IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 June 2022, accepted 16 July 2022, date of publication 17 August 2022, date of current version 25 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3199409

RESEARCH ARTICLE

Identification and Analysis of a Unique Cell Selection Phenomenon in Public Unlicensed Cellular Networks Through Machine Learning

SRIKANT MANAS KALA[1], (Graduate Student Member, IEEE), VANLIN SATHYA[2], KUNAL DAHIYA[3], TERUO HIGASHINO[4], (Senior Member, IEEE), AND HIROZUMI YAMAGUCHI[1], (Member, IEEE)

[1]Mobile Computing Laboratory, Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan
[2]The University of Chicago, Chicago, IL 60637, USA
[3]Indian Institute of Technology Delhi, Delhi 110016, India
[4]Kyoto Tachibana University, Kyoto 607-8175, Japan

Corresponding author: Srikant Manas Kala (manas_kala@ist.osaka-u.ac.jp)

This work was supported by the National Institute of Information and Communications Technology (NICT), Japan, under Grant 222C03.

ABSTRACT Cellular operators deploy 4G License Assisted Access (LAA) and 5G NR-U base stations in the unlicensed spectrum to enhance overall network capacity. This work highlights a unique phenomenon related to Physical Cell Id (PCI) that is observed in public LAA operator deployments. Notably, the licensed and unlicensed carriers of a device may have the same PCI or different PCIs. The phenomenon is triggered by the combined effect of unlicensed deployment architectures and cell selection mechanisms. Consequently, the phenomenon will intensify in the 5G NR-U, whose public deployment will soon begin. Unfortunately, the impact of this phenomenon on coexistence network performance is unexplored. It is also desirable to accurately identify the PCI scenarios at the device for improved cell selection and network performance. However, the data imbalance makes the classification problem challenging. This work addresses these problems through the following approach. Operator data from three LAA cellular providers is gathered and analyzed using machine learning algorithms. The impact of the phenomenon on LTE, LAA, and Wi-Fi components is demonstrated in three steps: First, the variation in network performance prediction accuracy in the PCI scenarios is examined. Second, the efficacy of numerosity reduction techniques used in data-driven cell selection is evaluated in both PCI scenarios. The third step entails a comparison of operator data analysis with network measurements. On-site experiments are conducted at the same PCI and different PCI sites to study differences in real-time network performance. A controlled LTE-WiFi coexistence environment is created and multiple traffic categories are considered. Finally, a class-weight-based solution is proposed for PCI scenario identification. F-score of 0.75 and AUC-ROC of 0.84 is achieved for LAA, with a minimalist feature set consisting of SINR and Throughput.

INDEX TERMS Unlicensed networks, cell selection, machine learning, imbalanced classification, LAA, NR-U, data analysis, network measurements.

I. INTRODUCTION

The increasing demand for mobile data has led to the utilization of unlicensed spectrum and the adoption of new cellular technologies for fair coexistence with Wi-Fi. Almost 500 MHz spectrum [1] is now available in the 5 GHz band for unlicensed operation. There has also been a rapid standardization of unlicensed coexistence cellular standards. A successful example is the Long Term Evolution-Licensed Assisted Access (LTE-LAA), prescribed in the 3GPP release 13 as

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Alatas.

an enhancement of LTE, to facilitate cellular operation in the 5GHz band. As defined by 3GPP, "Carrier aggregation with at least one Secondary Cell (Scell) operating in the unlicensed spectrum is referred to as Licensed-Assisted Access (LAA)" [2].

LTE-LAA deployments proliferated rapidly around the world, with 38 operators offering or planning to deploy LAA services in 21 countries [3]. Buoyed by the quick adoption and success of LAA coexistence deployments, 5G New Radio Unlicensed (NR-U) has been introduced as an evolutionary enhancement of LTE-LAA, under the 3GPP release 16 specifications. Furthermore, the Federal Communications Commission and the European Commission have both proposed rules for unlicensed coexistence in the 6 GHz (5925 MHz-7125 MHz) band for NR-U operation [4], [5].

The improved network performance offered by LTE-LAA depends on efficient cell selection by its Licensed and Unlicensed carrier components. Being a novel coexistence framework, LTE-LAA is characteristically different from traditional LTE heterogeneous networks (HetNets) in terms of deployment architecture, backhaul connectivity, and impact of external co-channel interference from Wi-Fi transmissions [6], [7]. Therefore, the cell selection mechanisms employed by cellular operators in LTE HetNets based on the signal-to-interference plus noise ratio (SINR) and transmission power are not suitable for LTE-LAA deployments. Although solutions for LTE HetNets/LTE-LAA cell association have been proposed, they are based on simulation or validated on network prototypes [8], [9]. More recently, a comprehensive analysis of the impact of cell selection on the performance of the LTE-LAA network and its Licensed and Unlicensed components has been performed using real-world LAA operator data [10], [11]. In addition, cell quality metrics derived from operator data have been proposed to improve the existing LAA cell selection mechanisms [10], [11].

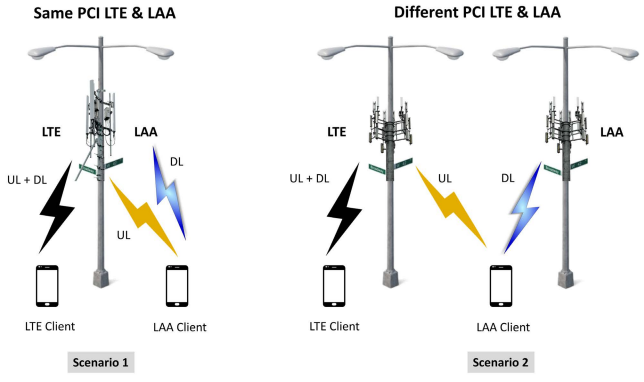


FIGURE 1. Coexistence PCI scenarios - A new phenomenon.

Despite the attention that LTE-LAA has received from industry and academia, a peculiar aspect of the current LAA deployment architecture is overlooked in the LAA literature. Cellular providers aim to minimize feedback delay from user equipment (UE) and carrier aggregation (dual connectivity), while planning coexistence deployments. Hence, macro and small cells are usually placed in close proximity, and may often be co-located on the same physical structure, such as a street-lamp/electricity pole. Thus, the LTE (Licensed) and LAA (Unlicensed) components of the network may share the same Physical Cell ID (PCI). But operator data reveals that the Licensed and Unlicensed carries of a UE may also be camped on different cells or cells with different physical locations. This creates a different PCI scenario, as shown in Figure 1. The existence of these PCI scenarios, *viz.,* "Same PCI" and "Different PCI," is a novel phenomenon that is unique to LTE-WiFi coexistence environment.

The two PCI scenarios differ in several respects, especially network performance. While LAA supports only Carrier Aggregation, NR-U will also support Dual Connectivity and Standalone unlicensed modes. Thus, this phenomenon is bound to intensify in NR-U deployments with respect to the frequency of occurrence and the multitude of PCI scenarios. Moreover, the amount of spectrum available at 6 GHz is higher compared to 5 GHz, and a higher rate of carrier aggregation per UE is possible in the 5G NR-U system. Consequently, the impact of unlicensed PCI scenarios in 5G NR-U will also be much greater than in LAA. Thus, their analysis is important not only from the perspective of optimizing existing LAA performance but also for the upcoming NR-U deployments.

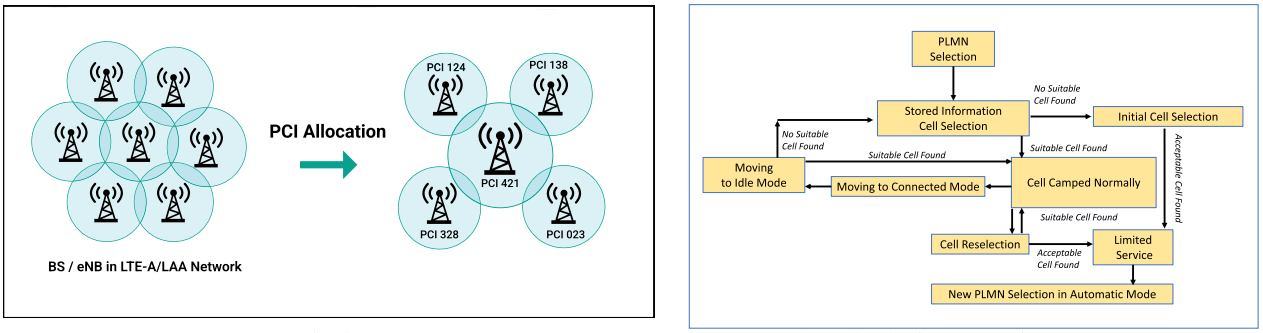
II. CHALLENGES AND CONTRIBUTIONS

From a Radio Resource Management (RRM) perspective, it is pertinent to study how the performance of the coexistence system varies when the Licensed and Unlicensed network components are attached to same or different PCIs.

A. RESEARCH PROBLEMS IN CELL SELECTION

Thus, cellular operator data needs to be analyzed to study the frequency of occurrence of the two scenarios and the distribution of important network variables in them. Data analysis is necessary to understand how the relationships between network variables differ in the two scenarios. More importantly, the variation in network performance and the ability to accurately predict expected capacity is of great interest. These tasks can typically be performed through machine learning (ML) algorithms. The PCI scenarios should also be investigated for other aspects of difference they cause in the network environment. The performance of data reduction mechanisms that are often used to help expedite data-driven cell selection by the cellular operator is one such example.

Additionally, differences in network performance should also be investigated through on-site measurements, i.e., by monitoring the LTE, LAA, and Wi-Fi networks in the two scenarios. Network performance may also vary depending on the type of traffic requested by the UE, QoS constraints, and interference from coexisting Wi-Fi access points (APs) in the unlicensed spectrum. The effect of PCI scenarios on network performance ought to be validated through real-time experiments, on both the cellular (Licensed and Unlicensed)



(a) PCI Allocation in LTE-A/LAA

(b) Cell Selection Process in LTE-A/LAA

FIGURE 2. PCI allocation and cell selection process in LTE-A/LAA.

and Wi-Fi side. It is relevant to see how Wi-Fi performance varies when coexisting with same PCI and different PCI LAA configurations. Moreover, on-site experiments should ideally consider multiple types of traffic for a comprehensive evaluation.

Equally important is the awareness of the PCI scenarios at the UE, as they are intricately associated with the cell selection decisions of the Licensed and Unlicensed components. Having an awareness of the PCI scenario will help an end-user device make informed cell selection decisions. The first step in that direction would be to accurately identify and differentiate between the two coexistence scenarios with a minimal feature vector. This is a challenging problem, given the inherent imbalance in the occurrence of the two scenarios in current LAA deployments.

These challenges must be addressed, as LAA deployments serve as a precursor to future NR-U deployments, where a lot more complex configurations are possible. For example, it is possible that NR-Licensed, NR-U, LTE, and LAA small cells are located on adjacent poles, leading to far more complex PCI scenarios. Thus, cell selection decisions resulting from awareness of PCI scenarios will ensure better network performance for the end-user.

However, to the best of our knowledge, no major study has investigated the current LAA deployments for the phenomenon of coexistence PCI scenarios and the challenges they present.

B. RESEARCH CONTRIBUTIONS

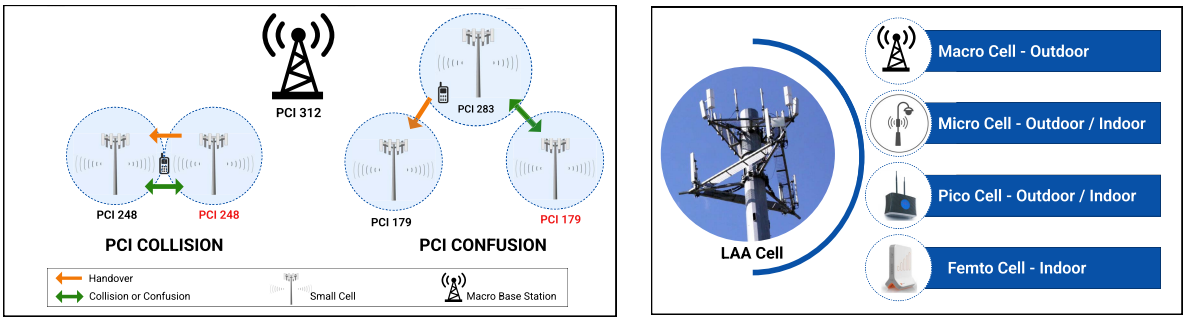
The proposed work analyzes LTE-LAA deployments through field experiments, measurements, and machine learning-based operator data modeling, with the vision to find solutions for upcoming NR-U deployments. The major research contributions are elucidated below.

- **Operator Data gathering and extraction**: To study the variation in the prediction of network performance in the two PCI scenarios, LTE-LAA operator data is gathered from the deployments of three cellular operators active in downtown Chicago, *i.e.,* AT&T, T-Mobile, and Verizon. Using computer vision and deep learning-based tesseract engine, network data is extracted from video logs to generate a data set of over 7000 data points.
- **Impact Assessment of PCI Scenarios** A PCI scenario specific distribution analysis of important network variables such as Throughput and SINR is performed. Further, the effect of PCI scenarios on network performance prediction and data reduction techniques in Licensed and Unlicensed components is explored and insightful inferences are drawn. This is done by analyzing the SINR-Capacity relationship through eight regression and supervised machine learning algorithms, and two popular numerosity reduction algorithms.
- **On-site Experiments**: Operator sites with Same and Different PCI configurations are identified using the Network Signal Guru (NSG) tool. Then, on-site experiments are conducted using multiple coexisting Wi-Fi hotspots. Variation in network performance is observed for both PCI scenarios by varying the type of traffic, the number of interfering Wi-Fi hotspots, the distance *etc.* Further, role of QoS constraints in LTE-LAA performance is explored through delay critical data traffic. Metrics such as Throughput, Signal Strength, Resource Block allocation, and Latency are measured. Experiments are conducted for both cellular (LTE/LAA) and Wi-Fi sub-systems.
- **Predicting PCI Scenarios** A binary classification model is designed that predicts the PCI scenario with high accuracy, in both Licensed and Unlicensed component. The proposed classification model requires a minimalist two-variable feature set and overcomes the challenge of data imbalance between the two scenarios.

Finally, open research problems related to the PCI scenarios in the unlicensed coexistence networks are identified and outlined.

III. CHALLENGES IN THE UNLICENSED BAND

This section briefly discusses the PCI allocation process and the challenges in unlicensed cell selection, followed by the unique PCI scenarios observed in LAA deployments.



(a) Challenges in PCI allocation

(b) Greater Competition for Limited PCIs

FIGURE 3. PCI allocation and challenges.

A. ROLE OF PCI IN CELL SELECTION

With the adoption and proliferation of new cellular standards, a wide variety of base stations, from Macro cells to low-power Femto cells, are being deployed. The consequent increase in network complexity presents new challenges in the planning, management, and maintenance of cellular networks. Efficient operation of cellular networks, especially LAA/5G-NRU deployments, requires reliable mechanisms for network configuration and reconfiguration with minimal human intervention. This served as motivation for the paradigm of Self-Organizing Networks (SON), where seamless and uninterrupted service is provided to the end-user through small cells capable of self-configuration [12].

The first and most important step in a cellular SON is *cell identification*, without which optimal cell selection and handover to user equipment (UE) is not possible. At the physical layer, the identification of a cell/eNB is done with the help of Physical Cell Id (PCI) in LTE-A/LTE-LAA and 5G NR-U. PCI allocated to a cell/eNB (Figure 2(a)) is used to discover and identify the cell/eNB when a cellular device initiates the association process, or during a handover [13]. The high-level cell selection process in LTE, LTE-A, and LTE-LAA is presented in Figure 2(b). In these cellular standards, the PCI of a cell is generated using the Primary Synchronization Signal (PSS) and the Secondary Synchronization Signal (SSS), using the expression ($PCI = 3 \times SSS + PSS$), where $PSS \in \{0, 1, 2\}$, and $SSS \in \{0, \ldots, 167\}$. Therefore, a typical PCI in LTE-LAA is always in the range of [0,503]. While PCIs in 5G NR/NR-U are generated in a similar fashion, 1008 unique PCIs are available, as $SSS \in \{0, \ldots, 335\}$.

B. NEW CHALLENGES IN THE UNLICENSED BAND

Traditional LTE/LTE-A cell selection process is generally dependent on parameters such as RSRP, RSRQ, SINR and transmission power received at the UE. Although there is ample research to suggest that parameters associated with traffic QoS (e.g., Guaranteed Bit Rate), traffic-load, and power/energy consumption at the UE ought to be considered, operators prefer practical cell selection and handover mechanisms that are SINR driven [14]. The simple approach works for cellular operators *viz.,* AT&T, T-Mobile, and Sprint as LTE/LTE-A HetNets do not suffer the adverse impact of interference from external sources [15]. LTE/LTE-A networks operate in dedicated bands of the Licensed spectrum that are exclusive to specific cellular operators. Further, the co-tier and cross-tier interference experienced within the network is usually mitigated and managed through centralized and distributed resource allocation solutions [16].

In sharp contrast, unlicensed cellular networks have to contend with transmission conflicts from coexisting Wi-Fi APs on the same channel, which can drastically degrade the SINR. Furthermore, interference from rogue mobile hotspots and other sources, hidden node problems, and dense Wi-Fi (802.11n/ac/ax) network deployments adversely affect SINR in an unlicensed coexistence system such as LAA. This is particularly true in dense scenarios, where radio resources may not always be allocated to LAA. SINR-based cell selection is also less suitable in LAA due to interference from external Wi-Fi transmissions, which typically causes fluctuation in SINR received at the LAA UE. As a result, cell selection and handover decisions made at the UE trigger frequent disconnections and re-attachments, degrading LAA performance. Thus, for efficient cell selection and radio resource allocation in LAA, additional information about the small cell is necessary to determine cell quality [10], [17].

Case in point, when LTE and LAA share a PCI, it implies that they share the same LTE backhaul, i.e., LTE Evolved Packet Core (EPC) with MME, P-GW, S-GW, HLR, etc., which ensures efficient splitting of resources. Additional signaling is required at the LAA secondary cell (Scell) for the UE to recognize that the component carrier belongs to the same LAA provider as the primary cell (Pcell), which in turn provides the Licensed carrier to the UE. In the Licensed band, PCI is sufficient for cell identification, since multiple operators share a carrier only if the operators: (a) provide coverage to spatially/geographically separated areas, or, (b) share the network. Despite these provisions, two major PCI related problems are encountered, which are illustrated in Figure 3(a) and briefly described below:

1) PCI collision: Neighboring cells are assigned identical PCI, which poses a challenge during UE attachment and handovers.

2) PCI confusion: Two different small cells within the same macro Base Station (BS) are assigned identical PCI, which poses a challenge, especially during handovers.

Since LTE-A/LAA has only 504 unique PCI values, strategic reuse and network planning provide reliable identification of cells/eNBs. However, with increasing network density and low-power cell deployment, as shown in Figure 3(b), an LAA cell has to compete with other small cells and macro cells for a limited number of PCIs, leading to a higher frequency of PCI collisions and PCI confusions [12]. An important reason is that cells/eNBs of different operators belonging to different Public Land Mobile Network (PLMNN) may assign the same PCI to their respective eNBs [2], [18]. These challenges were adequately addressed in LTE/LTE-A networks through smart PCI allocation using mechanisms such as graph coloring, clustering, and network-initiated SI reading, since operators controlled the Licensed bands allocated to them [12], [19]. However, the solutions prescribed for the unlicensed band, *viz.,* increasing the range of PCIs, broadcasting partly unique cell identities, and interoperator coordination are difficult to implement in a shared spectrum [18]. Furthermore, the complexity of these challenges will increase as cellular operators and enterprise network providers deploy more Citizens Broadband Radio Service (CBRS) small cells operating on the LTE protocol stack [20]. The upcoming small cell deployments in the C-band spectrum are also likely to intensify the contention in PCI ID allocation.

Finally, 3GPP release 15 may also prescribe additional specifications for LAA processes [2]. For example, an LAA handover process entails the following specific steps. First, the eNB/small cell configures the UE with one Discovery Signal Measurement Timing Configuration window on one frequency for the serving cell and for each available neighboring cell. Second, the UE utilizes the allocated window to identify and measure Discovery Reference Signals from the neighboring cells. Further, to overcome the challenges of hidden-nodes (unlike LTE/LTE-A), the UE may measure the Received Signal Strength Indicator (RSSI) during an RSSI measurement timing configuration window and share the average RSSI and channel occupancy with the serving cell in the next reporting interval [2].

Solutions to the problems anticipated in LAA networks have been prescribed in 3GPP release 15, in 3GPP minutes of meetings, and in recent research work [2], [6], [18]. Prior to the proliferation of public LAA deployments, the analysis and solutions for LAA were validated primarily through theoretical modeling, simulations, and experimental testbeds [21], [22], [23], [24]. After several cellular operators began offering LAA services globally, some studies have sought to conduct performance analysis of these networks through on-site measurements [6], [10], [25].

However, the new phenomenon observed in unlicensed deployments, which is presented in the next section, is not yet identified, nor has its impact been analyzed using ML modeling and operator data analysis.

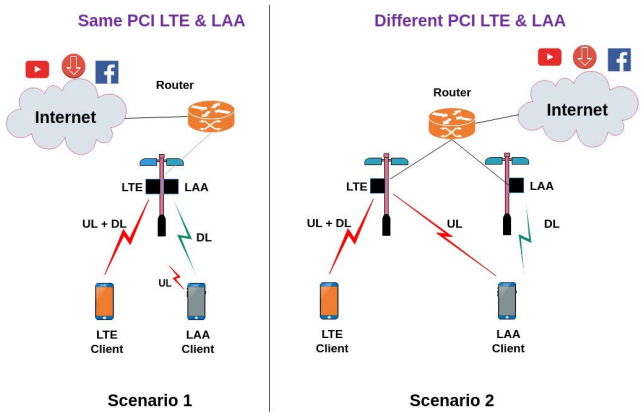


FIGURE 4. PCI scenarios in unlicensed networks.

IV. UNIQUE PCI PHENOMENON IN LAA DEPLOYMENTS

In the current coexistence deployments, the uplink transmissions are generally performed on the licensed spectrum and downlink transmissions on the LAA unlicensed spectrum. The strategy is designed to meet QoS requirements, reduce delay in ACK packets, and maximize power and spectral efficiency. In contrast, if the client is LTE-only, both uplink and downlink occur on the licensed LTE BS. However, cell selection in LTE-LAA coexistence networks involves additional technical nuances due to the system architecture.

A. THE TWO PCI SCENARIOS

Data collected from coexistence deployments of the three operators in Chicago, *viz.,* AT&T, Verizon, and T-Mobile reveals that a typical LTE-LAA coexistence deployment may be broadly categorized into two PCI scenarios, depicted in Figure 4. ''Same PCI'' represents the coexistence network architecture with Licensed (LTE) and Unlicensed (LAA) clients camped on the same cell, while ''Different PCI'' depicts the case when they are latched on to different cells with different PCIs. An architectural representation of the two scenarios is shown in Figure 4. Same PCI is the dominant scenario and is observed in over 80% of the sample space. It also seems to be the preferred scenario for maximal network throughput, which is why cell selection solutions often focus on it [10], [11]. However, the occurrence of Different PCI scenarios in LAA operator data is significant (18%) and cannot be overlooked [11]. Further, to the best of our knowledge, this is a hitherto unobserved phenomenon and unique to the coexistence paradigm. With the commencement of 5G-NRU operation in the 5GHz and 6GHz spectrum, taking the PCI scenarios into consideration will enhance coexistence network performance [6], [9].

These scenarios are the result of three network design factors, *viz.,* availability of space/structures for cell installation, availability of fiber optic-equipped backhaul connectivity, and operator deployment strategy. The LTE-LAA coexistence networks in the Chicago region have been deployed in both semi-indoor and outdoor environments. The operator data we

TABLE 1. Channel access priority class in LAA downlink.

DL CAPC	Initial CCA	CWmin	CWmax	TXOP	QCI
1 (Voice)	25 μs	30 μs	70 μs	2 ms	1, 3, 5, 65, 66, 69, 70
2 (Video)	25 μs	70 μs	150 μs	3 ms	2, 7
3 (Best Effort)	43 μs	150 μs	630 μs	8 ms, 10 ms	4, 6, 8, 9
4 (Background)	79 μs	150 μs	10.23 ms	8 ms, 10 ms	–

have gathered is for an outdoor setting, where the LTE-LAA cells have been installed on the street lights/electricity poles. The deployment plan is shaped not only by the availability of these poles but also by their type. Only some of these poles are structurally capable of providing a high-speed fiber optic connection to the backhaul. From the network planning perspective, this is an important consideration, because the Unlicensed carrier bandwidth is 3X that of the Licensed carrier, and the next-generation millimeter-wave networks offer up to 400 MHz bandwidth, which is 10X that of the current 5 GHz unlicensed bands [6]. Such speeds can only be serviced by a fiber optic connection to the backhaul. CapEx-QoS trade-off is another major consideration, as fiber optic cabling is expensive and is set to cost the 5G industry over $130 Billion in installation. These factors, coupled with market penetration of LAA equipped devices and subscriber density, determine operator deployment strategy and create the two PCI scenarios.

Preliminary on-site measurements indicated that LAA performance in the two scenarios is likely to be different. When LTE and LAA have the same PCI, they share the backhaul, which facilitates efficient resource splitting and packet aggregation. Consequently, the LTE-LAA network performance analysis is usually limited to the Same PCI scenario [10], [11]. However, the two PCI scenarios may impact LAA and Wi-Fi network performance differently.

Furthermore, the performance of the LAA uplink (UL) and the LAA downlink (DL) is likely to differ. This becomes important due to LAA mechanisms such as subframe and carrier (pool-based) resource allocation schemes and group UL grant. The two PCI scenarios may influence the use of Group ID or PCI for multiplexed transmissions on the subframe [2], [18]. Thus, Same PCI and Different PCI configurations will affect LAA UL and DL mechanisms as well. Similarly, these PCI scenarios will affect the network throughput, resource allocation, and latency for different types of data traffic depending on the respective Channel Access Priority Class. Equally importantly, the impact of LAA on the performance of coexisting Wi-Fi networks will also vary depending on the type of coexistence PCI scenario. In this work, we explore these aspects through a PCI scenario-specific analysis of LAA deployment data.

B. RELEVANCE TO 5G NR-U

LAA is the evolutionary antecedent of NR-U, and the first commercially successful cellular-WiFi coexistence standard.

TABLE 2. Access categories in 802.11ac.

Access Category	AIFS	CWmin	CWmax	TXOP
Voice (AC_VO)	18 μs	27 μs	63 μs	2.08 ms
Video (AC_VI)	18 μs	62 μs	135 μs	4.096 ms
Best Effort (AC_BE)	27 μs	135 μs	9.207 ms	2.528 ms
Background (AC_BK)	63 μs	135 μs	9.207 ms	2.528 ms

As such, the unresolved challenges encountered in LAA will assume a different dimension and scale in NR-U. This is especially true for the phenomenon of PCI scenarios that has come to attention only through the analysis of LAA operator data.

NR-U deployments will support three modes, viz., Carrier Aggregation, Dual Connectivity, and Standalone [26]. Apart from the fully unlicensed Standalone operation, the other two modes will operate in the 5GHz spectrum. This implies that the following can be present/co-located on a physical installation such as a pole or a lamppost:

- NR-U and NR Licensed
- NR-U alone in Standalone mode
- NR Licensed and LAA Unlicensed
- NR Licensed, NR-U and LAA
- LTE Licensed, NR Licensed, NR-U and LAA

Compared to the LAA deployment scenarios shown in Figure 4, the above small cell combinations present in physical installations close to each other can lead to more complex PCI scenarios than LAA. Thus, analysis and solutions presented in this work will certainly guide solutions for cell-selection and network optimization in NR-U.

V. UNLICENSED COEXISTENCE: RELEVANT ASPECTS

This section discusses the aspects of coexistence in the unlicensed band that are relevant to the experiments and analysis presented in this work. These include the bands and channels on which LAA/NR-U and Wi-Fi operate, the feasibility of bandwidth combinations, channel access mechanisms, and Channel Access Priority Class (CAPC).

A. THE 5 GHz UNLICENSED BAND

The 5 GHz Unlicensed National Information Infrastructure bands (U-NII bands) range from 5.15 to 5.85 GHz. Recently, a U-NII-4 band with 4 channels spanning 5.85 Ghz–5.925 GHz, has been notified by the Federal

Communications Commission through a Notice of Proposed Rulemaking [4]. Thus, a total of eight U-NII ranges have been prescribed by FCC, of which 1 to 4 lie in the 5 Ghz range and 5 to 8 are for 6 GHz. LAA coexists with IEEE 802.11a and higher Wi-Fi standards in the 5 Ghz spectrum. Each U-NII band is governed by certain guidelines and restrictions with respect to usage and operation. For example, U-NII 1 (5.150–5.250 GHz) is a low power band, which has no restrictions apart from being limited to indoor use with 50 mW of maximum transmission power. U-NII-2 is assigned mainly to radar systems. Therefore, unlicensed devices communicating in the band have to necessarily implement Dynamic Frequency Selection (DFS), giving priority to the radar signals. As a result, this band is seldom utilized by LAA and Wi-Fi. Therefore, although a substantial block of 560 MHz of spectrum seems available, a much smaller fraction of about 160 MHz belonging to U-NII-1 and U-NII-3 bands is fully harnessed. It is noteworthy that Wi-Fi channels are 20/40/80 MHz wide, while 3GPP has specified 20 Mhz channels for LAA, with the possibility of aggregating up to 3 LAA channels.

B. CHANNEL ACCESS PRIORITY CLASSES

An operator uses Channel Access Priority Classes (CAPCs) for specific types of traffic when UEs transmit data on the uplink (UL) and downlink (DL) over an LAA carrier. Four CAPCs are specified for LAA in 3GPP Release 15 [2]. Table 1 presents the details of the LAA DL CAPC classes along with specific values for each class for parameters *viz.,* QoS Class Identifier (QCI), Clear Channel Assessment (CCA) duration, maximum and minimum contention windows (CW_{Max}, CW_{Min}), and Transmission Opportunity (TXOP) duration. A data traffic category is assigned a specific TXOP duration based on its CAPC. For example, video and voice data have TXOPs of 3 ms and 2 ms, respectively, to meet end-user QoS guarantees. Given the low latency requirements of these data flows, small data packets are used. In contrast, for traffic categories such as data download in the background and best-effort, LAA employs large packets to optimize network throughput with a maximum TXOP of 8 ms. To meet QoS guarantees, 3GPP prescribes that operators use CAPCs according to the standardized QCI to which the data traffic belongs. However, operators are free to employ their own non-standardized QCIs so long as a suitable CAPC is used for data traffic. In uplink transmission, LAA eNB determines the CAPC of the traffic by considering the QCI with the least priority in a Logical Channel Group.

Similarly to LAA, Wi-Fi classifies traffic into Access Categories (ACs) depending on the type of traffic and priority. Wi-Fi (802.11ac) ACs for various types of traffic *viz.,* video, voice, background data, and best-effort data, with the corresponding values for parameters *viz.,* Arbitration Inter-frame Spacing (AIFS), CW_{Max}, CW_{Min}, and TXOP are presented in Table 2.

C. CHANNEL ACCESS MECHANISMS

LTE-LAA employs a Listen-Before-Talk (LBT) mechanism to access the unlicensed channel. The mechanism resembles the Carrier-sense Multiple Access with Collision Avoidance (CSMA/CD) MAC protocol of Wi-Fi. LBT facilitates a better coexistence of LAA with Wi-Fi than LTE-Unlicensed (LTE-U), which has a duty-cycle mechanism [27]. LAA eNB and UEs sense the channel prior to transmitting on the LAA SCell. The LAA eNB/UE LAA will defer for an initial CCA duration depending on the CAPC of the traffic. Subsequently, LBT triggers a random back-off period in the range of (0, CW_{size}), where CW_{size} is assigned a value between CW_{Min} and CW_{Max} according to channel conditions. If the back-off is successful, the LAA eNB/UE starts transmitting.

Wi-Fi makes use of the CSMA/CD mechanism, which ensures that an AP/device transmits only when the channel is idle and the station has not finished a successful transmission immediately prior to sensing the channel. The device waits for an initial sensing period of Distributed Coordination Function Inter-frame Space (DIFS) for IEEE 802.11n and below, or AIFS for amendment 11ac and above. If the Wi-Fi node is contending for access immediately after a successful transmission, it continues to sense the channel for an initial length of DIFS/AIFS, until the channel is idle again. If the channel is detected to be busy, a random back-off period in the range of (0, CW_{size}) is selected.

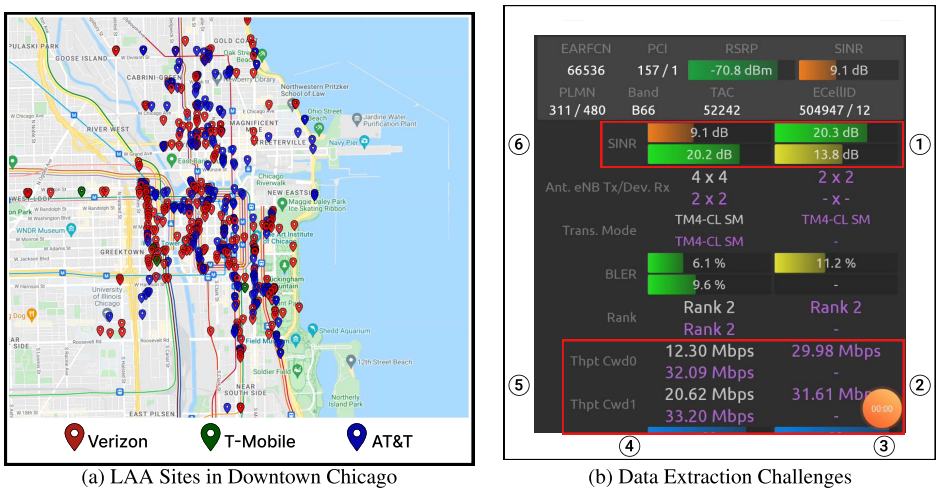
LAA's LBT is compatible with CSMA/CD of Wi-Fi, which reduces collisions during transmissions and leads to fair coexistence and enhanced network performance [28]. Furthermore, Wi-Fi has an energy threshold of −62 dBm and a preamble detection threshold of −82 dBm. Since the LAA LBT has an energy threshold of −72 dBm, its vulnerability to interference from Wi-Fi APs is reduced, which also facilitates fewer fluctuations in signal strength at the LAA UE. Finally, upon getting access to the channel, an LAA eNB/UE can transmit data for a TXOP length of up to 10 ms if it has prior information that a Wi-Fi device does not coexist in the unlicensed band. However, onsite measurements reveal that for most types of traffic (e.g., data, data plus video/streaming, etc.) LAA LBT generally allows for a maximum TXOP of 8 ms to the UE for transmission.

VI. DATA GATHERING AND EXPLORATORY ANALYSIS

Having discussed the new-found phenomenon and the technical aspects relevant to the analysis, this section describes the data gathering exercise, followed by data distribution analysis.

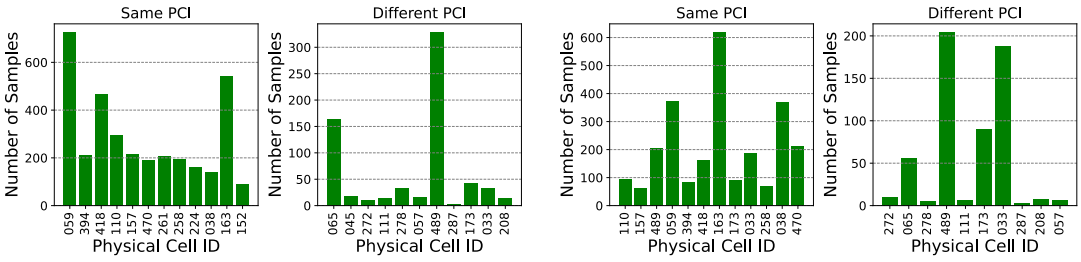
A. OPERATOR DATA COLLECTION AND EXTRACTION

Real-time LAA network data for three operators, *viz.,* AT&T, T-Mobile, and Verizon is gathered from downtown Chicago, as depicted in Figure 5(a), through the *Network Signal Guru* (NSG) app [29]. NSG is a subscription-based application developed and released by Qtrun Technologies. It offers detailed and precise information on network parameters such



(a) LAA Sites in Downtown Chicago (b) Data Extraction Challenges

FIGURE 5. LAA deployment sites and challenges in data extraction from NSG app.



(a) Unlicensed Data (b) Licensed Data

FIGURE 6. Sample sizes for the same PCI and different PCI scenarios.

as bandwidth, signal strength, throughput, resource allocation, block error rate, channel frequency, modulation coding scheme, and several others. A section of NSG's user interface is shown in Figure 5(b). It supports a wide variety of cellular standards (e.g., LTE, LAA, 5G, etc.) and provides more accurate and detailed information than other applications such as SigCap or FCC Speed Test. Unfortunately, NSG only permits monitoring the data and does not allow extraction for analysis in a suitable format such as ".CSV" or ".txt". Thus, operator data extraction involved a complex process that required screen-capturing NSG interface into a video and then processing it through computer vision techniques followed by a deep-learning-based OCR engine.

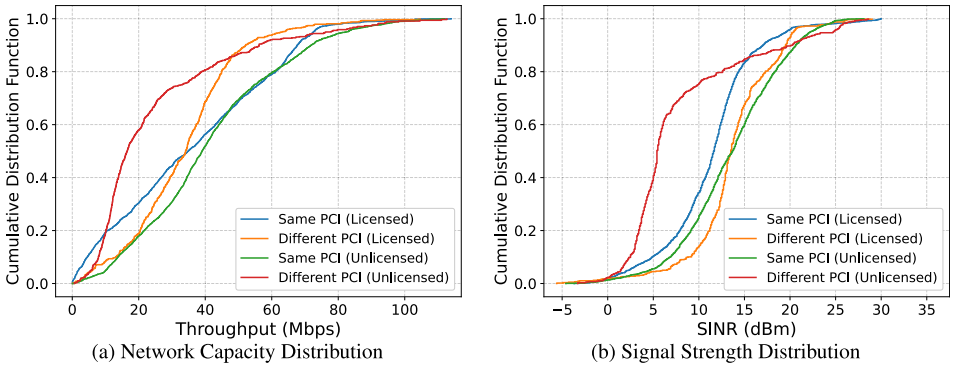
The video-to-text extraction process presented several challenges, some of which are shown in Figure 5(b), with the aid of numbered labels. First, the screen refreshes at intervals that are hard to predict, and the interface scrolls back to the top. This makes screen-capturing specific fields with perfect alignment for the computer vision (CV) and OCR pipeline extremely tedious. Second, values of the same metric (e.g., Throughput) are in different colors and need to be processed through different CV and OCR techniques for high accuracy. Third, elements in the UI mask portions of the values to be extracted, adversely impacting text recognition accuracy. Fourth, some fields/parameters have a compact placement, requiring additional steps in the image processing pipeline. Fifth, not all the fields are populated, which makes mapping of corresponding fields (e.g., Throughput with SINR) a non-trivial task. Finally, varying colors and extent of the backgrounds of fields (e.g., SINR) degrades recognition accuracy especially when dealing with negative values and, therefore, need a specially designed image processing pipeline. Our engineering solution solved these problems and close to 7500 samples were successfully extracted and analyzed in this work.

B. EXPLORATORY DATA ANALYSIS

1) SAMPLE SIZE DISTRIBUTION

The PCI-specific distribution[1] of extracted samples in the Same PCI and Different PCI scenarios, for Licensed and Unlicensed components is presented in Figure 6. As discussed earlier, Same PCI has a much higher proportion of samples in both LTE and LAA components. However, there is another dimension to this imbalance. Since the Different PCI LAA sites are less in number, the probability of

[1]Due to a large number of PCIs, only those with significant sample-sizes are displayed here.



(a) Network Capacity Distribution (b) Signal Strength Distribution

FIGURE 7. Distribution of network variables in the two PCI scenarios.

encountering them also reduces. Consequently, the samples are also concentrated within a few small cells, leading to an internal pattern imbalance in the feature set of the Different PCI sample space. These challenges add complexity to the intended objective of accurately identifying PCI scenarios, through imbalanced classification with a small feature set.

2) DENSITY DISTRIBUTION

The cumulative distribution plots of Throughput and SINR values in the operator data are presented in Figures 7(a) & 7(b), respectively. Both distributions exhibit the impact of the coexistence component and the PCI scenario within the component.

Beginning with component-specific analysis, for more than 85% Different PCI Throughput samples, the Unlicensed component has a lower magnitude, but this trend reverses for the last 15% samples. Consequently, 80% users in the Unlicensed Different PCI environment are likely to receive 10 Mbps less Throughout than Licensed Different PCI users. In contrast, Licensed Same PCI performs worse than Unlicensed Different PCI, for 80% users, but the performance of both is comparable for 20% users in the Throughput range of 45 Mbps–63 Mbps. The SINR trend for Licensed and Unlicensed Different PCI conforms to the Throughput trend. However, the SINR trend for the Same PCI does not correspond to the Throughput trend in entirety, and Unlicensed performance remains better than Licensed throughout. This shows a clear impact of PCI configurations on the performance of unlicensed network components. The allocation of radio resources and their availability play a key role in network performance. There seems to be better resource allocation in Licensed Same PCI that enables it to match up to Unlicensed Same PCI, despite lower SINR.

A PCI scenario-specific analysis also highlights a difference in resource allocation in the two scenarios. For example, Licensed Same PCI performs better than Licensed Different PCI in terms of Throughput, in 50% samples where Throughput is higher than 35 Mbps. However, the cumulative Licensed Same PCI SINR distribution remains consistently higher than the Licensed Different PCI SINR. This indicates that the variation in Throughput is a result of other network-related factors that differ in the two scenarios. In the Unlicensed component, the Same PCI Throughput is always better than the Different PCI Throughput. This is despite the fact that the cumulative distribution of Unlicensed SINR is not always better in Same PCI, and Different PCI performs better at higher SINRs ($\gtrsim 21.5\ dBm$).

Overall, both Licensed and Unlicensed Same PCI network capacity seems to be better than the corresponding Different PCI. This could be attributed to more efficient cell selection and backhaul.

VII. NETWORK PERFORMANCE ANALYSIS

The objective of the analysis presented in this section is to infer how network performance in the Licensed and Unlicensed components of the coexistence system varies in each PCI scenario. The discussion seeks to (a) highlight the relevance of PCI scenario-specific network analysis, (b) investigate the impact of PCI scenarios on the network performance in Licensed and Unlicensed components, and (c) examine the performance of data reduction techniques in both scenarios.

A. METHODOLOGY AND TECHNIQUES

1) METHODOLOGY

To study the variation in network performance in the two PCI scenarios, a data-driven approach leveraging machine learning algorithms is ideal [30], [31]. Thus, the prediction of network performance is done using feature relationship modeling. Operator data is subjected to eight supervised machine learning algorithms to determine the relationship between network feature points of primary importance, *viz.*, Signal and Interference plus Noise Ratio (SINR), network throughput (Capacity), and cell-allocation (PCI). It is a good practice to employ a broad set of machine learning algorithms consisting of different types *e.g.,* deterministic and stochastic, to ensure a robust coexistence performance prediction and feature relationship analysis through cross-verification of results. Further, for both PCI scenarios, *i.e.,* Same PCI and Different PCI, a fine-grained SINR-Capacity feature relationship analysis is presented by comparing models that consider

PCI as a categorical parameter (With PCI) with models that do not consider PCI as a categorical parameter (Without PCI). Since the analysis of the combined coexistence network data is shown to obfuscate the findings [10], [11], operator data is segregated into its Licensed and Unlicensed components. Finally, the R-sq of a model reflects the strength of the relationship between the network feature points. In this work, a multivariate relationship analysis is performed to determine the explanability of network throughput. Thus, the model R-sq is denoted as "**$\text{Throughput}_{\text{Exp}}$**" in this work. $\text{Throughput}_{\text{Exp}}$ serves as a metric for a comparative analysis of network performance prediction and feature relationship strength. A higher $\text{Throughput}_{\text{Exp}}$ implies a strong feature relationship and consequently a greater confidence in the model to predict the response variable, i.e., network throughput.

2) MACHINE LEARNING ALGORITHMS

Let N be the number of training points and D be the dimensionality of the feature vector. Then the LTE-LAA network data can be represented as $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^D$ is the feature vector and $y_i \in \mathbb{R}$ is the ground truth value for the i^{th} training point. The objective is to learn a mapping $f : \mathbf{x}_i \rightarrow y_i$ where x_i is the feature vector constructed using Signal & Interference plus Noise Ratio (SINR) and cell-allocation (PCI) and y_i is the target value, *i.e.,* network throughput (Capacity). A diverse set of regression algorithms are considered, which are discussed below:

(i) *Linear Regression*: This family of algorithms [32] aims to learn a linear mapping by solving, $\arg\min_{\mathbf{w},b} \sum_{i=1}^{N} ||(\mathbf{w}^\top \mathbf{x}_i + b) - y_i||_2^2 + \alpha \mathbf{w}^\top \mathbf{w}$. Here, $\mathbf{w} \in \mathbb{R}^D$ is the weight vector and $b \in \mathbb{R}$ is the bias term. Moreover, α is a hyperparameter used to control the importance/weightage of the l_2-regularization term. α is set as zero in the case of Ordinary Least Squares Linear Regression (OLS), whereas it is set via k-fold cross-validation (kCV) in the case of Ridge Regression (RR).

(ii) *Kernel Regression*: Kernel Ridge Regression algorithms [32] learn a non-linear mapping through a kernel function $K(a, b)$. The goal is to solve $\arg\min_{\mathbf{w},b} \sum_{i=1}^{N} ||K(\mathbf{w}, \mathbf{x}_i) + b - y_i||_2^2 + \alpha \mathbf{w}^\top \mathbf{w}$. Here, $\mathbf{w} \in \mathbb{R}^D$ is the weight vector, $b \in \mathbb{R}$ is the bias term, and α is a hyperparameter as defined earlier. Varying the kernel function as Radial Basic Function and Polynomial leads to Kernel RBF Regression (RBF) and Multi-variate Polynomial Regression (MPR), respectively. We considered 2–4 degree polynomials in MPR and the optimal degree was chosen based on the value of R-sq via kCV.

(iii) *Neural networks (NN)*: This family of algorithms learns a non-linear mapping via a sequence of feed-forward layers in contrast to hand-crafted kernel functions used by Kernel Ridge Regression [32]. A wide range of neural networks were considered with 1–2 hidden layers, 5–50 neurons per hidden layer, and ReLU, logistic & Tanh activation functions. The optimal configuration and hyperparameters were chosen on the basis of the R-sq metric via kCV.

(iv) *Decision Tree Regressor*: Decision Trees (DT) recursively partition the data points to minimize the mean squared error at each node [32]. We experimented with trees of depth 5–20 and the optimal depth was determined via kCV. The Minimal Cost-Complexity Pruning algorithm was implemented to prune the learned tree and avoid overfitting [32]. Ensembles with 3–100 base learners were considered, and the optimal number was determined by kCV. In particular, an ensemble of 25 decision tree regressors yielded the best R-sq value for both Random Forest Regressor (RF) and Gradient Boosting (GB) in the experiments.

B. PCI SCENARIOS AND LTE-LAA PERFORMANCE

SINR-Capacity feature relationship analysis is performed on LTE-LAA operator data. It is worth mentioning that for Licensed and Unlicensed, 18.61% and 16.04% data belongs to the Same PCI scenario, respectively. The rest falls under the Different PCI scenario. Linear algorithms (Linear and Ridge Regression), generate models with lower $\text{Throughput}_{\text{Exp}}$ than non-linear algorithms (remaining six algorithms). The latter seem to be more suited to fit the SINR-Capacity (and PCI) data which is also considered to be non-linear [10], [11]. However, the trend of predictive performance is fairly consistent across all algorithms. Hence, the discussion ahead focuses on network feature relationship patterns as a whole and not on the results of individual algorithms.

The results of data analysis are distilled in the form of inferences presented below along with the explanations.

(i) *Component and PCI scenario specific data segregation seems to be a pre-condition for accurate coexistence network performance prediction*. Although the need for component-specific data segregation has been highlighted earlier [10], [11], for reliable modeling of feature relationships, component-specific data segregation alone is not sufficient. Figure 8(a), presents the relationship analysis for data segregated based on LTE-LAA network components, with and without PCI as categorical parameter. It shows that the SINR-Capacity $\text{Throughput}_{\text{Exp}}$ for Licensed data models is always lower than the corresponding Unlicensed models. However, it will be demonstrated ahead that this is not the case. Similarly, segregating data based on PCI scenario alone does not yield an accurate network performance prediction. For example, Figure 8(b), shows that the $\text{Throughput}_{\text{Exp}}$ of Same PCI models is invariably lower than that of the different PCI models. Again, we now know that this is not true.

Thus, for reliable network prediction models, coexistence network data should be segregated based on both *network component* (Licensed/Unlicensed) and *PCI scenario* (Same/Different).

(ii) *Current LAA deployment architecture may not facilitate efficient Licensed operation in both PCI scenarios*. It is evident from Figure 9(a), that Licensed Same PCI model $\text{Throughput}_{\text{Exp}}$ is significantly higher than the corresponding Different PCI model $\text{Throughput}_{\text{Exp}}$. This implies that predicting network performance (e.g., network capacity) is

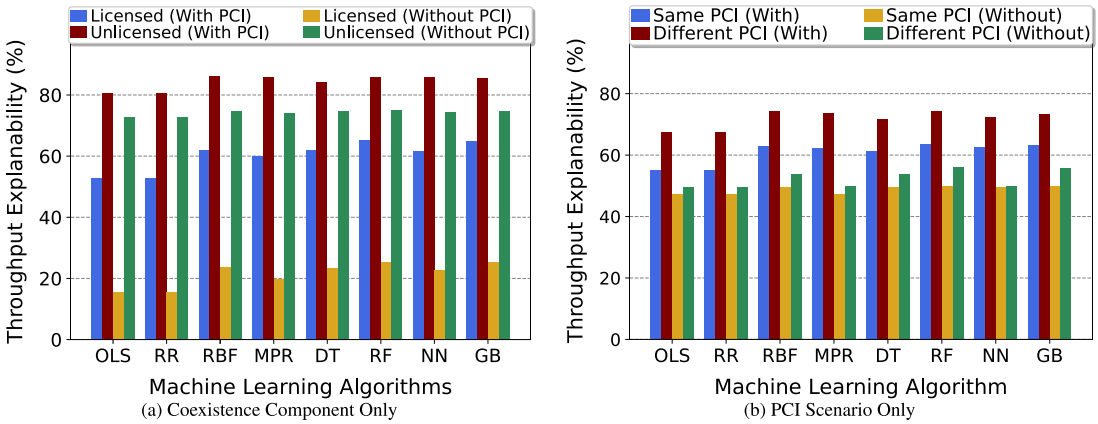


FIGURE 8. Data segregation for performance prediction through feature relationship analysis.

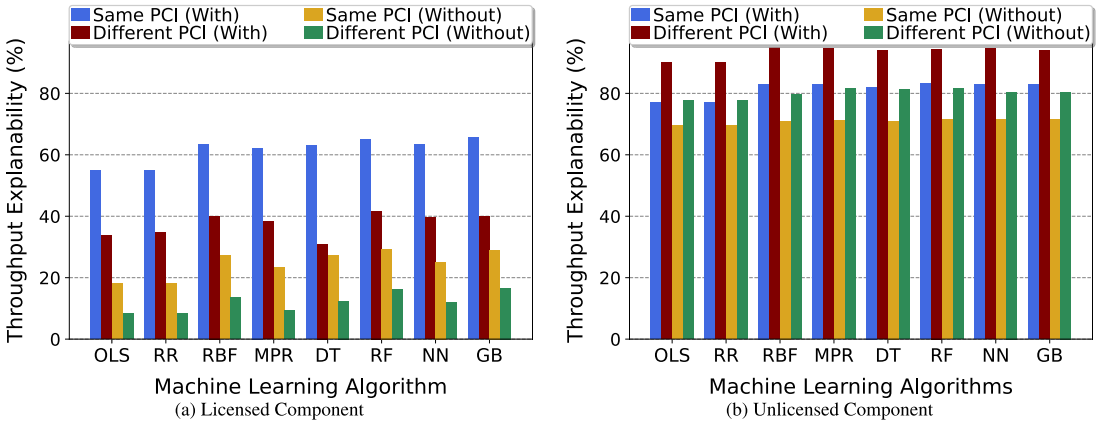


FIGURE 9. LAA PCI scenarios and network performance prediction.

more reliable in the former case. The trend holds for all regression and learning algorithms, regardless of whether PCI is considered a categorical parameter. However, without PCI as a categorical parameter, the difference in the model $\text{Throughput}_{\text{Exp}}$ for the two scenarios is exaggerated (109.72% on average) compared to when PCI is a categorical parameter (65.52% on average). Thus, the ''Without PCI'' modeling should be avoided for a reliable representation of the feature relationships. The more important inference is that for the Licensed component, feature relationships are rather poor in the Different PCI scenario. This implies that network performance cannot be predicted with confidence.

Thus, the current coexistence deployment architecture appears to be creating performance bottlenecks in the Different PCI configuration and may not facilitate optimal network performance in the Licensed component.

(iii) *Current LAA deployment architecture favors Unlicensed operation.* $\text{Throughput}_{\text{Exp}}$ of Unlicensed models demonstrates a trend exactly opposite to that observed in the Licensed component. Different PCI scenario models, across all machine learning algorithms, perform better than Same PCI scenarios. Further, there are three more points of difference when compared to the Licensed performance prediction patterns. First, the difference in $\text{Throughput}_{\text{Exp}}$ for the two Unlicensed PCI scenarios is not as pronounced as it was in the Licensed (only 14.68% on average for ''With PCI''). Second, the absolute magnitude of $\text{Throughput}_{\text{Exp}}$ for Unlicensed models, for both Same and Different PCI scenarios, is much higher than that of Licensed models. For example, the average Different PCI $\text{Throughput}_{\text{Exp}}$ for ''With PCI'' models is 37.39 for Licensed and 93.31 for Unlicensed. Third, not using PCI as a categorical parameter does not affect the relative difference in model $\text{Throughput}_{\text{Exp}}$ for the two PCI scenarios unlike Licensed models, although the absolute values decrease as expected in the case of ''Without PCI'' [10], [11].

The main takeaway is that the network feature relationships and performance prediction in the Unlicensed component are largely consistent in both PCI scenarios. This implies that the LTE-LAA site architecture leads to a relatively more efficient Unlicensed operation.

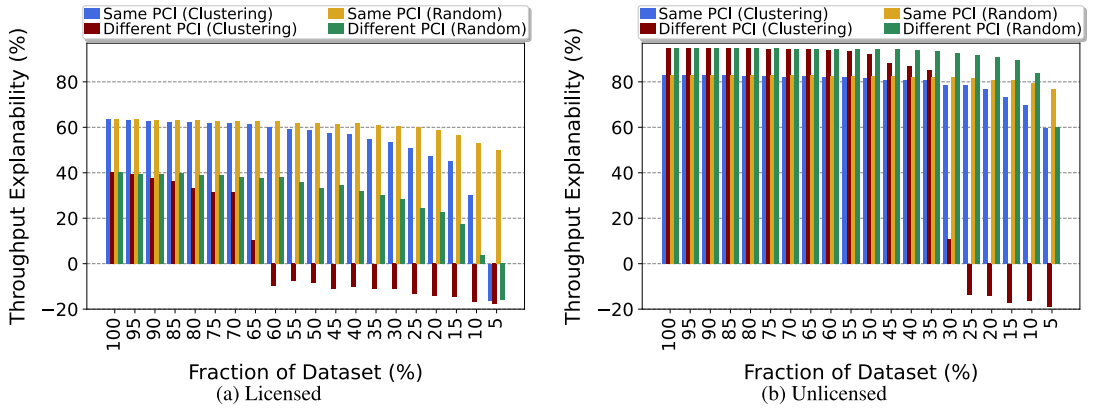


FIGURE 10. Impact of PCI architecture on numerosity reduction.

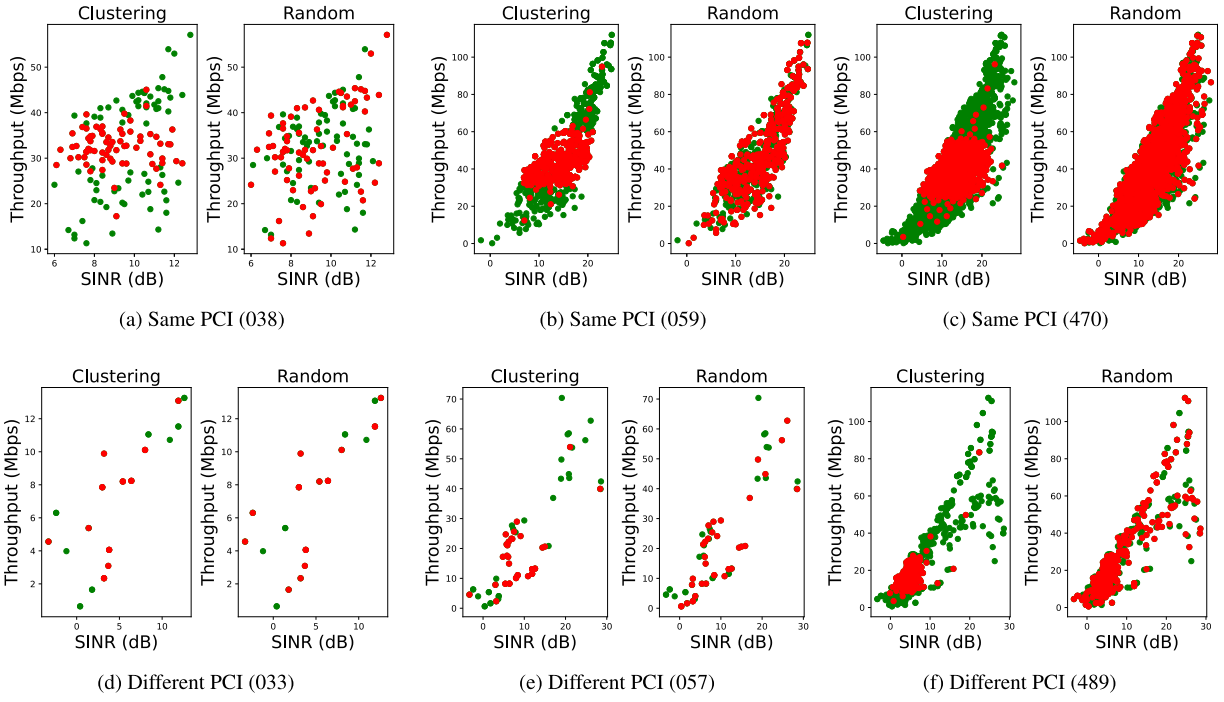


FIGURE 11. Cell-specific numerosity reduction for unlicensed PCIs.

C. PCI SCENARIOS AND NUMEROSITY REDUCTION

Numerosity reduction is a crucial tool for reducing the computational overhead of network data-driven cell selection and handover [33]. Sampling and clustering are two popular approaches to numerosity reduction [31]. The efficiency of these techniques depends on the selection of a smaller sample that resembles the characteristics and distribution of the original dataset. However, the nature and distribution of the data may vary in the two PCI scenarios, thus affecting the performance of the numerosity reduction techniques.

This problem is investigated by observing the performance of two nonparametric techniques *viz.,* Random Sampling and k-mediods Clustering, validated through kCV, for k=5 [31], [33].

1) MACRO LEVEL ANALYSIS

It can be discerned from Figures 10 (a) & 10 (b), that for both Licensed and Unlicensed, for all subsets of the original dataset, the $\text{Throughput}_{\text{Exp}}$ trends are as observed earlier. Further, Random Sampling seems to be a better numerosity reduction technique than k-mediods Clustering, for both PCI scenarios in both the coexistence components. Two differences are evident between the PCI scenarios. First, the Different PCI dataset is less than one-fourth of the Same PCI, which implies that as the reduced dataset gets smaller, the ability to predict the response variable will decrease at a faster rate in the Different PCI models than in the Same PCI models. For example, consider the average decrease in $\text{Throughput}_{\text{Exp}}$ for randomly sampled models in the Unlicensed component,

for both scenarios. In the Same PCI scenario, for data subsets ranging from 30% to 10% size of the original, the average decrease is six times that of subsets ranging from 100% to 30% of the original. In Different PCI, a similar comparison yields an 18-fold gap. Thus, for the Different PCI scenario, numerosity reduction techniques should be applied only if the original dataset has a large number of sample points. Second, PCI scenarios respond differently to the data reduction algorithms. The performance of both sampling and clustering on Same PCI data is consistent up to a 10% fraction of the original dataset for both Licensed and Unlicensed. However, for the Different PCI scenario, k-mediods Clustering does not seem to be a good choice. For Licensed, the Different PCI clustering models show negative $\text{Throughput}_{\text{Exp}}$ 60% fraction onward, while for Unlicensed, the $\text{Throughput}_{\text{Exp}}$ is negative after the 25% fraction. *This occurrence cannot be attributed to a smaller sample space of Different PCI data* as Random Sampling performs significantly better for subsets of identical sizes, and the model $\text{Throughput}_{\text{Exp}}$ is never negative even in the 10% subset.

2) PCI LEVEL MICRO ANALYSIS

At the level of individual PCIs, the impact of numerosity reduction techniques becomes more visible. Data from three different cells, with a low, medium, and large number of samples, is subjected to Clustering and Random Sampling. The results are presented in Figures 11(a), (b) and (c) for the Same PCI cells, and Figures 11(d), (e) and (f) for the Different PCI cells. Random Sampling seems to perform better, since at each stage the selected samples are more evenly distributed. When the data distribution is dispersed, as in Same PCI 038, Figure 11(a), both the algorithms yield similar results after distribution. Else, Clustering tends to capture only a subset of the overall distribution, especially for Different PCI. This changes the characteristics of the reduced sample vis-à-vis the baseline sample, leading to poor prediction performance.

Therefore, it is safe to conclude that k-mediods Clustering should be avoided in the Different PCI scenario. This is an interesting finding, as clustering algorithms are often used during the preprocessing phase of cell allocation and categorization [31]. It is worth noting that while the Same PCI scenario seems unaffected by the choice of data reduction approach, our evaluation is limited to two popular techniques, and a wider set of algorithms may yield a different view. More sophisticated algorithms including *local density-based instance selection* (LDIS) [34] can also be implemented. However, random sampling and k-mediods Clustering were chosen because an instance pruning algorithm should have a minimal computational footprint to ensure optimal performance in the downstream task (*e.g.,* learning a mapping between SINR and Throughput). A detailed discussion and analysis of popular instance pruning algorithms can be found in [35] and [34].

Nevertheless, it can be inferred from the findings that the PCI scenario influences the choice of numerosity reduction technique.

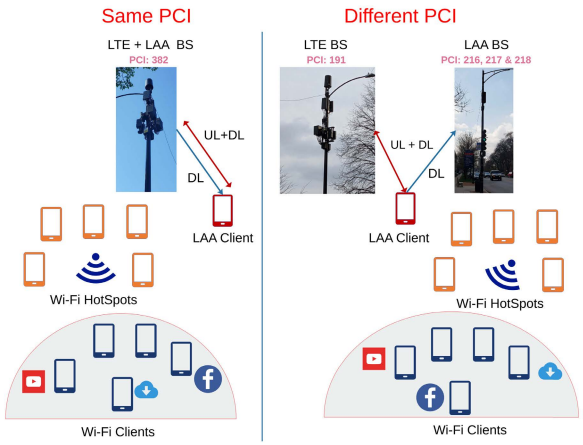


FIGURE 12. Illustration of the on-site experiment setup.

VIII. ON-SITE EXPERIMENTS, RESULTS, AND ANALYSIS

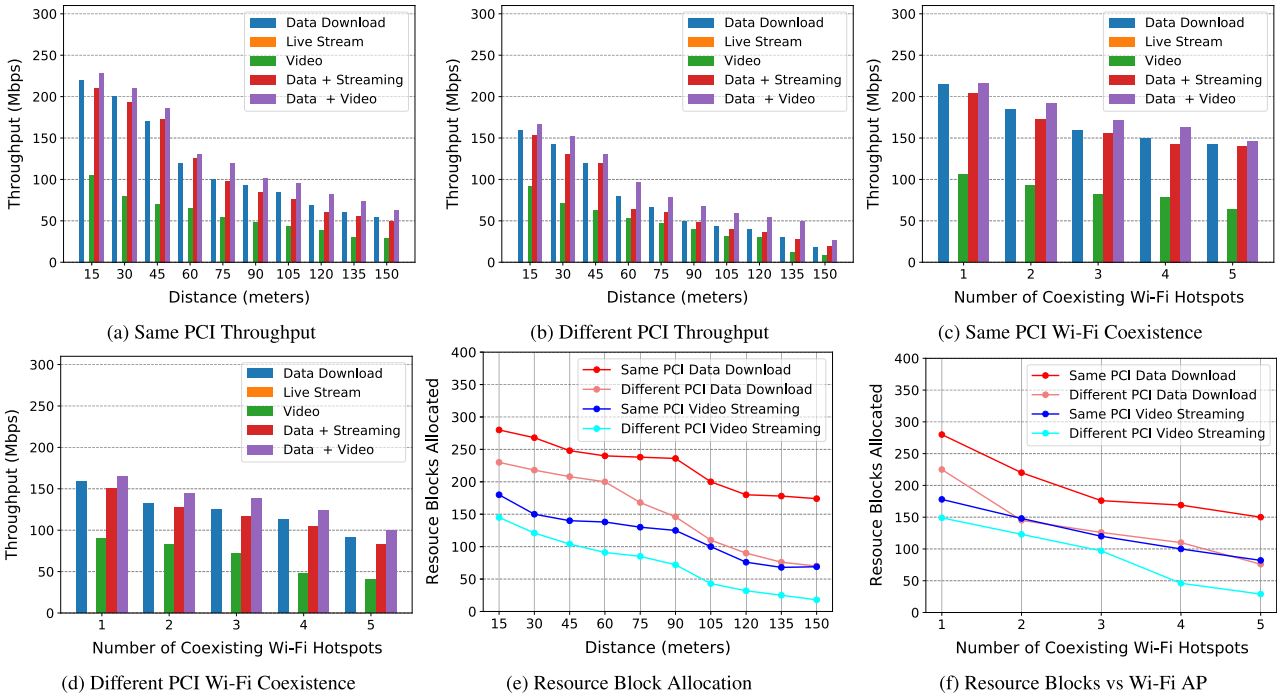
Next, a measurement-based analysis of real-time network performance is carried out for the two PCI scenarios. The onsite experiments and measurements involve several challenges, such as the identification of the two PCI scenarios in operator deployments, and creating a controlled LTE-WiFi coexistence environment. Where applicable, inferences drawn from operator data analysis are compared with measurement-based observations.

A. EXPERIMENT SETUP AND SITE SELECTION

An increasingly large number of Release 13 compliant LAA Base Stations (BS) are being deployed in downtown Chicago. Further, up to three Wi-Fi channels can be aggregated by an LAA BS in the U-NII 1 band (Channels 36, 40, & 44) or the U-NII 3 band (Channels 149, 153, & 157). It is capable of 2×2 MIMO transmissions with a maximum modulation coding scheme (MCS) of 256 QAM. Consequently, PCI scenario-specific site determination is a challenging task. The first step involves identifying the sites where the LAA & LTE BS are mounted on the same pole (Same PCI) and where the LAA & LTE BS are mounted on different poles (Different PCI), as shown in Figure 12. This is done using the NSG app, by monitoring the PCI assigned to the LTE and LAA components. Thereafter, the signal strengths (e.g., SINR and RSRP) are monitored and compared to verify the sources.

B. METHODOLOGY

The variation in real-time network performance in the two PCI scenarios is observed and compared to the findings of the data analysis. The traffic type and the LTE-WiFi coexistence environment are varied for a comprehensive evaluation of the impact on network performance. NSG app, with root permission, was used to observe performance metrics such as SINR, Throughput, Resource Block (RB) allocation for each channel, TXOP, *etc.* The observations were made at 15 minute intervals (on ground, not point-to-point) up to a distance of 150m from the LAA small cell.



(a) Same PCI Throughput (b) Different PCI Throughput (c) Same PCI Wi-Fi Coexistence

(d) Different PCI Wi-Fi Coexistence (e) Resource Block Allocation (f) Resource Blocks vs Wi-Fi AP

FIGURE 13. **Unlicensed on-site measurements.**

1) TRAFFIC TYPES AND QUALITY OF SERVICE

Several realistic traffic scenarios are considered, as LAA categorizes different types of data traffic into different *access classes*. Further, the Listen Before Talk (LBT) mechanism prescribes different duration of transmission opportunity (TXOP) to different traffic access classes, presented in Table 1. For example, background data (Access Class = 4) is offered a TXOP of 8ms, the maximum for any type of traffic, to maximize network capacity. Likewise, Wi-Fi defines specific *access categories* for different types of traffic, presented in Table 2. The Wi-Fi TXOP duration is smaller (2.528ms for background data), which implies that LAA can occupy the transmission channel for a longer duration on average. Most importantly, categorizing traffic is vital to meet QoS guarantees and is done through a QoS Class Identifier (QCI) which assigns a maximum packet delay threshold to each traffic class. The permitted delay determines which component (Licensed or Unlicensed) the traffic pipeline will be assigned to. The traffic types considered in the evaluation are:

- Data download: A fully bufferable large (>10 GB) YUV datafile from Derf Test Media Collection [36].
- Video: A 1920×1080 resolution Youtube video with a 12 Mbps bit-rate.
- Live Streaming: A 1280×720 resolution, 7.5 Mbps bitrate live stream on Youtube.
- Data & Video, Data & Streaming traffic combinations.

2) CONTROLLED COEXISTENCE ENVIRONMENT

LTE-LAA sites were identified where no interfering Wi-Fi signals were detected at the time of conducting the experiments. It was also verified (using NSG) that more than 95% of the available LAA RBs were allocated to the LAA client device. This can be attributed to fewer active users due to COVID-19, and availability of a limited number of LAA devices in the market which are also quite expensive. These factors helped us create a controlled coexistence environment. Measurements for Same PCI and Different PCI scenarios in Licensed and Unlicensed components were done with one active Wi-Fi hotspot to ensure true LTE-WiFi coexistence. Thereafter, different coexistence scenarios were created by incrementally introducing up to five mobile Wi-Fi hotspots. All Wi-Fi hotspots were operational on the same channel as LAA. Ten mobile devices including Google Pixel (2, 3, & 5), Motorola Edge+, and Samsung Galaxy S9 were used; five to create Wi-Fi hotspots and the remaining five to act as clients. A Google Pixel 5 served as the LTE-LAA client. For maximal impact, a dense network scenario was created with all Wi-Fi hotspots within a 10m range of each other and the target UE.

C. RESULTS AND ANALYSIS

1) UNLICENSED COMPONENT

The variation in Throughput with distance for five types of traffic and one coexisting Wi-Fi hotspot is shown in Figures 13(a) & 13(b). It is evident that the network performance in the Same PCI scenario far exceeds that in the Different PCI scenario (by 76.93% on average for background data). As expected, the Throughput for all types of traffic decreases with distance, due to the corresponding decrease in SINR. Likewise, with the incremental introduction of Wi-Fi hotspots (*i.e.,* Co-channel interference), the network Throughput consistently decreases, which is evident from

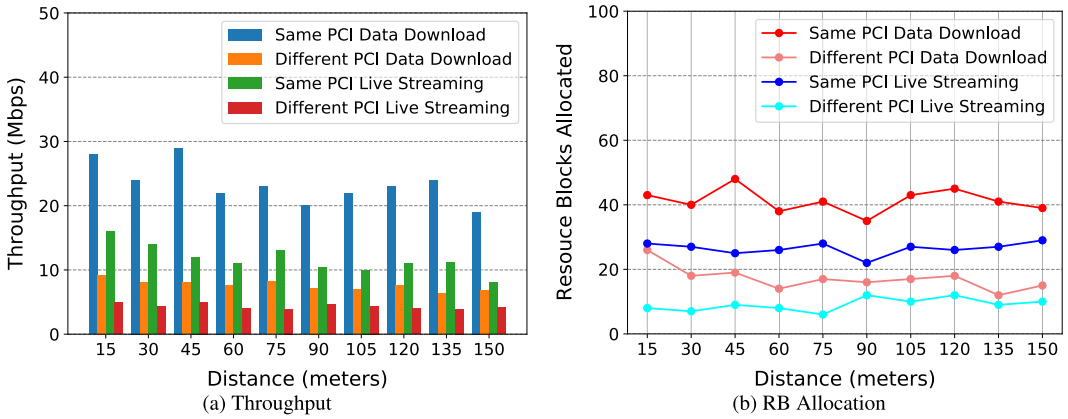


FIGURE 14. Licensed on-site experiments.

Figures 13(c) & 13(d). This can be attributed to the combined decrease in SINR and RB allocation, which will be discussed later. The following findings are noteworthy.

- *Correlating Throughput$_{Exp}$ with network performance*: Although the Throughput$_{Exp}$ of SINR-Capacity model for the Different PCI scenarios are higher than those of the Same PCI, the reverse is true for performance in terms of actual network Throughput. This cannot be attributed to a variation in SINR, which is comparable for both scenarios. From a network monitoring standpoint, it can be discerned from Figure 13(e) that RB allocation for data download and video streaming in Same PCI is higher than the corresponding Different PCI functions by 65.19% and 98.63%, respectively. With similar levels of SINR, a higher RB allocation explains the greater Throughput in Same PCI despite low model Throughput$_{Exp}$. However, from a data analysis perspective, the opposing trends imply that in Same PCI scenarios network variables other than SINR, Capacity, and PCI (e.g., RB allocated, QCI, etc.) play a more significant role in network model. Including these variables in the prediction of network performance may offer trends consistent with measurement analysis [10], [11]. They may also yield more robust network models with higher Throughput$_{Exp}$ and reliable feature relationship equations for use in network optimization [10], [37]. However, adding more network features will increase the computational overhead of training ML models.
- *Impact of coexisting Wi-Fi APs*: Figure 13(f) shows that increasing the coexisting Wi-Fi hotspots from one to five not only degrades the SINR due to increased interference, but also reduces the RB allocation due to increased contention for channel access.

2) LICENSED COMPONENT

Throughput and RB allocation results for the Licensed on-site experiments are presented in Figures 14(a) & 14(b), respectively. The Same PCI and Different PCI performance trends in the Licensed component are similar for both operator data analysis and network experiments. The Same PCI scenario outperforms the Different PCI, in terms of network performance prediction, network throughput, and RB allocation, for comparable levels of SINR. Compared to the Same PCI scenarios, the network performance drop (*e.g.,* Throughput & RB) in Different PCI conforms to the significantly lower model Throughput$_{Exp}$. As inferred earlier from the data analysis, Licensed operation in the current LTE-LAA deployments is not efficient in the Different PCI scenario.

3) PCI SCENARIOS AND QoS

An insightful observation on the LTE-LAA operation can be made by monitoring network performance metrics when Live Stream traffic is active. It can be discerned from Figure 13 that the Unlicensed component does not register any Throughput for Live Stream, in both Same PCI and Different PCI scenarios. Likewise, in both the Unlicensed scenarios, no RB is allocated to Live Stream.

Live Stream falls in QCI class 7, along with other delay critical and data-intensive services as online interactive gaming. The maximum permitted Packet Delay and Packet Loss Rate for QCI 7 are 100ms and $10^{-}3$, respectively. Live stream traffic is not assigned to the LAA component despite high RB availability and little contention from coexisting Wi-Fi APs, as the cellular operator cannot guarantee the QoS requirements in the Unlicensed spectrum. Consequently, as illustrated in Figure 14, the Licensed LTE component carries Live Stream traffic, for both Same PCI and different PCI scenarios. Thus, with respect to QoS guarantees, PCI Scenarios do not seem to have any bearing on the operator decision of assigning particular traffic pipelines to LTE or LAA interfaces.

4) RESOURCE BLOCKS AND LTE-LAA PERFORMANCE

A detailed comparison of Licensed and Unlicensed feature relationships is presented in [10] and [11]. However, the feature combination is limited to SINR, Capacity, and PCI. Measurement results reveal that RB allocation may significantly influence feature relationships and enhance network

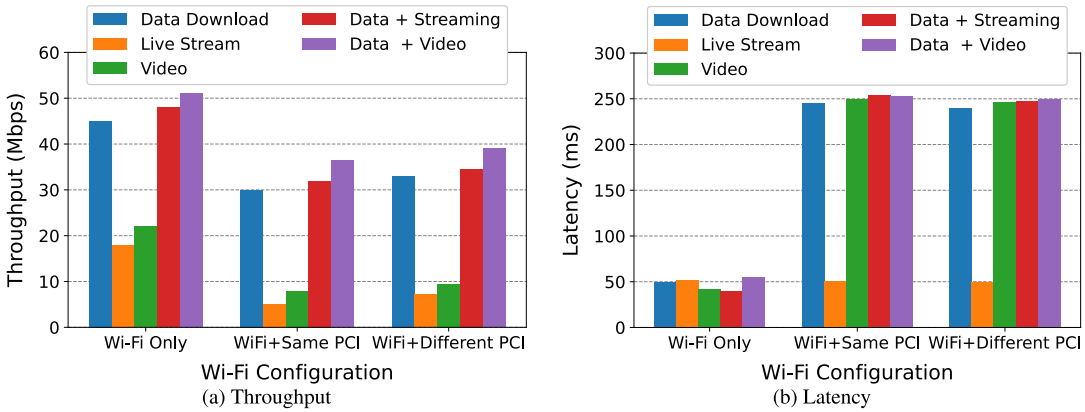


FIGURE 15. **Wi-Fi on-site measurements.**

performance prediction. In general, the Same PCI RB allocation is higher than Different PCI. However, this difference is more pronounced in the Licensed spectrum as compared to the Unlicensed. Further, despite the factors of resource sharing, contention, and back-off mechanisms in the Unlicensed spectrum, the observed RB allocation is much higher than the Licensed. This is because three 20MHz channels are available in the Unlicensed spectrum, leading up to a maximal RB allocation of 300. However, as user density increases, high availability of resource blocks may not be possible. Finally, Figure 13(e) shows that the RB allocation in the Unlicensed spectrum is adversely affected by distance, while there seems to be no evident pattern in the Licensed spectrum, as seen in Figure 14(b). However, the existence of a correlation can only be ascertained through feature relationship analysis.

5) IMPACT ON WI-FI PERFORMANCE

In this experiment, Wi-Fi AP 802.11 ac is made to coexist with LAA BS for the five different traffic flows. The closed authentication method is used so that outside Wi-Fi clients are not connected to the AP. The Wi-Fi performance is measured using Wireshark, which runs on a laptop in monitor mode so that it can capture all packets in the air corresponding to the locked channel. The Throughput and Latency for Wi-Fi baseline and coexistence scenarios is computed, and the results are presented in Figures 15(a) & 15(b). Wi-Fi performance is better for both scenarios when there is no LAA transmission. Upon activating LAA clients and initiating data traffic, it was observed that there is an immense adverse impact on Wi-Fi performance. While Throughput drops to half as compared to the baseline scenario, Latency increases by as much as four times for most traffic flows. Between Same PCI and Different PCI, the latter seems to coexist better with Wi-Fi, offering relatively better performance than the Same PCI scenario. This conforms to the LAA-side observations, as a better performance of Same PCI LAA, leads to greater contention with Wi-Fi, degrading Throughput, and increasing Latency, more than the Different PCI LAA.

The on-site experiments at LAA sites demonstrate the impact of coexistence PCI scenarios on both LAA and Wi-Fi sub-systems. The next section attempts to solve the important problem of recognizing the PCI scenario by analyzing the operator data.

IX. PREDICTIVE MODELING OF PCI SCENARIOS

Unlicensed network performance is influenced by the PCI scenario to which a UE belongs. Thus, awareness of PCI scenarios at the UE will be instrumental in implementing UE-initiated cellular operations such as cell selection, handover, broadcasts, etc. [38], [39]. A reliable solution will also be beneficial during vertical handovers in emerging 5G heterogeneous networks [40]. Thus, identification of the PCI Scenario at the UE will further improve its decision making process while performing these operations for enhanced end-user QoS.

However, there are two factors that make this task extremely challenging. First, there is the problem of power costs in UE-driven procedures, because battery drain is a major constraint in mobile devices. UE-driven cell selection and handover mechanisms must be *light-weight* with low power costs to avoid battery drain [38]. Traditionally, advanced power saving mechanisms, such as "sleep modes" are implemented to reduce battery drain [39]. Since the PCI Scenario identification requires ML based classification models that are inherently computationally expensive, the simplest feature vector conceivable in a wireless network comprising only SINR and throughput is considered in the proposed solution. The rationale being that cell selection mechanisms in LTE/LTE-A/LAA rely heavily on some measure of signal strength such as SINR or RSRP [15]. Furthermore, from the perspective of QoS guarantees, network throughput is the metric of primary interest. Learning the PCI scenario from a simple two-variable feature vector is particularly useful as it minimizes the energy budget of the computational requirements for classification, especially if it is being performed on a mobile device. However, it significantly increases the difficulty in achieving high

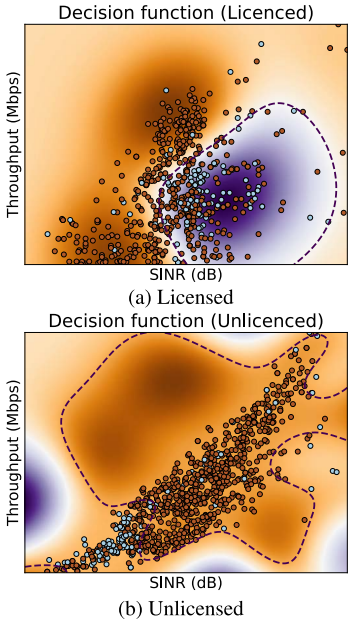


FIGURE 16. Decision boundaries.

prediction accuracy, an obstacle that the proposed solution overcomes.

Second, the inherent imbalance in the PCI scenario distribution (>80% LAA sites are Same PCI), is reflected in the network data, rendering the minority class to be a mere 16% in our data set. Together, the minimalist feature vector and extreme data imbalance make PCI scenario detection a difficult problem to solve.

A. PROBLEM SETUP

The identification of two PCI scenarios, i.e., Same or Different can be posed as a binary classification problem. In particular, the task is to learn a mapping $h : \mathbf{x}_i \rightarrow t_i$, where $\mathbf{x}_i$ is the feature vector with attributes SINR (dBm) & Throughput (Mbps). Further, t_i is 1 when the data point belongs to the Same PCI scenario and 0, otherwise.

B. LEARNING THE DECISION FUNCTION (H)

The Support Vector Machine (SVM) algorithm was deployed to learn h. In principle, other machine learning algorithms, including logistic regression, neural networks, and decision trees, can also be used with varying trade-offs in accuracy and compute time. Suppose that the data set has N training points, *i.e.,* $\{(\mathbf{x}_i, t_i)\}_{i=1}^{N}$, then the SVM algorithm aims to solve the following optimization problem:

$$\begin{aligned} &\arg\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^\top\mathbf{w} + C\sum_{i=1}^{N}\zeta_i \\ &\text{s.t.} \qquad t_i(\mathbf{w}^\top\phi(\mathbf{x}_i) + b) \geq 1 - \zeta_i, \forall i \\ &\qquad\quad \zeta_i \geq 0, \forall i \end{aligned} \tag{1}$$

where $\mathbf{w}$ and b are the weight vector and bias, respectively. C is a hyperparameter that is set through cross-validation. Moreover, ϕ is a function that allows us to learn non-linear boundaries. Figure 16 indicates that the data is not linearly separable. Thus, a non-linear decision function could help us make more accurate predictions as compared to a linear decision function. In Figure 16, the decision boundaries for both Unlicensed and Licensed data are presented. In the figure, 'orange' points refer to points belonging to the Same PCI scenario, whereas 'navy' points signify the Different PCI scenario. The SVM algorithm was used with the RBF kernel to learn the decision boundaries.

The variation of Macro f-scores for the task of PCI scenario classifications with different kernel functions used by the SVM algorithm is presented in Table 3. The scores are computed via k-fold cross validation, where $k = 5$. The Radial Basis Function (RBF) yields the best accuracy in both settings.

C. DATA IMBALANCE

Data analysis reveals that the data is imbalanced in favor of the Same PCI scenario – 83.51% data points are from the Same PCI scenario, whereas 16.49% data points are from different PCI. It is important that a classification model performs well on both classes and not just on the majority class. If this is not ensured, the Same PCI scenario will be detected with high accuracy, but a Different PCI sample may also be misjudged as Same PCI. From the perspective of UE initiated cell selection or handover, when a UE wishes to camp on a small cell with Different PCI architecture (e.g., to experience better Wi-Fi performance), it is very likely to erroneously camp on the Same PCI cell. Figure 17 shows that the standard classifier performs well only on the majority class. Therefore, it becomes crucial to explicitly handle the issue of data imbalance. Popular approaches handle data imbalance by introducing class weights or by down-sampling the majority class [41]. In particular, we used the inverse frequency as the class weight for both Unlicenced and Licenced data. In practice, class weights can be readily incorporated in to eq. 1. The class-imbalance requires us to adapt the evaluation metric in addition to the training pipeline as discussed below.

D. RESULTS AND ANALYSIS

A trivial algorithm that always predicts the majority class can lead to 83.51% accuracy on the classification task. Note that, the accuracy is defined as the percentage of points that were correctly classified by the model. However, such a model would be clearly undesirable, as it does well only on the majority class. Hence, it becomes important to look at f-score, confusion matrix, and receiver operating characteristic (ROC) curve for an exhaustive evaluation of the classification model.

The confusion matrices in Figure 17 demonstrate that class weights allow for more balanced predictions on both classes, compared to the standard classifier. Learning a classifier with class weights (second row) leads to accurate predictions on both classes, whereas the standard

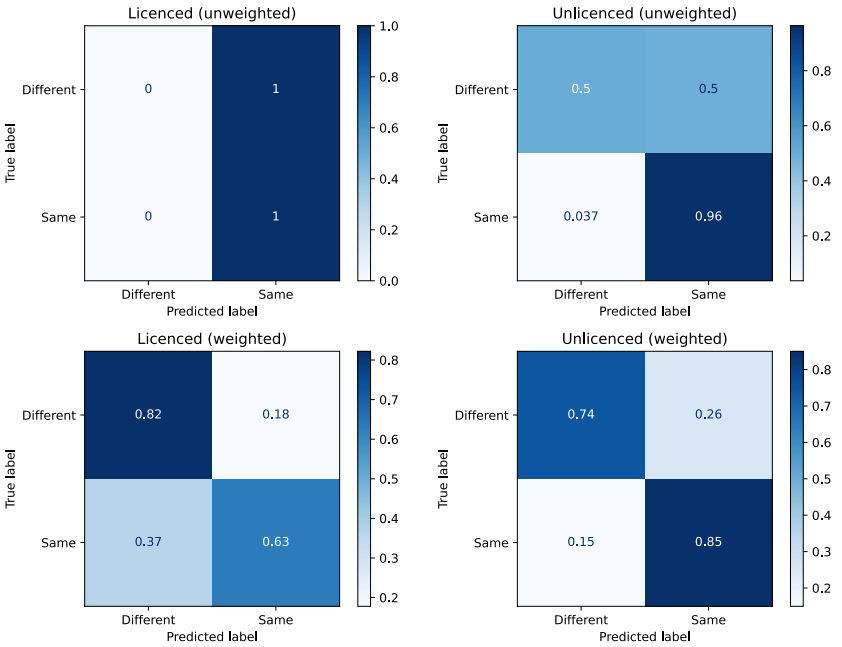


FIGURE 17. Confusion matrix.

TABLE 3. Macro f-scores of PCI scenario classification.

Network Component	Linear	Polynomial	RBF
Unlicenced	66.4	74.81	**75.0**
Licenced	59.0	56.2	**59.20**

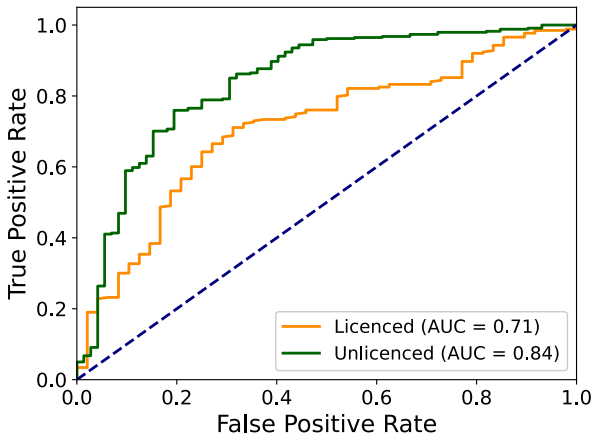


FIGURE 18. ROC curve for PCI scenario classification.

classifier (first row) performs well only on one class. In particular, the unweighted predictor yields 100% and 96% accuracy for the same PCI class for Licenced and Unlicenced data, respectively. However, the accuracy drops down to 0% and 50% on the Different PCI class for Licenced and Unlicenced data, respectively. On the other hand, the weighted predictor yields 82% and 74% accuracies on the Different PCI class for Licenced and Unlicenced data, respectively

Moreover, the ROC curve in Fig. 18 presents the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). Getting high TPR might be more important for certain applications, whereas a low FPR might be more important for others. So, the end-user may choose the configuration depending on the application based on the ROC curve. For example, when a user is mobile and needs higher performance in the Unlicensed spectrum by camping on a Same PCI cell, it is highly desirable to have a high TPR. Likewise, in an indoor setting, where the user needs higher Wi-Fi bandwidth for traffic such as Live-streams, it may be important to ensure a low FPR, for reliable identification of Different PCI Scenario. At times ensuring both high TPR and low FPR may also be necessary for latency-critical services such as augmented reality applications on the Licensed/Unlicensed component or live-streams on the Licensed
component.

The proposed PCI scenario classification solution is highly suitable for UE-driven procedures, especially in Unlicensed networks where its prediction performance is high (AUC=0.84). It is worth noting that the high accuracy is despite the use of a minimal feature set comprising only SINR and Throughput. Thus, the proposed solution offers strong prediction of the PCI scenario while ensuring energy savings at the UE, which is a prerequisite for UE-initiated handovers and cell attachments [39].

We would also like to add that for UE-driven cell selection, the distinction in the performance of Same PCI and Different PCI scenarios, will raise the questions of fairness in resource allocation. In recent past, game theory approaches have been shown to be particularly useful in solving problems of fair

coexistence [42]. Thus, to decide when a UE can attach to the Same PCI or Different PCI cell, a solution from game theory may be quite effective.

X. CONCLUSION AND WAY FORWARD

Unlicensed coexistence deployments are currently in a nascent stage and the two important factors that determine network characteristics and performance are deployment architecture and fiber optic backhaul availability. From the observations made at the LAA sites and the analysis of the dataset gathered from three major LTE-LAA coexistence service providers, we put forward some explanations and discuss a few open challenges with regard to the PCI Scenarios. First, the Same PCI for LTE and LAA components implies that the deployment has a common backhaul connectivity, which facilitates efficient splitting of Licensed and Unlicensed data. In this scenario, packet aggregation is likely to be more seamless and efficient. Consequently, the performance at the transport layer will improve, offering improved QoS to the end-user. Second, the impact of PCI as a categorical parameter when LTE and LAA components are camped on different cells is different from the Same PCI scenario. Third, resource block allocation not only differs in the Licensed and Unlicensed spectrum, but is also different for the two PCI scenarios within the same LTE-LAA component. Finally, successfully solving the PCI scenario identification problem with the minimal feature set will facilitate greater choice in the UE during the cell selection or handover process. These findings are extremely important in light of the upcoming 5G-NR Unlicensed coexistence deployments and the allocation of additional 1200MHz of Unlicensed spectrum in the 6 GHz band by the Federal Communications Commission.

Network performance measurements indicate that RB allocation is an important network feature point and considering it in the feature combination may explain the contrasting findings in the Unlicensed component. It may also improve the network performance modeling in both scenarios. We intend to extract RB data and pursue this aspect in the future.

REFERENCES

[1] V. Sathya, M. I. Rochman, and M. Ghosh, "Measurement-based coexistence studies of LAA & Wi-Fi deployments in Chicago," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 136–143, Feb. 2020.
[2] G. P. Project. (2015). *3GPP Release 13 Specification*, Accessed: Dec. 12, 2019. [Online]. Available: http://www.3gpp.org/release-13/
[3] G. M. S. Association. (2020). *LTE Unlicensed Reports*. [Online]. Available: https://gsacom.com/technology/lte-unlicensed/
[4] F. C. Commission. (2018). *Notice of Proposed Rulemaking on Unlicensed use of the 6 GHz Band*. [Online]. Available: https://docs.fcc.gov/public/attachments/FCC-18-147A1.pdf
[5] European Commision Radio Spectrum Committee. (2017). *Commission paper on a draft mandate to CEPT on RLAN in the 6 GHz band (5925–6725 MHz)*. [Online]. Available: https://circabc.europa.eu/d/d/workspace/SpacesStore/d63ea67f-8171-4619-a53d-8feb57387c27/RSCOM17-40_RLAN%206%20GHz.pdf
[6] V. Sathya, S. M. Kala, M. I. Rochman, M. Ghosh, and S. Roy, "Standardization advances for cellular and Wi-Fi coexistence in the unlicensed 5 and 6 GHz bands," *GetMobile, Mobile Comput. Commun.*, vol. 24, no. 1, pp. 5–15, Aug. 2020.
[7] S. M. Kala, V. Sathya, M. P. Kumar Reddy, B. Lala, and B. R. Tamma, "A socio-inspired CALM approach to channel assignment performance prediction and WMN capacity estimation," *J. Netw. Comput. Appl.*, vol. 125, pp. 42–66, Jan. 2019.
[8] N. Trabelsi, C. C. Chen, R. El Azouzi, L. Roullet, and E. Altman, "User association and resource allocation optimization in LTE cellular networks," *IEEE Trans. Netw. Service Manag.*, vol. 14, no. 2, pp. 429–440, Jun. 2017.
[9] V. Loginov, E. Khorov, A. Lyakhov, and I. F. Akyildiz, "CR-LBT: Listen-before-talk with collision resolution for 5G NR-U networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3138–3149, Sep. 2021.
[10] S. M. Kala, V. Sathya, E. Yamatsuta, H. Yamaguchi, and T. Higashino, "Operator data driven cell-selection in LTE-LAA coexistence networks," in *Proc. Int. Conf. Distrib. Comput. Netw.*, Jan. 2021, pp. 206–214.
[11] S. M. Kala, K. Dahiya, V. Sathya, T. Higashino, and H. Yamaguchi, "LTE-LAA cell selection through operator data learning and numerosity reduction," *Pervas. Mobile Comput.*, vol. 83, Jul. 2022, Art. no. 101586.
[12] A. Pratap, R. Singhal, R. Misra, and S. K. Das, "Distributed randomized k-clustering based PCID assignment for ultra-dense femtocellular networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 6, pp. 1247–1260, Jun. 2018.
[13] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.
[14] F. H. Khan and M. Portmann, "Joint QoS-control and handover optimization in backhaul aware SDN-based LTE networks," *Wireless Netw.*, vol. 26, pp. 1–23, Jun. 2019.
[15] S. Xu, A. Nikravesh, and Z. M. Mao, "Leveraging context-triggered measurements to characterize lte handover performance," in *Proc. Int. Conf. Passive Act. Netw. Meas.* Cham, Switzerland: Springer, 2019, pp. 3–17.
[16] V. Sathya, S. M. Kala, S. Bhupeshraj, and B. R. Tamma, "RAPTAP: A Socio-inspired approach to resource allocation and interference management in dense small cells," *Wireless Netw.*, vol. 27, pp. 1–24, Jan. 2020.
[17] V. Huang, A. Bertze, and S. Corroy, "Adaptive cell selection in heterogeneous networks," U.S. Patent 10 264 496, Apr. 16, 2019.
[18] *3GPP TSG-RAN WG2 Meeting89Bis*, document, 3GPP, Bratislava, Slovakia, Apr. 2015.
[19] A. Pratap, R. Misra, and U. Gupta, "Randomized graph coloring algorithm for physical cell ID assignment in LTE—A femtocellular networks," *Wireless Pers. Commun.*, vol. 91, no. 3, pp. 1213–1235, Dec. 2016.
[20] J. Jeon, R. D. Ford, V. V. Ratnam, J. Cho, and J. Zhang, "Coordinated dynamic spectrum sharing for 5G and beyond cellular networks," *IEEE Access*, vol. 7, pp. 111592–111604, 2019.
[21] S. M. Kala, V. Sathya, K. Dahiya, T. Higashino, and H. Yamaguchi, "Optimizing unlicensed coexistence network performance through data learning," in *Mobile and Ubiquitous Systems: Computing, Networking and Services*. Cham, Switzerland: Springer, 2022, pp. 128–149.
[22] J. Yi, W. Sun, S. Park, and S. Choi, "Performance analysis of LTE-LAA network," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1236–1239, Jun. 2017.
[23] A. M. Cavalcante, E. Almeida, R. D. Vieira, S. Choudhury, E. Tuomaala, K. Doppler, F. Chaves, R. C. D. Paiva, and F. Abinader, "Performance evaluation of LTE and Wi-Fi coexistence in unlicensed bands," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Jun. 2013, pp. 1–6.
[24] B. Bojovic, L. Giupponi, Z. Ali, and M. Miozzo, "Evaluating unlicensed LTE technologies: LAA vs LTE-U," *IEEE Access*, vol. 7, pp. 89714–89751, 2019.
[25] V. Sathya, M. I. Rochman, and M. Ghosh, "Measurement-based coexistence studies of LAA Wi-Fi deployments in Chicago," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 136–143, Feb. 2021.
[26] *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Release 16 Description; Summary of Rel-16 Work Items (Release 16)*, document, 3GPP, 2019.
[27] A. M. Baswade, M. Reddy, B. R. Tamma, and V. Sathya, "Performance analysis of spatially distributed LTE-U/NR-U and Wi-Fi networks: An analytical model for coexistence study," *J. Netw. Comput. Appl.*, vol. 191, Oct. 2021, Art. no. 103157.
[28] A. M. Baswade and B. R. Tamma, "Channel sensing based dynamic adjustment of contention window in LAA-LTE networks," in *Proc. 8th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2016, pp. 1–2.
[29] *Network Signal Guru*. (2020). [Online]. Available: https://play.google.com/ store/apps/details?id=com.qtrun.QuickTest&hl=en_US

[30] A. Abedi and T. Brecht, "Examining relationships between 802.11n physical layer transmission feature combinations," in *Proc. 19th ACM Int. Conf. Model., Anal. Simul. Wireless Mobile Syst.*, Nov. 2016, pp. 229–238.

[31] J. Moysen and M. García-Lozano, "Learning-based tracking area list management in 4G and 5G networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 8, pp. 1862–1878, Aug. 2019.

[32] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[33] S. Sundberg and J. Garcia, "Locating eNodeBs through sectorization inference—Sector fitting evaluated on a railway use case," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107945.

[34] J. L. Carbonera and M. Abel, "A density-based approach for instance selection," in *Proc. IEEE 27th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2015, pp. 768–774.

[35] D. R. Wilson and T. R. Martinez, "Instance pruning techniques," in *Proc. Mach. Learn., 14th Int. Conf. (ICML)*, 1997, pp. 404–411.

[36] *Derf's Collection*. [Online]. Available: https://media.xiph.org/video/derf/

[37] S. M. Kala, V. Sathya, S. Winston K. G., and B. R. Tamma, "CIRNO: Leveraging capacity interference relationship for dense networks optimization," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6.

[38] K. Chounos, S. Keranidis, A. Apostolaras, and T. Korakis, "Fast spectral assessment for handover decisions in 5G networks," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–6.

[39] I. Ashraf, F. Boccardi, and L. Ho, "Power savings in small cell deployments via sleep mode techniques," in *Proc. IEEE 21st Int. Symp. Pers., Indoor Mobile Radio Commun. Workshops*, Sep. 2010, pp. 307–311.

[40] R. Honarvar, A. Zolghadrasli, and M. Monemi, "Context-oriented performance evaluation of network selection algorithms in 5G heterogeneous networks," *J. Netw. Comput. Appl.*, vol. 202, Jun. 2022, Art. no. 103358.

[41] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[42] A. K. Bairagi, N. H. Tran, W. Saad, Z. Han, and C. S. Hong, "A game-theoretic approach for fair coexistence between LTE-U and Wi-Fi systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 442–455, Jan. 2019.

SRIKANT MANAS KALA (Graduate Student Member, IEEE) received the M.Tech. degree in computer science and engineering from IIT Hyderabad, India. He is currently a Doctoral Researcher with the Mobile Computing Laboratory, Osaka University, Japan. He has been awarded the Employee Excellence Award by Infosys and IIT Hyderabad Research Excellence Award, in 2016 and 2017. He led his startup team to the semifinals of Ericcson Innovation Awards 2020 and the Impact Summit of Hult Prize 2021. His research interests include the domain of extended reality, unlicensed and 5G networks, applied AI/ML, venture capital investment analysis, and thermal comfort prediction.

VANLIN SATHYA received the Bachelor of Engineering degree in computer science and the Master of Engineering degree in mobile and pervasive computing from Anna University, Chennai, India, in 2009 and 2011, respectively, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology (IIT) Hyderabad, India, in 2016. He continued his career at IIT Hyderabad, where he was a Project Officer for the converged radio access network (RAN) project. He is currently a System Engineer at Celona Inc., Cupertino, CA, USA. Prior to this, he was a Postdoctoral Scholar with The University of Chicago, USA, where he worked on the issues faced in 5G real time coexistence test-bed when LTE-unlicensed and Wi-Fi try to coexist on the same channel. His research interests include interference management, handover in heterogeneous LTE networks, device to device communication (D2D) in cellular networks, cloud base station and phantom cell (LTE-B), LTE in unlicensed, and private 5G (CBRS).

KUNAL DAHIYA received the B.Tech. and M.Tech. degrees from IIT Hyderabad. He worked on large-scale visual computing applications at IIT Hyderabad. He is currently a Research Scholar at IIT Delhi and a Research Intern at Microsoft Research, India, where he works on deep extreme multi-label learning. His work has not only led to publications in leading conferences, such as ICML, CVPR, and WSDM, but has found applications in various real-world applications, including query recommendations and ads benefiting millions of users and small businesses. His research interests include extreme multi-label learning, Siamese networks, representation learning, imbalanced classification, and 5G and LAA network operator data analysis.

TERUO HIGASHINO (Senior Member, IEEE) is currently a Professor and the Vice President of Kyoto Tachibana University, Japan. He is also a Specially Appointed Professor with the Graduate School of Information Science and Technology, Osaka University, Japan. He has been studying about algorithms, software and design methodologies concerning with localization/behavior estimation of pedestrians/crowds, development of ultra-low power consumption IoT devices, CPS research for future smart and connected communities, and IT technology for disaster mitigation. Since 2018, he has been serving on the PI of Society 5.0 Project of Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Society 5.0 is a motto of Japanese Government for constructing future super smart societies, and their project aims to contribute to life-design innovation through research and development. He was a member of Science Council of Japan (SCJ), from 2014 to 2020, and the Vice President of Information Processing Society of Japan (IPSJ), from 2016 to 2018. He is a fellow of IPSJ.

HIROZUMI YAMAGUCHI (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in information and computer science from Osaka University, Osaka, Japan, in 1994, 1996, and 1998, respectively. He is currently a Full Professor at Osaka University and leading Mobile Computing Laboratory. He has been working in mobile and pervasive computing and networking research areas and has published articles in top-quality journals, such as IEEE TRANSACTIONS and *Pervasive and Mobile Computing* (Elsevier). He has served on ICDCN2021 and Mobiquitous 2021 as the General Co-Chair, and many conferences, such as IEEE PerCom as a Technical Committee Member. He was awarded Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, in 2018.