YouTube goes 5G: QoE Benchmarking and ML-based Stall Prediction

Raza Ul Mustafa, Chadi Barakat, Christian Esteve Rothenberg

YouTube goes 5G: QoE Benchmarking and ML-based Stall Prediction

Raza Ul Mustafa
University of Campinas (UNICAMP)
Brazil
razaul@dca.fee.unicamp.br

Chadi Barakat
Inria, Université Côte d'Azur
France
chadi.barakat@inria.fr

Christian Esteve Rothenberg
University of Campinas (UNICAMP)
Brazil
chesteve@dca.fee.unicamp.br

***Abstract*—Given the dominance of adaptive video streaming services on the Internet traffic, understanding how YouTube Quality of Experience (QoE) relates to real 4G and 5G Channel Level Metrics (CLM) is of interest to not only the research community but also to Mobile Network Operators (MNOs) and content creators. In this context, we collect YouTube and CLM logs with 1-second granularity spanning a six-month period. We group the traces by their context, i.e., Mobility, Pedestrian, Bus/Railway terminals, and Static Outdoor, and derive key performance footprints of real 4G and 5G video streaming in the wild. We also develop Machine Learning (ML) classifiers to predict objective QoE video stalls by using past patterns from CLM traces. We release all datasets and software artifacts for reproducibility purposes.**

***Index Terms*—5G, QoS, QoE, Machine Learning, YouTube.**

I. Introduction

Mobile video traffic is continuously growing, thus adding an additional challenge for Mobile Network Operators (MNOs) to manage this exponential growth [1]. Applications utilizing social media, gaming, and recent advances in Augmented/Virtual Reality and UHD videos have accelerated the demands for the next generation of networks, 5G [2]. The New Radio (NR) of 5G technology is developed to address high bandwidth, low latency, and massive connectivity requirements of enhanced Mobile Broadband (eMBB) compared to Fourth Generation (4G) Long Term Evolution (LTE).

Evaluating Quality of Experience (QoE) of YouTube video streaming from MNOs perspective in hybrid 5G-4G/LTE networks is a challenging endeavor [3]–[7]. The research community is actively working on evolved approaches to deliver improved end-users' QoE and provide adequate methods to manage increased video traffic demands, including 5G-aware machine learning (ML) throughput prediction to aid applications in intelligent bitrate adaptation [6].

Both 4G and 5G technologies have different characteristics, so it is important to compare them experimentally in a fair and representative way. Moreover, 5G brings more benefits in video streaming than 4G due to higher data rates, low latency and improved connection stability, among other features. Furthermore, there is a rich variety of candidate workloads, such as 4K/8K video streaming, interactive 360 and volumetric video streaming, and Augmented Reality/Virtual Reality (AR/VR).

In this paper, we seek to understand based on evidence the performance of 5G compared to 4G when streaming YouTube videos of diverse type (e.g., Nature, Animation, Movie, Brand Promotions) at different Frame Per Second (FPS) rates and under varying context scenarios: (*i*) Mobility, (*ii*) Pedestrian, (*iii*) Bus/Railway terminals, and (*iv*) Static Outdoor. Next, we study the relationship between the Channel Level Metrics (CLM) and objective QoE scores of YouTube. This study helps us to propose a QoE interruption (Stall) prediction method based only on CLM metrics. We carry out a rich 4G and 5G dataset collection campaign using commercial 4G and 5G networks, where we consider YouTube as a baseline for video streaming to collect CLM and YouTube QoE logs with 1-second granularity. All videos are selected from different categories such as Sports, Animated, Movies, Nature, etc. In addition, we consider videos with 4K quality and some that coded at 60 FPS. We provide detail of each video in [8]. Our contributions can be summarized as follows:

- We collect 4G and 5G datasets with channel and context using YouTube as a baseline at the smallest granularity of 1-second in a rich set of use case scenarios.
- We derive a model relating CLM measurements to video stalls using a time-based method. We check for different observation time windows (1, 3, 5, 7, 9)-seconds to interrelate stalling events of YouTube streaming. We find that a 7-second and 9-second window is best for predicting stalling events, achieving high accuracy for the Binary Classification of Stall vs. No Stall scenarios.
- For reproducibility purposes, we publish on a public code repository an open source release of our dataset [8][1] as well as the functional artifacts that we used for collecting the dataset and producing the results.

The rest of the paper is organized in the following manner: Section II presents the required background for this work. In Section III, we introduce our methodology and dataset collection approach. Section IV provides a performance analysis of 4G and 5G using both CLM and objective QoE of YouTube. In Section V, we describe the use of CLM metrics to predict stalling events. Next, we cover related work in Section VI. We discuss limitations and future work in Section VII before concluding the work with use case perspectives in Section VIII.

[1] https://github.com/razaulmustafa852/youtubegoes5g

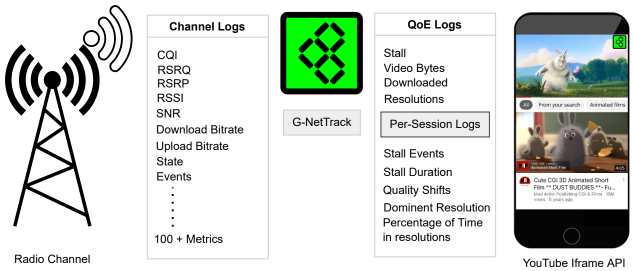


Fig. 1: Overview of 4G and 5G dataset collection tools.

II. Background

In order to provide 5G network services while addressing compatibility with previous cellular systems, there are two 5G deployment options, Non-Standalone (NSA) and Standalone (SA) and both have different mechanisms. In NSA, a pre-existing 4G core network is used, whereas in SA, a dedicated core network is required [9]. Both architectures require the deployment of a 5G NR Radio Access Network (RAN) composed of a set of Next Generation Node Bs (gNBs), i.e., the 5G equivalent of 4G Evolved Node Bs (eNBs).

Among the most important Radio Resource Management (RRM) in LTE are Channel Quality Indicator (CQI), Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ), where RSRQ and RSRP are crucial in Handoff (HO) decisions. When RSRP and/or RSRQ of the serving cell drop below a certain threshold of the neighboring cell by a predefined HO margin for a certain period of time, handover occurs [10]. Both metrics are mainly used to rank different candidate cells according to their signal quality. Signal to Noise Ratio (SNR) is measured by User Equipment (UE) on a Resource Block (RB) basis and then converted to CQI reports sent to the eNodeB. CQI is a quantized and scaled version of the experienced SNR [10] that gives an indication on the data rate that could be transmitted over a channel [3].

Looking into how a 5G-capable UE connects to the available 4G/5G Radio Access Technology (RAT), two Handover (HO) events can happen: i) Intra-RAT HO and ii) Inter-RAT HO. In Intra-RAT HO, UE switches from a 4G cell to another 4G cell or from a 5G cell to another 5G cell. However, it remains within the same technology. On the other end, Inter-RAT HO does the opposite, UE is instructed to rearrange its data plane from 5G to 4G or from 4G to 5G.

III. Data Collection Methodology

The two main software components used in our data collection methodology are presented in Figure 1: (i) YouTube Iframe API [2] to provide YouTube QoE logs, and (ii) G-NetTrack-pro, a 5G/4G/3G/2G network monitor and drive test application tool for Android UE.

A. YouTube Iframe API

The Iframe player API allows embedding the YouTube video player on web-based applications and controlling it using JavaScript. We design a custom web-based application and embed in it the YouTube Iframe. Then, using Javascript, we define `functions` to save player events in a MySQL database every 1-second interval. We collect player statistics such as Stalls and Quality shifts. Quality shifts refer to the change in resolution from lower to higher and vice versa. The application interface requires i) A unique ID to link CLM, and ii) a YouTube Video to play out. The resolutions available for the videos are 144p, 240p, 360p, 480p, 720p, 1080p, 1440p, 2160p for the first seven videos, and are the same for the remaining three videos, but this time with 60 FPS, i.e., 1080p/60FPS, 1440p/60FPS, 2160p/60FPS. YouTube Iframe API invokes `onStateChange` event, where the information of stall, along with other features, is available. Six states are available for the player: 0 – Ended, 1 – Playing, 2 – Paused, 3 – Stall, 5 – Cued, -1 – Unstarted. We write a script to save the QoE Key Performance Indicators (KPIs) of the YouTube player every 1-second using AJAX. For instance, we save i) Current Quality, ii) Video Bytes Downloaded (VBD), iii) Loaded Percentage (LP), iv) Available qualities, v) Time. These QoE KPIs can further provide per-session Objective QoE (i) Total stalling event, ii) Stalling ratio, iii) Stalling time, iv) Quality shifts or percentage of time in a single resolution, v) Dominant resolution, etc.

B. G-NetTrack Pro

We use G-NetTrack Pro[3] for the collection of Channel Level Metrics (CLMs). This tool allows the monitoring and logging of mobile network parameters without using specialized equipment. It provides 5G/4G/3G/2G serving and neighboring cells information and saves it in log files (text and kml format). We set a 1-second granularity for logs in the setting. The most valuable metrics include CQI, RSRQ, RSRP, SNR, and application download bitrate. An example of CLM metrics with their corresponding QoE player logs and events of YouTube is shown in Table I. We show an example of use case: 5G and Mobility, i.e., traces and metrics collected while in a bus and we have 5G coverage throughout the experiment. The experiment started at `17:43:17 - H:m:s` and lasted for six and a half minutes. The video is an animation named "Shadow of the Republic" with YouTube video id `JMbBjKnUoC4`. We show a few metrics of the whole session starting from (17:43:30) belonging to our three measurement categories, i.e., i) CLM, ii) Player logs and iii) Player events. In Table I, the first category presents six metrics, Time, RSRP, RSRQ, SNR, CQI, DL_bitrate (download bitrate) followed by the player logs category with four metrics, Time, Quality and VBD (video bytes downloaded), LP (loaded percentage). In the final part of the table, we show the player's events during this streaming session. In this sample of the dataset, we observe that the

[2] https://developers.google.com/youtube/iframe_api_reference

[3] http://bit.ly/3P3DBjK

TABLE I: CLM and their corresponding YouTube player logs and events for use case – Mobility, Technology – 5G

Channel metrics						Player logs				Player events		
Time	RSRP	RSRQ	SNR	CQI	DL_bitrate	Time	Quality	VBD	LP	Time	Quality	Event
17.43.30	-97	-3	20	13	5	17.43.30	hd2160	0.128010425	12.8	17:43:30	hd2160	buffering
17.43.31	-97	-3	20	13	61350	17.43.31	hd2160	0	0	17:43:32	hd2160	playing
17.43.32	-97	-3	20	13	86264	17.43.32	hd2160	0.08171254	8.2	-	-	-
17.43.33	-87	-3	21	13	94897	17.43.33	hd2160	0.102403732	10.2	-	-	-
17.43.34	-87	-3	21	15	4	17.43.34	hd2160	0.102403732	10.2	-	-	-

RSRP remains between (-97 — -87), followed by the RSRQ, which remains at (-3) for this window of observation. SNR values remain between (20-21) and CQI stays in the range (13-15). We experience a maximum of 94897-kbps current downlink data for the observed window. In the second section of the table, we show YouTube player logs for the time steps starting from (17:43:30) to (17:43:34) (H:m:s). During these time steps, the video quality remains hd2160, and the loaded percentage gradually increases after the first chunk. In the last section of the table, we show quality and buffering events. Our player experienced a buffering event at 17:43:30 with the video quality being set to hd2160.

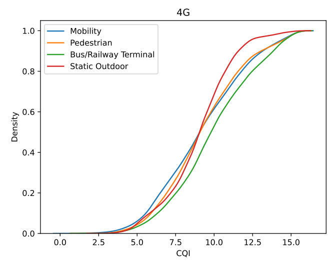
C. Data Collection Approach

We opted for one commercial 4G and 5G operator in France. The operator we selected provides low-band (3.4-3.8 GHz range) 5G service using NSA modes. The dataset collection is conducted in Nice area, France, for 6 months from June 2022 to November 2022. We covered approximately 1000+ km of mobility experiments. We used two smartphones of user equipment (UE) with 4G and 5G support, i) Samsung Galaxy S21 – 5G, and ii) Samsung Galaxy S8 – 4G. We selected a 15 km driving route with busy downtown regions and freeways with driving speeds (Mobility use case) ranging from 0 to 80 km/h. For the use case – Pedestrian, we ran 4G and 5G campaigns in busy downtown at different times and days. For the use case – Static Outdoor, we opted for measurements in Shopping Malls and in areas outside residential buildings. Finally, we carried out measurements in Bus/Railway terminals in the selected region for the final use-case (Bus/Railway terminals).

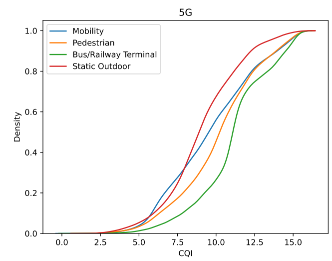
The data collection follows two methods, i) Standalone and ii) Comparison. For the first method, a single 4G device is used to collect the 4G dataset, and the same method applies for 5G. For the second method, two UE, 4G and 5G, were both used at the same time to draw a quick comparison of the two technologies. 5G experiments were done mostly in 5G-covered areas, i.e., downtown, malls, and bus/railway terminals.

During the dataset collection campaign, the G-NetTrack Pro application ran in the background, whereas we had to open our web application to collect video player logs. The process of stopping and playing the streaming session was manual. However, the logs and file-saving processes automatically upload data to the server. A demonstration of the whole process is explained in a public video.[4]

[4]https://bit.ly/3elgSkT

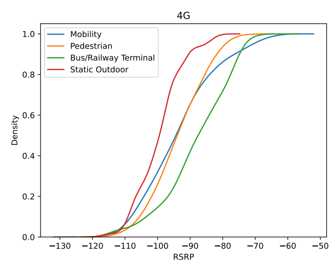


(a) 4G – CQI

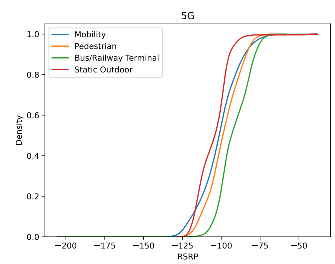


(b) 5G – CQI

Fig. 2: 4G vs. 5G CQI.

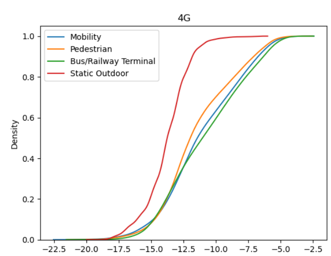


(a) 4G – RSRP

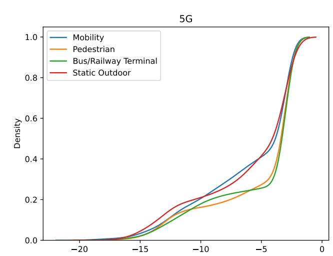


(b) 5G – RSRP

Fig. 3: 4G vs. 5G RSRP.



(a) 4G – RSRQ

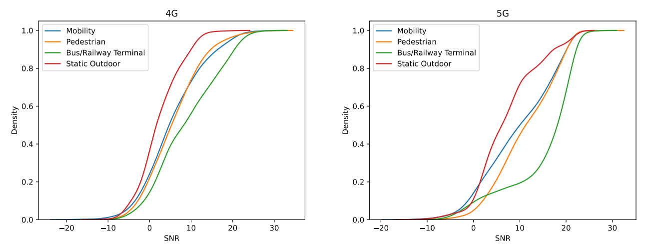


(b) 5G – RSRQ

Fig. 4: 4G vs. 5G RSRQ.

IV. 4G vs. 5G Performance Analysis

Channel Level Metrics (CLM). Figures 2, 3, 4, 5 show Cumulative Distribution Functions (CDFs) for the four Radio Resource Management (RRM) metrics we consider in our work. Figures plot results for the comparison method, where two UE devices are streaming the same video content at the same time with one UE as 4G and second UE as 5G. Each figure shows four CDFs corresponding to the four mobility use cases of our study. We observe in particular better CQI, RSRP, RSRQ and SNR at Bus/Railway terminals than in static outdoors for both 4G and 5G. Figures also confirm that

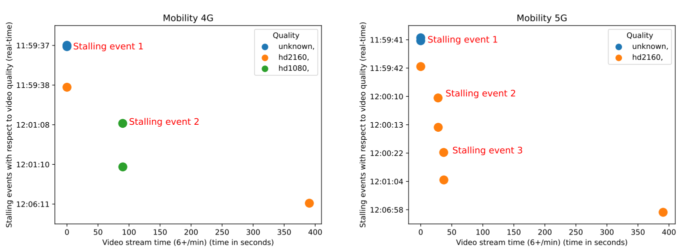


(a) 4G – SNR (b) 5G – SNR

Fig. 5: 4G vs. 5G – Use case dataset collection comparison.

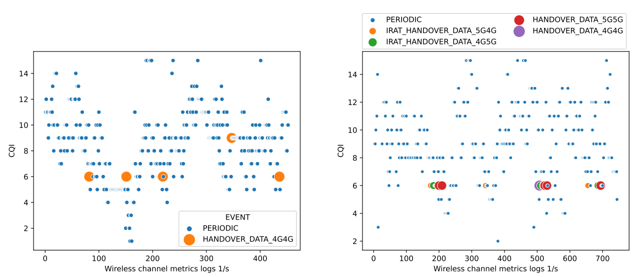
TABLE II: 4G vs. 5G percentage of player resolutions, case – mobility and pedestrian.

Case	480p	hd720	hd1080	hd1440	hd2160
			4G		
Mobility	7.7	-	3.2	25.9	63.2
Pedestrian	0.4	3.7	33.2	24.8	37.9
			5G		
Mobility				0.2	99.8
Pedestrian		0.6	3.7	0.4	95.3



(a) 4G – Stalls (b) 5G – Stalls

Fig. 6: 4G vs. 5G stalls and quality shifts in mobility.



(a) 4G – HO (b) 5G – HO

Fig. 7: 4G vs. 5G HO and CQI impact.

5G show in general better CQI and CLM than 4G networks, for all mobility scenarios, hence better QoS for 5G end users.
Stalling events. More stalling events can be surprisingly observed in 5G as compared to 4G under mobility, see an example of a video session in Figure 6 (a) and (b) for 4G and 5G, respectively. The x-axis represents time in seconds of video session, i.e., 6:30 minutes, while the y-axis represents occurrences of stalling events in real-time (from top to bottom). In 4G, we observe quality shifts between hd2160 and hd1080 during a video streaming session of 6:30 minutes. However, on the other end, 5G continues to stream at hd2160. For instance, in Figure 6 (b), the streaming session starts at 11:59:41 (H:m:s), and the first stalling event occurs at (11:59:41 – H:m:s) followed by the second event at (12:00:10 – H:m:s) and then third event at (12:00:22 – H:m:s). In 4G, the first stalling event occurs at (11:59:37 – H:m:s) see Figure 6 (a) and the second at (12:01:08 – H:m:s). Here, we argue that the first stalling event corresponds to the buffering process in both technology, 4G and 5G, which is 1-second. The video for Figure 6 (a) and (b) is an animation named "Shadow of the Republic" with YouTube video id `JMbBjKnUoC4`. A possible explanation for these results is that 5G under mobility exhibits more variable conditions than 4G leading to these stalls. In our experiments involving 5G mobility, we observe a `16.67 %` increase in stalling events compared to those in 4G. Note: We have a total of 8 events (Figure 6(b)) for 5G and 6 events (Figure 6(a)) for 4G. Here, events refer to moving from one state to another, i.e., stalling – to – streaming and streaming – to – stalling corresponds to 4 events. Finally, in Figure 7, we show HO events for the above use case. Most of the HO events occur when the CQI value is below 6 (Figure 7). We observe again better CQI in 5G as compared to 4G.
5G behaves greedily under mobility. Even with stalling events during mobility, the player remains in higher resolutions instead of choosing a segment with a lower resolution and bitrate to avoid stalls (see Figure 6 (b)).
Quality Shifts. We experience quality shifts for use cases – Mobility and Pedestrian in both technologies 4G and 5G. In other use cases – Indoor and Outdoor, there are negligible shifts. 4G experiences more quality shifting as compared to 5G both in mobility and pedestrian, see Table II. During the use case – High Mobility, 5G remains 99.8% in hd2160 resolution, whereas in 4G, this is equal to 63.2%. We also notice 95.3% of streaming in hd2160 in 5G whereas, it is 37.9% in 4G for low Mobility – Pedestrian. Static cases experience more stability in both technologies, i.e., hd1440 (10%) and hd2160 (90%) in 4G and in 5G hd2160 (100%).

V. Stall Prediction Using CLM

We now turn the attention to machine learning classifiers to predict Stall vs. No Stall using CLM metrics. For the proposed method, we use **Player Events** metrics, namely the player states, 0 – Ended, 1 – Playing, 2 – Paused, 3 – Stall, 5 – Cued, -1 – Unstarted. We also record the timestamps of stalling events. We start checking the previous *times (window)*, i.e., (1, 3, 5, 7, 9)-seconds to see if there is any correlation with the radio channel CLM. We take the previous n observations of CLM, i.e., CQI, RSRP, RSRQ, SNR, along with their distribution and standard deviation. Other features derived from these metrics include: i) Majority of a window, ii) Standard deviation, iii) 25, 50, and 75 percentile of a window.

We consider interruption player events as **Stall** class and differentiate them from other non-interruption events under the **(No Stall)** class. Thus, we have a binary classification

TABLE III: Accuracy of different classifiers.

KNN					Random Forest			
Window	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score	Accuracy
3s	0.84	0.83	0.83	83	0.89	0.88	0.88	88
5s	0.85	0.85	0.85	85	0.92	0.92	0.92	92
7s	0.86	0.86	0.86	86	0.93	0.93	0.93	93
9s	0.85	0.85	0.85	85	0.94	0.94	0.94	94
ANN					Decision Tree			
3s	0.87	0.87	0.87	87	0.83	0.83	0.83	83
5s	0.86	0.86	0.86	86	0.85	0.86	0.85	85
7s	0.93	0.93	0.93	93	0.86	0.86	0.86	86
9s	0.92	0.92	0.92	92	0.90	0.90	0.90	90

TABLE IV: Comparison of state-of-the-art work in 5G dataset collection.

Ref	**Use Case**	**Settings**	**YouTube Logs**	**Total Traces**	**Video Played**
[3]	Mobility – Car, Static	File Download Netflix Amzon Prime	No	83 Traces	animated (circa 200m) live-action (circa 400m)
[4]	Mobility – Car, Pedestrian	Online Video Gaming	No	7 Weeks	No
[5]	High Mobility – Train	Controlled Video Streaming DASH.js	No	6 Months	Custom Settings
[6]	Pedestrian	Controlled experiments TCP/IP stack and C++	No	20 Days	Custom Settings
[7]	Mobility – Pedestrian – Static	Controlled experiments Using 5G Traces	No	N/A	Custom Settings
This work	Mobility – Pedestrian – Indoor – Outdoor	YouTube	Yes	6 Months	Videos 4K 4K - 60FPS

problem where we have to estimate whether a player event is a stall event or not based on CLM measurements in previous time window. Next, we use different classification algorithms (Decision Trees – DT, Random Forests – RF, k-nearest neighbors – KNN, and Artificial Neural Networks – ANN) to predict the event class. The results of each classifier are listed in Table III. Regarding the settings of classifiers, we use 5 neighbours for KNN, 500 estimators for RF, depth of 3 for DT, and 3 layers for ANN.

On each layer of ANN we use 50, 100, and 150 neurons with RELU activation function on Hidden layers followed by Sigmoid Activation Function on the last layer with Binary Cross Entropy for Binary Classification. We use a batch size of 50 and 1000 epochs for training the classifier. To balance classes, we use Synthetic Minority Over-sampling Technique (SMOTE) from the scikit-learn library since the occurrences of stalling events rows are fewer in comparison to instances without stalling events. We show Precision, Recall, F1-score and accuracy of each classifier in Table III, defined as follows:

$$\text{Precision} = \frac{\text{True Positives – TP}}{\text{True Positives – TP + False Positives – FP}}$$

$$\text{Recall} = \frac{\text{True Positives – TP}}{\text{True Positives – TP + False Negatives – FN}}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$$

$$\text{Accuracy} = \frac{\text{True Positives – TP} + \text{True Negatives – TN}}{\text{Total Instances}}$$

From Table III we can conclude that RF with a 9-second time window is the best classifier to differentiate Stalling vs. No Stalling events when past 9 seconds CLM measurements are forwarded as input. After that, for a 9 seconds window, we observe better accuracy when we use ANN followed by DT and KNN. For a 7-second window, both RF and ANN provide an accuracy of 93% whereas 86% in KNN and DT. In a 5-second window, RF outperforms other classifiers with an accuracy of 92%, whereas ANN performs slightly better as compared to KNN and DT providing an accuracy of 86%. Finally on a 3-second window again RF and ANN perform better providing an accuracy of 88% and 87% respectively, whereas, 83% using KNN and DT.

VI. Related Work

Several studies analyze the performance of 5G NSA and SA from different angles, including HTTP vs. QUIC video streaming in DASH and its QoE, application level performance in both protocols among other performance metrics such as latency, power consumption and coverage [4]. The closest work related to our efforts is [3], where the authors provide CLM with a 1-second granularity in a total of 83 5G traces. Using two well-known videos – animated (circa 200m) and live-action (circa 400m), the key contribution is more related to the dataset for 5G Mobility and Static using file download, Netflix, and Prime video streaming.

Another work that performs a similar measurement campaign is [7]. The authors investigate the footprint of power consumption and the performance of state-of-the-art Adaptive Bitrate Streaming (ABS) algorithms under 5G compared to 4G, unveiling the major factors that impact ABS streaming performance over 5G. The conclusions include the need for new mechanisms to turn future ABS algorithms 5G-aware. In our case, we consider YouTube as a baseline for QoE KPIs, whereas, in [7], custom settings are used to run diverse experiments. We use 10+ 4K videos for streaming compared to a single video of 2.38-minutes duration.

Authors in [4] provide a dataset to understand the HO events by playing online video games. The total duration of the study is 7 weeks, but most of the work focuses on studying the HO while playing video games. They consider two use cases for collecting datasets, i) Pedestrian and ii) Mobility - Car. In another study [5], we see 5G dataset collection under very high mobility, such as Train. The authors analyze throughput, RTT, loss rate, physical resource utilization and signal quality for both technologies 5G and LTE. They provide a dataset for a duration of six months with custom settings using DASH.js player and controlled experimentation. Authors in [6] worked on 5G aware streaming to avoid stalls and predict throughput. However, they only consider the case of Pedestrians with 20 days of dataset collection. In comparison to previous work, we summarize the goal of this research work in Table IV.

VII. Limitations & Future Work

The dataset considered in this work is collected using a web-based application, which uses YouTube Iframe API. Moreover, the dataset collection is done using two Android devices, one for 4G and one for 5G. We do not consider multi-user streaming of the same content simultaneously. However, these experiments conducted in the wild, and so even if we did not generate a lot of video traffic, the other users at that particular time are generating traffic. Further, during the dataset collection campaign, we consider the full width of YouTube player, which automatically adjusts to the viewport of the device. However, we are aware that different screen sizes may influence QoE and lead to different results.

In our future work, we plan to tackle the limitations mentioned earlier by incorporating viewports and 360 videos into our QoE measurement. This will allow us to obtain a more accurate assessment of the user's experience.

VIII. Conclusions & Use Case Perspectives

MNOs strive hard to find relations between network KPIs and QoE using various methods. Therefore, the contributions of this work include: i) a rich dataset with various fine-grained features and metrics related 5G CLM and YouTube QoE, ii) a dataset with different use cases to run in `EFFECTOR` [11] replaying real 4G and 5G traces, iii) a new method based only CLM to predict YouTube stalling events. 5G suffers from stalling events, where YouTube player keeps streaming at high resolutions and bitrate. Overall, 5G outperforms 4G in YouTube streaming, as expected, but performance over currently deployed 5G networks is still not ideal.

The YouTube and CLM datasets with 1-second granularity can be relevant to the research community to exercise a variety of 4G and 5G use case scenarios, including: i) ML-based predictive models inputting CLM metrics to predict per-session objective QoE KPIs in a range (Low, Medium, High), or forecast next n-seconds radio resource management resources, i.e., CQI, RSRQ, RSRQ, SNR; ii) Video application developments to trigger pre-loaded contents (e.g. advertisement) to play during a predicted stall time or to continuously monitor location and mobility patterns, learn about past historic stalls and quality shifts, and hook in some strategy in the adaptive algorithm (e.g., increased buffer) to maximize QoE; iii) Education (Skill building and talent recruitment), where we leverage the proposed methodology in undergrad and grad classes to have students acquire knowledge and skills through 5G CLM collection and Youtube QoE analysis using their own devices.

Acknowledgment

This work was supported by the Innovation Center, Ericsson S.A., and by the Sao Paulo Research Foundation (FAPESP), grant 2021/00199-8, CPE SMARTNESS. This study was partially funded by CAPES, Brazil - Finance Code 001.

References

[1] K. Ragimova, V. Loginov, and E. Khorov, "Analysis of YouTube dash traffic," in *2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 1–5, IEEE, 2019.

[2] M. Ghoshal, Z. J. Kong, Q. Xu, Z. Lu, S. Aggarwal, I. Khan, Y. Li, Y. C. Hu, and D. Koutsonikolas, "An in-depth study of uplink performance of 5g mmwave networks," in *Proceedings of the ACM SIGCOMM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases*, pp. 29–35, 2022.

[3] D. Raca, D. Leahy, C. J. Sreenan, and J. J. Quinlan, "Beyond throughput, the next generation: a 5G dataset with channel and context metrics," in *Proceedings of the 11th ACM Multimedia Systems Conference*, pp. 303–308, 2020.

[4] K. Kousias, M. Rajiullah, G. Caso, O. Alay, A. Brunstrom, L. De Nardis, M. Neri, U. Ali, and M.-G. Di Benedetto, "Implications of handover events in commercial 5G non-standalone deployments in Rome," in *Proceedings of the ACM SIGCOMM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases*, pp. 22–27, 2022.

[5] Y. Pan, R. Li, and C. Xu, "The first 5G-LTE comparative study in extreme mobility," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 6, no. 1, pp. 1–22, 2022.

[6] E. Ramadan, A. Narayanan, U. K. Dayalan, R. A. Fezeu, F. Qian, and Z.-L. Zhang, "Case for 5G-aware video streaming applications," in *Proceedings of the 1st Workshop on 5G Measurements, Modeling, and Use Cases*, pp. 27–34, 2021.

[7] A. Narayanan, X. Zhang, R. Zhu, A. Hassan, S. Jin, X. Zhu, X. Zhang, D. Rybkin, Z. Yang, Z. M. Mao, *et al.*, "A variegated look at 5G in the wild: performance, power, and QoE implications," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, pp. 610–625, 2021.

[8] R. Ul Mustafa, C. Esteve Rothenberg, and C. Barakat, "Youtube goes 5G: Benchmarking youtube in 4G vs 5G," 2022.

[9] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang, and N. Li, "5G deployment: Standalone vs. non-standalone from the operator perspective," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 83–89, 2020.

[10] F. Afroz, R. Subramanian, R. Heidary, K. Sandrasegaran, and S. Ahmed, "SINR, RSRP, RSSI and RSRQ measurements in long term evolution networks," *International Journal of Wireless & Mobile Networks*, 2015.

[11] R. U. Mustafa, T. Islam, C. E. Rothenberg, and P. Henrique, "EFFECTOR: DASH QoE and QoS Evaluation Framework For EnCrypTed videO tRaffic," in *36th IEEE/IFIP Network Operations and Management Symposium (NOMS 2023)*, IEEE, 2023.