Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Enhancing VoLTE Quality using Unsupervised Learning

Mina Gameel Saddeek Girgis[1], Mohamed Shehata[1,*], Karim Hammad[1], Safa M. Gasser[1]

[1] Department of Electronics and Communications, Arab Academy for Science and Technology and Maritime Transport Cairo, Cairo 4471344, Egypt

ARTICLE INFO

Article history:
Received 2 March 2025
Received in revised form 20 July 2025
Accepted 20 August 2025
Available online 29 August 2025

Keywords:
LTE; VoLTE; mobile networks; machine learning; QoS; clustering; KPIs

ABSTRACT

The integration of machine learning (ML) algorithms with data analytics has become essential for optimizing telecommunication services due to the growing complexity of networks. Voice over LTE (VoLTE) is among other technologies that increased this complexity. Hence, optimizing VoLTE to ensure the highest quality-of-service (QoS) attained by end users presents a significant challenge. In this paper, we propose a framework that links key performance indicators (KPIs) with drive test data to gain more insights about the network utility and better manage its resources. The numerical evaluation for our proposed framework demonstrates 25% improvement in the VoLTE QoS compared to other existing approaches.

1. Introduction

With the explosive growth of 4G and 5G services and applications, the migration from the circuit-switched 2G and 3G voice services to the emerging LTE packet-switched voice services (also known as VoLTE) is inevitable [1]. VoLTE enables higher capacity and improved telecommunication services. It provides a foundation for improving business and enterprise collaboration services in combination with high-quality mobile voice services. Hence, mobile network operators (MNOs) pay much attention to enhance the voice quality to show its benefits and guarantee customer satisfaction.

Perceptual Objective Listening Quality Analysis (POLQA) is a widely known measure for the VoLTE quality. POLQA is the ITU-T P.863 standard that supersedes PESQ [2]. and addresses range of limitations of PESQ as well as improving the overall correlation with subjective to the mean opinion score (MOS). It allows for predicting overall listening speech quality in two modes: narrowband (300 to 3,400 Hz) and super wideband (50 to 14,000 Hz). POLQA [3] is an objective method for assessing the perceived voice quality of a VoLTE call. POLQA is widely used in the telecommunications industry

[*] *Corresponding author*
E-mail address: mkhshehata@aast.edu

https://doi.org/10.37934/ard.141.1.257271

to evaluate the voice quality of VoLTE calls and is a highly accurate and reliable method for measuring voice quality.

POLQA works by analysing the speech signal transmitted over the VoLTE network and comparing it to a high-quality and distortion-free reference speech signal. The difference between the transmitted signal and the reference signal is then used to calculate a Mean Opinion Score Listening Quality Objective (MOS-LQO) score. That difference represents the perceived voice quality of the call on a scale of 1 to 5, with 5 being the highest quality.

1.1 Related Work

Numerous studies have considered improving the LTE network performance using unsupervised machine learning (ML) techniques as Santos *et al.,* [4] they clustered their collected data for mobile broad band (MBB) to differentiate between good cluster performance and bad cluster performance. Then they took a proper action in terms of accessibility on the bad cluster sites area and got a clear gain which means the power of clustering was clear for network performance enhancement. Clustering is key to determine the affected KPIs areas (based on the category of chosen KPIs) to enhance the network's accessibility, retainability and integrity. The emergence of VoLTE has placed a significant shift in the approach to optimizing and enhancing the network's performance as it garnered considerable attention by MNOs for better spectral utilization and profitable considerations. Hence, detailing the new architecture and its operation is of a fundamental essence as described by El Wakiel *et al.,* [5]. In this paper they followed the default quality of service per VOLTE class identifier settings and collected some VOLTE calls traces for randomly distributed calls in good coverage, bad coverage, busy hours, non-busy hours, stationary UEs, moving UEs over congested or not loaded eNodeBs. From analysis a relation between MOS values with combination of bad coverage areas and Packet loss rate was figured which indicates bad coverage areas with high packet loss rates suffer from low MOS values. As a result, a new research space has been created for innovating new approaches that attain better QoS compared to traditional mobile broad band (MBB). In state-of-the-art voice-over-IP (VoIP) networks, the performance is evaluated using simulators to identify the main aspects affecting the QoS of end users [6]. In particular, the authors analysed VoLTE's end-to-end performance using OPNET simulator and determined the effect of different voice codecs on the MOS values as they got the relation of each voice codec to its MOS value and packets delay (End to End, Uplink and downlink). Nguyen [7] also studied the effect of the adopted channel coding scheme and its result shows that with an adaptation of joint source-channel code rate, the redundant bits generated by channel coding can be reduced up to 50% with a slight reduction of voice quality is 1% so a clear relation to voice quality obtained from this action. And another approach to enhance downlink scheduling mechanism which resulted in great voice quality enhancement. In a different context, the report by Rivas *et al.,* [8] had an experiment testbed and correlated (the packet loss rate (PLR) with the signal to noise ratio (SNR), PLR to Physical downlink shared channel (PDSCH) Block error rate (BLER) with correlation factor 0.5892, MOS to SNR, MOS to PLR with correlation factor 0.663, MOS to PDSCH BLER with correlation factor 0.6244) and they found usually a relation of those parameters change to MOS values. While Lipovac *et al.,* [9] the investigation of the impact of HARQ retransmissions on Quality of Experience (QoE) has been addressed and confirmed that increase in HARQ retransmissions decreases QoE. A more focused study on VoLTE's packet loss rate, jitter and packet end to end delay in conjunction with the channel bandwidth selection has been presented by Tabany *et al.,* [10] and according to simulation results it showed that higher bandwidth gives better voice quality. Vetoshko *et al.,* [11], the study investigated the relation between MOS values and signal power with the effect of delay and packet loss rate. They found a delay-dependence on the MOS in good coverage samples while in bad coverage samples the MOS is being affected by packet loss rate. The MOS has also been evaluated by Vyas *et al.,* [12] with different codecs based on the downlink path loss also they get the relation of the downlink pathloss change with packet loss rate for different codecs. Finally, Adhilaksono *et al.,* [13] presented a comprehensive investigation on several QoS metrics and arranged them from highly affecting parameter to be Jitter, packet loss, bandwidth and throughput.

From the above research we can confirm that many relations between some KPIs and voice quality were obtained while using simulated data, on the other hand the research that used a real data neither correlate nor use machine learning techniques to reach their conclusions. From this point of view, we collected the gain of real data availability with the power of machine learning techniques to reach a subjective powerful parameter that could affect the voice quality while also combining the evaluation between field tests and system collected KPIs in same study.

In this paper we strive to identify the main affecting KPIs on the VoLTE QoS from a practical point-of-view, as we used a real data from a local MNO, using unsupervised ML techniques and correlation approaches to confirm the expectations from a real field-based measured data.

The rest of the paper is organized as follows: Section 2 presents the proposed methodology; Section 3 summarizes the results for the network performance evaluation, while Section 4 presents the results verification. Finally, conclusions are drawn in Section 5.

2. Proposed Methodology

The traditional HO algorithm, as powerful as it is, still could not dynamically and flexibly address the changes occurring in the real time and hence, no guarantees to the user's QoE can be offered. On the contrary, NN models have the ability to learn complex patterns and correlations in the network data, which makes them more adaptive and context-aware in making HO decisions for VoLTE networks.

The analytical study presented in this work is based on a real cellular data provided by a local MNO. The KPIs were collected periodically from a live LTE network for LTE 1800 frequency during five days for 969 cells from two sources. The drive test (DT) team has collected POLQA each day from around 10 AM to 6 PM and performance management (PM) server KPIs for same days and same test hours. Prior to the feature engineering stage, the collected PM data went through a cleaning process where null values for any cell from DT or KPIs were removed.

We have used 27 KPIs related to different categories to be sure we have the effect of all KPIs that may affect the POLQA. KPIs categories are interference KPIs which indicates uplink interference for control and data channels. Capacity KPIs which take downlink and uplink physical resource blocks overall cell level and for specific QoS class identifier (QCI) related to VoLTE in addition to control channel elements into consideration. Coverage KPIs like the pathloss and the UE that use high uplink power to reach the sites. Integrity KPIs for signal to noise ratio, volte downlink control channel error rates and channel quality indication (CQI). Packet loss KPIs for downlink and uplink generally on cell level and specifically for VoLTE QCI. Finally delay KPIs which include many items like delay of packets that arrive out of delay budget, not acknowledge packets, latency per QCI, Discontinuous receiving time percentage, silent time in downlink and uplink in VoLTE calls, the inactive time in downlink and uplink as well in addition to number of scheduled user equipment (UE) in transmission time interval in downlink and uplink as well.

POLQA value for each cell that obtained from DT is added as well for each cell. The data is then processed in three different stages: correlation, applying minimum redundancy maximum relevance (MRMR) to detect the most affecting KPIs on POLAQ and clustering to confirm the difference in KPIs for the obtained clusters and act to enhance affected cells as shown in Figure 1.

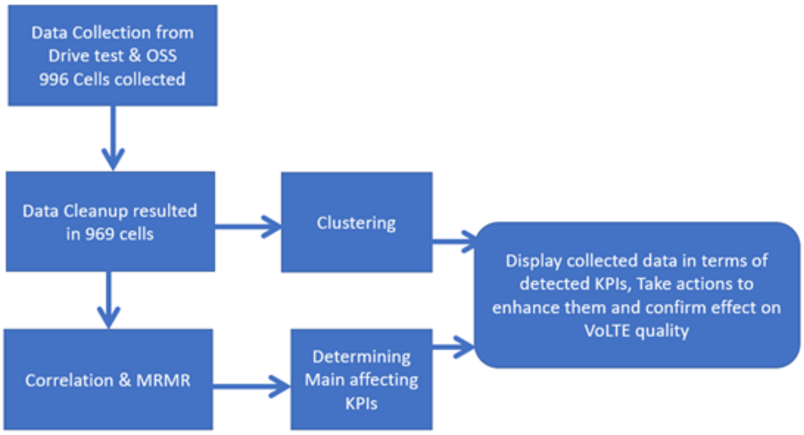

Fig. 1. Data processing architecture

2.1 Correlation Methods

We correlated all the previously listed KPIs to POLQA to get most affecting KPIs in POLQA using linear and nonlinear methods. Linear correlation refers to the relationship between two variables that can be represented by a straight line on a scatter plot. It is commonly used in statistical analysis and can provide insights into the relationship between two variables and thus, can aid in making predictions and developing models.

The strength of the linear correlation is measured by a correlation coefficient, which ranges from -1 to +1. A correlation coefficient of -1 indicates a perfect negative linear correlation, while a correlation coefficient of +1 indicates a perfect positive linear correlation. A correlation coefficient of 0 indicates no linear correlation between the two variables.

On the other hand, non-linear correlation is a type of correlation where the relationship between two variables is not linear. In particular, the variables do not have a constant rate of change. In non-linear correlation, the value of one variable changes in a non-proportional manner with the change in the other variable. Examples of non-linear functions include quadratic, logarithmic, exponential and power functions: we used linear correlation like Pearson correlation, some non-linear correlations like (Kendall and spearman correlation) , some correlations that measure both linear and nonlinear like (Maximal information coefficient and Distance similarity) and finally Cosine similarity.

2.1.1 Pearson correlation coefficient

Pearson correlation coefficient [14] is used to determine the linear relationship between two continuous variables. The Pearson correlation coefficient as defined in Eq. (1) is a descriptive statistic, meaning that it summarizes the characteristics of a dataset. Specifically, it describes the strength and direction of the linear relationship between two quantitative variables. We tried this method to confirm if our data has any linear characteristics or not.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{1}$$

Where r= Pearson Correlation Coefficient, x_i= x variable samples, y_i= y variable samples, $\bar{x}$= mean of values in x variable, $\bar{y}$= mean of values in y variable

2.1.2 Kendall Rank correlation coefficient

Kendall Rank correlation coefficient [15] expressed in Eq. (2) is used to measure the strength and direction of association between two variables where the data is in the form of ranks or ordinal data as shown in Figure 2. The advantage for this approach and why we have tried is its ability to deal with outliers which may bias the correlation results in other techniques.

$$r = \frac{(no.of\ concordant\ pairs) - (no.of\ discordant\ pairs)}{(number\ of\ pairs)} = 1 - \frac{2(no.of\ discordant\ pairs)}{(2^n)} \quad (2)$$

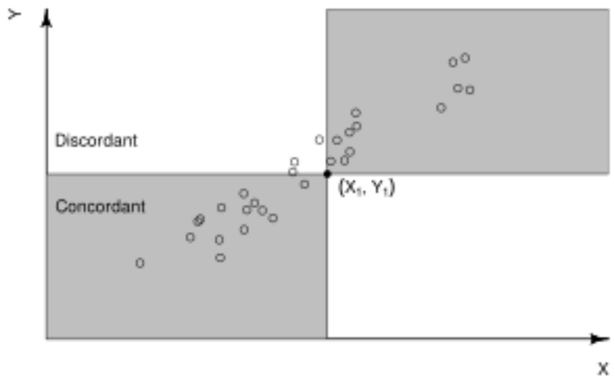

Fig. 2. Kendall correlation approach

2.1.3 Spearman Rank correlation coefficient

Data charts which are typically black and white, but sometimes include colour. Spearman Rank correlation coefficient [16,17] defined in Eq. (3) is like Kendall Rank correlation, but it measures the relationship between two variables with ordinal data or where variables are not normally distributed. Spearman's rank correlation measures the strength and direction of association between two ranked variables. It basically gives the measure of monotonicity of the relation between two variables i.e. how well the relationship between two variables could be represented using a monotonic function.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (3)$$

Where: ρ= Spearman's rank correlation coefficient, d_i= Difference between the two ranks of each observation, n= Numbers of observations

2.1.4 Maximal information coefficient (MIC)

MIC [18] is a measure of non-parametric correlation that measures the strength of the association between two variables. It can detect both linear and non-linear relationships between variables. MIC values range from 0 to 1. It is an information theoretic measure of association that can capture a wide range of functional and non-functional relationships between variables. MIC is equal to the coefficient of determination (R2), expressed in Eq. (4).

$$MIC(x, y) = max\left\{I(x, y)/log_2 min\{n_x, n_y\}\right\} \quad (4)$$

where n_x and n_y are the number of bins on the x-axis and y-axis, respectively. I (X, Y) denotes the mutual information under the grid.

2.1.5 Distance correlation

Distance Correlation is a measure of dependence between two variables that captures non-linear dependencies [19]. It measures the degree of association between two variables based on the distances between observations in a large-dimensional space. The values of distance correlation range between 0 and 1. Distance correlation is not the correlation between the distances themselves, but it is a correlation between the scalar products which the "double centred" matrices are composed of, as shown in Eq. (5). From its advantages it detects the casual relations and human behaviour, so we tried to check if our data matches any undefined correlation relations through that approach.

$$dCor(X,Y) = \frac{dCor(X,Y)}{\sqrt{dVar(X)dVar(Y)}} \tag{5}$$

where cov(X, Y) denotes the covariance of X and Y and for any random variable Z, Var(Z) denotes the variance of Z.

2.1.6 Cosine similarity

As illustrated in Eq. (6), it is a method of calculating the similarity between two vectors [20-22] by taking the dot product and dividing it by the magnitudes of each vector. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. This approach is a powerful one in computational efficiency and good retrieval performance. So, we used to get benefits of those differences of this approach's way of calculations rather than our techniques as we have some complexity in our data.

$$similarity(A,B) = cos(\theta) = \frac{A.B}{\|A\|\|B\|} \tag{6}$$

2.2 Minimum Redundancy Maximum Relevance (MRMR)

These MRMR [23,24] is a feature selection technique used in ML and data mining to select the most relevant and non-redundant features of a dataset. The main objective is to maximize the relevance of features to the target variable, while minimizing the redundancy among the features.

The MRMR algorithm works by calculating two scores for each feature in the dataset: relevance and redundancy. The relevance measures the relationship between the target variable and individual features, while the redundancy measures the similarity between features.

The algorithm then selects the features with the highest relevance and lowest redundancy scores until the desired number of features is obtained. The main advantages of using the MRMR algorithm are that it provides a set of relevant and non-redundant features that can improve the accuracy of models, reduce the dimensionality of the dataset and improve the interpretability of the results.

MRMR has been used in various applications, including gene expression analysis, text classification and image recognition. As expressed in Eq. (7) MRMR gets mutual information, try to minimize equation Eq. (8) and try to maximize equation Eq. (9).

$$I(x,y) = \sum_{i,j} p(x_i, y_i) log \frac{p(x_i,y_i)}{p(x_i)p(y_i)} \quad (7)$$

Where the mutual information I of two variables x and y is defined based on their joint probabilistic distribution $p(x, y)$ and the respective marginal probabilities $p(x)$ and $p(y)$

$$min\ W_I, \quad W_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j), \quad (8)$$

where we use $I(i,j)$ for notational simplicity and $|S|$ is the number of features in S.

$$max\ V_I, \quad V_I = \frac{1}{|S|} \sum_{i \in S} I(h,i), \quad (9)$$

we call h the classification variable and the gene expression *gi. I(h,gi)* quantifies the relevance of *gi* for the classification task, task, where we refer to *I(h,gi)* as *I(h,i).*

2.3 Clustering

Clustering is a technique used in unsupervised ML and data mining to group similar objects or data points based on their features or characteristics [25]. Clustering is used to identify patterns and structures in unlabelled data and to segment the data into meaningful groups or clusters. Clustering algorithms vary based on their approach and underlying assumptions about the data. Many types of clustering methods exist but we will briefly describe the ones that we used as follows:

2.3.1 K-means clustering

This algorithm partitions the data into K clusters based on the distance between the data points and the centroid of each cluster [25]. It aims to minimize the sum of squared distances Eq. (10) between the data points and their respective cluster centroids Eq. (11).

K-means silhouette is a metric used to evaluate the quality of clustering in K-means clustering algorithm [26,27]. It measures how well each data point fits into its assigned cluster. This is done by calculating the distance between a data point and other points in the same cluster compared to the distance between a data point and points in other clusters. A silhouette score ranges from -1 to 1, with a higher score indicating that the data point is well-matched to its assigned cluster and poorly matched to other clusters. The K-means silhouette score can be used to determine the optimal number of clusters for a given dataset. The number of clusters that produces the highest silhouette score is the best number of clusters for the data.

K-means is one of the recommended techniques to be used as it is able to deal with large data sets and usually reaches the convergence of the given data which suits our case.

$$Euclidean\ Distance\ = \sqrt{((X_1 - X_2)^2 + (Y_1 - Y_2)^2)} \quad (10)$$

$$Centroid\ = ((X_1 + X_2 + \cdots + X_n)/n, (Y_1 + Y_2 + \cdots + Y_n)/n \quad (11)$$

2.3.2 Hierarchical clustering

This algorithm creates a tree-like structure of clusters by recursively dividing the data into smaller clusters based on their similarity [24]. It can be agglomerative (bottom-up) or divisive (top-down) as shown in Figure 3. One common method to determine the best number of clusters in hierarchical clustering is to use the dendrogram [25] which shows the hierarchical relationships between the clusters. The advantage of this algorithm is, it shows all the data points in a separation by their distance which give us clear indication where and when to stop and determine which suitable number of clusters and this was needed in our case.

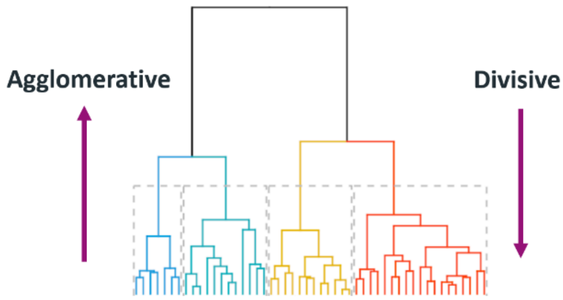

Fig. 3. Hierarchical clustering

2.3.3 Density-based clustering

This algorithm groups data points based on their density in the feature space [25]. It identifies clusters as areas of high density separated by areas of low density. The main drawback is that it doesn't perform as well as others when the clusters are of varying density. This is because the setting of the distance threshold ε and minPoints for identifying the neighbourhood points will vary from one cluster to another when the density varies. This drawback is more severe with very large-dimensional data since the estimation of the distance threshold becomes more challenging. We tried this algorithm to see the size of variation in our data and check if we got benefit from it or not.

2.3.4 Mean-shift clustering

Mean-Shift Clustering is a clustering algorithm that aims to find the centres of clusters in a dataset without the need for a specified number of clusters [28,29]. The algorithm works by iteratively shifting a kernel function towards the high-density regions of the data until it converges to a cluster centre. The kernel function is usually a Gaussian function and the algorithm identifies the cluster centres as the modes of the kernel density estimate. The mean-shift clustering algorithm has several advantages such as its ability to find any number of clusters which is one of our challenges in this work and handle non-linearly separable data that should match our data case. It is also computationally efficient since it does not require any pre-defined number of clusters. However, the algorithm may suffer from scalability issues when dealing with large datasets.

2.3.5 Expectation–maximization (EM) clustering using Gaussian mixture models (GMM):

Gaussian Mixture Models (GMMs) give us more flexibility than K-Means [30]. With GMMs, we assume that the data points are Gaussian distributed; this is a less restrictive assumption than saying they are circular by using the mean. Hence, we have two parameters to describe the shape of the clusters; the mean and the standard deviation. To find the Gaussian parameters for each cluster, we use an optimization algorithm called Expectation–Maximization (EM). EM is an iterative method which alternates between two steps, expectation (E) and maximization (M). For clustering, EM makes use of the finite Gaussian mixtures model and estimates a set of parameters iteratively until a desired convergence value are achieved [31]. We used this approach as it is considered the generic one from the K-means but with ability to capture the clusters in Gaussian shapes instead of expected circle ones from K-means.

3. Numerical Results

In this section we will show the output of all previously mentioned techniques. Starting with correlation and MRMR outputs going through clustering techniques. The aim is to reach the most affecting KPIs in the VoLTE POLQA.

3.1 Correlation and MRMR

Correlation results showed that all the used methodologies gave low correlation values except the cosine similarity technique which gave high correlation values for some KPIs. Table 1 shows the cosine similarity values for the highest correlated ten KPIs. Meanwhile, MRMR was applied as well and we detected the most affecting ten KPIs from the algorithm.

Table 1
Top ten correlated KPIs in Cosine similarity and their correlation values

KPI	Description	Cosine similarity
Uplink packet delay budget OK QCI1 %	The percentage of received uplink packets within the acceptable delay budget for VoLTE	0.993833712
PathLoss	Average of losses that the UL signal faces from UE till eNodeB	0.992467552
Discontinues receiving_Sleep_Time %	Percentage of sleep time of the cell	0.990863435
Channel Quality indicator	Indicates the channel quality on air interface to determine the scheduling resources	0.988425072
Scheduled_Ue/TTI_DL	Number of scheduled UE in downlink per transmission time interval	0.945852217
SNIR	Uplink signal to noise ratio	0.941901388
Scheduled_Ue/TTI_UL	Number of scheduled UE in uplink per transmission time interval	0.932578156
UE Power Restricted	Percentage of UE that send maximum power in uplink per cell	0.848754204
Latency QCI1	Waiting time for uplink VoLTE packets receiving	0.838300527
Silent time per VoLTE User UL ms	Received packets that couldn't be decoded cause silent time in uplink	0.834451173

Table 2 shows the top ten KPIs received from MRMR output. From the obtained results we selected the first ranked KPI from cosine similarity output (Uplink packet delay budget OK QCI1 %) and first ranked KPI from MRMR output (PL Rate UL QCI1) and considered them as the most affecting KPIs on voice quality.

Table 2
Top ten ranked KPIs in MRMR and their rank

KPI	Description	MRMR Rank
PL Rate UL QCI1	Uplink lost packets that have been transmitted and don't reach the receiver for VoLTE	1
PRB Utilization UL QCI1	Percentage of utilized uplink physical resource blocks for VoLTE	2
Inactive Gap UL QCI1	Number of released calls per cell due to inactivity longer than a certain threshold for VOLTE	3
PL Rate UL QCI1 ROCH Fail	Lost uplink VOLTE packets due to header decompression	4
Latency Per QCI	Waiting time for packet receiving per QCI	5
Uplink packet delay budget OK QCI1 %	The percentage of received uplink packets within the acceptable delay budget for VoLTE	6
DL_Packet_Loss	downlink lost packets that have been transmitted and don't reach the receiver for Mobile broad band and VoLTE	7
Silent time per VoLTE User UL ms	Received packets that couldn't be decoded cause silent time in uplink	8
PL Rate DL QCI1	Downlink lost packets that have been transmitted and don't reach the receiver for VoLTE	9
Silent time per VoLTE User DL ms	Received packets that couldn't be decoded cause silent time in downlink	10

3.2 Clustering

We have tested all the previously described clustering techniques and we got various results. For suitable clustering algorithms, they show two clusters are the optimum case.

We have applied the K-means silhouette score before applying K-means itself. We got K-means_silhouette= 0.42 as shown in Figure 4 for two clusters which is the highest value and hence, the optimum choice is two clusters.

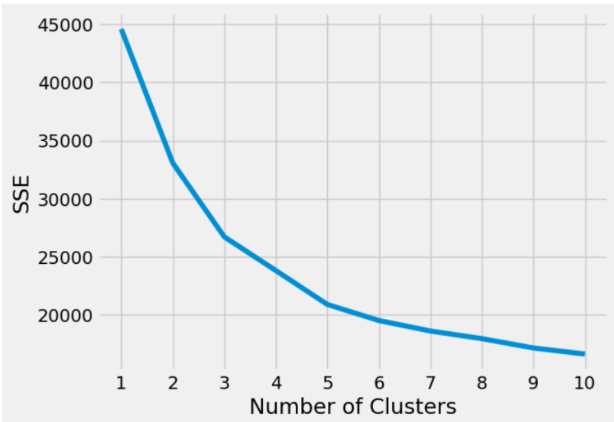

Fig. 4. K-means_silhouette

Also applying hierarchical clustering, we got this dendrogram as shown in Figure 5. It shows two clusters (i.e., two separate groups) are the optimum case then starts to divide small number of cells into bigger number of clusters which leads to misleading outputs.

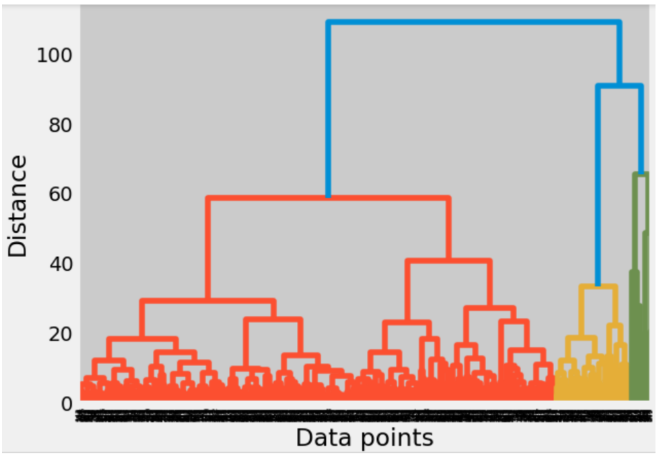

Fig. 5. HC dendrogram

The density-based clustering (with its default parameters), on the other hand, resulted in many -1 values in the cluster results make it unfavourable (as -1 stands for noise points). This means points that have less than min_sample neighbours in the eps (the radius of neighbourhood around a point x) neighbourhood. Hence, we have a single cluster (0) and some noise. In case of mean-shift clustering, the results show twelve needed clusters (i.e., not logic in our case). Finally, applying EM clustering using GMM on the data and testing different number of clusters we found two clusters are the optimum number. In sum, we have three techniques (out of five) that demonstrated acceptable results as shown in Table 3.

Table 3
The three clustering techniques output in matter of two clusters

Clustering technique	Number of Cells in cluster 1	Number of Cells in cluster 2
K-means	766	203
Hierarchical clustering	808	161
Expectation–Maximization (EM) Clustering	646	323

By plotting the selected two KPIs which determined most affecting KPIs for K-means & hierarchical clustering, methods we can confirm the difference between the two clusters in each KPI clearly which confirms our conclusion.

3.2.1 Packet loss rate uplink QoS class identifier 1

The packet loss rate in the uplink (UL) per QoS class indicator (QCI) measurement refers to packet losses for data radio bearers (DRBs). One packet corresponds to one packet data convergence protocol (PDCP) Service Data Unit (SDU) [32]. The measurement is done separately per QCI as shown in Figure 6 it illustrates the values for QCI1 which carriers VOLTE traffic. Good Cluster (GC) shows around 0.07% in UL PL Rate QCI1while Bad Cluster shows around 0.14% in UL PL Rate QCI1. So clear difference is observed upon K-means and HC clustering.

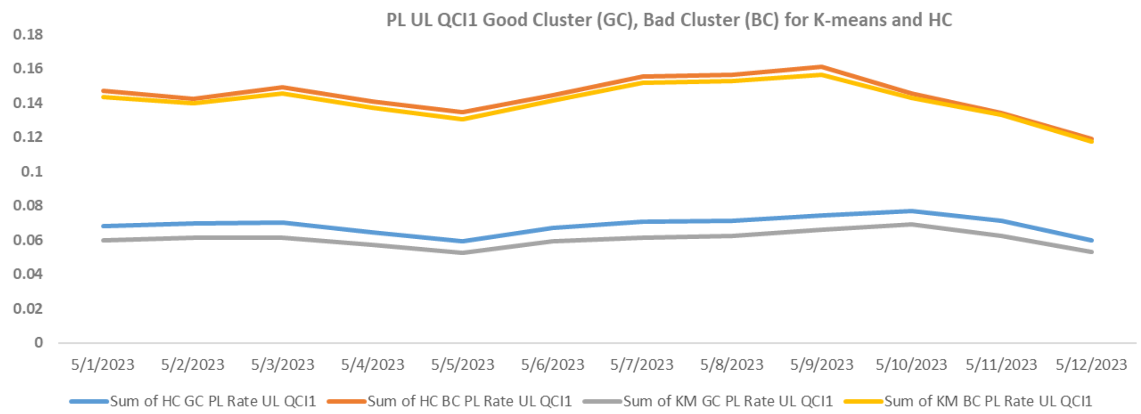

Fig. 6. Packet loss rate uplink QCI1 for K-means, HC clusters for good cluster and bad cluster

3.2.2 Uplink packet delay budget OK QCI1 % (ULPDB_QCI1)

Packet delay refers to the time it takes to transfer any packet from one point to another. In this KPI, we measure the percentage of UL packets that have been received within the packet delay budget and mark them as OK packets as shown in Figure 7. Good Cluster (GC) shows around 101.05% in ULPDB_QCI1 while Bad Cluster shows around 100.9% in ULPDB_QCI1. Note: If Robust header compression is active sometimes, we may have this ratio is higher than 100% and this is the case here in our cluster.

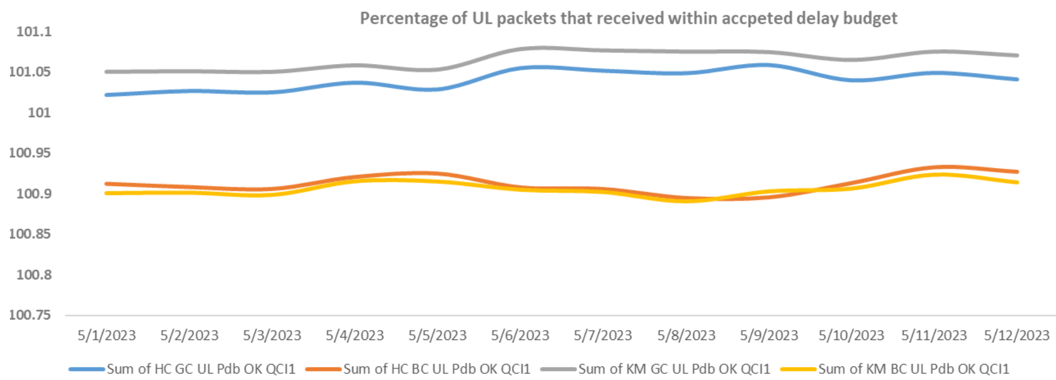

Fig. 7. ULPDB_QCI1 % for K-means, HC clusters for good cluster and bad cluster

3.3 Results Verification

We have taken an action on a cluster of cells which suffer badly from low VoLTE quality. The main target of this action is avoiding VoLTE calls to continue while the UL packet loss of QCI1 is higher than a certain threshold. Once the threshold of losses reached the system forces the call to move out of current frequency band which causes enhancement in UL packet loss for QCI1.

The action has been taken in 25th of October and we had a great result. As seen in Figure 8 we got a huge enhancement in UL Packet loss QCI1 from almost 3.5% to 2% only.

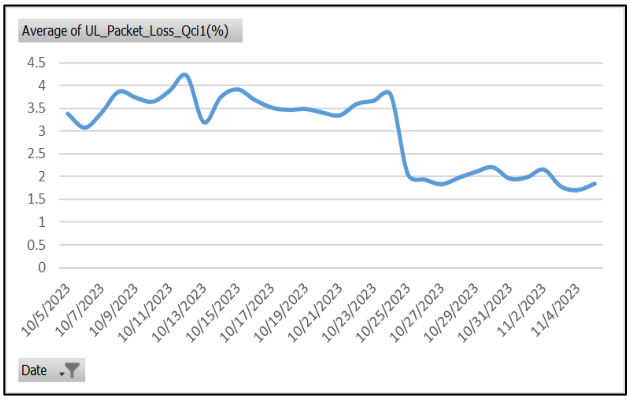

Fig. 8. UL packet loss QCI1 showing the effect of the taken action on its enhancement

Also, in Figure 9 we had a great enhancement in ULPDB_QCI1 from almost 98% to 99.5%.

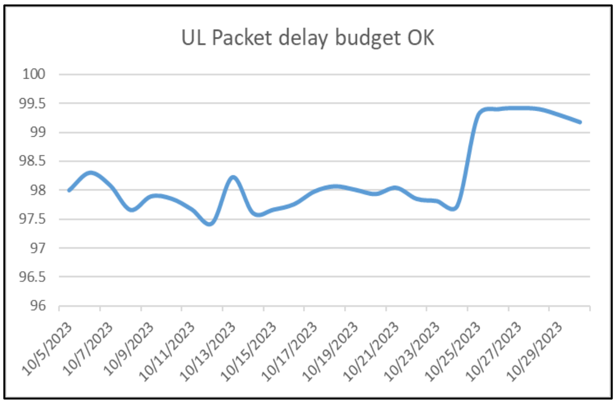

Fig. 9. UL delay budget OK QCI1 showing the effect of the taken action on its enhancement

Hence, the overall VoLTE QoS has attained a remarkable improvement as shown in Figure 10 from almost 60% till 75% which confirms our approach clearly.

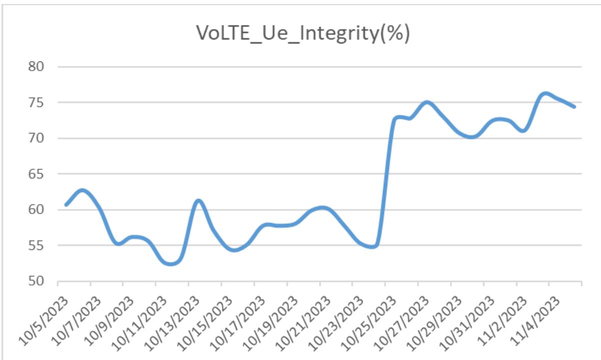

Fig. 10. VoLTE quality enhancement showing the effect of the taken action after enhancing the selected KPIs

4. Conclusion

This paper discusses VoLTE quality evaluation of a real LTE network using unsupervised ML techniques. Using a real network data and clustering algorithms were applied in addition to correlation with VoLTE quality to be used as proposed methodology of this work. The aim of using this methodology is to determine the main affecting KPIs in VoLTE QoS. Upon identifying those KPIs, we confirmed the results by applying some conditions which have a clear target to enhance the UL packet loss of QCI1 which in turn resulted in a clear gain in VOLTE quality which is our main aim exactly.

Acknowledgement
This research was not funded by any grant.

References

[1] Ericsson White Paper GFMC-284 23-3163 Uen Rev D July 2022 Voice and communication services in 4G and 5G networks
[2] Rix, Antony W., Michael P. Hollier, John G. Beerends and Andries P. Hekstra. "PESQ-the new ITU standard for end-to-end speech quality assessment." In *Audio Engineering Society Convention 109*. Audio Engineering Society, 2000.
[3] Hines andrew, Jan Skoglund, Anil Kokaram and Naomi Harte. "Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA." In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3697-3701. IEEE, 2013. https://doi.org/10.1109/ICASSP.2013.6638348
[4] Santos, Rui, Marco Sousa, Pedro Vieira, Maria Paula Queluz and Antonio Rodrigues. "An unsupervised learning approach for performance and configuration optimization of 4G networks." In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/WCNC.2019.8885524
[5] ELWakiel, Mohamed, Hesham ELBadawy and Hadia Elhennawy. "Design of quality of service parameters for voice over long term evolution "LTE" network." *International Journal of Sciences: Basic and Applied Research (IJSBAR)* 28, no. 2 (2016): 107-125.
[6] Vizzarri, Alessandro. "Analysis of VoLTE end-to-end quality of service using OPNET." In *2014 European Modelling Symposium*, pp. 452-457. IEEE, 2014. https://doi.org/10.1109/EMS.2014.68
[7] Nguyen, Duy Huy. "Enhancing and improving voice transmission quality over LTE network: Challenges and Solutions." PhD diss., Institut National des Télécommunications, 2017.
[8] Rivas, F. Javier, Almudena Díaz and Pedro Merino. "Obtaining More Realistic Cross-Layer QoS Measurements: A VoIP over LTE Use Case." *Journal of Computer Networks and Communications* 2013, no. 1 (2013): 405858. https://doi.org/10.1155/2013/405858
[9] Lipovac, Adriana, Vlatko Lipovac, Ivan Grbavac and Ines Obradović. "Practical cross-layer testing of HARQ-induced delay variation on IP/RTP QoS and VoLTE QoE." *EURASIP journal on wireless communications and networking* 2021, no. 1 (2021): 83. https://doi.org/10.1186/s13638-021-01970-y
[10] Tabany, Myasar R. and Chris G. Guy. "An end-to-end QoS performance evaluation of VoLTE in 4G E-UTRAN-based wireless networks." In *the 10th International Conference on Wireless and Mobile Communications*. 2014.
[11] Vetoshko, Ivan and Vyacheslav Noskov. "Evaluation of voice transmission quality in the lte networks." *Information and Telecommunication Sciences* 2 (2021): 22-26. https://doi.org/10.20535/2411-2976.22021.22-26
[12] Vyas, Purva, Sanjay Prasad, Shruti Mada and Yohannes Z. Tafesse. "A Comparative Study of Voice Quality and Coverage for Voice over Long Term Evolution Calls Using Different Codec Mode-sets."
[13] Adhilaksono, Bramantyo and Bambang Setiawan. "A study of Voice-over-Internet Protocol quality metrics." *Procedia Computer Science* 197 (2022): 377-384. https://doi.org/10.1016/j.procs.2021.12.153
[14] Sheugh, Leily and Sasan H. Alizadeh. "A note on pearson correlation coefficient as a metric of similarity in recommender system." In *2015 AI & Robotics (IRANOPEN)*, pp. 1-6. IEEE, 2015. https://doi.org/10.1109/RIOS.2015.7270736
[15] Emond, Edward J. and David W. Mason. "A new rank correlation coefficient with application to the consensus ranking problem." *Journal of Multi-Criteria Decision Analysis* 11, no. 1 (2002): 17-28. https://doi.org/10.1002/mcda.313
[16] Arsov, Nino, Milan Dukovski, Blagoja Evkoski and Stefan Cvetkovski. "A measure of similarity in textual data using Spearman's rank correlation coefficient." *arXiv preprint arXiv:1911.11750* (2019).

[17] Othman, Wan Roslina Wan, Syahrul Fahmy Abdul Wahab, Nurul Haslinda Ngah, I. I. A. Halim and S. S. S. A. Mutalib. "Factors Influencing the Selection of Private Higher Education in Malaysia: Spearman Rank Correlation." *J. Adv. Res. Appl. Sci. Eng. Technol* 41 (2024): 103-112. https://doi.org/10.37934/araset.41.1.103112
[18] Cao, Dan, Yuan Chen, Jin Chen, Hongyan Zhang and Zheming Yuan. "An improved algorithm for the maximal information coefficient and its application." *Royal Society open science* 8, no. 2 (2021): 201424. https://doi.org/10.1098/rsos.201424
[19] Edelmann, Dominic, Tamás F. Móri and Gábor J. Székely. "On relationships between the Pearson and the distance correlation coefficients." *Statistics & probability letters* 169 (2021): 108960. https://doi.org/10.1016/j.spl.2020.108960
[20] Kitasuka, T., Masayoshi Aritsugi and F. Rahutomo. "Semantic Cosine Similarity." In *The 7th International Student Conference on Advanced Science and Technology ICAST*. 2012.
[21] Lahitani, Alfirna Rizqi, Adhistya Erna Permanasari and Noor Akhmad Setiawan. "Cosine similarity to determine similarity measure: Study case in online essay assessment." In *2016 4th International conference on cyber and IT service management*, pp. 1-6. IEEE, 2016. https://doi.org/10.1109/CITSM.2016.7577578
[22] Hong, Foong Kin, Mohd Arfian Ismail, Nur Farahaina Idris, Ashraf Osman Ibrahim and Shermina Jeba. "Content-Based Filtering Technique using Clustering Method for Music Recommender Systems."
[23] Ding, Chris and Hanchuan Peng. "Minimum redundancy feature selection from microarray gene expression data." *Journal of bioinformatics and computational biology* 3, no. 02 (2005): 185-205. https://doi.org/10.1142/S0219720005001004
[24] Zhao, Zhenyu, Radhika Anand and Mallory Wang. "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform." In *2019 IEEE international conference on data science and advanced analytics (DSAA)*, pp. 442-452. IEEE, 2019. https://doi.org/10.1109/DSAA.2019.00059
[25] Xu, Rui and Donald Wunsch. "Survey of clustering algorithms." *IEEE Transactions on neural networks* 16, no. 3 (2005): 645-678. https://doi.org/10.1109/TNN.2005.845141
[26] Azad, Abdus Samad, Mudassar Iqbal, Mudasar Zafar, Atta Ullah, Musarat Elahi and Ahmad Abubakar Suleiman. "Exploring K-Means Clustering Efficiency: Accuracy and Computational Time across Multiple Datasets."
[27] Ismail, Puteri Zalikha Simya; Junoh, Ahmad Kadri; Wan Muhamad, Wan Zuki Azman; Mashor, Mohd Yusoff; Hassan, Rosline; Zulkeflee, Razan Hayati and Murnawan, Hery . “A Comprehensive Performance Comparison of Thresholding and the K-Means Clustering Algorithm in White Blood Cells Segmentation.” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 55(2), (2024): 203–213. https://doi.org/10.37934/araset.55.2.203213
[28] Carreira-Perpinán, Miguel A. "A review of mean-shift algorithms for clustering." *arXiv preprint arXiv:1503.00687* (2015).
[29] Anand, Saket, Sushil Mittal, Oncel Tuzel and Peter Meer. "Semi-supervised kernel mean shift clustering." *IEEE transactions on pattern analysis and machine intelligence* 36, no. 6 (2013): 1201-1215. https://doi.org/10.1109/TPAMI.2013.190
[30] Reddy, S. Nagarjuna, S. Sai Satyanarayana Reddy and M. Babu Reddy. "Enhanced Expectation–Maximization Clustering through Gaussian Mixture Models." *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 8, no. 8 (2019).
[31] Dempster, Arthur P., Nan M. Laird and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal statistical society: series B (methodological)* 39, no. 1 (1977): 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
[32] 3GPP TS 36.314 V13.1.0 (2016-03) : “Technical Specification Group Radio Access Network.”