Aspect-Aware Patient Feedback Analytics and Edge-Accelerated Time-Series Demand Forecasting for Remote Healthcare Delivery Platforms

Authors

James Abella, Christiana Sartorio, Sukamo Juliet, Jayar Tolosa, Perez Jude, Billy Elly, Bente Gonzale

Date; August 13, 2026

Abstract

The proliferation of remote healthcare delivery platforms has generated unprecedented volumes of unstructured patient feedback and high-frequency demand signals, yet existing analytics frameworks treat these data streams in isolation, limiting their utility for operational decision-making. This study addresses the research gap by proposing an integrated framework that combines aspect-aware sentiment analysis of patient feedback with edge-accelerated time-series demand forecasting. A DeBERTa-v3-based aspect classification model was developed to extract fine-grained sentiment across service quality dimensions from patient comments, achieving an accuracy of 89.4% on a corpus of 25,000 annotated patient reviews. Concurrently, a hybrid CNN-LSTM architecture was deployed at the edge to forecast service demand across 15-minute intervals, attaining 94.7% accuracy in predicting request volumes up to 48 hours in advance . The framework's key contribution lies in demonstrating that aspect-level sentiment patterns serve as leading indicators for demand fluctuations, enabling proactive resource allocation. Practical implications include reduced wait times, improved patient satisfaction, and bandwidth-optimized edge-cloud orchestration.

Keywords: Aspect-Aware Sentiment Analysis, Edge Computing, Time-Series Forecasting, Remote Healthcare, CNN-LSTM, Patient Feedback Analytics

1. Introduction

1.1 Background

Remote healthcare delivery platforms have experienced exponential growth, accelerated by technological advances in telemedicine and shifting patient preferences toward convenient, accessible care. These platforms generate two distinct but interconnected data streams: qualitative patient feedback expressed through reviews, surveys, and chat interactions, and quantitative demand signals captured as appointment requests, consultation volumes, and service utilization patterns . The ability to analyze these streams holistically presents a significant opportunity to enhance operational efficiency and patient experience.

Recent advances in natural language processing have enabled sophisticated analysis of unstructured patient feedback. Traditional approaches relied on rule-based sentiment tools such as VADER, which, while resource-efficient, struggle with context, ambiguity, and domain-specific terminology . More recent work has demonstrated the efficacy of transformer-based architectures for aspect-level sentiment classification, allowing healthcare providers to identify specific service dimensions—such as wait times, clinician communication, or facility cleanliness—that drive patient satisfaction or dissatisfaction . However, the application of these techniques to healthcare feedback remains limited, with most studies relying on online platforms rather than platform-generated survey data .

Simultaneously, demand forecasting has evolved from statistical methods to deep learning approaches capable of capturing complex temporal patterns. The hybrid CNN-LSTM architecture has proven particularly effective for time-series prediction, combining CNN's feature extraction capabilities with LSTM's sequential modeling strength . However, centralized cloud-based deployment introduces latency and bandwidth constraints that are problematic for real-time operational decisions. Edge computing offers a solution by enabling localized processing, reducing response times, and preserving data privacy—critical considerations in healthcare contexts where patient data protection is paramount .

1.2 Problem Statement

Despite advances in both sentiment analysis and demand forecasting, two critical gaps persist in the remote healthcare domain. First, existing approaches treat patient feedback and demand data as independent streams, failing to exploit the predictive relationship between patient experience signals and service demand. Research has shown that patient satisfaction correlates with service quality metrics, yet no validated framework exists that operationalizes aspect-level sentiment as a leading indicator for demand prediction . Second, the computational demands of modern deep learning models conflict with the latency requirements of real-time healthcare operations, while edge deployment of these models remains underexplored in healthcare contexts.

Current sentiment analysis of patient feedback is limited by the predominance of rule-based methods that are domain-independent and struggle with healthcare-specific terminology . When deep learning approaches are employed, they typically perform binary or ternary classification rather than aspect-aware analysis, obscuring actionable insights about specific service dimensions . Similarly, demand forecasting models face the challenge of balancing predictive accuracy with deployment feasibility, as high-performing deep learning models require significant computational resources that may not be available at the network edge .

The unsolved issue, therefore, is the lack of an integrated, edge-deployable framework that (1) performs aspect-aware sentiment analysis on patient feedback, (2) captures the predictive relationship between aspect-level sentiment and service demand, and (3) enables real-time operational decisions through edge-accelerated inference.

1.3 Objectives of the Study

General objective:
To develop and validate an integrated framework that combines aspect-aware patient feedback analytics with edge-accelerated time-series demand forecasting for remote healthcare delivery platforms.

Specific objectives:

1. To identify key aspect dimensions in patient feedback that serve as predictors of service demand fluctuations.
2. To design a DeBERTa-v3-based aspect classification model capable of fine-grained sentiment extraction from patient comments with high accuracy.
3. To develop a hybrid CNN-LSTM demand forecasting model deployed at the edge and evaluate its predictive performance against cloud-only and static baseline methods.
4. To validate the integrated framework using real-world patient feedback and service utilization data, assessing the contribution of sentiment features to forecasting accuracy.

1.4 Research Questions

1. What combination of aspect-level sentiment dimensions most accurately predicts fluctuations in remote healthcare service demand?
2. How does the proposed edge-deployed hybrid CNN-LSTM framework compare to cloud-based and traditional forecasting methods in terms of predictive accuracy and inference latency?

3. What are the practical implementation barriers and optimization strategies for deploying aspect-aware analytics and edge-accelerated forecasting in operational healthcare settings?

1.5 Significance of the Study

For healthcare administrators and practitioners: This research provides an actionable framework for proactive resource allocation, enabling platforms to anticipate demand surges triggered by patient experience signals. Administrators can monitor aspect-level sentiment trends as early warning indicators and adjust staffing, scheduling, and capacity accordingly, potentially reducing patient wait times and improving service quality.

For policymakers: The framework supports evidence-based decisions regarding telemedicine infrastructure investment, particularly in underserved areas where bandwidth limitations and latency concerns are acute. Demonstrating the viability of edge-accelerated analytics may inform policies promoting distributed computing models in healthcare.

For academic literature: This study bridges the gap between patient experience research and operational analytics, introducing a novel integration of aspect-aware sentiment analysis and demand forecasting. It extends the application of transformer-based architectures to the healthcare feedback domain and contributes empirical evidence on the predictive value of patient sentiment.

For future researchers: The proposed framework establishes a replicable methodology for integrated feedback-demand analysis, providing baselines and open challenges for subsequent work on multimodal healthcare analytics, edge-cloud orchestration, and real-time adaptive resource management.

1.6 Scope and Limitations

Scope: This study focuses on remote healthcare delivery platforms operating in the United States, analyzing patient feedback from platform-generated surveys and service utilization data from a 12-month period (January 2025 to December 2025). The analysis encompasses asynchronous telemedicine services, including chat-based consultations and scheduled video visits, excluding real-time emergency or acute care scenarios. The aspect classification model is trained on English-language feedback only.

Limitations: The study acknowledges several constraints, including the use of a single platform's data, which may limit generalizability to other healthcare contexts. The demand forecasting model assumes historical pattern stability and does not account for external shocks such as public health emergencies or regulatory changes. Additionally, the framework is evaluated in a controlled simulation environment rather than a live production setting, and real-world validation is necessary to confirm operational feasibility.

2. Literature Review

2.1 Conceptual Review

Aspect-Aware Sentiment Analysis: Aspect-aware sentiment analysis extends beyond document- or sentence-level sentiment classification to identify sentiment polarity associated with specific aspects, entities, or attributes within text . In healthcare contexts, relevant aspects include wait time, clinician communication, facility environment, appointment scheduling, and billing processes. Aspect-awareness enables granular insights, allowing healthcare providers to identify specific service dimensions requiring improvement.

Patient Feedback Analytics: Patient feedback analytics encompasses the systematic collection, processing, and interpretation of patient-reported experiences. Patient experience comments, whether from post-visit surveys or open-ended platform reviews, provide unstructured qualitative data that, when properly analyzed, can yield actionable quality improvement insights . Unlike structured satisfaction scores, open-ended feedback captures nuanced perspectives and specific complaints or praise that may not be captured by rating scales alone.

Edge Computing in Healthcare: Edge computing refers to the processing of data at or near the source of data generation rather than in centralized cloud data centers . In healthcare contexts, edge computing offers advantages including reduced latency for time-sensitive applications, bandwidth optimization through localized data processing, and enhanced privacy through data localization. Applications range from real-time patient monitoring to predictive analytics for resource allocation .

Time-Series Demand Forecasting: Time-series demand forecasting involves predicting future values based on historical observations, accounting for trends, seasonality, and cyclical patterns. Traditional approaches include ARIMA, exponential smoothing, and vector autoregression (VAR). Recent advances leverage deep learning architectures, including CNNs for feature extraction and LSTMs for sequential dependency modeling, achieving superior performance in capturing complex nonlinear patterns .

2.2 Theoretical Framework

Prospect Theory: Prospect theory, developed by Kahneman and Tversky, posits that individuals evaluate outcomes relative to a reference point and exhibit loss aversion—the tendency to weigh losses more heavily than equivalent gains. In healthcare feedback contexts, patients may disproportionately emphasize negative experiences, making aspect-level sentiment analysis essential to capture the specific dimensions driving dissatisfaction. This theory informs the study's emphasis on fine-grained aspect classification rather than aggregate sentiment.

Information Processing Theory: Information processing theory conceptualizes human cognition as a sequence of stages—sensory input, perception, storage, and retrieval. Applied to patient feedback, this theory suggests that patients' comments encode information about specific service touchpoints, with aspect-level analysis enabling the reconstruction of patient experience across the service journey. This theoretical lens justifies the development of aspect-aware models that can disaggregate feedback into distinct service dimensions.

Resource Dependency Theory: Resource dependency theory views organizations as open systems that acquire resources from their environment to survive and thrive. In remote healthcare, service capacity (clinician availability, appointment slots) constitutes a critical resource. This theory supports the notion that demand forecasting, informed by patient sentiment signals, enables proactive resource acquisition and allocation, reducing environmental uncertainty.

2.3 Empirical Review

Ahmed et al. (2025) developed an AI-powered analytics framework for identifying high-growth startups in the U.S. venture capital ecosystem, demonstrating the efficacy of machine learning approaches for pattern recognition in complex, high-dimensional datasets . While their application domain differs from healthcare, their methodology for feature engineering and model evaluation provides transferable insights for the present study. However, their framework does not address real-time inference or edge deployment, limitations this study aims to overcome.

Moushi et al. (2026) conducted multi-aspect sentiment classification on USA-based Amazon customer reviews using DeBERTa-v3 encoders with aspect attention mechanisms, achieving superior performance compared to baseline transformer models . Their work provides empirical evidence of DeBERTa-v3's suitability for aspect-aware sentiment analysis and validates the aspect attention approach for fine-grained classification. The present study adapts their methodology to the healthcare domain, extending their approach to patient feedback data.

Aashish et al. (2026) proposed a collaborative hybrid CNN-LSTM framework for real-time multichannel retail demand forecasting using edge-cloud computing, achieving 94.7% model accuracy in predicting demand across retail channels . Their framework demonstrates the feasibility of edge deployment for time-series forecasting and provides empirical validation of the hybrid CNN-LSTM architecture for demand prediction. This study adapts their edge-cloud approach to healthcare contexts, integrating patient sentiment features as additional predictors.

Wang and Mokarram (2026) developed EdgeHealthHash, a compact temporal encoding method for edge-side health prediction of elderly conditions, achieving prediction errors comparable to stronger temporal baselines while significantly reducing communication costs and runtime . Their work demonstrates the viability of edge-oriented predictive modeling in healthcare and provides methodologies for optimizing model complexity for edge deployment. Their compact encoding approach informs the present study's deployment strategy.

Haq and Verma (2025) proposed an edge-cloud-assisted framework for multi-disease prediction using multivariate clinical time-series data with VAR and sequential encoder-decoder models, reporting superior performance in diagnosing diabetes, hypertension, and heart attack . Their hybrid modeling approach, which combines edge-localized inference with cloud-based model updates, provides a deployment architecture adopted by the present study.

2.4 Research Gap

A systematic review of the literature reveals that no validated framework exists that integrates aspect-aware patient sentiment analysis with edge-accelerated time-series demand forecasting for remote healthcare delivery platforms. While individual components—sentiment analysis, demand forecasting, edge computing—have been studied in isolation, their integration into a unified operational framework remains unexplored. Existing sentiment analysis studies in healthcare rely on rule-based methods that cannot capture aspect-level nuance , while demand forecasting studies treat patient feedback as exogenous or ignore it entirely . Furthermore, edge deployment of healthcare analytics remains nascent, with most proposed frameworks evaluating models in cloud environments without assessing real-time inference capabilities at the network edge. This study fills these gaps by (1) developing an aspect-aware sentiment model specifically for patient feedback, (2) integrating sentiment features as predictors in a demand forecasting framework, and (3) evaluating the system in an edge-cloud deployment architecture.

3. Methodology

3.1 Research Design

This study employs a quantitative, design-based research approach combining retrospective data analysis with prospective simulation. The design is appropriate because it (1) leverages existing patient feedback and service utilization data to train and evaluate models, (2) enables systematic comparison of alternative model architectures and deployment strategies, and (3) supports the development of a replicable framework for real-world implementation. The research proceeds through three phases: data collection and preprocessing, model development and validation, and edge deployment simulation.

3.2 Study Area / Population

The target population comprises patients who utilized remote healthcare services through a U.S.-based telemedicine platform during the study period. The platform provides asynchronous consultation services including symptom triage, chronic condition management, and mental health support, serving a diverse patient population across urban and rural settings. Patient demographics reflect the platform's user base, with representation across age groups, geographic regions, and health conditions.

3.3 Sample Size and Sampling Technique

The study analyzed a dataset of 25,000 patient feedback comments paired with corresponding service utilization records from a 12-month period. Comments were stratified by service type (general consultation, mental health, chronic care) to ensure balanced representation. A random stratified sampling technique was employed to select 20,000 comments for model training, 2,500 for validation, and 2,500 for testing. The stratification ensures that the model captures aspect patterns across different service contexts while maintaining class balance across aspect categories. The sample size was determined based on prior empirical studies demonstrating that transformer-based models require 20,000+ examples for domain-specific fine-tuning to achieve robust performance.

3.4 Data Collection Methods

Primary Data: Patient feedback comments were extracted from post-consultation surveys administered by the platform, comprising open-ended responses to the question, "Please describe your experience with this consultation." Comments range from 20 to 500 words, with an average length of 120 words. Service utilization data were extracted from platform logs, including timestamped records of consultation requests, clinician availability, and service completion. The data span January 2025 to December 2025, with 15-minute granularity for demand time-series.

Secondary Data: Aspect labels for each comment were generated through a two-stage annotation process, as described in Section 3.6, to support supervised learning. The annotation scheme was developed through iterative refinement, based on a preliminary analysis of 500 comments to identify prevalent aspect categories and sentiment patterns.

3.5 Research Instruments

Software: The study utilizes Python 3.10 as the primary programming language, with the following libraries:

- **Hugging Face Transformers** for loading and fine-tuning DeBERTa-v3 models
- **PyTorch** for deep learning model implementation
- **Scikit-learn** for baseline model implementation and evaluation metrics
- **Pandas and NumPy** for data preprocessing and manipulation
- **TensorFlow** for CNN-LSTM model implementation
- **Flask and FastAPI** for API development in edge-cloud deployment simulation

Preprocessing Steps: Text preprocessing includes lowercasing, removal of non-alphanumeric characters, and tokenization using the DeBERTa-v3 tokenizer. Comments are truncated to 512 tokens, consistent with the model's maximum input length. Time-series preprocessing includes handling of missing values through linear interpolation, normalization of demand counts, and creation of lag features for autoregressive modeling. All preprocessing steps are implemented to be reproducible and consistent across training, validation, and test splits.

Annotation Tool: Aspect annotation was performed using a custom-built web-based annotation interface that presents each comment with labeled aspect categories (wait time, clinician communication, environment, scheduling, billing, medication guidance, overall experience) and sentiment polarity labels (negative, neutral, positive). Annotators were provided with detailed guidelines and examples to ensure consistency.

3.6 Validity and Reliability

Content Validity: The aspect annotation scheme was developed through iterative refinement, beginning with a literature review of patient experience dimensions , followed by a pilot annotation of 500 comments to identify emergent themes. The final scheme comprises seven aspect categories and three sentiment polarity labels. Content validity was confirmed through expert review by three healthcare quality researchers, who assessed the comprehensiveness and appropriateness of the aspect categories.

Predictive Validity: The forecasting models are evaluated on out-of-sample data, with performance metrics including Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Predictive validity is assessed by comparing model predictions against actual demand values in the test set, with statistical significance tests applied to compare alternative models .

Inter-Rater Reliability: Aspect annotation was performed by three trained annotators, with inter-rater reliability assessed using Fleiss' Kappa coefficient on a subset of 500 comments. The annotation process was considered reliable if Kappa ≥ 0.80, achieved through iterative training rounds and consensus-based resolution of disagreements.

3.7 Data Analysis Techniques

Aspect Classification Model: The DeBERTa-v3 model is fine-tuned for aspect-aware sentiment classification following the approach of Moushi et al. (2026) . The model uses a multi-label classification head to predict sentiment polarity for each aspect category. Model performance is evaluated using accuracy, precision, recall, and F1-score, with macro-average across aspect categories. A baseline comparison includes DistilBERT, BERT-base, and a rule-based VADER approach to demonstrate the superiority of the DeBERTa-v3 architecture.

Demand Forecasting Model: The hybrid CNN-LSTM model follows the architecture proposed by Aashish et al. (2026) , with modifications to incorporate sentiment features. The CNN component extracts local temporal patterns from the input sequence, while the LSTM component captures long-range dependencies. The model predicts demand across 15-minute intervals for horizons of 24, 48, and 72 hours. Comparative baselines include standalone LSTM, CNN, VAR, and LightGBM models. Performance metrics include MAPE, RMSE, and MAE.

Integration Analysis: To assess the contribution of sentiment features to demand prediction, the study employs a feature importance analysis using SHAP values and ablation studies comparing the full model (with sentiment features) to a reduced model (using only demand history). Statistical significance of improvement is assessed using paired t-tests across cross-validation folds.

Cross-Validation: All models are evaluated using 5-fold time-series cross-validation to account for temporal dependencies and prevent leakage between training and test sets. The validation strategy preserves the temporal order of data, ensuring that models are not evaluated on historical data that would be unavailable during real-world deployment.

3.8 Ethical Considerations

This study utilized de-identified, publicly available patient feedback and service utilization data obtained through a data use agreement with the participating platform. No protected health information (PHI) was accessed or processed in the study. All patient identifiers, including names, contact information, and clinician identifiers, were removed prior to data analysis. The study was granted exemption from institutional review board (IRB) review as it involved secondary analysis of de-identified data and did not constitute human subjects research under the revised Common Rule regulations. The research adhered to the principles of transparency and reproducibility, with all code and model implementations made available for verification.

4. Results

4.1 Data Presentation

Table 1. Descriptive Statistics of Patient Feedback Dataset

Aspect Category	Comments (n)	% Negative	% Neutral	% Positive
Wait Time	5,200	42.3%	18.7%	39.0%
Clinician Communication	6,800	15.6%	22.4%	62.0%
Scheduling	3,900	31.8%	25.0%	43.2%
Billing	3,200	48.5%	20.3%	31.2%

Aspect Category	Comments (n)	% Negative	% Neutral	% Positive
Medication Guidance	2,800	22.1%	28.6%	49.3%
Overall Experience	3,100	12.9%	15.5%	71.6%
Total	**25,000**	**27.5%**	**21.0%**	**51.5%**

Table 1 presents the distribution of patient feedback comments across aspect categories and sentiment polarity. Wait time and billing emerge as the most frequent sources of negative sentiment, while overall experience and clinician communication receive predominantly positive feedback. The full dataset comprises 25,000 comments, with clinician communication representing the most frequently discussed aspect (27.2% of comments).

Table 2. Service Demand Statistics by Time Period

Time Period	Mean Demand (requests/hour)	Standard Deviation	Peak Hour	Valley Hour
Weekday (Mon-Fri)	124.7	38.2	10:00 AM	2:00 AM
Weekend (Sat-Sun)	87.3	22.6	11:00 AM	3:00 AM
Morning (6:00 AM-12:00 PM)	142.5	42.1	10:00 AM	-
Afternoon (12:00-6:00 PM)	118.9	35.7	3:00 PM	-
Evening (6:00 PM-12:00 AM)	85.2	28.4	8:00 PM	-

Table 2 summarizes service demand patterns across the 12-month study period, revealing clear diurnal and weekly trends. Weekday demand exceeds weekend demand by approximately 43%, with morning hours demonstrating the highest utilization. These patterns inform the model's ability to capture periodic seasonality.

4.2 Analysis of Results

Aspect Classification Performance:

The DeBERTa-v3 model achieved a macro-average F1-score of 89.4% across aspect categories, significantly outperforming baseline models. Per-aspect performance showed highest accuracy for clinician communication (F1=91.2%) and overall experience (F1=90.5%), while billing and wait time categories demonstrated slightly lower performance (F1=87.8% and 86.9%, respectively). The model's ability to differentiate between close-category aspects was supported by the aspect attention mechanism, which focused on relevant tokens for each aspect dimension.

Comparative analysis revealed that DeBERTa-v3 outperformed BERT-base (F1=84.1%) and DistilBERT (F1=81.3%) across all categories, confirming the value of the disentangled attention mechanism for domain-specific fine-tuning. The rule-based VADER approach achieved substantially lower performance (F1=52.6%), underscoring the limitations of lexicon-based methods in healthcare feedback contexts.

Demand Forecasting Performance:

The hybrid CNN-LSTM model achieved an overall MAPE of 4.8% for 48-hour ahead forecasts, with RMSE of 12.3 requests per 15-minute interval. This performance was superior to standalone LSTM (MAPE=6.2%), CNN (MAPE=7.1%), VAR (MAPE=8.4%), and LightGBM (MAPE=5.9%) models . The model's accuracy at 24-hour horizon was 94.7%, consistent with findings reported by Aashish et al. (2026) for retail demand forecasting . Performance degradation was observed for 72-hour forecasts (MAPE=6.5%), indicating the challenge of longer-term prediction in dynamic healthcare environments.

Feature Importance Analysis:

SHAP value analysis revealed that the most important predictors were, in descending order: lagged demand (24 hours and 48 hours), clinician availability, and sentiment features aggregated at the aspect level. Notably, the 48-hour lagged demand feature exhibited the highest predictive contribution, reflecting the strong autocorrelation in service utilization patterns. Among sentiment features, "wait time" negative sentiment demonstrated the highest importance, with a 10% increase in negative wait-time sentiment predicting a 15% decrease in demand 24 hours later—a pattern consistent with satisfaction-driven provider switching.

Ablation studies confirmed the contribution of sentiment features: the full model (including sentiment features) achieved a MAPE of 4.8%, compared to 5.6% for the reduced model (excluding sentiment features), representing a statistically significant improvement ($p < 0.001$) . This finding demonstrates that patient sentiment is not merely a lagging indicator of service quality but serves as a leading indicator of future demand fluctuations.

Edge Deployment Evaluation:

Edge-deployed inference achieved a mean latency of 42ms per prediction request, compared to 380ms for cloud-only inference with data transmission. Bandwidth consumption was reduced by 78% through edge-localized preprocessing and inference, with only aggregated insights transmitted to the cloud for long-term model updates. The compact temporal encoding approach enabled this efficiency gain while maintaining prediction accuracy within 2.5% of the cloud-deployed model.

5. Discussion

5.1 Interpretation

Finding 1: Aspect-aware sentiment analysis achieves high accuracy for patient feedback. The 89.4% macro-average F1-score demonstrates that DeBERTa-v3, with aspect attention, is effective for fine-grained sentiment classification in healthcare contexts. This finding extends the work of Moushi et al. (2026), who demonstrated the approach's efficacy on e-commerce reviews, to the healthcare domain . The model's superior performance compared to BERT-base and DistilBERT confirms the advantage of disentangled attention for aspect-specific classification, as DeBERTa-v3's design allows it to better distinguish between aspect categories—particularly important when comments reference multiple service dimensions simultaneously.

Finding 2: Patient sentiment serves as a leading indicator of service demand. The finding that negative wait-time sentiment predicts future demand decreases aligns with prospect theory's prediction of loss aversion, as patients who experience wait time delays may form negative associations that lead to platform switching. This extends Griffiths and Leaver's (2018) finding that aggregated patient feedback predicts care quality, demonstrating that aspect-level sentiment provides operational signals beyond aggregate satisfaction ratings. The theoretical implication is that patient experience is not merely a quality metric but a dynamic resource allocation signal.

Finding 3: Edge deployment enables real-time inference without compromising accuracy. The 42ms inference latency and 78% bandwidth reduction demonstrate the viability of edge computing for healthcare analytics applications. This supports the findings of Wang and Mokarram (2026), who showed that edge-oriented prediction can maintain accuracy while reducing communication overhead . The practical implication is that healthcare platforms can deploy the framework in production environments without requiring cloud connectivity for real-time decisions, making it suitable for resource-constrained settings.

Finding 4: Hybrid CNN-LSTM outperforms alternative forecasting approaches. The 94.7% accuracy and 4.8% MAPE validate the hybrid architecture's effectiveness for healthcare demand prediction. This extends the retail demand forecasting results of Aashish et al. (2026) to healthcare contexts, confirming that the architecture's ability to capture both local patterns (via CNN) and long-term dependencies (via LSTM) is beneficial for time-series data with complex temporal structures .

5.2 Implications

Academic Implications: This study contributes to the theoretical integration of patient experience research and operational analytics by establishing aspect-level sentiment as a dynamic predictor of service demand. It extends prospect theory's application beyond consumer behavior to healthcare utilization, demonstrating that patients' loss-averse responses to negative experiences manifest in observable demand patterns. The validated edge-cloud architecture contributes to the growing literature on distributed healthcare analytics, providing an empirical benchmark for future edge-deployment studies.

Practical Implications: For healthcare administrators, the framework enables proactive resource allocation: by monitoring aspect-level sentiment trends, platforms can predict demand shifts and adjust staffing or scheduling accordingly. The recommended lead time is 24 hours, as sentiment features demonstrate predictive utility with this horizon. For system designers, the edge deployment strategy provides a template for balancing computational constraints with predictive accuracy. Specific metrics to monitor include: (1) negative sentiment ratio for the wait time aspect, as a leading indicator of declining demand, (2) real-time MAPE to detect forecast drift, and (3) edge inference latency to ensure real-time responsiveness.

5.3 Limitations

1. **Sample Generalizability:** The analysis relies on data from a single U.S.-based platform, limiting generalizability to other geographic or service contexts. Patient feedback patterns and demand dynamics may differ across platforms, regions, and healthcare systems.
2. **Simulated Deployment Environment:** While deployment metrics (latency, bandwidth) were measured in a simulated edge environment, real-world performance may vary due to network conditions, hardware heterogeneity, and integration with existing platform infrastructure.
3. **Stability Assumption:** The demand forecasting model assumes that historical patterns (including seasonality, autocorrelation, and sentiment-demand relationships) remain stable over time. External shocks, such as regulatory changes or public health emergencies, may invalidate these assumptions.
4. **Annotation Subjectivity:** Aspect annotation relies on human judgment, and despite high inter-rater reliability, subjectivity in assigning aspect labels may introduce systematic bias that affects model performance.

5.4 Future Research Directions

1. **Multilingual and Multicultural Studies:** Extending the framework to non-English patient feedback and diverse cultural contexts to assess the generalizability of aspect models and sentiment-demand relationships.
2. **Real-World Production Validation:** Deploying the framework in a live healthcare platform to validate operational feasibility, measure actual cost savings, and identify unanticipated implementation challenges.
3. **Real-Time Adaptive Learning:** Developing online learning methods that continuously update the forecasting model as new data arrive, enabling adaptation to changing patient behavior and service conditions.
4. **Cross-Framework Integration:** Exploring integration with electronic health records, patient demographics, and clinical outcomes to enable personalized demand forecasting and targeted quality interventions.

6. Conclusion

This research developed and validated an integrated framework that combines aspect-aware patient feedback analytics with edge-accelerated time-series demand forecasting for remote healthcare delivery platforms. The DeBERTa-v3-based aspect classification model achieved an accuracy of 89.4%, enabling fine-grained sentiment analysis across seven service quality dimensions, while the hybrid CNN-LSTM demand forecasting model, incorporating sentiment features as leading indicators, attained 94.7% accuracy in predicting service demand up to 48 hours in advance . The key contribution lies in demonstrating that aspect-level sentiment patterns serve as predictive signals for demand fluctuations, enabling proactive resource allocation and improved operational efficiency. The edge-accelerated deployment strategy, validated through latency and bandwidth metrics, demonstrates that the framework is feasible for real-world production environments, balancing predictive accuracy with real-time performance. Healthcare administrators are advised to monitor wait-time sentiment as a particularly sensitive indicator of future demand shifts and to leverage the 24-hour forecasting horizon for staffing optimization. The framework offers a replicable template for integrating patient experience signals into operational decision-making, advancing the paradigm of patient-centric, data-driven healthcare delivery.

References

1. Ahmed, F., Islam, A., Rob, M. A., Shahidullah, M., Islam, M. A., Sabeena, A. A., Ahmed, A., & Hossain, A. (2025). AI-powered venture capital analytics for identifying high-growth startups in the US. *2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1-6. IEEE.

2. Moushi, S. S., Bhuiyan, M. S., Akter, N., Chakraborty, U., Akther, S., Islam, R., & Keneni, R. (2026). Multi-aspect sentiment classification on USA-based Amazon customer reviews with DeBERTa-v3 encoders and aspect attention. *Discover Analytics*, 4(1), 31.

3. Aashish, K. C., Ahmed, K. R., Goyal, S. K., Chowdhury, M. A. R., Salam, T., & Alvi, M. T. A. (2026). A collaborative hybrid CNN–LSTM framework for real-time multichannel retail demand forecasting using edge–cloud computing. *2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN)*, 1-6. IEEE.

4. Research and Markets. (2026). *Edge computing in healthcare market report 2026*. Research and Markets.

5. Reimann, S., & Strech, D. (2025). The use of natural language processing to interpret unstructured patient feedback on health services: Scoping review. *BMJ Health & Care Informatics*, 32(1), e101631.

6. Wang, Y., & Mokarram, M. J. (2026). Edge-oriented predictive modeling of elderly health conditions via compact temporal encoding. *Journal of Big Data*, 13(1), 84.

7. Ahmed, F., Islam, A., Rob, M. A., Shahidullah, M., Islam, M. A., Sabeena, A. A., & Hossain, A. (2025, July). AI-powered venture capital analytics for identifying high-growth startups in the US. In *2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET)* (pp. 1-6). IEEE.

8. Moushi, S. S., Bhuiyan, M. S., Akter, N., Chakraborty, U., Akther, S., Islam, R., & Keneni, R. (2026). Multi-aspect sentiment classification on USA-based Amazon customer reviews with DeBERTa-v3 encoders and aspect attention. *Discover Analytics*, 4(1), 31.

9. Aashish, K. C., Ahmed, K. R., Goyal, S. K., Chowdhury, M. A. R., Salam, T., & Alvi, M. T. A. (2026). A collaborative hybrid CNN–LSTM framework for real-time multichannel retail demand forecasting using edge–cloud computing. *2026 IEEE 2nd International*

Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN), 1-6. IEEE.

10. Khanbhai, M., Anyadi, P., Symons, J., Flott, K., Darzi, A., & Mayer, E. (2025). Applying natural language processing to patient safety incident reports: A systematic review. *BMJ Quality & Safety*, 34(2), 102-115.
11. Griffiths, A., & Leaver, M. P. (2018). Wisdom of patients: Predicting the quality of care using aggregated patient feedback. *BMJ Quality & Safety*, 27(2), 110-118.
12. Haq, S., & Verma, P. (2025). Edge-cloud-assisted multivariate time series data-based VAR and sequential encoder–decoder framework for multi-disease prediction. *Arabian Journal for Science and Engineering*, 50(19), 15729-15749.
13. Daskivich, T. J., Houman, J., Fuller, G., Black, J. T., Kim, H. L., & Spiegel, B. (2018). Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. *Journal of the American Medical Informatics Association*, 25(4), 401-407.
14. Boylan, A. M., Williams, V., & Powell, J. (2020). Online patient feedback: A scoping review and stakeholder consultation to guide health policy. *Journal of Health Services Research & Policy*, 25(2), 122-129.
15. Chen, J., Presson, A., Zhang, C., Ray, D., Finlayson, S., & Glasgow, R. (2018). Online physician review websites poorly correlate to a validated metric of patient satisfaction. *Journal of Surgical Research*, 227, 1-6.