [4] M. Lopez-Benitez and F. Casadevall, "Space-Dimension Models of Spectrum Usage for Cognitive Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 306–320, 2017.
- [5] M. Wellens and P. Mähönen, "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," *Mob. Networks Appl.*, vol. 15, no. 3, pp. 461–474, 2010.
- [6] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," *IEEE Veh. Technol. Conf.*, pp. 3016–3020, 2007.
- [7] S. W. Ellingson, "Spectral occupancy at VHF: Implications for frequency-agile cognitive radios," *IEEE Veh. Technol. Conf.*, vol. 2, pp. 1379–1382, 2005.
- [8] J. Do, D. M. Akos, and P. K. Enge, "L and S bands spectrum survey in the San Francisco Bay area," *Rec. - IEEE PLANS, Position Locat. Navig. Symp.*, pp. 566–572, 2004.
- [9] F. H. Sanders, "Broadband spectrum surveys in Denver, CO, San Diego, CA and Los Angeles, CA: Methodology, analysis, and comparative results," *IEEE Int. Symp. Electromagn. Compact.*, vol. 2, pp. 988–993, 1998.
- [10] J. Xue, Z. Feng, and P. Zhang, "Spectrum Occupancy Measurements and Analysis in Beijing," *IERI Procedia*, vol. 4, pp. 295–302, 2013.
- [11] P. K. Dalela, A. Nayak, V. Tyagi, and K. Sridhara, "Analysis of spectrum utilization for existing cellular technologies in context to cognitive radio," *2011 2nd Int. Conf. Comput. Commun. Technol. ICCCT-2011*, pp. 585–588, 2011.
- [12] B. K. Engiz and Y. A. Rajab, "Investigation of spectrum occupancy in GSM band in Samsun, Turkey," *Proc. - 2019 6th Int. Conf. Electr. Electron. Eng. ICEEE 2019*, pp. 158–161, 2019.
- [13] A. A. Mirani, S. A. Memon, S. Hussain, M. A. Panhwar, and S. R. A. Shah, "Spectrum occupancy measurement of the cellular spectrum and smart network sharing in Pakistan," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 232–243, 2020.
- [14] F. Azmat, Y. Chen, and N. Stocks, "Analysis of spectrum occupancy using machine learning algorithms," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 6853–6860, 2016.
- [15] A. Agarwal, A. S. Sengar, and R. Gangopadhyay, "Spectrum Occupancy Prediction for Realistic Traffic Scenarios: Time Series versus Learning-Based Models," *J. Commun. Inf. Networks*, vol. 3, no. 2, pp. 44–51, 2018.
- [16] H. Anandakumar and K. Umamaheswari, "Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers," *Cluster Comput.*, vol. 20, no. 2, pp. 1505–1515, 2017.
- [17] A. Agarwal, S. Dubey, M. A. Khan, R. Gangopadhyay, and S. Debnath, "Learning-based primary user activity prediction in cognitive radio networks for efficient dynamic spectrum access," *2016 Int. Conf. Signal Process. Commun. SPCOM 2016*, 2016.
- [18] S. T. Gul, A. U. Omer, and A. Majid, "Machine learning techniques for wideband spectrum sensing in cognitive radio networks," *Networks Futur. Archit. Technol. Implementations*, vol. 31, no. 11, pp. 43–68, 2017.

- [19] M. Maalouf, "Logistic regression in data analysis: An overview," *Int. J. Data Anal. Tech. Strategy.*, vol. 3, no. 3, pp. 281–299, 2011.
- [20] M. Wasilewska and H. Bogucka, "Machine learning for LTE energy detection performance improvement," *Sensors (Switzerland)*, vol. 19, no. 19, 2019.
- [21] A. K. Mishra and P. Vijaykumar, "Analysis of spectrum occupancy using naïve Bayesian classifier," *Int. J. Recent Technol. Eng.*, vol. 7, no. 4, pp. 115–117, 2018.

### Authors' Profiles



**Muhammad Yasir**- received his B.E degree in Telecommunication Engineering from Dawood UET Karachi in 2015, currently enrolled in Master Communication System & Networks at Mehran UET Jamshoro, Sindh Pakistan. He has experience working as an RF DT ENGINEER at NETKOM Technologies Pvt. Ltd, Karachi in the duration of 2016-2018. His research interest in Cognitive radio networks and machine learning.



**Zafi Sherhan Shah** works as an Assistant Professor in the Department of Telecommunication Engineering at Mehran University, Jamshoro, Pakistan. He completed his bachelors from Mehran University, Masters from University of Strathclyde, UK, and PhD from Cardiff University, UK. He has previously worked as Secretary of IEEE Communications Society, Karachi Chapter.



**Sajjad Ali Memon** received the B.E. degree in telecommunication engineering and the M.E. degree in communication systems and networks from Mehran University of Engineering and Technology, Jamshoro, Pakistan, in 2007 and 2011, respectively. He has done his Ph.D. from Dalian University of Technology, Dalian, China, in 2017. He joined Telenor Pakistan as an Operation and Maintenance Engineer, in 2007. He then joined the Department of Telecommunication Engineering, MUET, as a Lab Lecturer, in 2008, became an Assistant Professor in 2011, and became an Associate Professor in 2018. His current research interests include machine learning, Massive MIMO and wireless communication systems.



**Zahid Ali** received his B.E degree in Telecommunication Engineering from Dawood UET Karachi and M.E in Communication Systems and Networks from Mehran UET Jamshoro, Pakistan in 2016 and 2020 respectively, currently enrolled in Ph.D. from Chongqing University China. He has various experiences relevant to his field working as Research Associate (from 2019) and Research Assistant (from 2018) in Mehran UET Jamshoro and 2018-2019 worked at Basecamp Data Solution in Multinational Company of Ohio, USA as a position of Business Process Associate (IT-Support) to support cooperate clients (Verizon stores of I-Phone) located in all states of America. His research interest in Microwave, Modern Antennas, Communication Systems and Networks.

**How to cite this paper:** Muhammad Yasir, Zafi Sherhan Shah, Sajjad Ali Memon, Zahid Ali, " Machine Learning Based Analysis of Cellular Spectrum", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.11, No.2, pp. 24-31, 2021.DOI: 10.5815/ijwmt.2021.02.03