

Article

# Multilayer Perceptron Neural Network-Based QoS-Aware, Content-Aware and Device-Aware QoE Prediction Model: A Proposed Prediction Model for Medical Ultrasound Streaming Over Small Cell Networks

Ikram U. Rehman <sup>1</sup>, Moustafa M. Nasralla <sup>2,\*</sup> and Nada Y. Philip <sup>3</sup>

<sup>1</sup> Department of Engineering, Coventry University Group, Coventry, CV1 5DL, UK; Ikram.Rehman@Coventry.ac.uk

<sup>2</sup> Department of Communications and Networks Engineering, Prince Sultan University, Riyadh, 11586, Saudi Arabia; mnasralla@psu.edu.sa

<sup>3</sup> Faculty of Science, Engineering, and Computing, Kingston University, London, KT1 2EE, UK; n.philip@kingston.ac.uk

\* Correspondence: mnasralla@psu.edu.sa; Tel.: +966114948042

Received: 8 January 2019; Accepted: 29 January 2019; Published: 7 February 2019



**Abstract:** This paper presents a QoS-aware, content-aware and device-aware nonintrusive medical QoE (m-QoE) prediction model over small cell networks. The proposed prediction model utilises a Multilayer Perceptron (MLP) neural network to predict m-QoE. It also acts as a platform to maintain and optimise the acceptable diagnostic quality through a device-aware adaptive video streaming mechanism. The proposed model is trained for an unseen dataset of input variables such as QoS, content features and display device characteristics, to produce an output value in the form of m-QoE (i.e. MOS). The efficiency of the proposed model is validated through subjective tests carried by medical experts. The prediction accuracy obtained via the correlation coefficient and Root Mean-Square-Error (RMSE) indicates that the proposed model succeeds in measuring m-QoE closer to the visual perception of the medical experts. Furthermore, we have addressed two main research questions: (1) How significant is ultrasound video content type in determining m-QoE? (2) How much of a role does the screen size and device resolution play in medical experts' diagnostic experience? The former is answered through the content classification of ultrasound video sequences based on their spatiotemporal features, by including these features in the proposed prediction model, and validating their significance through medical experts' subjective ratings. The latter is answered by conducting a novel subjective experiment of the ultrasound video sequences across multiple devices.

**Keywords:** Mobile health (M-health); Small cell networks; MLP Neural networks; Medical Quality of Service (m-QoS); Medical Quality of Experience (m-QoE)

## 1. Introduction

Presently, the utmost requirement of mobile healthcare (m-health) end-users and, in particular, healthcare professionals, is to have flexibility in their digital practice, where they have the provision to view clinical and diagnostic information regardless of location and viewing device. In order to satisfy this requirement, diagnostic multimedia content, especially medical video streaming applications, needs to be efficiently managed, which is a significant challenge, with a view to ensuring the desired medical Quality of Experience (m-QoE). This is because video contents are susceptible to various impairments from the time of capturing to preprocessing, compression, transmission

and postprocessing; each of these stages may deteriorate the video's Quality of Experience (QoE). Besides the aforementioned impairments, the perceived QoE is greatly influenced by the viewing content characteristics (e.g., spatial and temporal) and display device capabilities (e.g., screen size and resolution). Ignoring these and predicting QoE on the basis of Quality of Service (QoS) alone could lead to inadequate QoE levels, as well as the wastage of scarce network resources. From the literature, it can be construed that the predictive video QoE evaluation methods were previously carried out over 3G (i.e., Universal Mobile Telecommunications Service (UMTS)) networks, rather than the proposed 4G and beyond small cell networks, while evaluating the video quality. In addition, such predictive models rarely considered the combined consequential effects of network conditions, video content and viewing device capabilities [1–5]. It has been realised that end-users may react differently to the same video content while viewing it on TVs, laptops, tablets or smartphones, which can create significant variations in the QoE ratings.

The inadequacies in the QoE framework, as discussed, provide motivation for designing an m-QoE prediction model that incorporates the influencing parameters stated above, along with the validation of prediction accuracy through extensive subjective testing across multiple devices. From m-health's perspective, there is no such medical video quality assessment model that evaluates m-QoE, considering the network conditions, spatial and temporal features of the video content and display device characteristics, all combined together in a single framework.

In this paper, we adopt a hybrid approach to predict m-QoE using a multilayer perceptron (MLP) neural network. We present a video quality prediction model over small cell wireless networks that inputs QoS parameters, video content features and display device characteristics to predict video quality in a QoS-aware, content-aware and device-aware m-QoE framework. As a first step in this kind of study, the model seems efficient and suitable for all types of video content and devices so that it can be implemented at the receiver side to predict video quality and, if appropriate, control the end-to-end perceived quality by the use of adaptable mechanisms at the transmitting side. To the best of our knowledge, there has been no previous work in medical video quality prediction and control over small cell networks combining QoS, content and device related parameters. The proposed framework is our major contribution to the m-health domain, and can also be considered a suitable candidate for the prediction and automatic classification of perceived user quality for future 5G m-health systems. Our emphasis is not only on the prediction of m-QoE but also on the use of small cell networks to enhance the m-QoE in clinically sensitive m-health applications.

The rest of the paper is organised as follows. Section 2 presents the background on content awareness, device-awareness and machine learning-based QoE modelling, respectively. The methodology of the proposed MLP neural network-based m-QoE prediction model, including content classification, dataset generation, subjective assessment, implementation of the MLP neural networks and the proposed device-aware adaptive medical video streaming controller is presented in Section 3. In Section 4, we analyse the obtained results and discuss our research findings. Finally, Section 5 concludes this paper.

## 2. Background

This section presents the background of the medical ultrasound video streaming, small cell networks, content awareness, device-awareness, and QoE modelling and prediction.

### 2.1. Medical Ultrasound Video Streaming

The focus of this article is the streaming of medical ultrasound video as an example of an m-health application. Ultrasound videos are an important part of modern medical diagnosis procedures, and streaming such videos is vital for remote diagnosis and consultations [6,7]. Furthermore, medical ultrasound video streaming is known to be a demanding m-health application, and one that requires remote expert opinions, particularly in isolated areas where medical experts are not always available. In these cases, experts have nominal or limited access to the medical data (e.g., ultrasound video and radiology) to assess the level of severity in order to make important diagnostic decisions.

These decisions may include whether there is a need to send the patient by ambulance to the nearest hospital or whether it is appropriate to continue with the treatment at home [6]. According to a study, approximately 50% of the patients transported to a hospital for an ultrasound examination are sent back home within a few hours [8]. Furthermore, a study in Finland investigated the impact of teleradiology consultations on treatment costs and unnecessary patient transportation. The results suggested reductions of unnecessary patient transportation by 81%, hence reducing operational costs. In addition, 75% of patients transported to the hospital were treated immediately without the need for further radiological examination [8].

The earliest studies on ultrasound telemedical applications were carried out in the mid-1990s, when a 1.5 Mbps T-1 leased line communication link was used to transmit ultrasound images in real-time from remote patients' locations to a distant medical facility. The entire system was supervised by medical experts who would communicate in real-time with the technician acquiring the ultrasound images [9]. Since then, several studies have been carried out on tele-ultrasonography systems, mainly evaluating the visual quality of the transmitted ultrasound images/videos [10]. However, the studies concluded that limited available bandwidth, packet loss and delay are the major factors affecting accurate and robust diagnosis in the wireless ultrasound video streaming process.

In recent years, the advancements in wireless communications and network technologies have significantly contributed to tele-ultrasonography applications. Although the evolution of mobile communication technology aims to offer higher data rates and better exploitation of available network resources, it presents another challenge: coping with the rapidly growing internet traffic, particularly the influx of high bandwidth demanding video applications. According to a report by CISCO, internet video accounts for 70% of customer internet traffic and will reach 82% by 2020 [11]. The increase in video traffic has created high-pressure demands for spectral efficiency, energy consumption, high mobility, seamless coverage and the varying requirement levels of QoS and QoE. In a shared environment, the medical ultrasound video application will coexist and compete with regular video traffic to deliver healthcare services over mobile networks. To overcome the above-stated challenges, this article proposes the use of small cell networks as a communication platform for robust tele-ultrasound m-health systems.

## 2.2. Small Cell Networks

To boost the indoor coverage in an inexpensive way is a major challenge for wireless networks. It is an important issue to address because it is estimated that 80% of mobile traffic is generated indoors. Small cell networks are inexpensive, low-power and high-speed home-based solutions (HBs) that can be installed autonomously by end users through the 'plug and play' method indoors [12]. Small cell networks usually operate in the licensed spectrum and typically have a coverage area of ten meters. The working principle of small cell networks is to connect with standard user equipment such as cell phones, tablets and laptops through radio interfaces. Next, the user traffic is sent to the operator's core network using legacy broadband connections such as DSL/optical fibre networks [13,14]. From the user equipment's perspective, it recognises a small cell base station as a regular base station, and since the distance between the small cell base station and the users is short, the transmission power consumption of the user equipment is lowered, thus prolonging the battery life of the device [14]. Other benefits of small cell networks include better coverage, lower infrastructure requirements, reduced power consumption, improved Signal-to-Interference Noise Ratio (SINR), improved throughput and enhanced QoS and QoE [12–14].

Recently, small cells have attracted a great deal of interest from researchers, as is evident from the increase in the number of publications within this field [15–22].

The concept of small cells in the field of m-health is relatively new. Researchers are focusing on several concepts of small cell deployment in emergency and nonemergency medical scenarios [20]. Altman, Z. et. al. [16] proposed a small cell architecture option in order to achieve very low radio emissions in hospitals or the home environment. They have presented an approach under the

HOMESNET project, where paramedical staff can use a small cell network in the patient's apartment to provide telemedicine services. The performance of the proposed approach was compared to the conventional macrocell-based approach. The results indicated a significant reduction in the service outage rates using the proposed small cell network approach. These simulations were based on the indoor environment in a multidwelling unit building structure.

A hybrid small cell-based sensor network approach was presented by Maciuca, A. et. al. [21] to home monitor patients with chronic diseases. Sensors were grouped into three categories based on their functionality level, resource consumption and mobility. Experiments were conducted by placing sensors on the subject to identify different actions and to determine parameters such as humidity and temperature in the environment. Adaptive rates of communication were used to reduce energy consumption and levels of radiation. All of the data harvesting and processing was carried out at the small cell level.

A 5G-health use case for a small cell-based ambulance scenario for medical video streaming is proposed in [22]. The authors investigated the impact of small cell heterogeneous networks for medical video streaming as an example of m-health application in the uplink direction, they introduced an important m-health use case (i.e., an ambulance scenario) along with its system model and technical requirements and they provided a network performance analysis for medical video sequences affected by packet losses and different networks as a benchmark to enable researchers to test their medical video quality evaluation algorithms. The results of the proposed scenario showed in terms of medical-QoS key performance indicators that a mobile small cell-based ambulance scenario outperforms the traditional macrocell network scenario.

De, D. et. al. [23] presented a scheme based on mobile cloud computing and a small cell network. In this scheme, the data of the patient is transferred from the body sensors to the mobile device, and then further transmitted to the small cell network. This small cell contains a database server that verifies the normality of the data. In the case where any abnormality is detected, the data is sent to the cloud where it can be stored and accessed by healthcare professionals in order to take proper action to treat the patient. This small cell approach reduces the consumption cost of accessing a large amount of data on clouds, compared to a macrocell network.

A limited number of studies have been conducted on deploying small cells in the m-health environment. These studies mostly addressed low data rate m-health applications, such as wireless medical sensors [24]. Furthermore, the literature lacks any exploitation of small cells for high-bandwidth-demanding m-health applications (e.g., ultrasound video streaming), which is addressed in this paper.

### 2.3. Content Awareness

The studies with regard to video quality assessment reveal that video content parameters are likely to influence the evaluation of QoE, and different content types under the interaction of similar encoder parameters and network conditions may result in varying QoE levels [25,26]. Furthermore, the studies also show that the current video QoE prediction models are mostly content blind, that is, rather than considering different content types (e.g., slow motion and fast motion video clips) the prediction is based on averaging different video content types. This approach is not desirable and can affect the QoE prediction accuracy of a particular content type [27,28].

Various QoE prediction models based on QoS impairments and encoding parameters have been proposed in the literature [3,4,29–36]. The proposed models predicted video quality based on the network conditions and encoding parameters only, omitting the video content characteristics. The Anegekuh, L. et. al. [28] and Khan, A. et. al. [37] highlighted the importance of video content type by stating that it is the second most influential factor in predicting QoE, after QoS. To establish a content-aware QoE prediction model, the studies carried out in [37–44], extracted spatial and temporal features from impaired video sequences. These features were then used to categorise video sequences into groups based on their degree of movement, such as slow movement, moderate movement and/or fast movement. Finally, the mathematical expressions of the Sum of Absolute

Difference (SAD) were applied to predict QoE based on content classification. Likewise, the studies carried out in [28,39,45] utilised a video codec to predict content-aware video quality, where the authors extracted spatiotemporal features from the codec and proposed QoE evaluation frameworks based on objective video quality assessment metrics. All these studies provide a fairly accurate estimation of QoE considering QoS parameters, encoder settings, and video content types. However, such studies are limited—regarding the m-health domain for ultrasound video streaming, they are almost nonexistent. Moreover, the soundness of QoE prediction models proposed in these studies lack validation through extensive subjective testing.

#### 2.4. Device-awareness

In recent times, the influx of emerging devices, particularly the popularity of smartphones, has been a major breakthrough for end-users. End-users now have the flexibility to view any video content on their choice of device, such as laptops, tablets and smartphones, whereas earlier these were limited to TV viewing only [46]. It is observed that subjective testing carried out on the same video content but on different devices alters the entire viewing experience of the same video content, leading to varying results [47]. This is due to the fact that display device characteristics, such as system specifications (e.g., operating system), screen size, resolution, battery life and device mobility are unique to each type of device, which impacts the subjective ratings.

With the advancement in device capabilities, end-users now expect operators and content providers to enable the delivery of video applications with high levels of QoE anytime and anywhere, regardless of the device being used [48,49].

The earlier research studies mainly emphasised establishing a correlation between QoE and video acquisition, compression, network conditions, encoding methods and more recently video content types, while disregarding the impact of device capabilities on the perceptual video quality of end-users [50]. This disregard for the inclusion of display device features in predicting QoE has motivated this research, and has accordingly been addressed in this paper.

As mentioned above, the impact of display device preferences for video streaming applications has rarely been addressed in the literature, let alone in medical video streaming m-health applications. However, some studies have contemplated whether the display device really changes the end-user's visual experience, affecting the subjective rating. For example, Cermak, G. et. al. [51] investigated the impact of screen resolution on the subjective ratings. The selected screen resolutions under experiment were QCIF ( $176 \times 144$ ), CIF ( $352 \times 288$ ), VGA ( $640 \times 480$ ), and HD ( $1920 \times 1200$ ). Another study in [52], evaluated the quality of scalable video on handheld mobile devices, that is, a smartphone and a tablet, with screen sizes of 4.3 inches ( $800 \times 480$ ) and 10.1 inches ( $1280 \times 800$ ), respectively. Likewise, López, J.P et. al. [46] and Khan, A. et. al. [53] conducted subjective testing for video streaming application viewed over multiple devices with varying screen sizes, resolutions, and operating systems (e.g., PCs, laptops, PDAs and smartphones). In addition, the quality of a video sequence was evaluated for different bitrates (i.e., 200 kbps and 400 kbps) encoded using H.264/AVC video codec. Rodríguez, D. et. al. [54] proposed a subjective video quality assessment model, which included user preference as a significant parameter in assessing visual quality for different video contents. Lately, Rehman, A. et. al. [55] went on to explore the advantage of including display device features under various network conditions in objective video quality assessment methods. The proposed model was called Structural Similarity Index Metric plus (SSIMplus), and was based on a full-reference method. In a nutshell, the aforementioned studies concluded that the end-user QoE changes with the change in display device and, hence, it is suggested that display device characteristics be incorporated into QoE prediction frameworks. From the m-health perspective, the available literature does not contain any prior studies that investigate the influence of display device types and device preference on m-QoE for medical video applications.

### 2.5. QoE Modelling and Prediction: A Machine Learning Approach

Recently, there has been an increasing research interest in applying AI and machine learning methods to the development of video quality prediction models [56]. For instance, Mushtaq, M.S. et. al. [57] experimented on using a range of machine learning algorithms to establish the effect of QoS on QoE and formed a prediction model for cloud-based video streaming application. The authors achieved a fair degree of prediction accuracy with all six machine learning algorithms they considered.

Similarly, Pitas, C.N. et. al. [58] presented an experimental approach using quality prediction for mobile networks. In addition, the authors provided an empirical comparison of recent data mining approaches for audio as well as video quality prediction. Their proposed prediction models were generic and adaptable to various radio access network technologies with an aim to predict and enhance end-users' QoE.

Furthermore, Pitas, C.N. et. al. [59] proposed an Adaptive Neuro Fuzzy Inference System (ANFIS)-based end-to-end QoS prediction model over mobile networks. The authors obtained wireless key performance indicators, which influence the quality of the received audio and video contents. Besides, end-user QoE was estimated by utilising fuzzy-based approaches. The authors managed to achieve satisfactory prediction accuracy by employing their proposed methodology.

Likewise, a QoE prediction model based on two learning models is proposed in [34]. The first learning model is an ANFIS and the second model is a nonlinear regression analysis model. Their study showed that the two models were able to achieve a high degree of accuracy. However, both models require further validation through more subjective assessment.

Machine learning approaches can be exploited in many ways; for example, they can be used for classification, regression and clustering purposes [60]. In particular, the objective of regression analysis is to find the approximation of a nonlinear function. In this paper, we adopt an MLP regressor, which is a feedforward neural network model used to approximate any given function to a preset value.

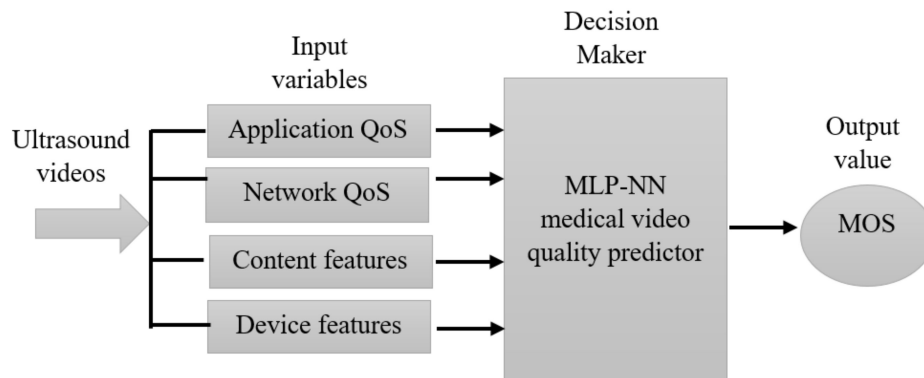
The majority of the objective video quality assessment methods explicitly map the mathematically computed objective values to the human visual system (HVS), which possess a nonlinear nature. Such direct mapping can be computationally simple but often leads to incorrect and unreliable results. On the other hand, QoE frameworks developed over machine learning algorithms avoid such direct nonlinear mapping; rather they mimic human perceptions of impairments [61].

The QoE prediction model based on machine learning algorithms usually functions in two phases. Firstly, the dataset of video impairments established upon feature representation is defined. Secondly, the machine learning algorithm learns and trains itself to the behaviour of earlier defined dataset and maps it to the QoE scores—that is, imitating the HVS. The success in imitating human perception of machine-learning-based QoE models depends on the rigorousness of its training phase. The more rigorously the model is trained, the higher the prediction accuracy.

## 3. Methodology

The methodology followed in this paper can be outlined as the prediction of m-QoE or Mean Opinion Score (MOS) value for a given set of input variables characterised by QoS parameters, content awareness and display device features. Figure 1 depicts the functional diagram of the proposed m-QoE prediction model.

In Figure 1, the application QoS parameters are Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and video bitrate; the network QoS parameters are Packet Loss Rate (PLR), delay and throughput; content features are the full or the subset of spatial and temporal features extracted from the ultrasound video sequences; device features are the device-related characteristics (e.g., screen size and resolution); and the medical video quality predictor is the MLP feedforward neural network.



**Figure 1.** Functional block diagram of proposed m-QoE model.

The methodology adopted in this paper starts with content classification of medical ultrasound sequences (i.e., slow movement, medium movement and high movement) and content features extraction (i.e., temporal and spatial). This is followed by dataset generation, next, the subjective tests are carried out by medical experts over multiple devices (i.e., laptop, tablet and smartphone). Then, the MLP neural network-based prediction model predicts m-QoE, and lastly, the device-aware adaptive ultrasound streaming controller uses the predicted m-QoE scores at the receiving side as a feedback to adapt the medical video content at the transmitting side.

### 3.1. Content Classification

For the medical ultrasound videos under consideration (given in Table 1), we chose cluster analysis, which is a multivariate statistical analysis used to categorise these videos into classes of similar characteristics. Cluster analysis is carried out by computing Euclidean distances, which can be defined as the square root of the sum of two corresponding points between different features in a video sequence [62]. This technique categorises medical ultrasound videos into classes based on the nearest Euclidean distance. We performed cluster analysis on multiple ultrasound video sequences and selected a video sequence that appropriately represents each category. The aforementioned categorisation of medical ultrasound videos is illustrated in Figure 2.

**Table 1.** Medical video content illustration.

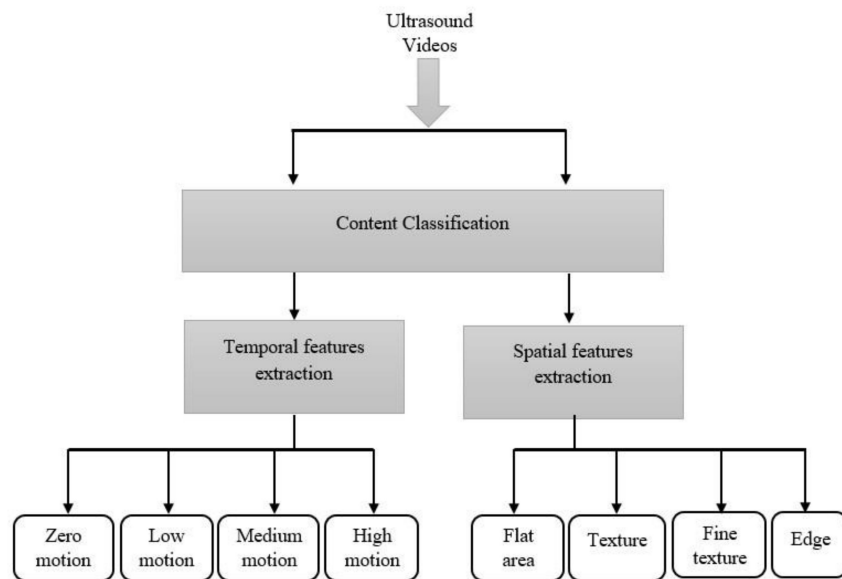
Ultrasound name	Activity level	Frames	Frame rate	Resolution	Duration
Thyroid	Slow Movement (SM)	250	30 fps	960 × 720	10 sec
Stomach	Medium Movement (MM)	250	30 fps	960 × 720	10 sec
Gallbladder	High Movement (HM)	250	30 fps	960 × 720	10 sec

#### 3.1.1. Temporal Features Extraction

Temporal features represent the movement in the video sequence, denoted by the sum of absolute difference (SAD) value. It is defined as the pixel wise sum and comparison between the reference and measured frames, which is mathematically computed as [37]

$$SAD_{n,m} = \sum_{i=1}^N \sum_{j=1}^M |B_n(i,j) - B_m(i,j)| \quad (1)$$

where,  $B_n$  and  $B_m$  denote the reference and measured frames, respectively,  $N$  and  $M$  represent the size of these frames and  $i$  and  $j$  symbolise the coordinates of the pixel.



**Figure 2.** Content classification methodology.

The method to extract temporal features is through block-based motion estimation, where a macroblock size of  $N \times M$  and the degree of prediction error represent the motion vector associated with it. Furthermore, every macroblock inside a video frame is categorised into zero prediction error, low prediction error, medium prediction error and high prediction error, based on the degree of prediction error. For example, the zero-motion macroblock represents zero motion features and nominal prediction error. The degree of prediction error can be mathematically computed by means of either SAD or mean squared error (MSE) values. These values are then compared to the threshold values.

The comparison between the motion prediction error and the threshold values aids with identifying whether the macroblocks should be classified as zero prediction error, low prediction error, medium prediction error or high prediction error.

Once the motion vectors are classified then the temporal features can be extracted from all the video frames of three ultrasound video sequences using the following mathematical expressions.

$$F_{T,zero} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(zero)}{M} \quad (2)$$

$$F_{T,low} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(low)}{M} \quad (3)$$

$$F_{T,medium} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(medium)}{M} \quad (4)$$

$$F_{T,high} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(high)}{M} \quad (5)$$

where the macroblocks classified as zero prediction error, low prediction error, medium prediction error and high prediction error in the  $k^{th}$  video frame are denoted by  $S^k(zero)$ ,  $S^k(low)$ ,  $S^k(medium)$  and  $S^k(high)$ , respectively.

It is important to note that the motion estimation involves computational complexities; therefore, a good practice is to select the key frames that contain the most scene changes rather than extracting motion features from the entire video sequence. This technique contributes towards reducing the computational complexities.



### 3.1.2. Spatial Features Extraction

Spatial features represent the spatial regions in the video sequence. Every macroblock inside a video frame is categorised into four classes—flat-area, texture, fine-texture and edge. For instance, the macroblocks categorised as ‘flat-area’ are smooth in appearance and have minimum variance between the pixel value and its mean. Further, the macroblocks categorised as ‘texture’ have a coarser texture in the spatial regions. The ‘fine-texture’ category falls in between the ‘flat-area’ and ‘texture’ categories, as its spatial regions are coarser than the ‘flat-area’ but finer than the ‘texture’. Lastly, the macroblocks categorised as ‘edge’ represent the edge pixels of a frame. In addition, the categorisation of the aforementioned classes is based on the texture masking energy (TME), which is mathematically expressed as

$$T_{E,m}^k = \left\{ \sum_{i=1}^N \sum_{j=1}^N [\hat{H}(i,j)]^2 [X_m^k(i,j)]^2 \right\}^{1/2} \quad (6)$$

where  $X_m^k(I, j)$  represents the pixel value of the  $m^{\text{th}}$  macroblock in the  $k^{\text{th}}$  frame and in the pixel coordinates of  $(i, j)$ ,  $\hat{H}(i, j)$  denotes the function of HVS sensitivity corresponding to the spatial frequency in the pixel coordinates of  $(i, j)$ . Additionally, to attain precise TME readings, it may be necessary to divide each macroblock into sub-blocks and eventually compute the value of TME for each of these sub-blocks.

Similar to the temporal feature extraction mentioned earlier, threshold values are set to determine which spatial regions are categorised as ‘flat-area’, ‘texture’, ‘fine-texture’ and ‘edge’.

Once the spatial regions are categorised, then the spatial features can be extracted from all the video frames of the three ultrasound video sequences using the following mathematical expressions.

$$F_{S, \text{flat}} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(\text{flat} - \text{area})}{M} \quad (7)$$

$$F_{S, \text{texture}} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(\text{texture})}{M} \quad (8)$$

$$F_{S, \text{fine}} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(\text{fine} - \text{texture})}{M} \quad (9)$$

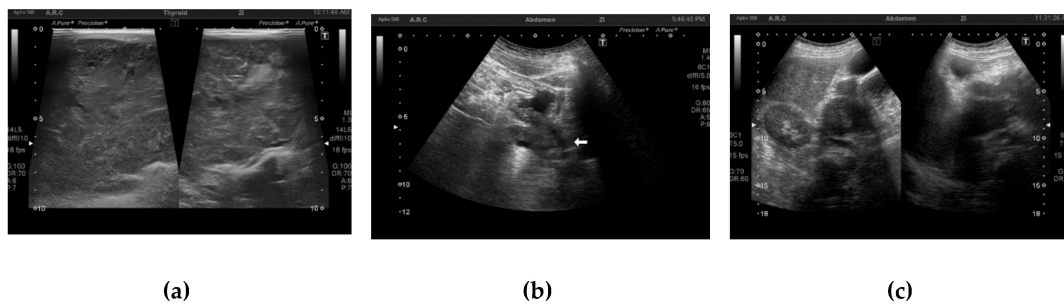
$$F_{S, \text{edge}} = \frac{1}{K} \sum_{k=1}^K \frac{S^k(\text{edge})}{M} \quad (10)$$

where the macroblocks categorised as flat-area, texture, fine-texture and edge in the  $k^{\text{th}}$  video frame are denoted by  $S^k(\text{flat} - \text{area})$ ,  $S^k(\text{texture})$ , and  $S^k(\text{edge})$ , respectively.

### 3.2. Dataset Generation

The MLP neural network requires training and therefore a dataset needs to be generated comprising all input variables and output value. Later on, this dataset is used for the validation of the proposed m-QoE prediction model. The vital aspect of this paper is the use of small cell (i.e., femtocell) networks to obtain the medical QoS (m-QoS) parameters in dataset generation for the ultrasound video streaming application.

The selected medical ultrasound video sequences (as shown in Figure 3) are converted into YUV format and encoded using H.264/AVC standard at the bitrates of 400 kbps, 500 kbps and 600 kbps. The frame rate and resolution are kept constant at 30 fps and  $960 \times 720$ , respectively.

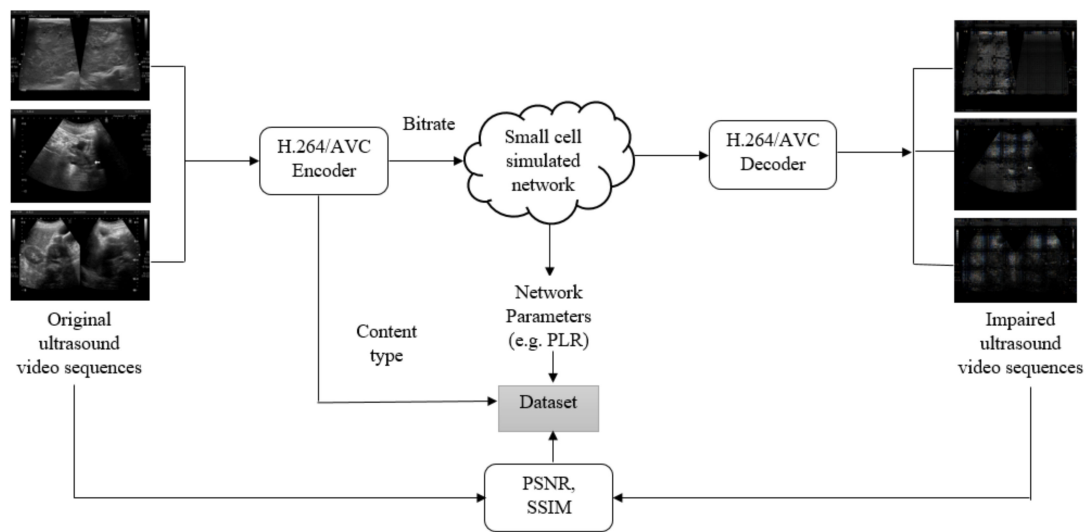


**Figure 3.** Screenshots of ultrasound video sequences. (a) Thyroid (SM), (b) stomach (MM) and (c) gallbladder (HM).

An LTE-Sim system level simulator [63,64] is used to transmit the ultrasound video sequences over small cell (i.e., femtocell) networks under different network conditions. On the receiving side, we created a crowded indoor hospital scenario surrounded by varying numbers of users generating different background traffic. A Proportional Fair (PF) type of downlink scheduling algorithm was chosen as it guarantees fairness as well as maintaining high throughput [65]. For the test ultrasound video sequences, we created realistic video trace files that send packets to the LTE-Sim simulator. The parameters used to create the simulation environment are presented in Table 2. Subsequently, the network QoS (e.g., PLR, delay and throughput) as well as application QoS metrics (e.g., PSNR, and SSIM) over small cell networks are obtained (refer to Table 3). These parameters formed a dataset for the three chosen ultrasound video content types in order to facilitate the training and validation of the MLP neural network. The pictorial representation of the dataset generation is shown in Figure 4.

**Table 2.** Downlink small cell (i.e., femtocell) simulation parameters.

Parameters	Values
Bandwidth	5 MHz
Cell radius	10 m
Frame structure	FDD
Macrocell BS power	43 dBm
Femtocell BS power	20 dBm
Maximum delay	100 ms
Video bitrate	400, 500, 600 kbps
Scheduler type	Proportional Fair (PF)
Video duration	10 sec
Simulation time	30 sec
UE speed	3 km/h
Path loss/channel model	Urban (Pedestrian-A propagation model)
Simulation repetitions	10



**Figure 4.** Experimental setup for dataset generation.

For each encoding bitrate (i.e., 400 kbps, 500 kbps and 600 kbps) and content type (i.e., SM, MM and HM), a total of 54 impaired ultrasound video sequences are generated for the training and validation of the m-QoE prediction model.

**Table 3.** m-QoS results over small cell network.

Ultrasound video (Thyroid-SM)																				
Bitrate 400 Kbps						Bitrate 500 Kbps						Bitrate 600 Kbps								
No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60
PLR (%)	0	0	0.83	7.25	10.54	14.03	PLR (%)	0	0	0.84	7.4	11.7	15.7	PLR (%)	0	0	0.86	7.6	14	16.5
Delay (ms)	8.78	17.22	25.72	30.6	33.71	37.45	Delay (ms)	9.12	17.6	26.4	31.3	34.1	37.4	Delay (ms)	9.6	18.8	28	32.8	35	38.2
Total Throughput (Mbps)	4.2	8.5	12.7	15.7	18	20.5	Throughput (Mbps)	5.2	10.4	15.5	19.4	22.8	26	Throughput (Mbps)	6.2	12.4	18.5	23.3	27.5	31.6
PSNR (dB)	37.1	36.6	34.9	30.7	27.3	24.8	PSNR (dB)	37.1	36.5	34.8	31.8	29.5	27.7	PSNR (dB)	37.1	36.4	34.4	32.2	30.8	30.2
SSIM	0.96	0.95	0.91	0.83	0.79	0.76	SSIM	0.96	0.94	0.89	0.82	0.78	0.76	SSIM	0.96	0.94	0.88	0.82	0.78	0.77
Ultrasound video (Stomach-MM)																				
Bitrate 400 Kbps						Bitrate 500 Kbps						Bitrate 600 Kbps								
No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60
PLR (%)	0	1	6	15	22.5	28	PLR (%)	0	1.3	6.8	16.1	23.4	28.7	PLR (%)	0	1.5	9.8	19.1	26.3	31.8
Delay (ms)	12.3	23.8	31.2	34.6	37	39.7	Delay (ms)	12.1	23.8	32.2	36.2	38.5	41.1	Delay (ms)	12.5	26.7	34.3	37.8	40.2	42.5
Total Throughput (Mbps)	4.5	8.8	12.4	14.7	16.6	18.4	Throughput (Mbps)	5.4	10.6	15.1	18	20.3	22.5	Throughput (Mbps)	6.4	12.5	17.1	20.2	22.8	25.1
PSNR (dB)	37.1	35.8	32	25.3	22.4	21.5	PSNR (dB)	37.2	35.9	32.4	26	22.9	21.8	PSNR (dB)	37.3	35.7	31.5	25.6	22.9	21.9
SSIM	0.96	0.95	0.89	0.81	0.77	0.75	SSIM	0.96	0.94	0.87	0.79	0.76	0.75	SSIM	0.97	0.93	0.85	0.78	0.76	0.75
Ultrasound video (Gallbladder-HM)																				
Bitrate 400 Kbps						Bitrate 500 Kbps						Bitrate 600 Kbps								
No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60	No. of users	10	20	30	40	50	60
PLR (%)	0	0.37	10.9	18.3	24.2	30.6	PLR (%)	0	0.43	12.2	20.4	26	31.1	PLR (%)	0	0.46	14.6	24.1	30.8	35.5
Delay (ms)	15.6	30.6	32.3	34.8	37.4	40.5	Delay (ms)	14.4	28.5	30.8	34.1	37.9	41	Delay (ms)	13.8	27.4	31	35.8	40.7	43.5
Total Throughput (Mbps)	4.6	9.2	11.5	13.3	15	16.6	Throughput (Mbps)	5.6	11.3	14.7	17.6	20.3	22.5	Throughput (Mbps)	6.7	13.5	17.9	21.6	25	27.2
PSNR (dB)	36.9	34.7	29	24.2	22.2	21.4	PSNR (dB)	37	35.5	30.6	25	22.6	21.7	PSNR (dB)	37.1	35.7	31	25.1	22.6	21.8
SSIM	0.96	0.94	0.88	0.80	0.77	0.75	SSIM	0.96	0.93	0.86	0.79	0.76	0.75	SSIM	0.96	0.92	0.85	0.78	0.76	0.75

### 3.3. Subjective Tests

The subjective tests carried out in this paper follow the ITU-T Rec.P.910 standard, which is specifically designed for video applications. The ultrasound video sequences varying in content types (i.e., slow motion, medium motion and high motion) were used for evaluation. These ultrasound video sequences were compressed at encoding bitrates of 400 kbps, 500 kbps and 600 kbps and impaired at PLR ranging from 0% to 35%. Before the start of the subjective assessment, brief training was provided to the medical experts regarding what to expect during the assessment process, in order to ensure a seamless running of the test. The original and impaired ultrasound video sequences were shown to 6 medical experts (3 males and 3 females) individually on three different devices (i.e., laptop, tablet and smartphone) in a closed room with natural ambience settings. The Absolute Category Rating-Hidden Reference (ACR-HR) type of subjective assessment method was used to record the MOS, which is a rating chart from 1 to 5 (refer to Table 4). It took the medical experts approximately 20 mins to complete the subjective assessment for 54 test sequences on a single device. A resting break of 10 minutes was given before moving on to the next device in order to reduce the possibility of fatigue as well as to avoid the memorising effect caused by viewing the same content repeatedly. In addition, they were given an option of replaying the ultrasound sequence in case of any ambiguity. The entire assessment process for one medical expert took approximately 90 mins to complete, including three resting breaks; this covered all 54 ultrasound sequences of three content types on three devices. To attain reliable MOS scores, viewing distances of approximately 20 inches, 15 inches and 10 inches were set for the laptop, tablet and smartphone, respectively. All three devices were installed with a VLC media player for the viewing of the ultrasound video sequences. The brightness was set as constant on all the devices, with sleep mode and portrait mode disabled to avoid any sort of interruption.

Following completion of the subjective assessment, we adhered to the ITU-R BT.500-13 recommendations [66] to screen for irregularities in the MOS results. Since the subjective assessment was performed by medical experts, no oddities were found in the MOS results. The MOS results from all the medical experts were averaged for each content type, PLR value, and device as shown in Figures 5–13. The analysis of the obtained MOS results can be found in the results and analysis section.

**Table 4.** Mean opinion score (MOS) for subjective video quality assessment.

Scale	Quality	Description
5	Excellent	Resolution: same as original, smooth and no jitters.
4	Good	Resolution: good, almost same as original, smooth. Very few jitters.
3	Fair	Resolution: good but occasionally bad, image jitters and breaks at periphery but is tolerable as long as region of interest (ROI) is not affected, obvious flow discontinuity of video due to image obstruction.
2	Poor	Resolution: poor, image jitters throughout the clip. ROI is significantly affected.
1	Bad	Resolution: bad, image jitters and breaks for longer intervals in various areas affecting ROI: not acceptable.

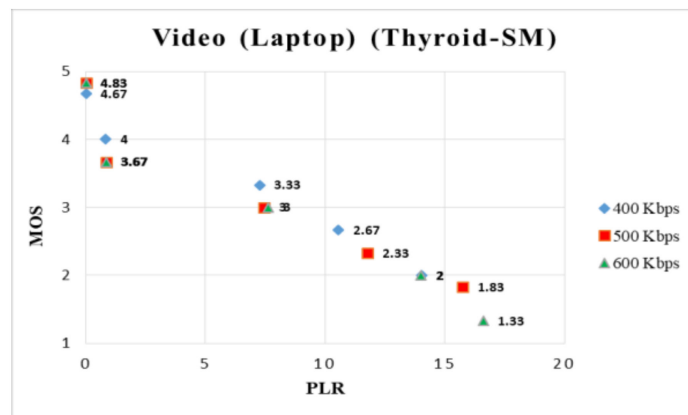


Figure 5. MOS ratings for thyroid ultrasound (SM) on laptop.

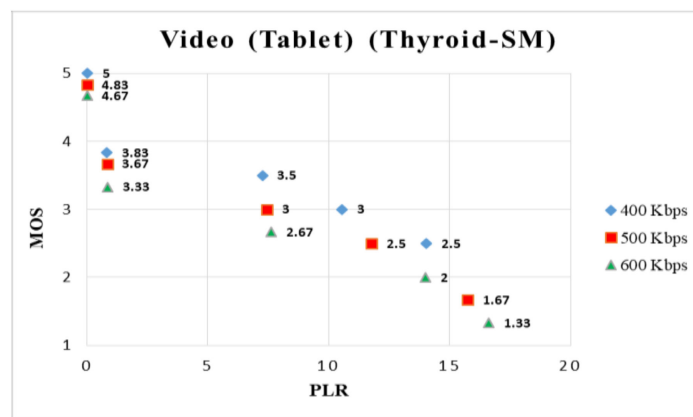


Figure 6. MOS ratings for thyroid ultrasound (SM) on tablet.

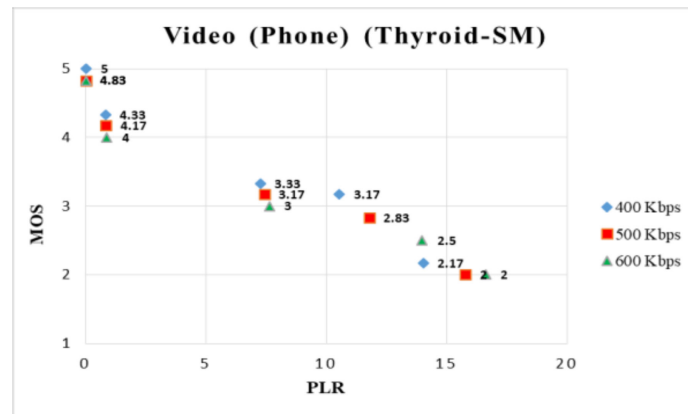


Figure 7. MOS ratings for thyroid ultrasound (SM) on smartphone.

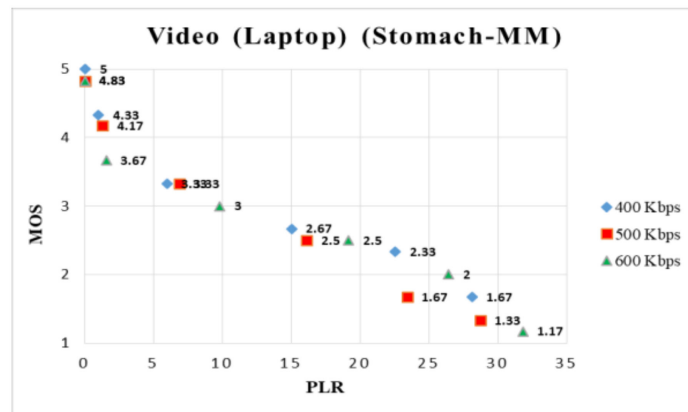


Figure 8. MOS ratings for stomach ultrasound (MM) on laptop.

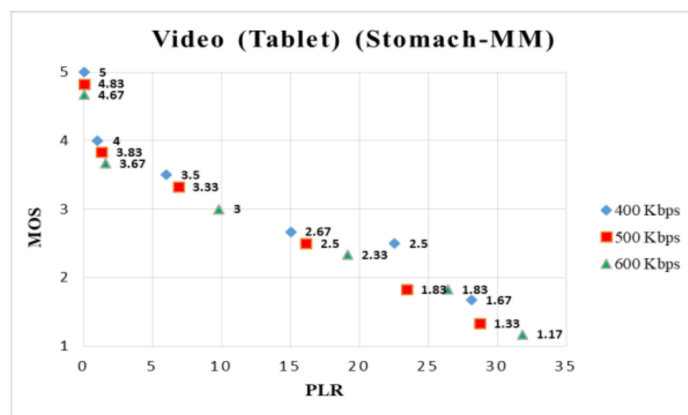


Figure 9. MOS ratings for stomach ultrasound (MM) on tablet.

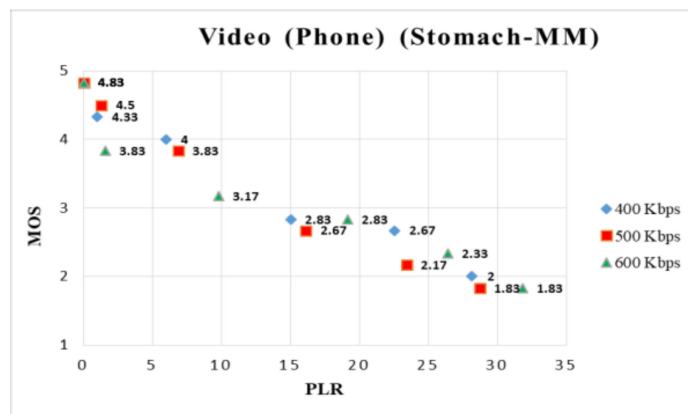


Figure 10. MOS ratings for stomach ultrasound (MM) on smartphone.

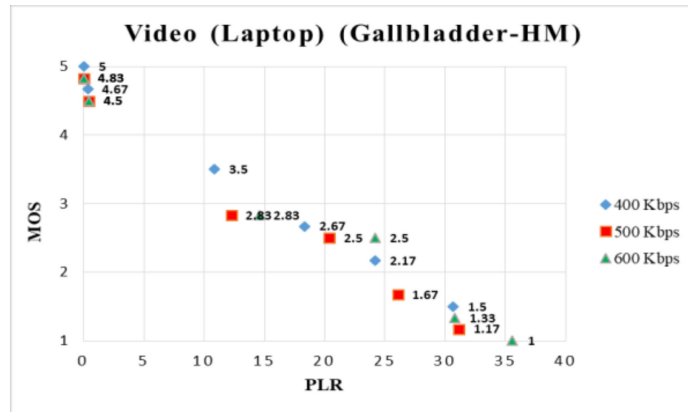


Figure 11. MOS ratings for gallbladder ultrasound (HM) on laptop.

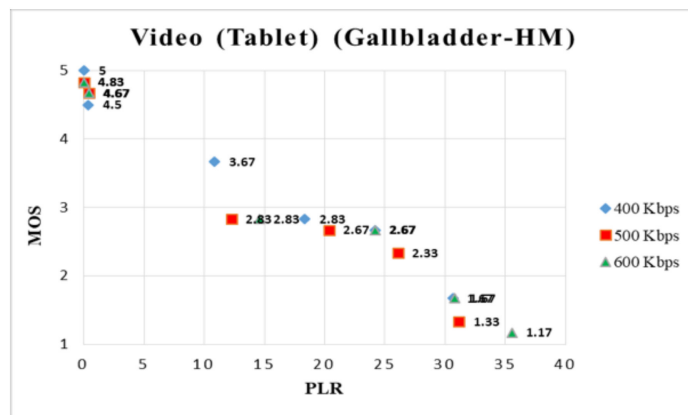


Figure 12. MOS ratings for gallbladder ultrasound (HM) on tablet.

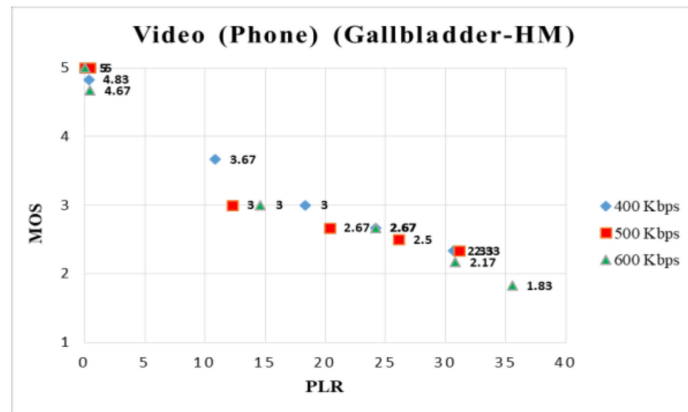


Figure 13. MOS ratings for stomach ultrasound (MM) on smartphone.

### 3.4. Medical QoE Prediction Model: MLP Neural Network Approach

In general, neural networks can be characterised into several categories, depending on the type of application they are used for. In practice, they are commonly utilised in pattern recognition applications, curve fitting problems, approximation of nonlinear functions and self-organising networks [60], to name just a few. The motive behind selecting MLP was its ability to establish a nonlinear relationship between input variables (e.g., m-QoS, content features and display device characteristics) and resultant scores (e.g., m-QoE) as well as its predictive efficiency in solving computational tasks for a large training dataset [67]. The structure of the three-layer MLP neural network-based m-QoE prediction model is depicted in Figure 14.



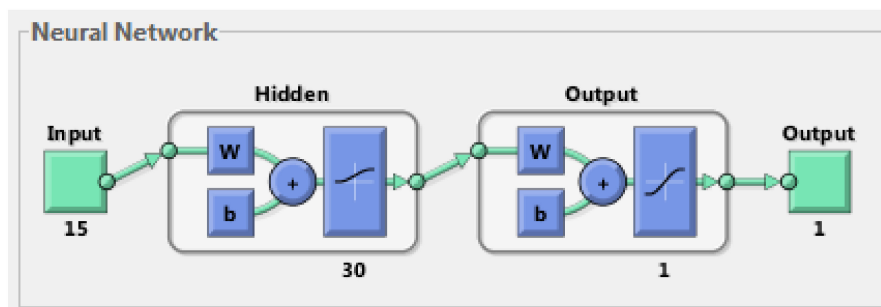


Figure 14. Proposed multilayer perceptron (MLP) neural network framework.

#### 3.4.1. Construction of Feature Vectors

In this research work, the MLP feedforward neural network is formed of three perceptron layers, namely the input layer, the hidden layer and the output layer, which consist of 15 neurons, 30 neurons and one neuron, respectively. The number of neurons in the input layer represents input variables, whereas a single neuron in the output layer represents the predicted output value (i.e., m-QoE). The hidden layer performs computational tasks and is used for cross validation of the prediction model, which is further comprised of two hidden units. One unit is used for training while the other is used for validation. The 15 dimensional input variables and a single output value used in the proposed m-QoE prediction model are given in Table 5.

Table 5. Input and output parameters.

Input Variables		
Device characteristics	Screen size, Resolution	
Content features	Temporal	Slow movement, Medium Movement, High movement
	Spatial	Flat-area, Texture, Fine-texture, Edge
Application QoS	PSNR, SSIM, Bitrate	
Network QoS	PLR, Delay, Throughput	
Output Value		
m-QoE	Predicted MOS	

After forming the feature vectors, the next step is to construct the training dataset; this is achieved by employing an appropriate mapping function, which maps the input variables to the output value. Since the relationship between the input variables and the resultant m-QoE is nonlinear, a nonlinear type of mapping function is chosen. Besides, nonlinear mapping functions are proven to provide better correlation than linear ones. There are several mapping functions available for nonlinear mapping, such as logistic, logarithmic, truncated exponential, power and cubic functions [68]. In our case, we used a truncated exponential mapping function to generate the training dataset, for the reason that this type of nonlinear mapping function is computationally amendable, efficient and less complex than the aforementioned. However, its performance capability is almost the same as that of the others.

#### 3.4.2. Training of MLP Neural Network

Once the training dataset is constructed then the MLP neural network is ready to be trained for m-QoE prediction through nonlinear regression. The training process is carried out for all the possible combinations between the training dataset and their corresponding output value(s). We used half of the dataset for training and the remaining half for validation and testing. The reason behind this division is to avoid overlearning or overfitting (which leads to losing the generalisation capabilities of

the neural network) which can occur when using a previously seen dataset. Therefore, the data used in the training phase should not be used in the validation and/or testing phases of the neural network.

The prediction accuracy of the training phase is analysed by comparing the error values between the subjective m-QoE and the predicted m-QoE. Subsequently, the weights and biases are assigned to ensure maximum performance optimisation of the MLP neural network. In order to find the weights and biases, training algorithms are used, which are gradient-based optimisation algorithms. There is an array of gradient-based training algorithms to choose from, ranging from less complicated gradient descent algorithms to the complex Levenberg–Marquardt algorithm [69]. We selected a scaled conjugate gradient algorithm based on its speed, memory efficiency, and suitability for all sizes of neural networks.

The performance of the training stage is usually measured by the sum of error squares (SSE), mean squared error (MSE), or absolute error (AE). We opted for MSE because it yields a better performance measurement when compared to SSE and AE [67].

The convergence was remarkably fast since it was an offline processing. A total of 13 epochs are recorded, where an epoch represents one complete scan/iteration of the training dataset. For example, in the case of the 54 impaired ultrasound sequences in the training dataset, each epoch represented a full scan of those 54 video sequences. For each epoch, the MLP neural network produced an output (predicted MOS), which is then compared to the subjective MOS results. This comparison produced an error (i.e., MSE), which formed the basis for adjusting the MLP neural network weights to the optimal value. In the subsequent epochs, the order of the training dataset is randomly changed and the process is repeated. A performance metric is obtained for each epoch until the value of MSE is minimised. Figure 15 shows the goodness of fit graphs for the training, testing and validation phases, where the goodness of fit (R) for all three phases is measured at 0.86. This implies that the closer the value of R is to 1, the better the MLP is trained.

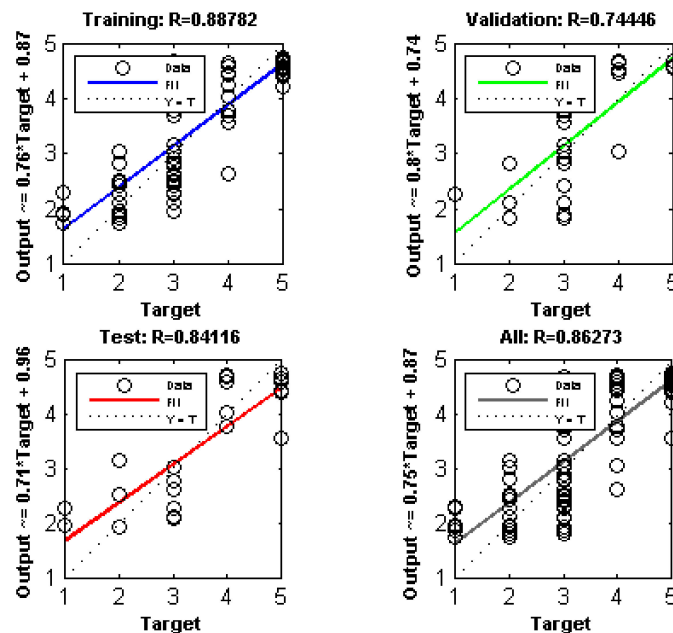


Figure 15. Goodness of fit for training, validation and testing data.

### 3.4.3. Predicted MOS scores

Following the training of the proposed m-QoE model, its prediction accuracy was validated through subjective ratings obtained from medical experts. The trained and validated proposed machine learning model can be considered as a multidimensional function predictor that maps any set of unseen input variables into a quality score or a MOS value for any video sequence.

The correlation graphs in Figures 16–19 represent the subjective MOS and the predicted MOS, for laptop, tablet, smartphone and overall MOS, respectively. The graphs indicate that a high degree of correlation is achieved between predicted and subjective MOS. The prediction accuracy of the proposed prediction model is measured in terms of the correlation coefficient R2 (92.2%) and RMSE (0.109), which implies that the MLP neural network-based m-QoE prediction model succeeds in learning and imitating the visual perception of medical experts.

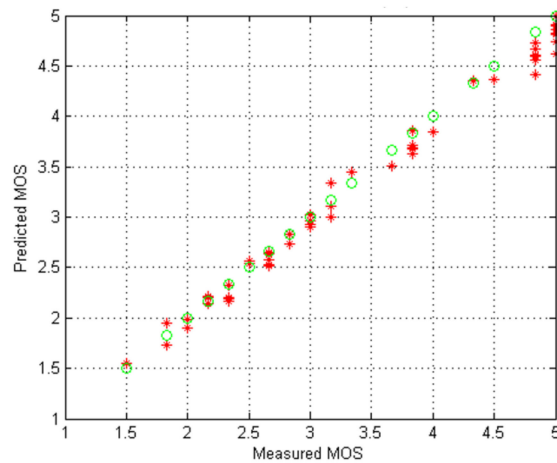


Figure 16. Subjective MOS vs. predicted MOS for laptop.

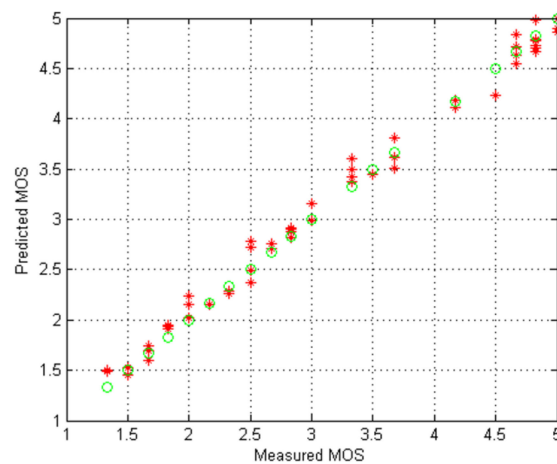


Figure 17. Subjective MOS vs. predicted MOS for tablet.

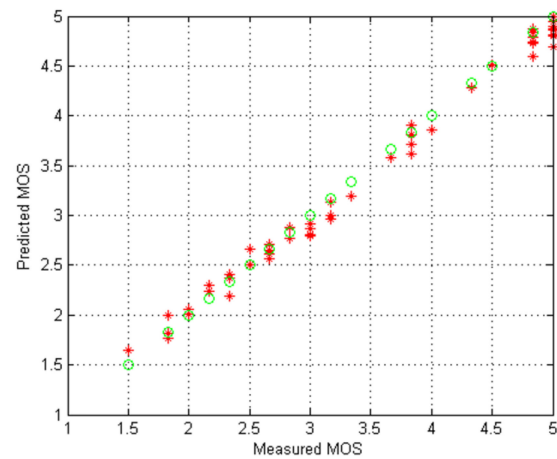


Figure 18. Subjective MOS vs. predicted MOS for smartphone.

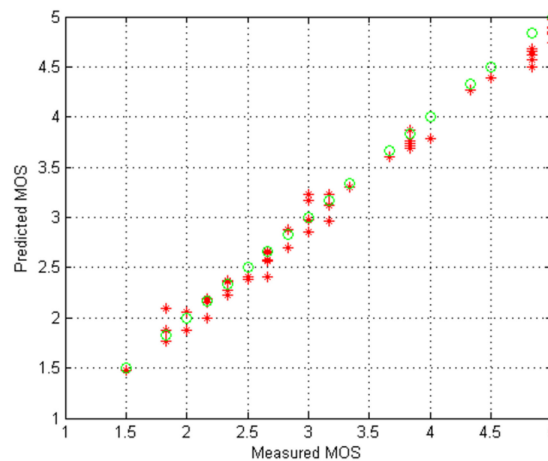


Figure 19. Subjective MOS vs. predicted MOS for all devices.

### 3.5. Device-Aware Adaptive Ultrasound Streaming

Device-aware systems refer to those systems that provide information to the requesting device in a format that is suitable to the device type. In order to achieve this, device-aware systems send information request to the device, which in turn detect the device type based on the feedback information request message. Hence, this helps in adapting the content for appropriate presentation on the detected device type without user interaction, which leads to an enhanced user experience [70].

As shown in Figure 20 below, the architecture of the device-aware system used in this paper consists of three layers:

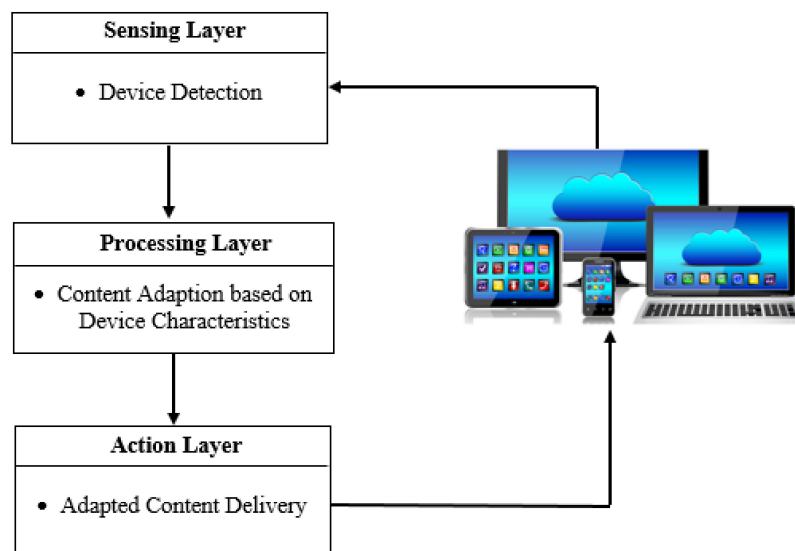


Figure 20. Device-aware system architecture.

#### (1) Sensing layer

The sensing layer acquires information about the end user’s device, such as screen size, device resolution, operating system, etc. This is achieved by sending request message to the device being used. The device responds back through a message, which is fed back into the device-aware system.

#### (2) Processing layer

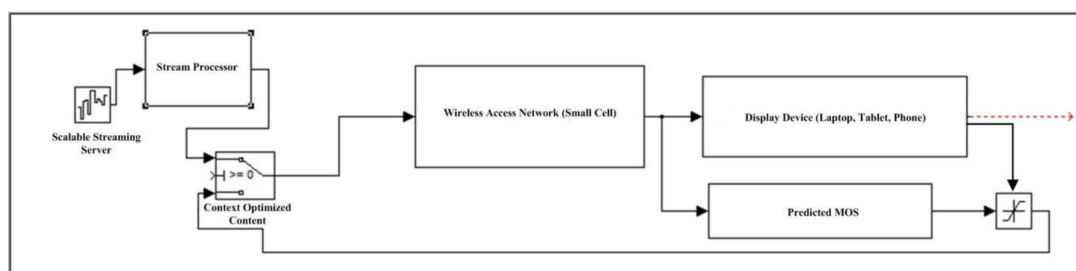
As the name suggests, the processing layer processes all the acquired information and create multiple versions of the video content by encoding and decoding the videos suitable for the device type.

### (3) Action layer

The action layer is the final decision-maker in the hierarchy that assigns the appropriate version of the video content to the requesting device and transmits the video from the adaptive video streaming server to the end user's device.

The aforementioned layers form a cycle, which needs to be repeated every time a new device is detected.

Furthermore, one of the objectives of this work is to exploit the m-QoE prediction model in order to optimise the end-to-end video quality experienced by medical experts. Therefore, we propose an adaptive controller that uses the predicted m-QoE scores at the receiving side as feedback to adapt the content at the transmitting side. To carry out this experiment, we set up a simulation scenario that includes a scalable video streaming server at the hospital, which stores and serves multiple versions of the same ultrasound video content at different encoding bitrates (i.e., 400 kbps, 500 kbps and 600 kbps) for different devices. The adaptive controller monitors the predicted MOS values at predefined time intervals. We used 1 second as the measurement interval, which means that at every 1 second the predicted MOS value, along with the PLR and device characteristics (e.g., screen size and resolution), are obtained. If the predicted MOS value drops lower than the predefined threshold value, which is set at 3.7, then this information is fed back to the scalable video streaming server. The scalable video streaming server changes the same content to a different version with a lower bitrate suitable for the device being used. Therefore, the predicted MOS in the next measurement interval is expected to increase. In this way, the predicted MOS, along with the viewing device information, can optimise the video quality by continuously monitoring the predicted MOS and subsequently changing the content version for either the viewing device or when the MOS falls below the threshold value. Hence, device-aware adaptive streaming can provide a better diagnostic experience to medical experts. Figure 21 shows the adaptive video streaming simulation blocks and Figure 22 shows the MOS values corresponding to the PLR with and without the adaptive mechanism. It can be observed that until time instance 10, the PLR is below 1% and the predicted MOS is 3.8, which is above the threshold value of 3.7, therefore no adaptation is enforced. However, at time instance 10, the PLR reaches 6% and the corresponding predicted MOS value falls to 3.6, which is below the threshold value. Thus, this information triggers the adaptive mechanism and at the next time instance the predicted MOS increases to 3.9.



**Figure 21.** Adaptive video streaming simulation blocks.

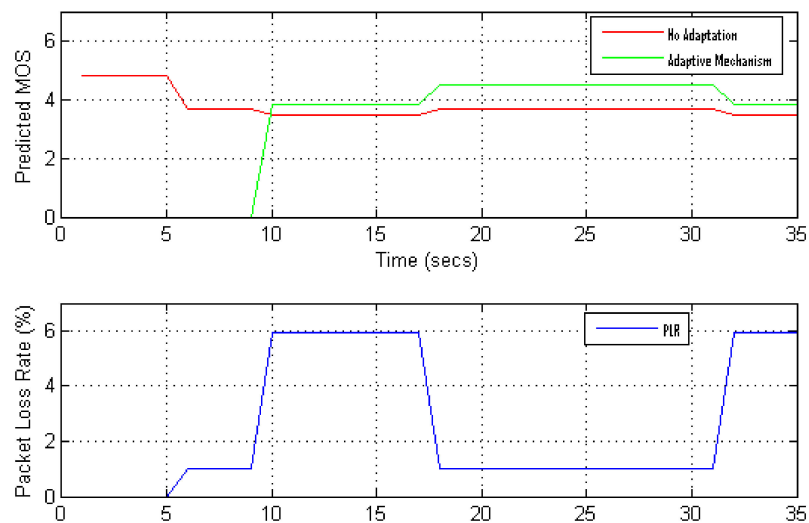


Figure 22. Device-aware adaptive streaming.

#### 4. Results and Analysis

In this section, we present the analysis of the proposed MLP neural network-based m-QoE prediction model over small cell (i.e., femtocell) networks for ultrasound video streaming m-health application.

To study the impact of m-QoS, content classification and the display device characteristics on m-QoE, we created a heterogeneous indoor hospital scenario including a macrocell with several randomly distributed small cells. A simulation on an LTE-Sim system level simulator is carried out for a medical ultrasound video streaming application, increasing the number of both macrocell and small cell users. To replicate a realistic environment, background traffic is generated, where a medical expert is streaming ultrasound video sequences. Three different ultrasound video sequences are selected, varying in content type, namely thyroid, stomach and gallbladder; these are then categorised into classes based on their spatiotemporal features.

From the obtained results, we established that a small cell network, compared to a traditional macrocell network, remarkably improves m-QoS and m-QoE, leading to a more reliable and accurate diagnosis of medical ultrasound videos delivered over error-prone wireless networks. Furthermore, we found that the influential parameters that affect m-QoE are m-QoS, content type and display device characteristics. These therefore need to be considered when designing a robust m-QoE prediction model. The m-QoS analysis showed that the KPIs are quite dependent on the spatiotemporal variations in the content. For instance, the slow-movement (thyroid) ultrasound video sequence gave better m-QoS results, followed by the medium-movement (stomach) video sequence and then the high-movement (gallbladder) video sequence. This is because with higher spatiotemporal complexities, the loss of information could be higher, leading to a greater impact on the QoS. The most significant KPI in the medical video streaming application is the PLR, which highly influences the degree of satisfaction or annoyance in the m-QoE prediction model. In addition, we observed that the MOS increases with the increase in bitrate – that is, when there is no packet loss. However, in our study, increasing bitrates leads to higher PLR values, which in turn lowers the MOS ratings. Furthermore, the subjective assessment by medical experts revealed that the MOS ratings are highly content dependent as the scores given to the slow-movement (thyroid) ultrasound test videos are better than the medium-movement (stomach) followed by the high-movement (gallbladder) sequences. This is because humans tend to overlook minor impairments in low spatiotemporal videos more than in video sequences with high spatiotemporal variations. Moreover, in this study, the influence of display device characteristics (e.g., screen size and screen resolution) on the medical experts' MOS ratings illustrated that the PLR values greater than 8%, 10% and 12% on laptops, tablets and smartphones, respectively, are not clinically favourable for the diagnosis of the small cell ultrasound video streaming scenario.

This implies that the smaller the screen size and screen resolution, the better the MOS results, as we found that the medical experts were more tolerant towards impaired ultrasound video sequences viewed on smartphones and preferred using them for diagnostics over the tablet and laptop for increasing PLR values. Therefore, an m-QoE prediction model must consider the display device characteristics when determining the end-user QoE.

In this paper, in addition to the m-QoE prediction model, we also proposed an adaptive video streaming controller to optimise predicted MOS values. This comprises a scalable video streaming server, which must be aware of the display device being used and must have multiple versions of the video content at different encoding bitrates. We found that decreasing the encoding bitrate resulted in reducing the PLR in the access network. Decreasing the encoding bitrate does not guarantee an excellent viewing experience, but at the same time it does not significantly affect the diagnostic visual quality.

The results showed that the appropriate version of the content transmitted to the receiver based on the feedback received about the predicted MOS and display device features can certainly help to maintain the m-QoE above the minimum diagnostically acceptable threshold value.

In conclusion, the underlying wireless communication platform has a major impact on m-QoS, which is the main influencing factor in predicting m-QoE. We established in [14] that the use of small cell networks is an innovative solution to enhance m-QoS for m-health end-users. Furthermore, when compared to the existing macrocell scenario, a small cell scenario for medical ultrasound video streaming outperforms it in terms of PLR, delay, and throughput. This contrast was also visually noticed in impaired ultrasound sequences, where the PLR values over traditional macrocell networks resulted in constant frame freezing and loss of entire GOPs.

The proposed m-QoE prediction model can achieve a high degree of correlation with subjective testing as long as the appropriate QoE influencing factors are precisely selected. In addition, the proposed adaptive video streaming controller performed well with the network conditions, content, and display device features and adapted predicted m-QoE, maintaining the diagnostic quality.

## 5. Conclusion

In this paper, we have presented a proposed m-QoE prediction model based on MLP neural network for ultrasound video streaming m-health application. The proposed prediction model utilises small cell networks as an end-to-end communication platform and takes into account encoding bitrates, m-QoS parameters for different content types of ultrasound videos and display device characteristics. The proposed model was trained for an unseen dataset and validated through subjective tests (MOS) by medical experts on three types of display devices (i.e., laptop, tablet and smartphone). The prediction accuracy of the proposed model was calculated in terms of the correlation coefficient (R2) and RMSE, which were measured as 92.2% and 0.109, respectively. These results showed that it is possible to achieve high prediction accuracy with the right combination of appropriate parameters. In addition, we found that the prediction accuracy is highly dependent on how well the dataset in the neural network is trained. Furthermore, we proposed a device-aware adaptive streaming controller that uses the MLP neural network predicted MOS, the encoding bitrates and display device characteristics as feedback to improve the end-to-end video quality for m-health end-users. Such an adaptive mechanism is beneficial when the predicted MOS falls below the minimum acceptable value. We can infer that MLP neural network succeeds in predicting m-QoE close to human visual perception without continuous end-user engagement and can be easily adapted in any wireless communication systems, from exiting 4G networks to the future 5G networks. In conclusion, the advent of innovative wireless solutions such as small cells will transform m-health systems, be it any service or application. In the case of medical video streaming applications, the use of small cells offers improved m-QoS, leading towards enhanced m-QoE for m-health users, which complements our objective of proposing an m-QoE framework that not only excels in predicting subjective m-QoE but also optimises predicted m-QoE for diagnostic precision.

**Author Contributions:** The authors of this article have contributed in building this research paper as follows. Conceptualization, I.U.R., M.M.N. and N.Y.P.; Methodology, I.U.R., M.M.N. and N.Y.P.; Software, I.U.R., M.M.N., and N.Y.P.; Validation, I.U.R., M.M.N. and N.Y.P.; Formal analysis, I.U.R., M.M.N. and N.Y.P.; Investigation, I.U.R., M.M.N., and N.Y.P.; Resources, I.U.R., M.M.N. and N.Y.P.; Data Curation, I.U.R., M.M.N. and N.Y.P.; Writing—Original Draft Preparation, I.U.R., M.M.N., and N.Y.P.; Writing—Review and Editing, I.U.R. and N.Y.P.; Visualization, I.U.R., M.M.N. and N.Y.P.; Supervision, N.Y.P.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to thank the radiologists at the Advanced Radiology Centre (ARC), Pakistan, in particular Rashid Ahmad for providing us consultation and these ultrasound videos.

Ikram Ur Rehman would like to acknowledge the CU Coventry for its support. Also, Moustafa Nasralla would like to acknowledge the management of Prince Sultan University (PSU) for the fund, the valued support and the research environmental provision which have led to completing this work. Finally, Nada Y. Philip would like to acknowledge Kingston University for its support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Suzuki, K. Survey of Deep Learning Applications to Medical Image Analysis. *Med. Imaging Technol.* **2017**, *35*, 212–226.
2. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **2017**, *10*, 257–273. [[CrossRef](#)] [[PubMed](#)]
3. Rivera, S.; Riveros, H.; Ariza-Porras, C.; Lozano-Garzon, C.; Donoso, Y. QoS-QoE Correlation Neural Network Modeling for Mobile Internet Services. In Proceedings of the 2013 International Conference on Computing, Management and Telecommunications (ComManTel), Ho Chi Minh City, Vietnam, 21–24 January 2013; pp. 75–80.
4. Lozano-Garzon, C.; Ariza-Porras, C.; Rivera-Diaz, S.; Riveros-Ardila, H.; Donoso, Y. Mobile network QoE-QoS decision making tool for performance optimization in critical web service. *Int. J. Comput. Commun. Control* **2012**, *7*, 892–899. [[CrossRef](#)]
5. Staelens, N.; Moens, S.; van den Broeck, W.; Marien, I.; Vermeulen, B.; Lambert, P.; van de Walle, R.; Demeester, P. Assessing quality of experience of IPTV and video on demand services in real-life environments. *IEEE Trans. Broadcast.* **2010**, *56*, 458–466. [[CrossRef](#)]
6. Razaak, M.; Martini, M.G.; Savino, K. A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1552–1559. [[CrossRef](#)] [[PubMed](#)]
7. Nasralla, M.M.; Razaak, M.; Rehman, I.; Martini, M.G. A Comparative Performance Evaluation of the HEVC Standard with its Predecessor H.264/AVC for Medical videos over 4G and beyond Wireless Networks. In Proceedings of the 8th International Conference on Computer Science and Information Technology (CSIT), Dubai, UAE, 24–25 November 2018; pp. 50–54.
8. Courreges, F.; Vieyres, P.; Istepanian, R.S.H.; Arbeille, P.; Bru, C. Clinical trials and evaluation of a mobile, robotic tele-ultrasound system. *J. Telemed. Telecare* **2005**, *11* (Suppl. 1), 46–49. [[CrossRef](#)] [[PubMed](#)]
9. Malone, F.D.; Athanassiou, A.; Nores, J.; D’Alton, M.E. Effect of ISDN bandwidth on image quality for telemedicine transmission of obstetric ultrasonography. *Telemed. J. Off. J. Am. Telemed. Assoc.* **1998**, *4*, 161–165.
10. Istepanian, R.; Philip, N.; Martini, M. Medical QoS provision based on reinforcement learning in ultrasound streaming over 3.5G wireless systems. *IEEE J. Sel. Areas Commun.* **2009**, *27*, 566–574. [[CrossRef](#)]
11. Cisco. *Cisco Visual Networking Index: Forecast and Methodology, 2014–2019*; Cisco: San Jose, CA, USA, 2015.
12. Rehman, I.U.; Philip, N.Y. M-QoE driven context, content and network aware medical video streaming based on fuzzy logic system over 4G and beyond small cells. In Proceedings of the 2015 EUROCON, Salamanca, Spain, 9 September 2015.
13. Philip, N.Y.; Rehman, I.U. Towards 5G health for medical video streaming over small cells. In Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing, Paphos, Cyprus, 31 March–2 April 2016; pp. 1093–1098.



14. Rehman, I.U.; Philip, N.Y.; Istepanian, R.S.H. Performance analysis of medical video streaming over 4G and beyond small cells for indoor and moving vehicle (ambulance) scenarios. In Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare—“Transforming Healthcare Through Innovations in Mobile and Wireless Technologies”, Athens, Greece, 3–5 November 2014.
15. Chou, S.F.; Chiu, T.C.; Yu, Y.J.; Pang, A.C. Mobile small cell deployment for next generation cellular networks. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014.
16. Altman, Z.; Balageas, C.; Beltran, P.; Ezra, Y.B.; Formet, E.; Hämäläinen, J.; Marcé, O.; Mutafungwa, E.; Perales, S.; Ran, M.; et al. Femtocells: The HOMESNET vision. In Proceedings of the 2010 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, 26–30 September 2010.
17. Quddus, A.; Guo, T.; Shariat, M.; Hunt, B.; Imran, A.; Ko, Y.; Tafazolli, R. Next Generation Femtocells: An Enabler for High Efficiency Multimedia Transmission. *IEEE ComSoc Multimed. Tech. Comm. E-Lett.* **2010**, *5*, 27–31.
18. Andrews, J.G.; Claussen, H.; Dohler, M.; Rangan, S.; Reed, M. Femtocells: Past, present, and future. *IEEE J. Sel. Areas Commun.* **2012**, *30*, 497–508. [[CrossRef](#)]
19. Chen, D.C.; Quek, T.Q.S.; Kountouris, M. Wireless backhaul in small cell networks: Modelling and analysis. In Proceedings of the 2014 79th IEEE Vehicular Technology Conference, Seoul, Korea, 18–21 May 2014.
20. Mutafungwa, E.; Zheng, Z.; Hämäläinen, J.; Husso, M.; Korhonen, T. Exploiting femtocellular networks for emergency telemedicine applications in indoor environments. In Proceedings of the 12th IEEE International Conference on e-Health Networking, Application and Services, Healthcom, Lyon, France, 1–3 July 2010.
21. Maciucă, A.; Popescu, D.; Strutu, M.; Stamatescu, G. Wireless sensor network based on multilevel femtocells for home monitoring. In Proceedings of the 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), Berlin, Germany, 12–14 September 2013; pp. 499–503.
22. Rehman, I.U.; Nasralla, M.M.; Ali, A.; Philip, N. Small Cell-based Ambulance Scenario for Medical Video Streaming: A 5G-health use case. In Proceedings of the 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT), Islamabad, Pakistan, 8–10 October 2018; pp. 29–32.
23. De, D.; Mukherjee, A. Femto-cloud based secure and economic distributed diagnosis and home health care system. *J. Med. Imaging Heal. Inform.* **2015**, *5*. [[CrossRef](#)]
24. Hindia, M.N.; Rahman, T.A.; Ojukwu, H.; Hanafi, E.B.; Fattouh, A. Enabling remote health-caring utilizing IoT concept over LTE-femtocell networks. *PLoS ONE* **2016**, e0155077. [[CrossRef](#)] [[PubMed](#)]
25. Nasralla, M.M.; Razaak, M.; Rehman, I.U.; Martini, M.G. Content-aware packet scheduling strategy for medical ultrasound videos over LTE wireless networks. *Comput. Netw.* **2018**, *140*, 126–137. [[CrossRef](#)]
26. Nasralla, M.M.; Khan, N.; Martini, M.G. Content-aware downlink scheduling for LTE wireless systems: A survey and performance comparison of key approaches. *Comput. Commun.* **2018**, *130*, 78–100. [[CrossRef](#)]
27. Rehman, I.U.; Nasralla, M.M.; Ali, A.; Maduka, I.; Philip, N.Y. The Influence of Content and Device Awareness on QoE for Medical Video Streaming over Small Cells: Subjective and objective quality evaluations. In Proceedings of the IEEE International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (IEEE 3ICT), Sakhir, Bahrain, 18–20 November 2018.
28. Anegekuh, L.; Sun, L.; Jammeh, E.; Mkwawa, I.H.; Ifeachor, E. Content-Based Video Quality Prediction for HEVC Encoded Videos Streamed Over Packet Networks. *IEEE Trans. Multimed.* **2015**, *17*, 1323–1334. [[CrossRef](#)]
29. Pokhrel, J.; Wehbi, B.; Morais, A.; Cavalli, A.; Allilaire, E. Estimation of QoE of video traffic using a fuzzy expert system. In Proceedings of the IEEE 10th Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2013; pp. 224–229.
30. Kim, H.J.; Choi, S.G. A study on a QoS/QoE correlation model for QoE evaluation on IPTV service. In Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT), Phoenix Park, Korea, 7–10 February 2010; Volume 2, pp. 1377–1382.
31. Cherif, W.; Ksentini, A.; Negru, D.; Sidibé, M. A-PSQA: Efficient real-time video streaming QoE tool in a future media internet context. In Proceedings of the IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011.
32. Kim, H.-J.; Yun, D.-G.; Kim, H.-S.; Cho, K.-S.; Choi, S.-G. QoE assessment model for video streaming service using QoS parameters in wired-wireless network. In Proceedings of the 2012 14th International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea, 19–22 February 2012; pp. 459–464.

33. Song, W.; Tjondronegoro, D.W. Acceptability-based QoE models for mobile video. *IEEE Trans. Multimed.* **2014**, *16*, 738–750. [[CrossRef](#)]
34. Khan, A.; Sun, L.; Ifeachor, E.; Fajardo, J.O.; Liberal, F. Video Quality Prediction Model for H.264 Video over UMTS Networks and Their Application in Mobile Video Streaming. In Proceedings of the 2010 IEEE International Conference on Communications, Cape Town, South Africa, 23–27 May 2010; pp. 1–5.
35. Chen, H.; Xin, Y.; Xie, L. End-to-end quality adaptation scheme based on QoE prediction for video streaming service in LTE networks. In Proceedings of the 11th International Symposium on Modeling Optimization in Mobile, Ad Hoc Wireless Networks (WiOpt), Tsukuba Science City, Japan, 13–17 May 2013; pp. 627–633.
36. Alreshoodi, M.; Woods, J. QoE prediction model based on fuzzy logic system for different video contents. In Proceedings of the UKSim-AMSS 7th European Modelling Symposium on Computer Modelling and Simulation, Manchester, UK, 20–22 November 2013; pp. 635–639.
37. Khan, A.; Sun, L.; Ifeachor, E. Content Clustering Based Video Quality Prediction Model for MPEG4 Video Streaming over Wireless Networks. In Proceedings of the 2009 IEEE International Conference on Communications, Dresden, Germany, 14–18 June 2009; pp. 1–5.
38. Ries, M.; Nemethova, O.; Rupp, M. Motion based reference-free quality estimation for H.264/AVC video streaming. In Proceedings of the 2007 2nd International Symposium on Wireless Pervasive Computing San Juan, Puerto Rico, 5–7 February 2007; pp. 355–359.
39. Staelens, N.; Van Wallendael, G.; Crombecq, K.; Vercammen, N.; De Cock, J.; Vermeulen, B.; Dhaene, T.; Demeester, P.; van de Walle, R. No-reference bitstream-based visual quality impairment detection for high definition H.264/AVC encoded video sequences. *IEEE Trans. Broadcast.* **2012**, *58*, 187–199. [[CrossRef](#)]
40. Seshadrinathan, K.; Bovik, A.C. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. Image Process.* **2010**, *19*, 335–350. [[CrossRef](#)]
41. Ries, M.; Crespi, C.; Nemethova, O.; Rupp, M. Content Based Video Quality Estimation for H.264/AVC Video Streaming. In Proceedings of the 2007 IEEE Wireless Communications and Networking Conference, Kowloon, China, 11–15 March 2007; pp. 2668–2673.
42. Joskowicz, J.; Sotelo, R.; Ardao, J.C.L. Towards a general parametric model for perceptual video quality estimation. *IEEE Trans. Broadcast.* **2013**, *59*, 569–579. [[CrossRef](#)]
43. Khan, A.; Sun, L.; Ifeachor, E. QoE prediction model and its application in video quality adaptation over UMTS networks. *IEEE Trans. Multimed.* **2012**, *14*, 431–442. [[CrossRef](#)]
44. Ries, M.; Nemethova, O.; Rupp, M. Video Quality Estimation for Mobile H. 264/AVC Video Streaming. *J. Commun.* **2008**, *3*, 41–50. [[CrossRef](#)]
45. Nightingale, J.; Wang, Q.; Grecos, C. HEVStream: A framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks. *IEEE Trans. Consum. Electron.* **2012**, *58*, 404–412. [[CrossRef](#)]
46. López, J.P.; Slanina, M.; Arnaiz, L.; Menéndez, J.M. Subjective quality assessment in scalable video for measuring impact over device adaptation. In Proceedings of the IEEE EuroCon, Zagreb, Croatia, 1–4 July 2013.
47. Qualinet. *Qualinet White Paper on Definitions of Quality of Experience Output from the Fifth Qualinet Meeting*; European Network on Quality of Experience in Multimedia Systems and Services: Lausanne, Switzerland, 2013.
48. Sidaty, N.O.; Larabi, M.C.; Saadane, A. Influence of video resolution, viewing device and audio quality on perceived multimedia quality for steaming applications. In Proceedings of the 2014 5th European Workshop on Visual Information Processing (EUVIP), Paris, France, 10–12 December 2014; pp. 1–6.
49. Zhao, T.; Liu, Q.; Chen, C.W. QoE in Video Transmission: A User Experience-Driven Strategy. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 285–302. [[CrossRef](#)]
50. Vucic, D.; Skorin-Kapov, L. The impact of mobile device factors on QoE for multi-party video conferencing via WebRTC. In Proceedings of the 13th International Conference on Telecommunications, Graz, Austria, 13–15 July 2015.
51. Cermak, G.; Pinson, M.; Wolf, S. The relationship among video quality, screen resolution, and bit rate. *IEEE Trans. Broadcast.* **2011**, *57* (Part 1), 258–262. [[CrossRef](#)]
52. Agboma, F.; Liotta, A. Addressing user expectations in mobile content delivery. *Mob. Inf. Syst.* **2007**, *3*, 153–164. [[CrossRef](#)]

53. Khan, A.; Sun, L.; Fajardo, J.O.; Taboada, I.; Liberal, F.; Ifeachor, E. Impact of end devices on subjective video quality assessment for QCIF video sequences. In Proceedings of the 2011 Third International Workshop on Quality of Multimedia Experience, Mechelen, Belgium, 7–9 September 2011; Volume 9, pp. 177–182.
54. Rodríguez, D.; Rosa, R.; Costa, E.; Abrahão, J.; Bressan, G. Video quality assessment in video streaming services considering user preference for video content. *IEEE Trans. Consum. Electron.* **2014**, *60*, 436–444. [[CrossRef](#)]
55. Rehman, A.; Zeng, K.; Wang, Z. Display device-adapted video quality-of-experience assessment. *Proc. SPIE* **2015**, *9394*, 939406.
56. Hu, J.; Lai, Y.; Peng, A.; Hong, X.; Shi, J. Proactive Content Delivery with Service-Tier Awareness and User Demand Prediction. *Electronics* **2019**, *1*, 50. [[CrossRef](#)]
57. Mushtaq, M.S.; Augustin, B.; Mellouk, A. Empirical study based on machine learning approach to assess the QoS/QoE correlation. In Proceedings of the 17th European Conference on Network and Optical Communications, Vilanova i la Geltru, Spain, 20–22 June 2012.
58. Pitas, C.N.; Panagopoulos, A.D.; Constantinou, P. Quality of consumer experience data mining for mobile multimedia communication networks: Learning from measurements campaign. *Int. J. Wirel. Mob. Comput.* **2015**, *8*. [[CrossRef](#)]
59. Pitas, C.N.; Charilas, D.E.; Panagopoulos, A.D.; Constantinou, P. Adaptive neuro-fuzzy inference models for speech and video quality prediction in real-world mobile communication networks. *IEEE Wirel. Commun.* **2013**, *20*, 80–88. [[CrossRef](#)]
60. Gelenbe, E. *Neural Networks: Advances and Applications*; Elsevier: Amsterdam, The Netherlands, 2014.
61. Rehman, I.U.; Philip, N.Y.; Nasralla, M.M. A hybrid quality evaluation approach based on fuzzy inference system for medical video streaming over small cell technology. In Proceedings of the IEEE 18th International Conference on e-Health Networking, Applications and Services, Healthcom, Munich, Germany, 14–17 September 2016.
62. Xu, C.; Zhang, P. Video Streaming in Content Centric Mobile Networks: Challenges and Solutions. *IEEE Wirel. Commun.* **2017**, *1*, 2–10. [[CrossRef](#)]
63. Piro, G.; Grieco, L.A.; Boggia, G.; Capozzi, F.; Camarda, P. Simulating LTE cellular systems: An open-source framework. *IEEE Trans. Veh. Technol.* **2011**, *60*, 498–513. [[CrossRef](#)]
64. Capozzi, F.; Piro, G.; Grieco, L. A system-level simulation framework for LTE femtocells. In Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques, Desenzano del Garda, Italy, 19–23 March 2012; pp. 6–8.
65. Nasralla, M.M.; Rehman, I. QCI and QoS Aware Downlink Packet Scheduling Algorithms for Multi-Traffic Classes over 4G and beyond Wireless Networks. In Proceedings of the IEEE International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (IEEE 3ICT), Sakhir, Bahrain, 18–20 November 2018.
66. The ITU Radiocommunication Assembly. *Methodology for the Subjective Assessment of the Quality of Television Pictures*; No. BT. 500-11; ITU: Geneva, Switzerland, 2002; pp. 1–48.
67. Pierucci, L.; Micheli, D. A Neural Network for Quality of Experience Estimation in Mobile Communications. *IEEE Multimed.* **2016**, *23*, 42–49. [[CrossRef](#)]
68. Heikki, N.K. NEURAL NETWORKS: Basics Using MATLAB Neural Network Toolbox. 2008. Available online: [http://staff.ttu.ee/~jjmajak/Neural\\_networks\\_basics\\_.pdf](http://staff.ttu.ee/~jjmajak/Neural_networks_basics_.pdf) (accessed on 10 October 2018).
69. Demuth, H.; Beale, M.; Hagan, M.; Chen, Q. Neural Network Toolbox™ 6. 2008. Available online: [http://cda.psych.uiuc.edu/matlab\\_pdf/nnet.pdf](http://cda.psych.uiuc.edu/matlab_pdf/nnet.pdf) (accessed on 10 October 2018).
70. Jiang, H.; Lee, H.C.; Basu, K.K.; Yu, C. Device aware internet portal. U.S. Patent No. 6741853, 25 May 2004.

