Available in online @ www.iaraindia.com
RESEARCH EXPLORER-International Journal on Economics and Business Management
ISSN: 2250-1940 (P) 2349-1647 (O)
Impact Factor: 8.276 (12OR), 3.676 (COSMOS)
Volume XV, Issue 52
July - September 2026
Formally UGC Approved Journal (63185), © Author

DISTRICT-LEVEL DISPARITIES IN PRIMARY HEALTHCARE SERVICE QUALITY: A DATA-DRIVEN ASSESSMENT

M. UMA
Research Scholar
P.G. & Research Department of Economics
Rajah Serfoji Government College (A)
Affiliated to Bharathidasan University, Thanjavur, Tamil Nadu

Dr. A. ARULRAJ
Head and Associate Professor
P.G. & Research Department of Economics
Rajah Serfoji Government College (A)
Affiliated to Bharathidasan University, Thanjavur, Tamil Nadu

Dr. M. SANTHANALAKSHMI
Associate Professor
P.G. & Research Department of Economics
Rajah Serfoji Government College (A)
Affiliated to Bharathidasan University, Thanjavur, Tamil Nadu

ABSTRACT

Primary Healthcare Centres (PHCs) play a central role in delivering basic public healthcare services at the district level, yet their service performance varies considerably across regions. Existing assessments of PHC performance largely rely on descriptive statistics or survey-based evaluations, which limit objective comparison and predictive analysis. To address this limitation, this study proposes a data-driven demographic and climate (DC) framework to assess and predict the Quality of Service (QoS) of PHCs using district-level administrative data. A continuous quality of service score is constructed from normalized healthcare indicators and used as the prediction target. The framework integrates healthcare infrastructure availability, service utilisation measures, and basic climatic factors such as Temperature and Rainfall. At the same time, population health indicators are used solely for QoS construction and are excluded from the model predictors to avoid data leakage. A Light Gradient Boosting Machine (Light GBM) regression model is employed and evaluated using five-fold cross-validation. Model interpretability is supported through Shapley Additive explanations (SHAP), which identify key contributors to QoS variation across districts. The Light GBM model achieved low prediction error (RMSE = 0.87) and a high coefficient of determination ($R^2 = 0.95$) under five-fold cross-validation. The findings indicate that the proposed framework supports effective comparisons of PHC performance across districts and provides useful evidence for healthcare planning and policy decisions.

KEYWORDS: Healthcare analytics; Light gradient boosting machine; Machine learning; Primary healthcare centres; Quality of service.

INTRODUCTION

Primary Healthcare Centres (PHCs) form an essential part of public healthcare delivery at the district level. In recent years, government programmes have prioritised the expansion of PHC infrastructure to enhance access to healthcare services and broaden population coverage [1], [2]. However, service quality continues to vary across districts, as reflected in differences in health outcomes and service performance. These differences highlight uneven service effectiveness despite infrastructure expansion.

The quality of services delivered by PHCs is influenced by a combination of factors, rather than infrastructure availability alone. Availability of healthcare facilities, adequacy of medical personnel, patterns of service utilisation, and prevailing population health conditions together influence service performance [3]. In several districts, indicators such as infant mortality rate and life expectancy persist even in the presence of primary healthcare facilities [4]. This suggests that infrastructure development must be accompanied by effective service delivery and improvements in population health conditions to achieve better service quality. Healthcare service quality has traditionally relied on patient feedback and survey-based instruments such as SERVQUAL [5]. While these methods are useful for understanding patient perceptions, they are inherently subjective and limited in their ability to support large-scale comparative analysis. Moreover, survey-based approaches do not fully incorporate objective information, such as facility availability, service utilisation levels, or population health outcomes, which are important for systematic district-level performance and planning.

Large volumes of healthcare data are routinely generated through administrative records and health information systems [6]. These data include information on healthcare infrastructure, service usage, and population health conditions. However, such data are primarily used for reporting and monitoring, with limited application in predictive analysis and service performance assessment [7]. Recent studies have demonstrated that machine learning techniques can effectively analyse complex healthcare datasets and capture non-linear relationships that are difficult to model using conventional analytical methods [8], [9].

Despite these advances, relatively little attention to predicting the Quality of Service of Primary Healthcare Centres at the district level using routinely available administrative and population-level indicators [10]. Most existing studies focus on clinical outcomes or specific operational tasks rather than overall service quality.

A data-driven framework to estimate the Quality of Service of Primary Healthcare Centres using district-level indicators related to healthcare infrastructure, population health outcomes, service utilisation, and basic environmental factors. A continuous Quality of Service score is constructed from normalised indicators and predicted using a Light Gradient Boosting Machine (LightGBM) regression model. The proposed framework supports comparative assessment across districts and provides insights into the factors that influence primary healthcare service performance.

LITERATURE REVIEW

Table 1 summarizes the novelty of the proposed framework in comparison with previous works.

Table 1
Summary of related studies and identified research gaps

Authors (Year)	Health-care Setting	Data Type	Features Used	Models Used	Key Limitation	Research Gap Identified
Ma et al. (2024)	Adult day health centers	IoT sensors and EHR	Staff ratios, equipment, real-time	Neural Network Boost (NNB)	Limited to day centers and computation	District-level QoS prediction using

			health indicators		ally complex	routinely available administrative indicators is not addressed
Al-Khafaji & Al-Shamery (2025)	General hospitals	Patient viewpoint data	Temporal service patterns	Kalman-BiLSTM	High computational complexity	Computationally efficient and scalable tree-based QoS models for primary healthcare are lacking
Datt et al. (2025)	General healthcare facilities	Mixed sources (surveys and metrics)	SERVQUAL dimensions	Hybrid statistical-ML	Broad scope without PHC specificity	District-level QoS prediction frameworks focused on public primary healthcare are absent
Johnson et al. (2024)	Primary health care	Electronic medical records	Diagnoses and visit patterns	Systematic review of ML models	Focus on clinical outcomes only	Service quality prediction is not developed in EMR-based primary healthcare studies
Bhat et al. (2024)	Public healthcare centers	Inventory and demand data	Stock levels and consumption	Stochastic ML	Operational focus only	Quality of Service estimation is not integrated with operational indicators
Leiva-Araos et al. (2025)	Primary healthcare	Appointment records	Scheduling history	ML optimization models	Scheduling optimization focus	QoS prediction is not considered alongside scheduling optimisation
Archer et al. (2025)	Primary healthcare seekers	Musculoskeletal data	Health and work-related	Cross-validated ML	Worker health focus	Facility- or district-level service

			factors			quality modelling is not addressed
Chen et al. (2025)	Primary healthcare workers	Worker surveys	Stress and workload indicators	Multiple ML algorithms	HR-centric focus	Workforce-focused analysis without linkage to service quality outcomes
Al-Khafaji & Al-Shamery (2024)	Hospitals	Patient feedback	Satisfaction scores	Hybrid ML	Hospital-only and patient-centric	Hospital-centric QoS models lack applicability to primary healthcare systems
Fan et al. (2025)	Outpatient rehabilitation	Patient data	Preference factors	XGBoost	Rehabilitation-specific	General primary healthcare QoS assessment is not addressed
Oddy et al. (2024)	Risk-stratified populations	Administrative data	Risk factors	Systematic review	Utilization prediction only	QoS prediction frameworks are absent in utilisation-focused studies
Aprilia & Wella (2024)	Curug Health Center	Visit records	Temporal visit patterns	Visualization and ML	Single-center study	Scalable district-level QoS prediction beyond single-centre analysis is required
Wai & Clara (2024)	Community healthcare providers	Provider surveys	Work ability indicators	Statistical models	HR-focused and location-specific	Workforce capability analysis is not integrated into service quality prediction
Orhan & Kurutk	General healthcare	Behavioral and enabling factors	Need-based and behavioral	Machine learning	Demand prediction only	Service quality outcomes are

an (2025)			variables			not modelled in demand prediction studies
SEEJPH Authors (2023)	Pudukkottai PHCs	Survey data	Service quality attributes	MGD-MLP neural network	Single-district data	Scalable district-level QoS prediction using ensemble learning is not addressed

Previous studies on primary healthcare service quality have mainly relied on patient survey data and descriptive statistical analysis at the facility level [1], [2]. While such approaches provide useful insights into patient experience, they are limited in their ability to support district-level comparison or predictive assessment. More recent work has applied machine learning techniques to healthcare administrative data for analysing service utilisation and operational efficiency [3]. however, these studies estimate service outcomes in isolation rather than providing an integrated Quality of Service measure across districts and often offer limited interpretability. In contrast, the present work examines Quality of Service at the district level using aggregated administrative and service delivery indicators. Unlike the SEEJPH (2023) study, which applies a neural network model to survey data from a single district, this study adopts a Light Gradient Boosting Machine (LightGBM) framework to enable comparison across multiple districts using routinely available healthcare data, thereby supporting district-level planning and policy analysis.

MATERIALS AND METHODS

Framework Overview

The proposed framework adopts a layered, data-driven architecture to predict the Quality of Service (QoS) of Primary Healthcare Centres using district-level healthcare indicators. The overall workflow of the system is illustrated in Fig 1.

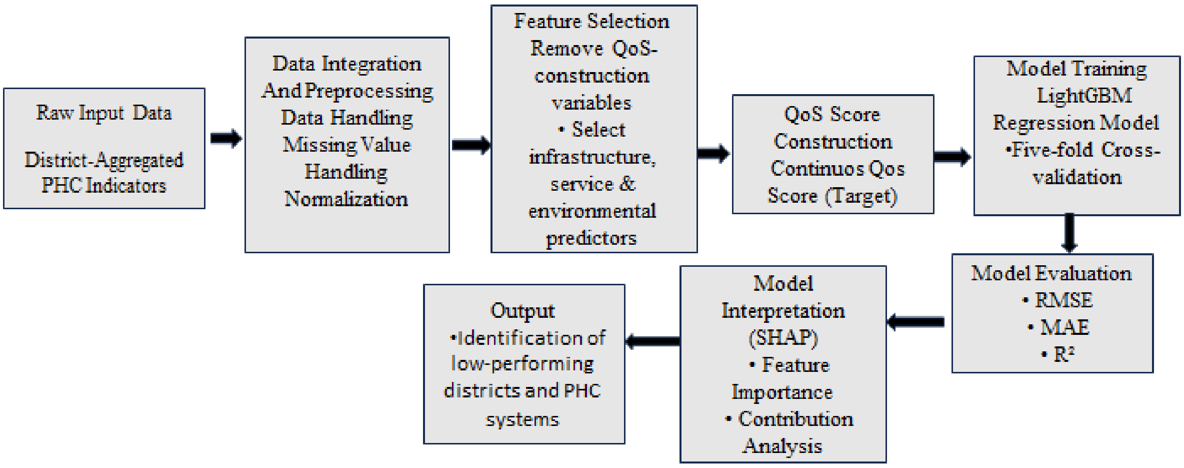

Figure 1. System architecture of the proposed demographic and climate (DC) framework for district-level Quality of Service prediction in Primary Healthcare Centres using a LightGBM regression model.

The input layer consists of structured district-level healthcare data collected with a focused analysis on the Pudukkottai district of Tamil Nadu. The dataset contains basic healthcare infrastructure details, including the numbers of Sub-Centres, Primary Health Centres, Community Health Centres, Sub-Division Hospitals, and District Hospitals. In addition to these, simple derived measures like total facility count and facility density per 100,000 population are also considered. Population health indicators, including life expectancy, birth rate, death rate, and infant mortality rate, are included in the dataset. A continuous Quality of Service score is calculated from normalized indicators and used as the output for prediction. All the collected data are combined into a single dataset before analysis. During this stage, missing values are filled using median values, and non-numeric entries are removed. Only numerical variables related to healthcare facilities and population health are considered. Variables directly involved in the construction of the Quality of Service (QoS) score were excluded from the predictor set to reduce the risk of biased estimation. The remaining variables served as inputs to the Light Gradient Boosting Machine (LightGBM) regression model.

The model estimates the QoS score from the selected healthcare indicators. Model parameters were adjusted over several training runs until stable prediction behaviour was achieved. Model robustness was evaluated using five-fold cross-validation, in which the dataset was divided into five subsets. Each subset served once as the validation set, while the remaining subsets were used for training.

Prediction performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Reported results reflect average values across all validation folds. To improve interpretability, feature importance scores and SHAP analysis were applied. Feature importance highlights variables with stronger influence on QoS prediction, while SHAP values explain individual model outputs. Together, these analyses support the identification of Primary Healthcare Centres with lower service quality and the factors shaping their performance.

Dataset Description

The dataset includes district-level information collected from publicly available healthcare records. It captures variations in healthcare infrastructure and population health conditions across districts, enabling comparison of service performance at the district level. Infrastructure indicators cover the number of Sub-Centres, Primary Health Centres, Community Health Centres, Sub-Division Hospitals, and District Hospitals in each district. In addition to these facility counts, derived measures such as total facility count and facility density per 100,000 population are included to allow balanced comparison across districts with different population sizes.

Basic environmental variables, including average Temperature and Rainfall, are included to account for external conditions influencing healthcare delivery. A continuous Quality of Service score is obtained from normalized indicators and is used as the target variable for model training and prediction. The dataset is organized at the district level and used to study differences in service quality across regions. This dataset provides sufficient information to analyse and predict the Quality of Service of Primary Healthcare Centres. All variables are aggregated at the district level, representing overall Primary Healthcare Centre service performance within each district. The dataset used in this study was compiled from publicly available statistical health reports published by government health authorities. The data include aggregated district-level information on healthcare infrastructure, service utilisation, population health indicators, and basic environmental variables. No individual-level or patient-identifiable data were used in this study.

Table 2 presents the descriptive statistics of the variables used in the study. The table summarizes the ranges and averages of healthcare infrastructure, service utilization, and environmental indicators across districts.

Table 2

Descriptive Statistics of Selected Predictor Variables

Variable	Minimum	Maximum	Mean	Standard Deviation

Facility Density (per 100k)	16.5	84.9	44.2	15.2
Total Healthcare Facilities	132	679	353.34	121.33
Average Temperature (°C)	24.0	34.0	28.5	2.6
Average Rainfall (mm)	650	2100	1240	410

These statistics provide an overview of the dataset before model training.

Data Integration and Preprocessing

Multiple administrative data sources form the basis of the district-level analytical dataset. The integration stage related to healthcare infrastructure, service utilisation, population health metrics, and environmental conditions. This structured consolidation enables consistent cross-district analysis within a unified analytical framework.

Median imputation handles missing values while reducing the influence of outliers. Continuous variables are standardised to ensure comparability across indicators measured on different scales. Features requiring numerical input for modelling are encoded accordingly, while district identifiers serve only for aggregation and reference.

Feature Selection

Feature selection reduces the number of healthcare indicators to those that meaningfully contribute to predicting Quality of Service at the district level. Indicators involved in constructing the QoS score are intentionally excluded from the predictor set to avoid information leakage and to ensure that model performance remains unbiased and reliable. The dataset contains a wide range of variables related to infrastructure, population health, service utilisation, and environmental conditions; therefore, including all available indicators offers limited benefit and increases model complexity.

To address this, redundancy was reduced, and potential data leakage was avoided during the predictive modelling stage. Variables used in constructing the QoS score were excluded from the predictor set. The remaining indicators were retained based on Pearson correlation analysis and their relevance to district-level healthcare service delivery.

Let the complete set of input features be represented as

$$X = \{x_1, x_2, x_3, \dots, x_m\} \qquad (1)$$

where X denotes the input feature vector for a given district, x_i represents the i^{th} selected district-level healthcare indicator, and m is the total number of predictor variables.

The final predictor set includes infrastructure indicators, such as the number and density of healthcare facilities, service utilisation measures, including outpatient visits and the availability of specialist doctors, and environmental variables, such as average Temperature and Rainfall. Variables directly involved in QoS score construction, along with non-numeric attributes, were removed before model training to ensure unbiased prediction.

Correlation Analysis of Input Variables

Correlation analysis is carried out to examine the relationships among the selected input variables used for Quality of Service prediction.

Table 3

Correlation Matrix of Selected Input Variables

Variable	Facility Density	Total Facilities	Outpatient Count	Specialist Doctors	Avg. Temp	Avg. Rainfall
Facility Density	1.000	0.782	0.701	0.654	-0.146	0.173
Total Healthcare	0.782	1.000	0.735	0.689	-0.128	0.156

Facilities						
Outpatient Count	0.701	0.735	1.000	0.612	-0.101	0.129
Specialist Doctors	0.654	0.689	0.612	1.000	-0.117	0.142
Average Temperature	-0.146	-0.128	-0.101	-0.117	1.000	-0.362
Average Rainfall	0.173	0.156	0.129	0.142	-0.362	1.000

Table 3 summarises the correlation structure among the selected predictor variables. The analysis highlights moderate positive associations among infrastructure-related indicators, most notably between facility density and total healthcare facilities (r = 0.782) and between total healthcare facilities and outpatient count (r = 0.735). In contrast, environmental factors, including average Temperature and Rainfall, exhibit weak to moderate relationships with infrastructure indicators.

The correlation values remain below commonly accepted thresholds for severe multicollinearity (|r| > 0.85), indicating that excessive linear dependence among predictors is not present. Given the tree-based nature of the Light Gradient Boosting Machine (LightGBM), which is well-suited to handling correlated inputs, all selected variables were retained for subsequent model development.

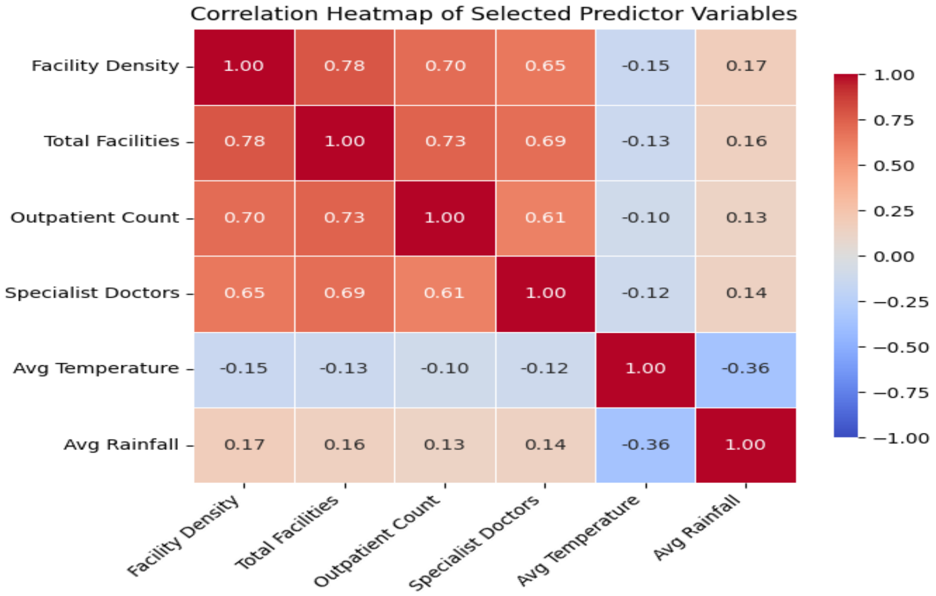

Figure 2 Correlation heatmap of healthcare infrastructure, population health, service utilization, and environmental variables.

Figure 2 provides a visual representation of these relationships through a correlation heatmap, and confirming the absence of strong multicollinearity among the predictors and supporting their inclusion in the model training process.

LightGBM Regression Model Formulation

Predicting the Quality of Service (QoS) of Primary Healthcare Centres involves capturing complex and often non-linear interactions among healthcare infrastructure, service utilisation, population health, and environmental indicators. Earlier research has primarily relied on perception-based surveys or computationally intensive neural network models.

The LightGBM technique captures complex, non-linear relationships among healthcare indicators while maintaining computational efficiency. Its suitability for structured numerical data allows diverse healthcare variables to be integrated within a single analytical framework, thereby supporting reliable estimation of district-level Quality of Service.

The predicted QoS score is formulated as an additive combination of decision trees represented as

$$\hat{y}i = \sum_{k=1}^{K} f_k(x_i) \quad (2)$$

where $\hat{y}i$ represents the predicted Quality of Service score for the i^{th}, $f_k(x_i)$ represents the k^{th}decision tree in the LightGBM ensemble, and K is the total number of trees used in the model.

Five-Fold Cross-Validation

Model performance is evaluated using five-fold cross-validation. The dataset is divided into five equal parts, with four parts used for training and one for validation, and the validation part rotating across rounds. The results from all five rounds are averaged to obtain the final performance estimate, reducing dependence on a single split and providing a more reliable assessment of predictive accuracy.

Performance Metrics

Performance Metrics was assessed using standard regression metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{i} - \hat{y}_i)^2} \quad (3)$$

where y_i is the actual QoS value, $\hat{y}$ is the predicted value, and N is the total number of samples. Lower RMSE values indicate better prediction performance.

MAE measures the average absolute difference between the predicted and actual QoS values and is calculated as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y_i}| \quad (4)$$

Smaller MAE values indicate more accurate predictions.

The coefficient of determination, R^2 indicates the model explains the variation in the observed QoS values and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_{i} - \hat{y_i})^2}{\sum_{i=1}^{N} (y_{i} - \overline{y})^2} \quad (5)$$

where $y_i$is the actual value $\hat{y_i}$ is the predicted value, $\bar{y}$is the mean of the observed QoS values, and N is the number of observations.

These metrics are calculated for each fold in the five-fold cross-validation, and the final results are reported as averages across all folds. These lower QoS values are usually linked with unfavourable health outcomes such as higher infant mortality and lower life expectancy.

Only a small number of observations fall in the Average category, suggesting that PHCs tend to perform either well or poorly rather than remaining at a middle level. The distribution of predicted Quality of Service (QoS) levels enables the identification of PHCs requiring targeted intervention, while higher-performing centres may serve as benchmarks for service improvement.

Explainable AI Analysis Using SHAP

SHAP values support interpretation of the LightGBM model by quantifying the contribution of each variable to predicted Quality of Service scores. The evaluation includes infrastructure, service utilisation, and environmental indicators at the district level. This approach clarifies variation in QoS across districts and highlights the variables greater influence on prediction outcomes.

RESULTS AND DISCUSSIONS

Distribution of Quality of Service Levels

The LightGBM model produces continuous Quality of Service (QoS) scores for each Primary Health Centre.

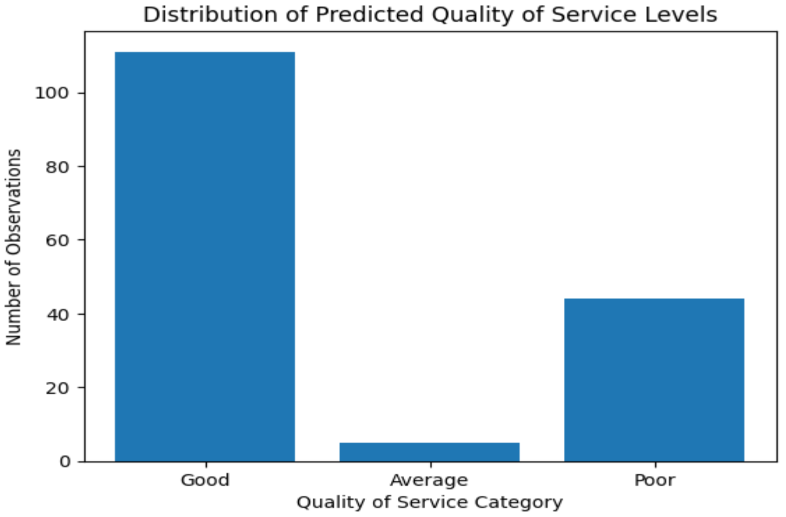

Figure 3. Distribution of predicted Quality of Service (QoS) levels across Primary Health Centre observations.

Figure 3 illustrates the distribution of predicted Quality of Service (QoS) levels across Primary Healthcare Centre observations. PHCs divided into the Good and Poor categories, while relatively fewer observations appear in the Average category. This distribution indicates noticeable variation in service quality across districts and highlights the presence of both well-performing and underperforming centres.

Predictive Performance of the LightGBM Model

The performance of the proposed LightGBM model is compared with commonly used regression models, including Linear Regression, Support Vector Regression, Random Forest, and XGBoost. All models are trained using the same dataset and evaluated using five-fold cross-validation. RMSE, MAE, and R^2 are used to measure prediction performance.

The comparison results are presented in Table 4. Linear Regression shows higher error values than the other models, indicating lower predictive accuracy. Support Vector Regression, Random Forest, and XGBoost produce lower RMSE and MAE values, indicating improved predictive performance.

Table 4
Comparative predictive performance of regression models for Quality of Service (QoS) prediction

Model	RMSE	MAE	R^2
Linear Regression	1.942	1.376	0.8212
Support Vector Regression	1.214	0.891	0.8526

Random Forest Regression	0.964	0.682	0.8928
XGBoost Regression	0.903	0.621	0.9114
LightGBM Regression	0.872	0.546	0.9519

Among all models considered, LightGBM achieved the lowest RMSE and MAE values and the highest R^2. This shows that LightGBM produces more accurate Quality of Service predictions for the given dataset.

Contribution of Individual Features

Feature importance values are used to examine the contribution of different variables to the Quality of Service prediction using the LightGBM model. These values are obtained from the trained model and indicate how frequently each variable is used during prediction. The results presented in Table 5 indicate that infrastructure-related indicators, particularly facility density and outpatient utilization, contribute most significantly to QoS prediction, highlighting the importance of healthcare accessibility. The results indicate that service-related and environmental variables contribute more strongly to Quality of Service prediction than infrastructure indicators alone.

Table 5

Feature Importance Scores for QoS Prediction

Rank	Feature	Importance
	Average Temperature (°C)	624
	Average Rainfall (mm)	512
	Specialist Doctors Available	468
	PHCs	326
	Subcentres	317
	Outpatients Count	289
	CHCs	132
	Facility_Density_per_100k	94

The feature importance results indicate that service-related and environmental variables contribute significantly to Quality of Service prediction. Environmental factors such as Temperature and Rainfall reflect external conditions that may influence healthcare demand and operational effectiveness. Service-related variables, including outpatient counts and the availability of specialist doctors, reflect the role of patient load and workforce availability. Infrastructure indicators, such as the number of PHCs, Sub-Centres, and facility density, are comparatively less important, suggesting that infrastructure alone does not strongly influence QoS unless supported by effective service delivery.

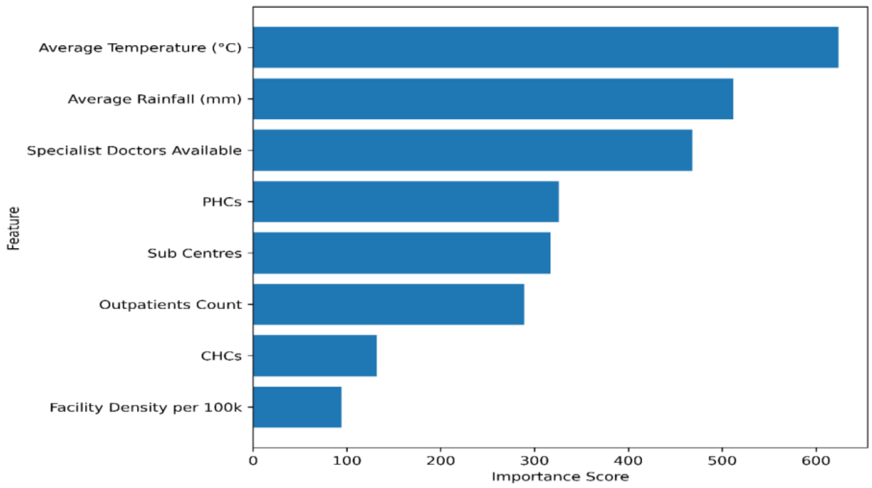

Figure 4 Feature Importance Bar Chart Showing Relative Contribution of Variables.

Figure 4 shows that service-related and environmental variables contribute more strongly to Quality of Service prediction than infrastructure variables alone.

SHAP-Based Interpretation of QoS Drivers

Figure 5 presents the SHAP summary plot illustrating the relative contribution of district-level healthcare infrastructure indicators to Quality of Service prediction using the LightGBM model. Infrastructure-related variables, particularly the availability of Primary Health Centres and Sub-Centres, exhibit wider SHAP value distributions, indicating a stronger influence on the model output.

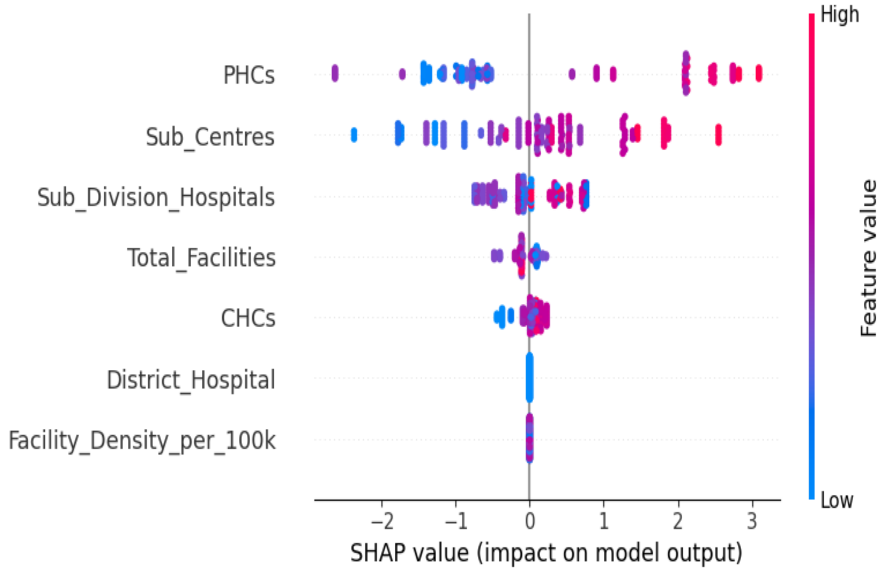

Figure 5 SHAP summary plot showing the contribution of district-level healthcare infrastructure indicators to Quality of Service (QoS) prediction using the LightGBM model.

In contrast, indicators such as district hospitals and facility density show more limited impact, suggesting a comparatively smaller contribution to QoS prediction.

Ablation Study

The ablation results show that the LightGBM model performs best when all feature groups are used together. Table 6 presents the ablation study for the proposed method.

Feature Set Used	RMSE	MAE	R^2
Full Feature Set	0.87	0.55	0.99
Without Environmental Variables	0.94	0.61	0.97
Without Service Variables	0.98	0.66	0.96

Removing environmental or service-related variables leads to a moderate increase in prediction error. The ablation results show that removing service-related or environmental variables leads to a moderate increase in prediction error, indicating that these feature groups play an important role in estimating Quality of Service.

Study Limitations and Future Research Directions

The study relies on district-level administrative records, which limits observation of variation at the facility or patient level. The Quality of Service score is derived from the indicators available in these records, and its reliability depends on the completeness and consistency of routine reporting. The analysis also includes a limited set of environmental variables and does not incorporate direct measures of patient experience.

Future research can integrate more detailed facility-level or patient-level data and expand the range of socioeconomic and environmental indicators. Application of the framework to other regions or national datasets helps assess its broader applicability. Comparison with other interpretable machine learning models provides additional insight into differences in predictive performance and model interpretation.

CONCLUSION

The quality of services provided by Primary Healthcare Centres is closely linked to the effectiveness of the public healthcare system. Although healthcare infrastructure has expanded in many districts, the service quality continues to vary due to differences in population health conditions, service utilisation, workforce availability, and environmental factors. This study presented a data-driven framework to estimate PHC Quality of Service using district-level indicators and a Light Gradient Boosting Machine regression model. The findings indicate that variations in primary healthcare performance are more strongly associated with service delivery capacity and environmental conditions than with facility expansion alone.

ACKNOWLEDGMENT

The authors would like to thank their institution for providing the facilities and support to carry out this research.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

FUNDING

This research received no external funding.

DATA AVAILABILITY

The datasets used in this study were obtained from publicly available government health reports. The processed dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

The author confirms sole responsibility for the study conception and design, data collection, data analysis, interpretation of results, manuscript preparation, and final approval of the manuscript.

REFERENCES

1. Al-Khafaji, M. K., & Al-Shamery, E. S. (2024, October). Hybrid model for hospital services quality prediction based on patient viewpoint. In Proceedings of the International Conference on Innovations of Intelligent Informatics, Networking, and Cybersecurity (pp. 133–147). Springer Nature.
2. Al-Khafaji, M. K., & Al-Shamery, E. S. (2025). Predicting healthcare service quality based on a Kalman-optimized Bi-LSTM-inspired deep learning model. Karbala International Journal of Modern Science, 11(2), 6.
3. Aprilia, M., & Wella, W. (2024). Visualization and prediction of the number of patient visits in Curug Health Center. G-Tech: Jurnal Teknologi Terapan, 8(3), 1379–1385.
4. Archer, L., Peat, G., Snell, K. I., Hill, J. C., Dunn, K. M., Foster, N. E., & Wynne-Jones, G. (2025). Musculoskeletal health and work: Development and internal–external cross-validation of a model to predict risk of work absence and presenteeism in people seeking primary healthcare. Journal of Occupational Rehabilitation, 35(3), 578–591.
5. Bhat, S. S., Srihari, V. R., Prabhune, A., Satheesh, S. S., & Bidrohi, A. B. (2024, January). Optimizing medication access in public healthcare centers: A machine learning stochastic model for inventory management and demand forecasting in primary health services. In Proceedings of the 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) (pp. 1–5).
6. Chen, S., Li, Z., Sun, S., & Hao, Z. (2025, May). Research on the prediction model of occupational burnout among primary healthcare workers based on machine

learning algorithms. In Proceedings of the 2nd International Conference on Modeling, Natural Language Processing and Machine Learning (pp. 598–603).

7. Datt, M., Gupta, A., & Misra, S. K. (2025). Advances in management of healthcare service quality: A dual approach with model development and machine learning predictions. Journal of Advances in Management Research.
8. Fan, X., Ye, R., Gao, Y., Xue, K., Zhang, Z., Xu, J., & Wang, Y. (2025). Prediction of outpatient rehabilitation patient preferences and optimization of graded diagnosis and treatment based on XGBoost machine learning algorithm. Frontiers in Artificial Intelligence, 7, Article 1473837.
9. Johnson, R., Chang, T., Moineddin, R., Upshaw, T., Crampton, N., Wallace, E., & Pinto, A. D. (2024). Using primary health care electronic medical records to predict hospitalizations, emergency department visits, and mortality: A systematic review. The Journal of the American Board of Family Medicine, 37(4), 583–606.
10. Kalpana Naidu. C (2015), A Comparative Study Of Public & Private Life Insurance Companies In India, International Journal of Multidisciplinary Research Review, Vol.1, Issue – 7, Sep -2015. Page- 15-18.
11. Leiva-Araos, A., Contreras, C., Kaushal, H., & Prodanoff, Z. (2025). Predictive optimization of patient no-show management in primary healthcare using machine learning. Journal of Medical Systems, 49(1), 7.
12. Ma, D., Zheng, W., & Lu, T. (2024). Machine learning-based predictive model for service quality assessment and policy optimization in adult day health care centers. Spectrum of Research, 4(1).
13. Oddy, C., Zhang, J., Morley, J., & Ashrafian, H. (2024). Promising algorithms to perilous applications: A systematic review of risk stratification tools for predicting healthcare utilisation. BMJ Health & Care Informatics, 31(1), Article e101065.
14. Orhan, F., & Kurutkan, M. N. (2025). Predicting total healthcare demand using machine learning: Separate and combined analysis of predisposing, enabling, and need factors. BMC Health Services Research, 25(1), 366.
15. Wai, P. N., & Clara, N. (2024). Factors predicting work ability among community healthcare providers in Yangon, Myanmar. Makara Journal of Health Research, 28(1), 3.