

Mean Opinion Score Estimation for Mobile Broadband Networks Using Bayesian Networks

Ayman A. El-Saleh¹, Abdulraqueb Alhammadi^{2,*}, Ibraheem Shayea³, Azizul Azizan⁴ and Wan Haslina Hassan²

¹College of Engineering, A'Sharqiyah University (ASU), Ibra, 400, Oman

²Communication Systems and Networks Research Lab, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, 54100, Malaysia

³Department of Electronics and Communication Engineering, Faculty of Electrical and Electronics Engineering, Istanbul Technical University (ITU), Istanbul, 34467, Turkey

⁴Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, 54100, Malaysia

*Corresponding Author: Abdulraqueb Alhammadi. Email: abdulraqueb.alhammadi@utm.my

Received: 25 October 2021; Accepted: 25 January 2022

Abstract: Mobile broadband (MBB) networks are expanding rapidly to deliver higher data speeds. The fifth-generation cellular network promises enhanced-MBB with high-speed data rates, low power connectivity, and ultra-low latency video streaming. However, existing cellular networks are unable to perform well due to high latency and low bandwidth, which degrades the performance of various applications. As a result, monitoring and evaluation of the performance of these network-supported services is critical. Mobile network providers optimize and monitor their network performance to ensure the highest quality of service to their end-users. This paper proposes a Bayesian model to estimate the minimum opinion score (MOS) of video streaming services for any particular cellular network. The MOS is the most commonly used metric to assess the quality of experience. The proposed Bayesian model consists of several input data, namely, round-trip time, stalling load, and bite rates. It was examined and evaluated using several test data sizes with various performance metrics. Simulation results show the proposed Bayesian network achieved higher accuracy overall test data sizes than a neural network. The proposed Bayesian network obtained a remarkable overall accuracy of 90.36% and outperformed the neural network.

Keywords: Quality of experience; quality of service; bayesian networks; minimum opinion score; artificial intelligence; prediction; mobile broadband

1 Introduction

As support for high-speed internet access, mobile broadband (MBB) networks have grown very quickly. Data demand has risen rapidly because of the large number of users using various mobile



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

technology and Internet services to access data. Hence, to enhance the quality of service, it is highly necessary to monitor network performance. Performance and quality assessment are becoming increasingly critical for mobile network operators with the rising demand on current networks. The current fourth-generation (4G) networks are insufficient for bandwidth-hungry and low latency applications. Enhanced MBB (eMBB) is the first of the three main categories in fifth-generation (5G), extending to existing 4G networks. The 5G eMBB offers high-speed data, low latency video streaming, and seamless mobility. Video streaming is the most used service by subscription users. It plays the main role in evolving the quality of services (QoS) and user experience for the MBB providers. Thus, monitoring network performance is very important to guarantee users satisfaction. Different network infrastructures support various video services. For example, video over Internet protocol (also known as AV-over-IP) is a technology that delivers video service over a conventional cable network infrastructure. In contrast, video over LTE (ViLTE) is an enhanced technology that supports voice services with a high-quality video channel over LTE network infrastructure. These services are improved continuously by mobile service providers to increase the number of satisfied users in cellular networks.

International telecommunication union (ITU) introduced a new term called mean opinion score (MOS), which assesses mobile network providers by evaluating the end-users' satisfaction by obtaining users opinions of a network's performance [1]. The MOS metric is widely used for several services and applications such as audio, audiovisual, and video. These services have various methods to assist in scoring their service quality [2]. It also includes entertainment services such as conversational, audiovisual, listening, talking, and video. Several MOS levels have been defined in ITU-T P.800.1, which can be categorized into five levels based on the network quality [2]. Various methodologies have also been used to examine network services, including audio, audiovisual, and video. While the use of "reference quality indicator" and its general acceptance has an obvious advantage, MOS is frequently used without enough assessment of its scope or limits. In [3], the common issues with various types of MOS are highlighted and a variety of alternative methods for media quality measurement are discussed. Numerous works have investigated and evaluated various types of service over wireless networks. In [4], the authors evaluated the perceived quality of the ViLTE service by performing an experimental testbed realized at a mobile provider. Several network QoS parameters such as packet loss and packet delay variation are utilized to predict differential MOS of the ViLTE services. In [5], the authors proposed a prediction model based on data collected from 3G and 4G networks to estimate the quality of experience (QoE) of web browsing and voice services. The model aimed to estimate the MOS scale based on the user-perceived web and video qualities. In [6], a simplified E-model based on a subjective MOS prediction model is proposed. This model aimed to enhance the objective measurement tool for specific codecs. In [7], random neural networks are utilized to objectively predict video quality, consisting of a three-layer feed-forward with a gradient descent training algorithm. The predicted MOS achieved accuracy with approximately 50% compared to other models. However, most proposed models deal with voice and web browsing services that are insufficient to evaluate the network performance. Video streaming and data become very important in people's daily life than voice because of the increase in smartphone video applications that require more mobile video traffic. Thus, more studies are required to evaluate video streaming performance over various MBB networks. Several optimization techniques, such as artificial intelligence (AI) can be used to assist mobile services providers to monitor and optimize their network performance. Currently, AI has become very popular and involved in many different technologies. It is considered one of the most promising solutions in the 5G networks that have been used to mimic human intelligence to support and improve a wide variety of applications. Bayesian networks are a combination of AI and several theories, such as a

graph, decision, and probability that uses Bayesian inference for probability computations. Bayesian inference is used in many studies such as [8,9] that help to infer missing or unknown data.

In this paper, we propose a Bayesian graphical model to evaluate MBB network performance by estimating the MOS from a dataset. The quality of video streaming service is considered, which depends on several parameters such as stalling time, end-to-end latency, downlink bite rate. The estimated MOS indicates service quality users' satisfaction, which helps mobile network operators maintain their network quality to satisfy users' requirements. The remainder of this paper is organized as follows. Section 2 provides a background on quality models used to determine services and applications' performance. Section 3 explains in detail the proposed Bayesian graphical model and procedure and performance metrics. Section 4 analyses and evaluates the performance of the proposed model. Section 5 concludes the paper.

2 Background

In this section, the background on the quality models and Bayesian networks are discussed. In the quality models, we present various quality models to determine or estimate a network performance based on the end-users preceptive. In the Bayesian networks, we introduce usage Bayesian networks through an experimental perspective on probabilistic reasoning to predict an unknown particulate variable.

2.1 Quality Models

Numerous services and applications, such as video streaming and real-time gaming suffer from buffering, dropped frames, and additional delay because of end-to-end transmission and video encoding and decoding. Moreover, some network constraints, such as limited bandwidth, network congestion, and packet loss can negatively affect the end user's QoE. Thus, several strategies for resource and network management are required to optimize user satisfaction. Various system configurations should consider the subjective user ratings to optimize the network performance.

ITU provides a recommendation of various quality models to determine or estimate network performance based on the end-users' preceptive. This recommendation specifies the terminologies to be used in conjunction with quality expressions of audio, audiovisual, and video in terms of MOS. Several identifiers are used to observe the MOS rate for quality of video (VMOS) service provided by mobile operators, such as the subjective, objective, and estimated, which are referred to as S-VMOS, O-VMOS, and E-VMOS, respectively [2]. In addition, various factors also influence the VMOS value application, viewing device display size, and display resolution [ITU-T P.800.2]. The S-VMOS refers to subjective video quality. It is based on the collected VMOS value laboratory test where it is determined by the arithmetic mean value of subjective judgments. Subjective tests were carried out according to obtain results in terms of MOS-VQS [10,11]. O-VMOS refers to the objective video quality that depends on an algorithmic quality real-time evaluation model for determining the MOS score. The real-time evaluation model utilizes various real-time objective metrics obtained from information carried in the video streams and corresponding networks. ITU presented several models to investigate bitstream assessment of video media streaming quality [12,13].

E-VMOS refers to the estimated video quality calculated by a network planning model. It aims to predict video quality over a network in non-real-time parameters. The E-VMOS is used mostly where the MOS values are not determined in real-time models. Some non-real-time models have been defined in [14,15]. Fig. 1 shows the relationship between VMOS identifiers. For the subjective model, the real-time parameters that come from the system feed the subjective test directly to determine the

S-VMOS. In the objective model, the value of S-VMOS depends on several factors, such as real-time video streams and system and calibration from the subjective test. However, in E-VMOS, video test library and non-real-time parameters feed the network planning model to predict the value of E-VMOS. However, these methods present a general concept of MOS models for various services. The S-VMOS and O-VMOS are required for real-time parameters, which lead to complicated processes and high costs, whereas E-VMOS requires only a sufficient database that can be collected at a specific time in the desired area for testing the network performance.

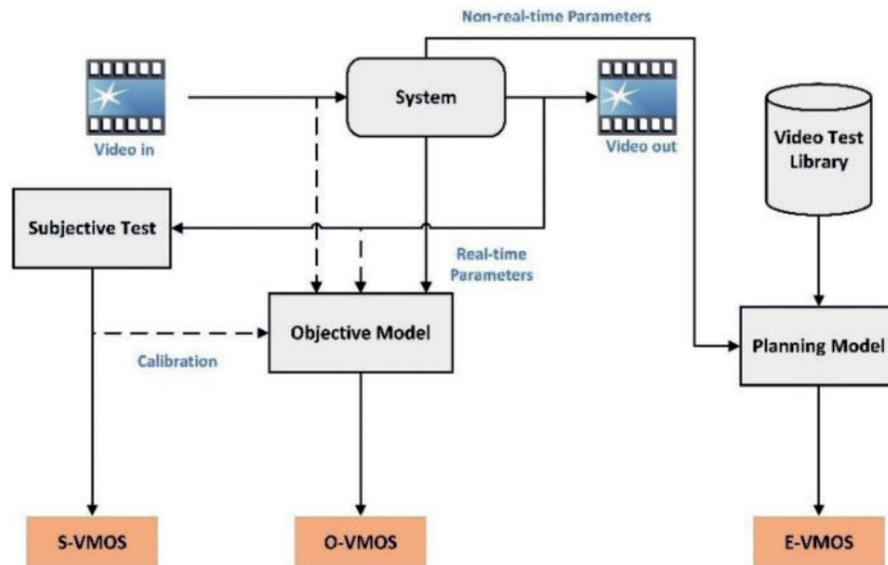


Figure 1: Relationship between VMOS identifiers [2]

2.2 Bayesian Networks

Bayesian networks are the method of choice for explainable reasoning in AI. This work aims to introduce a prediction model based on Bayesian networks through an experimental perspective on probabilistic reasoning. Bayesian networks are probabilistic graphical models that represent joint probability distributions of a set of random variables. The networks measure the conditional dependence structure of these variables according to Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

where $P(A|B)$ represents the posterior probability of an event A given prior knowledge of a condition B. $P(B|A)$ is a conditional probability of event B given prior knowledge of a condition A. $P(A)$ and $P(B)$ are the unconditional probabilities of observing events A and B.

The graphical model of the Bayesian network consists of several nodes N (also known as vertices) $N = \{x_1, x_2, x_3, \dots, x_n\}$ and connected through a set A arcs where $A = \{a_{yx_1}, a_{yx_2}, a_{yx_3}, \dots, a_{yx_n}\}$ (also known as an edge) as shown in Fig. 2.

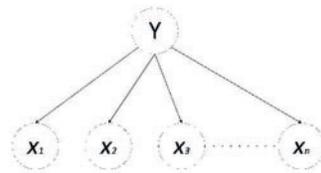


Figure 2: Basic concept of the Bayesian graphical model

Priors $P(Y)$ and conditionals $P(X_i|Y)$ for Naïve Bayes where it can be defined as a joint distribution:

$$P(x_1, x_2, x_3, \dots, x_n) = \prod_{i=1}^n P(x_i|Y). \tag{2}$$

Bayesian network is a type of network that contains data on the causal probability correlations between variables and is frequently used to help in decision making. The Bayes theorem usually updates the causal probability, where the Bayesian nodes indicate the inter-variable dependency structure and the conditional relationships depict the directed arcs in the form of a directed acyclic graph.

The conditional probabilities of $p(z)$ represented by directed acyclic graph can be mathematically expressed as follows:

$$P(Z) = \sum_{X,Y,H,F} P(X, Y, H, F), \tag{3}$$

$$P(Z) = \sum_{X,Y,H,F} P(Z|X, Y)P(X|H)P(H)P(Y|F)P(F). \tag{4}$$

The conditional probability of $p(z)$ represents the query stored in the graphical network in Fig. 3. Eq. (5) can be written efficiently in terms of summation into four separate summations, one over each variable. It requires enumerating all possible combinations of assignments to X, Y, H, and F, and then multiplying the factors for each node.

$$P(Z) = \sum_X \sum_Y \sum_H \sum_F P(Z|X, Y)P(X|H)P(H)P(Y|F)P(F) \tag{5}$$

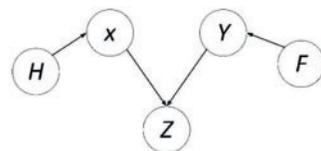


Figure 3: Directed acyclic graph

The variable elimination on Fig. 3 and Eq. (5) uses elimination order X, Y, H, F. The variable F can be eliminated by gathering the similar variable F (in this case $P(Y|F)P(F)$) of the two terms. The

product of this combination can be given in terms of factor f_1 of Y as follows:

$$\begin{aligned}
 P(Z) &= \sum_X \sum_Y \sum_H P(Z|X, Y)P(X|H)P(H) \underbrace{\sum_F P(Y|F)P(F)}_{f_1(y)} \\
 &= \sum_X \sum_Y \sum_H P(Z|X, Y)P(X|H)P(H)f_1(y)
 \end{aligned} \tag{6}$$

Similarly, the next variable H also can be eliminated because it has a similar term that can be computed in terms of factor f_2 of X as follows:

$$P(Z) = \sum_X \sum_Y P(Z|X, Y)f_1(y) \underbrace{\sum_H P(X|H)P(H)}_{f_2(x)} = \sum_X \sum_Y P(Z|X, Y)f_1(y)f_2(x) \tag{7}$$

Finally, the probability of variable Z is simplified as presented in Eq. (7) over variables X and Y.

3 Proposed Model and Performance Metrics

3.1 Proposed Model

This section presents the proposed Bayesian graphical model, which consists of several nodes and directed edges. This study is an extension of our previous work presented in [16]. The proposed model aims to estimate the MOS of video services provided by MBB networks. The estimation process of video quality requires non-real-time parameters which can be obtained from data measurements. Fig. 4 illustrates the proposed model, which comprises seven primary nodes. Three input data nodes, namely, stalling load, SL_i , round trip time, RT_i and bite rate, BR_i , three MOS nodes stalling MOS SM_i , latency MOS, LM_i and quality MOS, QM_i , and the estimated video MOS node MOS_i , where i represents the number of test data. The proposed model then learns and trains several features of a dataset obtained from the drive test in [17] to estimate the MOS_i node.

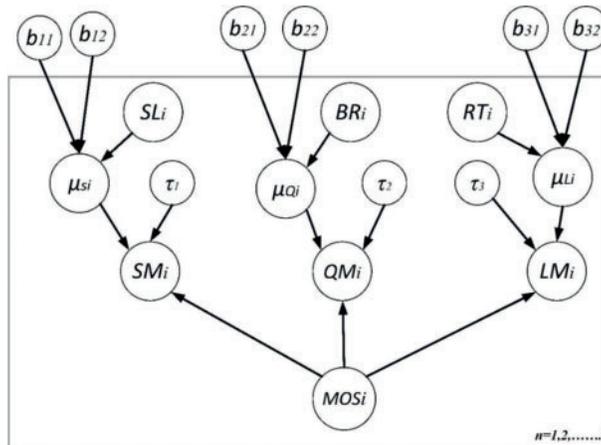


Figure 4: Proposed Bayesian network model for estimating MOS

Bayesian inference using the Gibbs Sampler (OpenBUGS) is a visual tool used to develop various graphical models [18]. It is an open-source software that assists in designing and simulating different Bayesian graphical models. The proposed model was built by utilizing the features of OpenBUGS, which estimate the probability distribution of unknown node-based Gibbs sampler. First three nodes

QM_i , SM_i , and LM_i related to the inputs data that represent the MOS of each input data and are defined as normal distributions with mean μ_{Si} , μ_{Qi} , and μ_{Li} and variance τ_1 , τ_2 , and τ_3 , respectively.

$$SM_i \sim N(\mu_{Si}, \tau_1) \quad (8a)$$

$$QM_i \sim N(\mu_{Qi}, \tau_2) \quad (8b)$$

$$LM_i \sim N(\mu_{Li}, \tau_3) \quad (8c)$$

where μ_{Si} , μ_{Qi} and μ_{Li} are computed as follows:

$$\mu_{Si} = b_{11} + b_{12}SL_i, \quad (9a)$$

$$\mu_{Qi} = b_{21} + b_{22}BR_i, \quad (9b)$$

$$\mu_{Li} = b_{31} + b_{32}RT_i, \quad (9c)$$

where SL_i , RT_i and BR_i refer to stalling, round trip time, and load bite rates, respectively. They are distributed uniformly in a defined range between minimum and maximum values as follows:

$$BR_i, SL_i, RT_i \sim Uniform(min, max) \quad (10)$$

The initial values (b_{11} , b_{12} , b_{21} , b_{22} , b_{31} , and b_{32}) were used to burn in for the first initial iterations at the beginning of the unknown variable inference. All the initial values generated samples during the burn-in period are discarded, except for the final state, which becomes the starting point for sampling. They are defined as a normal distribution with b_{kl} carry any random values:

$$b \sim N(b_{kl}, \tau_k), b_{kl} \sim N(0.001), \tau_k \sim \gamma(0.001, 0.001) \quad for \quad k = 1, 2, 3 ; l = 1, 2. \quad (11)$$

Then, the Bayesian model follows the Bayes' theorem to infer unknown nodes, which generates a posterior probability distribution based on the prior probability distribution and the likelihood function of the data. Thus, the conditional probability of the posterior distribution is given by the product of the prior and likelihood function.

$$posterior = prior \times likelihood$$

The framework of the proposed prediction model treated in this work is illustrated in Fig. 5. The framework consists of two main stages: real-time objective MOS and non-real-time estimated MOS. The former refers to the raw database collected from different mobile operators where the MOS is determined by an algorithmic quality real-time evaluation model. Meanwhile, the latter utilizes the input video parameters to estimate the MOS over a non-real-time network. These input parameters are considered necessary in QoE and QoS planning; thus, they feed the proposed Bayesian model to predict the MOS values.

The sampling distribution is based on the Gibbs sampler, which is one of the Markov chain Monte Carlo (MCMC) algorithms. The Gibbs sampling algorithm was employed to draw samples from the high dimensional joint distribution [19]. It draws a large number of samples denoted with $s_k^{(i+1)}$ from the given initial value $s = s^{(0)}$ and keep repeating the drawing samples k times. Eventually, these samples approximate the joint distribution of all variables in the proposed model and infer the unknown variable, MOS. Eqs. (12a)–(12e) describe the steps of the Gibbs sampling algorithm for each drawn sample. It proceeds as follows for the linear regression where all parameters initialize with starting values.

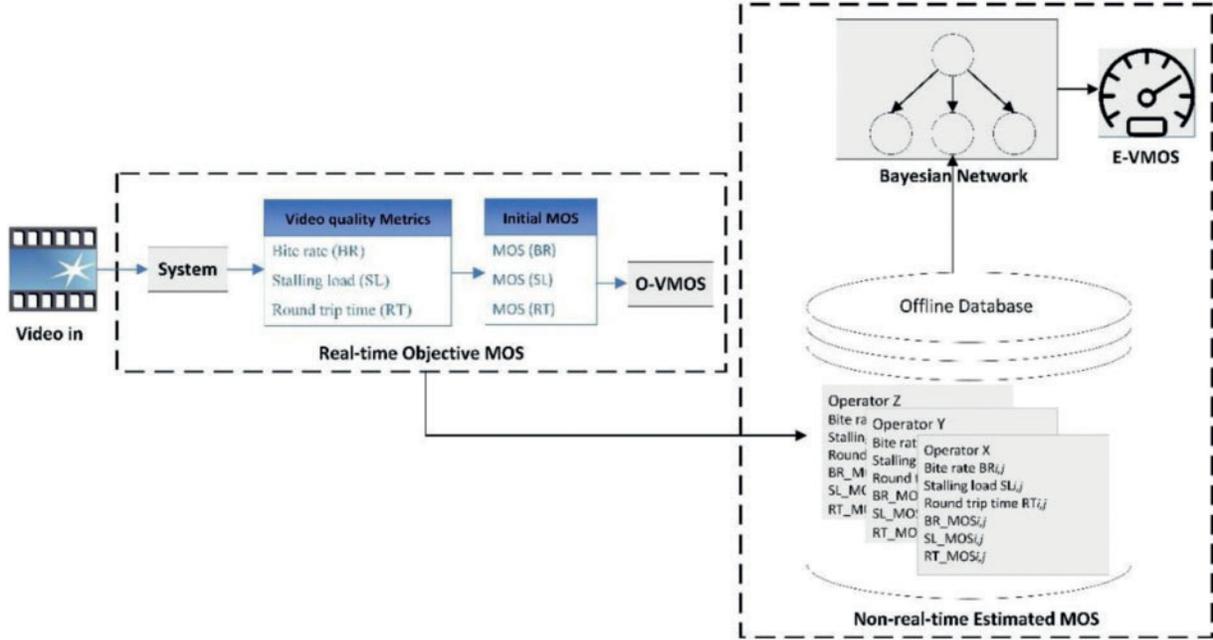


Figure 5: A general framework of estimated MOS process

$$s_i \sim P(S_k | s_{-i}) \quad (12a)$$

s_{-i} is an assignment to all $s_1, s_2, s_3 \dots s_k$ except s_i

Next loop through the following steps:

$$s_1^{(i+1)} \sim P(s_1 | s_2^{(i)}, s_3^{(i)}, \dots, s_k^{(i)}) \quad (12b)$$

$$s_2^{(i+1)} \sim P(s_2 | s_1^{(i+1)}, s_3^{(i)}, s_k^{(i)}) \quad (12c)$$

$$s_3^{(i+1)} \sim P(s_3 | s_1^{(i+1)}, s_2^{(i+1)}, \dots, s_k) \quad (12d)$$

⋮

$$s_k^{(i+1)} \sim P(s_k | s_1^{(i+1)}, s_{k-1}^{(i+1)}) \quad (12e)$$

These steps are repeated continuously with generated values to replace the initial values that burn the sample at the first loop. The chain of values produced by these procedures refers to the Markov chain (MC), which is a joint posterior distribution that converges to its equilibrium distribution.

3.2 Performance Metrics

In this section, the performance of the proposed model is evaluated based on an unknown estimated MOS (EMOS) node. The OpenBUGS tool is the most widely used software packages for fitting Bayesian models because it enables the user to specify a Bayesian model (prior and likelihood) in the R language. In this sense, this software was utilized to evaluate and analyze the proposed model based on MCMC simulations following Bayesian statistical theory. The specification of the proposed model targets a posterior distribution to infer the unknown values. OpenBUGS provides and displays

the numeric results of the estimation of the unknown MOS nodes and a summary of the posterior inference.

The EMOS was examined using different testing methods. The testing data were divided into four test groups: 20%, 40%, 60%, and 80%. Each testing data were examined separately to investigate the effectiveness of the testing data size. All tested data sizes used the same parameters during the inference phase. The proposed model was initialized, updated with 1000 iterations for burning-in samples, and then updated with 10000 iterations for estimating the EMOS node. Burning-in samples were used to discard the effect of initially generated samples on the posterior inference. The estimated EMOS node is represented by a value within a range of 1 to 5. These values indicate network quality and user satisfaction. Fig. 6 shows the different qualities and the lowest MOS Score limit for each one. The limit values are from the ITU-T G.107 standards [20]. Each range in the MOS scale indicates the network quality and user satisfaction. Several performance metrics were used to compare the difference between the actual AMOS and the EMOS values. In this work, four performance metrics were used: mean absolute error (MAE), mean squared error (MSE), mean percentage error (MPE), and mean squared error (RMSE) which can be given as follows:

$$\text{Estimated } E = (AMOS_i - EMOS_i), \quad i = 1, 2, \dots, n, \tag{13}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |AMOS_i - EMOS_i|, \tag{14}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|AMOS_i - EMOS_i|}{AMOS_i} \times 100, \tag{15}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (AMOS_i - EMOS_i)^2, \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (AMOS_i - EMOS_i)^2}, \tag{17}$$

where $AMOS$ and $EMOS$ are actual and estimated MOS, respectively. n represents the number of test data. Eq. (15) was used to calculate the error rate and feasibility of the proposed model. Finally, the proposed Bayesian model is compared with a neural network in terms of accuracy.

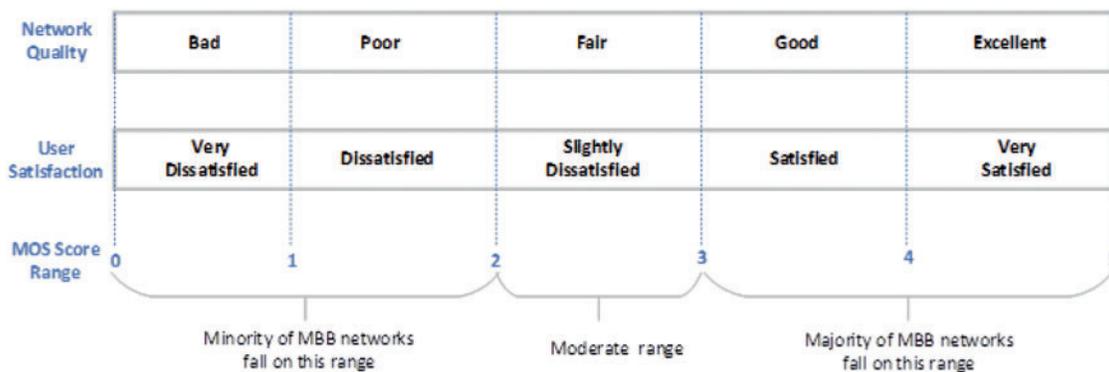


Figure 6: Relation between network quality and MOS value

4 Results and Discussion

In this section, the simulation results of the proposed model are presented and discussed. The test data are the same for the four cases to ensure a fair comparison. Fig. 7 shows an example of generated samples of a single MOS value of 3.5 during the sampling process for different iteration numbers. Most generated samples are bound between 3 and 4, considering the relative values to the actual MOS for all iterations. Increasing the number of iterations may result in a slight increase in system accuracy in such cases. Fig. 8 shows the dynamic tracing of generated values against iteration number for six estimated EMOS values during the inference process. These tracing figures analyze the pattern of the posterior distribution for obtaining estimated EMOS nodes. The results were obtained by running the MC chain for 10000 iterations with another 1000 iterations for the burn-in samples. Tab. 1 shows a statistical summary of the estimated EMOS nodes with present MC error. The MC represents the error margin when the MCMC samples were used to estimate the posterior mean. The table shows that the MC error achieved very low values for all EMOS nodes.

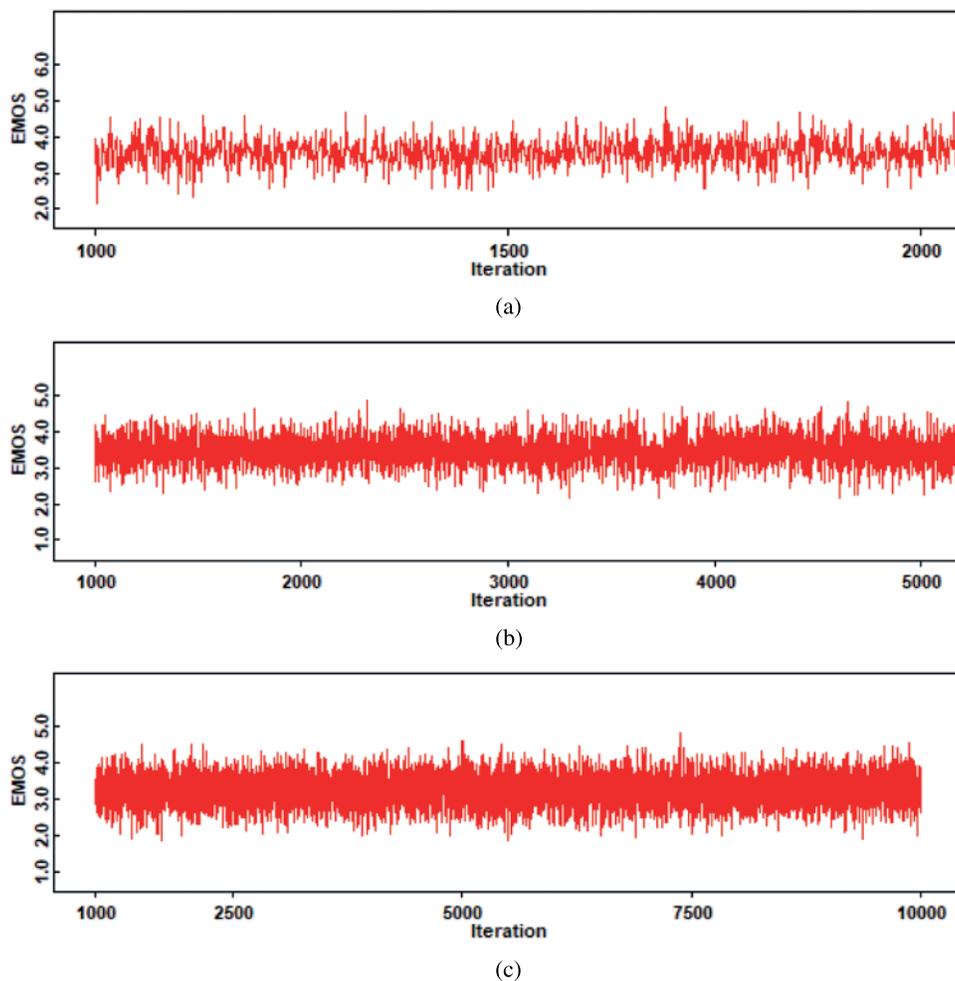


Figure 7: Generated samples of EMOS value (a) 2000 iterations (b) 5000 iterations (c) 10000 iterations

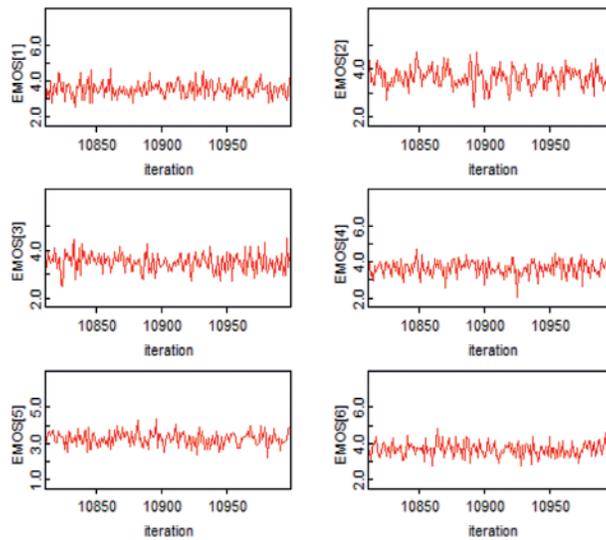


Figure 8: Dynamic tracing of six EMOS nodes

Table 1: Statistics summary of estimated EMOS nodes

	AMOS	Mean	Standard deviation	MC_error	val2.5pc	Median	val97.5pc
EMOS [1]	3.34	3.539	0.3795	0.003954	2.802	3.535	4.291
EMOS [2]	3.74	3.629	0.3786	0.003527	2.885	3.63	4.373
EMOS [3]	3.30	3.482	0.3781	0.003793	2.738	3.485	4.207
EMOS [4]	3.80	3.698	0.3783	0.003969	2.963	3.701	4.441
EMOS [5]	3.12	3.258	0.3809	0.004077	2.52	3.259	3.999
EMOS [6]	3.80	3.703	0.3809	0.003535	2.957	3.706	4.45

Note: [] indicates the node index in the dataset *val2.5pc* and *val97.5pc* indicate the 2.5% and 97.5% quantiles, respectively.

Fig. 9 shows the smoothed kernel density of six estimated MOS values. This figure is similar to a smoothed histogram where each iteration is distributed around the estimated node using a Kernel function, such as normal distributions instead of counting the estimates into bins of particular widths. The kernel density estimation is a means of estimating the posterior probability density function of a random variable. The high density concentrates between 3 and 4, which are the most accurate values of actual MOS. Bell-shaped posterior distributions indicate that the MC chain has reached the convergence level.

Fig. 10 visualizes the various plotting types of estimated MOS values over several points of data. Fig. 10a shows the boxplot of the posterior distributions of all estimated MOS values were summarized side by side. The interquartile range and anticipated mean value of MOS are represented by the green boxes and the center lines, respectively. The box arms extend to cover the center 95% of the distribution, with their ends corresponding to the 2.5% and 97.5% quantiles. More informative statistics and visualizations are provided in Fig. 10b, which depicts the full posterior distribution through shading, with the strip blackness at each data point defined as proportionate to the estimated density. Similarly,

Fig. 10c represents the “caterpillar” plot, which is conceptually very similar to Fig. 10a. The x-axis for each distribution is summarized by a horizontal line representing the 95% interval, whereas the dots indicate the mean location.

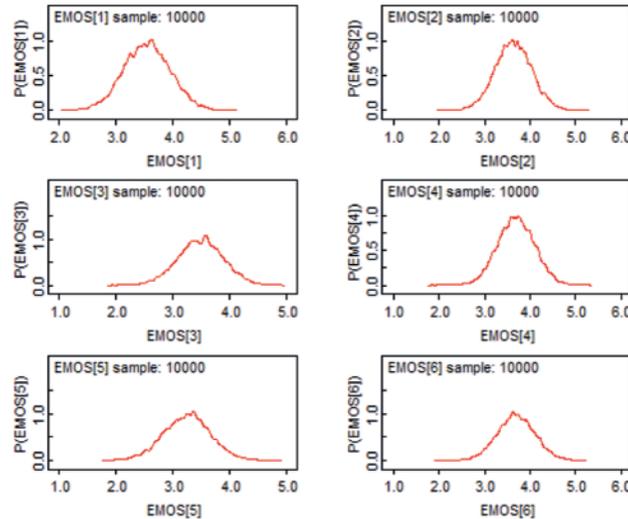
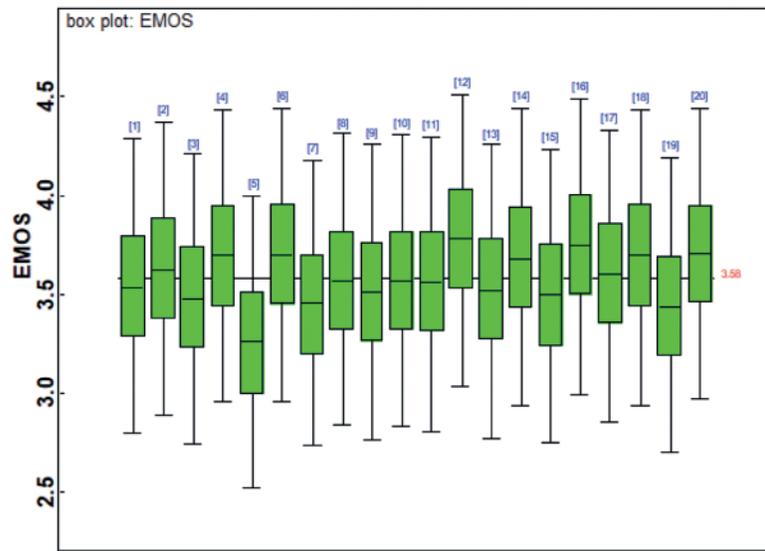


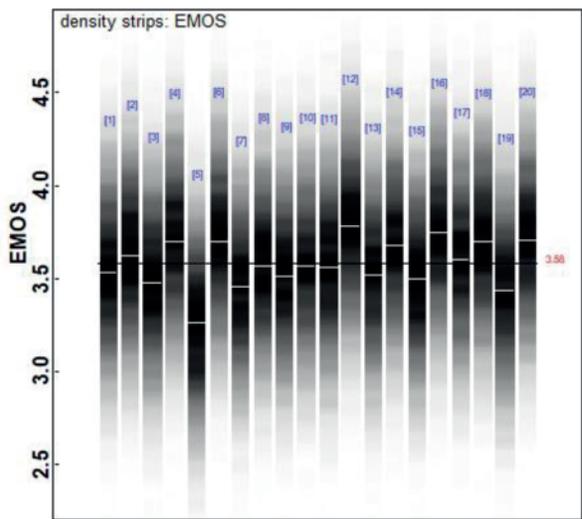
Figure 9: Kernel density of six EMOS nodes

Fig. 11 shows a bar graph of AMOS, EMOS with absolute and squared error against random testing points. The absolute error of estimating MOS with AMOS of greater than 3 is lower compared to the MOS smaller than 3 because the data contain MOS greater than 3, and the majority of MBB networks fall in this score with a good and satisfying network. Thus, the model trains well with these MOS compared with the small data of MOS = 1. The maximum absolute average errors for MOS greater and smaller than 3 are 0.20 and 0.45, respectively.

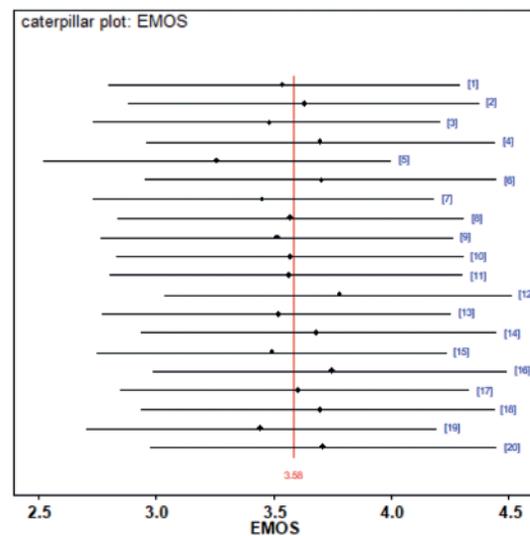
Fig. 12 illustrates the comparison between the proposed Bayesian and neural networks in terms of MAE, MSE, and RMSE over test data sizes. In this comparison, these test data sizes are considered (20%, 40%, and 60%) where a higher test data size cannot be implemented practically. The dataset is divided into 20% test data and 80% training data in most common works. However, the proposed model is examined for two more test data sizes. The figure shows that the MAE, MSE, and RMSE increase when the size of test data increases. In the Bayesian network, the test data size of 20% achieved the lowest MAE and MSE of 0.148 and 0.025, respectively, whereas the test data size of 60% achieved the highest MAE and MSE of 0.259 and 0.110, respectively. Similarly, in the neural network, the test data size of 20% achieved the lowest MAE and MSE of 0.21 and 0.10, respectively, whereas the test data size of 60% achieved the highest MAE and MSE of 0.68 and 1.21, respectively. The results are because when the test data size is large, the training data will be small, which is insufficient to train the model. Thus, the estimated nodes have greater variance with less training data, while the performance statistic will have greater variance with less testing data. The proposed Bayesian network outperforms the neural network in all test data sizes. Tab. 2 summarizes different performance metrics over test data sizes.



(a)



(b)



(c)

Figure 10: Visualization of estimated MOS (a) Boxplot (b) Caterpillar (c) Density strip

Fig. 13 demonstrates the accuracy comparison between the Bayesian and neural networks. The proposed Bayesian network achieved a high accuracy of 95.72%, 93.87%, and 81.49% for the test data sizes of 20%, 40%, and 60%, respectively. The neural network achieved a low accuracy of 88.07%, 81.00%, and 35.42% for test data size of 20%, 40%, and 60%, respectively. Both networks obtained the highest accuracies when using 20% of the test data size. However, the neural network had poor accuracy of 35.42% when using 60% of the test data size compared with the Bayesian network. Overall, the Bayesian and neural networks achieve 90.36% and 68.16%, respectively. Therefore, the proposed Bayesian network achieved remarkable accuracy and outperformed the neural network.

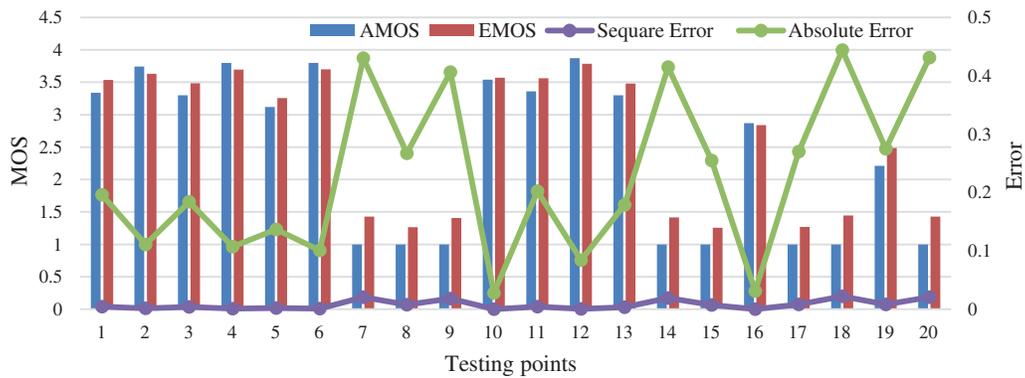


Figure 11: Graph bar of AMOS, EMOS with absolute and squared error against random test points

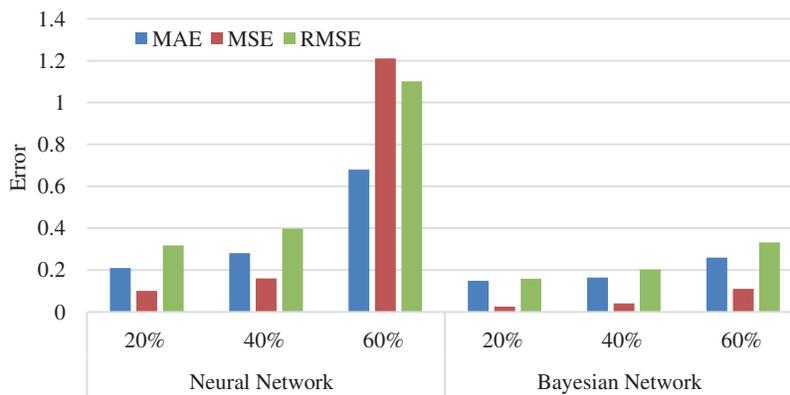


Figure 12: Comparison between bayesian and neural networks over test data sizes

Table 2: Summary of different performance metrics over test data sizes

	Test data size	MAE	MSE	RMSE	MAPE
Bayesian network	20%	0.148	0.025	0.158	4.28%
	40%	0.164	0.041	0.202	6.13%
	60%	0.259	0.110	0.331	18.51%
Neural network	20%	0.21	0.10	0.317	11.93%
	40%	0.28	0.16	0.397	19.00%
	60%	0.68	1.21	1.102	64.58%

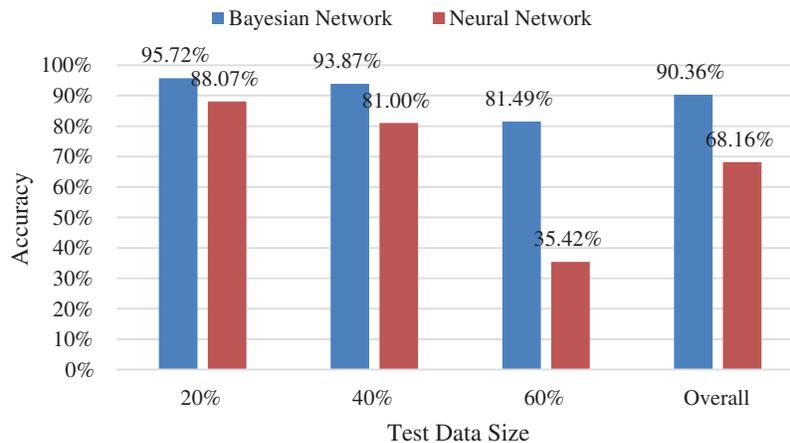


Figure 13: Accuracy comparison between bayesian and neural networks

5 Conclusion

In the present work, a Bayesian network based on a probabilistic graphical model to estimate MOS in MBB networks is proposed. Video streaming data have been utilized for training and testing the proposed model to estimate accurate MOS. Several performance metrics have been used to evaluate the proposed model with various implementing scenarios. The results showed that the proposed Bayesian network achieved high accuracy overall test data size compared with the neural network. Overall, the Bayesian network obtained an accuracy of 90.36% overall test data sizes. Therefore, the proposed Bayesian network achieved remarkable accuracy and outperformed the neural network. An interesting path for future investigation would be to explore the MOS of several services, such as video streaming with multiple resolutions, gaming, and web browsing. Multiple databases of several mobile network providers can be also be investigated further.

Acknowledgement: This research study was sponsored by the Universiti Teknologi Malaysia through the Professional Development Research University Grant (No. 05E92). It is also funded using the Collaborative Research Grant National (Grant reference PY/2019/02791).

Funding Statement: The research leading to these results has received funding from The Research Council (TRC) of the Sultanate of Oman under the Block Funding Program with Agreement No. TRC/BFP/ASU/01/2019.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] ITU, "Series P: Telephone transmission quality, telephone installations, local line networks, vocabulary for performance, quality of service and quality of experience. T P.10/G.100," in *International Telecommunication Union Telecommunication Standardization Sector*, Geneva, Switzerland: International Telecommunication Union (ITU), 2017.
- [2] ITU, "Series P: Terminals and subjective and objective assessment methods: Methods for objective and subjective assessment of speech and video quality: Mean opinion score (MOS) terminology, ITU-T Rec. P.800.1," in *Telecommunication Standardization Sector ITU*, Geneva, Switzerland: International Telecommunication Union (ITU), 2016.
- [3] R. C. Strejtl, S. Winkler and D. S. Hands, "Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives," *Multimedia Systems*, vol. 22, pp. 213–227, 2016.
- [4] C. Callegari, R. G. Garroppo, S. Giordano, C. C. Labruzzo, G. Procissi *et al.*, "Experimental analysis of ViLTE service," *IEEE Access*, vol. 6, pp. 21129–21139, 2018.
- [5] V. Pedras, M. Sousa, P. Vieira, M. P. Queluz, A. Rodrigues *et al.*, "A no-reference user-centric QoE model for voice and web browsing based on 3G/4G radio measurements," in *2018 IEEE Wireless Communications and Networking Conf. (WCNC)*, Barcelona, Spain, pp. 1–6, 2018.
- [6] T. Daengsi and P. Wuttidittachotti, "QoE modeling for voice over ip: Simplified e-model enhancement utilizing the subjective MOS prediction model: A case of G. 729 and Thai users," *Journal of Network and Systems Management*, vol. 27, pp. 837–859, 2019.
- [7] T. Ghalut and H. Larijani, "Non-intrusive method for video quality prediction over lte using random neural networks (rnn)," in *2014 9th Int. Symp. on Communication Systems, Networks & Digital Sign (CSNDSP)*, Manchester, UK, pp. 519–524, 2014.
- [8] Y. Zang, T. Hu, T. Zhou and W. Deng, "An automated penetration semantic knowledge mining algorithm based on bayesian inference," *Computers, Materials & Continua*, vol. 66, pp. 2573–2585, 2021.
- [9] A. -I. Al-Omari, A. -S. Hassan, H. -F. Nagy, A. -R. -A. Al-Anzi, L. Alzoubi *et al.*, "Entropy bayesian analysis for the generalized inverse exponential distribution based on URRSS," *Computers, Materials & Continua*, vol. 69, pp. 3795–3811, 2021.
- [10] ITU, "P.912 subjective video quality assessment methods for recognition tasks," in *Int. Telecommunications Union Telecommunication Sector*, Geneva, Switzerland: International Telecommunication Union (ITU), pp. 72400Z, 2016.
- [11] ITU, "913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," in *International Telecommunication Union*, Geneva, Switzerland: International Telecommunication Union (ITU), 2021.
- [12] ITU, "P.1202. Parametric non-intrusive bitstream assessment of video media streaming quality," in *International Telecommunications Union Telecommunication Sector*, Geneva, Switzerland: International Telecommunication Union (ITU), 2013.
- [13] ITU, "J.343: Hybrid perceptual bitstream models for objective video quality measurements," in *International Telecommunications Union Telecommunication Sector*, Geneva, Switzerland: International Telecommunication Union (ITU), 2014.
- [14] T. ITU, "1070 multimedia quality of service and performance—generic and user-related aspects," in *Opinion Model for Video Telephony Applications*, Geneva, Switzerland: International Telecommunication Union (ITU), 2012.
- [15] ITU-T, "G.1071: Opinion model for network planning of video and audio streaming applications," in *International Telecommunications Union Telecommunication Sector*, Geneva, Switzerland: International Telecommunication Union (ITU), 2016.
- [16] A. Alhammadi, A. El-Saleh, and I. Shayea, "MOS prediction for mobile broadband networks using Bayesian artificial intelligence," in *2021 Int. Conf. on Artificial Intelligence and Computer Science Technology (ICAICST)*, Yogyakarta, Indonesia, pp. 47–50, 2021.

- [17] I. Shayea, M. Ergen, M. H. Azmi, D. Nandi, A. A. El-Salah *et al.*, “Performance analysis of mobile broadband networks with 5G trends and beyond: Rural areas scope in Malaysia,” *IEEE Access*, vol. 8, pp. 65211–65229, 2020.
- [18] OpenBUGS (Access: 31-05-2021). “Bayesian inference using gibbs sampling,” [Online]. Available: <https://www.openbugs.net/w/FrontPage>.
- [19] A. E. Gelfand and A. F. Smith, “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, vol. 85, pp. 398–409, 1990.
- [20] ITU-T, “ITU-T recommendation G. 107: The E-model, a computational model for use in transmission planning,” in *Series G: Transmission Systems and Media, Digital Systems and Networks*, Geneva, Switzerland: International Telecommunication Union (ITU), 2015.