Continuous Prediction and Assessment of Video Quality of Experience Using Deep Learning

A Dissertation

Submitted to the Council of the College of Information Technology, University of Babylon in Partial Fulfillment of the Requirements for the Degree of Doctorate of Philosophy in Information Technology-Software

Ahmed Salih Ghafil Hmad

Supervised by

Prof. Dr. Israa Hadi Ali Hussein

2022 A.D.   **1444 A.H.**

Dedication

I am dedicating this dissertation to four beloved people who have meant and continue to mean so much to me. First of all, my wife, Zainab. she has been a constant source of support and encouragement during the challenges of study and life. I am truly thankful for having her in my life. This work is also dedicated to my wonderful children, Hassan, Hussain, and Roqaih, who has always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

There were two persons to that I would like to dedicate this dissertation, my loving parents. My father, Salih Ghafil your unwavering support from my infancy and for the constant emphasis on the importance of education, I will forever remember your inspiring words. Because of all that, I am what I am today. My tender mother your incisive thinking and decisiveness that has motivated and encouraged me for as long as I can remember. To the loving memory of my late father and mother, may God bless them; I dedicate this dissertation to them.

Ahmed Salih Ghafil

Acknowledgement

Ahmed Salih Ghafil

Abstract

The explosive development of streaming media applications is driving a high demand for quality of experience (QoE) measurement. The concept of quality works in telecommunications include the definition of dynamic adaptive streaming (DASH) over HTTP. This work focuses on the content features. The quality of the video provided to the customer remains variable over time, depending on end-to-end network conditions. Changing network conditions may also cause the video client to run out of playback content, resulting in rebuffering events (stilling, freezing). These features affect user satisfaction and cause the deterioration of the user experience quality. The problem is the lack of user satisfaction with the continuous self-assessment with the service or application provided to him.

The proposed QoE measurement system for a No- Referenced video stream predicts the subjective opinion of the viewer. This dissertation presents GRU-QoE, a recurrent neural network-based QoE prediction model using a Gate Recurrent Unite (GRU) network. The GRU-QoE is a network of cascaded GRU blocks to capture the nonlinearities and the complex temporal dependencies involved in the time-varying QoE. The proposed model is inspired by the mechanism of the non-Markovian dynamics process for QoE, which is used to predict the current state based on previous cases. In addition, the model shows that the final layer is dense with Softmax activation. The softmax function is used as an activation function in the output layer of neural net models that predict polynomial probabilistic models.

The proposed model is compared to state-of-the-art QoE prediction models and proves to outperform them across the LIVE Mobile Video Stall Database-II, the LIVE QoE, the LFOVIA QoE, and the LIVE Netflix datasets. The evaluation model looks at the linear correlation coefficient (LCC), the Spearman order

correlation coefficient (SROCC), the root standard mean error (RMSEn), and the outage rate (OR) to figure out how well the QoE prediction works. Statistical comparisons show that the proposed model outperforms the state-of-the-art QoE prediction models. Several experiments have been conducted on video content and various standard networks. The results showed an improvement in the quality value expectation close to that of the actual viewer. Calculation using the proposed method improves the quality of the NARX model using the LCC evaluation via the LIVE Netflix database improved over 25.4%, SVR-QoE was improved over the LFOVIA QoE database by more than 20.4%, the Hammerstein-Wiener model via the LIVE QoE database improved over by 15 %, and the NLSS-QoE model by the LIVE Mobile Video Stall Database-II over by 21%.

Declaration Associated with this dissertation

Published Articles Dissertation

1- Ghafil Ahmed Sileh, and Israa H. Ali. "Video Streaming Forecast Quality of Experience-A survey." 2021 1st Babylon International Conference on Information Technology and Science (BICITS). IEEE, 2021.

2- Ghafil Ahmed Sileh, and Israa Hadi Ali. "State-of-the-Art in QoE for Video streaming." 2021 International Conference on Advanced Computer Applications (ACA). IEEE, 2021.

Papers in progress

1- Quality of Experience Evaluation for Streaming Video Using GRU

2- Selecting the optimal value for QoE result

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

Abbreviation	Description
ACR	Absolute Category Rating
BLSTM	Bidirectional Long Short Term-Memory
CDN	Content Delivery Networks
CDVL	Consumer Digital Video Library
CN	Core Network
CNN	Convolutional Neural Network
DASH	Dynamic Adaptive Streaming Over Http
DCR	Degradation Category Rating
FR	Full Reference
GRU	Gate Recurrent Unit
HAS	Human Audio System
HD	High Definition
HTTP	Hypertext Transfer Protocol
HVS	Human Visual System
ISPs	Internet Service Providers
ITU	International Telecommunication Union
KPIs	Key Performance Indicators
LSTM	Long Short-Term Memory
ML	Machine Learning
MMF	Multimeric Fusion
MOS	Mean Opinion Scores
MS-SSIM	Multi-Scale, Structure Similarity
OR	Outage Rate

OTT	Over-The-Top
PC	Pair Comparison
PDM	Perceptual Distortion Metric
PESQ	Perceptual Evaluation Of Speech Quality
PI	Playback Indicator
LCC	Linear Correlation Coefficient
PSQA	Pseudo-Subjective Quality Assessment
QA	Quality Assessment
QoE	Quality of Experience
QoS	Quality of Service
QP	Quantization Parameter
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RR	Reduced Reference
RTP	Real-Time Transport Protocol
SI	Spatial Information
SROCC	Spearman's Rank Order Correlation Coefficient
SSIM	Simple Implementation of Structural Similarity
STRRED	Spatial-Temporal Reduced Reference Entropic Differences
STSQ	Short Time Subjective Quality
TCP	Transmission Control Protocol
TI	Temporal Information
TR	Time Elapsed Since Last Rebuffering
TVSQ	Time-Varying Subjective Quality

List of Symbols

Symbols	Description
R	linear correlation coefficient (r) has a value between 1 and -1
RMSE(A,B)	Root Mean Square Error between A and B
SROCC(A,B)	evaluated by SROCC (whether linear or not) function between A and B
PLCC	PLCC measures the accuracy of the forecast,
E(Ø)	Outage Rate
$\mathbf{H}_{t-1}$	Hidden state
$\mathbf{R}_t$	reset gate
$\mathbf{Z}_t$	update gate
$\mathbf{W}_x$	weight parameters
$\mathbf{b}_r, \mathbf{b}_z$	biases.
$\odot$	is the Hadamard (elementwise) product operator

Chapter One

General Introduction

<h1 style="text-align: center">Chapter One</h1>
<h1 style="text-align: center">General Introduction</h1>

1.1   Introduction

The exponential development of streaming media apps leads to a high demand for measuring Quality of experience (QoE) and video supply technologies powered by QoE. It is becoming a leading term in operational networks for performance assessment, network management, and Internet services Demand. Video streaming, in particular, has increased due to the COVID-19 pandemic. It is the most significant contributor to worldwide internet traffic. The percentage of global IP video traffic is predicted to increase from 73 % in 2016 to 82% in 2021. Similarly, it is anticipated that mobile video traffic would climb from 56 % in 2017 to 73 % in 2023 [1]. Video data consumption and adoption from 2017 to 2022. The growth shows that Internet video encourages network operators to proactively estimate the Quality of Experience (QoE) of video streaming customers in their networks. The Internet video quality of experience (QoE) is influenced by subjective QoS (Quality of Service) aspects such as initial loading time (starting delay or join time), rebuffering frequency, stalling occurrences, and playout quality (spatial resolution) and its variations [2].

Quality of Service (QoS) is the traditional approach for measuring network services quality. It reflects the performance of the network and its components. It measures the network's ability to satisfy the needs of the service. That is, it is a network-centric metric [3]. The common QoS parameters are throughput, bandwidth, packet loss, delay, and jitter [4]. These QoS parameters work as QoS metrics that have been more suitable to measure the performance and reliability of network elements. However, QoS metrics do not capture the user's feeling of satisfaction with

the service. The service provided to the user is affected by the network components and several other confounding end-to-end factors like terminal, client hardware and operating system, system load, and the user's psychological and environmental settings. To comprehend the user's happiness with a service, it is crucial to evaluate the influence of these end-to-end confounding variables on the end user's impression of quality. QoS can be defined as the network's ability to provide an assured service level with a focus on parameters that exist in the network and application level [5].

The concept of QoE is more global and broader than the concept of QoS because it includes the network with the service or application.

Most of the studies utilized perceived quality and (QoE) terminology interchangeably. They defined the QoE as the total acceptability of an application or service, as viewed subjectively by the end-user [6]. In [7] QoE is defined as "the ability of an application to make a person happy or upset. It meets his or her expectations in terms of the current condition of utility and enjoyment of the app or service. Content, network, device, application, user expectations and goals, and the context of usage were all significantly impacted by the communication service [8].

QoE factors include network, application, and device performance, as well as, content characteristics [9]. (QoE) is mostly a reflection of how an individual perceives the (QoS). (MOS) is a common way for customers to rate the quality of a service they have received [10]. However, gathering subjective input from a large number of people takes a lot of time and money. Therefore, It is not always possible to do so objectively. Instead, objective assessment methods continuously monitor the quality of experience (QoE) of a service. The use of objective evaluation models can reduce the need for repeating subjective assessments [11]. The metric commonly used to quantify the quality of particular pieces of

multimedia content has been Mean Opinion Scores (MOS). It is the relative measurement used for the comparative measure of quality [12]. Even though there are some other methods as well to estimate the QoE, such as e-model [13], instrumental metrics (PESQ) [14], and neural approach, they all are used to estimate voice and video [15].

Video streaming QoE assessment is done under a controlled environment in the laboratory by using subjective humans as viewers to assess the quality of the video that has been tested. To overcome the high cost and the off-line nature of subjective tests, by using methods that are automated, objective, no-referenced, of low complexity, relying only on the information that the user side can provide, and running in real-time without any delay in the video's streaming process [16].

1.2   Research Motivation

The main goal of QoE is to meet user expectations. To satisfy this goal, network management methods need to be user-centric and provide services according to understanding the user's expectations. This change in network usage leads to the rapid shift from technology-centric to user-centric management of networks.

The user-centric QoE has overtaken the role of network-centric QoS by increasing the importance of user-perceived quality instead of the technical performance of the overall service. In Figure (1.1) the QoS-QoE domains are sketched [17]. QoE depends on many user-related factors, technical and context, while the QoS focuses on the network perspective only. In the evaluation of QoE, technical groups focus on network and service performance based on QoS models.
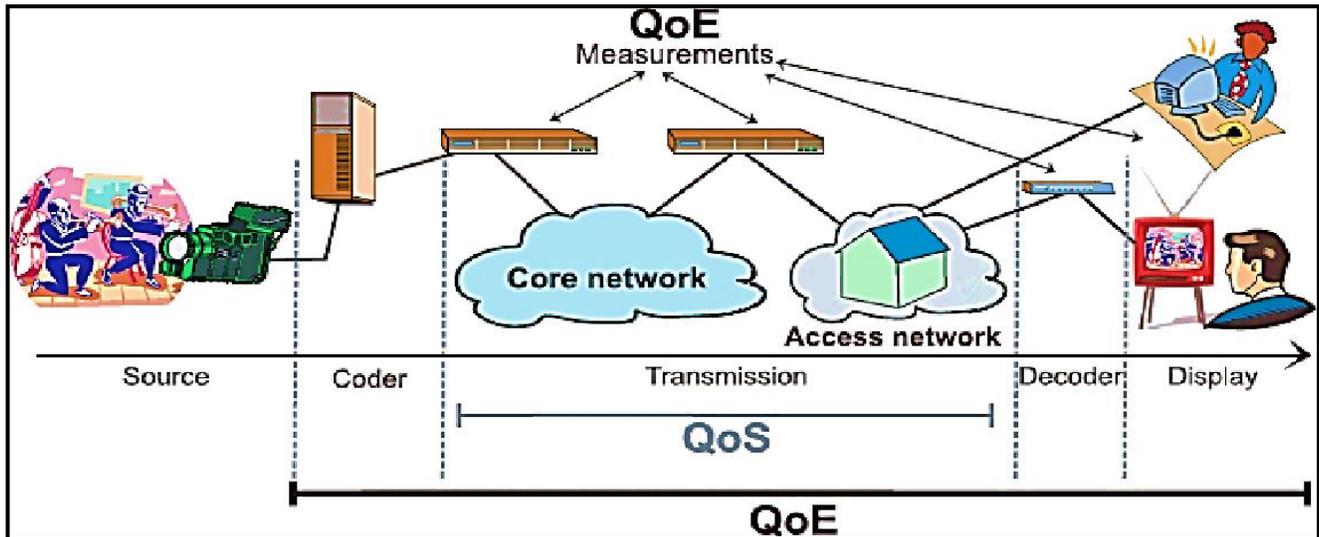
Figure (1.1) The QoS-QoE domains [17].

Several areas of application are considered within the domain of QoE including telecom services, assistive technology, cloud computing, multimedia learning, and gaming [18].

It is clear that QoE considers the perceptual side of the QoS. The QoS is measured through the parameters of coding and network monitors which are not suited to measure the end-user perception of video content because user perception is subjective and varies among environments [2].

.

1.3   Related Works

In this section, the researcher discusses related work that measures QoE for the video stream and the machine learning implemented to measure the QoE objectively.

1-   G. Dimopoulos in 2016 [19]**,** focused on privacy. Many online streaming services adopt end-to-end encryption, leaving providers with limited indicators to identify difficulties with user experience. This presents a technique for identifying quality concerns in streaming video from encrypted communications.

2-   Tsolkas, et al. in 2017 [20], state that the Key Performance Indicators (KPIs) consist of a collection of measurable indicators that are used to assess the overall long-term performance of a company. The KPIs help, in particular, to depart from other companies in the same market, and to assess the political, financial, and operational success of a company.

3-   Zhengfang et al.in 2017 [21], In this work, first build a streaming video database and carry out a subjective user study to the human responses to the combined effect of video compression, initial buffering, and stalling. The authors propose a QoE prediction approach named Streaming QoE Index that accounts for the instantaneous quality degradation due to perceptual video presentation impairment. Experimental results show that the proposed model is in close agreement with subjective opinions and significantly outperforms existing QoE models. The proposed model provides highly effective and efficient meanings for QoE prediction in video streaming services. The authors used Spearman's Rank order

Correlation Coefficient (SRCC) is employed for performance evaluation by comparing MOS and objective QoE scores. The average value of SRCC that can be obtained is 0.78. Moreover, the frameworks proposed by the dissertation use same (Waterloo QoE III) video dataset using the new feature based on both bitrate and pixel mode with highly correlated to end-user QoE, using this research to discuss the proposed system result with it.

4-   Demirbilek and Gregoire 2017 [22], developed multiple no referenced ML models to compare accuracy performance among ensemble decision

tree, genetic programming, and deep learning by using quality of audiovisual dataset obtained from Institute National de la Recherché Scientific (INRS), the dataset design specially to include impairment and network parameters showed in real-time transmissions. The result of using INRS dataset compared with other datasets to find that Random Forests achieved accuracy more than other prediction models. Correspondingly decision tree overcome both deep learning and genetic programming in RMSE and values of Pearson correlation terms. Demirbilek and Gregoire achieve result RMSE for all model range from (0.340) to (0.469). Moreover, the frameworks proposed by the dissertation use same (INRS) video dataset using the new feature based on both bitrate and pixel mode with highly correlated to end-user QoE, using this research to discuss the proposed system result with it.

5-   Shallal in 2018 [23], the author proposed a method to analysis video Streaming and extracts two types of features (bitrate and pixel-based

features), then used this features in decision tree C4.5 machine learning method to assess the QoE. In this proposed system used three machine learning: Artificial Neural Network (ANN), AdaBoost, and Random forest, to compare the accuracy performance. The author employed the Root Mean Score Error (RMSE) for performance evaluation that gets 0.072 and 98% in Cross Validation. The frameworks proposed by the dissertation use INRS and ReTRiEVED video database to test the system. The researcher on his work depends on the small number of features with limited videos.

6-   Khokhar, Muhammad Jawad in 2019 [24] shows that Machine Learning (ML) is used to model objective QoE metrics such as buffer alert (low buffer, high buffer), video status (an increase of buffer, decay of buffer, stable or stall).

7-   Nagabhushan Eswara et al. in 2019 [25]**,** present LSTM-QoE, a novel dynamic model for predicting the continuous QoE based on Long-Short-Term-Memory(LSTM). The proposed model involves a network of LSTMs for capturing the complex temporal dependencies present in the continuous QoE.

8-   Belmoukadam et al. in 2020 [26], the critical effects of this work are connected to the detection of problems concerning QoE and consumer interaction. The fundamental causes of weak QoE are three problems. They are network performance changes, CDN performance problems, and Internet provider overloading. Besides, it is observed that the rebuffering ratio has the most significant influence on consumer interaction.

9-   Chunhe, et al. in 2021 [27] indicate that standard approaches store all videos in the cloud and distribute them to users through competing distribution networks or the Internet by Prefetching and storing short videos. The researchers focus on Network edge servers.

10-   Liyana, et al. in 2022 [28], highlighted the benefits of employing QoE Optimization for Live Video Streaming Unmanned Aerial Vehicles Communications using a sophisticated actor-learner architecture based on Deep Reinforcement Learning for delivering video streaming services in IoT.

11-   Ghosh, et al. in 2022 [29], DeSVQ is a deep learning approach that combines various feature processing stages with an integrated framework consisting of Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) networks, each of which captures the intricate dependencies underlying the QoE prediction process.

Table (1.1): The summary of the related work for QoE video streaming

No	Method name	Method Type	Factor	Feature	Dataset
1-	identifying quality concerns in streaming video from encrypted communications	Subjective	User and system	Encrypted channel	Real time video streaming
2-	Key Performance Indicators (KPIs)	Subjective	Context	long-term performance	operational success of a

				of	company
				a company	
3-	Using Spearman's Rank order Correlation Coefficient (SRCC) as a measurement	subjective	User	initial buffering, and stalling	build a streaming video database
4-	developed multiple no referenced ML models	Objective	network parameters	real-time transmissions.	quality of audiovisual dataset obtained from Institute National Recherché Scientific (INRS)
5-	method to analysis video Streaming and extracts two features then used in decision tree C4.5 machine learning	Objective	System	bitrate and pixel-based features	INRS and ReTRiEVED video database
6-	Machine Learning is used to model objective QoE metrics	objective	System	low buffer, high buffer, video status	Institute National Recherché Scientific (INRS)
7-	dynamic model for predicting the continuous QoE based on Long-Short-Term-	objective	Content	Number of buffer, playback	Netflix database II

	Memory(LSTM)				
8-	connected to the detection of problems concerning QoE and consumer interaction.	Objective	user	network performance changes, CDN, and ISP overloading. rebuffering	Planform real time video streaming
9-	standard approaches for Network edge servers.	Objective	Network	by Prefetching and storing short videos.	store all videos in the cloud and distribute them to users through CDN or the Internet
10-	a sophisticated actor-learner architecture based on Deep Reinforcement Learning for delivering video streaming services.	hybrid	Network and system	Internet of Thinks feature	Live Video Streaming Unmanned Aerial Vehicles Communications
11-	using DeSVQ is a deep learning approach and (CNN) and (LSTM) networks	objective	user	that combines various feature	LEVIO video streaming

1.4   Problem Statement

The ability of telecommunication companies and service providers to meet the needs of their customers and create a positive user experience is quickly becoming their most important differentiators. In this competitive market, bad (QoE) leads to a chain reaction of bad word of mouth, driving clients into the arms of waiting rivals. Furthermore, customers are willing to pay for a better (QoE) with multimedia services [2] But if multimedia service providers fail to provide the promised quality, this will lead to a chain reaction of negative word of mouth, as reported in [4] that on average nearly 9 in 10 customers globally told the people around them about their bad experiences. And this poor customer experience remains the number one force pushing customers into the arms of waiting competitors.

Therefore, service providers must offer a high (QoE) to prevent client disloyalty and a bad reputation. QoE is a quickly evolving multi-disciplinary discipline are focusing on understanding total human quality needs from many aspects such as technology, business, and context of usage.

1.5   Challenges of the Dissertation

Every new assessment has a set of limitations, and one of the most important of these challenges facing service or application providers is knowing the quality of the products or services they provide to their customers. In the field of measuring service quality, first there is no global data set for all the factors that affect the quality or contain as many subject variations as possible and cover all influence factors. Second, providers usually do not see high-quality (4K) video content streaming. So quality cannot be measured directly. Third, there is the lack of accurate and continuous self-evaluations. Hence, there is a lack of a variety in intensity and patterns of distortion that would broadly reflect the different bandwidth limitations

must be addressed by video streaming services such as Netflix, YouTube, and Amazon Video.

1.6   Objectives of the Dissertation

The main objectives of this dissertation are to:

1- Design and implement a model for Objective Assessment Metric for Video Streaming based on content parameters that affect QoE video streaming.

2- Improve the accuracy of objective assessment metrics to be close to the expectations of the viewer. The results will then be compared with those of the state-of-the-art methods.

3- Enhance the QoE to near to opinion viewers by prediction accuracy while reducing the computational complication to support a variety of devices for accurate predictive No Reference (NR) assessment metrics.

1.7 Dissertation Contributions

The following are the primary contributions of this dissertation:

1. Proposing **Recurrent Neural Network, which is a Gate Recurrent Unit (GRU)**, a deep learning QoE model which can reliably forecast the overall and continuous scores of video streaming services.

2. Proposing a technology called GRU-QoE, with which build a network of **interconnected modules**. By connect each unit to another in the form of a bidirectional network to calculate the overall quality of the video stream. And finally collection result of **Time Distribution for SoftMax layer** for time distributed processing employed as the output layer to predict accurate overall QoE value.

1.8 Dissertation Outlines

The chapters are arranged as follows:

- **Chapter two** describes the essential principles of presenting and reviewing the theoretical background of Quality of Experience system factors, as well as the methodology for assessing their value.

- **Chapter three** proposes and achieves a framework for the QoE Assessment System. It presents a model that uses four datasets and effective features to learn that system.

- **Chapter four** presents the reported results that have been documented, analyzed, and dissected.

- **Chapter Five** summarizes the findings and makes recommendations for future research.

Chapter Two

Theoretical Background

Chapter Two

Theoretical Background

2.1 Overview

Historically, the quality of a network was established by the objective evaluation of a number of parameters. Quality of Experience is the metric of quality that takes into account the user's view point for (QoE). QoE is a subjective statistic that contains human factors; It measures network quality based on user perceptions, expectations, and experiences of application network performance [30].

Adopting a more holistic understanding of quality as perceived by end-users (QoE) is becoming an active area of research. When a customer gets bad service, the service provider cannot wait for the customer to say something. A survey by Accenture found that about 90% of users who get bad service do not complain about it. Instead, they leave their current provider and look for better service elsewhere. So, It is important that the service provider has a way to keep track of the QoE and improve it as needed [31].

A number of variables, including network dependability, the content production process, and terminal performance might influence the perceived quality of service. Many interconnected aspects influence the quality of service (QoS) of multimedia streaming services delivered over IP networks. Some of these characteristics, such as bandwidth picture resolution, may be changed, while others, such as packet loss latency, cannot. In order to maximize the happiness of the end user, these missing criteria must be taken into consideration. The literature provides a

14

number of various video quality measuring techniques, each of which has its own set of computational and operational demands. This chapter provides and elaborates on the central concept of Video Streaming Quality of Experience. In addition, it covers the Machine Learning techniques that are utilized to forecast new Quality of Experience (QoE) values from the data that were used in this dissertation [31].

2.2 Fundamental of Quality of Experience

The exponential development of streaming media applications leads to high demand for measuring Quality of experience (QoE) and video supply technologies powered by QoE. It is becoming a leading term in operational networks for performance assessment network management, the Internet services Demand. Video streaming has increased especially due to the COVID-19 pandemic. In 2013 there was more than one definition of the concept of QUALINET including has been characterized as [32]. The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations for the utility /or enjoyment of the application or service in the light of the user's personality current state.

Here "personality" defines the characteristics of a person that account for consistent patterns of feeling thinking behaving whereas "current state" refers to situational or temporal changes in the feeling thinking or behavior of a person.

The latest definition for QoE in 2021 the totality of end-user experience of the delivered service. They present another definition by the International Telecommunication Union (ITU-T P:10/G.100) recommendation for QoE. It is the

15

overall acceptability of an application or service as perceived subjectively by the end-user [33]. As far as QoE is concerned, Pearson states that user satisfaction is the sum of feelings or affective responses to distinguishable factors of the computer-based information products and services that are provided within the organization[34].

The impact of network performance is on user perception, which is associated with the failure or success of a service. Service providers focus on specific network-level QoS parameters like (delay, loss ratio, bit-rates and jitter) which are measured technically on network nodes but for the users [35].

But the fact that there are so many confusing acronyms and ideas that overlap around the idea of quality measures makes it hard to understand how important Quality of Experience is in telecommunication services. In [36] explain more about terms like user experaince (UX , QoX and QoE), Quility of servise (QoS). All of these are useful for figuring out both the qualitative and quantitative aspects of how a user interacts with a system or service. However, experts have recently come to agree that Quality of Experience should be given the most weight (abbreviated either as QoE or QoX). Let us follow this idea to find out why it is so popular and what comes from it. This is by far the most thorough way to find out how happy customers are. QoE has been used in a lot of different consumer-related industries and settings. But since its start, the Quality of Experience paradigm has mostly been used in the field of telecommunications. This is because this is where its research tools and measurements are the sharpest and most accurate. In some ways, Quality of Experience (QoE) in telecommunications is similar to Quality of Service (QoS), which aims to measure, improve, and guarantee software and hardware

characteristics. However, QoE in telecommunications is much more comprehensive and covers a much wider area [37].

From the user's point of view, Quality of Experience is the overall level of satisfaction with a service. Now comes the hard part, though. What true QoE is really trying to measure is not just how well the system works (like QoS does). It aims to include the subjective experiences of the service user with all of their complexities and human-dependent variables like physical, temporal, social, and economic factors.

2.3 Definition of the Overall Video's Quality

Video quality is used to define two main parts:

➢ Picture Quality is a measure of how well the picture can be seen.
➢ Audio quality is a measure of how well the sound can be heard.

In the end, there is only one way that has been shown to work for judging video quality, and that is Subjective Testing. But this is very expensive, takes a lot of time, and is often not a good idea [38]. Degradation Category Rating (DCR), Pair Comparison (PC), and Absolute Category Rating (ACR) are the three main ways to measure subjective quality. People are shown two sequences (the original and the processed version) and are asked to rate the overall quality of the processed sequence compared to the original sequence (which is used as a reference). Being the most for each session, a number of dummy sequences are added and used to train the human subjects but are not counted in the final score. The subjects give

the processed video sequence a score (usually between 5 and 9) based on how good they think it is. This score is called the Mean Opinion Score (MOS) [38].

On a scale from 1 to 5, these are the MOS scores:

1. Unacceptable

2. Unsatisfactory

3. Acceptable

4. Good

5. Excellent

PSNR and MSE are two types of metrics. They measure the difference between two signals, which can be measured exactly. This is very important in quality assessment (QA) monitoring, where the perceived quality has already been measured in the lab. What is needed is a PASS/FAIL indicator.

Traditional perceptual video quality index methods are based on a bottom-up approach that tries to simulate how the early parts of the human visual system (HVS) and human audio system (HAS) work. Most of the time, these methods involve[39]:

> ➢ Video/Audio range.
> ➢ Calculating the differences that affect the human eye/ear.
> ➢ Blurriness.
> ➢ Lack of Dynamic Range.
> ➢ Classifying the types of distortions and adding up the scores.
> ➢ Applying these scores to the Subjective MOS.

➢ Low pass filtering (to simulate the eye – video only).

➢ Blackness.

➢ Hight pass filtering.

Even though these bottom-up approaches can easily use many known psycho-physical characteristics of the HVS/HAS (human vision system/human auditory system), it is important to be aware of their limitations. In particular, the HVS and HAS been complex highly nonlinear systems. The complexity of natural images and sounds is also very important. Most models are based on linear or almost-linear operators that have been defined using limited, simple stimuli. Here are some examples that fit into this category [40]:

➢ Sarnoff/Picture Quality (PQR) was the First Well-Known HVS Metric.

➢ Video Quality Metric (VQM).

➢ Perceptual Evaluation of Video Quality (PEVQ).

➢ Perceptual Evaluation of Audio Quality (PEAQ).

The based similarity approach is a different way to look at the problem of figuring out how excellent a video is. It is predicated on the idea that the HVS is very good at figuring out the structure of a scene, So a way of measuring structural similarity should be a close estimate of how good an image looks to the human eye. Even if portion of a shape is missing, the eye can still see it. It has been shown that a simple implementation of structural similarity (SSIM) works better than the best perceptual image quality metrics that are currently available. But the SSIM index works best when it is used with the right scale (i.e. viewer distance/screen height). The hardest part of this method is figuring out how to set parameters like viewing

distance and picture resolution. Multi-Scale Structure Similarity (MS-SSIM) has been created to fix this. In MS-SSIM, the picture is evaluated at different resolutions, and the result is the average of these calibrated steps. It has been shown that MS-SSIM does a better job than simple SSIM, even when the SSIM is calibrated to the environment dataset correctly [39].In either case, the model comes up with a score, which then needs to be matched up with the subjective MOS. Using the Clear View Analyzer, there are two ways to do this [39]:

➢ Differential Mean Option Score (DMOS index with MS-SSIM algorithm).

➢ Just Noticeable Differences (JND Index with Sarnoff / PQR algorithm)

DMOS is the difference between "reference" and "processed" Mean Opinion Score in full reference testing as shown in Table (2.1).

TABLE 2.1: A general guideline for DMOS scoring.

Range	Description
0.1-2.0	Most customers Unsatisfied
2.1-4.0	Many customers Unsatisfied
4.1.-6.0	Some customers Satisfied
6.1-8.0	Most customers Satisfied
8.1-10	Every customer Satisfied

2.4 QoS and QoE

This section describes the distinctions between QoS and QoE. This will assist us in comprehending the operator's and end- user's needs. QoS is defined as

the network's capacity to provide a service at a specified service level. QoS is made up of all the processes, techniques, and procedures in the cellular network and terminal that make sure the user equipment (UE) and the core network (CN) deliver the service quality agreed-upon. QoE is how a user sees the usefulness of a service while using it how satisfied he or she is with a service in terms of, for example, usability, accessibility, retainability, and integrity. Service accessibility refers to the unavailability, security (authentication, authorization, and accounting), activation, access, coverage, blocking, and setup time of the related bearer service. Service retainability refers to connection losses in general [41].

QoE relates to the user's assessment of the quality of a certain service or network. It is represented via human emotions such as good, great, bad, etc. In contrast, QoS is fundamentally a technological idea. It is measured, represented, and comprehended in terms of networks and network components, which are often incomprehensible to the user. The ideas of QoE and QoS are shown in Figure (2.2). Although a higher network QoS will often result in a higher QoE, meeting all traffic QoS requirements does not ensure user satisfaction. Even if there is good throughput in one part of a network, it may not help if there is no coverage in other parts [41].

As for measurements, these numbers provide relatively little regarding the degree of client satisfaction for an operator. The flawless delivery of corrupted packets does not result in satisfied consumers. Therefore, it may not always be appropriate to conclude that QoE is enhanced when QoS methods are employed to minimize jitter or average packet delivery delay. The quality of user experience QoE is important, and the purpose of quality of service should be to give a high QoE.

Delivering a high QoE is contingent upon acquiring an understanding of the elements that contribute to the user's perception of the target services and utilizing this information to define the operational requirements. This method lowers development costs and the chance that a customer will reject or complain about an item or system because it does not meet their needs [41].
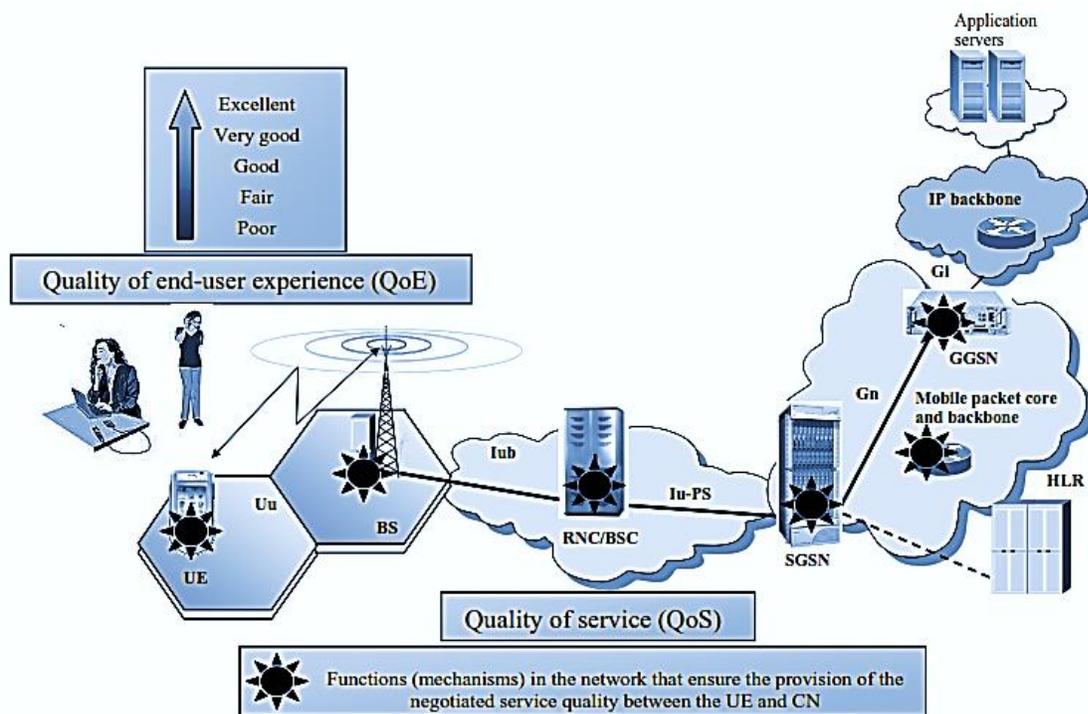
Figure (2.2): QoE is represented via feelings as opposed to statistics [41].

QoS refers to the network and terminal methods, operations, and procedures that implement the quality characteristics (carrier service) agreed between the UE and CN.

22

2.5 The Concept of Video Streaming

A video stream is a data flow supplied by a service provider (such as YouTube, Dailymotion, or Vimeo) and shown in real time by the end customer. The key concept behind streaming video is that the user does not need to receive the whole file or streaming data in order to play it. The media material is instead sent in a continuous stream of packets and is played as it arrives at the client interface (i.e. mobile, tablet, laptop...). To play a stream of data and see a video, the user needs a player, which is a piece of software that decompresses and transmits video data to the display and audio data to the speakers. A player may be a component of a web browser or a downloadable program. Figure (2.3) explains that when a mobile device requests a video, an interoperability between the OSI model's transport layer and application layer is created. The distribution of video through video streaming seeks to circumvent the issues associated with file downloads and gives a number of extra features. The fundamental concept behind video streaming is to divide the video into segments, send these segments sequentially, and allow the receiver to decode and play the video as it is received, without having to wait for the complete video to be delivered. The video content is pulled from the server when the end user asks for it, and the channel encoder changes the video stream to meet the QoS needs of the network. After the video stream has been encoded, it is broken up into packets and sent over a network [42]. At the end-user's terminal, the source decoder converts the incoming digital data into a continuous waveform that may be seen using various players at the application layer [43].
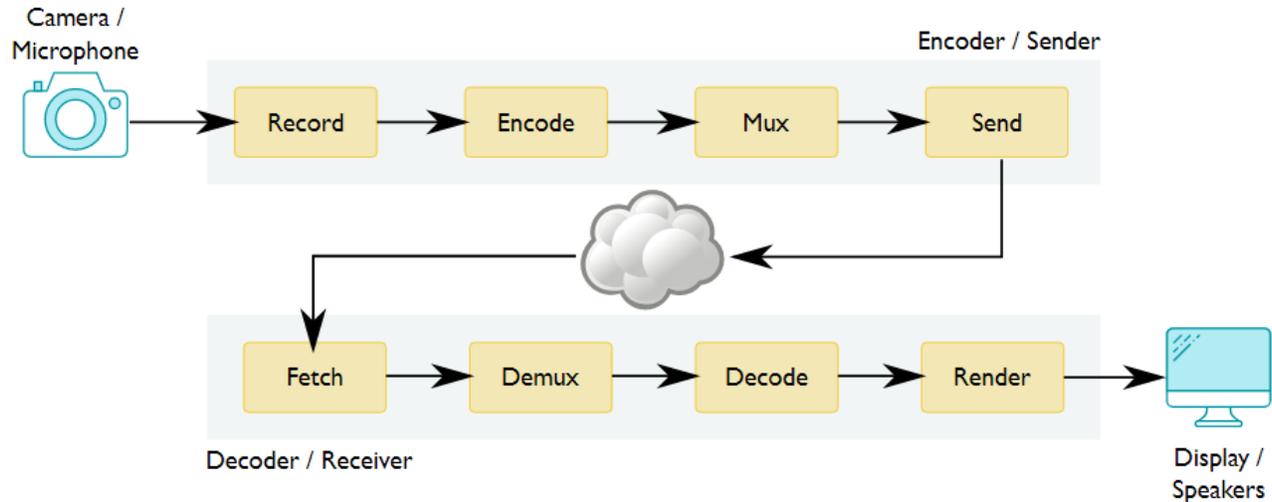
Figure (2.3): Encoder and decoder video streaming [44].

Unlike conventional services, such as online browsing, where network delivery quality is not crucial, video streaming services must transmit material with minimal video quality distortion from the user's perspective. Due to network congestion, the video stream may be subject to a variety of disturbances, including packet loss and latency variance. The QoS and QoE evaluation range for video services is shown in Figure (2.3). The QoS evaluation considers data management at the access point and across the network. In contrast, the QoE measurement encompasses the whole route followed by packets from their source to their terminal distribution in various modes (viz. display, streaming, and download modes) [45].

2.6 Video Streaming Protocols

One of the most common video streaming protocols for data transmission is used to manage the data flow between the video server and clients. The most common protocols used to communicate between network devices are Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). UDP is the transport protocol used by the majority of real-time video services. Compared to TCP, UDP

does not include any retransmission mechanism, which makes it desirable for delay-sensitive applications. However, UDP is an unstable technology and the video streams might suffer packet loss, which can cause distortion of the multimedia information [46]. In order to broadcast the video in real time, the Real-time Transport Protocol (RTP) must operate on top of UDP to make advantage of RTP services. RTP does not ensure QoS or reliable delivery, but offers assistance for applications with time limitations by providing a standardized framework for common features like time stamps, sequence numbering, and payload definition. RTP provides detection of missing packets [47,48].

TCP and on-demand video streaming are similar in nature. First, smooth video playback is achieved by pre-fetching and buffering in video streaming. TCP offers this buffer in addition to the reliable transmission assurance of no frame loss; nonetheless, a frame may still miss the play-out time and be rejected [49].

Second, TCP's bandwidth probing and congestion management will strive to utilize all of the available bandwidth between the server and client to get material as rapidly as feasible while maintaining compatibility with other TCP traffic on the same lines. Live video streaming, on the other hand, uses UDP because live streaming allows for less pre-fetching - the more buffering, the longer the video playout delay. Due to the fact that UDP only provides the most fundamental transport layer capability, it is often combined with additional application-layer and video-specific protocols such as RTSP to provide video streaming. Unfortunately, firewalls (from companies and ISPs) despise these protocols (unlike TCP, these protocols may use precious bandwidth incessantly; see TCP congestion management), making it impossible to circumvent their video traffic (dropped or throttled) [50].

The majority of Content Delivery Networks (CDN) servers were initially and already designed to primarily support web services. Consequently, video streaming via HTTP works out of the box without the need to configure dedicated servers, and the majority of firewalls do not block HTTP traffic. In reality, DASH (Dynamic Adaptive Streaming over HTTP) has become widespread. Although HTTP may be encased in other protocols in principle, these protocols must nevertheless offer reliable transmission (as standard HTTP assumes; Hypertext Transfer Protocol), hence excluding by UDP.

2.7 Management of the Application Layer of Traffic

Videos on YouTube are requested via HTTP over TCP. TCP is a transport protocol that guarantees reliable data flow by retransmitting dropped data packets and implementing congestion management to prevent the network from being overloaded. TCP's error correction and congestion management may both result in significant delay and jitter [52].

Rao et al. [53], have conducted a comprehensive analysis of YouTube's distribution approach. The authors demonstrate that the distribution approach is dependent on the video container (Flash, Flash High Definition, or HTML5), the client device (PC or mobile devices such as smart phones or iPads), and the browser (Internet Explorer, Chrome, or Firefox). The delivery strategy must balance a variety of potentially contradictory objectives such as:

1. Smooth playout throughout the length of a watching session;

2. Efficient use of server resources, such as timer management

3. Avoiding transmitting an excessive quantity of data in advance of consumption in order to (I) limit the amount of buffering at the client, which is especially important for mobile devices, and (II) prevent the waste of network and server resources by delivering data that are never utilized.

Finamore et al. [54] found that 60% of requested videos were seen for less than 20% of their overall time, leading to a transmission of (25–39)% of the data that was superfluous. The impact of degraded playback is the principal cause of the video transmission delay. As the video transmission is performed through HTTP over TCP, there is no assurance that the data will be supplied to the client at a pace at least equal to that at which it is consumed. Rresearchers concluded here that improve the chances of a smooth playback, YouTube conducts aggressive buffering when a video is requested. During the first starting stages, the server broadcasts as quickly as possible to fill up the client's initial playout buffer. When Internet Explorer is used as a browser, this playout buffer has about 40 seconds of Flash content and 10–15 MB of HTML5 content, which is often much more than 40 seconds of video [54].

2.8 Traditional Streaming vs. Adaptive Streaming

In standard streaming, a single video with a single resolution is downloaded gradually to your cache and then played by the video player. If you have a high-quality video but a poor Internet connection with little bandwidth, the movie will constantly halt and buffer since your connection cannot manage the Realtime

download. That is why they established adaptive streaming protocols such as Apple HLS and MPEG Dash. It is the ISO standard developed by MPEG for video streaming over HTTP-based networks. There is a movie with multiple resolutions that is separated into small pieces. Therefore, there is a video with four distinct resolutions (bandwidths) and each film is segmented into 2-second chunks. The player then receives a file including all resolutions, chunk lengths, media codecs, etc. When you play a video, the player evaluates the bandwidth of your Internet connection and selects the best possible resolution. The movie is divided into smaller segments so that the player may quickly transition (every 2 seconds in our example) to a different resolution when the internet's bandwidth changes as shown in Figure (2.4). YouTube has lately adopted adaptive streaming. With the standard HTTP progressive download mechanism, files can only be downloaded in their entirety. With the new mechanism, the YouTube server provides chunk requests only when the user is actively viewing the video. Consequently, server loads are greatly reduced, allowing YouTube to serve more people with fewer servers [55]. That video

streaming behavior and recorded videos are saved in a cloud server by the video streaming service provider. When a user clicks on a video in the online video player, the cloud server receives the request to transmit video content to the user's device. The video file is broken down into streaming audio and video data packets containing bite-sized data.

Transmission Control Protocol (TCP) is used to send data across a network. It is essential for the user to have a high internet speed for data to be transmitted easily so that the video can be watched without any buffering. The user can watch the
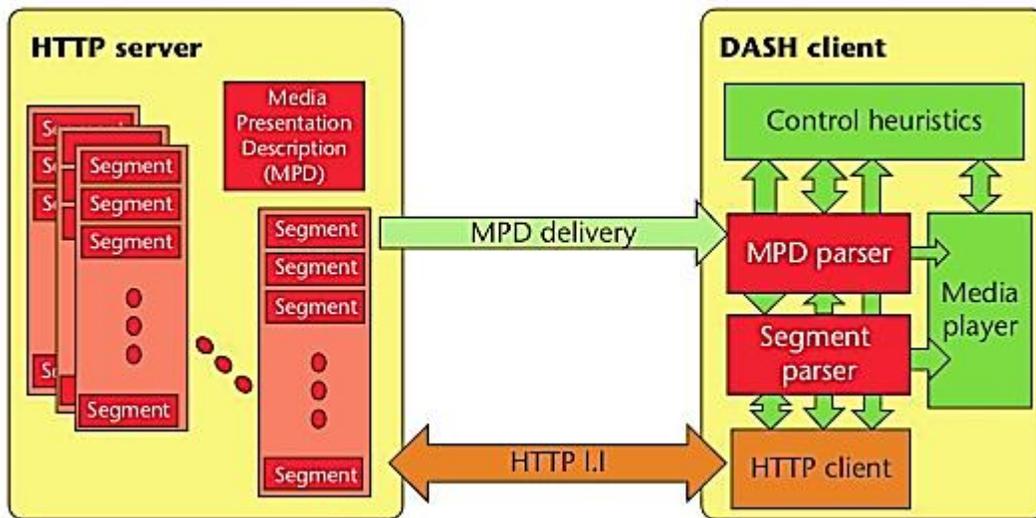
videos at low quality with medium speed interest.

Figure (2.4): DASH Framework [56].

2.9 QoE Influence Factor

Various variables impact the quality perception of Internet-based video streaming services among users. As a common denominator, four distinct kinds of

impact factors are distinguished: context, user, system, and content level influence factors [57].

I.   **The context level** takes into account factors such as the setting in which the user is consuming the service, the user's social and cultural background, and the reason for using the service, e.g., for entertainment or information retrieval.

II. **The user-level** comprises psychological elements such as user expectations, memory and recency effects, and program use history.

III. **the system level**, the technological influencing elements are abstracted. They include effects of the transmission network, devices, and displays.

IV. The **content level** for video distribution handles not only the video codec, format, and quality, but also the video's length, content, and motion patterns. Figure (2.5) demonstrates the influence factors.
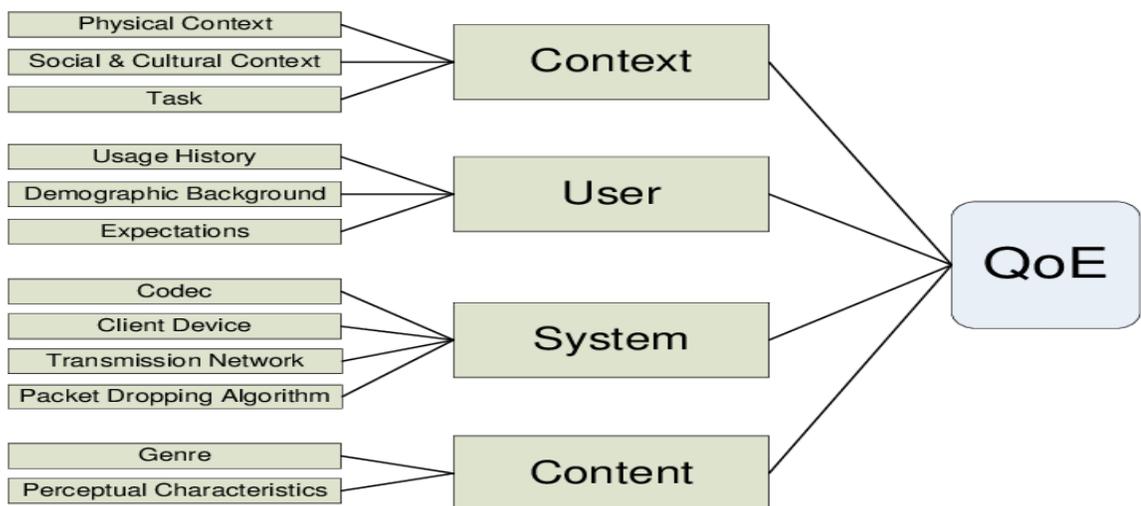
Figure (2.5): The Diagram of Influence Factors on QoE [58].

2.10 A Taxonomy for Measuring QoE

It is feasible to measure the QoE rate of a service using a variety of techniques. The primary classification of the potential techniques is determined by whether the QoE is evaluated manually by people or automatically by technical factors. In the first case, certain evaluation methods, also called subjective measures, are used. In the second case, quantitative frameworks are used for statistical calculations or algorithms.

Typically, subjective assessments are based on observational studies conducted in the real world with individuals to evaluate their application or service expertise firsthand. These users may assess the effectiveness of any stimuli in a passive (just see or listen) or active (interactive) manner (by participating in a conversation). Participants will be asked to evaluate how effectively a telephone service listens or speaks; the quality of the recording; etc. These evaluations must be meticulously organized in advance, and the user population must be selected based on the guidelines and suggestions of standards organizations [59].

There may be several subjective evaluation techniques incorporated. In addition, clients may evaluate the quality using an absolute scale, from best to worst, or by comparing successive images or photographs. The results measure customer views, previous impressions, preferences, user understanding, evaluation, explanation, etc., as well as performance, effectiveness, and total service engagement[60].

Such subjective evaluations are regarded as the most accurate since they comprise both explicit and implicit aspects of the study of human quality that are not otherwise obtained. Using just perceptual consistency experiments, the human

element may be portrayed legitimately and reliably. However, such subjective techniques might be considered accurate if they are constructed with care and the customers are objective and critical.

One disadvantage of the aforementioned technique is that the results of these tests are only relevant for assessing a specific service in the lab and not for improving QoE in the real world. Implementing a real QoE evaluation in which customers quantify their experience of a service or after the closure of a service is one technique to resolve the issue (post-service). One example is the OneClick paradigm, which may be used for QoE monitoring and input in real time, and hence for QoE power. This technique is activated by clicking a certain key when the user is dissatisfied with the program's consistency. An example of a post-service evaluation is provided by Skype, which provides users with the Mean Opinion Score (MOS) scale computation [60].

2.10.1 Mean Opinion Score (MOS)

The MOS is a numerical representation of the average outcome of an event or experience as determined by people. MOS is a measurement of voice and video quality. In regulated laboratories, subject techniques must be meticulously prepared in accordance with standard criteria. These directives describe everything, such as space conditions (e.g., control room, noise-free room), audio headsets and typically specific audio equipment, techniques, panel selection recommendations, etc. These last contain instructions for the quantity, age, background (experts or non-experts), prior involvement in similar research, randomization of their experiment selection, etc. In this sort of trial, streaming or download approaches are used to evaluate a service. These methods are seen as more applicable and accessible to a broader

audience than laboratory research, allowing for better management. In actuality, a substantial proportion of participants were able to identify very trustworthy and applicable CE values [60].

2.10.2 Crowdsourcing methods

Crowdsourcing methods are referred to as alternatives to this model's implementation in [59] since they eliminate subjective, anonymous internet users from the job of quality monitoring. This is shown by the Google Microworkers website and the Amazon Mechanical Trunk, in which internet users may execute Quality of Experience (QoE) tests for third parties (e.g., researchers) that need a big audience to complete an evaluation task. Ultimately, the distinction between immediate and everything consistency measures is a major challenge in subjective research methodology. The first strategy proposes that the subjective quality of the end user is continually assessed during the course of a single trial, while the second approach includes just a composite score for the user's impression at the conclusion of a single experiment. The first strategy provides device designers with a clearer concept, since instantaneous output may be connected with current technical requirements in the network. The latter, however, describes the user interface more comprehensively, Figure (2.6) Approaches to QoE modeling classification.
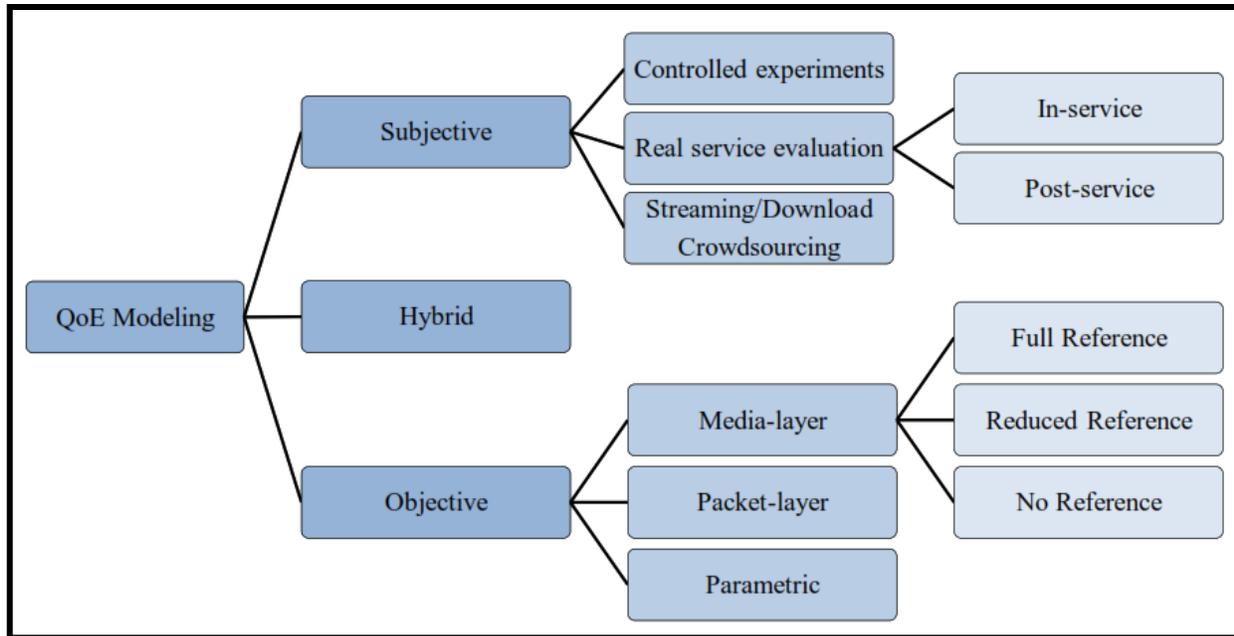
Figure (2.6): Approaches to QoE Modeling Classification [61].

2.11 Memory Effect and Degree of Interest

Many studies have been done on models of how video quality can be seen. Some studies attempt to quantify the effect of perceptual factors [62]. But authors usually ignore the memory effects [63], which are the temporal dynamics and historical experience of the user's satisfaction. Other studies are beginning to investigate the role of primacy and recency effects [64,65]. As a result, QoE predictions are very accurate. Most of the time, the primacy and recency effects [66,67] show how impairments that happen at the beginning and end of the streaming session affect memory.

In addition, bad things that happen in the middle of a session have a big effect on how good the video looks [67,68]. Theoretically, these effects can be shown by the forgetting curve for events that do not happen very often, or by repetition for impairments that happen over and over again. Existing QoE models [69,70], on the other hand, do not take into account how people forget and how often they do

things. So, along with the primacy and recency effects, the forgetting curve and repetition should also be discussed in order to fully explain how human memory affects QoE assessment. Aside from that, things that have to do with video content also have a clear effect on how QoE is seen. These things could be the type of video, how complicated it is, etc. Further, some studies such as [71] have found that the user's interest in video content may affect how he or she rates the QoE. More specifically, the user tends to give higher QoE scores for video content that is more interesting. This kind of behavior is affected by something called the degree of interest (DoI), which shows how interesting different video content is or how well it can catch and keep the user's attention [72]. Existing studies, on the other hand, often do not take this into account because these numbers might be different for different users based on their own interests [73].

2.12 Model of Objective Quality

To provide generally accurate QoE forecasts while avoiding the need for subjective testing, researchers build objective quality models. Models of objective quality calculate a measure based on QoS settings and external inputs. The output measure must have a strong correlation with the subjective test results, which serve as the QoE benchmark:

I.   The psychophysical and engineering approaches [74] are also referred to as the vision-based model and the signal-driven model respectively. Characterizing the processes of the human visual system (HVS), such as the masking effect, contrast sensitivity, and adaptation to color and light is the primary focus of the psychophysical approach. The engineering

method, on the other hand, is based on identifying and evaluating certain distortion patterns or video characteristics, such as statistical data, structural similarity (SSIM), and compression artifacts (e.g., blockiness and edginess).

II. Classification technique based on references [74]. The objective quality models are categorized as Full Reference (FR) model, Reduced Reference (RR) mode, and No Reference (NR) model based on whether or not a reference to the source video is required.

- ✓ Full Reference (FR) Model. Complete access to the video source is needed.
- ✓ Reduced Reference Model (RR). Partial video source information is needed.
- ✓ No Reference (NR) Model. No reference model requires access to the video source.

For quality comparison and evaluation, the complete reference and reduced reference models must refer to the original video, making them less suited for online QoE measurement. In the sense that they contribute an extra burden to the network or service, they are intrusive models [75]. No reference model is nonintrusive, adding no strain to the network or service. It is thus more appropriate for online QoE measurement and system adaptability. When deciding on a model or measure for online QoE assessment without a reference point, real-time performance and speed are also very important factors.

III.   The method for categorization is based on input data [76]. On the basis of the kind of input data, there are five model types:

- ✓ Models of the media layer whose input is the media signal.
- ✓ Parametric packet-layer model using packet header information as input.
- ✓ Parametric planning model using quality design factors as inputs.
- ✓ Bitstream layer model using packet header and payload information as input.
- ✓ Hybrid model, the mixture of several models.

The first two categorization systems are the most often used and frequently complement one another. In general, the psychophysical approach is associated with the FR, while the RR and NR rely mostly on the technical method. Numerous survey articles mention both categorization strategies, but often only use one. For instance, [77] and [78] mostly use the psychophysical/engineering approach classification technique. in [79], on the other hand, primarily use the reference-based classification method.

combines both classification methods. The third way of categorization is suggested in [80], the goal models are categorized as pixel-based models (e.g., PSNR and MSE), vision-based single-channel models, vision-based multi-channel models, and specialized models. However, this classification approach is not widely utilized. Figure (2.7) provides a review of the objective quality models on which they primarily focus. The researchers use FR/RR/NR as the first-tier classification, the psychological/engineering method as the second-tier classification, and other particular criteria as the third-tier classification. Note that some classifications are

not exclusive. For instance, similarity structural (SSIM) is an engineering method, but other variants of SSIM also include psychophysical characteristics in the design. As their primary foundation is SSIM, they will continue to identify these variants as engineering approaches. The investigators think that as research on objective quality models improves, classification algorithms will need to evolve.
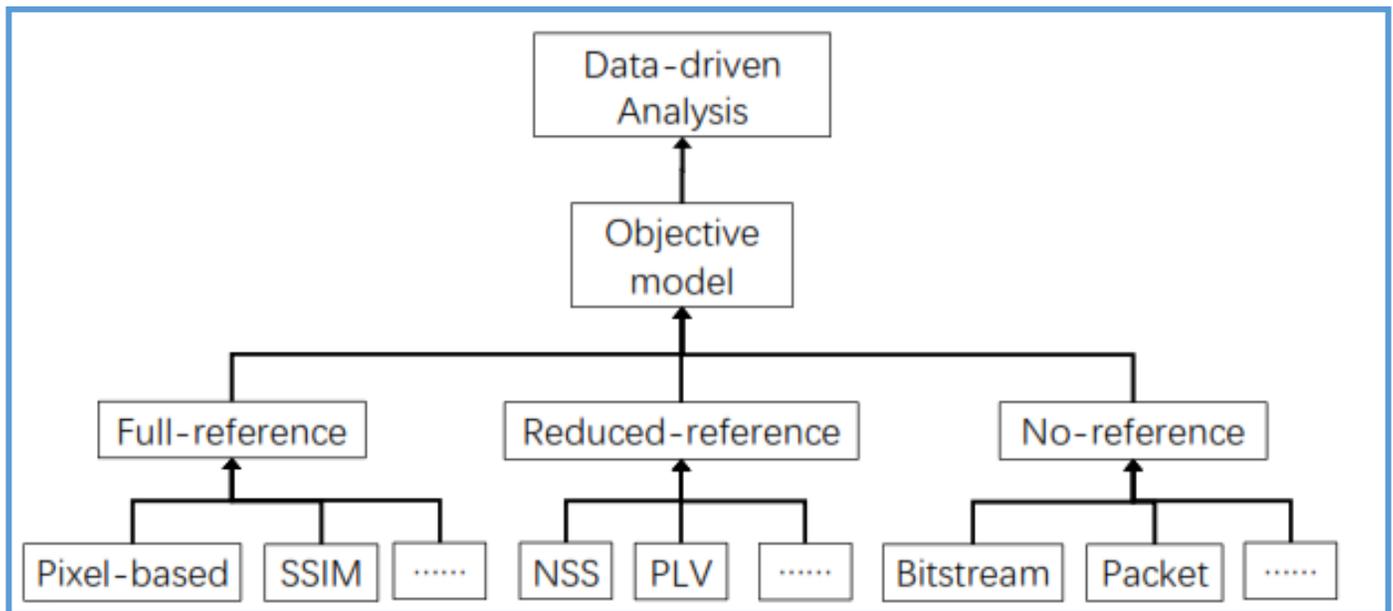
Figure (2.7): A General Introduction to The Objective Quality Models [74].

2.13 Video Quality Assessment (VQA)

Reduced Reference Spatio-Temporal Entropic Differencing Both Spatio Entropic Differencing (SRRED) and Temporal Entropic Differencing (TRRED) indices are based on conditional entropic differences using the Gaussian scale mixture (GSM) model for wavelet coefficients. The entropy is conditioned on the realization of the premultiplier random variable of the GSM model being its ML estimate. The conditional entropy calculations here require the GSM model and it does not appear to be possible to evaluate these using just empirical distributions.

The SRRED indices are computed in a manner similar .The STRRED indices combine the SRRED and the TRRED indices. The SRRED and TRRED indices operate individually on data obtained by separate processing of the spatial and temporal frequency components, relevant and important in the context of QA.

Interestingly, while the SRRED indices are obtained using only spatial frequency information, the TRRED indices are obtained using spatial and temporal information (the spatial information is used to weigh the temporal information). As a result, only the TRRED indices are influenced by temporal distortions, while both SRRED and TRRED indices are affected by spatial distortions. The computation of the quality index from the spatial and temporal information concerns the processing that occurs in the later stages of human visual processing, where there is evidence of interactions between the two . While we are inspired by these observations, and have used them in constructing VQA models.

The STRRED index is obtained as a product of the SRRED and TRRED indices and is expressed as (STRRED = SRRED . TRRED)

Although the TRRED index, indeed, does involve the local spatial variances through the scaling factor, it still only computes temporal entropic differences while only the SRRED index computes spatial entropic differences. Thus, it is the combination of both SRRED and TRRED indices that renders the STRRED index an effective VQA algorithm [81].

2.14 Statistical Performance Metrics

There are a number of statistical metrics used for testing the feasibility and robustness of the proposed system, In the next section it is explained briefly.

2.14.1 Evaluation Criteria for Assessing Metric

Performance evaluation is the most important part of all machine learning model development. So the question is how to measure the proficiency of the performance of the machine learning model. How would they know when to stop training and evaluating and call the project done [82].

For classification and regression, different metrics are used. Some metrics, like accuracy-recall, can be used for more than one job. Regression and classification are two examples of supervised learning, which is used in most machine learning tasks. To improve our model's ability to predict before put it into use on data not seen yet by using different ways to measure performance. If you do not make a proper evaluation of the Machine Learning model by using different evaluation metrics instead of just focusing on how well it works, you may end up with bad predictions when the model is used on data it has never seen [83].

The performance of an objective Visual Quality Assessment (VQA) is determined by correlating objective quality scores with subjective MOSs. Typically, correlation is measured using four parameters: linear correlation coefficient (LCC), root mean square error (RMSE), Outage Rate (OR), and Spearman's rank order correlation coefficient (SROC).

I.   The linear correlation coefficient, also known as the LCC:

(X) represent the actual value and the (Y) represent the predication value ,It illustrates linear correlations between (x) actual and (y) predication , with increasing degrees of strength in the relationship between the two variables. The linear correlation coefficient is a number that is derived directly from the data and indicates the degree to which (x) and (y) are associated with one another.

The value (r) that is the result of the following measurement is the linear correlation coefficient for a set of (x) pairs (y) of numbers that are included in a sample, as shown in equation (2.1):

$$r = \frac{SSxy}{\sqrt{SSxzSSyy}} \quad (2.1)$$

The linear correlation coefficient has the following properties:

 A. (r) has a value between 1 and -1, inclusive.

 B. The direction of the linear connection between (x) and (y) is indicated by the sign of (r)

 C. The magnitude of |r| represents the linear connection between (x) and (y):

  ✓ The linear connection between (x) and (y) is strong if |r| is close to 1 (that is, if r is close to either 1 or 1).

  ✓ If |r| is close to 0 (that is, if r is near 0 and of either sign). The linear connection between (x) and (y) is thereafter weak [85].

 II. RMSE between A and B may be calculated as:

$$RMSE(X,Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Xi - Yi)^2} \qquad (2.2)$$

The elements of **X** and **Y**, provided by $\mathbf{X}_i$ and $\mathbf{Y}_i$ (I = 1, 2..., n), are ranked for the two data sets X and Y, as shown in equation (2.2).

 III. SROCC is a non-parametric measure of rank correlation (statistical dependence between the rankings of two variables). How

accurately the link between two variables may be stated is evaluated by SROCC (whether linear or not). SRCC is then specified as:

$$SROCC(X,Y) = \frac{\sum_{i=1}^{n}(Xi - Y\check{})(Xi - Y\check{})}{\sqrt{\sum_{i=1}^{n}(Xi - X\check{})}\,\sqrt{\sum_{i=1}^{n}(Yi - Y\check{})}} \qquad (2.3)$$

X and **Y** represent the intermediate ranks of **X**$_i$ and **Y**$_i$ with regard to their magnitudes, where **X**$_i$ and **Y**$_i$ are the ranked data with respect to their magnitudes. Thus, as shown in equation (2.3), LCC measures the accuracy of the forecast, RMSE measures the error during the fitting process, and SROCC measures the monotonicity of the prediction, i.e. the degree to which the predictions of a metric coincide with the relative magnitudes of MOSs. Larger LCC and SROCC values suggest a stronger connection between objective image quality ratings and MOSs. Lower RMSE values, on the other hand, indicate less prediction error and, hence, improved performance. As an error measure, to use the outage rate, which was also employed in [86]. Specifically, the outage rate of a time-varying subjective quality (TVSQ) model is defined as the frequency with which the estimated TVSQ deviates by at least twice the observed TVSQ's confidence interval. Specifically, the outage rate may be expressed as shown in equation (2.4) :

$$E(\theta) = \frac{1}{NT}\sum_{n=1}^{N}\ \ \sum_{t=1}^{T}(|q\hat{}n(t,\theta)| > 2 \in n[t]) \qquad (2.4)$$

2.14.2 K-fold Cross Validation

Cross-validation is a process to estimate the quality of an (RNN) model by selecting the best set of parameter values (such as the number of hidden nodes, learning rate, and so on) applied to several neural networks. Unreliable evaluation for the Deep Learning model may occur with the unfair division of dataset to testing and training sets such that they are not representative of all data. The k-fold cross-validation reports the performance measure then computes the average of the values in the loop [87]. This approach repeats the whole process (i.e. training and testing) many times according to the K value with different samples of data. Computationally, this approach is expensive but actually does not waste too much data, such as where the number of samples is very small. In K-fold cross-validation, the prediction model repeats K times and in each iteration 1-K of data is selected for testing and the remainder used for training. Then the overall error rate is calculated by computing the average error of all iterations [87,88]. The procedure of k-fold cross-validation with training and testing is shown in Figure (2.8).
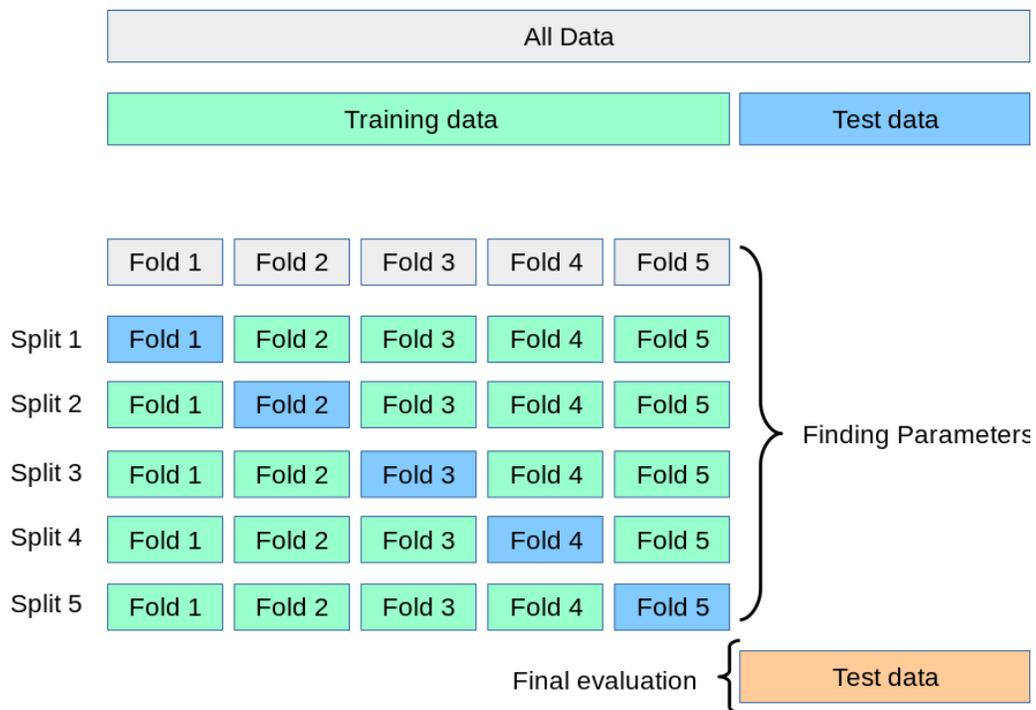
Figure (2.8): Cross-Validation Procedure With k-Folds [89].

2.15 The Effect of Hysteresis

The effect of hysteresis is the dependence of a system's current state on its past. This indicates that the nature of the QoE process is non-Markovian, as there is a memory of a chain of events beyond the first occurrence, which are explained in more detail below. The QoE process can be subject to a range of long- and short-term dependencies due to its continual evolution. However, training RNNs to capture long-term dependencies is challenging due to the fact that gradients tend to either fade (most of the time) or explode (rarely, but with severe effects) [90]. Consequently, a non-Markov understanding was used to predict the continuous video quality while preserving the previous prediction for each video view and utilizing the gated recurring unit (GRU) so that these requirements can be modeled using a higher-order procedure, as shown in Figure (2.9), where the current customer experience has an effect on the quality values of the past experience.
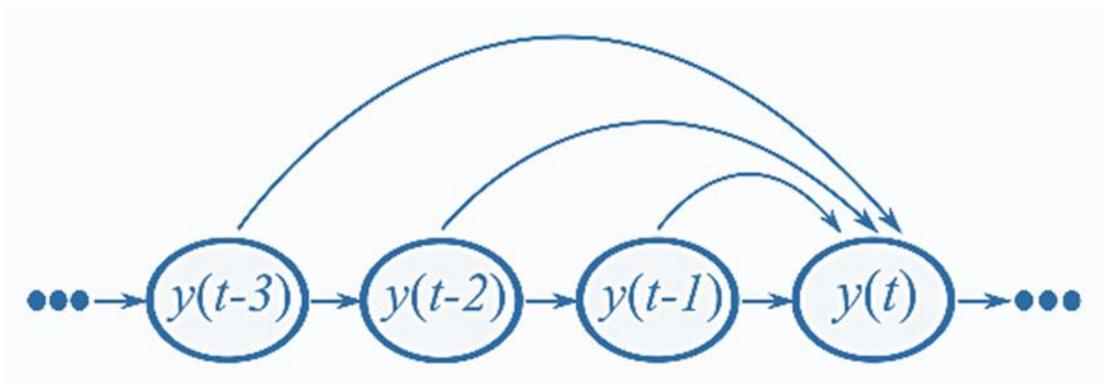
Figure (2.9): Non-Markovian Dynamic of QoE Process With Long Stat-Term State [91].

Non-Markovian processes are stochastic processes that lack the Markov property. The Markov property, often known as the memoryless property, says that the conditional probability of a future state depends solely on the current state and is independent of all previous states. So the Gates Recurrent Unit (GRU) architecture was used to predict the value of Continuous Quality.

2.16 Machine Learning Algorithms (ML)

The advancement in computational applications and frameworks provides solutions to our everyday problems. Machine learning is one of the computational applications of algorithms and statistical models to carry out a task without explicit instructions, but with the use of patterns to give inference. Machine learning refers to the use of computer algorithms that support systems operation in training to automatically learn and enhance data to predict or classify the nature of such data through the use of patterns [92]. Generally, machine learning is a subfield of artificial intelligence that allows the systems to make decisions autonomously with

no external support. The decision is made by finding valuable hidden layers of patterns within the complex data.

The machine-learning approach depends on the data type for input and output operation and problem type which is based on the applications on data for decision making and an embedded instruction to carry out the assignment with minimum supervision from the programmers [92]. Machine learning is classified as supervised learning, semi-supervised learning, and unsupervised and reinforcement learning while there are few hybrid approaches and other common methods [93].

2.16.1 Machine learning techniques

The categories of Machine Learning Techniques are mainly divided into four categories: Supervised learning, Unsupervised learning, Semi-supervised learning, and Reinforcement learning [92]. The techniques are discussed further according to their applicability for solving real-world problems

.

A. Supervised learning

In the supervised learning category of machine learning, the algorithms (step by step method of solving a problem in a particular format) operate in such a way that they will develop a mathematical model (translating or encoding a problem into mathematical formulations) of the data which comprises the input (data sent to a computer system) and the expected output (processed information sent out from a computer) [133]. The data supplied is categorized as the training data which comprises the sets of training examples with one or more inputs. The mathematical modeling is applied in the supervised learning using array vector (feature vector for

extraction) and the data to be trained by matrix. The algorithm that enhances and improves the outcomes in the accuracy of the outputs for classification or prediction purposes has learned the task and therefore it can give a good outcome [95].

B. Unsupervised learning

Unsupervised learning algorithms operate in such a way that they takes a set of data and detect the patterns in it for grouping or clustering purpose. Unsupervised learning algorithms identify resemblance in the data and react based on the presence or absence of such identity in each new piece of data. The algorithms learn from test data that is not labeled, classified, or categorized. Unsupervised learning analyzes unlabeled datasets without the need for human interference, i.e., a data- driven process [96]. The common unsupervised learning tasks are anomaly detection, dimension reduction, clustering, density estimation, feature learning, finding association rules, etc. [97].

C. Semi-supervised learning

The semi-supervised learning is situated between unsupervised learning (with no labeled training data) and supervised learning (with labeled training data). It is a hybrid form of machine learning techniques because it operates on labeled and unlabeled data which brings a better accuracy. The major aim of unsupervised learning is to give better outcomes for prediction than the ones done with labeled data. The application areas of semi-supervised learning are text classification, fraud detection, machine translation, etc. [95].

D. Reinforcement learning

Reinforcement learning in machine learning parlance is concerned with the use of software agents and machines to make the decision automatically in an environment to improve efficiency. Generally, reinforcement learning is used in operation research, game theory, information theory, swarm-intelligence, genetics algorithms, etc. The learning uses the reward or penalty system, and the primary goal is to use leading obtained from environmental parameters to validate the reward or to minimize the risk involved. The algorithms are used in autonomous vehicles or in learning to play a game against a human opponent, It is an effective tool in training AI models to increase automation which is used in robotics, autonomous driving tasks, manufacturing, and supply chain logistics [92].

2.16.2 Machine learning and deep learning

Deep learning refers to a distinctive subtype variant in the machine learning, It is also a subclass in the domain of artificial intelligence (AI). Furthermore, Machine learning primarily means a computer that learns from data and makes predictions using algorithms. Machine learning yields to some environmental parameters. Conversely, deep learning operates in a quick manner and adapts to it using constant feedback in building on the models. Deep Learning system leverage on the Neural Networks which imitates the human brain with an embedded multiple-layer architecture. It also learns through the data to carry intelligent decisions [98].

2.16.3 Neural networks

Neural Networks are an aspect of machine-learning that comprises different node layers which include the input, hidden, and output layers. The Network is used in most deep learning architectures. Neural Networks work in a manner that the nodes connect with their different weight and threshold. More so, if a node for

instance has an output that is more than that of the threshold, then it will be triggered and will send the data involved to the layer that is next and if not, there will be no data activated to the succeeding layer in the network [99].

2.16.4 Convolutional neural networks

The traditional manual process employed in the prediction and classification of images convincingly wastes time. Wrong diagnosis is another major problem attributed to it. The convolutional neural network provides a better and more scalable method in the computer vision process. The CNN involves the identification of images through a computational approach that combines linear algebra and matrix multiplications. CNN outperformed other networks in applications like image processing and speech recognition. The CNN has three parts: the convolutional, pooling, and fully-connected layer. The convolutional part is where the major computation happens to be the building block among the three and it comprises the data, filter, and feature area. The pooling layer is responsible for the data sample dimension reduction known as down-sampling. The pooling layer also holds a filter and it moves over the input, but may not have weight. Pooling is sub-divided into Max and Average pooling with the functions of calculating the maximum and or average value respectively. In the fully-connected layer, output layers are fully joined via a node to former layer and do classification tasks through the feature extracted from the preceding layer [99].

2.16.5 Recurrent Neural Networks (RNN)

Backpropagation was first used to train recurrent neural networks, which was one of the most exciting things about it (RNNs). RNNs are often better for tasks like speech and language where the inputs come in that certain order. RNNs

process a sequence of inputs one item at a time. They keep a state vector in their hidden units that contains implicit information about the history of all the previous items in the sequence. When work with the outputs of the hidden units at different discrete time steps as if they were the outputs of different neurons in a deep multilayer network, it becomes clear how backpropagation can be used to train RNNs.

RNNs are very powerful dynamic systems, but training them has been hard because the backpropagated gradients either grow or shrink at each time step, So after a lot of time steps, they usually either explode or disappear [100],[101].

A. Gated recurrent units (GRUs)

Several strategies for dealing with timely successive events have been introduced. Long-term memory is one of the first [Hochreiter & Schmidhuber, 1997 [102]]. Both strategies will be explaining, and why the gated repetition unit (GRU) was used instead of the standard recurring unit (SRU) or long-term memory (LSTM). Details of this strategy will be reviewed.

B. Hidden State Gated

The main difference between RNNs and regular GRUs is that GRUs have a gate hidden state. This means that it has special ways to know when to change the hidden state and when to reset it. For example, if the first second is very important, the GRU gate architecture learns not to change the hidden state after the first observation. In the same way, hidden state gates will learn to skip unimportant temporary notes. The gates work automatically, learning how to

reset the base state whenever they need to. This will be explained in more depth below.

C. Reset Gate and Update Gate

Resetting the gateway and the update gateway are the first things. They must be set up so that they are vectors with entries in the range [0,1], so the shape is like a convex combination. For example, the reset gate determines how much of the previous state you want to keep. In the same way, the update gateway controls how much of the new state is just a copy of the old state. Figure (2.10) describes the inputs for both the reset and refresh gates of the GRU, considering the inputs for the current time step and the hidden state of the previous time step. Two fully connected layers with sigmoidal activation functions give gate outputs.
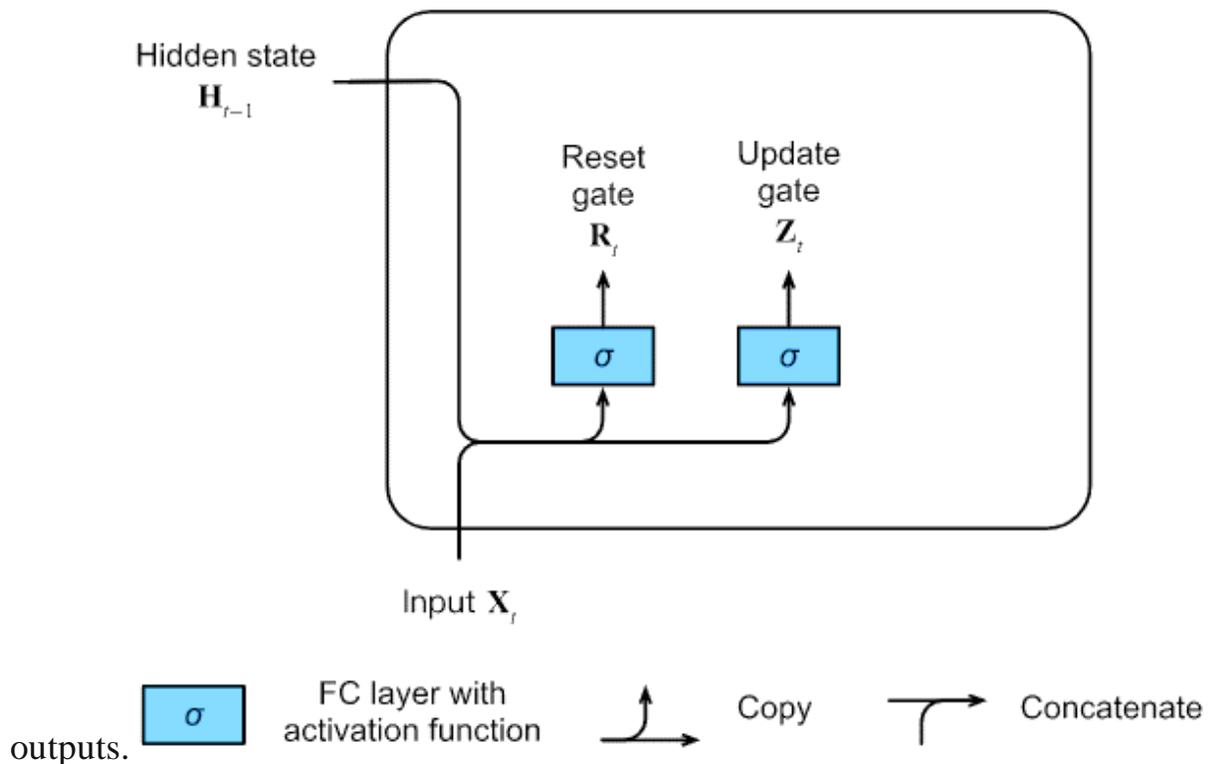
Figure (2.10): The Reset Gate and Update Gate in A GRU Model.

Mathematically, for a given time step t, suppose that the input is a minibatch $\mathbf{X}_t \in \mathbf{R}^{n\times d}$ (number of examples: n, number of inputs: d) and the hidden state of the previous time step is $\mathbf{H}_{t-1} \in \mathbf{R}^{n\times h}$ (number of hidden units: h). Then, the reset gate $\mathbf{R}_t \in \mathbf{R}^{n\times h}$ and update gate $\mathbf{Z}_t \in \mathbf{R}^{n\times h}$ are computed as follows :

$$\mathbf{R}_t = \sigma(\mathbf{X}_t\mathbf{W}_{xr}+\mathbf{H}_{t-1}\mathbf{W}_{hr}+\mathbf{b}_r), \qquad (2.5)$$

$$\mathbf{Z}_t = \sigma(\mathbf{X}_t\mathbf{W}_{xz}+\mathbf{H}_{t-1}\mathbf{W}_{hz}+\mathbf{b}_z),$$

where $\mathbf{W}_{xr}, \mathbf{W}_{xz} \in \mathbf{R}^{d\times h}$ and $\mathbf{W}_{hr}, \mathbf{W}_{hz} \in \mathbf{R}^{h\times h}$ are weight parameters and $\mathbf{b}_r, \mathbf{b}_z \in \mathbf{R}^{1\times h}$ are biases., As shown in the equation (2.5), it uses sigmoid functions to transform input values to the interval (0,1) [103].

D. Candidate Hidden State

Integrate the reset gate $\mathbf{R}_t$ with the regular latent state updating mechanism leads to the following *candidate hidden state* $\mathbf{H}_t \in \mathbf{R}^{n\times h}$ at time step *t*,, as show in the equation (2.6):

$$\mathbf{H}_t = \tanh(\mathbf{X}_t\mathbf{W}_{xh}+(\mathbf{R}_t \odot \mathbf{H}_{t-1})\mathbf{W}_{hh}+\mathbf{b}_h) \quad (2.6)$$

where $\mathbf{W}_{xh} \in \mathbf{R}^{d\times h}$ and $\mathbf{W}_{hh} \in \mathbf{R}^{h\times h}$ are weight parameters, $\mathbf{b}_h \in \mathbf{R}^{1\times h}$ is the bias, and the symbol $\odot$ is the Hadamard (elementwise) product operator. Then using a nonlinearity in the form of tanh to ensures that the values in the candidate hidden state remain in the interval $(-1,1)$.

The result is a candidate since it still needs to incorporate the action of the update gate whenever the entries in the reset gate $\mathbf{R}_t$ are close to 1. For all entries of the reset gate $\mathbf{R}_t$ that are close to 0, the candidate hidden state is the result of (multilayer perceptron's) MLP, as previously explained, with $\mathbf{X}_t$ as the input. Any pre-existing hidden state is thus *reset* to its defaults.

E. Hidden State

It needs to incorporate the effect of the update gate $\mathbf{Z}_t$. This determines the extent to which the new hidden state $\mathbf{H}_t \in \mathbf{R}^{n \times h}$ is just the old state $\mathbf{H}_{t-1}$ and by how much the new candidate state $\mathbf{H}_t$ is used. The update gate $\mathbf{Z}_t$ can be used for this purpose, as shown in the equation (2.7), simply by taking elementwise convex combinations between both $\mathbf{H}_{t-1}$ and $\mathbf{H}_t$. This leads to the final update equation for the GRU:

$$\mathbf{H}_t = \mathbf{Z}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}_t) \odot \mathbf{H}_t. \qquad (2.7)$$

Whenever the update gate $\mathbf{Z}_t$ is close to 1, by simply retain the old state. In this case, the information from $\mathbf{X}_t$ is essentially ignored, effectively skipping time step t in the dependency chain. In contrast, whenever $\mathbf{Z}_t$ is close to 0, the new latent state $\mathbf{H}_t$ approaches the candidate latent state $\mathbf{H}_t$. These architectures may help overcome the issue of vanishing gradients in RNNs and capture dependencies for sequencing with large time step intervals more accurately [103].

For example, if the update gate has been near to (1) for all time steps of a subsequence, the old concealed state at the time step of the subsequence's commencement will be readily kept and transferred to its conclusion, regardless of the subsequence's duration.

In brief, the GRUs are distinguished by the two characteristics listed below:

✓ In sequences, the use of reset gates is helpful in capturing short-term dependencies.

✓ Using update gates to their full potential may assist in capturing long-term dependencies in sequences.

✓ GRUs train faster and perform better than LSTMs on less training data if you are doing language modeling (not sure about other tasks).

✓   GRUs are simpler and thus easier to modify, for example adding new gates in case of additional input to the network. It is just less code in general [103].

2.17 Normalization

Min-max normalization one of the most used data normalizing techniques. For each feature, the lowest value is converted to a zero, the highest value is converted to a one, and all other values are converted to a decimal between zero and one. As shown in the equation (2.8), the application of the Min-Max Normalization approach is as follows:

$$A = \left(\frac{A - \min\ value\ of\ A}{\max\quad value\ of\ A - \min\ of\ A} \right) \qquad (2.8)$$

Where, A includes Min-Max Normalized data. Then. if A is the original data range, then to make the dataset well-organized or structured, the maximum value of the continuous score should be (100) and the lowest value should be (0).

As shown in equation (2.9), use the formula below to execute z-score normalization on each item in a dataset:

$$New\ Value = (x - \mu) / \sigma \qquad (2.9)$$

Where x is the initial value, the mean of data, and the standard deviation of data, it was not able to compute temporal Differential Mean Opinion Scores (DMOS) since the created video patterns are of varying lengths due to the inclusion of rebuffering events [104].

54

Chapter Three

Proposed QoE Measurement System

Chapter Three
Proposed QoE Measurement System

3.1 Overview

This chapter is devoted to describing the proposed QoE measurement system for a No-Reference video stream by predict the subjective opinion of the viewer. To achieve this, by use a combination of the methods mentioned in the previous chapter.

3.2 The Architecture of the GRU-QoE for Continuous Video

The scheme in Figure (3.1) explain the layout architecture of the proposed model. The general framework of this model is implemented through the following stages:

1. Using four separate data sets on the same model to measure its efficiency, it is subdivided into four main parts: Live Mobile Stall Video Database II, LIVE-Netflix mobile VQA database, LFOVIA QoE Database, and LIVE QoE Database II. including different videos (category, size, resolution and device).

2. The preprocessing stage has several steps for each database, including filtering and normalization using two algorithms depending on the type of the dataset (Min-max and Z-Score normalization).

3. In the feature extraction stage, features which have the greatest impact on the video quality are extracted through the use of correlation and p-value algorithms. A new feature is also created from features in the video database.

4. In Train and Test Stages, the GRU-QoE algorithm is used in the training and validation process. The proposed model train consists of all training datasets available for learning on the set determined by the video datasets from the total of all videos in the database. The rest of the dataset is transferred to the testing stages. The appropriate k-fold algorithm is used to ensure that the training and testing process is fair for all data.

5. Assessing outcomes in order to calculate the median and mean rate and use measurements, the performance of an objective visual quality measure was determined by correlating objective quality scores with subjective MOSs. Most of the time, correlation is measured using four parameters: Linear correlation coefficient (LCC), root mean square error (RMSE), outage rate (OR), and Spearman's rank-order correlation coefficient (SROC).

Dataset

Feature Selection By
(Correlation)-(p-value)

Normalization Preprocessing By

(z-score)

Feature extraction
Generate new feature

Training

Bidirectional RNN
Layer with GRU-QoE
Model with features

Testing

Save trained Model

Predication of Evaluation
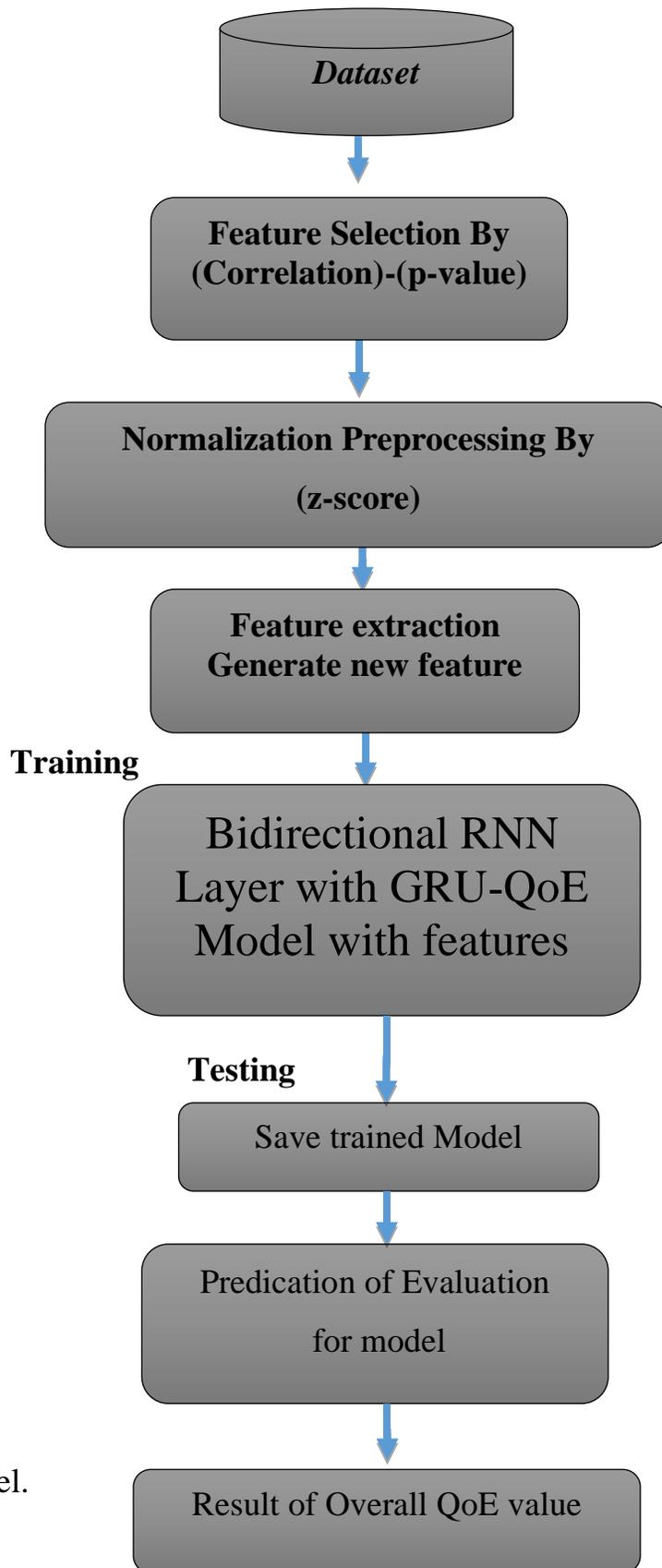
for model

Figure 3.1:

The Proposed

Framework Model.

Result of Overall QoE value

3.3 Dataset

To evaluate the performance of the presented QoE assessment approach, there are four comprehensive QoE databases available, These were used to implement the proposed model and the reason for its adoption because of the diversity and variance in the data used. as shown in the Figure (3.2). The details of this datasets are described along with training and testing procedure followed for desaturation as shown in the figure below together with their ratio.
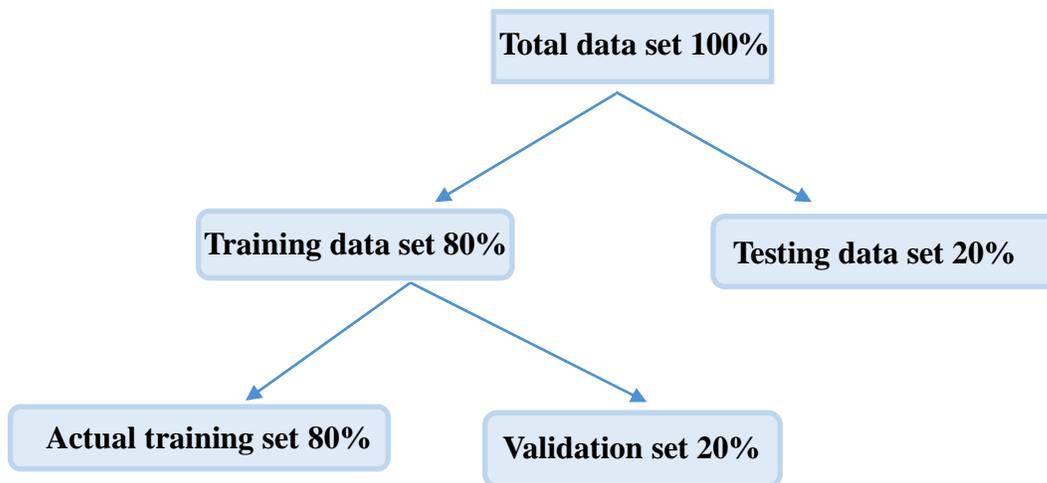
Figure (3.2): The dataset is divided into percentages for training, Testing and evaluation.

3.3.1 The Live Mobile Stall Video Database II

The LIVE Mobile Stall Video [105] collection is comprised of 174 distorted videos created from 24 reference videos, including 26 distinct stall occurrences and rich continuous-time personal opinion ratings derived from 54 subjects. It includes 24 creative high definition (HD) videos licensed

from YouTube and Vimeo (with audio). The original resolutions of these generic films were 1280 x 720, 1280 x 640, 480 x 360, 484 x 360, 490 x 360, 540 x 360, and 640 x 360. All videos are 30fps. Any visual distortions were considered low or undetectable. In order to concentrate only on network vulnerability, movies with stuttering or lag were omitted; As a result, each of (24) detected video sequences included limited spatial distortions or unexpected camera shaking, as shown in Table (3.1). A video was selected from each of these video sequences that was linguistically consistent, experimental, and lengthy enough to be relevant on its own. After twice reloading, durations of these films range from 29 to 134 sec.

TABLE (3.1): Displays the categorization of 24 video materials into five main groups.

Sports	Talk shows / Documentaries	Music	Advertisements Newscasts	Newscasts
9	8	2	3	2

This dataset aims to comprehend the impact of dynamic network impairments, (such as varying patterns of rebuffering events and startup delays) on the QoE of mobile video viewers. It also analyzes the continuous-time subjective behavior of the subjects in order to better comprehend the temporal fluctuations in perceived QoE caused by variables such as stall placements, the number and duration of stalls, and changing video content.

Subjects appear to fetch their experiences under the impact of recency, primacy, and repetition activating when making decisions about the overall perceived quality of a video while viewing it, Subjective QoE scores recover to lower values at the end of a stall event as the number of stalls increases.

Table (3.2) Description of four stall parameters (left column) and different values of these parameters considered constructing the stalling patterns in our database. L refers to the total length of a given video.

Table (3.2): Contains information on the range of values for these four parameters for various stalling patterns.

Number of stalls	Few ($1 - 3$ stalls)	Many ($4 - 7$ stalls)		
Stall length	Short ($2 - 4$ sec.)	Medium ($5 - 9$ sec.)	Long ($10 - 15$ sec.)	
Position of stalls	Beginning (between $0 - L/2$ sec.)	Middle (between $L/4 - 3L/4$ sec.)	End (between $L/2 - L$ sec.)	Uniformly throughout (between $0 - L$ sec.)
Startup delay	Short ($0 - 7$ sec.)	Long ($8 - 20$ sec.)		

Many stall patterns were created by modifying four distinct characteristics:

➢ start-up delay duration

➢ stall lengths

➢ stall placements

➢ number of stalls.
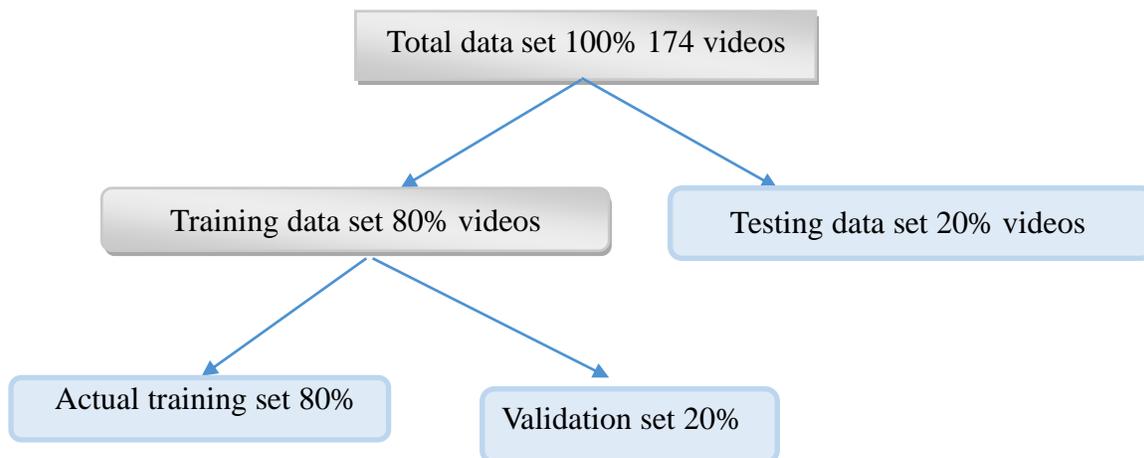
Figure (3.3): The whole collection consists of 174 videos with distortion.

Each dataset includes 24 reference movies with 26 distinct halting incidents, as well as detailed, continuous-time subjective opinion ratings from 54 individuals. The whole dataset partition is illustrated in Figure 3.3.

3.3.2 LIVE Netflix dataset

The LIVE-Netflix mobile VQA dataset [106] consists of 112 distorted movies assessed by more than 55 human participants on a mobile device. Subjective ratings were applied to 14 video contents of spatial resolution 1080p at (24, 25 and 30) fps by imposing a set of (8) alternative playout patterns, which included dynamically varying H.264 compression rates, rebuffering events, and combinations of both. While more contemporary compression standards like H.265/HEVC and VP9 are in the process of being developed, H.264. It is now the most used format.

Moreover, while H.265 achieves greater efficiency than H.264, it is theoretically identical to H.264. It employs the same hybrid motion-compensated/transform/lossless entropy coding architecture and virtually the same coding tools. Hence, coding artifacts are perceptually identical across these two codecs, The results of this research will also apply to H.265-based broadcasts with lower encoding bit rates correctly.

The database covers 12 distinct categories of Netflix material (drama, action, comedy, anime, etc.) and three publicly available categories. This dataset was built based on frequent playout circumstances while streaming under realistic bandwidth and buffer capacity limits. It was a collection of fourteen video test items with a broad range of spatiotemporal features. Eleven of the fourteen contents are Netflix titles, featuring action sequences, drama, adventure, anime, and cartoons. The following three clips were collected from the Consumer Digital Video Library (CDVL) [109]. There is a range of frame rates and resolutions in the

test material. It boasts 4K resolution (4096 x 2160) and a frame rate of 60 fps, while most Netflix movies have 1080p resolution (1920 x 1080) and frame rates of (24, 25, or 30 fps).

In order to appease this difference, the sequence was resized to (1080p) and the frame rate was reduced from 60 fps to 30 fps. Spatial and temporal complexity measurements provide an approximation of the content variability of a subjective database. The total dataset division is illustrating in Figure (3.4).
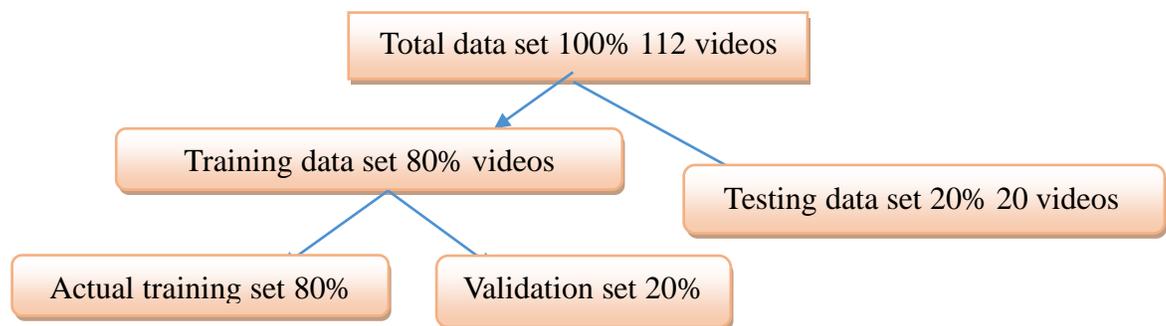
Figure (3.4): The dataset divisions' ratio according to training and testing process.

3.3.3 LFOVIA Quality of Experience Database

The LFOVIA QoE dataset [107], which contains 18 uncompressed reference films and 36 distorted 120-second video sequences, is the only dataset which includes films with resolutions up to ultra hd. The dataset accounts for both rebuffering events (rebuffering frequency and rebuffering duration) and quality switching (various resolution-bitrate pairs) under realistic conditions (though the ideal fixed duration up and down switching may not be too realistic). A dataset like this, which has both continuous and total ratings, is enough to build and test applications for the real world. In the plot indicates the reference video as per the Table (3.3) and Figure (3.5).

TABLE (3.3): Identifier for the database's reference videos.

Video	Name	Resolution	Frame Rate	Format
(a)	Surfers	FHD	30	YUV422
(b)	Boxing	FHD	30	YUV422
(c)	Football	FHD	30	YUV422
(d)	Goose Park	FHD	30	YUV422
(e)	Tulip	FHD	25	YUV422
(f)	Euro Truck	FHD	60	YUV420
(g)	Under Water	FHD	30	YUV422
(h)	Big Buck Bunny	FHD	24	YUV420
(i)	Skating	FHD	30	YUV422
(j)	Restaurant	UHD	30	YUV420
(k)	Theme Show	UHD	30	YUV420
(l)	Flight	UHD	30	YUV420
(m)	Disney World	UHD	30	YUV420
(n)	Hostel	UHD	30	YUV420
(o)	Dining Block	UHD	30	YUV420
(p)	Highway	UHD	30	YUV420
(q)	Vegetable Market	UHD	30	YUV420
(r)	Taj Mahal	UHD	30	YUV420

Bitrate switching and rebuffering events have affected the playout patterns of these videos. The overall QoE and time-varying instantaneous QoE scores for those videos were acquired from this database, and they are in the range [0, 100], with 0 being the worst and 100 being the greatest. As shown in the Figure (3.5), the collection of distorted movies was separated into training and testing sets with an 80:20 training : testing ratio. As a result, the training set had (25) movies and the testing set included 10 films. In [110], there were separate training and testing sets, separately.
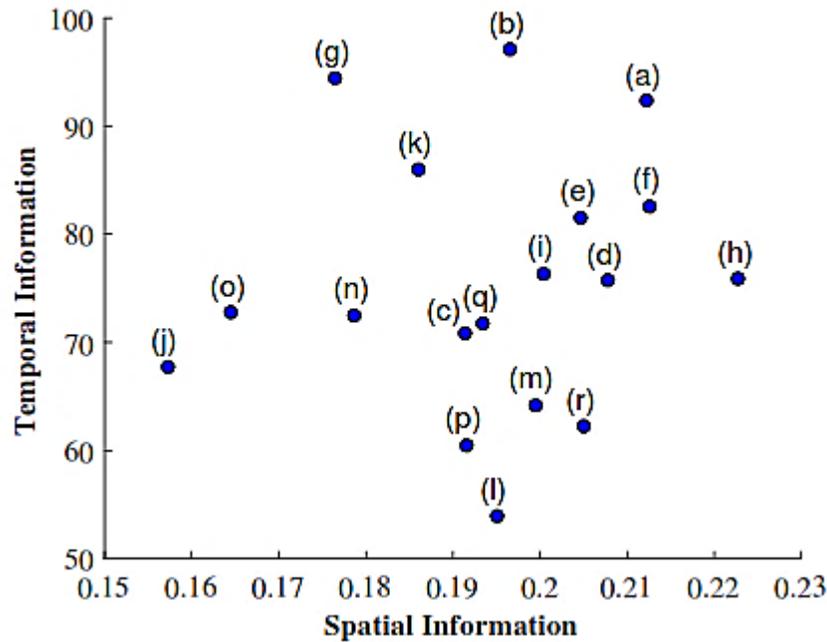
Figure (3.5): Spatial and temporal information indices.

3.3.3.1 Development of Reference Videos

The reference films at each resolution are created by concatenating source videos of varying lengths. The concatenation of small video clips into larger video sequences was used to examine the time-varying quality.

During concatenation, care is taken to retain the resolution, frame rate, and decoded video format. For the Big Buck Bunny animated video sequence, the nearly 10-minute-long reference footage was shortened to (120) seconds from the beginning. The raw YUV video format is kept for all reference video sequences. depicts a typical frame from each reference video sequence in the database. The first nine videos are in FHD, while the remaining videos are in UHD.

The terminology and characteristics of these movies are shown in Table (3.3). The films picked for the research incorporate a broad range of topics, including nature, wildlife, outdoor, marine, sports, animation, and gaming. The availability of a wide range of information in the reference movies guarantees that the videos

stimulate and maintain the participants' attention as they watch, enabling them to react effectively to distortions. To examine the effects of rate adaptation and rebuffering on the user's QoE, these distortions were intentionally added into the reference movies to produce distorted video sequences. The total dataset consists of 36 distorted videos. The total dataset division is illustrated in the Figure (3.6).
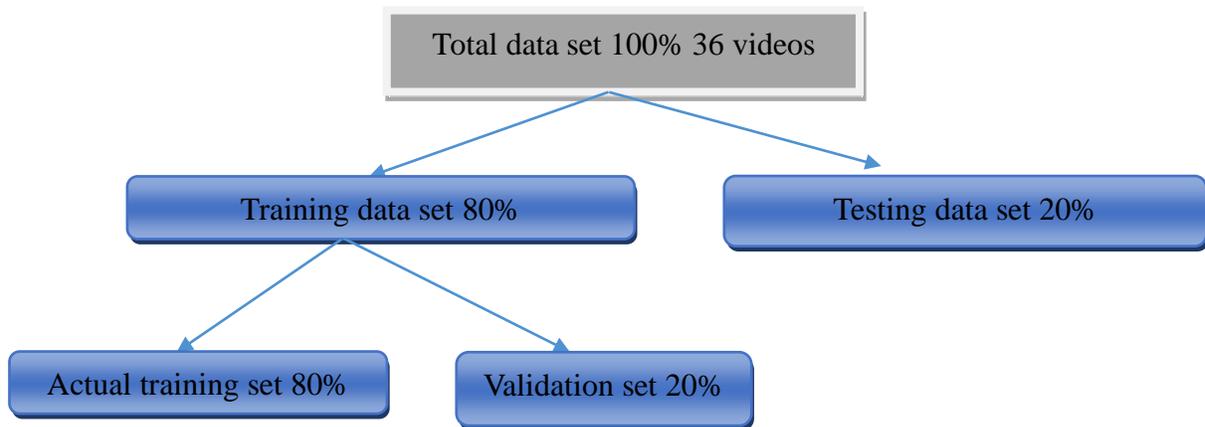
Figure (3.6): The dataset divisions' ratio according to the training and testing process

3.3.4. Live QoE Video Database II

LIVE QoE Database II [108] is a dataset that employs a Hammerstein-Wiener model to forecast the time-varying subjective quality (TVSQ) of rate-adaptive movies. To get data for model calibration and validation, a library of longer-duration movies with changing distortions over time was made, The subjective TVSQs of videos were measured in a the large-scale subjective study.

Quality-varying the 15 movies of varied quality were produced in such a way that their STSQs fluctuated randomly throughout time. The selection consists of eight uncompressed, high-quality video clips with distinct content. These videos have a 720p resolution (1280 * 720) and the frame rate of 30 frames per second. The QoE scores for the videos in the database vary from 0 to 100, with 0 being the

poorest score and 100 being the highest. In Table (3.4), a brief summary of these clips is given.

TABLE (3.4):A BRIEF DESCRIPTION OF THE VIDEO CLIPS IN DATABASE.

Name	Abbreviation	Description
Fountain	ft	Still camera, shows a fountain.
Turtles	tu	Still camera, a girl is feeding turtles.
Stick	st	Still camera, a man is waving a stick.
Bulldozer	bu	Camera span, a man is driving a bulldozer.
Singer & girl	sg	Camera zoom, a man is singing to a girl.
Volleyball	vo	Still camera, shows a volleyball game.
Dogs	do	Camera span, two dogs play near a pool.
Singer	si	Camera zoom, a singer is singing a song.

Three reference videos are produced. Each reference video was created by stringing together five or six segments see Figure (3.7). This is how the reference video was prepared since extended movies with repetitive information might bore subjects. This might negatively impact the TVSQ measurement accuracy in the subjective research. Each film is five-minutes length, making it a more effective tool for comprehending the long-term behavioral reactions of the human visual system.
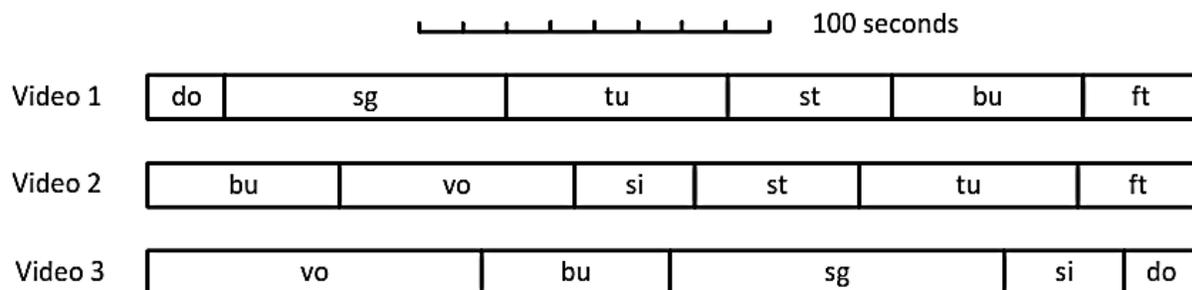
Figure (3.7) represents the development of the reference videos. The abbreviations for the clip names are included in Table (3.4).

The dataset includes 28 distorted copies of each reference video produced. Specifically, every source video sequence was encoded using the H.264 encoder into 28 constant bitrate streams, which were then decoded. Encoding bitrates ranging from hundreds of kilobytes per second to several megabits per second were used to provide a broad variety of video quality examples.

The total dataset consists of 15 distorted videos. The total dataset division is illustrated in Figure (3.7).

Total data set 100% 15 videos

Training data set 80% 10 videos

Testing data set 20% 5 videos

Actual training set 60%

Validation set 20%
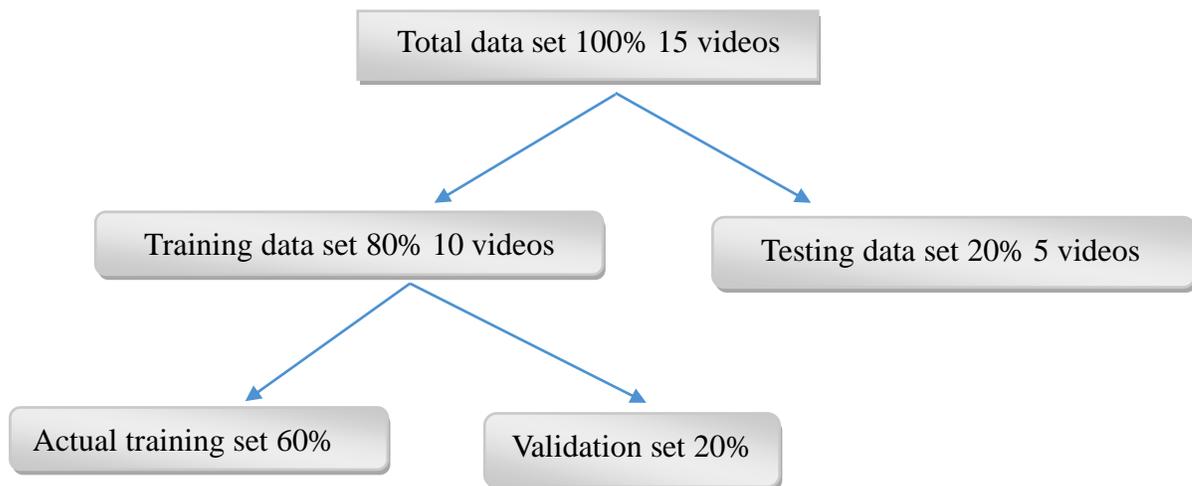
Figure (3.7): The dataset divisions' ratio according to the training and testing process.

By analyzing each database separately, containing a number of features for all videos, the sum of the total number of features and the number of videos for each database is summarized in Tables (3.5):

Table (3.5): A summary of four databases used in the dissertation.

Dataset	No. of Videos	Number of Features	Resolution	Time-varying Quality	Rebuffering	QoE range
Live Netflix[106]	112	102	1920X1080	Yes	Yes	(-2.6,1.9)
LFOVOA QoE[107]	36	17	3840X2160 1920X1080	Yes	Yes	(0,100)
LIVE QoE[108]	15	8	1280X720	Yes	No	(0,100)
LIVE Mobile[105] Video Stall-II	174	14	Diverse	No	Yes	(0,100)

3.4 Feature Selection

The Selection of features from the streaming video for containing QoE database is one of the challenges of the proposed model. To reduce the number of input variables to both reduce the computational cost of modeling and to improve the performance of the model. The estimation stage is where it must be done without any delay during the streaming process. Statistical measures for feature selection must be carefully chosen based on the data type of the input variable and the output or response variable. Feature selection is primarily focused on removing non-informative or redundant predictors from the model. Using Correlation and p-value algorithms helps to feature selection by removing of different features from the dataset that will have different effects on the dataset. These measured p-values can be used to decide whether to keep a feature or not.

3.4.1 Correlation algorithms

The selection of features is based on the link between correlation and high-correlation characteristics that are more linearly dependent and so have almost the same impact on the dependent variable. Therefore, when two characteristics have a strong correlation, than exclude one of features together with an assessment measure that ranks the various subsets of features. The relationship between features and getting rid of one of two features when the correlation is higher than 0.90 is shown in the correlation algorithm for feature selection (3.1).

Name : Correlation based Feature Selection

Input F: The Original list of feature {f0,f1,f2,…,fn} and the start
          position K.
          N: Size of All features in Dataset.

Output: Optimal Feature Subset

Begin

Initials: Feature Subset ←f0

Step 1: Best_Feature_Selection←FeatureSubSet   K ← 0

Step 2: While K <= N:

 If Correlation {f (K) , f(K+1)}

         FeatureSubSet ← FeatureSubSet ∪ {Next Feature}

 Else

   Best_Feature_Selection ← FeatureSubSet

 Endif   K =K+1

End while

       Return Best_Feature_Selection

Algorithm (3.1): correlation-based feature selection.

3.4.2 P-value algorithms

P-value asymptotic significance is a model's likelihood. A P-value shows the likelihood of an observation if a hypothesis is correct. This probability determines whether the hypothesis is true. The choice of different dataset characteristics removes different p-values. Using p-values removes distinct aspects and measures. These p-values help determine whether to preserve a feature if the correlation coefficient is statistically different from 0. A coefficient of 0 indicates that there is no linear relationship. If the p-value is less than the significance threshold, the correlation is different from 0. Dataset variables may have complicated and unknown connections. This information could help us prepare data for GRUs algorithms, which don't work as well when there are dependencies between them. This defines P-Value mathematically, see equation (3.1). P-Value is the chance of seeing E if null is true.

$$P\,Value = P(\,E\,|\text{H0}) \qquad (3.1)$$

The P value represent feature choice, and E represent the condition state. the It is essential to note that the p-value pertains not just to what we saw but also to more extreme findings. This is why the official definition of p-value states that it would be equal to or more extreme than its observed value. Correlation refers to the statistical link between two variables.

A positive correlation means that both variables move in the same direction. A negative correlation means that the value of one variable goes down as the value of the other variable goes up, as shown in the P-value method for feature algorithm (3.2).

After the implementation of the feature selection stage, To obtain some notices: Using correlation and p-value to choose features will have different effects on the dataset depending on which features are taken out. These measured p-values can

be used to decide whether to keep a feature or not. High-correlation features are more dependent on each other in a straight line, so they have almost the same effect on the dependent variable. Therefore, when two features have a high correlation, it can drop one of the two features along with an evaluation measure that scores the different feature subsets. The relationship between correlation and p-value

---

Name: P-value for feature selection

INPUT: N features form dataset

OUTPUT: Subset of more important feature

Step 1: Select a significance level

Step2: compute p value $P\,Value = P(\,E\,|\text{H0})$ for all feature

Step 3: Fit the model with feature [i]. # i represent the index feature

      If p-value is higher than significance level, remove the feature.

Step 4: if i <= N Repeat step 3 with feature [i]+1 to reduced features till only

the features having p-values <= significance level remains.

End if

Algorithm (3.2): P-value for feature selection

After completing feature selection stages and applying them to four of the databases, the eight features were obtained that have a great impact on the viewer's reaction. Five of these features are generated by a new feature, that is, Short Time Self Quality (STSQ), The rest of the features are Playback Indicator (PI) and Time elapsed since last rebuffering (TR).

3.5 Preprocessing Stages

There are many normalization techniques and in the dissertation the following were used:

➢ Min-Max normalization.

➢ Z-score normalization.

Min-Max normalization is applied to each of the three data sets. Another algorithm is applied due to the nature of the data used. as shown in the Min-Max Normalized Algorithm (3.3). Additionally, by use z-score normalization approaches to handle the data (LIVE-Netflix mobile VQA database). Persession and per-subject z-score normalization was used to account for variances in rating scale usage by each subject. Z-score normalization is the process of normalizing each value in a dataset such that the mean of all values is 0 and the standard deviation is (1). The greatest number indicates a continuous score of (1.9), while the smallest value represents a continuous score of (-2.6). z-score normalization has benefits. When comparing the raw results of different tests, z-score normalization takes into account both the average value and the range of raw values.

Name: Min-Max Normalized.

Input: Dataset $\{Yi\}_{i=1}^{t}$; the length of dataset k; and the significance level α.

Output: Normalized dataset $\{Y`i\}_{i=1}^{t}$.

Step 1: calculate Max, Min using VaR Max_{var}, Min_{var} = VaR($\{Yi\}_{i=1}^{t}$; k ; α)

Step 2: choose the highest value Max global = max(Max_{var}, max($\{Yi\}_{i=1}^{t}$))

Step 3: choose the lowest value Min global = min(Min_{var}, min($\{Yi\}_{i=1}^{t}$))

Step 4: normalizing dataset using standard Min-Max

$$Y`i = \left(\frac{Yi - \text{Min } var \ global}{\text{Max } var \ global - \text{Min } var \ global}\right)$$

Return: Normalized dataset $\{Y`i\}_{i=1:\ 1}^{t}$

Algorithm (3.3) : Min-Max Normalized.

3.6 Feature Generation for QoE Prediction

It aims to reduce the number of features in a dataset by generating features from the existing ones and then discarding the original features. This new reduced set of features should then be able to summarize most of the information contained in the original set of features. It increases the accuracy of learned models by extracting features from the input data. This stage of the general framework reduces the dimensionality of data by removing redundant data, via creating a new feature that still captures most of the useful information.

Video Quality Assessment (VQA) for STSQ, by using the STRRED (spatial-temporal reduced reference entropy differences) was employed to map features extracted from each video frame to the TVSQ, the STRRED as explain in chapter two. It has an inverse association with video quality, which means that a lower STRRED number suggests greater video quality and vice versa. The STRRED metric is applied, which in turn is implemented by the generation of the STSQ feature, which is one of the four features of the proposed system that are not

included in the dataset, but generated by using five features, including the dataset as shown in Table (3.7). This feature is selected from a dataset by correlation and p-value algorithm. It depends on the spatio-temporal reference of the video content, which has been shown by a number of studies to be a major factor in determining the quality of the video the user sees. As for the remaining three features (TR, IP, and number of buffers), which are determined directly from the database, where the selection algorithms were used from Table (3.5), Where the algorithm was applied to each of the four databases, it was proven that these features are the most influential among other features. Each database has been shown as in Table 3.6.

A. Short Time Subjective Quality (STSQ)

Using Time Varity Subject Quality (TVSQ) is to begin by predicting the short-time subjective quality (STSQ) of videos. TVSQ is a continuous-time record of viewers' judgments of the quality of the video as it is being played and viewed. The TVSQ depends on many elements of the video including spatial distortions and temporal artifacts. whereby the TVSQ of a video at a particular moment depends on the viewing experience before the moment. The quantitative nature of this dependency is critical for efficient rate adaptation. An STSQ predictor operates by extracting perceptually relevant spatial and temporal features from videos and then using these to form local video quality predictions. The basic premise is that STSQ of videos is a relatively stationary phenomenon. These so-called Video Quality Assessment (VQA) models do not capture long-term variations in STSQ nor do they predict human behavioral responses to these variations.

By partitioning every compressed version into one-second-long video chunks and predicting the Differential Mean Opinion Score (DMOS) of STSQ. DMOS scores range from 0 to 100 and a higher value indicates worse quality. To represent STSQ more naturally, so that higher numbers indicate better STSQ, By use a

Reversed DMOS (RDMOS) given by RDMOS = 100 − DMOS. Broadly, an RDMOS score of less than 30 on the LIVE database indicates bad quality, while scores higher than 70 indicate good quality. The QoE evaluator (STSQ) is employed by predicting its value through using per-frame distortions, which were predicted by spatial and temporal features extracted from the video in the following five features extraction, as shown in Table (3.6):

TABLE (3.6): Using Five Features to Generate The STSQ

Dynamic Input	Brief Description
Stall Length	The duration of the stall
Number of stalls	
Time since previous stall	Time that elapsed since a stall hasended the time the user had to recoverfrom an unpleasant stall experience
Frequency of stalls	Density of stall occurring in video
Rebuffering Rate	The fraction of stall time in a given video

In video QoE evaluation, the user immediately perceives perceptual aspects such as visual quality, rebuffering frequency, and rebuffering length. Typically, the user is sensitive to the quality of the current video segment, or short-time subjective quality (STSQ). STSQ is defined as the perceived quality of the user-rendered video segment. STSQ may be anticipated using any of the dependable video quality evaluation (VQA) indicators.

B. Playback Indicator (PI)

Since re-caching leads in a substantial fall in QoE, this feature has a substantial influence during the feature setup stages. By use the binary indicator

variable PI to signal if the video is actively playing or rebuffering. It is a continuous binary time variable that indicates the current state of playback: (1) for rebuffering and (0) for regular playback.

C. Number of Rebuffering Events (NR)

As the user's annoyance increases whenever a rebuffering event occurs, the number of rebuffering events (NR) happening from the beginning to the current time instant of the session is considered. Intuitively, a perceived video quality tends to decrease when distorted events occur, and gradually recovers since the end of those events.

D. Time Rebuffering (TR)

The fourth input which refers to the Time Elapsed Since Last Rebuffering since the last video impairment (TR) (i.e., bitrate switch or rebuffering occurrence) is utilized.

All considered perceptual factors are fed into an GRU-QoE model to predict the instantaneous QoE as in equation(3.4):

$$\hat{y}(t) = \text{GRU-QoE}^0(x(t),\ c\ (t-1)\ ) \tag{3.4}$$

where $\hat{y}(t)$ represents the predicted instantaneous QoE at the time instant t, x(t) is the input features, and c(t) is the update cells which encode the knowledge of the inputs that have been observed up to the time t, the update gate in a GRU decides which data to keep in the cell or to be clear. The update gate in a GRU decides which data to keep in the cell. In other words, it decides when to be close to (1) and when to be close to (0). GRU-QoE provides two functionalities: GRU-QoE 0 for output QoE prediction and GRU-QoE c for update cells, as in equation (3.5):

$$c(t) = \text{GRU-QoE}^c\ (\ c\ (0:t-1),\ \hat{y}(0:t-1)),\ \forall t \geq 1 \tag{3.5}$$

where c(0: t − 1) and y^(0 : t − 1), respectively, refer to the past cells and the past predicted QoE. The examples of four features (including STSQ, PI, NR, and TR) and the architecture of GRU-QoE model are illustrated in Figure (3.9). It can be observed that while STSQ tracks the time-variation in the quality, PI and TR are responsive to the rebuffering events.

However, that only a limited set of features are available as part of the LIVE Netflix QoE Database upon which evaluate the performance of the proposed model. Further, only a few videos of the database are made publicly available. This restricted us from the exploration and the investigation of further QoE features, as the database is not completely available.

3.7 Train and Test Stages

This process is implemented in two stages: learning and testing mode. The suggested system implementation modes across the complete dataset constituted a general design of the proposed system in the learning mode. For objective measurement. In proposing a GRUs-based strategy, it is recognized that, in the content of video streaming, calculating the perceptual quality of video frames often necessitates a large number of resources and the availability of reference material. The multi-layered, multi-unit GRU network in the model is a cascade of numerous GRU units stacked to generate GRU layers. When dealing with sequential input, GRU improves RNN, although it has more parameters and takes more training time.

The period of learning: The training and validation processes are included in these stages. The suggested model learns using all accessible training datasets; The set contains 80% of all movies in the database, with the remainder going to the testing stage. Supervised learning is achieved by giving the model the training examples with their appropriate labels. The goal is to utilize training data to progressively

enhance the model's predictions. One training step is one cycle of updating the weights and biases. The weights for each layer are updated throughout the training stages depending on the difference between the network model's actual output and the planned output, as illustrated in Figure (3.9). Weights are continuously updated until the network converges to the lowest error using the loss function.

The mean squared error (MSE) is selected because it is sensitive to outliers and gives numerous samples with the same input feature values. The best prediction is their mean target value. When employing means squared error with regression data, assume that objective is normally distributed and wants big mistakes to be substantially (quadratically) more punished than small ones.

The model is divided into two parts according to the number of features entered. In the first model, 4 features used in 3 databases are employed. The second model, however, uses only 2 features and is applied to the mobile database, after which the two models meet in the evaluation stage.

The testing stages: In these stages, the proposed trained model implements to predict the output on unseen (untrained samples) inputs and represents the application of the system, considering that every feature is in fact a single neuron with multiple instances.

Let x (t) and $\hat{y}(t)$ represent the actual and expected Quality of Experience (QoE) at a given time (t). Let $x(t) \in \geq 0$, The feature set that may take values from an m-dimensional space of positive real numbers is denoted by x(t). Here, the feature set x(t) reflects the elements that influence or determine QoE development. To predict the current QoE $\hat{y}(t)$ at any given time instant t, in this model By use a time-indexed feature vector $x(t) = [x_1(t), x_2(t), ..., x_m(t)]$. Given that the QoE $\hat{y}(t)$ is non-Markovian, as in formula (3.6):

$$p(y(t) \mid y(t-1), \; y(t-2), \cdots, \; y(1)) \tag{3.6}$$

where the conditional probability $p(y(t)|y(t-1), y(t-2),..., y(1))$ demonstrates that the QoE includes temporal dependencies of higher degree. The complexity of these relationships may prevent a single GRU from successfully capturing them. As seen in Figure 3.9 of the proposed multi-layer and multi-module GRU network for QoE prediction, a number of GRUs learn these dependencies involved in the QoE process. The proposed network consists of many GRU modules layered to build GRU-QoE layers.

Let us refer to the suggested GRU network as GRU-QOE (l,d), where (l) is the number of levels and (d) is the number of units in each layer. The design parameters (l) and (d) must be modified depending on the nature of the underlying process and the complexity of its relationships. Using input attributes $x(t)$, the GRU-QOE network continually computes the QoE estimate $y^{\wedge}(t)$ at each time instant (t) using the QoE estimate $y^{\wedge}(t)$.

The predicted QoE Y^(t)
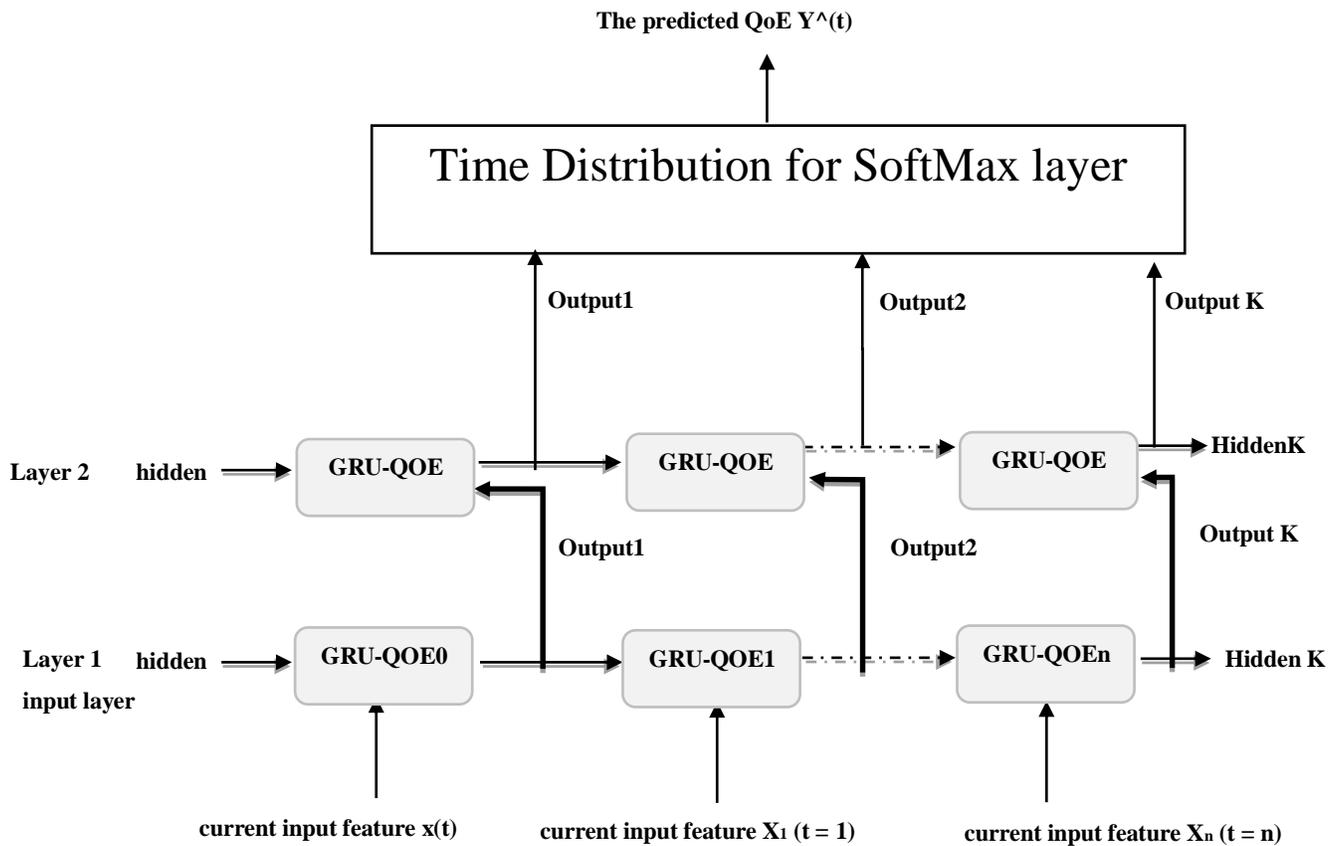
Figure (3.8): Proposed architecture of a bidirectional GRU-QOE network for QoE prediction.

Each GRU unit monitors the stochastic process by keeping an internal cell state, known as the latent state, and the state shifts are controlled by the input attributes $x(t)$. Let $c(t)$ denotes the network's set of GRU cell states as equation (3.8). The set of GRUs are modeled in order to understand the underlying complicated distribution regulating the state transition and estimate the QoE at every time instant.

GRU(l,d) provides two functionalities:

    1) GRU^o (l,d) for the output QoE prediction

    2) GRU^c (l,d) for the cell state update.

    The predicted QoE $\hat{y}(t)$ is given by

$$y^{\wedge}(t) = \text{GRU}^{o}{}_{l,d}(x(t), c(t-1)). \tag{3.7}$$

The cell state update for the GRU network is given by

$$c(t) = \text{GRU}^{c}{}_{l,d}(c(1:t-1), y^{\wedge}(1:t-1)), \forall t > 1. \tag{3.8}$$

The cell state $c(t)$ is a deterministic function of the previous QoE $y^{\wedge}(1: t_{l-1})$ and the prior cell states as represented by the $\text{GRU}c$ (l, d). This enables the state vector $c(t)$ to monitor complex interrelationships in the QoE process and GRUs to represent sequence data. The estimated QoE $y^{\wedge}(t)$ is computed using the current input feature x (t) and the cell state before the update $c(t_{l-1})$ according to formula (3.8). In addition, the non-linearities associated with the QoE prediction are taken into account, since GRUs are naturally nonlinear. The input feature selection step x (t) is essential for continuous QoE prediction, as shown by equations (3.7) and (3.8). The particular input characteristics must adequately capture and include the numerous forces that determine the development of QoE throughout the GRU states. The final layer is a dense layer with Softmax activation. The softmax function is used as the activation function in the output layer of neural net models that predict polynomial probabilistic models. As show the algorithm (3.4) of GRU-QoE for QoE result value.

Name: GRU-QoE for QoE result value

Input: number of GRU layers , number of GRU units , timesteps , number of epochs, batch size # batch size fixed to 1 as the GRU network is stateful

Dataset information , no feature, Normalizing score max and min

Output : Result of Overall QoE

Initial: path for dataset name for example LIVE Mobile Stall II

Step 1: Load training and validation sets

Step 2: Begin evaluation to create test video indexes

      all videos = range (0, no videos - 1)

      all videos = np.array (all videos)

  For test video no in range(0, no videos - 1)

      aa = all videos[0:test video no - 1]

      bb = all videos[test video no:,] # collect all video indexes except the test video index

  End if

Step 3: make permutate

  perm videos = np.random.permutation(Training video)

Step 4: define the training and validation sets to print('test video:(test video no))

Step 5: import the Training stage.

Step 6: reshape for timesteps

  training features2 = training features[0:len(training features) // timesteps * timesteps, :]

  training labels2 = training labels[0:len(training labels) // timesteps * timesteps, :]

  training features ts = np.reshape(training features2, (-1, timesteps, training features2.shape[1]))

  training labels ts = np.reshape(training labels2, (-1, timesteps, training labels2.shape[1]))

 Step 7: import the Validation stage

 Step 8: import the Testing stage.

 Step 9: save the predicted QoE result

 Step 10: plot the predicted QoE and the actual QoE

 Step 11: print the output Test Performance for LCC test mean, LCC test median, SROCC test mean, SROCC test median, RMSE test mean, RMSE test median, OR test mean and OR test median

Step 12: saving filename← GRU QoE , dataset name , GRU layers , GRU units, epochs , loss Function optimizer and saving filename LCC test, SROCC test, RMSE test and OR test

      Return Result of QoE value

algorithm (3.4): GRU-QoE for QoE result value.

3.8 The Model-Evaluation Stage

After training has been completed, it is time for assessment. Here comes into effect the dataset that placed aside before. The proposed model can be tested with information that has not been used to train it. (80/20, 70/30, or 50/50) is a reasonable rule of thumb for the ratio between training and assessment. The specifics of these three techniques may be found in the next chapter, and the 80/20 ratio yields effective outcomes. Reliable QoE models must be built in order to accurately measure the end-user quality of experience (QoE) for adaptive streaming services and then control and manage the network based on QoE, as shown in Figure (3.10).

On the LIVE Netflix dataset, STSQ is the highest performing video quality assessment (VQA) metric for the assessment (LCC), Yet on the LFOVIA QoE Database, the assessment (OR) is the highest performing VQA metric for STSQ. PSNR has the lowest score of all VQAs since it is not a perceptual VQA/QoE metric. Varying VQA metrics lead to varying QoE performance between the two databases. This may be due to the ability of VQA measurements to predict video

quality at different resolutions. When assessing video quality, video resolution is an important factor to evaluate. All of the STSQ measurements evaluated operate well at resolutions lower than high definition. Their VQA performance on movies with higher dimensions, such as 1920x1080 pixels and 3840 x 2160 pixels, is unknown (which is the situation in the databases under consideration). This shows how important it is to have VQA measures that can be relied on and always give good prediction performance at all resolutions.

For performance assessment of the LIVE QoE Database, By use the leave p-out cross validation (LpOCV) approach. It is mostly utilized in situations where forecasting is the objective and p = 5 is used to evaluate how well a predictive model would work in practice. Consequently, for the assessment of each test video,

training is conducted using ten movies that differ in video content and time-varying quality pattern from the test video.
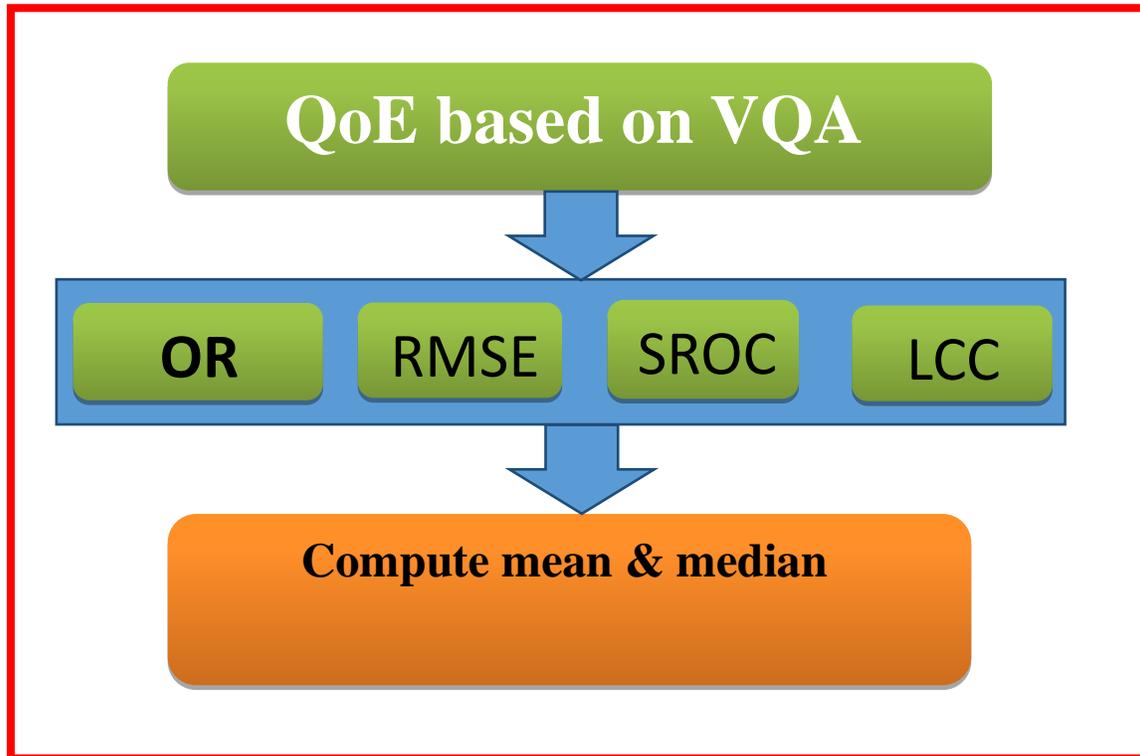
Figure (3.9): The Evaluation stage for the QoE model

Therefore, ten of the fifteen movies in the database are used for training in each training-test split. This database only contains movies with time-dependent quality artifacts and no rebuffering. As a result, By use STSQ as the only input feature for this database's suggested model. Alternatively, in our suggested model for this database, the feature Playback Indicator (PI) may be changed to (1) indicate ON and Time Rebuffering (TR) can remain constant throughout the length of movies. In this stage, the proposed system is assessed using the testing dataset. To assess the given models and the whole proposed system, the Linear Correlation Coefficient (LCC), Spearman Rank Order Correlation Coefficient (SROCC),

Normalized Root Mean Squared Error (RMSEn), and Outage Rate (OR) are used. Finally, Training and testing operations are performed for each and every video in every dataset. In addition, the training and testing methods are performed five times utilizing 5-folds.

The cross-validation process measures the performance of a machine learning algorithm on unseen data by testing machine learning models on a specified resampling data set using a single parameter termed (k). The limited sample is used to see how well the model is thought to be able to predict data that was not used to train the model.

The suggested models are given k =5 ; For each value of K, the dataset is divided into Training (80%) + Validation (10%) = 90% and Testing (10%), noting the testing results according to the measure chosen (adopted accuracy). Finally, the efficiency average is calculated to indicate the eventual result.

3.9. Evaluation of Training and Test Stages

After the network has stabilized, the validation procedure is carried out to verify the model (which is regarded a completed portion of the learning stage) using the available validation dataset with appropriate labels. The weights preserved from the training stages are used as a formula to demonstrate the performance and accuracy of the proposed model and to determine the model preparation to predict with fresh data.

The system used a different approach for selecting features. This method is performed on every video clip in the dataset and was applied every second in both training and testing steps. An explanation of the result of selecting video features for datasets will be presented. A large number of experienced quality assessment studies focus on using machine learning to assess the quality of a video experience without worrying about the quality of the video itself. The proposed system uses

several features and generates a new feature out of five features, which is a way to measure the quality of the experience. The STRRED metric is applied, which in turn is implemented by the generation of the STSQ feature, It is one of the four features of the proposed system. It depends on the spatio-temporal reference of the video content, which has been shown by a number of studies to be a major factor in determining the quality of the video the user sees.

As for the remaining four features, which are determined directly from the database, where the selection algorithms were used as mentioned in Table 3.5, where the algorithm was applied to each of four databases, it was proven that these features are the most influential among other features. Each database has been shown in Table 3.6. To determine the parameters of the GRU-QoE network, By investigated the effect of the number of layers and units in the proposed GRU-QoE network for QoE prediction.

At this point, the following action is taken: all the videos in the dataset will be captured and put into an array of an equal number of videos with their features; namely (STSQ, PI, NR and TR). During this process, normalization is calculated using its mathematical equation. Details can be referred to in Section (3.4). From the input feature vector and the output QoE vector

Next, a chart showing the expected quality curve is drawn and compared with the objective quality. Through the use of these features, the result is sought to be close to the user's expectations, which of course depend on the video measured by its type, content and the conditions in which it was captured.

Then the overall mean of all the previously mentioned metrics is measured. These metrics are the (Linear Correlation Coefficient (LCC), Spearman Order Correlation Coefficient (SROCC), Root Mean Square Error (RMSEn), and discontinuity rate (OR)) for the whole video because the quality is captured continuously at the end of the video, as mentioned earlier, using a non-Markov

concept in predicting the current state based on the memory for the previous time. As a result, this process takes place cyclically and continuously for the entire video and its playback, which contains in some examples illustrating the performance of the prediction model across 3 samples of a test video of the GRU-QoE model via the LIVE Mobile Stall II QoE dataset. In order to further evaluate after each video prediction, the overall quality will be measured for it, then plotted the so-called quality overall, then the mean and median are measured for each video.

To determine the parameter of the GRU-QoE network, the effect of the number of layers and units in the proposed GRU-QoE network for QoE prediction was investigated. The number of layers and GRU-QOE units are then altered to examine prediction performance using STSQ, PI, and TR features in the data sets. STRRED was used to be the STSQ in this investigation. First of all, GRU-QOE performance is checked and then the number of modules and layers in the GRU-QOE network is increased by investigating the functioning of the STSQ feature using a VQA scale called STRRED. During training, I override the time-varying video quality scale, so timestep is set = 4. During testing, $y(t)$ is predicted with an accuracy of about 1 second. Hence, during testing, predictions were made at each time step, for example, timestep = 1 at the end of the last layer, i.e. time-distributed. As show the algorithm (3.5) of training stage. dense layer, also called SoftMax layer.

Name: Training stage algorithm.

Input: Dataset information

　　　Number videos ← number of all video in dataset

　　　Number training videos ← number of training videos in dataset

output: Training dataset model

step 1: for i in range(no training videos):

　　add video features after feature selection

　　add video features ← array(video features)

　　training video number ← TrV[i]

　　Number buffers ← np.transpose(Nrebuffers all[training video no - 1, 0])

　　score continuous ← np.transpose(continuousQoE all[training video no - 1, 0])

step 2 : calculation the score continuous normalization ← (score continuous -

score　　continuous min) / (score continuous max - score continuous min)

　　video features ← Number buffers

step 3: *constitute the input feature vector and the output QoE vector*

　　if i equals to 0: training features ← video features

　　　training labels ← score continuous normalization

　　else:

　　　training features ← np.vstack((training features, video features))

　　　training labels ← np.vstack((training labels, score continuous norm))

function is used to stack the sequence of input arrays vertically to make a single

array

　　end if

Return Training dataset

algorithm (3.5): Training stage

3.10. Model Fitting

Model fitting is a metric that indicates how effectively a machine learning model generalizes to data that is comparable to the data on which it was trained. A well-fitted model generates more accurate results. An overfitted model fits the data too closely. A model that is under-fitted does not fit enough. Each algorithm for machine learning has a fundamental set of parameters that may be altered to increase its accuracy.

Name: Model fitting algorithm

Input: Define the shape of the input dataset (number GRU layers, number GRU units, number of epochs, batch size, timesteps, no feature, loss function, optimizer)

Output: test fit model to measure the accuracy of QoE results

Begin

Step 1: Define the dropout rate, which is used to prevent overfitting by Compare the predication features with testing features

Step 2: Specify the output layer to have a linear activation function.

End

Algorithm 3.6: Model fitting algorithm.

During the process of fitting, execute an algorithm on data that knows the target variable, often known as labeled data, and generate a machine learning model. The correctness of the results is then determined by comparing them to the actual values of the target variable. Next, utilize this knowledge to alter the algorithm's standard parameters to lower the degree of error, thereby improving the algorithm's ability to identify patterns and links between the remaining characteristics and the goal. Keep doing this until the algorithm finds the settings that give you the most accurate, useful, and relevant information for that own

practical challenge. The example fitting the core of machine learning is significant. If the proposed model does not properly match research data, the generated results will not be precise enough for practical decision-making. A well-fitted model contains hyperparameters that represent complicated interactions between known factors and target variables, enabling it to discover pertinent insights or make accurate forecasts. Fitting is an automated process that makes sure your machine learning models have the right parameters to solve specific problems in the real world, as shown in the model fitting algorithm (3.6).

Chapter Four

Experimental Results and Discussion

Chapter Four

Experimental Results and Discussion

4.1 Introduction

This chapter presents the implementation details and the discussions of results for the proposed GRU-QoE measurement model. The tested results are discussed and compared in this chapter through various figures and tables that present the proposed system performance and show the system behavior to achieve the objective of QoE assessment system. The evaluation testing process by using Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks architecture specifically designed to address the vanishing gradient problem. The key to the GRU solution to technical problems was the specific internal structure of units used in the model for the QoE assessment system. The effectiveness of the proposed system illustrated in the Chapter 3 is tested with different parameter values, and results of its implementation will be analyzed in this chapter. It is worth mentioning that the evaluation of the proposed system is implemented separately. Four sets of public data are applied as a case study to determine the behavior of this model. The following sections describe the system requirements, and present results which obtained from each stage of the proposed system.

4.2 System Requirement

First of all, to perform machine learning and deep learning on any dataset, the software/program requires a computer system powerful enough to handle the necessary computing power. The proposed system is implemented using the following:

The experimental tests are conducted on different Dataset types, varying in their size, resolution, display device type, and content type (news, movies, cartoons,...). The proposed QoE assessment system has been built using Python 3.6 version 2022 programming language and its Toolbox, under the Windows 10 operating system. The model is running on a Dell personal computer (Intel (R) Core (TM) i7 CPU @ 2.60 GHz, RAM of 16GB, and a 64-bit operating system). The next section presents an evaluation of the results obtained from the proposed system.

4.3. The Proposed QoE Measurement Implementation

In this section, illustrations of system performance results in experience quality assessments are presented. The proposed system is a No-Referenced (NR) video stream QoE measurement. QoE factors and their effects on video quality and QoE metrics were discussed in the previous chapter. At the same time, video content features and their effects on video quality QoE, which call (GRU-QoE), and the use of extracted features to create new ones are also discussed. The following sections describe all the results obtained by the proposed system, supported by figures that explain the results and details of the work step by step.

4.3.1. Evaluation of Feature Selection Stages

As mentioned earlier in Section 3.5, the system used a different approach for selecting features. This method was performed on every video clip in the data set and was applied every second in both the training and testing steps. An explanation of the result of selecting video features for data sets will be presented in chapter 3. A large number of quality assessment studies focus on using machine learning to

assess the quality of a video experience without worrying about the quality of the video itself.

4.3.2. Overall Evaluation of GRU-QoE

As a metric to verify the objective score obtained by running the model, the relationship between the subjective QoE and objective QoE of the videos is presented in the previous sections, and the relationship between them is evident. The proposed system provides statistical methods to measure the correlation between its results and those of other systems. The GRU-QoE system uses four statistical terms: SRCC, LCC, RMSE, and OR. The reader can find the details of each measure in chapter 2, section 2.13. The proposed GRU-QoE network is trained for the evaluation of each dataset using Keras. In all our reviews, By use the GRU network configuration. By use all four features (STSQ, PI, TR, and Number buffer) to predict QoE.

4.4 Comparison of Results

In this step, the quality of outcome predictions for each video clip in the dataset of the four databases will be evaluated. The mean and median for each of the four databases used in this assessment (LCC, SROCC, RMSEn, and OR) are calculated according to the assessment of each stage at the end of the learning and training stages processing as mentioned in Chapter 3. Our approach is evaluated using a recently released trial quality dataset that contains continuous subjective scores for time.

The average QoS prediction performance results for each video in the four databases are demonstrated in Table 4.1. that our model is superior to the time

series modeling approaches such as the Autoregressive Nonlinear External Model (NARX). It is distinguished from other nonlinear autoregressive models by the presence of external inputs. This indicates that the model establishes a correlation between the current value of a time series and the historical values of the same series as well as other values. [69] in the LIVE Netflix Database [106] is shown in Table 4.2. Similarly, the support vector regression comparison (SVR-QoE) in the LFOVIA QoE database [107] gives good results, as shown in Table 4.3. HW [108] in the LIVE QoE II database is a QoE model suggested in their databases. HW's approach has the potential to reliably estimate TVSQ for modified videos. This method is suitable for online TVSQ prediction over HTTP-based broadcasting since the Hammerstein-Wiener model has a fairly easy-to-order organizational structure. From Table 4.4, also note that GRU-QoE is superior to NLSS-QoE (a space model). Non-linearity can predict the quality of user experience on a continuous-time basis in an accurate and efficient manner. It can also be seen that GRU-QoE outperforms current QoE models in all performance measures. The best performance results are shown in red color.

Table (4.1): QoE prediction performance of the GRU-QoE model over the LIVE QoE Database.

QoE Model	VQA	LCC	SROCC	RMSEn(%)	OR(%)
GRU-QoE	STRRED[109]	**0.8927**	**0.8726**	**15.14**	45.3546
NLSS-QoE [111]	STRRED[109]	0.723	0.707	7.04	26.22
	MS-SSIM [110]	0.883	0.871	4.58	**11.36**
	NIQE [113]	0.211	0.189	9.23	43.47
HW [112]	STRRED [109]	0.742	0.732	7.40	32.02
	MS-SSIM [110]	0.727	0.705	6.70	29.11
	NIQE [113]	0.511	0.509	8.34	36.02

Table (4.2): The QoE prediction performance of the GRU-QoE model over the LIVE Netflix Database.

QoE Model	VQA	LCC	SROCC	RMSEn(%)	OR(%)
GRU-QoE	STRRED[109]	**0.875**	**0.863**	**15.5**	46.009
	STRRED[109]	0.655	0.483	16.09	69.16
NLSS-QoE [111]	MS-SSIM [110]	0.583	0.420	18.22	73.74
	NIQE [113]	0.527	0.300	14.50	53.33
	STRRED [109]	0.621	0.557	8.52	**23.84**
NARX [115]	MS-SSIM [110]	0.598	0.549	10.27	25.95
	NIQE [113]	0.605	0.537	9.82	30.66

Table (4.3): The QoE prediction performance of the GRU-QoE model over the LFOVIA QoE Database.

QoE Model	VQA	LCC	SROCC	RMSEn(%)	OR(%)
GRU-QoE	STRRED[109]	**0.8902**	**0.8745**	**14.74**	43.9012
	STRRED[109]	0.767	0.685	7.59	8.47
NLSS-QoE [111]	MS-SSIM [110]	0.781	0.680	7.37	6.78
	NIQE [113]	0.825	0.794	6.97	6.51
	STRRED [109]	0.686	0.648	10.44	22.87
SVR-QoE [114]	MS-SSIM [110]	0.737	0.683	9.48	18.25
	NIQE [113]	0.797	0.750	8.32	13.64

Table (4.4): The QoE prediction performance of the proposed model over the LIVE Mobile Video Stall Database-II.

QoE Model	LCC	SROCC	RMSEn(%)	OR(%)
GRU-QoE	**0.8700**	**0.8552**	**14.72**	44.6793
NLSS-QoE [111]	**0.680**	**0.590**	**9.52**	**42.40**

While STRRED as the STSQ measure perform best on the LIVE Netflix [106] and the LIVE QoE datasets, it must be noted that the LFOVIA QoE Dataset consists of videos at FHD and UHD resolutions. Although the VQA performance of STRRED has been demonstrated over the resolution 768×432, its performance at higher resolutions such as FHD and UHD is not well studied. Hence, It has been hypothesized that the difference in QoE performance across the different STSQ scales is due to the dependence of the STSQ metrics on video resolution. This indicates that there is scope for more efficient VQA algorithms that consistently perform well across levels of accuracy. However, from Tables 4.1 to 4.4, it can be inferred that STRRED can serve as a good metric for measuring STSQ.

Furthermore, the proposed model provides the flexibility to choose the appropriate VQA for QoE prediction. The results also demonstrate the effectiveness of the chosen features for QoE prediction. Besides, we would like to highlight that the best performing STSQ metrics observed in the proposed model concur with those of the QoE models reported over the respective datasets.

It was also noted that during testing, the QoE calculation using the proposed GRU network is performed in an automated manner, similar to that performed in recent QoE models such as NARX [115]. The QoE computational complexity in

these models is determined by the computational complexity of STSQ, which is in turn determined by the VQA method employed for computing the STSQ. However, for applications such as on demand video streaming, the STSQs could be computed offline and could be readily made available in order to facilitate QoE computation in real time.

4.5 The experimental results

The experimental results were validated and the effectiveness of the proposed system in assessing the QoE value was demonstrated on four datasets. Experimental quality prediction performance for GRUs and different on four QoE datasets in terms of LCC, SROCC, and RMSEn measures but a very poor score for the OR scale, in Figures 4.1, 4.2, 4.3, 4.4 and 4.5 illustrate the performance of the prediction model over sample videos from GRU-QoE model.

Also, for further configuration of the GRU input, the input data must have three dimensions. (batch size, number of eras, and time steps). Batch size is the number of samples sent to the model at a time. In this work, By using the batch size of 1, but you can take (4, 8, 16, 32, 64,...) depending on memory. But using a large value of batch size will affect the learning and training time significantly. Therefore, the optimal number of periods to train for most data sets depends on the situation. Monitor loss values without using the early reconnection function: train the model up to (50) epochs and plot the training loss and validation loss values against the number of periods.

The time step is one iteration of the cell. For example, in the first step, the system produces output 1. In the second step, it produces output 2, and so on. So the optimal number of training and testing time steps for most data sets is 4.

Adam is the common optimizer. Use Adam to train the neural network in less time and with more efficiency. For sparse data, make use of the optimizer with a dynamic learning rate.

Through the use of the proposed method QoE-GRU and by comparing it with the LSTM, it has been found that video with a long duration has encountered evaluation problems and has been treated using the proposed method.
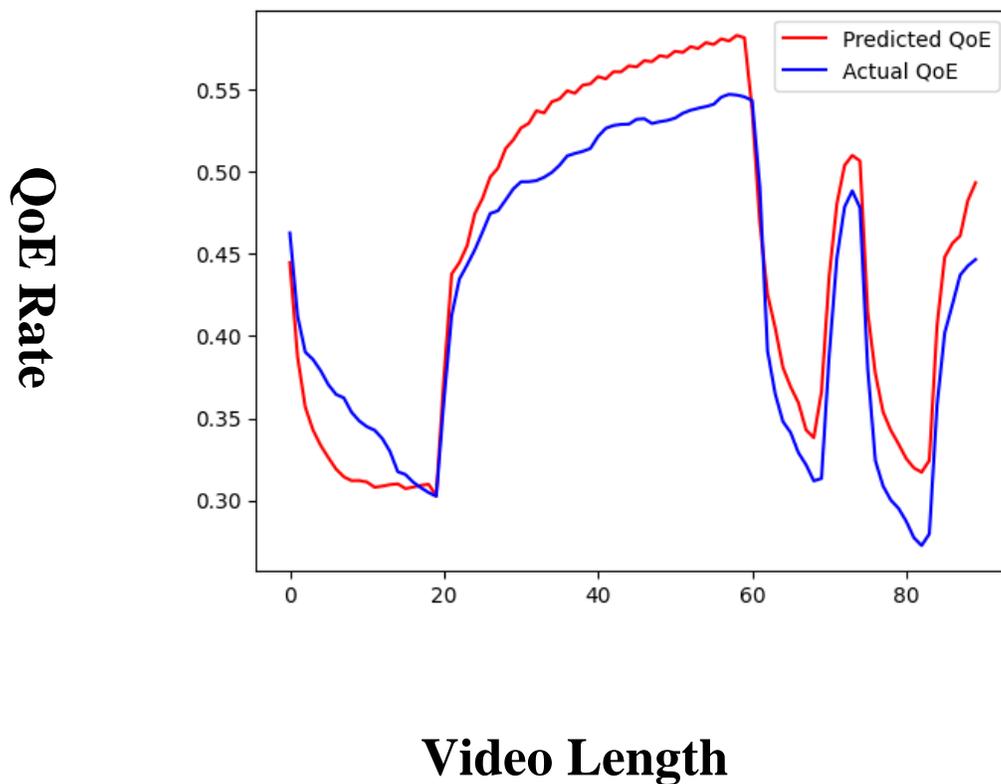

Video Length

Figure (4.1): The graph illustrates the performance of the prediction model over sample videos from the GRU-QoE model over the LIVE QoE Database II.
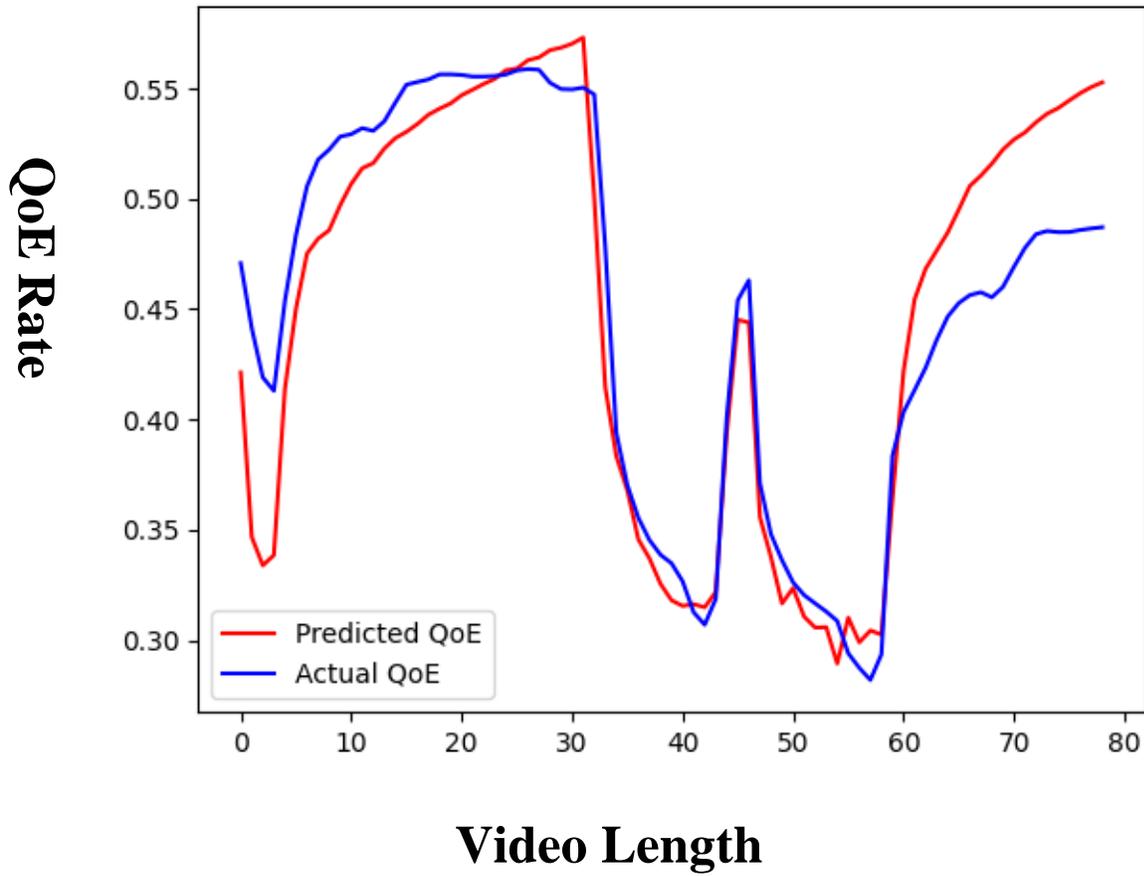
Figure (4.2): The graph illustrates the performance of the prediction model over sample videos from the GRU-QoE model over the LFOVIA QoE dataset.
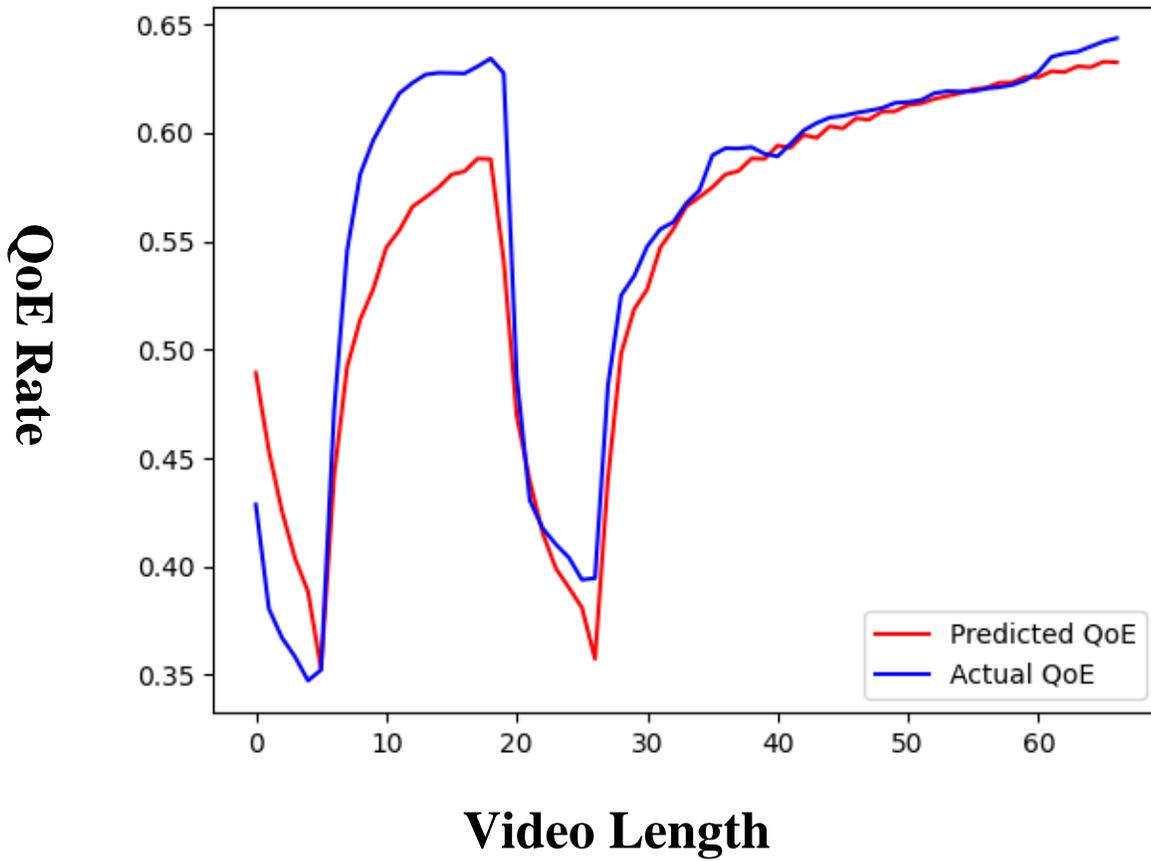
.

Figure (4.3): The graph illustrates the performance of the prediction model over sample videos from the GRU-QoE model over the LIVE-Netflix mobile VQA dataset.
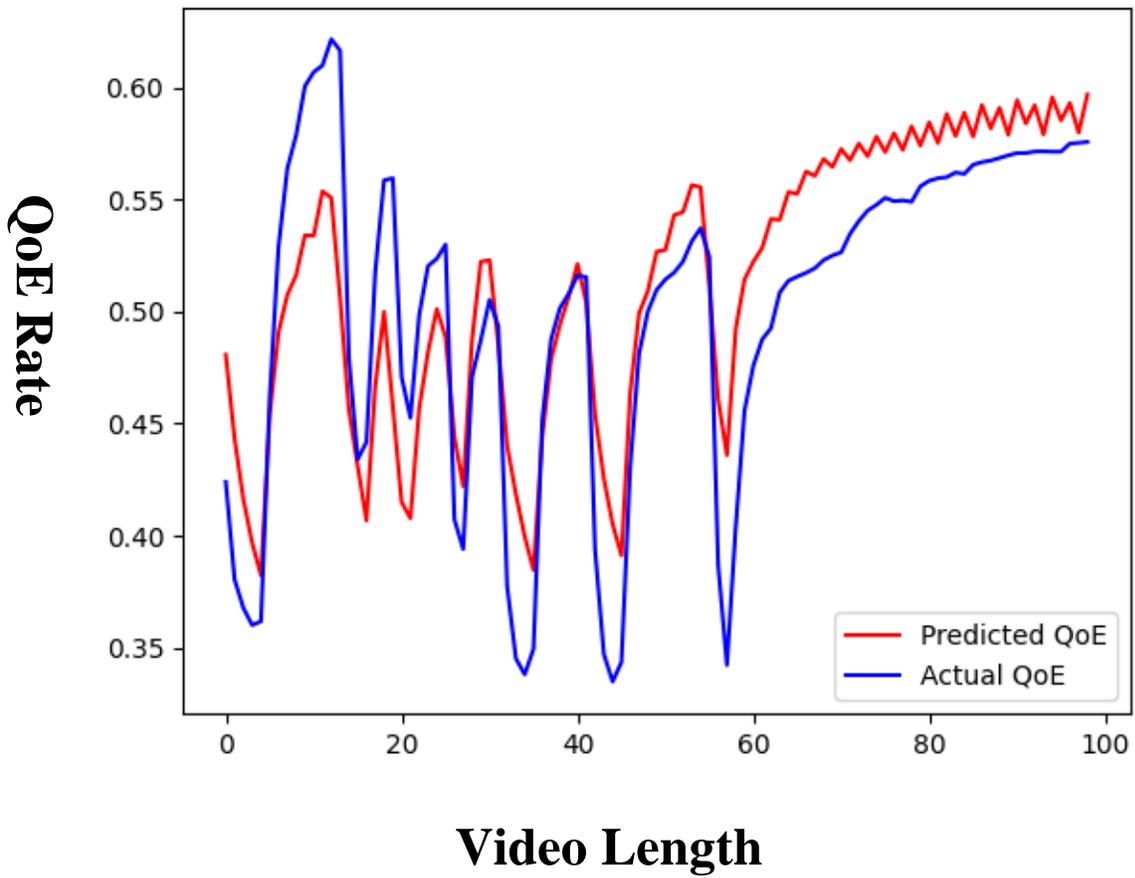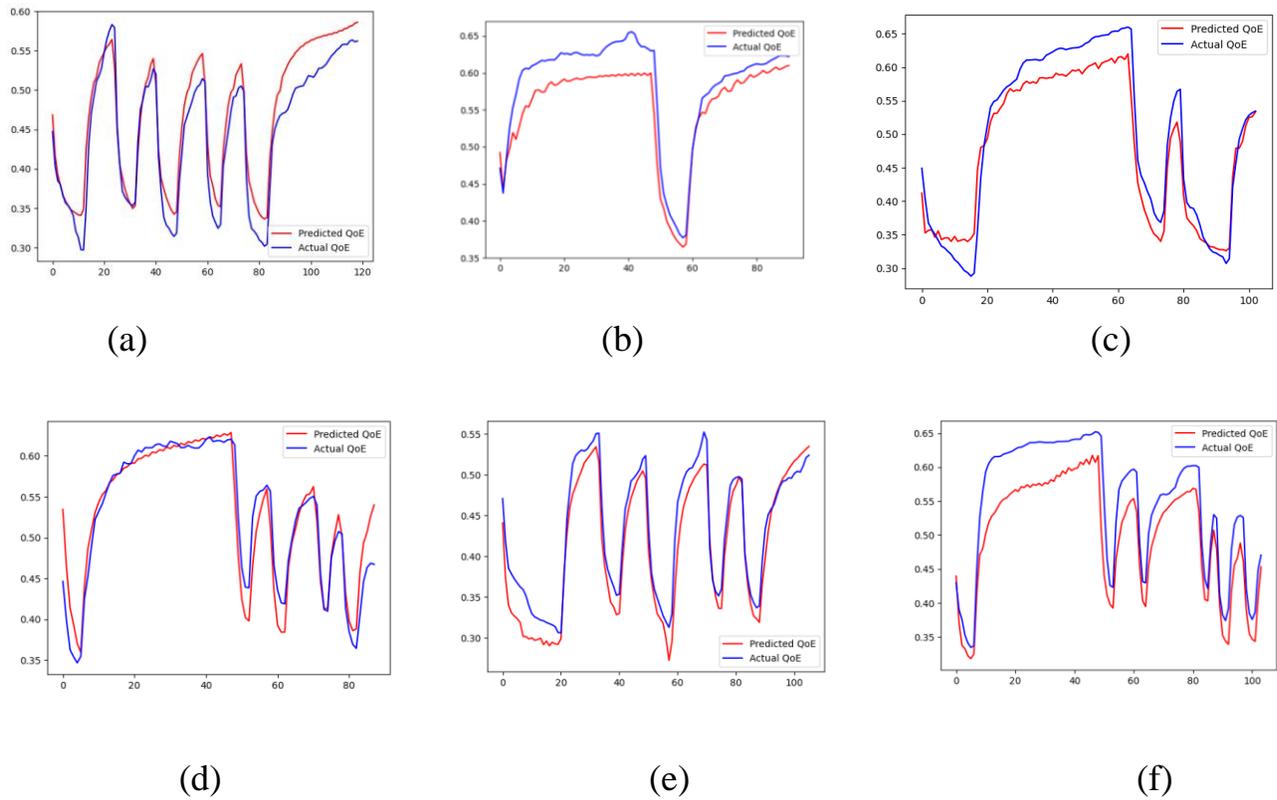
Figure (4.4): The graph illustrates the performance of the prediction model over sample videos from the GRU-QoE model over the LIVE_Mobile Stall II QoE Dataset.

(a) LIVE Netflix by LCC. (b) LIVE Netflix by OR. (c) LFOVIA QoE by LCC.

(d) LFOVIA QoE by OR.

Figure (4.5): QoE prediction performance of various GRU-QoE network.

Figure. a,b,c and d illustrate the QoE prediction performance for different GRU-QoE upon the LIVE Netflix and the LFOVIA QoE Databases in terms of LCC and OR measures.

      The red curve outline represents the expected quality value, The blue color is the objective quality value, which, as mentioned earlier, is intended to approximate the curve and the target behavior of the viewer.

4.6 K-fold Cross-Validation (CV)

K-fold cross-validation is utilized in order to further enhance the network against overfitting. By doing K iterations of training-validation-testing on independent, non-overlapping, equally-proportioned Training (Tr = 80 %), Validation (Va = 10 %), and Testing (Te = 10 %) sets, this technique employs the simple technique. The 10-fold CV implemented with proposed GRUs-QoE models and experimental results can be represented through Figure 4.5 for GRUs-QoE model.
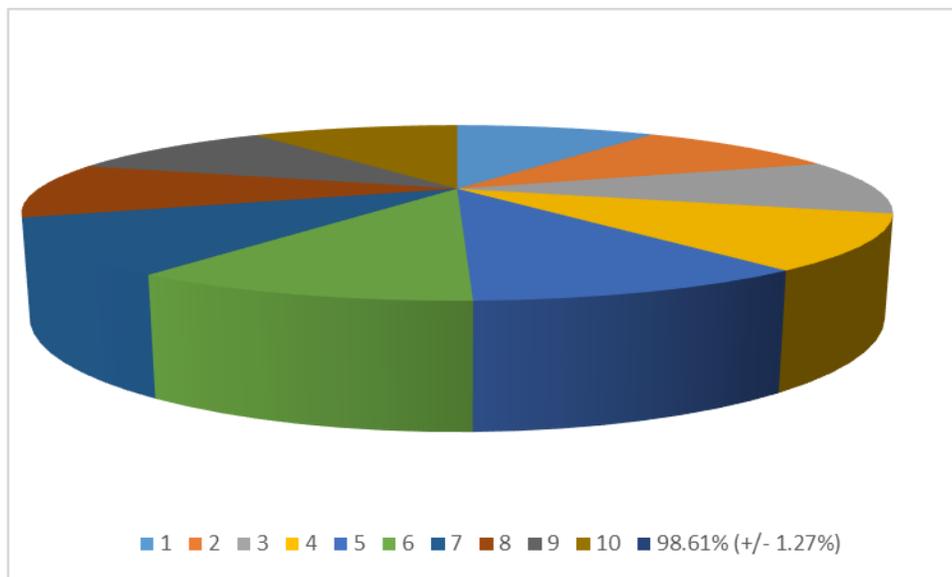

Figure (4.5): 10-Fold CV for GRUs-QoE models

4.7 Discussion results

There are a number of results that reached through the phases of work, and they will be mentioned in the following form:

1- A comprehensive evaluation of the proposed model is performed on four publicly available continuous QoE databases, and it was found that GRU-QoE provides good prediction performance across four databases. The reason for diversity is to ensure that the system operates in different environments, networks, adaptive bit-rate algorithms, and resolutions.

2- A number of layers and a number of GRU-QOE units are tested to examine prediction performance using STSQ, PI, NR, and TR features in the datasets.

3- A trial-and-error technique is used to determine the number of units and layers, such as (4,8,11,20,23,25) for units and using (2,3,4,5,6,7) for layers. By notice the number of calibration and adjustment processes, The use of 23 units gives the best result, with the use of two hidden layers. When the number of layers was increased, there was no appreciable change in the quality of the evaluated result, and when reaching seven layers, with the increase in the training and learning time, the result become worse for the value of the expected quality of each processed video.

4- It is noted that there is a significant improvement in prediction performance in terms of LCC and RMSEn with the addition of GRUs and layers to the network. For network configurations with two layers or higher, there is a steady increase in performance up to 10 GRUs, after which LCC performance begins to saturate. Prediction performance in terms of OR continues to show improvement after 10 units, and improvement begins to saturate after 19 GRU.

5- In this work, there is a batch size of 1, but it can be 4, 8, 16, 32, 64, etc. depending on memory. Therefore, the optimal number of periods for training depends on the data set.

6- The time step is one iteration of the cell. Therefore, the optimal number of training and testing time steps for most data sets is 4.

12- To train the neural network in less time and with more efficiency using Adam is the common optimizer. use the optimizer with a dynamic learning rate in the body of code

4.8 Work Limitation

One of the most important limitations of the work are having a comprehensive database containing all the events such as freeze, sleep, rejection, within different data transmission networks for the same video content, such as mobile networks (5G cellular phone), Fiberoptic and Wi-Fi.

In addition to adding audio evaluation as additional features to the video content that must be dealt with within different technologies

Chapter Five

Conclusions and Future Work

<p align="center">**Chapter Five**</p>

<p align="center">**Conclusions and Future Works**</p>

5.1 Conclusions

Quality of Experience for video streaming services is an important factor in understanding the user's perception of the video services. This dissertation has mainly concentrated on present solutions for the measurement and improvement of QoE for video streaming services.

As a solution to measure the QoE of video streaming services proposed a method to assess video QoE that depends on selecting and generating different types of features and using machine learning techniques to address assessment problems in the QoE system. Experimental tests have been applied in different sizes, categories and types of videos. Different features are selected from the videos and these features are used in the training and assessment process. This chapter enumerates the main findings of the dissertation as follows:

1- The system deals with different types of features. The QoE metric depends on the video streaming parameters that have a great effect on the human opinion that depends on video content.

2- A overall QoE prediction performance analysis shows that the mean and the median continuous QoE pooling strategies are effective for quantifying the overall QoE of the users.

3- When applying the LSTM model, it has been noticed a problem evaluating videos whose length was more than 200 ms, where the expected values are reduced or vanished value, so using our proposed method has solved the

<p align="center">108</p>

problem and demonstrated superiority due to the different architecture of the GRU-QoE.

4- In this dissertation, different types of video contents are considered. They broadly covered video sequences from slow-moving like animal life environment to fast-moving like sports type.

5- The effect feature's the video QoE is robust (by using STRRED) to generate a new feature called STQS which is more accurate than each feature alone.

6- Most QoE systems depend on just QoS features without looking at the video quality content and this matter is unacceptable because the video may have a bad quality that looks like the network effect and vice versa.

7- By evaluating the data set in order to draw different observations. It has been noticed that the subject's experience was negatively affected by the frequency of stilling, regardless of their content. Even when the displayed bitrate is low, users prefer a temporary drop in bitrate over re-caching of low-complexity information. To prevent buffering, users cannot tolerate a persistently low bitrate.

8- Although adding more than 2 GRU layers makes little difference, notice that performance starts declining after 5 GRU layers.

9- Based on the careful examination of the performance of LCC and OR, it was determined that the two-layer and 23-unit GRU configuration is the optimal choice of the GRU-QoE network for Overall result prediction.

5.2 Future Works

This dissertation has made to predict of QoE, the assessment depending on video content parameters. It is believed that the proposed system can be development in the following directions:

1. The focus on using the video QoE model will be in the area of user-behavior prediction in order to prefetch information for the video streaming, taking into account new parameters such as voice searchability and scalability.

2. Developing the machine learning algorithm by using Content-Adaptive Recurrent Unit that the design of CARU contains all the features of GRU but requires fewer training parameters in order to predict a common output for speed up the predication in less complexity.

References

[1] Khokhar, Muhammad Jawad, Thibaut Ehlinger, and Chadi Barakat. "From network traffic measurements to QoE for internet video." 2019 IFIP Networking Conference (IFIP Networking). IEEE, 2019.

[2] Seufert, Michael, et al. "A survey on quality of experience of HTTP adaptive streaming." IEEE Communications Surveys & Tutorials 17.1 (2014): 469-492.

[3] Juluri, Parikshit, Venkatesh Tamarapalli, and Deep Medhi. "Measurement of quality of experience of video-on-demand services: A survey." IEEE Communications Surveys & Tutorials 18.1 (2015): 401-418.

[4] Talwalkar, Rohit A., and Mohammad Ilyas. "Analysis of quality of service (QoS) in WiMAX networks." 2008 16th IEEE International Conference on Networks. IEEE, 2008.

[5] Ameen, M. A., Ahsanun Nessa, and Kyung Sup Kwak. "QoS issues with focus on wireless body area networks." 2008 Third International Conference on Convergence and Hybrid.

[6] Shin, Dong-Hee. "Conceptualizing and measuring quality of experience of the internet of things: Exploring how quality is perceived by users." Information & Management 54.8 (2017): 998-1011

[7] García, Boni, et al. "Understanding and estimating quality of experience in WebRTC applications." Computing 101.11 (2019): 1585-1607.

[8] Chang, Danni, et al. "A user-centric smart product-service system development approach: A case study on medication management for the elderly." Advanced Engineering Informatics 42 (2019): 100979.

[9] Song, Wei, and Dian W. Tjondronegoro. "Acceptability-based QoE models for mobile video." IEEE Transactions on Multimedia 16.3 (2014): 738-750.

[10] Hoßfeld, Tobias, et al. "From QoS distributions to QoE distributions: A system's perspective." 2020 6th IEEE Conference on Network Softwarization (NetSoft). IEEE, 2020.

[11] Olsson, Sofie, Mario Stroppiana, and Jamal Baina. "Objective methods for assessment of video quality: state of the art." IEEE transactions on broadcasting 43.4 (1997): 487-495.

[12] Chikkerur, Shyamprasad, et al. "Objective video quality assessment methods: A classification, review, and performance comparison." IEEE transactions on broadcasting 57.2 (2011): 165-182.

[13] Reichl, Peter, et al. "Towards a comprehensive framework for QoE and user behavior modelling." 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX). IEEE, 2015.

[14] Moore, Alastair H., P. Peso Parada, and Patrick A. Naylor. "Speech enhancement for robust automatic speech recognition: Evaluation using a baseline system and instrumental measures." Computer Speech & Language 46 (2017): 574-584.

[15] Hou, Jen-Cheng, et al. "Audio-visual speech enhancement using multimodal deep convolutional neural networks." IEEE Transactions on Emerging Topics in Computational Intelligence 2.2 (2018): 117-128.

[16] Serral-Gracià, René, et al. "An overview of quality of experience measurement challenges for video applications in IP networks." International Conference on Wired/Wireless Internet Communications. Springer, Berlin, Heidelberg, 2010.

[17] Rifai, Hala et al. "A brief synthesis of QoS-QoE methodologies." 2011 10th International Symposium on Programming and Systems (2011): 32-38.

[18] Barakabitze, Alcardo Alex, et al. "QoE management of multimedia streaming services in future networks: a tutorial and survey." IEEE Communications Surveys & Tutorials 22.1 (2019): 526-565.

[19] Dimopoulos, G., Leontiadis, I., Barlet-Ros, P. and Papagiannaki, K., 2016, November. Measuring video QoE from encrypted traffic. In Proceedings of the 2016 Internet Measurement Conference (pp. 513-526).

[20] Tsolkas, D., Liotou, E., Passas, N., & Merakos, L. (2017). A survey on parametric QoE estimation for popular services. Journal of network and computer applications, 77, 1-17.

[21] D. Zhengfang, Z. Kai, M. Kede, Abdul Rehman, and W. Zhou, "A Quality-of-Experience Index for Streaming Video", IEEE Journal of Selected Topics in Signal Processing, Vol. 11, No. 1, 2017.

[22] E. Demirbilek, and J.Ch. Gregoire, "Perceived Audiovisual Quality Modelling based on Decision Trees, Genetic Programming and Neural Networks, "Journal ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol. 133, No. 2, USA, 2017.

[23] O.F.M. Shallal, "A Proposed Technique for Measuring Video Quality of Experience", Msc. Thesis, Iraqi Commission for Computers and Informatics, Informatics Institute for Postgraduate Studies, 2018.

[24] Khokhar, Muhammad Jawad. Modeling quality of experience of internet video streaming by controlled experimentation and machine learning. Diss. COMUE Université Côte d'Azur (2015-2019), 2019.

[25] Eswara, Nagabhushan, et al. "Streaming video QoE modeling and prediction: A long short-term memory approach." IEEE Transactions on Circuits and Systems for Video Technology 30.3 (2019): 661-673.

[26] Belmoukadam, Othmane, Muhammad Jawad Khokhar, and Chadi Barakat. "On excess bandwidth usage of video streaming: when video resolution mismatches browser viewport." 2020 11th International Conference on Network of the Future (NoF). IEEE, 2020.

[27] Song, Chunhe, et al. "QoE-driven edge caching in vehicle networks based on deep reinforcement learning." IEEE Transactions on Vehicular Technology 70.6 (2021): 5286-5295.

[28] Burhanuddin, Liyana, et al. "QoE Optimization for Live Video Streaming in UAV-to-UAV Communications via Deep Reinforcement Learning." IEEE Transactions on Vehicular Technology (2022).

[29] Ghosh, Monalisa, Dr Chetna Singhal, and Rushikesh Wayal. "DeSVQ: Deep Learning Based Streaming Video QoE Estimation." 23rd International Conference on Distributed Computing and Networking. 2022.

[30] Alreshoodi, Mohammed, and John Woods. "Survey on QoE\QoS correlation models for multimedia services." arXiv preprint arXiv:1306.0221 (2013).

[31] Estrela, Vania V., et al. "Transmedia Ecosystems, Quality of Experience and Quality of Service in Fog Computing for Comfortable Learning." 2021 IEEE Global Engineering Education Conference (EDUCON). IEEE, 2021.

[32]Brunnström, K., Beker, S. A., De Moor, K., Dooms, A., Egger, S., Garcia, M. N., ... & Zgank, A. (2013). Qualinet white paper on definitions of quality of experience.

[33] Brooks, Peter, and Bjørn Hestnes. "User measures of quality of experience: why being objective and quantitative is important." IEEE network 24.2 (2010): 8-13.

[34] Pearson, Sammy W., and James E. Bailey. "Measurement of computer user satisfaction." ACM SIGMETRICS Performance Evaluation Review 9.1 (1980): 59-68.

[35] Varela, Martín, Lea Skorin-Kapov, and Touradj Ebrahimi. "Quality of service versus quality of experience." Quality of experience. Springer, Cham, 2014. 85-96.

[36] AVSystem. "Quality of Experience (QoE) in Telecommunications Services." Quality of Experience (QoE) in Telecommunication Services, AVSystem, 27 Mar. 2020, https://www.avsystem.com/blog/qoe-quality-of-experience-telecommunication/.

[37] Ries, Michal, Peter Froehlich, and Raimund Schatz. "QoE evaluation of high-definition IPTV services." Proceedings of 21st International Conference Radioelektronika 2011. IEEE, 2011.

[38] Ahmad, Toufeeq, and Shahryar Shafique Quershi. "The full reference quality assessment metrics for super resolution of an image: Shedding light or casting shadows." 2010 International Conference on Electronics and Information Engineering. Vol. 2. IEEE, 2010.

[39] Chikkerur, Shyamprasad, et al. "Objective video quality assessment methods: A classification, review, and performance comparison." IEEE transactions on broadcasting 57.2 (2011): 165-182.

[39] Hulusic, Vedad, et al. "Acoustic rendering and auditory–visual cross-modal perception and interaction." Computer Graphics Forum. Vol. 31. No. 1. Oxford, UK: Blackwell Publishing Ltd, 2012.

[40] You, J., Reiter, U., Hannuksela, M. M., Gabbouj, M., & Perkis, A. (2010). Perceptual-based quality assessment for audio–visual services: A survey. Signal Processing: Image Communication, 25(7), 482-501.

[41] Soldani, David, et al. "QoE and QoS Monitoring." QoS and QoE Management in UMTS Cellular Systems (2006): 315-384.

[42] Alreshoodi, Mohammed AM. Prediction of Quality of Experience for Video Streaming Using Raw QoS Parameters. Diss. University of Essex, 2016.

[43] Kozamernik, Franc. "Media Streaming over the Internet–an overview of delivery technologies." EBU Technical Review 10 (2002): 39.

[44] Daoust, François. "W3C Workshop on Web and Machine Learning." Media Processing Hooks for the Web, MEDIA PROCESSING HOOKS FOR THE WEB, 2022, https://www.w3.org/2020/Talks/fd-media-processing/

[45] Rifai, Hala, Samer Mohammed, and Abdelhamid Mellouk. "A brief synthesis of QoS-QoE methodologies." 2011 10th International Symposium on Programming and Systems. IEEE, 2011.

[46] Zheng, Haitao, and Jill Boyce. "An improved UDP protocol for video transmission over internet-to-wireless networks." IEEE Transactions on Multimedia 3.3 (2001): 356-365.

[47] Schulzrinne, Henning, et al. "RFC3550: RTP: A transport protocol for real-time applications." (2003).

[48] Apostolopoulos, John G., Wai-tian Tan, and Susie J. Wee. "Video streaming: Concepts, algorithms, and systems." HP Laboratories, report HPL-2002-260 (2002): 2641-8770.

[49] Liu, Chenghao, Imed Bouazizi, and Moncef Gabbouj. "Rate adaptation for adaptive HTTP streaming." Proceedings of the second annual ACM conference on Multimedia systems. 2011.

[50] De Cicco, Luca, Saverio Mascolo, and Vittorio Palmisano. "Skype video congestion control: An experimental investigation." Computer Networks 55.3 (2011): 558-571.

[51] Herrero, Rolando. "Network and Transport Layers." Fundamentals of IoT Communication Technologies. Springer, Cham, 2022. 77-109.

[52] Matijasevic, Maja. Data Traffic Monitoring and Analysis: From Measurement, Classification, and Anomaly Detection to Quality of Experience. Springer Berlin Heidelberg, 2013.

[53] Wu, H. R., K. R. Rao, and Ashraf A. Kassim. "Digital Video Image Quality and Perceptual Coding" Boca Raton." Journal of Electronic Imaging- 16 (2005).

[54] Finamore, A., Mellia, M., Munafo, M. M., Torres, R., & Rao, S. G. (2011, November). Youtube everywhere: Impact of device and infrastructure synergies on user experience. In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (pp. 345-360).54

[55] Mondal, Abhijit, et al. "Candid with youtube: Adaptive streaming behavior and implications on data consumption." Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video. 2017.

[56] Sodagar, Iraj. "The mpeg-dash standard for multimedia streaming over the internet." IEEE multimedia 18.4 (2011): 62-67.

[57] Sumby, William H., and Irwin Pollack. "Visual contribution to speech intelligibility in noise." The journal of the acoustical society of america 26.2 (1954): 212-215.

[58] Liang, Yi J., John G. Apostolopoulos, and Bernd Girod. "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames." IEEE Transactions on Circuits and Systems for Video Technology 18.7 (2008): 861-874.

[59] Hoßfeld, Tobias, et al. "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing." IEEE Transactions on Multimedia 16.2 (2013): 541-558.

[60] Chen, K-T., C-C. Tu, and W-C. Xiao. "Oneclick: A framework for measuring network quality of experience." IEEE INFOCOM 2009. IEEE, 2009.

[61] Tsolkas, Dimitris, et al. "A survey on parametric QoE estimation for popular services." Journal of network and computer applications 77 (2017): 1-17.

[62] Liu, Yao, et al. "Deriving and validating user experience model for DASH video streaming." IEEE Transactions on Broadcasting 61.4 (2015): 651-665.

[63] Rodríguez, Demóstenes Zegarra, et al. "Video quality metric for streaming service using DASH standard." IEEE Transactions on broadcasting 62.3 (2016): 628-639.

[64] Chen, Chao, et al. "Modeling the time Varying subjective quality of HTTP video streams with rate adaptations." IEEE Transactions on Image Processing 23.5 (2014): 2206-2221.

[65] Garcia, Marie Neige, Werner Robitza, and Alexander Raake. "On the accuracy of short-term quality models for long-term quality prediction." 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX). IEEE, 2015.

[66] Hoßfeld, Tobias, et al. "The memory effect and its implications on Web QoE modeling." 2011 23rd international teletraffic congress (ITC). IEEE, 2011.

[67] Shen, Yun, et al. "A method of QoE evaluation for adaptive streaming based on bitrate distribution." 2014 IEEE International Conference on Communications Workshops (ICC). IEEE, 2014.

[68] Bampis, Christos G., and Alan C. Bovik. "Learning to predict streaming video QoE: Distortions, rebuffering and memory." arXiv preprint arXiv:1703.00633 (2017).

[69] Bampis, Christos G., Zhi Li, and Alan C. Bovik. "Continuous prediction of streaming video QoE using dynamic networks." IEEE Signal Processing Letters 24.7 (2017): 1083-1087.

[70] Eswara, Nagabhushan, et al. "A continuous QoE evaluation framework for video streaming over HTTP." IEEE Transactions on Circuits and Systems for Video Technology 28.11 (2017): 3236-3250.

[71] Eswara, Nagabhushan, et al. "Modeling continuous video QoE evolution: A state space approach." 2018 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2018.

[72] Greene, Anthony J., Colin Prepscius, and William B. Levy. "Primacy versus recency in a quantitative model: Activity is the critical distinction." Learning & Memory 7.1 (2000): 48-57.

[73] Nguyen Duc, Tho, et al. "Modeling of cumulative QoE in on-demand video services: Role of memory effect and degree of interest." Future Internet 11.8 (2019): 171.

[74] Wu, H. R., K. R. Rao, and Ashraf A. Kassim. "Digital Video Image Quality and Perceptual Coding" Boca Raton." Journal of Electronic Imaging- 16 (2005).

[75] Cole, Robert G., and Joshua H. Rosenbluth. "Voice over IP performance monitoring." ACM SIGCOMM Computer Communication Review 31.2 (2001): 9-24.

[76] Takahashi, Akira, David Hands, and Vincent Barriac. "Standardization activities in the ITU for a QoE assessment of IPTV." IEEE Communications Magazine 46.2 (2008): 78-84.

[77] Lin, Weisi, and C-C. Jay Kuo. "Perceptual visual quality metrics: A survey." Journal of visual communication and image representation 22.4 (2011): 297-312.

[78] Winkler, Stefan, Animesh Sharma, and David McNally. "Perceptual video quality and blockiness metrics for multimedia streaming applications." Proceedings of the international symposium on wireless personal multimedia communications. 2001.

[79] Engelke, Ulrich, and Hans-Jurgen Zepernick. "Perceptual-based quality metrics for image and video services: A survey." 2007 Next Generation Internet Networks. IEEE, 2007.

[80] Chikkerur, Shyamprasad, et al. "Objective video quality assessment methods: A classification, review, and performance comparison." IEEE transactions on broadcasting 57.2 (2011): 165-182.

[81] Soundararajan, R., & Bovik, A. C. (2012). Video quality assessment by reduced reference spatio-temporal entropic differencing. IEEE Transactions on Circuits and Systems for Video Technology, 23(4), 684-694.

[82] Amershi, Saleema, et al. "Modeltracker: Redesigning performance analysis tools for machine learning." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015.

[83] Agrawal, Sumeet Kumar. "Evaluation Metrics for Classification Model: Classification Model Metrics." Analytics Vidhya, Sumeet Kumar Agrawal, 20 July 2021, https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/.

[85] Libretexts. (2022b, January 6). 10.2: The Linear Correlation Coefficient. Statistics LibreTexts. https://stats.libretexts.org/Bookshelves/Introductory_Statistics/.

[86] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," IEEE Transactions on Neural Networks, vol. 17, no. 5, pp. 1316 –1327, sept. 2006.

[87] Curteanu, Silvia, and Hugh Cartwright. "Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks." Journal of Chemometrics 25.10 (2011): 527-549.

[88] J. Brownlee, Statistical Methods for Machine Learning: Discover how to Transform Data in to Knowledge with python.

[89] Pedregosa, F. (n.d.). scikit-learn: machine learning in Python — scikit-learn 0.16.1

[90] Barman, Nabajeet, and Maria G. Martini. "QoE modeling for HTTP adaptive video streaming–a survey and open challenges." Ieee Access 7 (2019): 30831-30859.

[91] Capela, Matheus, et al. "Monogamy of temporal correlations: Witnessing non-Markovianity beyond data processing." Physical Review Research 2.1 (2020): 013350.

[92] Padhy, Neelamadhab, and Rasmita Panigrahi. "Data Mining: A prediction Technique for the workers in the PR Department of Orissa (Block and Panchayat)." arXiv preprint arXiv:1211.5724 (2012).

[93] Mandal, Subrata Kumar. "Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree." International Journal Of Engineering And Computer Science 6.2 (2017): 20388-20391.

[94] Hassani, Hossein, and Emmanuel Sirimal Silva. "Forecasting with big data: A review." Annals of Data Science 2.1 (2015): 5-19

[95] Razali, Nazim, et al. "Predicting player position for talent identification in association football." IOP Conference Series: Materials Science and Engineering. Vol. 226. No. 1. IOP Publishing, 2017.

[96] Pippal, Sanjeev, et al. "Data mining in social networking sites: A social media mining approach to generate effective business strategies." International Journal of Innovations & Advancement in Computer Science 3.2 (2014): 22-27.

[97] Thomas EH, Galambos N. What satisfies students Mining student-opinion data with regression and decision tree analysis. Research in Higher Education, 2004; 45(3), pp.251-269.

[98] Mandal, Subrata Kumar. "Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree." International Journal Of Engineering And Computer Science 6.2 (2017): 20388-20391.

[99] Alkhatib, Mohammad, Amir Talaei-Khoei, and Amir Ghapanchi. "Analysis of research in healthcare data analytics." arXiv preprint arXiv:1606.01354 (2016).

[100] Hochreiter, Sepp. "Investigations into dynamic neural networks." Diploma, Technical University of Munich 91.1 (1991).

[101] Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult." IEEE transactions on neural networks 5.2 (1994): 157-166.

[102] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735–1780.

[103] "References — Dive Into Deep Learning 1.0.0-alpha1.post0 Documentation." *References — Dive Into Deep Learning 1.0.0-alpha1.post0 Documentation*, 4 Jan. 2022, d2l.ai/chapter_references/zreferences.html.

[104] Sabarivani, A., Ramadevi, R., Pandian, R., & Krishnamoorthy, N. R. (2021). Effect of data preprocessing in the detection of epilepsy using machine learning techniques.

[105] Ghadiyaram, Deepti, Janice Pan, and Alan C. Bovik. "A subjective and objective study of stalling events in mobile streaming videos." IEEE Transactions on Circuits and Systems for Video Technology 29.1 (2017): 183-197.

[106] Bampis, Christos George, et al. "Study of temporal effects on subjective video quality of experience." IEEE Transactions on Image Processing 26.11 (2017): 5217-5231.

[107] Eswara, Nagabhushan, et al. "A continuous QoE evaluation framework for video streaming over HTTP." IEEE Transactions on Circuits and Systems for Video Technology 28.11 (2017): 3236-3250.

[108] Chen, Chao, et al. "Modeling the time—Varying subjective quality of HTTP video streams with rate adaptations." IEEE Transactions on Image Processing 23.5 (2014): 2206-2221.

[109] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 4, pp. 684–694, Apr. 2013.

[110] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in Proc. Asilomar Conference on Signals, Systems and Computers, vol. 2, Nov. 2003, pp. 1398–1402.

[111] N. Eswara, H. P. Sethuram, S. Chakraborty, K. Kuchi, A. Kumar, and S. S. Channappayya, "Modeling continuous video qoe evolu- tion: A state space approach," in IEEE International Conference on Multimedia and Expo (ICME), Jul. 2018. [Online]. Available: https://arxiv.org/abs/1805.06335.

[112] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, "Modeling the time-varying subjective quality of http video streams with rate adaptations," IEEE Trans. Image Process., vol. 23, no. 5, pp. 2206–2221, May 2014.

[113] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[114] N. Eswara, K. Manasa, A. Kommineni, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "A continuous qoe evaluation framework for video streaming over http," IEEE Trans. Circuits Syst. Video Technol., vol. PP, no. 99, pp. 1–1, 2017.

[115] C. G. Bampis, Z. Li, and A. C. Bovik, "Continuous prediction of streaming video qoe using dynamic networks," IEEE Signal Process. Lett., vol. 24, no. 7, pp. 1083–1087, Jul. 2017.

121

Published Paper (1)

Video Streaming Forecast Quality of Experience- A survey

Publisher: IEEE   Cite This   PDF

Ahmed S. Ghafil ; Israa H. Ali   **All Authors**

183
Full
Text Views

Feedback

Abstract

Document Sections

I. Introduction

II. Algorithms for video streaming

III. Related work

Abstract:
This paper provides a subjective analysis of streaming video to understand human visual Quality of Experience (QoE). The emergence of the Corona pandemic has prompted a rise in demand for internet services, particularly for video streaming. A limited number of companies in Europe choose to reduce video transmission quality to escape Internet strain. The QoE markers are defined as either satisfactory if the video loads in less than 10 seconds and plays out efficiently or undesirable if the loading time is above 10 seconds or the playout suffers from stands. Online media will display various kinds of high motion sports and music videos. Such metrics would enable internet providers to assess the degree to which the QoE

Published Paper (2)

State-of-the-Art in QoE for Video streaming

Publisher: IEEE   Cite This   📄 PDF

Ahmed Sileh Ghafil ; Israa Hadi Ali   **All Authors**

81
Full
Text Views

Ⓡ   ⋘   Ⓒ   📂   🔔

Document Sections

I. Introduction

II. HTTP Dash

III. Taxonomy for measuring
QoE

IV. Estimate of the objective
QoE

V. Literature Review for Quality
of experience

Show Full Outline ▾

Abstract:
Videos have been the most commonly used source of entertainment on the Internet, accounting for 65% of the world's Internet traffic in 2020. Efficient internet delivery of multimedia with low bandwidth requirements using DASH (Dynamic adaptive streaming HTTP) means better quality for the end-user and lower costs for the server. With HTTP streaming, it is possible to provide a weak quality of service (QoS), but a better quality of experience (QoE), because HTTP offers a stable meaning of transportation. In this paper, a comprehensive survey on the use of Machine learning (ML) for predicate QoE for user progress of this research area. Furthermore, what are the current trending algorithms in frontier ML research, and discuss why you prefer dealing with "HTTP DASH" describes the benefits of using QoE with HTTP webpage.

الخلاصة

يؤدي التطور الهائل لتطبيقات الوسائط المتدفقة إلى زيادة الطلب على قياس جودة التجربة (QoE). يتضمن مفهوم جودة الأعمال في الاتصالات تعريف البث الديناميكي التكيفي (DASH) عبر HTTP. يركز هذا العمل على ميزات المحتوى. تظل جودة الفيديو المقدم للعميل متغيرة بمرور الوقت ، اعتمادًا على ظروف الشبكة الشاملة. قد يؤدي تغيير ظروف الشبكة أيضًا إلى تدهور محتوى تشغيل الفيديو ، مما يؤدي إلى أحداث إعادة التخزين المؤقت (السكون والتجميد). تؤثر هذه الميزات على رضا المستخدم وتتسبب في تدهور جودة تجربة المستخدم. المشكلة هي عدم رضا المستخدم عن التقييم الذاتي المستمر للخدمة أو التطبيق المقدم له.

يتنبأ نظام المقترح لقياس جودة الخدمة لدفق فيديو بدون مرجع. تقدم هذه الأطروحة مفهوم GRU-QoE ، وهو نموذج تنبؤ QoE قائم على الشبكة العصبية المتكرر باستخدام شبكة Gate Recurrent Unite (GRU). GRU-QoE عبارة عن شبكة من GRU المتتالية لالتقاط اللاخطية والاعتماديات الزمنية المعقدة المتضمنة في QoE المتغيرة بمرور الوقت. النموذج المقترح مستوحى من آلية عملية الديناميكيات غير الماركوفية لـ QoE ، والتي تُستخدم للتنبؤ بالحالة الحالية بناءً على الحالات السابقة. بالإضافة إلى ذلك ، يوضح النموذج أن الطبقة النهائية كثيفة مع تنشيط Softmax. تُستخدم وظيفة softmax كدالة تنشيط في طبقة الإخراج لنماذج الشبكة العصبية التي تتنبأ بالنماذج الاحتمالية متعددة الحدود.

تتم مقارنة النموذج المقترح بأحدث نماذج التنبؤ بجودة الخدمة ويثبت أنه يتفوق عليها عبر LIVE Mobile Video Stall Database-II و LIVE QoE و LFOVIA QoE و LIVE Netflix datasets. ينظر نموذج التقييم إلى معامل الارتباط الخطي (LCC) ، ومعامل ارتباط ترتيب سبيرمان (SROCC) ، والخطأ المعياري الجذر (RMSE) ، ومعدل الانقطاع (OR) لمعرفة مدى جودة عمل التنبؤ بجودة الخدمة. تظهر المقارنات الإحصائية أن النموذج المقترح يتفوق على أحدث نماذج التنبؤ بجودة التجربة. تم إجراء العديد من التجارب على محتوى الفيديو والشبكات القياسية المختلفة. أظهرت النتائج تحسنًا في توقعات قيمة الجودة قريبة من تلك الخاصة بالمشاهد الفعلي. يعمل الحساب باستخدام الطريقة المقترحة على تحسين جودة نموذج NARX باستخدام تقييم LCC عبر قاعدة بيانات LIVE Netflix التي تم تحسينها بأكثر من 25.4٪ ، وتم تحسين SVR-QoE على قاعدة بيانات LFOVIA QoE بأكثر من 20.4٪ ، ونموذج -Hammerstein Wiener عبر تحسنت قاعدة بيانات LIVE QoE بأكثر من 15٪ ، ونموذج NLSS-QoE بواسطة LIVE Mobile Video Stall Database-II أكثر من 21٪.

التنبؤ المستمر وتقييم جودة الفيديو باستخدام التعلم العميق

أطروحة

مقدمة إلى مجلس كلية تكنولوجيا المعلومات في جامعة بابل والتي هي جزء من

متطلبات الحصول على درجة الدكتوراه في تكنولوجيا المعلومات – البرمجيات

مقدمة من قبل

احمد صالح غافل حمد

باشراف

أ.د. اسراء هادي علي حسين

2022 م   **هـ 1444**