Lumos5G: Mapping and Predicting Commercial mmWave 5G Throughput

Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand A. K. Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, Feng Qian, Zhi-Li Zhang
Department of Computer Science & Engineering, University of Minnesota – Twin Cities, U.S.A.
fivegophers@umn.edu*

ABSTRACT

The emerging 5G services offer numerous new opportunities for networked applications. In this study, we seek to answer two key questions: i) is the throughput of mmWave 5G predictable, and ii) can we build "good" machine learning models for 5G throughput prediction? To this end, we conduct a measurement study of commercial mmWave 5G services in a major U.S. city, focusing on the throughput as perceived by applications running on user equipment (UE). Through extensive experiments and statistical analysis, we identify key UE-side factors that affect 5G performance and quantify to what extent the 5G throughput can be predicted. We then propose Lumos5G – a composable machine learning (ML) framework that judiciously considers features and their combinations, and apply state-of-the-art ML techniques for making context-aware 5G throughput predictions. We demonstrate that our framework is able to achieve 1.37× to 4.84× reduction in prediction error compared to existing models. Our work can be viewed as a feasibility study for building what we envisage as a dynamic *5G throughput map* (akin to Google traffic map). We believe this approach provides opportunities and challenges in building future *5G-aware apps.*

CCS CONCEPTS

• Networks → Mobile networks; **Network measurement**; **Network performance analysis**; **• General and reference → Measurement**; **Estimation**;

KEYWORDS

5G, machine learning, deep learning, throughput prediction, mmWave, prediction, bandwidth estimation, Lumos5G

ACM Reference Format:
Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand A. K. Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, Feng Qian, Zhi-Li Zhang. 2020. Lumos5G: Mapping and Predicting Commercial mmWave 5G Throughput. In *ACM Internet Measurement Conference (IMC '20), October 27–29, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3419394.3423629

*Corresponding authors: {arvind,eman,zhzhang@cs.umn.edu}, fengqian@umn.edu.

IMC '20, October 27–29, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8138-3/20/10…$15.00
https://doi.org/10.1145/3419394.3423629

1 INTRODUCTION

With new radio (NR) specifications (5G NR [19]) that cover a wide spectrum of frequencies from low-band, to mid-band and high-band with flexible waveform, the 5th generation (5G) cellular technology is envisaged to offer a whole gamut of new services from Ultra-Reliable Low-Latency Communication (URLLC) and massive Machine Type Communication (mMTC) to enhanced Mobile Broadband (eMBB) services[1]. Exciting new applications enabled by these services include (Industrial) Internet-of-Things (IoT), autonomous driving, Augmented/Virtual Reality (AR/VR), and ultra-HD interactive video services.

2019 saw the commercial deployment of 5G services in US and worldwide, with majority using 5G NR mid-band and low-band frequencies and a few using 5G NR millimeter wave (mmWave) high-band frequencies (*e.g.,* Verizon in US). We are particularly interested in mmWave 5G performance for several reasons. First of all, the ultra-high bandwidth (theoretically up to 20 Gbps) of mmWave 5G offers exciting new opportunities to support a variety of emerging and future *bandwidth-intensive* applications expected of the 5G eMBB service. On the other hand, from theoretical analysis, simulation studies, controlled experiments and limited field testing, it has been widely known that there are many technical challenges facing mmWave radios (see, *e.g.,* [31, 33, 39, 40, 51, 66, 68, 69] and the references therein), making designing and managing 5G services based on mmWave radio a daunting task. For example, due to the directionality and limited range of mmWave radio and its high sensitivity to obstructions (*e.g.,* surrounding buildings, moving bodies, foliage, *etc.*), establishing and maintaining a stable communication link with user equipment (UE) can be difficult, especially when the UE is moving around [17].

Indeed, our very recent measurement study of commercial 5G services in US [47] has shown both the exciting new opportunities offered by mmWave 5G but also the difficult challenges involved: **(i)** commercial mmWave 5G services can offer ultra-high bandwidth (up to 2 Gbps) which makes it possible to support new classes of applications with high bandwidth requirements; **(ii)** the challenges lie in that 5G performance can fluctuate wildly over time and from one location to another, reaching as high as 2 Gbps but sometimes quickly dropping below that of 4G or to nearly zero (5G "dead zones"). For illustrative purposes, Figs. 1 and 2 summarize these

[1]In terms of bandwidth, 5G low-band offers similar capacity as 4G (less than 100 Mbps). 5G mid-band offers bandwidth usually in the range of 100–400 Mbps similar to that of advanced 4G LTE. Both low-band and mid-band radio signals are largely omni-directional, thus providing large coverage areas without requiring *line-of-sight* (LoS) to user equipment (UE). However, with flexible waveform and numerology, they are expected to be more reliable and capable of providing 1ms latency to support URLLC and mMTC services. URLLC and mMTC, together with eMBB supported primarily by 5G NR high-band, form the three new 5G services as envisioned by the ITU-R.

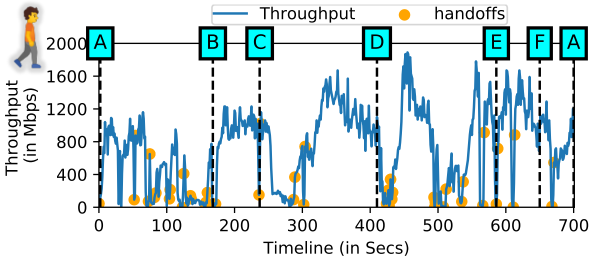


Figure 1: Walking in a Loop.

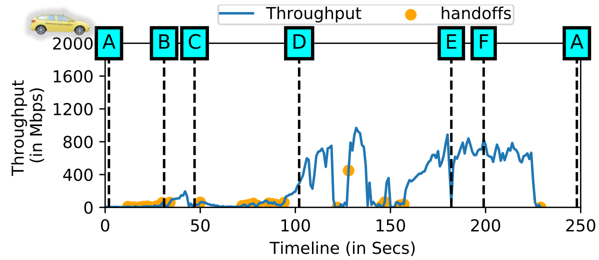


Figure 2: Driving in a Loop.

observations. Given these findings, it is therefore natural to ask: what new tools and mechanisms are needed to help emerging eMBB applications to effectively leverage the ultra-high bandwidth offered by 5G while mitigating its challenges, thus making them *5G-aware*?

In this paper, we conduct a systematic measurement study to characterize and map mmWave 5G throughput performance, with the goal to develop effective tools for predicting 5G throughput. As further articulated in §2, we focus on 5G throughput measurement (as opposed to, *e.g.,* "low-level" signal strength measurement) as this is what matters to emerging applications such as 4K/8K, 360° and volumetric video streaming, cloud gaming, AR/VR, which require ultra-high bandwidth. Tools for throughput predictions [20, 26, 44, 45, 54, 58, 63, 70] are essential to these applications, *e.g.,* in aiding them in bandwidth adaptation to maximize user quality-of-experience (QoE). The diverse impact factors and their complex interplay on 5G performance also call for data-driven, machine learning (ML) tools for throughput prediction.

Our work can be viewed as a "feasibility" study for building what we envisage as a (dynamic) *5G throughput map* (akin to a Google traffic map) that not only depicts 5G coverage but also feeds variegated throughput performance information to mobile applications over time; furthermore, it captures and incorporates key impacting factors specific to a user's environs and context *in the form of downloadable ML models*. Such a throughput map augmented with the ML models can then aid a 5G-aware application to, *e.g.,* select the initial bitrate for video streaming [27, 62, 70], and predict future throughput for rate adaptation (see §2.2 for potential use cases). We recognize that 5G deployment is still in its infancy, the measurement findings and tools developed in this study will need to be revised and evolved over time. Nonetheless we believe that this is also the perfect time to consider the design and development of tools that can be incorporated in user-side systems and apps, *making them 5G-aware*. While we focus on the user side, our findings and ML models can also help 5G carriers in improving their 5G services (see §8 for details).

Key Contributions and Roadmap: In the following we summarize the key contributions of our work, which also serve as a roadmap to the technical part of the paper.

• Due to lack of publicly available tools/APIs for 5G measurements, we develop our own 5G service monitoring and throughput measurement platform (**§3.1**). We use our tools to conduct extensive, and repeated on-field experiments for 5G throughput data collection in a large U.S. city. While ensuring data quality remains high, we carefully design systematic measurement methodologies to measure 5G throughput under various settings (indoor/outdoor, mobility scenarios, *etc.*). After cleaning, our dataset contains throughput samples captured by walking over 331 km and driving over 132 km (**§3.2**), part of which have been made publicly available on our website: `https://lumos5g.umn.edu`.

• To understand their potential impact on 5G throughput, we identify several UE-side factors and decompose them into quantifiable factors. In **§4**, we conduct numerous empirical and statistical analysis over the factors individually to understand their impact on 5G throughput behavior and its predictability. We find that 5G throughput performance is driven by a wide spectrum of factors and their interplay that are much more complex compared to traditional cellular technologies such as 3G and 4G.

• Based on our measurement findings, we develop Lumos5G – a holistic and robust ML framework that predicts 5G throughput both qualitatively (via classification) and quantitatively (via regression). Our framework is "composable" in that it judiciously considers different feature groups (geographic location, mobility, tower-based, radio connection) as well as their combinations. This is to our knowledge the first study taking a look at the predictability of commercial 5G performance using real-world data.

• Using Lumos5G, we conduct extensive evaluations and demonstrate that it achieves accurate and reliable 5G throughput prediction, and that using 5G-specific features significantly improves the prediction results (**§6.1-§6.2**). Powered by judicious feature and ML model selection, our framework achieves an overall weighted average F1 score of up to 0.96 (with three prediction classes), and 1.37× to 4.84× reduction in throughput prediction error compared to existing approaches designed for 3G/4G (**§6.3-§7**). Finally, we reveal other interesting research opportunities consequential to our work (**§8**).

2 BACKGROUND & OVERVIEW

We first provide a quick background on the current commercially available 5G services and their measurement. We then present the rationale for our work and potential use cases.

2.1 5G Deployment & Measurement

Today's commercial 5G services are all deployed in the so-called NSA (*non-standalone*) mode: namely, 5G NR is deployed with its own antennas, but shares the 4G packet core infrastructure. As such, 5G "towers" are either co-located with or are close to 4G towers. With NSA, much of the touted 5G benefits come from 5G NR. 5G NR encompasses a wide spectrum from low-band (<1 GHz), mid-band (1-6 GHz) to high-band (>24 GHz) frequencies. Low-band and mid-band 5G form the basis of most of today's initial 5G service deployment in the world – they offer only moderately higher bandwidth than existing 4G LTE or advanced LTE services.

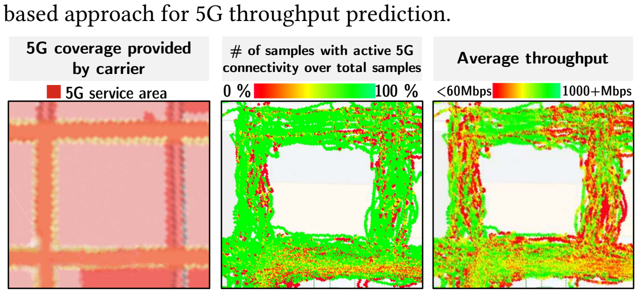
In contrast, high-band 5G – which covers the mmWave frequency bands – offers bandwidth as high as 20 Gbps *theoretically*, but considerably lower bandwidth in practice[2].

The commercial deployment of 5G services offers a new opportunity to conduct "in the field" measurement of 5G performance, especially mmWave 5G that is known to be highly sensitive to various radio signal quality impairments and environmental factors [17, 31, 33, 39, 40, 66]. Our recent work [47] conducted a measurement study of commercial 5G deployment, including mid-band and mmWave 5G services offered by several carriers in the US. It shows that commercial mmWave 5G services can deliver up to 2 Gbps bandwidth per UE, but their performance is subject to various environmental and other factors.

Building upon [47] which provides a broad, general measurement study of 5G performance from various aspects, this paper focuses on understanding the key user-side factors (*i.e.,* features) affecting mmWave 5G throughput performance, and how to build good machine learning models that can utilize such user-side features to predict 5G throughput performance. Hereafter when not explicitly stated, 5G refers to mmWave 5G. To help illustrate and motivate the problems we study in this paper, Figs. 1 & 2 show two sample 5G throughput traces under two mobility scenarios from our measurement study. We see that 5G throughput performance can vary widely and wildly from as high as 2 Gbps to as low as close to 0; user mobility and obstructions also create frequent handoffs (see §4 for further discussion). Such high variability poses many challenges for new applications that rely on the ultra-high bandwidth offered by mmWave 5G eMBB services. Our study therefore focuses on characterizing and mapping 5G throughput performance, with the goal of identifying the key (especially, UE-side) impact factors and quantifying the (short- & long-term) predictability of 5G throughput performance via repeated experiments.

2.2 Case for Mapping & Predicting mmWave 5G Throughput

As the rationale for our work, we answer two key questions: **(i)** why 5G throughput mapping is important; and **(ii)** why take an ML-based approach for 5G throughput prediction.



(a) By Carrier [13]. **(b) 5G Coverage Map.** **(c) 5G Throughput Map.**

Figure 3: 5G Performance.

• **Why 5G Throughput Mapping?** Signal strength, spectrum and channel state measurements have been widely studied in wireless and cellular networks, many from the perspective of a cellular provider, *e.g.,* for 3G/4G cellular channel scheduling. Several studies have shown that even in the case of 3G/4G networks, location alone cannot provide a good prediction of signal strength or throughput. As confirmed by our measurement results, there are far more factors affecting 5G performance. We therefore focus on 5G throughput measurement directly by building a measurement platform (an app) that can run on 5G mobile handsets (§3.1). The ability of *predicting 5G throughput with a reasonable accuracy* can help improve transport-layer mechanisms [22–24, 42] needed to address new challenges posed by 5G. It can also benefit many applications, *e.g.,* adaptive video bitrate streaming [27, 58, 62, 70]. For example, it is shown in [58] that with a prediction error $\leq$ 20%, the QoE of adaptive video streaming can be improved close to optimal (> 85)%. We believe that such an ability is more critical to emerging 5G eMBB applications that require ultra-high bandwidth. Conventional methods adopted by applications for throughput estimation and prediction have been mostly *"in situ"* in that applications either use past data transmissions or generate a few probes to estimate and predict (immediate) future throughput [45]. Some of these approaches also heavily rely on having access to PHY-layer information [20, 43, 54]. However, to address modern-day security concerns, mobile OS developers have increasingly started to restrict third-party app developers from having access to OS-level APIs which earlier provided easy access to low-level PHY-layer information [54]. Going forward, we believe that conventional methods would be inadequate for 5G applications to estimate throughput performance. Moreover, in order to predict 5G throughput with a reasonable accuracy, it is also important to capture and account for various environment, contextual, and other exogenous factors. We show that the carrier's 5G coverage map (see Fig. 3a) as well as the 5G coverage mapped by us (see Fig. 3b that shows the percentage of 5G connectivity) are insufficient to understand 5G throughput. We thus advocate building *5G throughput maps* (*e.g.,* see Fig. 3c) based on user-led (collaborative) 5G throughput measurement data. Such throughput maps not only show 5G coverage and depict 5G throughput variability over time and across different locales, but more importantly, they also incorporate (*mmWave-specific*) environmental and contextual factors (in the form of ML models) to help apps better utilize 5G's high-throughput.

• **Why ML Models for 5G Throughput Prediction?** For a long time, ML has been used for throughput prediction not only in wireless networks but also in wired networks [32, 46]. However, due to the vagary of wireless signals and the recent advancements in ML, data-driven machine learning (ML) models have become popular for 3G/4G cellular network management (see §7). Given the diverse array of impact factors and their complex interplay, the need for ML models for 5G networks is more acute. However, instead of blindly applying machine learning to the problem of 5G throughput prediction, we seek to answer a few basic questions: **(i)** Is mmWave 5G throughput predictable, and to what extent? **(ii)** What key UE-side factors (or features) most affect 5G throughput? **(iii)** In order to capture these key factors, what types of ML models are best suited for 5G throughput prediction? In particular, can we develop ML models that are *explainable*? To this end, we carefully design our measurements under various settings (*e.g.,* selecting indoor and outdoor areas, considering both stationary and mobility scenarios of various moving speeds), and conduct extensive and *repeated*

[2]High-band (especially mmWave) 5G radio signals are known to be highly directional, require line-of-sight (LoS), and have limited ranges. Particularly, they are sensitive to the environment and can be blocked by concrete structures, tinted glass, human bodies, and other moving objects [47, 66].

experiments for data collection, throughput characterization and factor analysis (see §4). Based on these results, we motivate and present our proposed ML models in §5.

2.3 Potential Use Cases

We conclude this section by illustrating some potential use cases of Lumos5G framework when in action, and its proposed 5G throughput maps and ML models.

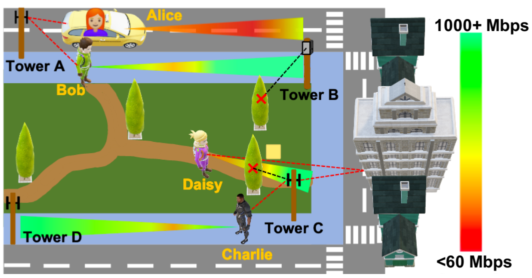


Figure 4: Lumos5G in Action.

Consider four users Alice, Bob, Charlie, and Daisy are all streaming high-resolution videos (see Fig. 4). Alice is taking a ride inside a taxi, while Bob is walking on the pedestrian street in the same direction as Alice's ride. Charlie is walking on the other side, while Daisy is walking inside the park. With Lumos5G, their UEs automatically downloads 5G throughput maps with ML models based on their geographic locations; the video streaming app interacts (via appropriate APIs) with the ML models which take into account the context and various factors such as location, moving speed & direction, type of available service[3] to predict 5G throughput (shown as a *conical heatmap*). Accordingly, the app can make intelligent decisions (*e.g.,* bitrate adaptation) to improve user QoE.

For instance, user mobility has a significant impact on 5G performance. Hence, Alice who is taking a taxi ride at a relatively high speed should expect to experience degraded performance compared to Bob who is walking along the same trajectory. Similarly, when Charlie is about to walk across a handoff patch (as learned by the model), there will be a momentary degradation in performance which the app can anticipate and prepare for. Daisy who is walking in the park does not have a clear line of sight to the 5G tower; however due to the concrete high-rise buildings around her, signals may reflect back, providing degraded 5G performance. Thus, 5G carriers can incorporate Lumos5G and its ML models to supply apps with throughput prediction by taking into account the key factors based on the user context, and aid the apps (*a la* service or content providers) in making intelligent decisions. UE can also provide feedback information to help carriers in making resource allocation and scheduling decisions based on application needs.

3 DATA COLLECTION & QUALITY

We focus on 5G throughput measurement with the goal of identifying and characterizing the key UE-side factors impacting 5G throughput performance. In the following, we list the challenges in collecting 5G throughput data, present our measurement platform, key considerations we make to ensure data quality remains high, and summarize the details of our datasets.

[3] *e.g.,* mmWave, mid-band or low-band 5G, LTE, LTE-A or LTE-CA

3.1 Our Measurement Platform

Measurement App. At the time of this study, the state-of-the-art Android OS (version 10) claims to provide access to 5G-NR related APIs [2–4]. However, none of the 5G carriers provide any meaningful responses to these APIs. With no mature tools available to collect 5G information and absence of 5G datasets, we have developed our own suite of Android app and tools for 5G performance monitoring and throughput measurement. We parse raw-string representation of Android's `ServiceState` & `SignalStrength` objects to get information about phone state, service state, and signal strength. Our app logs information sampled every second such as the UE's geolocation, orientation (*i.e.,* compass direction), tower (or panel) ID, moving speed, active radio type (*e.g.,* 5G-NR or LTE), *etc.*

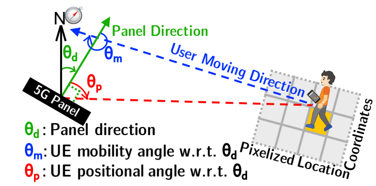


Figure 5: {Panel, Positional, Moving} Angles.

With the knowledge of the 5G panel[4] location and orientation (identified by manually surveying the area), we compute additional fields of the UE with respect to each panel to study their impact on 5G throughput. As depicted in Fig. 5, the UE-Panel distance is shown with the red line between pixelized location of UE and tower panel. The green arrow indicates the panel direction with respect to (w.r.t.) the North pole. *UE-Panel positional angle* θ_p is the angle of the UE w.r.t. panel irrespective of moving direction. *UE-Panel mobility angle* θ_m is the angle between the line normal to the front-face of the panel and the UE's trajectory. Table 1 lists all the fields that our app collects. They will be used in our subsequent measurement analysis and features for ML.

Obtaining Throughput Ground Truth. To get the throughput ground truth, our tool measures the bulk transfer throughput over 5G. We cross-compile `iPerf 3.7` [8] and integrate it into our app such that a UE is periodically downloading content from a backend server. This enables us to not only collect vital statistics about the network state, but also evaluate 5G throughput performance under different settings such as mobility mode, geolocation, *etc.* To ensure we fully saturate the available bandwidth provided by the 5G carrier, we establish 8 parallel TCP connections with the backend server, as the UE was not able to fully utilize 5G's downlink bandwidth using 1 TCP connection [47].

Prevent Internet being Bottleneck. With the ultra-high bandwidth offered by mmWave 5G, the bottleneck of an end-to-end path between a UE and the backend server (*i.e.,* the content server) may shift from the radio access network or carrier's infrastructure to the Internet. To avoid this and ensure more accurate 5G throughput measurement results, we have conducted extensive measurements using a variety of servers hosted by multiple public and private cloud providers at diverse geographical locations. We observe that factors such as server location and cloud service provider affect 5G performance. Taking cues from our prior work [47], we conduct several experiments (at least 5× 60-second runs) using servers from

[4] We observed each mmWave 5G tower deployment had one to three 5G panels or transceivers (often installed on poles) facing different directions.

Table 1: Fields Recorded by Our 5G Monitoring Tool (* Fields With an Accuracy % Provided by Android).

Field	Description
	Raw values / objects obtained from Android APIs
timestamp	logs the date and time every sec
latitude*, longitude*	UE's fine-grained geographic coordinates (*i.e.,* geolocation) & its estimated accuracy reported by Android API
detected Activity*	reports if user is walking, still, driving, *etc.* using Google's Activity Recognition API
moving speed*	reports UE's moving speed using Android API
compass direction*	The horizontal direction of travel of the UE w.r.t. North Pole (also referred to as azimuth bearing) & its accuracy
	Values obtained after post-processing or from other sources
throughput	Downlink throughput reported by `iPerf 3.7`
radio type	UE connected to 5G or 4G, identified by parsing it from raw ServiceState object
cell ID	mCid (tower identity) the UE is connected to, parsed from raw ServiceState object
signal strength	Signal strength of LTE (rsrp, rsrq, rssi) & 5G (ssrsrp, ssrsrq, ssrssi) respectively, parsed from raw SignalStrength object
horizontal handoff	UE switches from one 5G panel (cellID) to another
vertical handoff	UE switches between radio type (*e.g.,* 4G to 5G)
UE-Panel Dist.	distance between the UE and panel it is connected to
Positional Angle (θ_p)	angle between UE's position relative to the line normal to the front-face of 5G panel (see Fig. 5 for illustration)
Mobility Angle (θ_m)	angle between the line normal to the front-face of 5G panel and UE's trajectory (see Fig. 5 for illustration)

multiple cloud service providers (three public and one private). We then choose servers using the following filtering criteria: **(1)** downloading from these servers yields the highest 5G throughput (statistically) compared to servers in other locations and/or providers; and **(2)** downloading from these servers using other wired (non-mobile) hosts yields at least 3 Gbps throughput, well beyond the peak 5G throughput. Finally, to confirm the accuracy of our measurements, we also use the commercial Ookla Speedtest [11] tool to test the throughput and ensure that their results match ours, with a difference less than 5%.

Ensuring High Data Quality. GPS coordinates, compass direction, and moving speed reported by Android APIs are often inaccurate enough especially when fine granularity matters. Hence, direct usage of these values can be misleading. To ensure data quality remains high, we: **(1)** repeatedly conduct multiple measurements per trajectory on different dates and times of day to ensure the collected data is statistically representative (see §3.2), **(2)** discard data where the average GPS error (reported by the Android Location API) is greater than 5 meters along the trajectory, **(3)** add a "buffer period" at the beginning of each walk/drive test waiting for the UE to perform GPS/compass calibration, and **(4)** reduce the localization noise by discretizing raw GPS coordinates to the nearest known (pre-calculated) pixelized coordinates. The pixel coordinates are defined by Google Maps Javascript API [9] for each zoom level a Google map is viewed at. This helps create a grid over the geographic map. For instance, at zoom level 17, each pixel's spatial resolution ranges between 0.99 to 1.19 meters (~1 meter for this paper) [6, 12]. In our study, we use 17 as the zoom level as this resolution provides a nice balance without being overly precise as GPS coordinates are but at the same time represents a geographic location with a reasonable spatial resolution. Pixelized coordinates also help reduce the sparseness that exists in high resolution GPS-based coordinates. In the rest of the paper, geolocation coordinates refer to pixelized *(X, Y)* coordinates at zoom level 17.

3.2 Datasets

Spanning across a duration of 6 months, we use our 5G monitoring tool (see Table 1 for details of the recorded fields) to measure Verizon's 5G service in Minneapolis (a large metropolitan city in the U.S.) using 4× Samsung Galaxy S10 5G smartphones.

Table 2: Details About Areas.

Area	Intersection	Airport	Loop
Description	Outdoor 4-way traffic intersection	Indoor mall-area w/ shopping booths	w/ railroad crossings, traffic signals, and open park restaurants
Trajectories	12	2	2
Traj. Length	232 to 274 m	324 to 369 m	1300 m

We judiciously select three urban areas with mmWave 5G coverage (see Table 2 for summary). **(1)** **`Intersection`**: an outdoor four-way traffic intersection at the heart of Minneapolis downtown region consisting of 3 dual-panel faced 5G towers, **(2)** **`Airport`**: representing an indoor mall-area inside Minneapolis-St.Paul (MSP) International Airport with two head-on single-panel 5G towers ~200m apart, and **(3)** **`Loop`**: a 1300-meter loop near U.S. Bank Stadium in Minneapolis downtown area that covers roads, railroad crossings, restaurants, coffee shops, and recreational outdoor parks. These areas are representative as they cover indoor and outdoor environments in an urban setting.

Table 3: Full Dataset Statistics.

Data Points	563,840 (per-sec. throughput w/ feature) samples
Mobility Modes	Walking (331 km), Driving (132 km), Stationary
Data	38,632 GBs of data downloaded over 5G
Duration	6 months

For each area, we select several trajectories and perform multiple walking passes per trajectory (at least 30×). For instance, the 4-way intersection had 12 different walking trajectories. In addition to walking, we also conduct driving tests at the `Loop` area with speeds ranging between 0 to 45 kmph. Our full dataset covers 331 km walking and 132 km driving (see Table 3 for other statistics).

Ethical Consideration. This study was carried out by paid and volunteer graduate and undergraduate students. No personally identifiable information (PII) was collected or used, nor were any human subjects involved. Our study complies with the customer agreements of the wireless carrier.

4 5G THROUGHPUT MEASUREMENT & IMPACT FACTOR ANALYSIS

Using our collected dataset, we investigate how a wide range of factors affect 5G throughput. This provides insight for feature selection of our ML-based framework in §5. A summary of our findings is shown in Table 4.

4.1 Impact of Geolocation

In 3G/4G networks, geographic location is the dominant factor for indicating throughput performance [20, 25, 26, 53] or their coverage. However, as shown earlier, our initial experiments on 5G networks indicate that the throughput performance wildly fluctuates even for areas known to have 5G service. Next, we study the impact of geolocation (*i.e.,* pixelized latitude, longitude information, see

Table 4: Summary of Factors Affecting 5G Throughput and Its Predictability for the Indoor (Airport) Area (See Appendix A.1 for More Details and the Results for the Other Areas).

Results ⇒	Statistical Analysis					Simple Pred. Models				
	CV		Norm. Test	Sp. Coeff.		KNN		RF [20, 54]		Key Observations
⇓ UE-Side Factors	*(mean*	*±std. dev.)*	*(p-val. > 0.001)*	*(mean*	*±std. dev.)*	*MAE*	*RMSE*	*MAE*	*RMSE*	
(1) Geolocation	57.60%	±22.24	51.56%	0.021	±0.19	240	326	228	313	Geolocation alone is insufficient to characterize & predict 5G throughput, but it still remains a key factor.
(2) Mobility + (1) ⊦ UE-Panel Distance ⊦ UE-Panel Positional Angle ⊦ UE-Panel Mobility Angle ⊦ Moving Speed	40.24%	±20.94	78.05%	0.68	±0.14	167	247	135	201	Along with geolocation, accounting for mobility-related factors decreases variation in 5G throughput and improves its predictability.

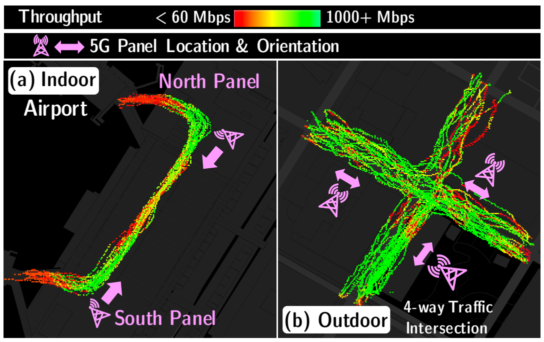
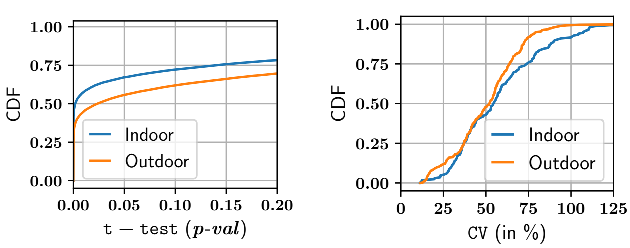


Figure 6: 5G Throughput Maps. (a) Indoor (Airport) *v/s* (b) Outdoor (Intersection)

Table 1 for details) on 5G throughput. We exemplify using data from the **Airport** and **Intersection** areas.

Fig. 6 shows the 5G throughput map visualized as a (scatter-plot) heatmap, where each point represents a grid of $2m \times 2m$ area. For each grid, we calculate the *mean* of all throughput measurements color-coded to represent different throughput levels: dark red for below 60 Mbps and lime green for above 1 Gbps. 5G panel locations (pink arrow) are also marked indicating the direction of coverage. We make the following observations: **(1)** in certain patches 5G throughput is consistently high; **(2)** in other patches 5G throughput is consistently poor, *e.g.,* due to frequent horizontal and/or vertical handoffs caused by obstructions in and around the environment; **(3)** finally, there are patches where the throughput is uncertain.



(a) Similarity of throughput samples between geolocations using pairwise t-test.

(b) Variability of throughput samples within the same geolocation using CV.

Figure 7: Measuring 5G Throughput Similarity & Variability

To quantify the throughput differences across different geolocations, we perform *pairwise* t-test and Levene test of throughput measurements for every pair of geolocation (or grid) at the airport.

Table 5: Statistical Analysis to Show the Percentage of Geolocations Whose Throughput Significantly Differs From Each Other (p-val < 0.1).

	Indoor	Outdoor	Key Observation
Pairwise t-test	70.86%	69.66%	Geolocation still matters for
Pairwise Levene test	64.26%	61.06%	5G throughput prediction.

The *p-value* results (see Fig. 7a) show that considering a significance level of 0.1, on average, the mean throughput measurements of 70.86% of geolocation pairs for indoor area significantly differ from each other. These numbers imply that geolocation is one of the key factors to capture throughput differences. Similar results for Outdoor (4-way intersection) are included in Table 5.

Next, we study the throughput variability at the same geolocation. The normality test results in Table 4 show that throughput measurements of roughly 48% of geolocations (*i.e.,* almost half the area) at the airport do *not* follow normal distribution. To reduce the false positives in detecting normal distributions, we use two types of normality tests: **(1)** D'Agostino-Pearson test [28, 29], and **(2)** Anderson-Darling test [21]. We consider the measurements associated with a geolocation as normal if they pass any of the two types. We also calculate the *mean* and *coefficient of variation (CV)* of throughput samples at each geolocation. We find that approximately 53% of geolocations have CV values ≥ 50% (see CDF in Fig. 7b), (the plots for the other tests are in Appendix A.1.1). This confirms our observation that 5G throughput varies significantly even at the same geolocation. Indeed, when we attempt to build ML models using geolocation as the only feature, we find that the models (we use KNN and Random Forest, see Table 4) yield poor accuracy – an average *MAE* and *RMSE* of ~234 Mbps and ~320 Mbps, respectively. The results indicate that ***geolocation alone is insufficient to characterize or predict 5G throughput.***

4.2 Impact of Mobility Direction

Besides geolocation, we now investigate how mobility direction affects 5G throughput. We select mobility direction as a factor since unlike omnidirectional signals in 3G/4G, 5G mmWave signals are highly directional, and sensitive to obstructions such as human body or structures [47, 55, 57, 67]. For instance, walking away from a 5G panel will naturally obstruct the UE's LoS to the 5G panel due to user's body, thus needing to acquire a NLoS reflective path.

We exemplify our finding using the **Airport** area data. We filter data representing two walking trajectories: NB (north-bound) and SB (south-bound). The data represents throughput traces collected by walking each of the two trajectories repeatedly for over 30 times. Each of the ~340-meter long walking sessions captured

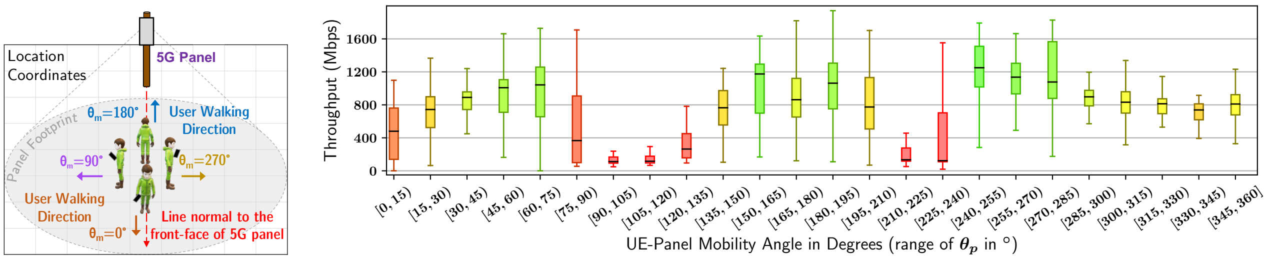


Figure 8: Impact of UE-Panel Mobility Angle θ_m on 5G throughput. See left side for illustration of different values of θ_m.

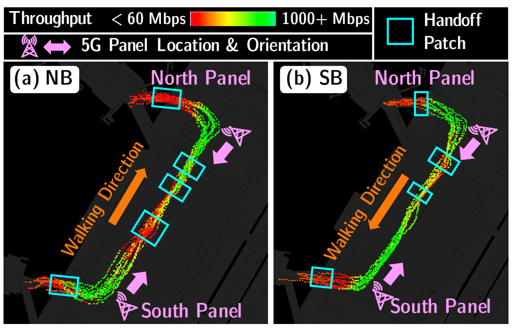


Figure 9: (a) NB *v/s* (b) SB: Airport Throughput Maps.

a ~200-second throughput trace. Fig. 9 shows both these trajectories as well as the location of the head-on 5G panels on either sides of the mall-area. We also annotate the maps with patches where handoffs usually occur (see *cyan* rectangular patches). We select airport area because both panel locations were equipped with single-sided 5G panel unlike dual-panel installations seen in outdoor environments. This ensures that we are connected to only one side of the panel, thus allowing us to understand the impact of mobility direction. 5G throughput maps for trajectories NB and SB are shown in Figs. 9a and 9b, respectively. We find that although NB and SB are in opposite directions (with partial overlap in their coverage footprints), their heatmaps are highly different, indicating that mobility direction has a significant impact on 5G throughput performance. Similar observations were made in other areas.

To further quantify the above observation, we use `Spearman's rank` correlation coefficient to measure the monotonic trend (*i.e.,* a consistent upward or downward trend) between throughput traces. The average Spearman coefficients of throughput traces belonging to NB and SB are 0.61 and 0.74, respectively. In other words, with values above 0.5, throughput traces in the same direction shows a consistent trend in increase or decrease of throughput values along the trajectory. However, the average Spearman coefficients between throughput traces belonging to different directions is only 0.021. Fig. 10 further shows the drastic increase in Spearman's coefficients by grouping traces according to their mobility directions. Similarly, 29.76% of geolocations have throughput samples with CV values greater than 50% – a decrease of 23% (see Appendix A.1.2 for extended results).

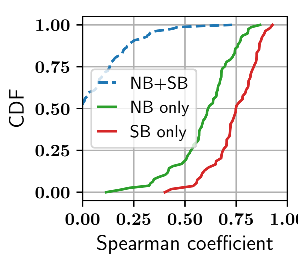


Figure 10: Impact of Mobility Trajectory.

Recall from our previous analysis that we build KNN and RF models using only the geolocation feature to predict the throughput, and got poor accuracy. Based on those models, by additionally accounting for mobility direction, we are able to reduce RMSE by 24% and 36% for KNN and RF, respectively. The results indicate that ***in addition to the absolute geolocation, further considering the movement direction leads to improved 5G throughput prediction.***

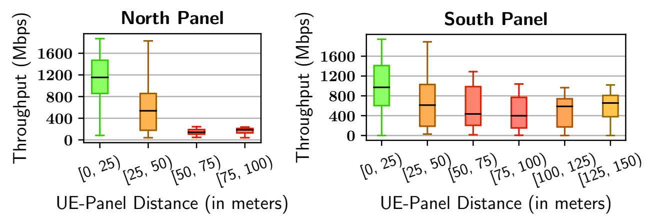


Figure 11: Varying Impact of UE-to-Panel Distance.

4.3 Impact of UE-Panel Distance

Inspired by our findings in §4.2, we now take a more detailed look at the geometric relationship among 5G panel, UE, and moving direction. We identify three geometric factors: **(1)** the UE-panel distance, **(2)** the UE-panel mobility angle (θ_m), and **(3)** the UE-panel positional angle (θ_p). We quantify their impact on 5G throughput in this subsection, §4.4, and §4.5, respectively. Due to its high frequency, mmWave signals bear high attenuation as they propagate. Therefore, as shown in Fig. 11a (the north panel at **`Airport`**), the throughput degrades fast as the distance increases. However, the detailed, quantitative distance-throughput relationship differs from one location to another due to the environmental impact. For example, Fig. 11b for the south panel at **`Airport`** shows that the throughput first (statistically) goes down and then ramps up as the distance increases. This is because there is NLoS between 50 and 100m due to obstacles (caused due to open-space restaurants and information booths) in the mall-area. The UE regains LoS beyond 100m, and the regained throughput outweighs the penalty incurred by distance increase.

4.4 Impact of UE-Panel Mobility Angle

We define the UE-panel mobility angle (θ_m) as the angle between the line normal to the front-face of 5G panel and UE's trajectory. It represents UE's movement with respect to the face of the 5G panel. As illustrated in Fig. 8, when $\theta_m = 180°$, the UE is moving head-on towards 5G panel, and 0° when walking along the same direction as the 5G panel's facing direction. Thus, if a UE is hand-held by a walking-user, $\theta_m = 0°$ will make the user's body obstruct the

LoS between UE and the 5G panel (the case in our experiments), causing performance degradation. We indeed observe this high-level trend in all three areas. However, again, different geolocations exhibit discrepancy. For example, we identify one "outlier" where $\theta_m \in [30^\circ, 75^\circ)$ at the south panel (see Fig. 18 in Appendix A.1.3 for the analysis of each panel separately). Despite the user moving away from the 5G panel, the throughput appears to be high. This is likely because the signal is properly deflected by the environment, mitigating any severe performance degradation incurred by NLoS.

4.5 Impact of UE-Panel Positional Angle

We define the UE-panel positional angle (θ_p) as the angle between the line normal to the 5G panel and the line connecting the UE to the panel. When θ_p is close to 0° ("F" in Fig. 12), the UE is in front of the panel; when θ_p is around 180° ("B" in Fig. 12), the UE is on the back side of the panel, creating a NLoS situation leading to potential performance degradation. Similarly, we can define positions such as left ("L") and right ("R"). A general trend we find is that the F position exhibits far better performance compared to the L, R, and B positions, in particular when the UE-panel distance is short, as exemplified in Fig. 13 (the south panel at **Airport**). There is a subtle difference between θ_p and θ_m. A UE with $\theta_m = 180^\circ$ need not necessitate that it is in front of the 5G panel. For instance, a UE with $\theta_p = 180^\circ$ positioned at the back ("B") of 5G panel can also have $\theta_m = 180^\circ$. In other words, as shown earlier in Fig. 5, θ_p differs from θ_m as the former considers the UE's absolute position instead of its moving direction. Thus, both these angles (θ_p and θ_m) coupled with the UE-panel distance is useful in capturing the UE's location from the 5G panel's perspective (more about these features in §5).

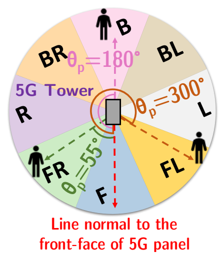


Figure 12: UE-Panel Positional Angle θ_p.

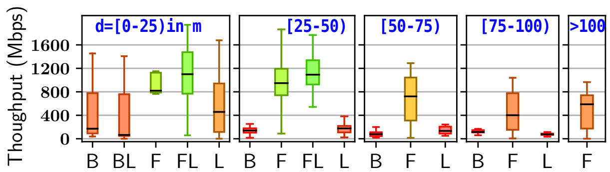


Figure 13: Impact of {Positional Angle & Distance} factors between UE and South Panel on 5G throughput.

4.6 Impact of Mobility Speeds

Mobility is a major technical challenge in mmWave 5G due to the physical layer characteristics of mmWave that make its signals highly fluctuating thus causing wild variations in performance [17, 47]. Next, we conduct experiments in the wild to investigate the impact of mobility speeds on 5G throughput. We repeatedly conduct walking and driving tests on the 1300m **Loop** area (at least 30× times). For the driving tests, we mounted the phone on the car's windshield while for the walking tests, we hand-held the phone in front of us. Located in the Minneapolis downtown region, this area covers a number of traffic/pedestrian lights, public transit rail crossings, restaurants and popular joints, high rise buildings, and a public park. Driving speeds on the loop ranged between 0 to 45 kmph while walking speeds hovered between 0 to 7 kmph. Fig. 14 shows the throughput distributions of different ground speeds (reported by Android API [16]), where each box represents 1-second samples measured for a given speed range. In the upper plot (Fig. 14a), we find that mobility under driving mode has a significant impact on 5G throughput. Statistically, the throughput decreases as the driving speed increases. Under no-mobility to very low moving speeds (<5 kmph) representing times when the car is about to stop/start or stationary (due to a traffic stop sign or a red light), the throughput peaks at ~1.8 Gbps with a median throughput of ~557 Mbps. Beyond 5 kmph, 5G performance takes a huge degradation as the median 5G throughput falls to 4G-like performance, ranging between 164 Mbps and 60 Mbps. At the same time, peak throughput for moving speeds between 5 and 30 kmph are above 850 Mbps suggesting other factors might still boost the throughput performance.

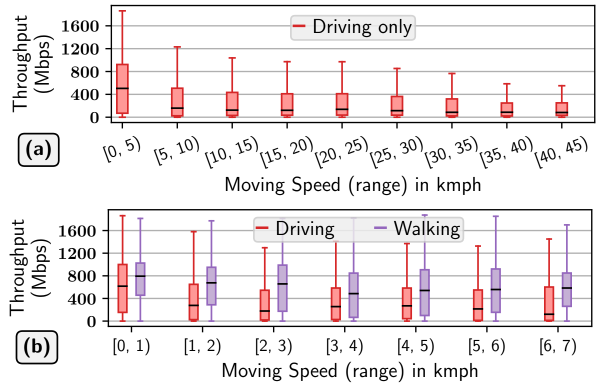


Figure 14: Impact of Mobility Speed on 5G throughput.

However, this is not the case while walking. To investigate further, Fig. 14b considers the mode of transport and shows a side-by-side throughput distribution comparison of walking *v/s* driving with a finer-grained speed range (1 kmph) per box. Compared to driving mode, we clearly find that there is little to no significant degradation in 5G throughput for walking as the speed increases. Peak throughput while walking is able to reach high levels of above 1.8 Gbps across the entire range of moving speed (*i.e.,* 0 to 7 kmph). At the same time, we also see the median throughput while walking is consistently better (by 148 to 457 Mbps) than that while driving. Such poor performance while driving is not surprising as mmWave signals need to reach the UE by propagating through the car's body (*e.g.,* windshields or side windows) that attenuates the signal strength causing throughput degradation.

Thus, this study shows that 5G throughput is also affected by a combination of effects caused not only by ground mobility speeds but also the mode of transport further highlighting the complex interplay of factors impacting 5G throughput.

Summary. Through in-the-field experiments, we reveal that numerous factors impact 5G throughput: geolocation, mobility direction, UE-Panel orientation, UE-Panel distance, UE's mobility speed, *etc.* – far more sophisticated than those impacting 4G/LTE. Instead of independently affecting the performance, these factors may cause complex interplay that is difficult to model analytically. Table 4 summarizes the statistical findings and the 5G throughput prediction accuracy using existing models. It clearly shows that accounting for UE-side mobility-related factors in addition to UE's geolocation is able to better characterize 5G throughput (thus leading to better prediction accuracy) compared to using geolocation alone. This motivates us to seek for a learning based approach for 5G throughput prediction.

5 LUMOS5G: CONTEXT-AWARE ML MODELS FOR 5G THROUGHPUT PREDICTION

Building on the insights obtained in §4, in this section we discuss the key considerations and criteria we employ for developing ML models in Lumos5G framework for 5G throughput prediction. In particular, we introduce the idea of *feature groups* to account for diverse sets of impact factors at the UE-side, and develop "composable" ML models that employ different sets of features depending on the availability of the features and usage context.

5.1 Feature Selection & Grouping

As discussed, there are a whole gamut of diverse factors that impact 5G performance, many of them, *e.g.,* channel state, various radio impairments that may be sensed by the 5G base station, are not readily available to applications running on the UE. Hence we focus on UE-side features that can be measured and collected. We will also take advantage of additional features, *e.g.,* radio type, signal strength, handoff information from the PHY layer, when available. We introduce the notion of *feature groups* by classifying features into several categories. This notion offers several benefits. **(1)** It helps account for the collective effects and interplay of similar features. **(2)** It allows us to select available and relevant features, and compose feature sets depending on the usage case (*e.g.,* stationary *v/s.* mobile scenarios). **(3)** It enables us to compare ML models with different feature combinations to investigate the importance of various feature groups under diverse settings and develop *explainable* ML models for 5G throughput prediction.

Table 6 lists four *primary* feature groups we consider in our study. `L` represents the basic *location-based* feature group which contains (pixelized) geographic location coordinates. `M` represents the basic *mobility-based* feature group which includes moving speed and compass direction (*i.e.,* azimuth angle) that can be measured using sensors on the UE. In place of location-based features, `T` represents the (more advanced) *tower-based* feature group which contains features such as the distance from a UE to the 5G panel, positional (θ_p) and mobility (θ_m) angles to the 5G panel (see Fig. 5 for illustrations). These features can be collected by the UE but rely on *exogenous* information obtained, *i.e.,* via the 5G tower location/direction information measured by us or supplied by the carrier. Despite that, ML models trained using them are likely more *transferable* to other areas with similar geolocation characteristics as the features do not depend on the absolute locations of the UEs, *i.e.,* being *location-agnostic.* `C` represents the *connection-based* feature group which includes, *e.g.,* (the immediate) *past* throughput values measured by an application and/or various low-level PHY-layer features provided by the UE, when available.

Next, in Table 6, we list four feature group *combinations* "composed" of multiple primary feature groups: **(i)** `L+M` (the *Location+Mobility* model); **(ii)** `T+M` (the *Tower+Mobility* model); **(iii)** `L+M+C` (the *Location+Mobility+Connection* model); and **(iv)** `T+M+C` (the *Tower+Mobility+Connection* model). We choose these

Table 6: Feature Groupings.

	Feature Group	List of Features
Primary	L	Pixelized Longitude & Latitude coordinates
Primary	M	UE Moving Speed + UE Compass Direction
Primary	T	UE-Panel Distance + UE-Panel Positional Angle + UE-Panel Mobility Angle
Primary	C	Past throughput measurements + (PHY features: Radio Type + LTE Signal Strength + 5G Signal Strength + Horizontal Handoff + Vertical Handoff)
Combinations	L+M	(L) + UE Moving Speed + UE Compass Direction
Combinations	T+M	UE Moving Speed + UE-Panel Distance + UE-Panel Positional Angle + UE-Panel Mobility Angle
Combinations	L+M+C	(L+M) + Radio Type + LTE Signal Strength + 5G Signal Strength + Horizontal Handoff + Vertical Handoff
Combinations	T+M+C	(T+M) + Radio Type + LTE Signal Strength + 5G Signal Strength + Horizontal Handoff + Vertical Handoff

four combinations to compare the performance of ML models using different feature groups under mobility scenarios, and to study the feature group importance in 5G throughput prediction. We consider ML models with and without connection-based features for different use cases as connection-based features require a 5G connection to be established for collecting measurement data. ML models without connection features are still useful, for example, for initial bitrate selection in adaptive video streaming. In addition to the above four combinations, other feature group combinations may also be formed to support other usage scenarios. Other primary feature groups such as *"static features"* containing information about the UE device model and specifications are also important for 5G throughput prediction. However, our study is limited to only one device model, hence we do not consider this feature group. This is a limitation of our study and we discuss more about this in §8.

5.2 Proposed ML Models

Before we present our ML models proposed as part of the Lumos5G framework, we first describe the basic settings. We formalize the 5G throughput prediction either as a *classification* problem or as a *regression* problem. We also consider the *short-term* versus *long-term* prediction problems. These settings are motivated by different use cases for the ML models.

5G Throughput Prediction as a Classification Problem. In many settings, we are interested in knowing the "level" or range of throughput a user may receive, *e.g.,* low throughput (*e.g.,* 100 Mbps) or high throughput (*e.g.,* 700 Mbps and above) or somewhere in between, given her current location and usage context. This reduces the 5G throughput prediction problem to a classification problem: given a set of features/feature groups, predict the level of 5G throughput a user can be expected to receive (similar to the the signal bars on a cellphone). This information can be used, *e.g.,* for initial bitrate selection for various applications. We consider three throughput classes: *low* (below 300 Mbps), *medium* (from 300 Mbps to 700 Mbps), and *high* (above 700 Mbps)[5].

5G Throughput Prediction as a Regression Problem. In many settings, however, we may have access to, *e.g.,* a trajectory of

[5]These levels are chosen partially based on our analysis of 5G throughput variability. As shown in §4, 5G throughput often fluctuates ± 200 Mbps, due to various "uncontrollable" random effects. Our ML models also work well with other choices of throughput classes; the results for which are omitted.

feature values measured or calculated by the UE as a user is moving along a route. Given such data, we want to predict the expected throughput value at the next time slot (*e.g.,* 1 second) or next k time slots (*e.g.,* 30 seconds). Regression-based 5G throughput prediction can aid many applications in making fine-grained decisions in the duration of an ongoing session, *e.g.,* to predict and select the quality levels for adaptive video streaming.

Short-Term *v/s.* Long-Term Prediction. In the examples cited above, throughput prediction is *short-term*, *i.e.,* in the time scales of seconds; they utilize current (or recent past) measured feature values to predict the immediate future throughput. Such short-term prediction is most useful for dynamic application decision making; ML inference must be relatively light-weight. For general 5G throughput mapping, we are also interested in longer-term prediction problems (*e.g.,* in the time scales of minutes, hours, or even days). Longer-term prediction will allow us to employ more datasets and devote more computation resources for training and inference; which can be valuable for network management and planning applications, among others.

In Lumos5G, we consider two classes of ML models, one based on a classical machine learning method – *gradient decision boosted trees* (GDBT) [30], and another based on a deep learning technique – *sequence-to-sequence* (Seq2Seq) [59] which is particularly suited for time-series/trajectory-based regression problems. We now briefly describe these two classes of ML models.

• **GDBT ML Models.** Gradient boosting is a class of ML algorithms that produces a strong prediction model in the form of a weighted combination of weak learners which optimize a differentiable loss function by gradient descent in functional space. It follows an additive multi-stage approach in which weak learners are added one at a time and gradient descent procedure is used to minimize the loss when learners are added. The weak learners are typically depth-bounded decision trees. We choose GDBT for several reasons. First, it is *lightweight*, requiring little computation power. Second, it is *composable*, allowing different sets of features (feature groups) to be easily added and combined as weak learners. Third, it can be used for both classification and regression. Fourth, it is *interpretable* as its predictive power has strong mathematical justifications and provides us with the ability to compute and analyze the *(global) feature importance*. Last but not the least, as will be shown in §6.3, it outperforms other classical machine learning methods such as Random Forest (RF) and k-Nearest Neighbors (KNN) which have been proposed for 3G/4G signal strength/bandwidth prediction problems in the literature [20, 34, 54, 60].

• **Seq2Seq ML Models.** Initially devised for natural language processing and machine translation, Seq2Seq learning has now become ubiquitous for solving various high-dimensional time series prediction problems [49, 50, 61]. Unlike the standard long short-term memory (LSTM) models [35], *Seq2Seq* allows us to model an *arbitrary* length of the predicted output sequence instead of an immediate one-time prediction, thus capable of predicting over a longer horizon into the future. Formally, let $X_t = \{x_1, x_2, \ldots, x_t\}$ be a sequence of inputs known *a priori* at time t where each x_t is a feature vector. Let $Y_t = \{y_1, y_2, \ldots, y_k\}$ be a sequence of k outputs to be predicted. In our case, Y_t is a sequence of future throughput values to be predicted over the future k time slots. The time slots are defined based on the prediction problem at hand (*e.g.,* seconds for short-time prediction, or minutes/hours for long-term prediction). In our design of the *Seq2Seq* ML models (see Fig. 15 for its illustration), we incorporate an *encoder-decoder* architecture using an LSTM-type network. Our models can work with different feature groups represented as a sequence of high-dimensional inputs.

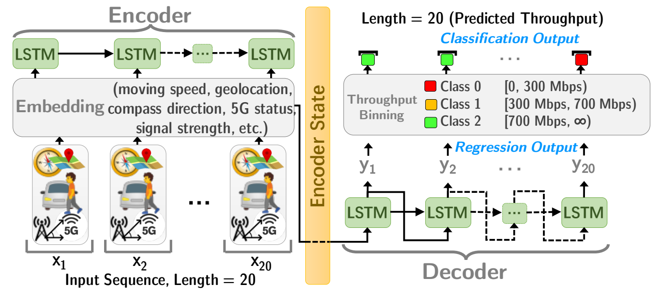


Figure 15: Seq2Seq w/ Encoder-Decoder Architecture.

6 PERFORMANCE EVALUATION

Using the proposed Lumos5G framework for 5G throughput prediction, we evaluate the performance of GDBT and Seq2Seq models using different feature groups and their combinations. We also compare our models with several other analytical and ML models proposed in the literature for 3G/4G signal strength/throughput prediction.

6.1 Evaluation Framework

We start by presenting the model setups and evaluation metrics used in our evaluation framework.

Model Setups for GDBT & Seq2Seq. We perform grid search for tuning the hyperparameters for both Seq2Seq and GDBT models using throughput traces representing a new area, thus not part of the training or testing data. Although the models were fairly robust to multiple hyperparameter values, we select a set that provided best performance. For GDBT models, we use a gradient boosting regressor (and classifier) with 8000 estimators, bounded by depth of size 8 and with 0.01 learning rate. For Seq2Seq models, we use a two-layer LSTM Encoder-Decoder architecture with 128 hidden units. We run Seq2Seq experiments for 2000 epochs, where the batch size is set to 256. The input and output sequence length is set to be 20. We keep the hyperparameters fixed throughout all our experiments. To obtain classification results, during post-processing, we additionally associate our predicted throughput with throughput class. For both GDBT and Seq2Seq, we randomly split our datasets using a 70/30 ratio for training and testing, respectively. We consider mean-squared-error (MSE) as the loss function. All experiments are run on a single machine with Intel Core i7-6850K (12-core) CPU and 2× NVIDIA TITAN V GPUs. Time to train each of the Seq2Seq and GDBT models varied depending on the area or its dataset size. The number of data points representing each area are governed by the trajectory length (see Table 2 for details). Seq2Seq took 6 to 44 hours for training each model while GDBT was comparatively much quicker taking 10-30 minutes.

Evaluation Metrics. For regression, we evaluate using standard metrics – *Mean Average Error (MAE)* and *Root Mean Squared Error (RMSE)*. For classification, we consider the *weighted average F1 score* as the main metric for evaluation. In addition, we also use

Table 7: Classification Results: Comparison of Models Using `Weighted Average F1 Score` ↑ and `Recall` ↑ Metrics.

Feature Groups ⇓ / Areas Models	4-way Intersection (Outdoor) GDBT	4-way Intersection (Outdoor) Seq2Seq	1300m Loop (Outdoor) GDBT	1300m Loop (Outdoor) Seq2Seq	Airport (Indoor) GDBT	Airport (Indoor) Seq2Seq	Global GDBT	Global Seq2Seq
L	0.79 0.60	0.86 0.71	0.58 0.74	0.65 0.56	0.79 0.88	0.83 0.85	0.78 0.73	0.73 0.46
L+M	0.91 0.85	0.94 0.89	0.79 0.88	0.89 0.92	0.91 0.95	0.91 0.94	0.90 0.89	0.93 0.92
T+M	0.91 0.85	0.95 0.93	–	–	0.91 0.96	0.93 0.96	0.91 0.89	0.94 0.93
L+M+C	0.92 0.87	0.97 0.95	0.89 0.93	0.96 0.98	0.92 0.96	0.91 0.95	0.92 0.92	0.96 0.95
T+M+C	0.93 0.87	0.96 0.94	–	–	0.92 0.97	0.91 0.95	0.92 0.93	0.95 0.95

Metrics. ↑ Weighted Average F1-score ↑ Recall of low-throughput class [0, 300)

Table 8: Regression Results: Comparison of Models Using `Mean Average Error` ↓ and `Root Mean Square Error` ↓ Metrics.

Feature Groups ⇓ / Areas Models	4-way Intersection (Outdoor) GDBT	4-way Intersection (Outdoor) Seq2Seq	1300m Loop (Outdoor) GDBT	1300m Loop (Outdoor) Seq2Seq	Airport (Indoor) GDBT	Airport (Indoor) Seq2Seq	Global GDBT	Global Seq2Seq
L	236 347	151 218	313 395	234 327	170 283	133 223	225 314	208 273
L+M	121 188	68 137	220 293	81 147	79 146	67 133	127 186	74 144
T+M	117 181	58 120	–	–	76 142	57 126	115 173	52 109
L+M+C	114 177	54 116	130 192	28 65	72 139	71 138	109 166	49 112
T+M+C	107 166	67 131	–	–	69 131	70 147	100 154	57 119

Metrics. ↓ Mean Absolute Error (MAE) ↓ Root Mean Square Error (RMSE)

recall to evaluate the low-throughput class (*i.e.,* below 300 Mbps) prediction. The recall is defined as *True Positives/(True Positives + False Negatives).* The rationale of using recall for the low-throughput class is that, misclassifying low-throughput as high-throughput may often times incur more QoE degradation (*e.g.,* a video stall) compared to misclassifying high-throughput as low (*e.g.,* only video quality degradation without a stall). Therefore, in most cases, we prefer that the low-throughput class gets a high recall value.

6.2 Results and Observations

Table 7 shows the classification results for both GDBT and Seq2Seq models under different feature groupings, while Table 8 show the regression results. Datasets collected from three areas under stationary+walking (4-way `Intersection` & `Airport`) and stationary+walking+driving (1300m `Loop`) mobility scenarios are used for training and testing. We additionally build a model by combining data from all areas with known 5G panel locations into a single dataset – referred to as `Global`. In the case of GDBT, the prediction is based only on the *current* feature values, whereas in the case of Seq2Seq, recent feature history values (*i.e.,* a sequence of feature values) are used for prediction. The *classification results* of each model in Table 7 contain two values in each cell: the *weighted average F1-score* and *recall of low-throughput class* [0, 300) Mbps – as indicated at the bottom of the table. For 1300m `Loop`, no results are reported for `T+M` and `T+M+C`, as we are unable to reliably obtain the 5G panel location information. In Table 8, we show the *regression* results of GDBT and Seq2Seq models of all the areas. Additionally, Fig. 16 shows sample regression prediction plots for `L+M+C` feature group on `Global` dataset using GDBT and Seq2Seq, with ± 200 Mbps error bounds shaded.

Key Observations. The results in Tables 7 and 8 clearly demonstrate that both Seq2Seq and GDBT are able to achieve overall good prediction results especially under feature group combinations that account for additional UE-side features beyond geolocation. Location-based feature group alone is in general inadequate to achieve high prediction accuracy, especially under

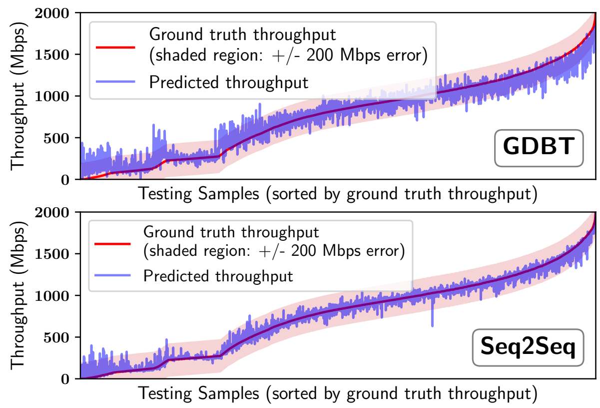


Figure 16: Regression plots for Seq2Seq and GDBT using `L+M+C` feature groups on `Global` dataset.

high mobility. By combining additional features from mobility and/or connection-related feature groups, the weighted average F1 scores for both GDBT and Seq2Seq throughput class predictions are consistently above 0.89 except for one `L+M` result for GDBT at the `Loop` area. The Seq2Seq model produces slightly better prediction results over GDBT for possibly two reasons: **(i)** in the case of throughput class prediction, Seq2Seq uses a sequence of past feature values, which indicates the benefits of incorporating history data for prediction; and **(ii)** as an LSTM-based general-purpose encoder-decoder, Seq2Seq is known to have stronger representation power [37, 59] compared to GDBT. This is best demonstrated in the regression results shown in Table 8 where for most cases Seq2Seq has far lower *MAEs* and *RMSEs.*

Transferability Analysis. Comparing feature groups – `L+M` *v/s.* `T+M` and `L+M+C` *v/s.* `T+M+C`, we see that the prediction results obtained using tower-based (`T*`) features, which are *location-agnostic,* match those using location-based (`L*`) features. A key advantage in using the `T`-based feature groups is that ML models trained on one area may potentially be *transferable* to another area if both share similar environments. To demonstrate that, at the

Table 9: Performance Comparison With Baseline Models on `Global` Dataset - Both Regression and Classification Setups.

Feature Group ⇓	KNN	RF [20]	OK[6] [26]	GDBT	Seq2Seq
Regression *(Metrics* – MAE RMSE)					
L	285 362	300 378	316 442	**225 314**	208 273
L+M	229 303	256 330	NA	127 186	**74 144**
T+M	252 326	173 253	NA	115 173	**52 109**
L+M+C	223 311	162 241	NA	109 166	**49 112**
T+M+C	228 320	163 241	NA	100 154	**57 119**
Classification *(Metric* – Weighted average F1-score)					
L	0.67	0.61	0.63	**0.78**	0.73
L+M	0.74	0.68	NA	0.90	**0.93**
T+M	0.73	0.70	NA	0.91	**0.95**
L+M+C	0.75	0.72	NA	0.92	**0.96**
T+M+C	0.73	0.75	NA	0.92	**0.95**

Model: History based Harmonic Mean (HM) [38, 64]	
Regression *(Metric* – MAE RMSE)	
Past Throughput	231 340
Classification *(Metric* – Weighted average F1 score)	
Past Throughput	0.73

`Airport` area, using the data collected from UEs connected to North panel, we train a `T+M` model. We then use that model to test the features associated with the South panel, and achieve a decent weighted average F1-score (`w-avgF1`) of 0.71. When the UE-Panel distance is less than 25m, the `w-avgF1` further increases to 0.91 as there exists high environmental similarity between the North and South panels within this range.

Feature Importance Analysis. We use GDBT's capability of reporting the features' global importance to understand how each individual feature contributes to the final prediction outcome. Overall, we find that no single feature or feature group alone dominates in predicting 5G throughput. We include a more detailed analysis of global feature importance in Appendix A.2. The results further support our argument that various factors and their complex interplay collectively affect 5G throughput.

6.3 Comparison with Existing Models

We now compare the performance of our ML models used in Lumos5G with several baseline models that have been proposed in the literature for 3G/4G performance prediction: **(1)** ***Classic ML:*** Random Forest (RF) [20], KNN; **(2)** ***Analytical:*** Ordinary Kriging (OK) [26], Harmonic Mean (HM) [38, 64]. While HM is used for short-term predictions, others have been used in the short and long term prediction contexts.

To compare classification-based models, we again use *weighted average F1-score* (`w-avgF1`) as the metric, while *MAE* and *RMSE* are used for regression. We combine all the data (*i.e.,* the `Global` dataset discussed earlier) and evaluate our models against these baselines. Table 9 shows a summary of the results. The results clearly show the superiority of GDBT and Seq2Seq models over the baseline models across all the feature groups. For instance, our regression models are able to achieve 27% to 79% reduction in MAE, while classification models show an improvement of 9% to 37% in the weighted average F1-score. History-based models such as Harmonic Mean (HM) – that typically use the immediate past throughput observations to make future predictions in real-time – also suffer due to the wild and frequent fluctuations in mmWave 5G throughput. The superiority of Lumos5G mostly stems from two reasons: **(1)** judicious feature selection by considering diverse impact factors affecting 5G throughput, and **(2)** the expressiveness of the ML models themselves, *e.g.,* the "deep" nature of the Seq2Seq model (§5.2). Results for other areas can be found in Appendix A.3.

[6]Ordinary Kriging (OK) is a grid interpolation algorithm, hence can only work on `L` feature group. For other feature groups, OK is therefore not applicable (NA).

7 RELATED WORK

Various ML-based or analytical models have been proposed for 3G/4G cellular networks. For example, Margolies *et al.* [43] incorporate UE mobility prediction in channel state estimation for 3G resource scheduling. Schulman *et al.* [56] consider UE signal strength measurement for energy-aware scheduling of user data sessions. Chakraborty *et al.* [26] employ Ordinary Kriging (OK)-based geospatial interpolation that relies on strong spatial correlations to build spectrum maps, whereas Alimpertis *et al.* [20] use Random Forest (RF) models to predict LTE signal strengths and build campus-wide (or city-wide) 4G signal strength maps. A similar effort is also seen in [54] which studies the key information needed for 4G throughput prediction. An LSTM-based deep learning model is proposed in [45] for predicting 3G/4G throughput at the immediate next time slot only. Several of these studies have pointed out that location-alone is insufficient to predict 3G/4G signal strengths/throughput performance; other factors such as mobility, indoor/outdoor, etc. must be accounted for. In the case of mmWave 5G throughput prediction, there are far more complex factors at play, and 5G throughput prediction is far harder than 3G/4G prediction. For example, due to various obstructions in an environment, there are far less spatial correlations. We cannot rely on geospatial interpolation alone to build 5G throughput maps. As we have shown earlier, the existing ML models proposed in the literature do not perform as well as our methods. To demonstrate the key differences between 5G and 3G/4G performance prediction, we have conducted a comparative study details of which can be found in Appendix A.4.

Our work further differs from existing 3G/4G ML models in several other aspects. All existing models use a fixed set of features for prediction (some of which may be missing or inaccessible by UE). Instead, by introducing *primary* and *composed* feature groups, Lumos5G framework enables to select and compose feature groups that can be readily collected and relevant to the current use case and context. Furthermore, we consider two classes of ML models in conjunction with feature grouping. This allows us to take advantage of the more powerful Seq2Seq for higher prediction accuracy, while employing *light-weight, interpretable* GDBT to investigate the feature importance and build best "explainable" ML models for 5G throughput prediction. In addition, by considering location-agnostic tower-based features, we have shown there is potential in developing *transferable* ML models that are location-independent.

8 DISCUSSION

Next, we discuss the limitations of our work and also highlight the possible future extensions.

8.1 Limitations of Our Work

With the goal to build ML models for throughput prediction, our study relies primarily on the measurement of key UE-side factors affecting 5G throughput performance, as well as other exogenous information (*e.g.,* 5G tower and panel locations) that can be gathered. We do not heavily utilize PHY-layer features due to two reasons: **(1)** no ability yet to unlock bootloaders (thus no root access) for mmWave-based 5G smartphone models supported by US carriers, **(2)** Android 10's 5G-NR APIs for accessing signal strength (*e.g.,* `getCsiRsrp()` or `getDbm()`, see [4] for API details) did not always provide meaningful data, hence not highly reliable.

Our study is also limited to one smartphone device model. Ideally, UE device model should also be included as an input feature (say, part of a feature group called *"Static Features"*). Undoubtedly, different device models and their specifications such as processor/RAM, 5G modem capabilities [7, 10, 18], antenna design [36], *etc.* will likely have high impact on the 5G throughput performance. Studying the impact of device model will be left as part of our future work. We are also unable to account for the impact of other "uncontrollable" factors such as radio resource contention[7] at RAN or congestion at the wired core network/Internet. Our study also reveals other research opportunities, including: **(1)** transferability of our proposed ML models across different areas; **(2)** temporal generalizability of such models over ultra-short (daily), short (seasonal) and long (yearly) timescales; and **(3)** sensitivity of the models to inaccuracies in input feature values. A more comprehensive study on these aspects is left as part of future work.

While we have conducted extensive measurements and experiments at various locations (both indoor and outdoor) in a large U.S. city and downloaded over 35,000 GBs of data using 5G network, the coverage of our study is still limited. Clearly, there is a need for a much larger corpus of data with increased user participation in data collection – highlighting the importance of crowdsourced platforms [48]. Nonetheless, our study demonstrates the *feasibility* of predicting mmWave throughput performance with a reasonable accuracy based primarily on UE-side factors.

8.2 Need for Collaborative Efforts

Lumos5G and its ML models are designed to predict 5G throughput with an emphasis on aiding applications using 5G services on UE. Many of the features required by our proposed models such as user location, mobility speeds, *etc.* might not always be available to application developers. This raises questions on how such models can be built or who will build them. Mobile carriers are plausible candidates, as they already collect UE-side data such as the tower UEs are connected to, radio signal strength, user data usage, *etc.* for billing and resource management purposes. Additionally, as mmWave signals are highly directional in nature [41, 52], carriers providing mmWave-based 5G service need to track user movement for beam forming and mobility management purposes. More importantly, carriers also have knowledge about their 5G network such as the location/properties of 5G panels/services, bands supported by their 5G network (*e.g.,* low-/mid-/high- band or multiband), carrier's back-haul capacity, UE congestion around the tower, *etc.* With all such information already available, 5G carriers can clearly adopt/adapt Lumos5G framework and build similar ML models.

User-Carrier Collaborative & Crowdsourced Platforms. We further suggest a *user-carrier* collaborative approach to tackle the challenges posed by 5G networks to reap the benefits offered by 5G. For example, channel state, handoff, and other information obtained at 5G towers can be used by 5G carriers for better throughput prediction which can be fed back to 5G-aware applications through APIs provided by mobile OSes such as Android, e.g., via the MOWIE mechanism proposed in [65]. Likewise, UE can provide application information such as its demands to 5G towers that can aid carriers in resource allocation and scheduling. Recognizing such needs, several formal efforts [1, 5] are underway, but are not yet fully operational. Conventionally, cellular providers have relied on their own radio signal quality testing, congestion, and coverage mapping to help configure and manage their radio access networks (RANs). The complexity of mmWave 5G and future multiband 5G [14, 15] makes such operations far more costly, if not impractical. We believe a *user-carrier collaborative*, *crowdsourced* platform is the most promising avenue to realize our envisaged 5G throughput mapping, bringing benefits to all stake holders – from 5G carriers to users/customers to application developers and providers.

Building 5G-Aware Apps. Our study points out both the opportunities and challenges in building 5G-aware apps. In particular, to tackle high bandwidth variability, new mechanisms are called for. Our preliminary study shows that existing adaptive bitrate adaption (ABR) algorithms based on throughput measurement alone do not work well for ultra-HD (*e.g.,* 8K) video streaming over 5G. Using Lumos5G for throughput prediction, we propose new rate adaptation algorithms with layered video coding, "content bursting" and multi-radio switching mechanisms. Clearly, fully exploring the issues in developing 5G-aware apps warrants another paper.

9 CONCLUDING REMARKS

We have conducted a first-of-its-kind study on understanding the predictability of mmWave 5G throughput. Despite mmWave 5G's fast attenuation and its sensitivity to environment/mobility, we find that it is indeed feasible to predict its throughput, both qualitatively and quantitatively, via a carefully designed ML framework – Lumos5G. Our experiences indicate the importance of selecting composable 5G-specific features discovered from our extensive measurements, as well as the benefits of using expressive deep learning architectures that can mine the complex interplay among the features. Our study hints the potential of developing a city-level or even country-level fine-grained "performance map" of 5G services, which can benefit numerous applications over 5G.

ACKNOWLEDGMENTS

We thank our shepherd Neil Spring and the anonymous reviewers for their insightful suggestions and feedback. We also thank Glenn Hutt, Jeff Bjorklund, Metropolitan Airports Commission and MSP Airport authorities to aid and allow us conduct our measurement study at the Minneapolis-Saint Paul International (MSP) airport. This research was in part supported by NSF under Grants CNS-1903880, CNS-1915122, CNS-1618339, CNS-1617729, CNS-1814322, CNS-1831140, CNS-1836772, and CNS-1901103.

[7] A simple and preliminary study on the impact of the number of UEs over individual UE's 5G data rate is included in Appendix A.1.4.

REFERENCES

[1] 2017. Mobile Throughput Guidance Inband Signaling Protocol. (2017). Retrieved June 2020 from https://tools.ietf.org/html/draft-flinck-mobile-throughput-guidance-04

[2] 2019. CellIdentityNr | Android Developers. (2019). Retrieved June 2020 from https://developer.android.com/reference/kotlin/android/telephony/CellIdentityNr

[3] 2019. CellInfoNr | Android Developers. (2019). Retrieved June 2020 from https://developer.android.com/reference/kotlin/android/telephony/CellInfoNr

[4] 2019. CellSignalStrengthNr | Android Developers. (2019). Retrieved June 2020 from https://developer.android.com/reference/kotlin/android/telephony/CellSignalStrengthNr

[5] 2019. Developers: It's your 5G wake-up call - time to start designing it into your apps. (2019). Retrieved June 2020 from https://www.qualcomm.com/news/onq/2019/05/10/developers-its-your-5g-wake-call-time-start-designing-it-your-apps

[6] 2019. Google Employee | Google Groups. (2019). Retrieved June 2020 from https://groups.google.com/d/msg/google-maps-js-api-v3/hDRO4oHVSeM/osOYQYXg2oUJ

[7] 2019. Huawei Launches 5G Multi-mode Chipset and 5G CPE Pro. (2019). Retrieved June 2020 from https://www.huawei.com/en/press-events/news/2019/1/huawei-5g-multi-mode-chipset-5g-cpe-pro

[8] 2019. iperf3 – iperf3 3.7 documentation. (2019). Retrieved June 2020 from https://software.es.net/iperf/

[9] 2019. Map and Tile Coordinates | Maps JavaScript API | Google Developers. (2019). Retrieved June 2020 from https://developers.google.com/maps/documentation/javascript/coordinates

[10] 2019. Snapdragon X50 5G modem-RF system. (2019). Retrieved June 2020 from https://www.qualcomm.com/products/snapdragon-x50-5g-modem

[11] 2019. Speedtest by Ookla. (2019). Retrieved June 2020 from https://www.speedtest.net/

[12] 2019. Tiles - Google Maps: Coordinates, Tile Bounds and Projection - conversion to EPSG:3785 and EPSG:4326 (WGS84) - MapTiler. (2019). Retrieved June 2020 from https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection/

[13] 2020. 5G Coverage Map: This is 5G Built Right | Verizon. (2020). Retrieved June 2020 from https://www.verizon.com/5g/coverage-map/

[14] 2020. 5G Low Latency Requirements. (2020). Retrieved June 2020 from https://broadbandlibrary.com/5g-low-latency-requirements/

[15] 2020. 5G spectrum: strategies to maximize all bands. (2020). Retrieved June 2020 from https://www.ericsson.com/en/networks/trending/hot-topics/5g-spectrum-strategies-to-maximize-all-bands

[16] 2020. getSpeed() | Location | Android Developers. (2020). Retrieved June 2020 from https://developer.android.com/reference/android/location/Location#getSpeed()

[17] 2020. Keysight Technologies | Engineering the 5G World Just Got Easier. (2020). Retrieved June 2020 from https://www.keysight.com/us/en/assets/7119-1223/ebooks/Engineering-the-5G-World.pdf

[18] 2020. Snapdragon X55 5G modem-RF system. (2020). Retrieved June 2020 from https://www.qualcomm.com/products/snapdragon-x55-5g-modem

[19] 3rd Generation Partnership Project. 2019. Release 15. (April 2019). Retrieved June 2020 from https://www.3gpp.org/release-15

[20] Emmanouil Alimpertis, Athina Markopoulou, Carter Butts, and Konstantinos Psounis. 2019. City-Wide Signal Strength Maps: Prediction with Random Forests. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 2536–2542. https://doi.org/10.1145/3308558.3313726

[21] Theodore W Anderson and Donald A Darling. 1952. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics* (1952), 193–212.

[22] A. Bakre and B. R. Badrinath. 1995. I-TCP: indirect TCP for mobile hosts. In *Proceedings of 15th International Conference on Distributed Computing Systems*. 136–143.

[23] Hari Balakrishnan, Srinivasan Seshan, Elan Amir, and Randy H. Katz. 1995. Improving TCP/IP Performance over Wireless Networks. In *Proceedings of the 1st Annual International Conference on Mobile Computing and Networking (MobiCom '95)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/215530.215544

[24] Kevin Brown and Suresh Singh. 1997. M-TCP: TCP for Mobile Cellular Networks. *SIGCOMM Comput. Commun. Rev.* 27, 5 (Oct. 1997), 19–43. https://doi.org/10.1145/269790.269794

[25] Nicola Bui, Matteo Cesana, S Amir Hosseini, Qi Liao, Ilaria Malanchini, and Joerg Widmer. 2017. A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques. *IEEE Communications Surveys & Tutorials* 19, 3 (2017), 1790–1821.

[26] A. Chakraborty, M. S. Rahman, H. Gupta, and S. R. Das. 2017. SpecSense: Crowdsensing for efficient querying of spectrum occupancy. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2017.8057113

[27] Jiasi Chen, Rajesh Mahindra, Mohammad Amir Khojastepour, Sampath Rangarajan, and Mung Chiang. 2013. A Scheduling Framework for Adaptive Video Delivery over Cellular Networks. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom '13)*. Association for Computing Machinery, New York, NY, USA, 389–400. https://doi.org/10.1145/2500423.2500433

[28] RALPH D'AGOSTINO and Egon S Pearson. 1973. Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b}$. *Biometrika* 60, 3 (1973), 613–622.

[29] Ralph B d'Agostino. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 2 (1971), 341–348.

[30] Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29, 5 (10 2001), 1189–1232. https://doi.org/10.1214/aos/1013203451

[31] Signals Research Group. 2019. A Global Perspective of 5G Network Performance. (2019). Retrieved June 2020 from https://www.qualcomm.com/media/documents/files/signals-research-group-s-5g-benchmark-study.pdf

[32] Qi He, Constantine Dovrolis, and Mostafa Ammar. 2005. On the Predictability of Large Transfer TCP Throughput. *SIGCOMM Comput. Commun. Rev.* 35, 4 (Aug. 2005), 145–156. https://doi.org/10.1145/1090191.1080110

[33] K. Heimann, J. Tiemann, D. Yolchyan, and C. Wietfeld. 2019. Experimental 5G mmWave Beam Tracking Testbed for Evaluation of Vehicular Communications. In *2019 IEEE 2nd 5G World Forum (5GWF)*. 382–387. https://doi.org/10.1109/5GWF.2019.8911692

[34] J. D. Herath, A. Seetharam, and A. Ramesh. 2019. A Deep Learning Model for Wireless Channel Quality Prediction. In *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*. 1–6.

[35] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[36] W. Hong, K. Baek, Y. Lee, Y. Kim, and S. Ko. 2014. Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices. *IEEE Communications Magazine* 52, 9 (2014), 63–69.

[37] Kurt Hornik, Maxwell Stinchcombe, Halbert White, et al. 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2, 5 (1989), 359–366.

[38] Junchen Jiang, Vyas Sekar, and Hui Zhang. 2012. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming with FESTIVE. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT '12)*. Association for Computing Machinery, New York, NY, USA, 97–108. https://doi.org/10.1145/2413176.2413189

[39] Wooseong Kim. 2019. Experimental Demonstration of MmWave Vehicle-to-Vehicle Communications Using IEEE 802.11 ad. *Sensors* 19, 9 (2019), 2057.

[40] K. Larsson, B. Halvarsson, D. Singh, R. Chana, J. Manssour, M. Na, C. Choi, and S. Jo. 2017. High-Speed Beam Tracking Demonstrated Using a 28 GHz 5G Trial System. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. 1–5. https://doi.org/10.1109/VTCFall.2017.8288043

[41] X. Liu, J. Yu, H. Qi, J. Yang, W. Rong, X. Zhang, and Y. Gao. 2020. Learning to Predict the Mobility of Users in Mobile mmWave Networks. *IEEE Wireless Communications* 27, 1 (2020), 124–131.

[42] Feng Lu, Hao Du, Ankur Jain, Geoffrey M. Voelker, Alex C. Snoeren, and Andreas Terzis. 2015. CQIC: Revisiting Cross-Layer Congestion Control for Cellular Networks. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*. Association for Computing Machinery, New York, NY, USA, 45–50. https://doi.org/10.1145/2699343.2699345

[43] Robert Margolies, Ashwin Sridharan, Vaneet Aggarwal, Rittwik Jana, NK Shankaranarayanan, Vinay A Vaishampayan, and Gil Zussman. 2016. Exploiting mobility in proportional fair cellular scheduling: Measurements and algorithms. *IEEE/ACM Transactions on Networking (TON)* 24, 1 (2016), 355–367.

[44] R. Margolies, A. Sridharan, V. Aggarwal, R. Jana, N. K. Shankaranarayanan, V. A. Vaishampayan, and G. Zussman. 2016. Exploiting Mobility in Proportional Fair Cellular Scheduling: Measurements and Algorithms. *IEEE/ACM Transactions on Networking* 24, 1 (2016), 355–367.

[45] Lifan Mei, Runchen Hu, Houwei Cao, Yong Liu, Zifa Han, Feng Li, and Jin Li. 2019. Realtime mobile bandwidth prediction using lstm neural network. In *International Conference on Passive and Active Network Measurement*. Springer, 34–47.

[46] Mariyam Mirza, Joel Sommers, Paul Barford, and Xiaojin Zhu. 2007. A Machine Learning Approach to TCP Throughput Prediction. In *Proceedings of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '07)*. Association for Computing Machinery, New York, NY, USA, 97–108. https://doi.org/10.1145/1254882.1254894

[47] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 894–905. https://doi.org/10.1145/3366423.3380169

[48] Arvind Narayanan, Eman Ramadan, Jacob Quant, Peiqi Ji, Feng Qian, and Zhi-Li Zhang. 2020. 5G Tracker – A Crowdsourced Platform to Enable Research Using

Commercial 5G Services. In *Proceedings of the ACM SIGCOMM 2020 Conference Posters and Demos (SIGCOMM Posters and Demos '20)*. Association for Computing Machinery, Virtual Event, USA. https://doi.org/10.1145/3405837.3411394

[49] Arvind Narayanan, Saurabh Verma, Eman Ramadan, Pariya Babaie, and Zhi-Li Zhang. 2019. Making Content Caching Policies 'Smart' Using the DeepCache Framework. *SIGCOMM Comput. Commun. Rev.* 48, 5 (Jan. 2019), 64–69. https://doi.org/10.1145/3310165.3310174

[50] Rohit Prabhavalkar, Kanishka Rao, Tara Sainath, Bo Li, Leif Johnson, and Navdeep Jaitly. 2017. A Comparison of Sequence-to-Sequence Models for Speech Recognition. http://www.isca-speech.org/archive/Interspeech_2017/pdfs/0233.PDF

[51] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun. 2015. Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design. *IEEE Transactions on Communications* 63, 9 (2015), 3029–3056.

[52] Maryam Eslami Rasekh, Zhinus Marzi, Yanzi Zhu, Upamanyu Madhow, and Haitao Zheng. 2017. Noncoherent MmWave Path Tracking. In *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications (HotMobile '17)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/3032970.3032974

[53] Haakon Riiser, Tore Endestad, Paul Vigmostad, Carsten Griwodz, and Pål Halvorsen. 2012. Video streaming using a location-based bandwidth-lookup service for bitrate planning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 8, 3 (2012), 1–19.

[54] A. Samba, Y. Busnel, A. Blanc, P. Dooze, and G. Simon. 2017. Instantaneous throughput prediction in cellular networks: Which information is needed?. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. 624–627.

[55] Mathew K Samimi and Theodore S Rappaport. 2016. 3-D millimeter-wave statistical channel model for 5G wireless system design. *IEEE Transactions on Microwave Theory and Techniques* 64, 7 (2016), 2207–2225.

[56] Aaron Schulman, Vishnu Navda, Ramachandran Ramjee, Neil Spring, Pralhad Deshpande, Calvin Grunewald, Kamal Jain, and Venkata N Padmanabhan. 2010. Bartendr: a practical approach to energy-aware cellular data scheduling. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 85–96.

[57] Sumit Singh, Federico Ziliotto, Upamanyu Madhow, E Belding, and Mark Rodwell. 2009. Blockage and directivity in 60 GHz wireless personal area networks: From cross-layer model to multihop MAC design. *IEEE Journal on Selected Areas in Communications* 27, 8 (2009), 1400–1413.

[58] Yi Sun, Xiaoqi Yin, Junchen Jiang, Vyas Sekar, Fuyuan Lin, Nanshu Wang, Tao Liu, and Bruno Sinopoli. 2016. CS2P: Improving Video Bitrate Selection and Adaptation with Data-Driven Throughput Prediction. In *Proceedings of the 2016 ACM SIGCOMM Conference (SIGCOMM '16)*. Association for Computing Machinery, New York, NY, USA, 272–285. https://doi.org/10.1145/2934872.2934898

[59] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[60] N. Theera-Ampornpunt, T. Mangla, S. Bagchi, R. Panta, K. Joshi, M. Ammar, and E. Zegura. 2016. TANGO: Toward a More Reliable Mobile Streaming through Cooperation between Cellular Network and Mobile Devices. In *2016 IEEE 35th Symposium on Reliable Distributed Systems (SRDS)*. 297–306.

[61] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko. 2015. Sequence to Sequence – Video to Text. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 4534–4542.

[62] Xiufeng Xie, Xinyu Zhang, Swarun Kumar, and Li Erran Li. 2016. PiStream: Physical Layer Informed Adaptive Video Streaming Over LTE. *GetMobile: Mobile Comp. and Comm.* 20, 2 (Oct. 2016), 31–34. https://doi.org/10.1145/3009808.3009819

[63] Qiang Xu, Sanjeev Mehrotra, Zhuoqing Mao, and Jin Li. 2013. PROTEUS: Network Performance Forecast for Real-Time, Interactive Mobile Applications. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '13)*. Association for Computing Machinery, New York, NY, USA, 347–360. https://doi.org/10.1145/2462456.2464453

[64] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM '15)*. Association for Computing Machinery, New York, NY, USA, 325–338. https://doi.org/10.1145/2785956.2787486

[65] Yunfei Zhang, Gang Li, Chunshan Xiong, Yixue Lei, Wei Huang, Yunbo Han, Anwar Walid, Y. Richard Yang, and Zhi-Li Zhang. 2020. MoWIE: Toward Systematic, Adaptive Network Information Exposure as an Enabling Technique for Cloud-Based Applications over 5G and Beyond (Invited Paper). In *Proceedings of the 2020 Workshop on Network Application Integration/CoDesign, NAI@SIGCOMM 2020, Virtual Event, USA, August 14, 2020*. ACM, 20–27. https://doi.org/10.1145/3405672.3409489

[66] Hang Zhao, Rimma Mayzus, Shu Sun, Mathew Samimi, Jocelyn K Schulz, Yaniv Azar, Kevin Wang, George N Wong, Felix Gutierrez Jr, and Theodore S Rappaport. 2013. 28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city.. In *ICC*. 5163–5167.

[67] Kun Zhao, Jakob Helander, Daniel Sjöberg, Sailing He, Thomas Bolin, and Zhinong Ying. 2016. User body effect on phased array in user equipment for the 5G mmWave communication system. *IEEE antennas and wireless propagation letters* 16 (2016), 864–867.

[68] K. Zhao, J. Helander, D. Sjöberg, S. He, T. Bolin, and Z. Ying. 2017. User Body Effect on Phased Array in User Equipment for the 5G mmWave Communication System. *IEEE Antennas and Wireless Propagation Letters* 16 (2017), 864–867.

[69] Yibo Zhu, Zengbin Zhang, Zhinus Marzi, Chris Nelson, Upamanyu Madhow, Ben Y. Zhao, and Haitao Zheng. 2014. Demystifying 60GHz Outdoor Picocells. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom '14)*. Association for Computing Machinery, New York, NY, USA, 5–16. https://doi.org/10.1145/2639108.2639121

[70] Xuan Kelvin Zou, Jeffrey Erman, Vijay Gopalakrishnan, Emir Halepovic, Rittwik Jana, Xin Jin, Jennifer Rexford, and Rakesh K. Sinha. 2015. Can Accurate Predictions Improve Video Streaming in Cellular Networks?. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*. Association for Computing Machinery, New York, NY, USA, 57–62. https://doi.org/10.1145/2699343.2699359

A APPENDICES

A.1 5G Throughput Performance Impact Factor Analysis: Extended Results

A.1.1 Impact of Geolocation. Fig. 17 shows extended results for the `normality` and `Levene` tests for §4.1. Using a significance value of (0.001), the `normality test` shows that throughput measurements of roughly 48% of geolocations (*i.e.,* almost half the area) at the indoor (`Airport`) do *not* follow normal distribution; similarly for the outdoor (`Intersection`) the percentage of geolocations is ∼33%. Using a significance value of (0.1), the `Levene test` shows that the variances of throughput measurements of 64.26% and 61.06% of the geolocation pairs significantly differ from each other for the Indoor (`Airport`) and the outdoor (`Intersection`), respectively.

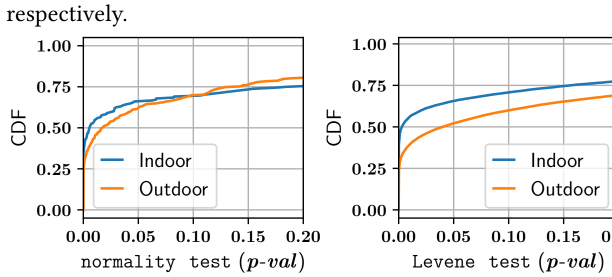


Figure 17: Indoor *v/s* Outdoor: `Normality` & `Levene Tests`.

A.1.2 (Indoor) Impact of Mobility Direction. With the same mobility direction, 78.05% of geolocations have throughput samples that are normally distributed – an increase of over 25%, compared to the case when mobility direction is ignored as shown in Table 4. As for the pairwise `t-test`, with the consideration of the mobility direction information, 80.87% of the geolocation pairs have significantly different throughput means (*i.e.,* an increase of 10.01%). Moreover, 29.76% of geolocations have throughput samples with CV values greater than 50% – a decrease of 23% when mobility direction was ignored. This indicates that the variances of throughput samples at a given geolocation decreases when the mobility direction is accounted for. Fig. 19 shows extended results for the impact of mobility direction of the indoor (`Airport`) area in §4.2 of both trajectories (NB, SB) as well as the combined trajectories when the mobility direction is ignored (NB+SB).

Fig. 20 shows the 12 trajectories of the outdoor (`Intersection`) area as well as the impact when the mobility direction is accounted

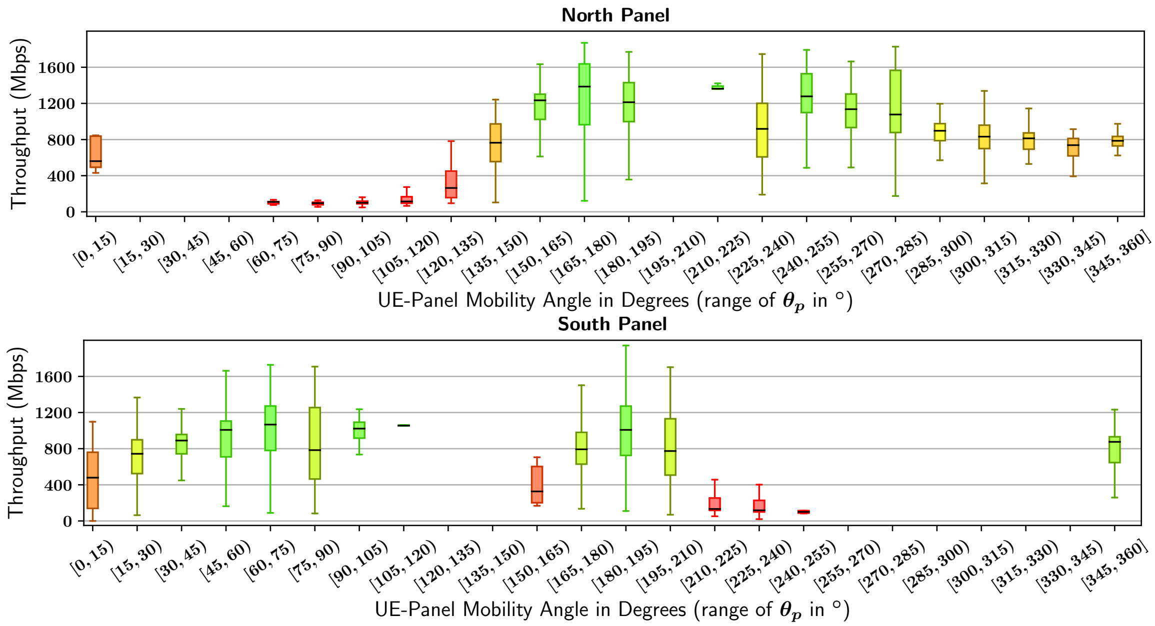


Figure 18: Impact of UE-Panel Mobility Angle θ_p by individual 5G panels on 5G throughput at the `Airport` area.

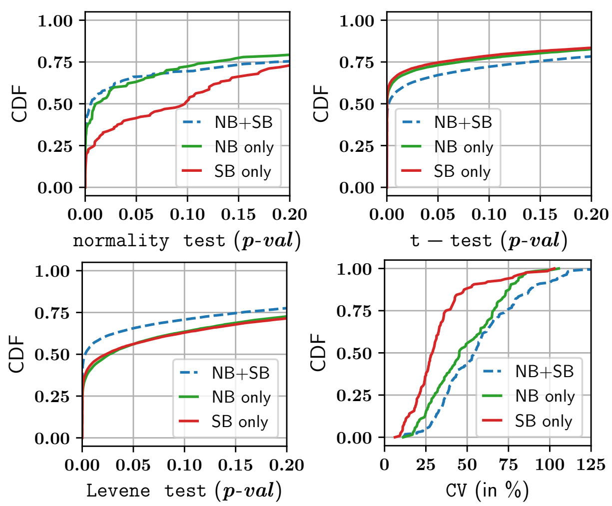


Figure 19: (Indoor) Impact of Mobility Trajectory on `Normality Test`, `t-test`, `Levene Test` & `CV`.

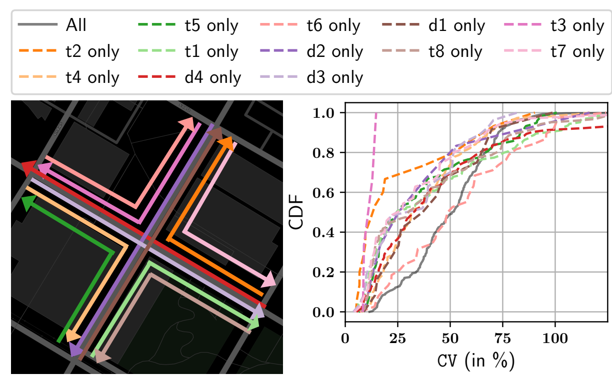


Figure 20: (Outdoor) Impact of Mobility Trajectory.

for *v/s.* when it is ignored. The same observation still holds: considering mobility direction information significantly reduces the variances of throughput samples at a given geolocation.

A.1.3 Impact of UE-Panel Mobility Angle. Fig. 18 shows the effect of the UE-panel mobility angle θ_m on 5G throughput *w.r.t.* the south and north panels. We can notice that, for both panels, throughput is high when the UE is moving straight towards either panels – *i.e.*, θ_m is in the range [165°, 180°) (see illustration of θ_m values in Fig. 8). South panel seems to have a better coverage as the UE can achieve relatively good throughput even while moving away from the panel (*i.e.*, when θ_m in [30°, 75°]). For certain ranges such as [210°, 240°) for the south panel and [90°, 120°) for the north panel), poor throughput is observed most likely due to poor NLoS path in the airport's mall-area.

A.1.4 Congestion with other UEs. When multiple UEs connect to the same 5G panel, how does 5G throughput get affected? To answer this question, we perform an experiment in the `Airport` area by placing multiple UEs side-by-side such that they all fall in the coverage footprint of a single 5G panel. We decide on a location that is at a distance of ~25m from the 5G panel with clear line of sight (LoS). We use 4 UEs – UE_1, UE_2, UE_3, UE_4 – where each of the UEs is scheduled to start an `iPerf` session such that the session start timestamps of each device are separated by a gap of 1 minute and the end timestamps are the same for all devices. Each `iPerf` session on a UE is at least 1 minute long (thus, each set of experiment with all the 4 UEs was 4-minute long, see Fig. 21). This allowed us to overlap `iPerf` sessions running on separate UEs that are connected to the same 5G panel and observe the impact of the *"artificially induced"* congestion. `iPerf` servers are running on VMs provisioned in different public clouds (*e.g.*, Google Cloud, Amazon

Table 10: Summary of Factors Affecting 5G Throughput and Its Predictability for the Outdoor (`Intersection`) Area.

Results ⇒	Statistical Analysis					Simple Pred. Models			
	CV		Norm. Test	Sp. Coeff.		KNN		RF [20, 54]	
⇓ UE-Side Factors	*(mean*	*±std. dev.)*	*(p-val. > 0.001)*	*(mean*	*±std. dev.)*	*MAE*	*RMSE*	*MAE*	*RMSE*
(1) Geolocation	52.83%	±20.28	66.54%	0.17	±0.36	297	376	258	336
(2) Mobility + (1) ⊦ UE-Panel Distance ⊦ UE-Panel Positional Angle ⊦ UE-Panel Mobility Angle ⊦ Moving Speed	43.08%	±20.30	84.17%	0.49	±0.08	258	337	238	310

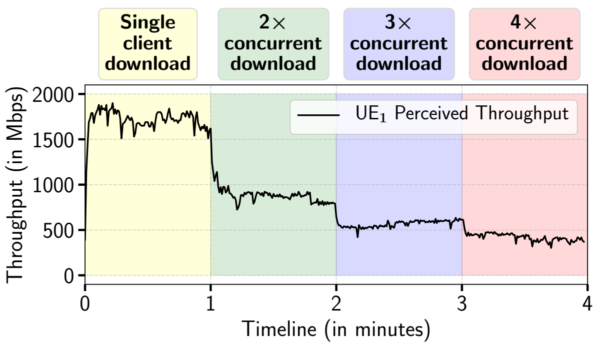


Figure 21: Impact on 5G throughput perceived by a single UE when multiple UEs are connected to the same 5G panel.

Web Services, Microsoft Azure). To avoid Internet being the bottleneck, we take 2 approaches. (1) We repeatedly run the experiment sets. (2) We shuffle the UE-VM combination in each iteration to ensure no device-side or server-side issues affect the experiment, and additionally to reduce the impact of Internet being the bottleneck. Fig. 21 shows a representative run of this experiment and reports the throughput perceived by UE_1 – the device that ran for the entire 4-minute long duration. In the first minute when UE_1's `iPerf` session was running alone, we see great throughput performance of more than 1.5 Gbps. However, as soon as UE_2 started its session, UE_1's throughput nearly halved. Although not shown, UE_2 also experienced similar throughput as that seen on UE_1. Since these experiments were done during late night hours, we believe there was little to no impact on our experiments by other passengers in the airport who might also potentially be using 5G service in this area. We see similar behavior when the `iPerf` sessions of UE_3 and UE_4 started. This experiment highlights another UE-side factor impacting 5G throughput which practically reflects a *"dynamic"* factor or the *time-of-day* effect. Since we did not have the information on the number of users actively served by different 5G towers at a certain time, we could not account for this factor in this paper. However, due to the expandable nature of feature groups in Lumos5G, we believe 5G carriers can extend and adapt the feature groups and use the number of subscribers connected to a 5G panel as another input feature to further improve 5G throughput prediction in real deployments.

A.1.5 Statistical Analysis for Intersection Area. Similar to Table 4 which represented the Indoor/`Airport` area in main text, Table 10 summarizes the statistical analysis of the Outdoor (`Intersection`)

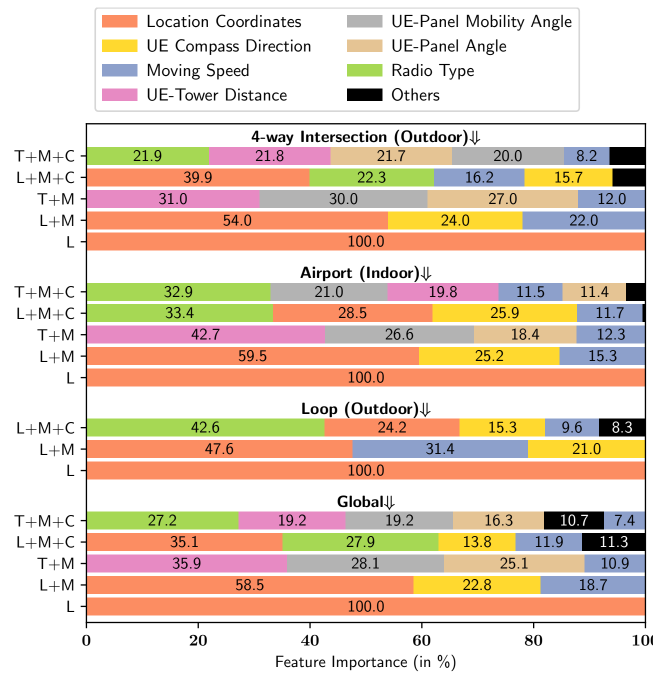


Figure 22: Feature Importance using GDBT.

area. The observations made for Indoor area also hold true for the Outdoor area: *(i) Geolocation alone is insufficient to characterize & predict 5G throughput, but it still remains a key factor; (ii) Along with geolocation, accounting for mobility-related factors decreases variation in 5G throughput and improves its predictability.*

A.2 Feature Importance

Using GDBT's ability to report global feature importance, Fig. 22 shows the feature importance score in 5G throughput prediction. This score is a value between 0 and 100% where scores of all features sum up to 100%. We make the following key observations: **(1)** In general, there is no single feature that dominates the throughput prediction problem in 5G. For instance, in the case of `T+M+C`, we see connection status (Radio Type/Strength), UE-Panel mobility angle, UE-Panel distance, UE-Panel positional angle, and UE's moving speed all show significant importance in predicting the throughput. This observation is additional evidence that accounting for interplay between different types of features yields better performance than only considering location-based features. **(2)** It is also interesting to find that the performance of both `L+M` and `T+M` feature groups are comparable to each other. When considered within a single region, this observation is intuitive as both `L+M` and `T+M` are the same, it is just that the former's features are from the UE's perspective while the latter is purely from the 5G panel's perspective.

A.3 Performance Improvement of Lumos5G Over Existing Baselines

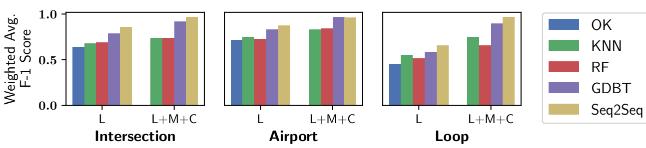


Figure 23: Performance Comparison with existing models on `Intersection`, `Airport` & `Loop` Areas.

In Fig. 23, we compare the performance of our models to existing approaches in different areas using feature groups. Approaches using naive location-based models (`L`) and spatial interpolation methods (OK) perform poorly compared to our models which account for mobility and connection information. Our models achieve 16% to 113% higher `w-avgF1` than pure-location based Kriging method, and achieve 5% to 88% higher `w-avgF1` than pure-location based KNN and RF models. This shows the importance of mobility and connection features for 5G throughput prediction. Our results clearly indicate the superiority of both Seq2Seq and GDBT models over existing throughput prediction methods.

A.4 4G *v/s.* 5G Throughput Prediction

In 3G/4G, location alone is known to be useful for predicting cellular performance [43, 56]. To further investigate, we construct a dataset by holding two 5G smartphones side-by-side, one connected to 4G network and the other to 5G, and walk the 1300m loop mentioned earlier for over 30 times spanning across multiple days, and log the perceived throughput traces. We then apply existing approaches such as KNN classifier, OK [26], RF [20] which are known to work well for 4G throughput estimation on 5G traces. Results show that the mean absolute error (MAE) on 4G traces is about [29.01, 69.13, 25.94] Mbps for KNN, OK and RF, respectively, while the same approaches on 5G traces show the MAE to be 10× higher – [325.95, 625.83, 339.57] Mbps, respectively. These results exemplifies that while existing models work well for predicting 4G throughput, but are unable to predict 5G throughput. This is because such methods are unable to account for the sensitivity of mmWave-based 5G to the environment – a small perturbation (*e.g.,* device orientation, moving direction, moving speed) affects 5G performance as discussed earlier in §4. Thus, geolocation alone is infeasible to estimate mmWave based 5G performance as shown in §6. In this paper, we propose Lumos5G framework that generalizes the classical location-based cellular performance prediction into context-aware prediction problem. The framework shows that in future, a data driven model could potentially use a wide range of contextual and environmental data such as location, time, mobility level, moving orientation, traffic information, *etc.* to model and predict 5G (all bands) + LTE + other lower band performance to account for several challenges faced by mmWave.