IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

pDate of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

A Hybrid Ensemble-Based QoE Prediction from QoS and User Satisfaction Data: A case study on Cameroon's 3G/4G Mobile Networks

Kinge Mbeke Theophane Osee[1], Valery Nkemeni[1], Michael Ekonde Sone[2], Godlove Suila Kuaban[3]

[1]Faculty of Engineering and Technology, Department of Electrical and Electronic Engineering, University of Buea, P.O. Box 63, Buea, Cameroon

[2]College of Technology, Department of Electrical and Electronic Engineering, University of Buea, P.O. Box 63, Buea, Cameroon

[3]Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Baltycka 5, 44-100 Gliwice, Poland

Corresponding author: Kinge Mbeke Theophane Osee (email: kinge.mbeke@yahoo.com)

ABSTRACT The rapid growth of mobile internet and social media usage in Sub-Saharan Africa has amplified the need for accurate Quality of Experience (QoE) assessment in resource-constrained network environments. This paper introduces a novel hybrid Machine Learning (ML) framework for predicting user QoE in Cameroon's Third Generation and Fourth Generation (3G/4G) networks, leveraging a unique real-world dataset that integrates crowdsourced Quality of Service (QoS) measurements with subjective user satisfaction surveys. Addressing the limitations of existing QoE studies that focus on well-resourced regions, our approach proposes *QoE Predictor (QoEPredict)*. This stacking ensemble combines eXtreme Gradient Boosting (XGBoost) and Random Forest classifiers with an XGBoost meta-learner. A key innovation is the use of disagreement features to capture divergences between base model predictions, allowing the meta-learner to resolve conflicts and enhance predictive accuracy. The proposed five-stage pipeline incorporates data preprocessing, feature engineering via Uniform Manifold Approximation and Projection (UMAP), unsupervised clustering, and Bayesian hyperparameter optimisation using Hyperopt, ensuring a robust and transferable methodology. Explainable AI (XAI) is integrated through SHapley Additive exPlanations (SHAP) analysis to provide feature-level interpretability and actionable insights for network operators. An experimental evaluation of 1,934 user sessions demonstrates that *QoEPredict* achieves a 90% F1 Score and accuracy, outperforming single-model baselines across all metrics. This work represents one of the first large-scale, interpretable QoE prediction frameworks for mobile social media applications in Sub-Saharan Africa. By combining ensemble modelling with explainability and contextualised insights, the study offers both methodological advances and practical guidance for implementing QoE-aware network management strategies in developing regions facing infrastructural and operational constraints.

INDEX TERMS Hybrid Ensemble Model, Mobile Networks, Machine Learning, Quality of Experience (QoE), Quality of Service (QoS), eXplainable Artificial Intelligence (XAI)

I. INTRODUCTION

The last decade has witnessed a surge in global mobile internet usage, with social media applications being the primary driver of data traffic. Like many Sub-Saharan African countries, Cameroon has seen a rapid growth in mobile connectivity. By early 2025, there were 12.4 million internet users (41.9% penetration), and 5.45 million active social media users in the country. Most of these users use these services via Third Generation and Fourth Generation (3G/4G) cellular networks, reaching roughly 70% of the population [1]. In this landscape, ensuring a high Quality of Experience (QoE) for mobile apps is both a commercial imperative and a technical challenge. QoE, which is defined as the overall level of user satisfaction or dissatisfaction with an application or service, as perceived subjectively by the end-user [2], is a holistic metric that expands upon network Quality of Service (QoS) indicators by incorporating the user's end-to-end perspective, including human factors, and context.

In practical terms, good QoS, e.g., high throughput, is ideal but not always sufficient to guarantee good QoE as expectations, application design, and individual perception all influence user satisfaction. Research has shown that QoS alone is insufficient to capture the subjective dimension of the user experience. Instead, QoE assessment must integrate technical metrics with human factors like user expectations, context of use, etc [3]. In this light, recent works have started

adopting Machine Learning (ML) to predict users' QoE. For example, Panahi et al. [4] presented ML-based QoE prediction frameworks that achieved high accuracy (~95%) for video streaming services. However, these frameworks largely focused on controlled or well-resourced environments, highlighting the potential of data-driven QoE prediction. Also, these frameworks reveal current limitations regarding real-world QoE modelling, especially in developing regions.

Assessing and improving QoE in Cameroonian mobile networks, like other African mobile networks, remains challenging due to limited local research. Regional factors such as frequent power outages, legacy infrastructure, limited bandwidth, and inconsistent coverage exacerbate these challenges [5]. A glance through the social media handles of local mobile operators further underscores the need for studies like this. These platforms reveal a daily influx of subscriber complaints and widespread dissatisfaction with the services provided. A phenomenon that has drawn government scrutiny and hefty fines for the operators involved [6]. These subscriber complaints underscore the practical importance of QoE. Poor QoE directly affects both users and businesses, highlighting the need for more effective QoE monitoring tools in this context.

A. MAIN CONTRIBUTIONS OF THE PAPER

We introduce a novel hybrid QoE prediction framework tailored to an African 3G/4G scenario with social media usage as the use-case, validated on a real-world dataset. To our knowledge, this is the first known ML-based QoE study in this setting, using a hybrid approach that combines crowdsourced network QoS data and user satisfaction surveys with explainable Artificial Intelligence (XAI). Our approach introduces a stacking ensemble, named QoE Predictor (QoEPredict), that combines eXtreme Gradient Boosting (XGBoost) and Random Forest classifiers with an XGBoost meta-learner. We not only fed the meta-learner with the base models' contributions but also with disagreement features. This permitted the ensemble to capture any conflicting base model predictions and resolve them via learned compensatory patterns. Hyperparameter optimisation was performed using the Hyperopt Bayesian engine to find the best configurations for all models, ensuring a fair and optimised comparison. In addition, we incorporate model interpretability via SHapley Additive exPlanations (SHAP), enabling us to explain the predictions of the ensemble in terms of feature importance and contributions.

This hybrid of high-performance ensemble modelling with XAI is a novel contribution in the QoE domain. This paper makes several significant contributions to the field of QoE prediction, particularly within the context of African 3G/4G networks and social media usage:

1. **Hybrid QoE Prediction Framework for an Understudied Region:** We propose the first known ML-based QoE prediction framework tailored to the Cameroonian 3G/4G environment, combining crowdsourced QoS measurements from user devices with subjective user satisfaction surveys. This provides unique insights into QoE within an underrepresented market.
2. **Novel Stacking Ensemble with Disagreement Features:** We introduce *QoEPredict*, a hybrid ensemble model that integrates XGBoost and Random Forest classifiers with an XGBoost meta-learner. Unlike conventional stacking, our approach incorporates *disagreement features*, quantifying divergences between base model predictions, to allow the meta-learner to resolve conflicting outputs through learned compensatory patterns, thereby enhancing prediction accuracy.
3. **Modular Five-Stage Machine Learning Pipeline:** We develop a comprehensive and reusable ML pipeline comprising (i) data preprocessing, (ii) feature engineering with Uniform Manifold Approximation and Projection (UMAP), (iii) unsupervised clustering, (iv) Bayesian hyperparameter optimisation using Hyperopt, and (v) stacked ensemble modelling. This modular structure supports reproducibility and adaptability to other QoE contexts.
4. **Integration of Explainable AI (XAI):** We embed model interpretability into the framework using SHAP analysis, enabling domain experts to understand feature contributions and the decision mechanisms of the ensemble. Combined with cluster profiling, this provides actionable insights for network operators to link QoE drivers to operational strategies.
5. **State-of-the-Art Performance and Regional Relevance:** The proposed framework achieves a peak F1 Score and accuracy of 90%, outperforming single-model baselines across all evaluation metrics. Beyond performance, it offers region-specific value by uncovering QoE determinants in Cameroon, thereby supporting QoE-aware network management strategies in emerging markets.
6. **Creation of a hybrid dataset for QoE prediction:** A unique dataset from Cameroonian 3G/4G users, collected from user devices via a network measurement app and follow-up user surveys, illustrating QoE in an understudied environment.

Collectively, this work advances QoE prediction research by delivering a high-performance, interpretable, and regionally contextualised framework. It serves as a methodological contribution through the novel integration of stacking and XAI techniques. This study also provides a practical blueprint for applying predictive analytics to enhance user experience in resource-constrained mobile network environments.

B. ORGANISATION OF THIS PAPER

The remainder of the paper is structured as follows: Section II presents the related works associated with this study, and Section III depicts the methodology employed in this research. Section IV presents the interpretable insights and performance results of the models. This is followed by a discussion of the implications of these findings, the limitations of this study, and the constraints faced during this research in Section V. The paper concludes with Section VI,

highlighting the significance of this study within the African context.

II. RELATED WORKS

Globally, early works on QoE focused on subjective measurement techniques such as user surveys and Mean Opinion Score (MOS) experiments to map QoS metrics to perceived QoE [7]. While these methods were regarded as the "ground truth" for user experience, they were labour-intensive, costly, and lacked scalability for large-scale deployment. In response, research shifted toward objective ML-based QoE prediction models trained on network performance indicators. In this regard, Alreshoodi and Woods [8] provided a comprehensive review of such objective and subjective QoS-to-QoE mapping efforts. They observed that although many models exist, each only partially addresses the challenge of robust, real-world QoE prediction. This study emphasised the importance of integrating subjective measurements with objective metrics to develop more reliable and comprehensive hybrid QoE models.

Building on this, Casas et al. [9] compared single-model predictors to ensemble methods using smartphone-collected QoE data and found that while decision-tree-based models, e.g., RF, performed well individually, ensemble methods yielded superior predictive accuracy. These ensemble approaches, including bagging, boosting, and stacking, enhance performance by aggregating outputs from multiple base learners. Stacking, in particular, is advantageous because it allows the integration of heterogeneous base models, leveraging their complementary strengths for improved generalisation and stability [10]. In parallel, explainable ML techniques have gained attention in QoE research; methods such as SHAP are now being used to elucidate how specific features impact user-perceived QoE [11]. These tools enhance the interpretability and trustworthiness of ML-based QoE systems, especially in operational or regulatory contexts [12].

At the national level in Cameroon, research on QoE has followed a similar evolutionary trajectory, albeit with distinct local challenges. Molem et al. [13] analysed the impact of technological innovations on customer satisfaction using survey data from 363 long-term MTN subscribers in Buea. This study employed descriptive statistics and classification analysis to examine the relationship between the rollout of 3G and 4G networks and user satisfaction and loyalty. However, it relied solely on subjective perceptions and lacked integration with network KPIs or predictive modelling. Expanding on this, Kum and Austin [14] proposed a theoretical framework called QoE-Incorporation Feedback Mechanism (QoE-IFM), which combined technical (e.g., optimal network coverage), regulatory (e.g., compliance with Net Neutrality), and business (e.g., cost), dimensions into a mathematical model. While their approach provided a high-level view of QoE integration within network operations, it remained conceptual. It did not utilise granular user-level metrics or field data from end-user devices.

The most technically advanced work within the Cameroonian context is that of Abana et al. [15], who developed an ML and Deep Learning-based platform for predicting customer satisfaction at Orange Cameroon (OCM). Their model was trained on internal KPIs such as Call Success Rate, SMS Hit Rate, TCP Session Counters, and MOS, supplemented by a very small-scale internal user satisfaction survey of OCM's Customer Experience Department employees. Their system was limited to internal network logs and lacked external validation via large-scale crowdsourced data. Moreover, interpretability and actionable feedback mechanisms were absent. Together, these works reflect a logical progression: from survey-based perception studies to theoretical modelling, and recently, to predictive QoE modelling using network KPIs.

However, a critical gap remains: the lack of an integrated framework that combines large-scale crowdsourced QoS data from user devices with structured user-rated QoE feedback, supported by explainable and optimisable ML architectures. Hence, the following questions: how can we accurately predict users' QoE in Cameroon's mobile networks, with a focus on social media applications, by combining large-scale crowdsourced network performance metrics with user feedback? Which factors have a greater influence on user experience in this local context?

III. METHODOLOGY

To achieve our objectives, we followed a 4-step methodology encompassing data collection, data fusion, modelling, and evaluation. An overview of the process is depicted in Fig. 1.

A. DATA COLLECTION

Over 3 months, we gathered data points from volunteer smartphone users of all the local networks across 6 regions of Cameroon. The South West Region (Buea, Limbe), the Centre Region (Yaoundé), the Littoral Region (Douala), the East Region (Bertoua), the North Region (Garoua), and the North West Region (Bamenda) are the towns with the most participants. Each data point consists of: (a) Objective network performance features collected via the *SpeedTest Master Pro* application, and (b) Subjective user satisfaction feedback collected via a *Google Forms* questionnaire. The *SpeedTest Master Pro* app; a mobile app used for measuring QoS metrics over WI-FI and mobile networks, was chosen due to its ease of use and ability to capture the network QoS metrics of interest for our case.

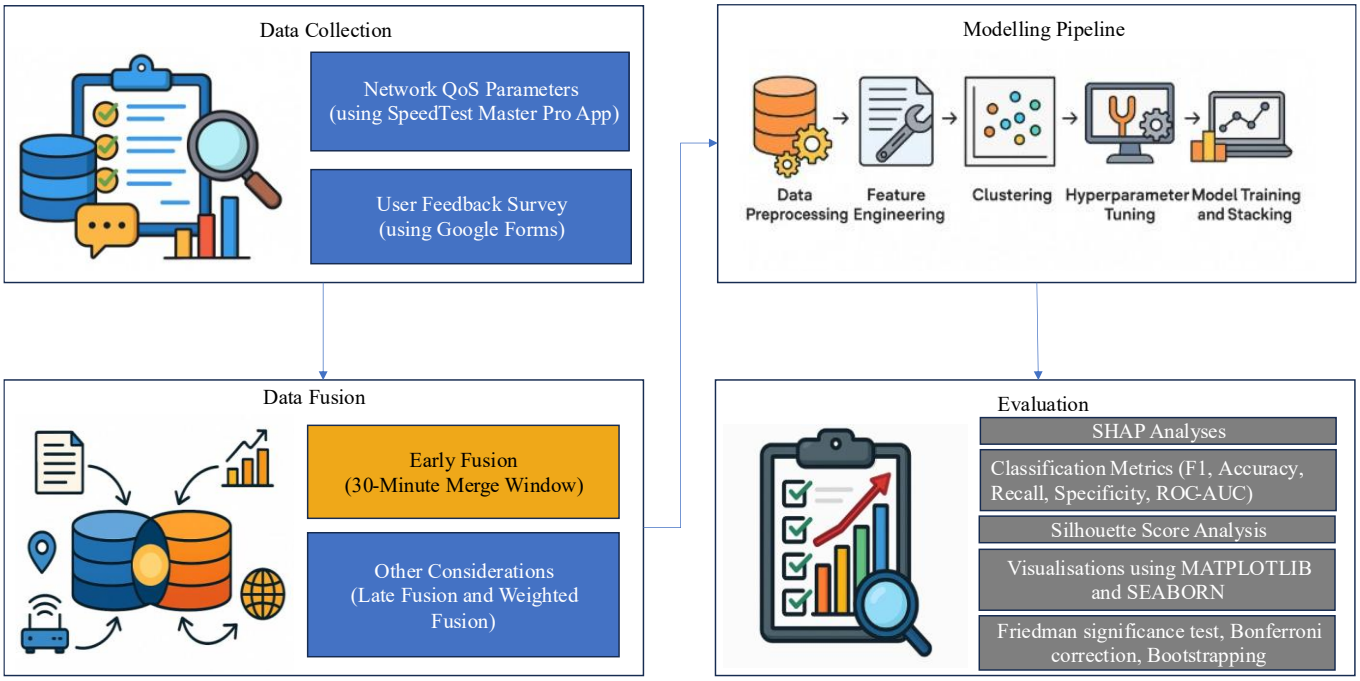


FIGURE 1. Four-step research method outlining the workflow from data collection, fusion, model development and performance evaluation of our developed model

The QoS metrics measured on the user's mobile device using the SpeedTest Master Pro app included: Downlink throughput (Mbps), Uplink throughput (Mbps), Ping/Latency (ms), Jitter (ms), Packet Loss (%), and Device (Operating System) type, along with test timestamp. These were automatically logged on the user's device after executing the network tests under everyday conditions, thus reflecting realistic network performance. The *Google Forms* survey was administered immediately after each network Speedtest to capture the user's subjective satisfaction ratings and provide context like the primary app in use, e.g., Facebook, Content type, e.g., text messaging, any observed issues, e.g., video buffering, location, and time of day. Basic user demographics, e.g., age and technical literacy, network type, i.e., 3G or 4G, were also recorded to enrich the analysis. Since the app measures both mobile and WI-FI network ratings, we filtered out any WI-FI readings from the app's metadata. We retained only records where the app's metadata field for network type explicitly indicated a mobile network connection (Mobile, LTE, 4G, 3G). This systematic Wi-Fi exclusion ensures the dataset focuses solely on mobile network QoE measurements. Each data entry was accompanied by a user-reported QoE score, a MOS-style rating on a 1-5 scale. This served as the ground truth target for our predictive models regarding social media app usage experience at that moment, presumably influenced by the network performance they just experienced.

We also collected users' actual phone numbers as user IDs to serve a dual purpose: first, to ensure we were collecting data from real users of local mobile networks, and second, to use them as the pivot for data fusion. Also, these user IDs could be used to identify different network operators, especially for studies that would be interested in evaluating the QoE of particular networks. All users who participated in this data collection did so willingly and gave their informed consent by signing a consent form for their data to be used anonymously for the aggregate analysis. To ensure anonymity, we hashed the user IDs in both QoS and the User feedback files using SHA3-256. The hashed identifiers were stored on a secure, access-controlled system. All raw data were retained only for the duration of model development before being permanently deleted, in compliance with General Data Protection Regulation (GDPR) standards.

B. DATA FUSION TECHNIQUE

For the scope of this paper, we used early fusion as the baseline fusion technique to merge all collected features into a single dataset. Early fusion simplified our pipeline and is a common approach in related QoE prediction works. Early fusion means that we merged the network QoS and user data from the very beginning, before feeding it into the ML models [16]. Concretely, we treated each QoE score as the dependent variable and all corresponding QoS and user feedback metrics as independent variables in a single feature vector. This way, the model directly learns the mapping from technical parameters plus context to the user's QoE. E.g., our model could discover if *"high throughput is only beneficial to QoE when latency is also low",* by examining those features together. All data lived in one table, and standard classification algorithms could be applied directly.

To achieve a successful merge, we implemented the following precautions: Firstly, since QoS data files came from different devices with possibly different versions of excel/csv and also since there is the possibility of having

special characters in the user feedback file, we encoded the compiled network QoS and user feedback data files using ISO 8859-1 encoding to guarantee uniformity. Secondly, we cleaned and forced the User ID on both files to strings to eliminate any spaces or '-'s between user ID digits, which may lead to possible mismatches. Thirdly, we harmonised the column names for uniformity. Also, we harmonised the Timestamps on both files to the 'dd/mm/yyyy hh:mm' format. Then, we fused both files, using the User ID (one entry per user ID) and the nearest Timestamp within a 30-minute tolerance window, into one dataset. This strict deduplication process guarantees that each entry is linked to a unique user experience rating. The 30-minute window was necessary to balance user convenience with our data fusion needs. However, it may introduce noise if a user's network conditions change significantly between the speed test and survey completion. Future work employing real-time triggering or continuous monitoring could further enhance label fidelity. Next, we removed all users' personal identifiable information to anonymise the dataset, adhering to relevant ethical GDPR guidelines. Finally, we cleaned the dataset for outliers, e.g., we filtered extremely low speeds likely due to test errors. We also dropped entries with no corresponding user survey responses.

The final merged dataset (1934, 38), which served as the basis for training and evaluating our hybrid QoE prediction models, has 1934 user-collected data points (from 1934 different users) with 38 feature columns and the QoE target variable. We selected and designed these 38 features to represent 14 of the most documented categories of factors influencing QoE in the literature. Table I below presents the taxonomy of these QoE influencing factor categories and their references, the dataset features that represent these categories, and brief descriptions of these features.

TABLE I. Taxonomy of QoE influencing factors from literature captured in the hybrid dataset features (adapted from references [17] - [18])

Dataset Feature	Feature Description	QoE Influence Factor Category
Improve Experience Increase Use	Likelihood of increased usage if QoE improves	User Behaviour/Preference [17] [18]
Platform Loyalty	Willingness to switch platforms	
User engagement tendency	Impact of poor performance on willingness to engage with social media content	
Usage per day	Time spent daily on social media	
Platform Preference	Social media platforms frequently used	
Content Type	Most consumed content type	Content Type [19] [20]
Congestion Impact on Experience	Impact of network congestion on social media experience	User Experience [21] [22]
Network instability	Frequency of interruptions during social media use	
Difference during Peak Hour	Noticing performance differences at peak hours	
Peak Hour Rating	Network rating during congested times	
App Speed Rating	Satisfaction with social media app speed/responsiveness	
Network Reliability	How important a good network is	
Data Throttle	Data exhaustion or throttling during congestion	
Video Loading Time	Time to start video playback	Usability [23] [24]
Message Delivery Speed	Delay in sending/receiving messages	
Notification Latency	Delay in receiving app notifications	
Video Buffer duration	How long do delays usually last	
Video Reliability	Interruption frequency while watching social media videos	
Message Reliability	Messages not being delivered or arriving out of order	
Video Buffering	Pauses during video playback	Product Quality Degradation [25] [26]
Video Quality	Experience of poor visual resolution	
Image Resolution	Blurry or failed image loads	
Image Quality	Quality of images displayed on social media platforms	
Video Stuttering	Lag in videos or live streaming while using social media apps	
Network Type	Network connection	Demographics and Environmental Context [27] [28]
Age Group	Age group of the respondent	
TimeStamp	Time of day record of data collection	
User Environment	User location	
Pay for QoE	Willingness to pay for a better experience	Price and Value [29] [30]
QoS awareness/Expectations	Belief that social media prioritisation would help	User Expectation [31] [32]
Suggestions	User opinion to improve the network	
Latency	Round-trip time for packets	Latency [33] [34]
Packet Loss	Packet loss results in content not loading	Packet Loss [32] [35]
Jitter	Variation in network delays	Jitter [36] [37]
Throughput (Downlink and Uplink)	Both download and upload speeds	Throughput [38] [39]
Bandwidth	Available Bandwidth to user	Bandwidth [36] [39]
Device type	Operating system type	Device Capability [34] [38]
Note: This categorisation is the authors' synthesis of referenced literature and not results from new experiment		

The QoE scores in our data spanned from 1 (Very poor), 2 (Poor), 3(Average), 4(Good) to 5 (Excellent), with a roughly symmetric distribution around the middle. This indicates many moderately satisfied users, coupled with a significant number of dissatisfied users, as seen in Fig. 2. The QoS metrics varied widely, e.g., latency ranged from less than 2ms up to more than 1000ms. These variations highlight the heterogeneity of network conditions captured in the country.

The combined dataset is multidimensional, comprising both numerical and categorical variables, which necessitates careful feature engineering.

To enrich our understanding of local context, we additionally asked 20 randomly selected participants an open-ended question: *"Please describe any issues or factors affecting your social media experience.", "Would you increase your social media usage if these issues or factors affecting your social media experience were resolved?"* Users mentioned issues like *"images took too long to load", "videos usually froze for a moment", "the connection was fine, no problems." or "I may enjoy spending more time on my social media."*

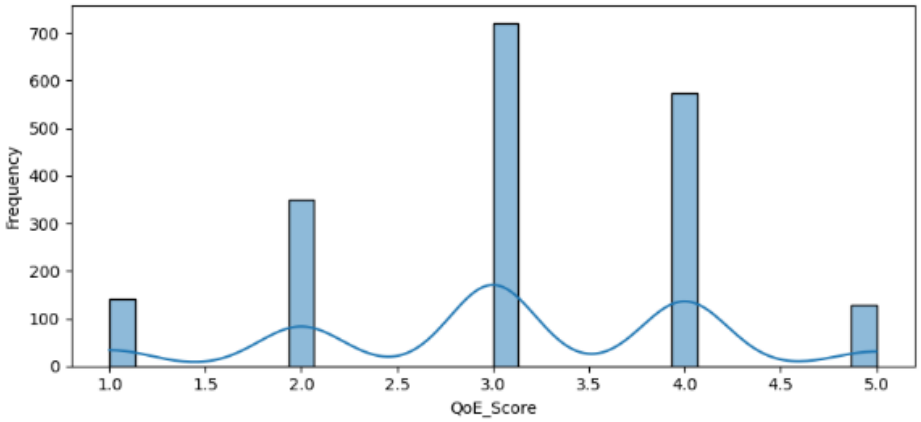


FIGURE 2. Distribution of user-reported QoE scores across the sampled population of 1934 users. The figure shows a slightly right-skewed pattern, indicating that most users experienced moderate to good perceived quality

We did not directly use these comments as inputs to the quantitative model. However, we analysed them to confirm that factors such as speed, loading time, and network stability were noticeable to users. Such insights underscore the importance of certain QoS metrics, e.g., many complaints about *"slow loading"* correlate with high latency or low throughput measurements.

It is important to note that early fusion is not the only fusion strategy; we also considered alternatives for completeness:

1) LATE FUSION

In this scheme, separate submodels are trained on different datasets. E.g., one model predicts QoE purely from objective QoS metrics, and another predicts QoE from other subjective user-factor datasets. Then, these outputs are combined via another learning layer. Late fusion is useful if the relationship within each data type is complex and distinct [40]. If we had additional subjective inputs, such as a user's qualitative sentiment, late fusion could have been applicable. As a conceptual exercise, one could imagine training one model to estimate QoE using only network QoS data and another using user survey responses about their expectations, and then fusing the two models. We recognise this as a potential extension of this work if more user-centric features become available.

2) WEIGHTED FUSION

Another approach is to explicitly weight the contributions of objective versus subjective features in the model [41]. E.g., if one suspects that network QoS explains, say, 65% of QoE and other factors explain 35%, one could adjust the input representation accordingly. This approach is more relevant in neural networks, where one could design a custom architecture, e.g., separate input layers that are later merged with certain weights. We did not implement a custom weighted fusion in our current study. We effectively let the learning algorithm determine the weights via feature importance or learned parameters.

C. MODULAR ML PIPELINE ARCHITECTURE

We implemented a five-stage pipeline as shown in Fig. 3, with each stage encapsulated in a Python module:

1) PREPROCESSING STAGE

This module handles data cleaning and normalisation. We parsed, cleaned, and transformed users' multisuggestions and multiplatform preferences into dummy variables for inclusion as categorical features. We systematically imputed missing data; less than 3% of our dataset, using column-type-aware strategies: we filled numerical fields using mean imputation via *SimpleImputer*, while we filled categorical fields using mode values with Pandas. Then, we extracted temporal information from Timestamps, yielding additional 'hour of day' and 'day of week' features. For feature transformation, we employed a *ColumnTransformer* to apply *StandardScaler* to numeric columns and *OneHotEncoder* to categorical ones. The resulting pre-processed dataset (1934, 3264), passed onward for downstream modelling, contained standardised numeric values, encoded features, temporal attributes, and a target vector.

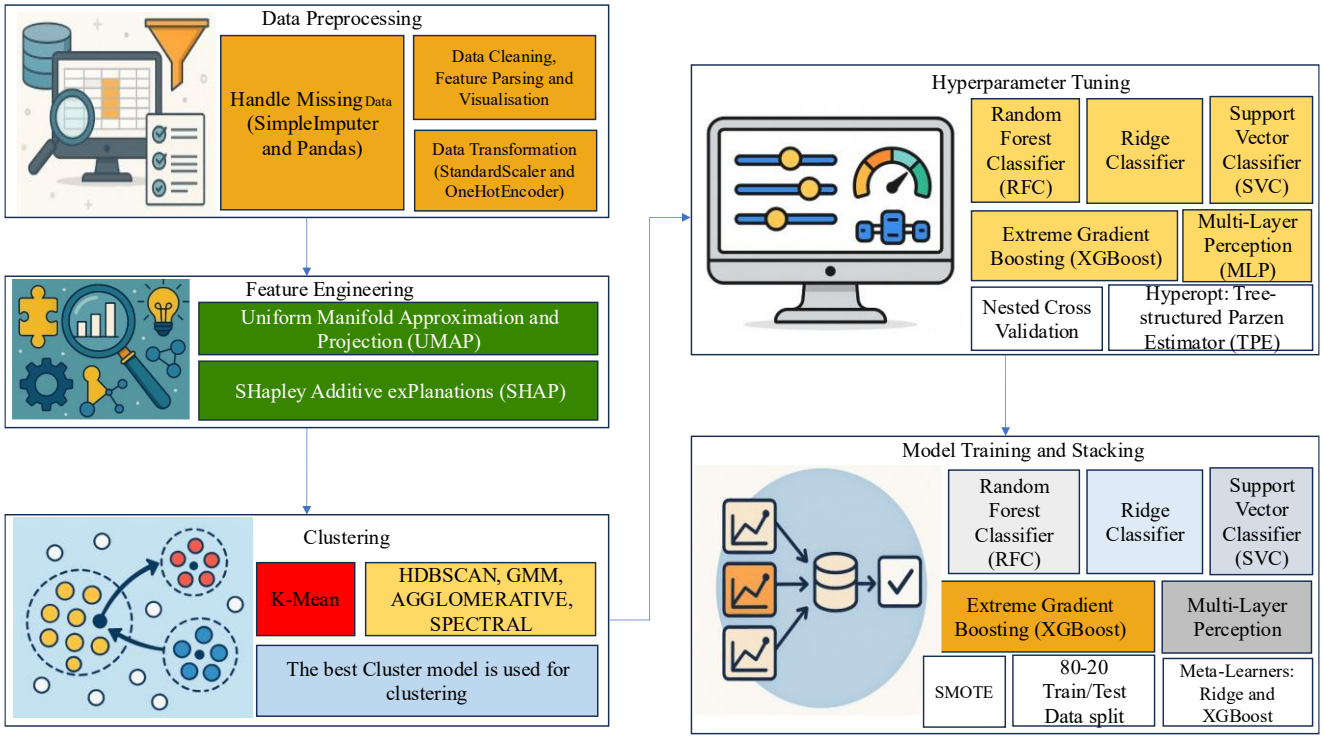


FIGURE 3. Five-stage modular ML modelling pipeline illustrating data preprocessing, feature engineering, clustering, hyperparameter tuning, and model training for QoE prediction

2) FEATURE ENGINEERING STAGE

This module performs comprehensive feature selection and dimensionality reduction to refine the pre-processed dataset for optimal model performance. To achieve a robust feature selection and avoid overfitting, we employed a 5-fold Cross-Validation (CV) feature selection procedure [42]. In each fold, features were first ranked by their Mutual Information (MI) with the QoE score, following the maximal dependency/relevance criterion [43]. An approach known to capture nonlinear QoS-QoE relationships. Then, we computed the SHAP values for feature importance; SHAP provides model-agnostic estimates of each feature's contribution to predictions [44]. By examining SHAP bar plots of mean absolute feature contributions and the Feature Heatmap, we could transparently assess which factors most strongly affect the predicted QoE. This approach addresses the core challenge of quantifying each factor's influence in telecommunications QoE models.

Next, to reduce redundancy and noise, we applied UMAP for nonlinear dimensionality reduction. UMAP projects the data into a low-dimensional embedding while preserving its intrinsic structure. UMAP was chosen over alternatives like PCA and t-SNE because it better retains global structure in the data manifold, which improves cluster separability and visualisation [45]. Since UMAP is sensitive to feature scaling, we standardised all features before applying UMAP. Table II contains the UMAP parameters used in this study.

As an exploratory exercise to justify our use of UMAP, we experimented with PCA, a well-known linear dimensionality reduction method [46]. PCA-reduced datasets failed to achieve a superior accuracy performance over our No PCA dataset as shown in Fig. 4, illustrating the nonsuitability of linear dimensionality reduction techniques for our context.

This feature engineering stage serves as a critical intermediary in enhancing data quality and interpretability before model training and stacking.

TABLE II. UMAP parameter configuration (adapted from [18])

Parameter	Value	Purpose
n_components	10 and 2	Number of dimensions for the output embedding: 10 for grouping, 2 for visualisation.
n_neighbors	20	Local neighbourhood size used for manifold approximation
min_dist	0.1	Controls how tightly UMAP packs points together; lower values preserve local structure.
Metric	Euclidean	Distance metric used to compute similarity between points.
Random State	42	Handles UMAP's stochasticity

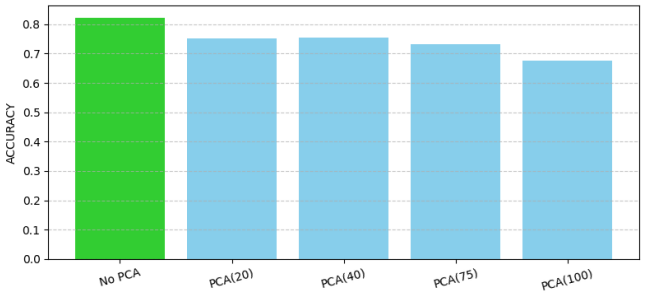


FIGURE 4. Accuracy comparison of different PCA dimensions of the dataset to investigate PCA suitability for this study

3) CLUSTERING STAGE

This module seeks to enhance explainability and provide empirical justification for the choice of clustering method. It also offers insights into how different user or network behaviour groups relate to QoE levels. The first step was to provide a robust comparative analysis of the unsupervised clustering techniques, using the feature-engineered dataset to identify natural user experience groupings. In particular, we compared partitioning methods: K-Means [47] and Gaussian Mixture Models (GMM) [48], graph-based Spectral Clustering [49], Ward's linkage hierarchical Agglomerative Clustering [50] and the density-based HDBSCAN method [51]. Evaluating a broad set of algorithms is important because each makes different assumptions about cluster shape and can reveal different structures. For example, K-Means and GMM assume roughly convex, Gaussian clusters and require the number of clusters, k, to be specified a priori. In contrast, HDBSCAN automatically determines k based on density and identifies outliers as noise.

For a fair comparison, we optimised each method's hyperparameters using internal validation metrics to obtain its best configuration, as presented in Table III. For K-Means and GMM, we varied k and applied the elbow method and silhouette analysis to choose an optimal cluster count. We used the Silhouette Coefficient as a primary internal metric. Higher silhouette values indicate more coherent and well-separated clustering. We also calculated the Davies–Bouldin (DB) Index for each result, with lower DB index values indicating low intra-cluster distance relative to inter-cluster distance [52]. Similarly, we tuned hyperparameters such as the number of neighbours in Spectral Clustering or the minimum cluster size in HDBSCAN, maximising silhouette scores or minimising density-based cluster validity indices.

TABLE III. Configurations used for tuning Clustering Algorithms (adapted from standard clustering practices [17] - [51])

Parameter	Purpose	K-Means	GMM	Spectral Clustering	Agglomerative	HDBSCAN
n_clusters	Number of clusters to form or evaluate	2–10		2–10	2–10	
n_components	Number of components (clusters/distributions)		2–10			
affinity	Similarity metric for graph construction			nearest_neighbours		
linkage	Linkage criterion				Ward	
min_cluster_size	Minimum number of samples per cluster					[15, 20, 25, 30]
min_samples	Minimum number of core samples per cluster					[5, 10, 15]
gen_min_span_tree	Hierarchical visualisation of clusters					True

Next, we used a comprehensive set of metrics and statistical tests to investigate cluster quality and stability. These include: Silhouette Score and Davies-Bouldin Index for the final clustering solutions as global indicators of cluster quality, one-way ANOVA tests on the QoE scores across clusters to ensure the clusters represent significantly different QoE populations and Kruskal-Wallis tests to account for nonnormal score distributions, adding rigour to the clustering validity. Cluster Purity was also computed to measure how well each cluster aligned with known categories [53]. More so, we computed the inter-cluster effect size for QoE differences to assess practical significance beyond statistical significance [54].

To compare the performance of the different clustering methods, we defined a composite score that aggregates different validation metrics [55]. We combined metrics emphasising different aspects: silhouette for cohesion/separation, a penalty for high DB, external validity through cluster purity, and statistical effect size by using (1). Each component is weighted to reflect its relative importance, forming a robust and interpretable composite index for ranking the clustering solutions in our context (See Table IX). This multicriteria evaluation ensured that the "best" clustering method was chosen based on a balanced consideration of internal consistency, separation, and stability.

$$\text{Composite Score} = 0.4 \times \text{Silhouette} + 0.2 \times (1-\text{Davies-Bouldin}) + 0.2 \times \text{Purity} + 0.2 \times \text{Effect Size} \quad (1)$$

Finally, for interpretability and practical insights on the clusters, we generated a suite of visualisations: Violin plots of QoE variation within clusters (see Fig. 12), Silhouette plot depicting the distribution of silhouette values per cluster to verify that most points have high within-cluster similarity and are appropriately assigned (see Fig. 13) and UMAP embedding to visually inspect how well the clusters separate in feature space (see Fig. 14), of the QoE distributions per cluster to reveal differences in central tendency and spread of user experience in each group. Additionally, to explain the cluster characteristics in terms of original features, we investigated the key QoS features influencing each cluster, effectively revealing why a data point belongs to that cluster. Thus, by combining statistical rigour: significance tests, effect size, sound engineering practices: multiple algorithms and tuning, and XAI techniques: SHAP plots and visual analytics, we yield a reproducible and transparent clustering module for QoE modelling in telecommunications. Also, we

identify distinct user experience profiles and reveal the key factors driving those differences, thereby offering both theoretical and practical value for network QoE optimisation. The final clustered dataset is forwarded downstream for supervised modelling.

4) HYPERPARAMETER TUNING STAGE

In this stage of our prediction pipeline, we employed an automated optimisation strategy using the Hyperopt library with the Tree-structured Parzen Estimator (TPE) algorithm [56], chosen for its efficiency in exploring high-dimensional, conditional search spaces. The goal was to maximise the predictive performance of six base classifiers: RF, SVC, XGBoost, Ridge, MLP, and KNN [56][57], by systematically searching for optimal parameter configurations. Table IV contains all the parameter configurations used for tuning all models. We chose the base classifiers to represent a mix of linear, nonlinear, and ensemble methods commonly used in QoE modelling. Also, we employed Synthetic Minority Oversampling TEchnique (SMOTE) to address the class imbalance in our dataset, a typical issue in real-world QoE classification tasks.

TABLE IV. Hyperparameter tuning space configurations (adapted from standard ML optimisation practices [17] - [18])

Parameter	RFC	SVC	XGBoost	MLP	KNN	Ridge	Meta-Ridge	Meta-XGBoost
n_estimators	100–400 (step 50)		100–500 (step 50)					50–300 (step 50)
max_depth	5–20 (step 5)		3–10 (step 1)					2–10 (choice index)
min_samples_split	2–10							
min_samples_leaf	1–4							
class_weight	'balanced'	'balanced'						
C		log-uniform [0.1, 10]						
gamma (γ)		log-uniform [0.001, 0.1]						
kernel		'rbf', 'linear'						
learning_rate			log-uniform [0.01, 0.3]	'constant', 'invscaling', 'adaptive'				log-uniform [0.01, 0.2]
subsample			uniform [0.6, 1.0]					
colsample_bytree			uniform [0.6, 1.0]					
hidden_layer_sizes				(50,), (100,), (100, 50)				
activation				'relu', 'tanh'				
solver				'adam', 'sgd'				
alpha (α)				log-uniform [1e-4, 1e-2]		log-uniform [0.1, 10]	log-uniform [0.01, 10]	
tol						log-uniform [1e-4, 1e-2]	log-uniform [1e-4, 1e-2]	
n_neighbors					3–15 (int)			
weights					'uniform', 'distance'			
algorithm					'auto', 'ball_tree', 'kd_tree'			
smote_k	3, 5, 7	3, 5, 7	3, 5, 7	3, 5, 7	3, 5, 7	3, 5, 7	3, 5, 7	3, 5, 7

To ensure robust performance estimates of model variability and generalisation, as summarised in Table V, we employed a nested CV approach [58]. We tuned each model across 30 trials within an outer 5-fold stratified CV loop, yielding 180 tuning evaluations per fold. Then, we selected the top three base models, ranked by F1 Score, and investigated all four possible stacking combinations. This approach reflects a balanced trade-off between computational feasibility and statistical robustness, consistent with best practices in ensemble learning research. For each stacking combination, we further tuned two types of meta-learners: Ridge Classifier and XGBoost, using an inner 3-fold CV and 15 evaluations each. This dual meta-learning approach was necessary to compare linear and nonlinear meta-modelling paradigms for our context.

TABLE V. Synthesis of total trials performed in hyperparameter tuning (adapted from [59])

Component	Per Fold Trials	Total Folds	Total Trials
Base Models	6 × 30 = 180	5	900
Meta-Learners	4 combinations × 2 × 15 = 120	5	600
Grand Total			1500

The tuning process, though a critical step in ensuring each model component within the stacking framework was configured to achieve optimal performance, was computationally intensive. Each Hyperopt trial took ~2 minutes, totalling a runtime of ~24 hours for all models. By automating this step and independently optimising each model-fold combination, we avoided the biases of manual tuning, information leakage, and ensured fair model comparison on unseen holdout data.

5) MODEL TRAINING AND STACKING STAGE

Each model was initialised using its best parameters, based on CV performance across all folds, and retrained on the entire 80% dataset. Then, we observed the test performance on the 20% validation dataset. For the proposed QoEPredict, as illustrated in Fig. 5, we implemented a two-level StackingClassifier using scikit-learn's built-in stacking framework. In Level 1, XGBoost and Random Forest as the base learners. To explicitly capture instances of base model uncertainty, we computed a binary disagreement feature on the out-of-fold base learner predictions. This feature was set to '1' if the standard deviation of the base models' class predictions for a sample was greater than zero, indicating disagreement and 0 otherwise (indicating consensus). We passed through the disagreement features to the meta-learner alongside the base models' predictions, to provide the meta-learner with more information to distinguish when the base models have conflicting predictions [59]. These disagreement features contributed to a 1.07% boost in QoEPredict's performance (an F1 increase from 88.59% to 89.66%). In Level 2, XGBoost is used as the meta-learner, leveraging its flexibility and strong performance in blending predictions from base models. Although XGBoost includes default regularisation mechanisms to mitigate overfitting, we did not explicitly tune the L1 (alpha) and L2 (lambda) regularisation parameters (to minimise computational cost) in this case. L1 regularisation can effectively perform feature selection by reducing the impact of less important features. It achieves this by adding a penalty proportional to the absolute values of the leaf weights, which encourages sparsity in the model. L2 regularisation, on the other hand, adds a penalty proportional to the square of the leaf weights. This penalty shrinks the weights and stabilises predictions, reducing variance and making XGBoost less sensitive to noisy data. While these parameters play an important role in controlling model complexity and improving generalisation, tuning them was beyond the scope of the current study. Future work could explore systematic tuning of L1 and L2 to further enhance QoEPredict's performance and mitigate potential overfitting.

.

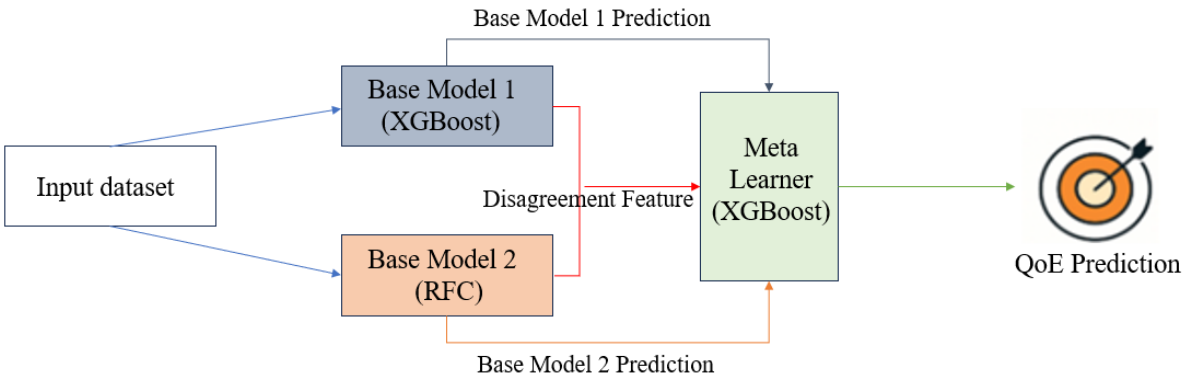


FIGURE 5. Architecture of the QoEPredict ensemble framework. The model combines XGBoost and Random Forest Classifier(RFC) and as base learners, whose outputs feed into a meta-level XGBoost model for final QoE prediction. Disagreement-based features are incorporated to improve robustness, and SHAP-based explainability is integrated for model transparency

The model's performance relied solely on boosting and its ability to capture complex relationships, with regularisation potentially applied but not fully optimised. We recognise this regularisation parameter tuning as a future direction to possibly improve the model's performance.

The meta-learner was trained on the outputs of Level 1 models for each training example. This stacked model, dubbed QoEPredict, essentially forms a hybrid function approximator that leverages the strengths of each base model.

6) MATERIALS USED

We configured the development environment using Python 3.13.2 and Visual Studio Code (v1.100.0, Electron 34.5.1, Node.js 20.19.0) on a 64-bit Windows 10 (Build 26100) system. For model implementation, the main libraries used include pandas 2.2.3, NumPy 2.1.3, scikit-learn 1.6.1, XGBoost 2.1.4, and SHAP 0.47.1 for model explainability. We performed hyperparameter tuning using Hyperopt 0.2.7, and data visualisations were carried out using Matplotlib 3.10.1 and Seaborn 0.13.2. The integrated environment leveraged the Chromium 132.0.6834.210 engine and V8 JavaScript engine v13.2.152.41 for rendering and extension support. We used random state 42 for all model configurations and experiments in our pipeline. Table VI contains the libraries and tools used for this study, with their versions to aid reproducibility.

TABLE VI. Materials used in this study

Library/Tool	Version
Python	3.13.2
Pandas	2.2.3
Numpy	2.1.3
Scikit-learn	1.6.1
Xgboost	2.1.4
Matplotlib	3.10.1
Seaborn	0.13.2
Shap	0.47.1
Hyperopt	0.2.7
VSCode	1.100.0
OS	Windows 10 x64 (Build 26100)
Processor	Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz, 2101Mhz, 4 Core(s), 8 Logical Processor(s)
Node.js (VSCode)	20.19.0
Electron (VSCode)	34.5.1
Dataset	Cameroon Hybrid QoS/QoE Dataset [60]

D. EVALUATION

We evaluated QoEPredict's prediction performance against the six other models using standard classification metrics, using (2) to (7), to capture different aspects of predictive accuracy and reliability. We used accuracy as a general indicator of correct predictions across all classes. Accuracy can be inflated by large QoE classes due to an imbalance in classes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP+FN}} \quad (2)$$

Where: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

However, acknowledging the imbalance in class distribution, we employed weighted precision, recall, and F1 Score, which account for class frequency, thereby ensuring fair performance assessment across all classes. We used F1 Score as the primary metric for model comparison, as it provides a harmonic balance between precision and recall. Unlike accuracy, the weighted F1 Score provides a more robust measure of a model's overall classification performance, especially for correctly identifying both satisfied and dissatisfied users in QoE prediction.

$$\text{Precision}_{\text{weighted}} = \frac{\sum_{s=1}^{C} \frac{\text{TP}_s}{\text{TP}_s + \text{FP}_s}(\omega_s)}{\sum_{s=1}^{C} \omega_s} \; ; \; \omega_s = \frac{n_s}{n} \quad (3)$$

Where: n_s = *number of samples in class s, n = total number of samples*

$$\text{Recall}_{\text{weighted}} = \frac{\sum_{s=1}^{C} \frac{\text{TP}_s}{\text{TP}_s + \text{FN}_s}(\omega_s)}{\sum_{s=1}^{C} \omega_s} \quad (4)$$

$$\text{F1 Score} = \frac{2 \text{ x Precision x Recall}}{\text{Precision + Recall}} \quad (5)$$

To measure the model's capability to correctly reject negative instances, we computed macro-averaged specificity, which evaluates true negative rates for each class and averages them, making it especially informative in multiclass imbalanced scenarios.

True Negative Rate per class:

$$\text{Specificity}_s = \frac{\text{TN}_s}{\text{TN}_s + \text{FP}_s} \quad (6)$$

Macro-average over all classes:

$$\text{Specificity}_{\text{macro}} = \frac{1}{C} \sum_{s=1}^{C} \text{Specificity}_s \quad (7)$$

Additionally, we incorporated Receiver Operating Characteristic-Area Under Curve (ROC-AUC) as a threshold-independent metric to evaluate the separability of classes, using a One-vs-Rest strategy for multiclass settings using (8). This diverse metric suite ensures that the model's performance is not only accurate but also robust, fair, and generalisable, particularly crucial for QoE modelling, where misclassifications can have varying user impact.

$$\text{ROC-AUC}_{\text{ovr}} = \frac{1}{C} \sum_{s=1}^{C} \text{AUC}_s \quad (8)$$

Where each AUC_s is for class s versus all others

We computed these metrics during CV hyperparameter tuning to select the best single models' parameters, and during holdout testing of the data not seen by the models during training. This simulates the model's performance on new users or new instances. To provide explainability that links directly to the original input features, we computed SHAP values for each base model separately using its respective explainers. Then we combined the SHAP values by averaging the absolute contributions across base models, while preserving the directional effects of XGBoost (the better-performing base learner). This approach produces a consensus feature importance ranking that guarantees the closest approximation of what the meta-learner sees. Avoiding the complex interactions of applying SHAP to the entire ensemble stacking. Also, to validate QoEPredict's performance's statistical significance, we performed a Friedman significance test ($p < 0.05$); to control the Family-Wise Error Rate (FWER) inherent in multiple comparisons, we applied a Bonferroni correction. We divided the significance threshold ($\alpha = 0.05$) by the number of comparisons (n) against the top-performing model to establish a corrected alpha ($\alpha\_corrected = 0.05/n$). This is a conservative method that ensures only strong, meaningful differences are flagged as significant. Next, we performed a bootstrap analysis ($\alpha = 0.05$, n=100 iterations). For each model compared to QoEPredict, we calculated the difference

in F1-score on 100 resampled versions of the holdout set. Then, we observed the mean difference with a 95% confidence interval (CI). We considered a difference statistically significant if the 95% CI did not cross zero. Additionally, we plotted a learning curve to check for high bias or variance and a confusion matrix to inspect any systematic prediction errors.

IV. RESULTS

In this section, we present the performance results of our models, including quantitative metrics, comparative evaluations, and visual analyses of errors and interpretations.

A. PERFORMANCE OF MODELS AND ENSEMBLES

Table VII shows the four best-performing cross-validation stacking configurations ranked by F1 Score. F1 Score is appropriate in QoE classification, where there is class imbalance, and the need for sensitivity to both satisfied and dissatisfied users cannot be captured via accuracy alone. The top-ranked model, QoEPredict, combines XGBoost and Random Forest as base learners with XGBoost as the meta-learner, achieving the highest F1 Score, 78.63%. This indicates a well-balanced trade-off between precision and recall across all QoE classes. Also, XGBoost was present in all top-performing stacks, indicating the unsuitability of a linear meta-learner like Ridge for this problem.

TABLE VII. Top 4 Stacking configurations based on cross-validation performance, ranked by F1 Score

Rank	Base Model	Meta-Learner	F1 Score (%)	Accuracy (%)
1 (QoEPredict)	XGBoost + Random Forest	XGBoost	78.63	79.03
2	XGBoost + MLP	XGBoost	75.72	75.81
3	XGBoost + MLP + Random Forest	XGBoost	75.09	75.16
4	MLP + Random Forest	XGBoost	68.91	68.93

Fig. 6 illustrates QoEPredict's performance based on F1 Score against the 6 single models on the Holdout (test set), as a surrogate for real-world performance on unseen data. QoEPredict is the top performer. Notably, among single models, XGBoost > Random Forest > KNN > MLP > SVC > Ridge, aligning with expectations that ensemble and nonlinear models do better for this complex task.

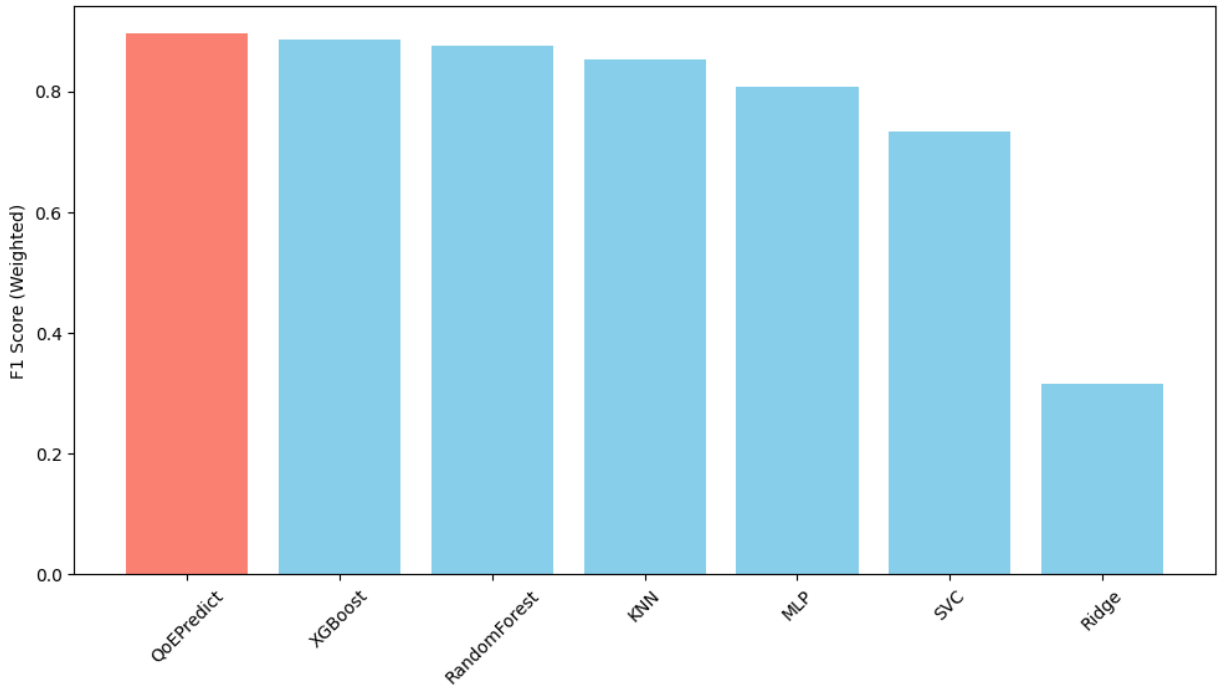


FIGURE 6. Comparison of the F1 Scores of QoEPredict and six baseline models on the holdout test set. QoEPredict outperforms all single-model baselines, demonstrating the benefit of our stacking ensemble

Additionally, as seen in Table VIII, QoEPredict outperformed every individual model on all evaluation metrics (F1, accuracy, etc.), e.g., by 22.17% F1 over SVC. While a 1-2% gain over XGBoost is competitive. In practical terms, even a 1-2% improvement in accuracy and F1 score significantly enhances QoE prediction reliability for telecom operators [61]: enabling more precise identification of user experience issues, optimised network resource allocation, proactive customer support, and improved service quality. All of which ultimately fosters higher customer satisfaction and retention in mobile network services.

TABLE VIII. Model Performance Comparison (Test Set)

Model	F1 Score (%)	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
QoEPredict	89.66	89.66	89.89	89.66	97.02
XGBoost	88.60	88.63	88.85	88.63	96.66
Random Forest	87.55	87.60	87.82	87.60	96.45
KNN	85.36	85.27	85.86	85.27	96.07
MLP	80.91	80.88	81.26	80.88	94.47
SVC	73.39	73.39	73.53	73.39	92.52
Ridge	31.57	32.04	47.30	32.04	84.10

We also confirm that our stacking is beneficial by experimenting on a plain averaging ensemble of the base models. It yielded an F1 Score of 89.3%, which is lower than the 89.66% F1 Score of our stacked model. Thus, the meta-learning approach effectively learned the optimal weighting and interaction of base model outputs. This aligns with findings in other studies that stacking can outperform weighted averaging [62]. Additionally, the Friedman test results (Statistic: 42.6873, p-value = 0.0000) confirmed that QoEPredict's performance is statistically significant and not due to chance. The Bootstrapping results further revealed that QoEPredict's performance was significantly superior compared to KNN, MLP, SVC and Ridge, while being very competitive with consistently higher mean F1 compared to XGBoost and Random Forest; although its superiority was not found to be statistically significant after our implemented strict correction, as shown in Table IX below.

TABLE IX. QoEPredict's Bootstrapping results

Model	Mean F1 Difference	Lower CI	Upper CI	Significant difference
XGBoost	0.01065816	-0.011478646	0.029049682	FALSE
RandomForest	0.023843873	-0.002338558	0.050741837	FALSE
KNN	0.039838856	0.010446974	0.080060814	TRUE
MLP	0.085139933	0.045899627	0.116775816	TRUE
SVC	0.165032493	0.128220595	0.201732536	TRUE
Ridge	0.577259038	0.526716299	0.633201388	TRUE

1) LEARNING CURVE AND GENERALISATION CAPABILITY

We examine the learning curve for the QoEPredict model in Fig. 7 to gain critical insights into the model's learning behaviour as a function of the training set size. The plot shows the F1 Score on the y-axis and the number of training examples on the x-axis, with separate lines for the training score and the CV score. The model achieved near-perfect performance, 0.98-1.0, on the training data regardless of the sample size, indicating high capacity and strong memorisation ability. More importantly, the validation F1 Score displayed a consistent upward trend from 0.50 upwards as the number of training examples increased from 100 to 1200. This continuous improvement without a clear plateau suggests that the model has not yet reached its performance ceiling and could benefit further from additional data.

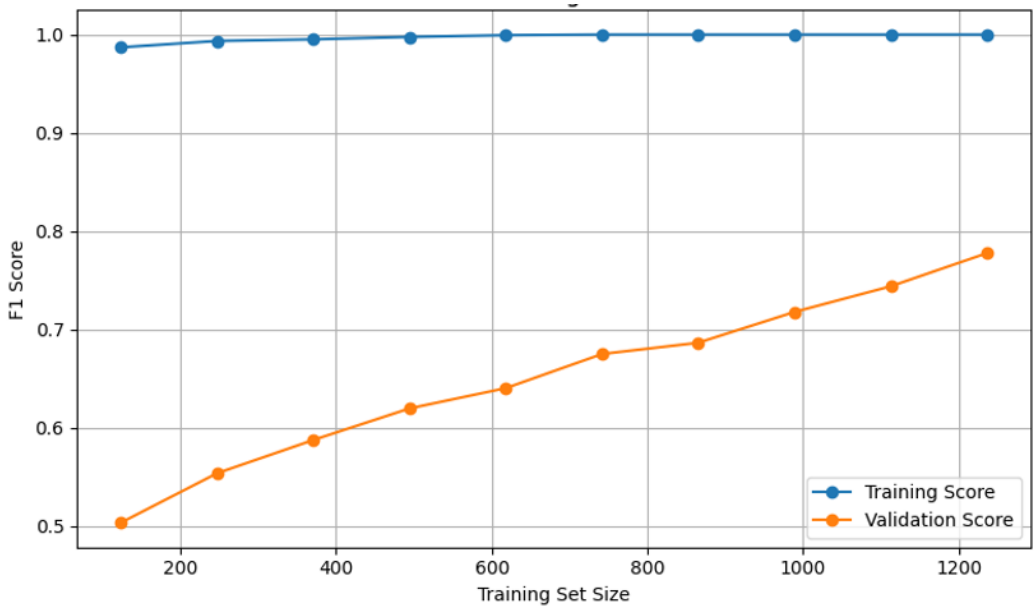


FIGURE 7. Learning curve of QoEPredict showing training and cross-validation F1 Scores as a function of training sample size. The gap between curves reduces as the training set size increases. This indicates that our ensemble's overfitting to sample data decreases as data volume increases.

However, the persistent but narrowing gap between the training and validation scores highlights a moderate variance issue. This is expected in expressive models and can be effectively mitigated to enhance generalisation by increasing data volume and/or introducing additional model refinement. These observations confirm that the classifier not only fits the data well but also generalises increasingly better with more training samples. This learning curve also justifies the use of an ensemble: with limited data, a single complex model might overfit, but our stacked approach manages complexity well.

2) CONFUSION MATRIX

To further investigate QoEPredict's performance, we assessed its class-wise predictive performance via a confusion matrix, Fig. 8, using predictions on the held-out test set. Overall, the matrix reveals strong alignment between predicted and actual classes, with dominant diagonal values across all categories. Notably, 133 correctly classified samples were in class 3 and 107 in class 4. Also, we observed minimal misclassification, with most errors occurring between adjacent QoE levels, e.g., class 3 is misclassified as class 2 and vice versa. This pattern is consistent with the ordinal nature of the QoE task, where semantic similarity between adjacent classes can naturally lead to borderline predictions. Extreme classes 1 and 5 exhibited lower values in terms of the magnitude of correct classifications, i.e., 28 for class 1 and 20 for class 5. However, the equally high predictive accuracy for these classes, 1 and 5, similar to the other classes, justifies our use of SMOTE during training to mitigate the effects of class imbalance. Although SMOTE generates synthetic feature-space samples without explicit consideration of the ordinal QoE target, its use is justified pragmatically to improve model attention to minority classes.

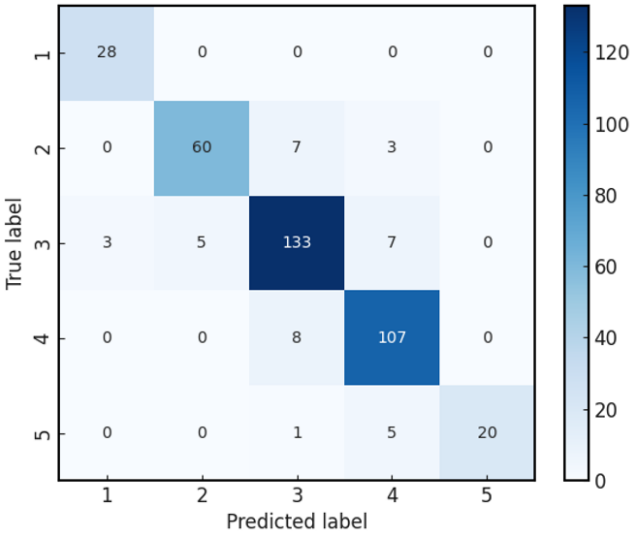


FIGURE 8. Confusion matrix of QoEPredict on the holdout test set. Diagonal dominance across all five QoE classes (1–5) indicates strong predictive accuracy with minimal inter-class confusion

Thus, the confusion matrix confirms not only the classifier's robustness across varying QoE levels but also the model's practical utility in distinguishing between subtly different user experiences with high fidelity, though not perfectly.

Table X presents a class-wise breakdown of performance across all five QoE categories. This clearly illustrates how our model performs on critical extremes: very poor (class 1) and excellent (class 5). As shown, recall is slightly reduced for these underrepresented classes (0.86 and 0.77), which is consistent with class imbalance. Precision remains very high (0.96 and 0.95), indicating that when extreme QoE is predicted, it is almost always correct. This analysis confirms that the model is not simply optimised for average performance but captures critical QoE states with reasonable reliability.

TABLE X. Class-wise Classification report on the Hold-out test set

Class	F1-score(%)	Precision(%)	Recall(%)	Precision(%)	Specificity(%)	Support
1	91	96	86	96	99	31
2	88	85	90	85	92	65
3	90	88	92	88	89	149
4	92	93	90	93	92	132
5	85	95	77	95	99	20
Macro Avg	89	91	87	91	94	397
Weighted Avg	90	90	90	90	91	397

This near-perfect class separation was further confirmed by the registered ROC-AUC value of 95.52%.

B. SHAP FEATURE IMPORTANCE AND INTERPRETATION

A key benefit of our approach is the integration of SHAP explainability, which helps us understand why the model makes certain predictions. For the interpretability of QoEPredict, we examined the contribution of each input feature by creating two complementary visualisations: a bar chart of average SHAP values and a heatmap. Fig. 9 displays the top 20 features ranked by mean absolute SHAP value across the five QoE classes (QoE=1 to QoE=5), quantifying each feature's contribution to the model's output in the original feature space. The multicoloured bars illustrate the additive impact of each feature on different QoE levels, highlighting that high-impact features consistently influence both low and high QoE predictions.

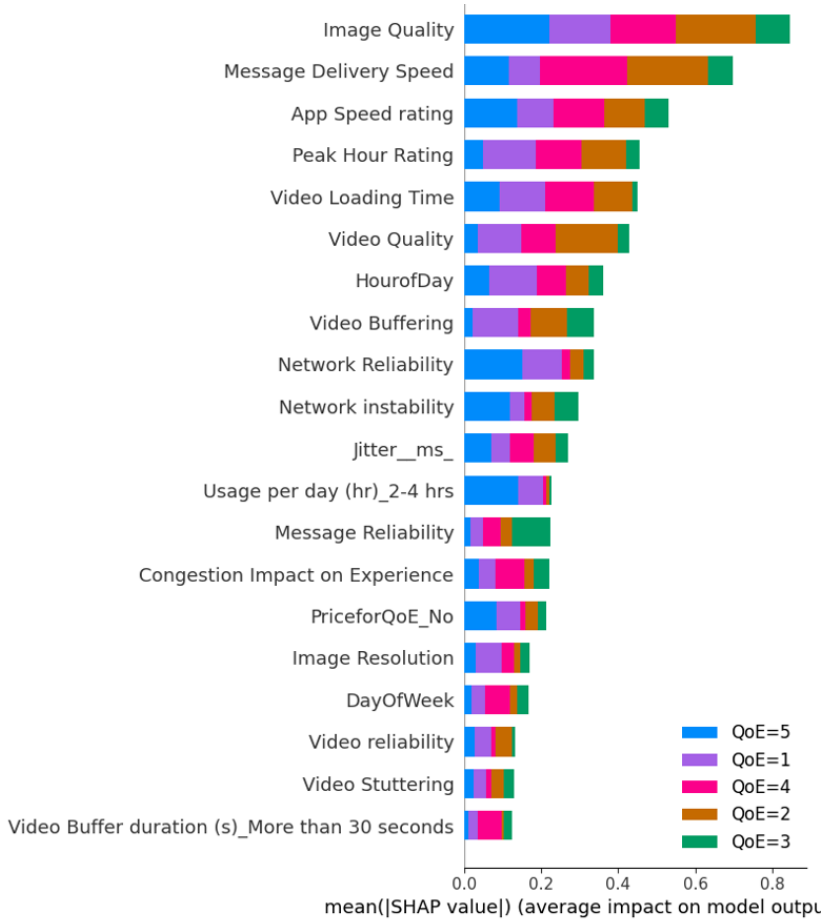


FIGURE 9. Top 20 features ranked by mean absolute SHAP values across all QoE classes. Higher SHAP values indicate stronger influence on predicted user experience

Notably, Image Quality, Message Delivery Speed, App Speed Rating, Network Reliability, and Video Quality emerge among the top predictors, each demonstrating substantial mean SHAP values across all QoE classes. These results suggest that perceived multimedia fidelity and responsiveness are primary drivers of user experience in social media applications. This confirms domain intuition: Faster Message delivery speed contributes positively to QoE, confirming that users strongly prefer instantaneous communication and are frustrated by delays [63][64]; Image quality and Video Quality demonstrate a high positive impact on QoE, indicating that preserving high resolution and clarity in shared images is crucial for user satisfaction e.g., reducing image resolution was found to have a significantly larger negative effect on perceived quality than minor losses in colour fidelity [63][65]. Network reliability was also among the top contributors, with reliable network conditions yielding higher predicted QoE, confirming that better network quality, e.g., stable connectivity and prompt data transfer, leads to quicker responses, fewer errors, and an overall enhanced user experience [63][18]. The model's emphasis on both user-centric, e.g., the influence of time-of-day (HourOfDay and DayOfWeek), user habits (Usage per Day(h)_2-4h), and network-level indicators (e.g., jitter_ms) justifies the effectiveness of the hybrid QoS-QoE modelling approach, reinforcing previous findings that both objective metrics and subjective perceptions are essential in accurate QoE estimation [66].

These interpretable insights have practical significance: For mobile network providers, the findings highlight the importance of infrastructure that minimizes delay and maximizes consistency, since even subsecond message delays can erode user satisfaction; For app developers, the results suggest focusing on optimisations like efficient image compression and content delivery techniques such as streaming or background pre-loading of media, to ensure high visual quality and instantaneous feedback in the user interface.

Additionally, the correlation heatmap, Fig. 10, provides further insight by visualising the pairwise linear relationships among the numeric features within the dataset. Notably, QoE_Score exhibits strong positive correlations with perceptual indicators such as Message Delivery Speed ($r \approx 0.5$), Image Quality ($r \approx 0.5$), and Network Reliability ($r \approx 0.5$), further underscoring their critical role in shaping user satisfaction. Conversely, negative correlations were observed with impairments such as Network Instability ($r \approx -0.2$) and Video Buffering ($r \approx -0.2$), which highlights the detrimental impact of service interruptions on QoE. Few feature pairs showed internal consistency, such as Video Stuttering and Video Buffering ($r \approx 0.4$), confirming the opportunity for dimensionality reduction.

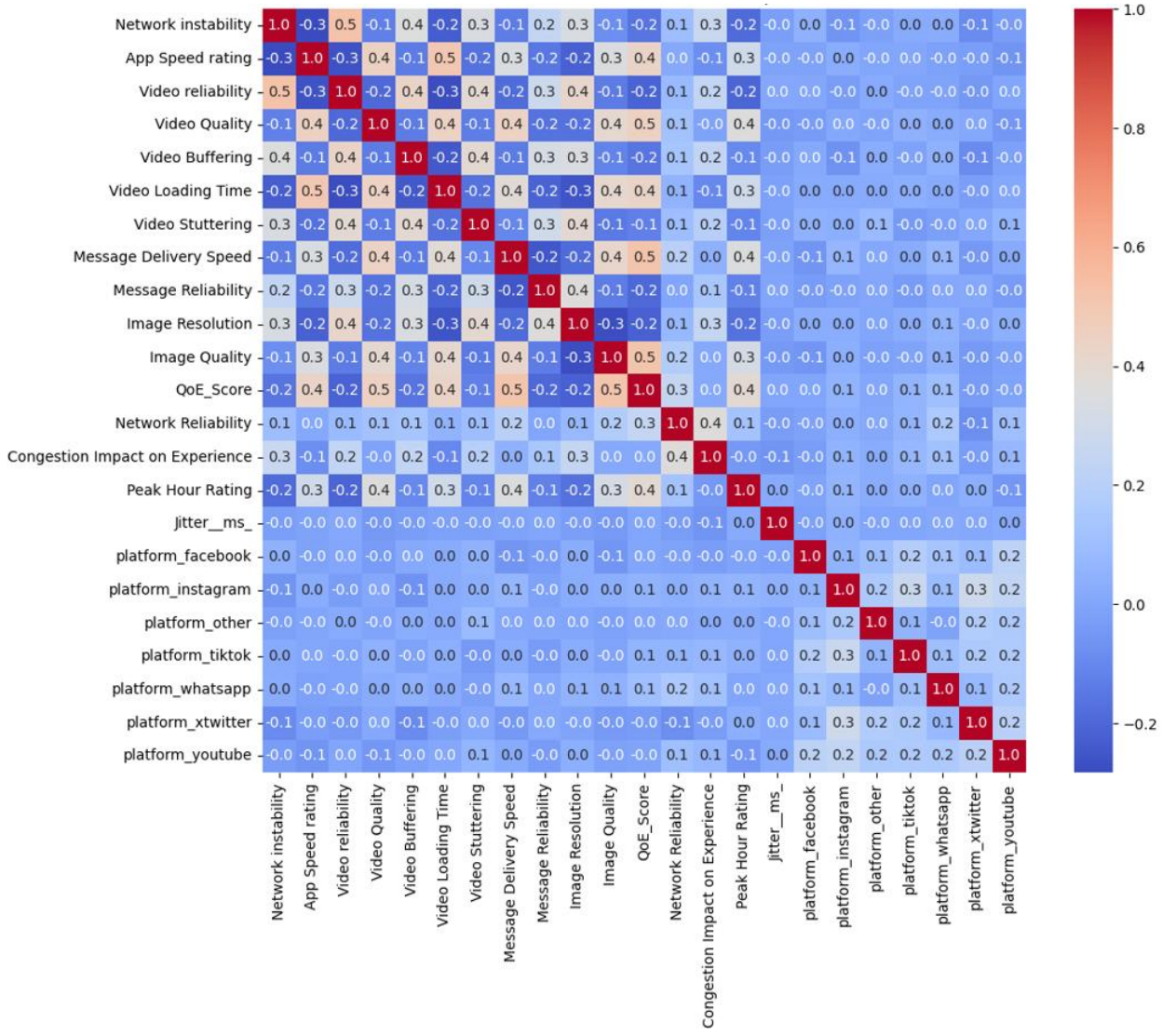


FIGURE 10. Correlation heatmap of key features in the hybrid dataset. Strong positive correlations with QoE_Score are observed for Image Quality, Message Delivery Speed, and Network Reliability, while negative correlations is observed for Video Buffering

Interestingly, platform-specific dummy variables, e.g., platform_whatsapp and platform_instagram, exhibited minimal correlation with other numeric predictors and QoE_Score, implying limited direct influence on overall QoE when evaluated independently. However, this does not preclude their relevance when explored via interaction effects or segment-specific modelling. The heatmap thus reinforces the significance of media performance and network reliability features over platform identity in driving perceived QoE. These findings align with previous studies,

e.g., that emphasise service responsiveness, delivery fidelity, and network stability as key determinants of user experience [29] [67]. Insights that can guide both feature selection strategies in ML pipelines and practical prioritisation for network operators seeking to optimise service delivery.

Using SHAP to probe our model not only validates that its predictions are grounded in known determinants of mobile QoE, but also bridges the gap from model outputs to actionable guidance, emphasising that improvements in message delivery speed, media quality, and network reliability can tangibly boost user experience in mobile and social applications, as consistently reported in QoE-focused studies. This convergence of model interpretability with domain knowledge greatly enhances trust in the model. It provides a clear roadmap for stakeholders seeking to leverage these insights to deliver improved user experiences.

C. UMAP AND K-MEANS CLUSTERING

Table XI presents the empirical summary of the Clustering methods comparison results. K-Means achieved superior performance across all key indicators, including the highest Silhouette score of 0.9242, lowest Davies-Bouldin index of 0.0749, and a competitive cluster purity score of 0.3842, indicating well-separated and compact clusters. It also produced the highest composite score of 0.6985, matching GMM and Agglomerative Clustering. While GMM and Agglomerative showed comparable statistical performance, we proceeded with K-Means due to its computational efficiency, scalability, and robustness in high-dimensional spaces without relying on strong distributional assumptions. Spectral Clustering, in contrast, demonstrated significantly weaker clustering quality; Silhouette = 0.5288, DB = 1.4196, rendering it unsuitable for the present application.

TABLE XI. Clustering methods comparison results

Metric	K-Means	GMM	Spectral	Agglomerative
Clusters	5	5	2	5
Silhouette	0.9242	0.9242	0.5288	0.9242
DB	0.0749	0.0749	1.4196	0.0749
Purity	0.3842	0.3842	0.3811	0.3842
EffectSize	0.3345	0.3345	0.1822	0.3345
ANOVA p	0.0000	0.0000	0.0082	0.0000
Kruskal p	0.0000	0.0000	0.0054	0.0000
Composite Score	0.6985	0.6985	0.2403	0.6985

We visually illustrate the quality of the K-Means clustering using multiple supporting plots. Fig. 11 shows an imbalanced distribution of samples across the five clusters, with Cluster 2 holding the largest proportion with ~1000 samples, confirming the actual imbalanced nature of QoE classes in the sampled context.

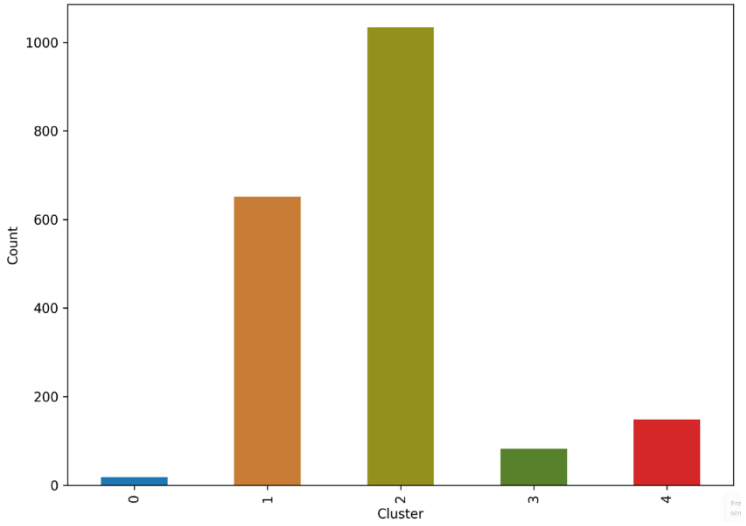


FIGURE 11. Distribution of sample counts across five K-Means clusters derived from UMAP embeddings. Cluster 2 contains the largest group (~1000 users), reflecting real-world QoE imbalance

Furthermore, the trio of visualisations contained in Figs. 12, 13, and 14 further strengthen the validation of the K-Means clustering outcome, revealing meaningful segmentation in user QoE profiles. Fig. 12 illustrates distinct QoE distributions across the five clusters using violin plots, highlighting heterogeneity in user experience levels and supporting actionable QoE differentiation. Cluster 0 exhibits a broader, higher median QoE, while clusters 1 and 3 skew lower.

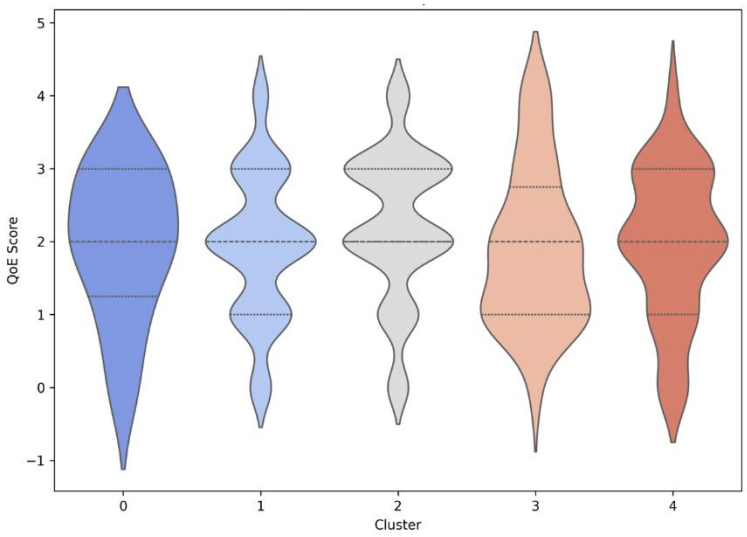


FIGURE 12. Violin plots of QoE distributions per cluster, illustrating varying median QoE levels across user groups

The silhouette plot, Fig. 13, confirms strong intra-cluster cohesion and inter-cluster separation, with most samples achieving positive silhouette scores, and a considerable number of samples exceeding 0.75 (dotted red line). This implies that most points are assigned to their respective clusters with minimal confusion.

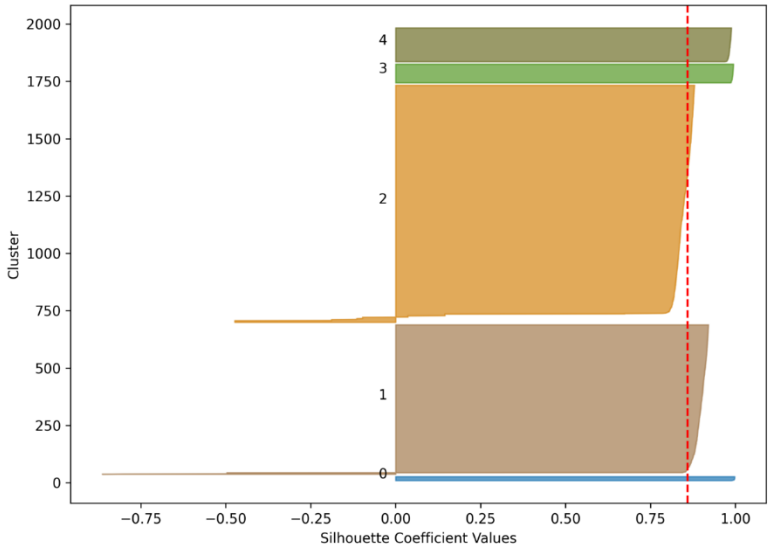


FIGURE 13. Silhouette plot showing intra-cluster cohesion and inter-cluster separation. Most samples achieve positive and high silhouette values, indicating stable clustering

The silhouette structure reinforces that the clusters are internally cohesive and reasonably well-separated, even if not perfectly as depicted in Fig. 14. The UMAP-based cluster separation scatterplot, Fig. 14, illustrates well-defined spatial separation among the clusters, with distinct groupings evident despite some peripheral dispersion; likely introduced by the nonlinear dimensionality reduction. This validates the model's ability to segment users meaningfully in the reduced feature space, preserving local structure and enhancing the interpretability of cluster-specific QoE patterns.

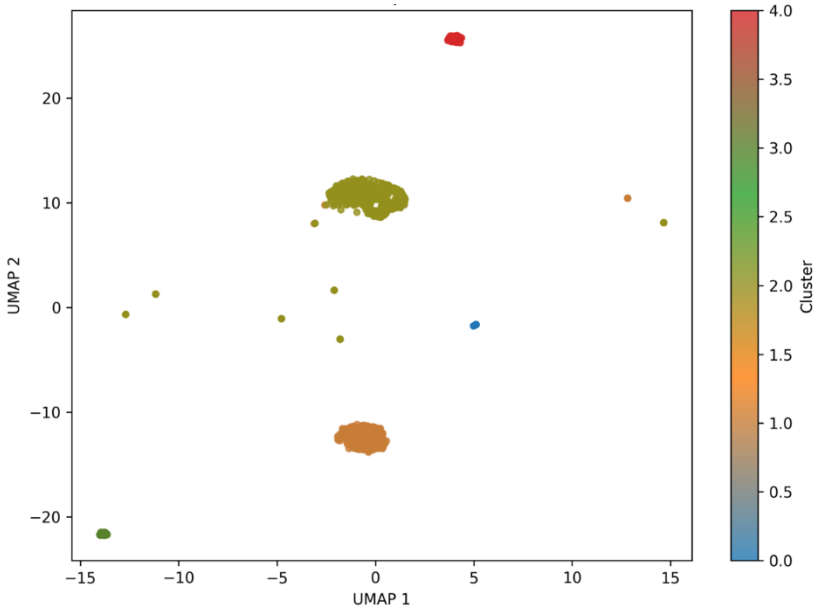


FIGURE 14. 2-D UMAP projection showing spatial separation among the five K-Means clusters. Distinct groupings highlight meaningful segmentation in real-world user experience profiles

Together, these figures demonstrate that the selected K-Means clustering yields interpretable, behaviourally distinct user groups. E.g., identifying the top QoS factors influencing the different clusters, Table XII, as revealed by SHAP, is essential for developing targeted QoE improvement strategies and supporting the deployment of adaptive QoE-aware service or network interventions.

TABLE XII. Taxonomy of Key QoS factors with the highest influence on different clusters

Cluster Label	Mean	St. D.	Cluster user count	Key QoS factor
0	1.944	0.998	18	Upload Speed < 1.18Mbps
1	1.949	0.996	652	Upload Speed < 4.43Mbps
2	2.237	1.005	1034	Bandwidth > 4.4Mbps
3	1.902	1.061	82	Latency >200ms
4	1.966	1.020	148	Download Speed < 2.35Mbps

D. HYBRID VS. QOS-ONLY VS. USER-ONLY MODELS

While our current results focus on the hybrid model, it is worth proposing how one would rigorously compare our hybrid versus single-source models in future work: Train a model using only the network QoS features as inputs and evaluate performance; Train another model using only user-centric features as inputs and evaluate performance; and compare these to the hybrid model that uses all features. We expect that the hybrid model would outperform either model alone, demonstrating the benefit of combining network QoS and user feedback data. This is consistent with the findings of other researchers who note that QoE is multidimensional and purely objective or purely subjective models are incomplete. E.g., a study in video streaming QoE demonstrated that adding user interaction data to network data improved QoE prediction accuracy by a notable margin [68].

In summary, the results validate that our hybrid approach is effective for the given context. We have shown quantitatively that it can predict user satisfaction to a useful degree of accuracy. Qualitatively, we have interpreted what the model learned and found it aligns with domain knowledge, enhancing the understanding of QoE drivers in Cameroon's mobile social media usage. With the core results established, the next section will discuss their implications: what they mean for network operators and policy, what limitations exist, and how future work can build upon this foundation.

V. DISCUSSION, INSIGHTS AND FUTURE WORKS

The findings of this study offer several revelations into an ensemble, hybrid QoE modelling and carry important practical implications for mobile network stakeholders in Cameroon and similar contexts. We discuss these insights, reflect on the local constraints, assess this study's limitations, detail the ethical considerations we navigated, and suggest directions for future research and deployment.

A. STUDY INSIGHTS

This study uncovers contextual insights that enhance the understanding of Mobile Network QoE in a developing country like Cameroon.

1) SUPERIOR MODEL PERFORMANCE

Our results reinforce the concept that QoE is best understood through a combination of network and user perspectives.

Neither QoS metrics nor user feedback alone paints the full picture; it is their interplay that matters. The hybrid ensemble model's success demonstrates that much of what a user feels about service quality can indeed be predicted from technical parameters, provided we train the model on actual user opinion data. This marries the subjective and objective approaches found in QoE literature. The clear improvement of the stacking ensemble over any single model confirms the benefit of ensemble methods for QoE prediction. By aggregating different aspects of threshold-driven learning approaches, the ensemble likely captured a broader range of patterns. For instance, we suspect XGBoost's boosting mechanism captured dominant global thresholds, while Random Forest's feature randomness detected context-specific local thresholds. The meta-learner successfully learned to trust the appropriate model in the appropriate scenario. This aligns with previous research that noted ensembles improve QoE regression accuracy [9], and extends it to a 5-class classification setting with a novel stacking scheme. The improvement is not just statistical but practical.

We also note that the ensemble reduced overfitting. The stacking, by using CV predictions and combining models, smoothed out the idiosyncratic biases of each model. In effect, the ensemble behaves like a regulator; it is harder for random noise to consistently fool all base models in the same way. This is reflected in the increasingly reduced gap between training and test curves for the ensemble as the dataset size grows. That said, stacking itself could overfit at the meta-level if not using proper blending; our use of out-of-fold predictions mitigated that.

2) MODEL INTERPRETABILITY AND TRUST

A key aspect of our work is the use of SHAP for interpreting the QoE model. This addresses the common criticism that advanced ML models are "black boxes" [69]. By applying SHAP, we were able to verify that the model's behaviour makes sense. These insights are valuable to network providers and reveal that managing QoE is not a one-size-fits-all solution, but should appropriately consider context. A specific insight is the primacy of App speed rating, service latency, service quality and reliability for social media QoE. This suggests social apps, while not "real-time" in the way video calls are, still demand responsiveness; people scroll quickly, tap on links, and expect snappy feedback. Even loading a photo thumbnail can feel frustrating if there is a long delay after a click. Thus, operators aiming to improve social media experience should focus as much on reducing latency, through network optimisation, edge caching, etc, as on increasing throughput. High bandwidth is beneficial up to a point, especially for video-heavy content on platforms like Instagram, but if the network is unreliable, users notice that first. This insight is somewhat in line with the general QoE theory that *"quality is as good as the weakest link"*. If any dimension is particularly bad, it will dominate the user's perception. Furthermore, being able to explain individual predictions means operators could investigate specific complaints: e.g., if a user report is predicted as very dissatisfied, the model might highlight high Jitter as the cause, directing engineers where to look.

Another insight is the role of contextual factors: time-of-day emerged as a surrogate for network load and possibly user mood. The fact that experiences during peak hours impact QoE indicates that network resource contention is impacting QoE. For Cameroon's operators, this may highlight a need for capacity planning. If, e.g., users in urban areas consistently face reduced QoE at night, interventions like offloading traffic, deploying additional small cells, or optimising backhaul could be targeted at those hours. This interpretability also adds credibility to the model in a regulatory context. Regulators in Africa are increasingly interested in QoE but also cautious due to its subjectivity. A model that can demonstrate *why* it thinks QoE is below threshold for an area, e.g., "*because average speeds are only 1 Mbps and users expect at least 3 Mbps for acceptable video streaming*", can facilitate dialogue between operators and regulators on quality standards.

3) REGIONAL NOVELTY

This study is, to our knowledge, among the first to rigorously model QoE using actual user data from Africa. Past studies in Africa often relied on small surveys or focused on either QoS technical KPIs or qualitative assessments. Our work provides a quantitative baseline for what factors drive user experience in African mobile networks. One might wonder if the determinants of QoE are the same as elsewhere; we found that speed and quality issues reign supreme, but there could be subtle differences. For instance, one could imagine that in regions where users are new to mobile internet, their expectations might be lower; so QoE might not dip as sharply until QoS is very poor.

Our model indirectly captures the collective expectations of the user base we sampled. If those expectations differ from, say, a European user base, the model's learned threshold for "acceptable delays" would differ. In practice, our SHAP analysis indicated that many users tend to rate QoE poorly when Video Buffer Durations exceed 30 seconds. This threshold might shift as networks improve and content demands increase.

Another regional aspect is device variety; emerging markets often have older or lower-end smartphones, which might themselves limit performance. We did not explicitly model device type in this study, but it might be partially reflected in throughput, as older devices might not achieve high speeds even on good networks. In any case, future work should incorporate more region-specific aspects like power outages affecting networks. Also, we did not include 5G in our study, largely because 5G is currently scarce in Africa. As 5G rolls out, new QoE studies will be needed to see, for example, if ultralow latency and high bandwidth change the QoE game or simply raise user expectations further.

B. PRACTICAL IMPLICATIONS FOR MOBILE OPERATORS IN CAMEROON

Practically, our QoEPredict model can be integrated into network operators' analytics platforms as a decision-support module. Our model outputs (predicted QoE classes and SHAP-based feature importance scores) can feed into real-time dashboards that visualise, for example, regional or cell-level user experience trends. Low predicted QoE zones can automatically trigger alerts, e.g., for bandwidth reallocation, signal optimisation, or targeted maintenance. In parallel, the feature importance insights can guide strategic interventions by identifying which QoS factors (e.g., jitter or throughput) most affect user satisfaction. This integration enables a dynamic feedback loop between predictive analytics and network management, thereby supporting proactive QoE-aware decision-making. For telecom companies and internet service providers in Cameroon, this research provides a prototype framework for QoE-centric network management:

1) PROACTIVE TROUBLESHOOTING

Instead of waiting for customers to call and complain about poor service, an operator could use a model like ours to predict in real-time which users or areas are likely to have a bad experience. For instance, by inputting the live QoS stats from base stations or user equipment into the model, the operator's dashboard could highlight "Zone X has an estimated QoE of 2/5 right now, likely due to high latency." This allows technicians to prioritise investigating Zone X; maybe a cell is overloaded or a backhaul link is congested, before user churn occurs. Over time, this could improve customer satisfaction and loyalty by reducing sustained poor experiences.

2) RESOURCE ALLOCATION

Knowing what affects QoE most helps in network optimisation decisions. Our results suggest that reducing latency could have a bigger bang-for-buck on QoE than marginally increasing throughput. So, an operator might invest in technologies like optimising routing paths, local caching of content, to reduce round-trip time to popular social media servers, or upgrading old transmission equipment that adds delay. In summary, a QoE model provides a user-centric lens as illustrated in Fig. 15, for network upgrades: rather than just looking at utilisation stats, it tells where users are hurting, as reflected in their suggestions.

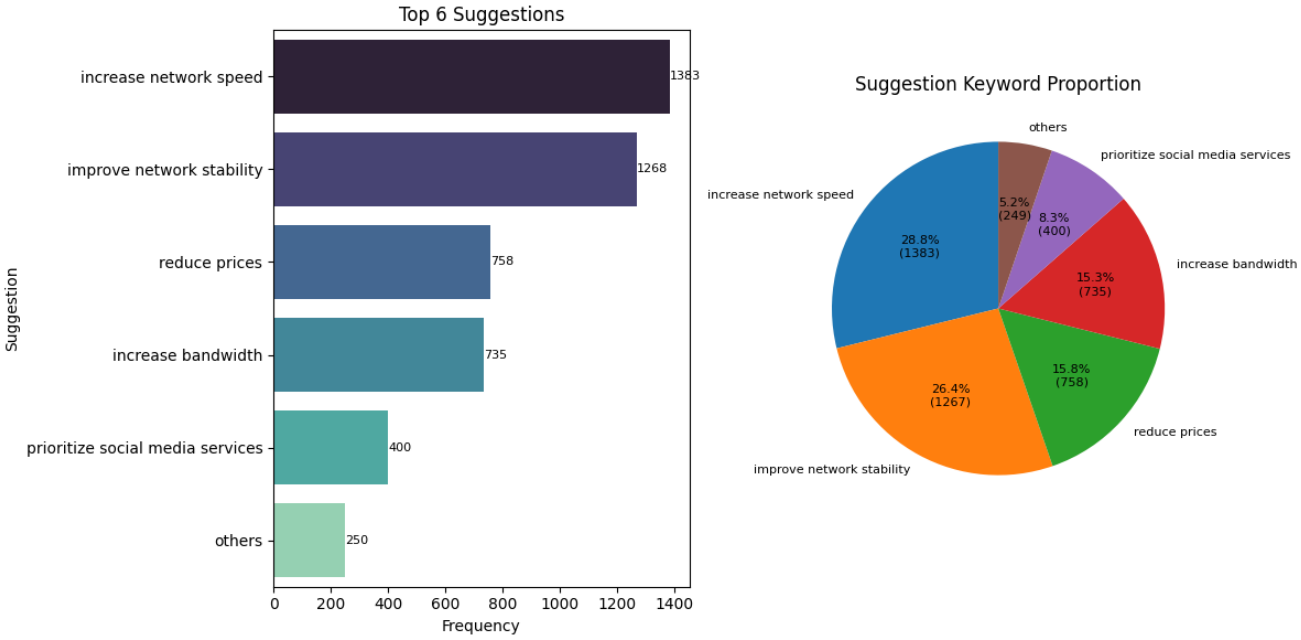


FIGURE 15. User Suggestions frequency ranking across the sampled population

3) MARKETING AND SERVICE TIERS

Another interesting implication is in *customer relationship management*. Operators could identify segments of users who consistently have lower QoE, maybe those in fringe coverage areas or with older devices. They can then target those users with upgrades. E.g., offer a special promotion for a 4G signal booster, or a discounted device upgrade, or simply prioritise network improvements in those locales. Conversely, if QoE is generally good, it can be used in marketing, especially on the social media platforms that subscribers use the most, as captured in Fig. 16. E.g., *"90% of our customers enjoy HD video and fast social media browsing without issues, as indicated by our QoE scores!"*. Caution is needed here to ensure the QoE scores are truly representative and not used misleadingly.

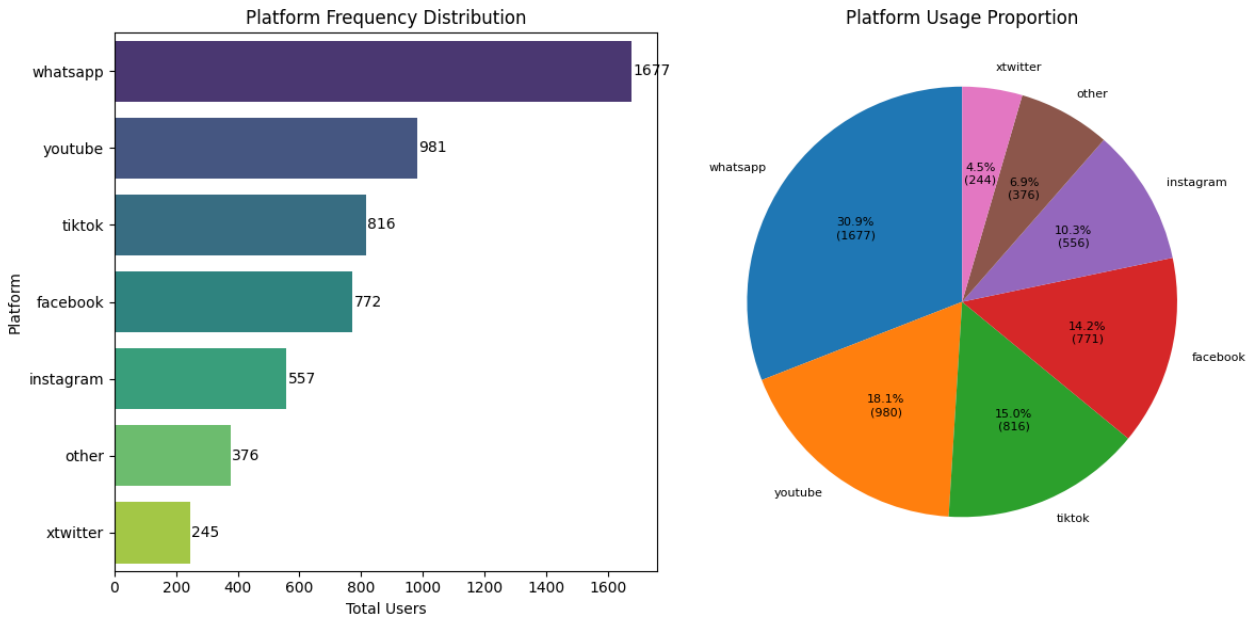


FIGURE 16. Social Media Platforms usage frequency distribution of the sampled population

4) REGULATORY REPORTING

Regulators in some countries are starting to consider QoE in addition to QoS for assessing operators. Cameroon's regulator could benefit from models like this to complement their QoS audits. E.g., instead of just checking if an operator meets 4G speed targets, they could use a QoE model on crowd-sourced data to see if users are satisfied. Our approach, being data-driven and local, is more reflective of ground truth user experience than generic thresholds. Thus, it could contribute to more consumer-centric regulations.

C. LOCAL CONSTRAINTS

During this study, we encountered a few Cameroon-specific constraints worth noting:

1) INFRASTRUCTURE VARIABILITY

The quality of 3G/4G in Cameroon is not uniform. Some areas, especially rural or certain carriers, might not consistently deliver what one would consider "3G" or "4G" speeds. We had to ensure our data collection included a broad sample, which it did; though urban users were more represented, given smartphone penetration biases. This variability means the model has to cover from very poor to decent network conditions, which it did. But it also means our model is somewhat specific to the range of QoS present in Cameroon's networks. In an environment with ubiquitous high-speed broadband, the QoS-QoE curve might look different since only extreme conditions degrade QoE.

2) DEVICE DIVERSITY

Cameroon's smartphone market includes many low-cost android devices, some with older 3G-only support. Lower-end devices might produce more jitter in performance, independent of the network. A country where high-end phones are common might see QoE limited more by the network, whereas here, occasionally, the phone could be a bottleneck.

3) USER EXPECTATIONS

Cultural and usage pattern differences can affect how people rate their experience. It is possible that Cameroonian users, many coming from a history of slower internet or limited connectivity, might rate an average network experience as good/excellent because it meets their expectations. This can contrast with users in markets who expect ultrafast, always-on connectivity. This expectation gap might partially explain why we had relatively few ratings in the very low end, since some users might just accept certain slowness as normal. Over time, as people get exposed to better service, their expectations may rise, and QoE ratings for the same QoS could become harsher. Operators and models will need to adapt to that moving target.

4) DATA ACCESS AND COSTS

Getting widespread user participation required careful consideration of data costs and incentives. We had to ensure the speedtest or survey did not consume too much of users' data bundles, for ethical and practical reasons, by sometimes providing data bundle incentives. This constraint is somewhat unique to developing regions where unlimited data plans are rare and users are cost-sensitive. It means any large-scale QoE monitoring solution should ideally work with minimal overhead or be zero-rated by the operator to encourage participation.

D. STUDY LIMITATIONS

We recognise a few limitations to this study as thus:

1) SIZE OF DATASET

The dataset with 1934 datapoints, while real, is modest. It covers selected towns, largely urban, in Cameroon. These results might vary if more samples and the entirety of the country are fully represented. We mitigated overfitting through CV and saw consistent performance, but a larger dataset would always be preferable to further enhance model generalisability. Secondly, there may be unobserved variables that affect QoE, which we did not include. For instance, QoS metrics like signal-to-noise ratio (SNR), and congestion/cell load measures, because we did not have access to operator probes. We did not have data on users' current pricing or data caps; a user nearing their data limit might have lower QoE independent of instantaneous speed. Likewise, we did not explicitly model intrinsic device type or content type parameters beyond a rough platform category; more granular QoE models might consider video resolution, web page complexity, etc.

Our model currently provides an offline prediction, based on collected data; it is not implemented as a real-time system. For practical deployment, one would integrate it into a network monitoring tool to predict QoE continuously, which might require optimisation for speed. Though our ensemble is reasonably fast in inference, monitoring millions of users would need further scaling strategies. Also, the scope of applications is limited to social media apps. Our model might not directly apply to other use cases, like online gaming, without retraining. Different app types have different QoE models. So, while our framework is general, the numerical model is specific to social media usage patterns.

2) GROUND TRUTH SUBJECTIVITY

We relied on user ratings as ground truth QoE. Human ratings have inherent noise; one user's "4" might be another's "5". We tried to mitigate this by having a decent sample size and a robust model design. Random Forest and XGBoost are relatively robust to label noise. Furthermore, the stacking itself helps to smooth out idiosyncratic errors that might be present in the training labels, validating model performance across folds.

However, the evaluation metrics themselves are based on these possibly noisy labels. In an ideal scenario, we might do multiple ratings or MOS averaging per condition to get a more stable ground truth. That was not feasible in our crowdsourced approach, but it is a trade-off. Also, to compensate for a formal inter-rater reliability test, such as Fleiss' Kappa, which requires multiple independent ratings for the same sample, one of our expert authors re-rated a random subset of 150 samples. This spot-check revealed a 75% exact agreement and a 94% agreement (within ±1 class) between the expert and the original crowd-sourced labels. This high level of agreement provided strong confidence that the users largely applied the QoE scale as intended.

3) STATIC Vs. DYNAMIC QoE

Our analysis is static and point-in-time. We did not explicitly model how QoE evolves during a session because we did not record continuous QoE over a long session, just snapshots. So, we extracted the temporal features (HourOfDay and DayOfWeek) as numeric attributes and scaled. We did not use cyclical encoding because each record was independent and session-based; making a discrete contextual representation more appropriate. However, QoE can have memory, and a bad spike at the start of a session can taint the whole experience rating. Our data can not directly capture that effect since each rating is independent. This is a limitation if one tries to use the model in a scenario that requires time-series predictions. Nevertheless, cyclical encoding may be explored in future work when sequential temporal patterns are of interest.

4) GENERALITY IN CAMEROON

While we included users from various regions, our sample may still be slightly skewed to tech-savvy or engaged users, i.e., those who were willing to install a speed test app and take the survey. Although we collected data from all local networks, operator representation was volunteer-driven and thus uneven, potentially biasing our results toward the dominant operators within the sampled population. Also, our data collection period of three months may not have captured the longer-term network dynamics, such as network upgrades or evolving user expectations. So, the general mobile user population may have different characteristics. Therefore, if an operator has truly random sampling via network probes and occasional SMS surveys, they might gather a more representative dataset. We believe our sample is reasonably representative of the urban user base, which constitutes the majority of 3G/4G users; however, rural users may be underrepresented. We recognise that partnering with operators to access these rural populations, e.g., via incentivised campaigns, will boost future data collection endeavours. Thus, caution is warranted in assuming these exact metrics hold for 100% of users.

5) REGULARISATION AND OVERFITTING TRADE-OFF

We deliberately traded off fine-grained XGBoost regularisation tuning to minimise model computational cost. This decision may have contributed to the mild overfitting observed in Fig. 7, where training F1 Scores (≈ 0.98 - 1.0) slightly exceed validation F1 Scores (≈ 0.89). This observed gap suggests that our ensemble may have marginally over-specialised on training patterns due to the untuned XGBoost regularisation parameters (L1 and L2). Nonetheless, QoEPredict's performance remains stable across CV folds, with a standard deviation of ±1.12 in F1 Scores, indicating consistent generalisation. Future work will incorporate systematic regularisation tuning to further reduce variance and enhance robustness.

Despite these limitations, the study provides proof-of-concept that hybrid QoE modelling using crowdsourced user data is feasible and useful in the Cameroonian context. It

establishes a baseline for further refinement. Therefore, while the model is calibrated to Cameroon, the modular pipeline is designed for replication. Generalisation to new regions will primarily require the collection of local QoS-QoE data for re-training, allowing the framework to discover region-specific clusters and drivers. Thus, providing a scalable blueprint for data-driven QoE management across emerging markets.

E. FUTURE DIRECTIONS

Moving forward, we see several avenues to extend this work:

1) OBTAINING MORE DATA

Additional data, both in volume and variety, would allow validation and possible model improvement. E.g., integrating additional fine-grained radio network KPIs, such as Signal-To-Noise Ratio, modulation scheme, and congestion indicators, etc, which operators have. Our model used relatively high-level QoS metrics that a user app can measure. Future work will aim to collaborate with operators to integrate the broader socio-technical ecosystem influencing QoE, e.g., data pricing, data cap status. Additionally, collecting multi-country datasets could also help ensure that the model generalises well to different network conditions and user behaviours, especially within the Sub-Saharan region. This will also permit us to adapt the model to real-time QoE variations for optimised predictions. Furthermore, modelling the impact of external factors like power outages on network stability could provide a more complete picture of the user's experience. Capturing these elements is essential to achieving a truly context-aware QoE model that reflects the economic and infrastructural realities of users in emerging markets.

2) INTEGRATION INTO OPERATOR DASHBOARDS

We envision a system where the operator's network management software includes a *QoE dashboard* showing real-time predicted satisfaction scores across cells or regions. Implementing this would require robust real-time data pipelines from the network and possibly periodic user feedback collection integrated into an operator's mobile app or via SMS. There would also need to be threshold-based alarms, e.g., *"QoE in the Northwest region dropped below 3.0 for >15 minutes"*, to prompt action. Our model could be a starting algorithm for such a system, but deployment would involve engineering efforts and validation.

3) REFINE THE CLUSTERING APPROACH

Future work could turn our unsupervised analysis into a semi-supervised approach where cluster identity is used to train specialised models, e.g., one model for urban versus rural users. This might improve predictions further by allowing varying model parameters for each segment. The challenge is ensuring enough data per segment. In our analysis of linear versus non-linear clustering methods, UMAP achieved a significantly higher silhouette score (0.92) compared to PCA's 0.21, suggesting that it preserves the intrinsic structure of the dataset more effectively. However, when benchmarked directly against the no-reduction baseline, UMAP did not consistently outperform it in terms of predictive accuracy or F1 score; a similar behaviour to PCA is shown in Fig. 4. But unlike PCA, which showed decreasing performance with higher dimensions, the no-UMAP reduction baseline achieved an accuracy of 82%, while the UMAP(40), UMAP(75) and UMAP(100) configurations achieved 55.5%, 56.3% and 56.8% respectively.

This indicates that although UMAP provides superior cluster separation and interpretability of latent data structure, this improvement in representation does not automatically translate into higher predictive performance in our QoE classification setting. This provides a more complete view of UMAP's role: it excels in revealing structure and supporting interpretability, but does not necessarily surpass the no-reduction approach in predictive accuracy.

4) COLLABORATION WITH CONTENT PROVIDERS

Social media companies also measure QoE from their side, with more granular metrics like time to load content. A collaboration between network operators and content providers could lead to richer datasets, combining network QoS with application-level QoE metrics. Such combined data could improve model accuracy and also help pinpoint whether issues are network or app-related. While this might be beyond the scope of a single research study, it is a direction the industry is moving in.

5) REGULATORY FRAMEWORKS

As a future consideration, if the regulator in Cameroon or regionally were interested, they could sponsor larger studies to continuously monitor QoE using approaches like ours. This could feed into a public scorecard for operators, adding a competitive drive to improve QoE. We believe research like this can inform those policy-level initiatives by providing methodology and evidence of what works.

6) EXTENSION INTO 5G, 6G AND NEW APPLICATION DOMAINS

A key future step is to implement and test our QoE modelling approach in the context of newer network generations (5G and beyond) and emerging applications. Our work has mainly focused on 3G/4G networks and social media usage in Cameroon. However, as 5G networks become more widespread (and 6G approaches), it is essential to explore how the significantly different features of these networks influence QoE. 5G's ultra-low latency and high bandwidth enable applications like cloud gaming, VR, or telemedicine with exceptional quality, but they could also raise user expectations considerably. Future research should investigate whether the factors that were critical for QoE in 3G/4G in developing countries continue to be relevant in

5G/6G, or if new factors appear when, say, holographic communication or tactile internet services are introduced.

F. ETHICAL CONSIDERATIONS FOR FUTURE WORKS

Any future work should maintain the focus on user benefit. QoE prediction should ultimately serve to enhance user experience, not to justify charging more for *"premium QoE"*. As this is a potential misuse, where an operator might say *"pay more to get better QoE"*. That would be problematic if the baseline QoE is intentionally kept low. Transparency with users, offering opt-outs, and using the data responsibly will be key as this field progresses.

In summary, the discussion highlights that hybrid QoE modelling is a promising tool that, when applied thoughtfully, can bridge the gap between network engineering and user satisfaction. By focusing on Cameroon's context, we have ensured that the model and insights are grounded in local reality, a step towards more inclusive global research where regions with different usage patterns and constraints are represented. Our proposed model also demonstrated a successful case of applying state-of-the-art ML techniques to a practical engineering problem in a developing region context. This approach can serve as a reference for similar QoE studies or be extended as mentioned above. We hope this work spurs more data collection and research on QoE in Cameroon, Africa, and other underserved regions, as improving QoE is key to user satisfaction and broader digital inclusion.

VI. CONCLUSION

This paper introduced QoEPredict, a novel hybrid ML framework designed to predict QoE in Cameroon's 3G/4G networks, using social media usage as a case study. By integrating crowdsourced QoS data with user-reported satisfaction scores, we built a regionally contextualised, data-driven model capable of capturing both technical and perceptual dimensions of QoE. Our stacking ensemble, which combines XGBoost, Random Forest, and disagreement features, achieved a high predictive performance (F1 score and accuracy of 90%) and outperformed traditional baselines. Beyond its predictive strength, QoEPredict contributes methodologically through its modular pipeline and interpretability via SHAP-based explainability. These features can enable mobile network operators to identify actionable QoE drivers, such as jitter, service delivery delays, and contextual user feedback, informing more targeted and user-centric optimisation strategies. As one of the first large-scale ML-based QoE studies in Cameroon, this work fills a critical gap in the literature and sets a precedent for similar efforts in other resource-constrained regions. Future work can extend this framework to broader service types and geographies, further bridging the gap between technical performance and real user experience.

ACKNOWLEDGEMENT

The authors would like to thank all the participants in Cameroon who provided their network data and feedback; this work would not have been possible without their cooperation.

CONFLICT OF INTEREST

None to report

REFERENCES

[1] Datareportal, 'Digital 2025: Cameroon'. Accessed: Apr. 14, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-cameroon#:~:text=The%20%E2%80%9Cstate%20of%20digital%E2%80%9D%20in%20Cameroon%20in,to%2018.5%20percent%20of%20the%20total%20population

[2] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad, and Z. Fei, 'Subjective QoE of 360-Degree Virtual Reality Videos and ML Predictions', *IEEE Access*, vol. 8, pp. 148084–148099, Aug. 2020, doi: 10.1109/ACCESS.2020.3015556.

[3] S. Zeng, W. Yang, Z. Wang, Q. Zeng, Y. Wang, and X. Liu, 'Study on User Satisfaction Prediction and Influencing Factors based on GBDT Modelling', *TCSISR*, vol. 1, pp. 39–46, Oct. 2023, doi: 10.62051/t05r6x47.

[4] P. H. S. Panahi, A. H. Jalilvand, and A. Diyanat, 'Enhancing Quality of Experience in Telecommunication Networks: A Review of Frameworks and ML Algorithms', Sept. 2024. Accessed: Oct. 10, 2024. [Online]. Available: http://arxiv.org/abs/2404.16787

[5] P. K. Guma and M. Mwaura, 'Infrastructural configurations of mobile telephony in urban Africa: vignettes from Buru Buru, Nairobi', *Journal of Eastern African Studies*, vol. 15, no. 4, pp. 527–545, Oct. 2021, doi: 10.1080/17531055.2021.1989138.

[6] Business in Cameroon, 'Consumer Group Demands Action Against Telecom Giants for Failing Service', *Business in Cameroon*. Accessed: May 20, 2025. [Online]. Available: https://www.businessincameroon.com/public-management/2908-14100-consumer-group-demands-action-against-telecom-giants-for-failing-service

[7] H. O. Hamidou, J. P. Kouraogo, O. Sie, and D. Tapsoba, 'Machine learning based Quality of Experience (QoE) Prediction Approach in Enterprise Multimedia Networks', in *Proc. of the 5th edition of the Computer Science Research Days, JRI 2022,* Ouagadougou, Burkina Faso: EAI, Nov. 2022, doi: 10.4108/eai.24-11-2022.2329806.

[8] M. Alreshoodi and J. Woods, 'Survey on QoE\QoS Correlation Models For Multimedia Services', *IJDPS*, vol. 4, no. 3, pp. 53–72, May 2013, doi: 10.5121/ijdps.2013.4305.

[9] P. Casas, M. Seufert, N. Wehner, A. Schwind, and F. Wamser, 'Enhancing Machine Learning Based QoE Prediction by Ensemble Models', in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, Vienna: IEEE, pp. 1642–1647, Jul. 2018, doi: 10.1109/ICDCS.2018.00186.

[10] Khoshkroodi A., Parvini H., and Aajami M., 'Stacking Ensemble-Based Machine Learning Model for Predicting Deterioration Components of Steel W-Section Beams', *Buildings,* vol. 14, no. 240, Jan. 2024, doi :10.3390/buildings14010240.

[11] N. Wehner, A. Seufert, T. Hoßfeld, and M. Seufert, 'Explainable Data-Driven QoE Modelling with XAI', *in 2023 15th International Conference on Quality of Multimedia Experience (QoMEX),* Ghent, Belgium: IEEE, pp. 7–12, Jun. 2023, doi: 10.1109/QoMEX58391.2023.10178499.

[12] S. Wang, M. A. Qureshi, L. Miralles-Pechuán, T. Huynh-The, T. R. Gadekallu, and M. Liyanage, 'Explainable AI for 6G Use Cases: Technical Aspects and Research Challenges', *IEEE Open J. Commun. Soc.*, vol. 5, pp. 2490–2540, Apr. 2024, doi: 10.1109/OJCOMS.2024.3386872.

[13] C. Molem, A. D. Akume, and B. P. Bihkongnyuy, 'Customer Satisfaction in Cameroons' Mobile', vol. 6, no. 1, pp. 70-81, Jan. 2018, Paper ID: 16011801.

[14] Kum Bertrand Kum and Dr. Austin, 'Quality of Experience (QoE) in LTE GSM UMTS Mobile Networks', *IJLTEMAS*, vol. 13, no. 9, pp. 136–146, Oct. 2024, doi: 10.51583/IJLTEMAS.2024.130914.

[15] A. Binele Abana, P. D. Bavoua Kenfack, P. S. Ngohe Ekam, E. Tonye, and L. Nyobe Makani, 'Modelling the Customer Experience using Machine Learning for Optimizing the Performance of Telecommunications Networks: Case of Mobile Networks', *IJAEMR*, vol. 09, no. 05, pp. 151–171, Oct. 2024, doi: 10.51505/ijaemr.2024.9511.

[16] Y. Chen, P. Yu, Z. Zheng, J. Shen, and M. Guo, 'Modelling feature interactions for context-aware QoS prediction of IoT services', *Future Generation Computer Systems*, vol. 137, pp. 173–185, Dec. 2022, doi: 10.1016/j.future.2022.07.017.

[17] F. Metzger *et al.*, 'An Introduction to Online Video Game QoS and QoE Influencing Factors', *IEEE Commun. Surv. Tutorials*, vol. 24, no. 3, pp. 1894–1925, May 2022, doi: 10.1109/COMST.2022.3177251.

[18] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, 'A Survey on Quality of Experience of HTTP Adaptive Streaming', *IEEE Commun. Surv. Tutorials*, vol. 17, no. 1, pp. 469–492, Sept. 2014, doi: 10.1109/COMST.2014.2360940.

[19] I. Sousa, M. P. Queluz, and A. Rodrigues, 'A survey on QoE-oriented wireless resources scheduling', *Journal of Network and Computer Applications*, vol. 158, p. 102594, May 2020, doi: 10.1016/j.jnca.2020.102594.

[20] M. Chen, L. Wang, J. Chen, and X. Wei, 'QoE-Driven D2D Media Services Distribution Scheme in Cellular Networks', *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–10, Jul. 2017, doi: 10.1155/2017/8754020.

[21] P. Ke and F. Su, 'Mediating effects of user experience usability: An empirical study on mobile library application in China', *EL*, vol. 36, no. 5, pp. 892–909, Nov. 2018, doi: 10.1108/EL-04-2017-0086.

[22] M. A. F. Kamil, 'User Experience Analysis of LinkedIn Social Media Using Usability Metric for User Experience (UMUX)', *JIEET*, vol. 7, no. 2, pp. 78–82, Dec. 2023, doi: 10.26740/jieet.v7n2.p78-82.

[23] T. L. Mitzner *et al.*, 'Older adults talk technology: Technology usage and attitudes', *Computers in Human Behavior*, vol. 26, no. 6, pp. 1710–1721, Nov. 2010, doi: 10.1016/j.chb.2010.06.020.

[24] Universiti Teknikal Malaysia Melaka et al., 'A Review on Usability and User Experience of Assistive Social Robots for Older Persons', *IJIE*, vol. 14, no. 6, Nov. 2022, doi: 10.30880/ijie.2022.14.06.010.

[25] S. Borsci, S. Federici, S. Bacci, M. Gnaldi, and F. Bartolucci, 'Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience', *International Journal of Human-Computer Interaction*, vol. 31, no. 8, pp. 484–495, Aug. 2015, doi: 10.1080/10447318.2015.1064648.

[26] M. E. Dimaro, 'Service Quality for Customers' Satisfaction: A Literature Review', *EMSJ*, vol. 7, no. 1, pp. 267–276, May 2023, doi: 10.59573/emsj.7(1).2023.24.

[27] J. Ruan and D. Xie, 'A Survey on QoE-Oriented VR Video Streaming: Some Research Issues and Challenges', *Electronics*, vol. 10, no. 17, p. 2155, Sept. 2021, doi: 10.3390/electronics10172155.

[28] M. A. Habibi, M. Ulman, J. Vaněk, and J. Pavlík, 'Measurement and Analysis of Quality of Service of Mobile Networks in Afghanistan – End User Perspective', *AOL*, vol. 8, no. 4, pp. 73–84, Dec. 2016, doi: 10.7160/aol.2016.080407.

[29] K. Bouraqia, E. Sabir, M. Sadik, and L. Ladid, 'Quality of Experience for Streaming Services: Measurements, Challenges and Insights', *IEEE Access*, vol. 8, pp. 13341–13361, Jan. 2020, doi: 10.1109/ACCESS.2020.2965099.

[30] Nitin Liladhar Rane, Anand Achari and Saurabh P. Choudhary, 'Enhancing customer loyalty through quality of service: effective strategies to improve customer satisfaction, experience, relationship, and engagement', *IRJMETS*, vol. 5, no. 5, pp. 427-452, May 2023, doi: 10.56726/IRJMETS38104.

[31] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, 'A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction', *IEEE Access*, vol. 10, pp. 19507–19538, Feb. 2022, doi: 10.1109/ACCESS.2022.3149592.

[32] T. Spetebroot, S. Afra, N. Aguilera, D. Saucez, and C. Barakat, 'From network-level measurements to expected quality of experience: The Skype use case', in *2015 IEEE International Workshop on Measurements & Networking (M&N)*, Coimbra, Portugal: IEEE, pp. 1–6, Oct. 2015. doi: 10.1109/IWMN.2015.7322989.

[33] B. Baldovino, 'An Overview of the Networking Issues of Cloud Gaming: A Literature Review', *jinita*, vol. 4, no. 2, pp. 120–132, Dec. 2022, doi: 10.35970/jinita.v4i2.1581.

[34] A. A. Barakabitze, I.-H. Mkwawa, A. Hines, L. Sun, and E. Ifeachor, 'QoEMultiSDN: Management of Multimedia Services using MPTCP/SR in Softwarized and Virtualized Networks', *IEEE Access*, vol. 13, pp. 123151 – 123168, Nov. 2020, doi: 10.1109/ACCESS.2020.3039953.

[35] A. Colarieti, A. Marotta, M. Mpervarakis, L. Pomante, and V. Tsolkas, 'QoE provisioning over mobile networks: The CASPER perspective', in *2017 IEEE 22nd International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, Lund, Sweden: IEEE, pp. 1–5, Jun. 2017, doi: 10.1109/CAMAD.2017.8031639.

[36] W. T. Vambe, 'Fog Computing Quality of Experience: Review and Open Challenges', *International Journal of Fog Computing*, vol. 6, no. 1, pp. 1–16, Jan. 2023, doi: 10.4018/IJFC.317110.

[37] P. Juluri, V. Tamarapalli, and D. Medhi, 'Measurement of Quality of Experience of Video-on-Demand Services: A Survey', *IEEE Commun. Surv. Tutorials*, vol. 18, no. 1, pp. 401–418, Feb. 2015, doi: 10.1109/COMST.2015.2401424.

[38] J. Algar *et al.*, 'A Quality of Experience Management Framework for Mobile Users', *Wireless Communications and Mobile Computing*, vol. 2019, pp. 1–11, Jan. 2019, doi: 10.1155/2019/2352941.

[39] A. Sackl, R. Schatz, and A. Raake, 'More than I ever wanted or just good enough? User expectations and subjective quality perception in the context of networked multimedia services', *Qual User Exp*, vol. 2, no. 1, p. 3, Dec. 2017, doi: 10.1007/s41233-016-0004-z.

[40] Xuanyi Dong , Yan Yan , Mingkui Tan , Yi Yang , and Ivor W. Tsang, 'Late Fusion via Subspace Search With Consistency Preservation', *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 518–528, Aug. 2018, doi: 10.1109/TIP.2018.2867747.

[41] S. Zhou, D. Huang, C. Liu, and D. Jiang, 'Objectivity meets subjectivity: A subjective and objective feature fused neural network for emotion recognition', *Applied Soft Computing*, vol. 122, p. 108889, Jun. 2022, doi: 10.1016/j.asoc.2022.108889.

[42] N. D. Duran and R. Fusaroli, 'Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement', *PLoS ONE*, vol. 12, no. 6, pp. 1–25, Jun. 2017, doi: https://doi.org/10.1371/journal. pone.0178140.

[43] Hanchuan Peng, Fuhui Long, and C. Ding, 'Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy', *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[44] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, 'Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods', *J Big Data*, vol. 11, no. 1, p. 44, Mar. 2024, doi: 10.1186/s40537-024-00905-w.

[45] L. McInnes, J. Healy, and J. Melville, 'UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction', *arXiv*: arXiv:1802.03426, Sept. 2020, doi: 10.48550/arXiv.1802.03426.

[46] S. Schwarzmann, C. Cassales Marquezan, M. Bosk, H. Liu, R. Trivisonno, and T. Zinner, 'Estimating Video Streaming QoE in the 5G Architecture Using Machine Learning', in *Proc. of the 4th Internet-QoE Workshop on QoE-based Analysis and Management of Data Communication Networks*, Los Cabos Mexico: ACM, pp. 7–12, Oct. 2019, doi: 10.1145/3349611.3355547.

[47] L. R. Jimenez, M. Solera, M. Toril, C. Gijon, and P. Casas, 'Content Matters: Clustering Web Pages for QoE Analysis with WebCLUST', *IEEE Access*, vol. 9, pp. 123873–123888, Sept. 2021, doi: 10.1109/ACCESS.2021.3110370.

[48] Y. Zhang *et al.*, 'Gaussian Mixture Model Clustering with Incomplete Data', *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 1s, pp. 1–14, Jan. 2021, doi: 10.1145/3408318.

[49] Z. Kang, C. Peng, Q. Cheng, and Z. Xu, 'Unified Spectral Clustering With Optimal Graph', *AAAI*, vol. 32, no. 1, pp. 3366 -3373, Apr. 2018, doi: 10.1609/aaai.v32i1.11613.

[50] O. Eric U. and O. Michael O., 'Overview of Agglomerative Hierarchical Clustering Methods', *British Journal of Computer, Networking and Information Technology*, vol. 7, no. 2, pp. 14–23, Jun. 2024, doi: 10.52589/BJCNIT-CV9POOGW.

[51] M. F. Rahman, W. Liu, S. B. Suhaim, S. Thirumuruganathan, N. Zhang, and G. Das, 'HDBSCAN: Density based Clustering over Location Based Services', *arXiv*: arXiv:1602.03730, Feb. 2016, doi: 10.48550/arXiv.1602.03730.

[52] Scikit-Learn, 'Clustering'. Scikiet-Learn. Accessed: Mar. 10, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html

[53] Alex Dekhtyar, 'Clustering: Cluster Evaluation and Analysis', *California Polytechnic State University*, San Luis Obispo, CA, 2025. Accessed: Mar. 10, 2025. [Online]. Available: https://users.csc.calpoly.edu/~dekhtyar/466-Winter2025/lectures/lec09.466-new.pdf

[54] E. S. Dalmaijer, C. L. Nord, and D. E. Astle, 'Statistical power for cluster analysis', *BMC Bioinformatics*, vol. 23, no. 1, p. 205, Dec. 2022, doi: 10.1186/s12859-022-04675-1.

[55] Sarah Lee, '7 Metrics and 5 Strategies for Clustering Validity', *Number Analytics*. Accessed: Apr. 29, 2025. [Online]. Available: https://www.numberanalytics.com/blog/7-metrics-and-5-strategies-clustering-validity

[56] S. Watanabe, 'Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance', *arXiv*: arXiv:2304.11127, May. 2023, doi: 10.48550/arXiv.2304.11127.

[57] F. Pedregosa *et al.*, 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research,* vol. 12*,* pp. 2825-2830, Oct. 2011, HAL Id: hal-00650905.

[58] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.

[59] P. Charonyktakis, M. Plakia, I. Tsamardinos, and M. Papadopouli, 'On User-Centric Modular QoE Prediction for VoIP Based on Machine-Learning Algorithms', *IEEE Trans. on Mobile Comput.*, vol. 15, no. 6, pp. 1443–1456, Jun. 2016, doi: 10.1109/TMC.2015.2461216.

[60] J. Wang, L. Wang, Y. Zheng, C.-C. M. Yeh, S. Jain, and W. Zhang, 'Learning-From-Disagreement: A Model Comparison and Visual Analytics Framework', *IEEE Trans. Visual. Comput. Graphics*, vol. 29, no. 9, pp. 3809–3825, Sept. 2023, doi: 10.1109/tvcg.2022.3172107.

[61] K. M. Theophane Osee, V. Nkemeni, M. E. Sone, 'Cameroon Hybrid QoS/QoE Dataset'. Zenodo, Jun. 2025, doi: 10.5281/ZENODO.15730111.

[62] E. Boz, B. Finley, A. Oulasvirta, K. Kilkki, and J. Manner, 'Mobile QoE prediction in the field', *Pervasive and Mobile Computing*, vol. 59, p. 101039, Oct. 2019, doi: 10.1016/j.pmcj.2019.101039.

[63] K. B. Ajeyprasaath and P. Vetrivelan, 'A Hybrid Machine Learning Approach for Improvised QoE in Video Services over 5G Wireless Networks', *CMC*, vol. 78, no. 3, pp. 3195–3213, Mar. 2024, doi: 10.32604/cmc.2023.046911.

[64] A. A. Barakabitze *et al.*, 'QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey', *IEEE Commun. Surv. Tutorials*, vol. 22, no. 1, pp. 526–565, Dec. 2019, doi: 10.1109/COMST.2019.2958784.

[65] K. F. Mccoy, J. L. Bedrosian, L. A. Hoag, and D. E. Johnson, 'Brevity and Speed of Message Delivery Trade-offs in Augmentative and Alternative Communication', *Informa*, vol. 23, no. 1, pp. 76 – 88, Jul. 2009, doi: 10.1080/07434610600924515.

[66] M. Pedersen, '*Image quality metrics for the evaluation of printing workflows*', Ph.D. dissertation, University of Oslo, Oslo, Norway, 2011.

[67] K. M. Theophane Osee, V. Nkemeni, M. E. Sone, 'Quality of Experience Management in Mobile Networks: Techniques, Constraints, and Emerging Trends for Value-Added Services', *Computer Networks and Communications*, vol. 3, no. 2, pp. 21–58, Jul. 2025, doi: 10.37256/cnc.3220256766.

[68] M. M. Yahaya, Y. Surajo, and A. H. Rawayau, 'An enhanced resource allocation scheme for Long Term Evolution advanced network', *FJS*, vol. 7, no. 3, pp. 7–12, May. 2023, doi: 10.33003/fjs-2023-0703-1766.

[69] O. B. Maia, H. C. Yehia, and L. D. Errico, 'A concise review of the quality of experience assessment for video streaming', *Computer Communications*, vol. 57, pp. 1–12, Feb. 2015, doi: 10.1016/j.comcom.2014.11.005.

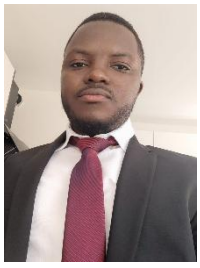
[70] M. Krishnan, 'Against Interpretability: A Critical Examination of the Interpretability Problem in Machine Learning', *Philos. Technol.*, vol. 33, no. 3, pp. 487–502, Sept. 2020, doi: 10.1007/s13347-019-00372-9.

KINGE MBEKE THEOPHANE OSEE received the B.Eng. degree in electrical and electronics engineering in 2015, and the M.Eng. degree in telecommunications and networks engineering in 2017, both from the University of Buea, Buea, Cameroon, where he is currently pursuing the PhD degree in telecommunications and networks engineering. He has over ten years of experience in the telecommunications industry, occupying various roles, including Account Manager at Huawei Technologies, and currently, Business Solutions Manager at Cameroon Telecommunications, Cameroon. His work has focused on digital transformation, Customer Satisfaction, Business and Operations Support Systems integration, network optimisation, and strategic planning. His current research interests include mobile network quality of experience, telecom customer satisfaction, edge computing, telecoms revenue assurance, Internet of Things, Machine Learning/Artificial Intelligence applications in telecoms, digital service delivery, and ICT innovation in emerging markets.

VALERY NKEMENI received the B.Eng. degree in computer engineering in 2015 and the M.Eng. degree in telecommunications and networks engineering in 2017, both from the University of Buea, Buea, Cameroon, and the PhD degree in computer engineering from the Université de Lyon, Lyon, France, in 2021. He is currently a Lecturer with the Department of Computer Engineering and the Head of the Department of Electrical and Electronic Engineering, Faculty of Engineering and Technology, University of Buea, Cameroon. His research interests include green networks, edge computing, telecommunication network optimisation, engineering education, technology adoption, the Internet of Things, and artificial intelligence, with applications in agriculture, healthcare, transportation, and energy systems.

MICHAEL EKONDE SONE received the BSc. from the University of Yaoundé, Cameroon in 1988, the MSc. in electrical engineering from the University of Lagos, Nigeria in 1996, and the PhD in telecommunications engineering from the Ecole Nationale Supérieure Polytechnique, University of Yaoundé, Cameroon in 2012.

He is currently an Associate Professor in Telecommunications and Networks Security and serves as the Deputy Vice-Chancellor at the University of Buea, Buea, Cameroon. His research focuses on designing communication systems that enhance security and throughput across the physical, data link, and network layers of the TCP/IP model. He developed a novel cryptographic algorithm that integrates wavelet lifting schemes (subband coding), non-linear convolutional coding, and RSA public-key cryptography. He has published extensively in peer-reviewed journals and contributed book chapters to the field of cybersecurity, including "New Perspectives in Behavioural Cybersecurity" by Wayne Patterson and "Computer and Network Security" by J. Sen.



GODLOVE SUILA KUABAN received the B.Eng. degree in Electrical and Electronics Engineering, with a speciality in Telecommunications, from the University of Buea, Cameroon, in 2014, and an MSc. Degree in Inter-disciplinary Studies of Automatic Control, Robotics, Electronics, Telecommunications, and Computer Science, and specialised in Computer Science from the Silesian University of Technology, Gliwice, Poland, in 2017. He obtained a PhD degree in Telecommunications and Technical Computer Science from the Silesian University of Technology, Gliwice, Poland, in 2023.

From 2017 to 2023, he served as a Research Assistant at the Institute of Theoretical and Applied Informatics, Polish Academy of Sciences (IITiS-PAN), Gliwice, Poland. He is currently an Assistant Professor at IITiS-PAN. His research interests include computer systems modelling and performance evaluation. Specifically, modelling and evaluations of Software Defined Networking (SDN) and IoT networks, and energy performance of green networks (e.g., IoT and cellular mobile networks, linear wireless sensor networks). He has participated in six EU-funded research grants: three in IoT security, one in building an IoT laboratory testbed, one in developing online education resources on Assembly Language Programming, and one on Reliable Electronics for Tomorrow's Active Systems. He is currently working on the Celtic-Next project RAI-6Green: Robust and AI-Native 6G for Green Networks, with partners in France. Additionally, he received the Best Paper Award at the International Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2023) in Stony Brook, NY, USA, and at MASCOTS 2024 in Krakow, Poland (published by IEEE).