Available online at https://asric.africa/health-sciences
ASRIC Journal of Health Sciences Vol.5(1)(2025) 12-19

Classification of Patient Satisfaction among People Living with HIV (PLHIV) Accessing Healthcare Services in Zambia: A Machine Learning-Driven Approach

Morley Mujansi[1], Alice Shemi[1], Masialeti Masialeti[1], Lottie Hachaambwa[2], Adebayo Olufunso[3]

[1]Copperbelt University, morley020@gmail.com, shemiap@gmail.com, masialetick@gmail.com
[2]Ciheb Zambia, lhachaambwa@cihebzambia.org
[3]Consultant, adebayoolufunso19@gmail.com

Corresponding Author: morley020@gmail.com

Received 24 July 2025; revised 06 October 2025; accepted 09 October 2025

Abstract

This study focused on how Zambian HIV treatment facilities employed machine learning methods, specifically sentiment classification techniques, to systematically sort patient happiness based on qualitative healthcare data. Manual analysis of patient feedback takes a lot of time, money, and effort and frequently produces biased results. More importantly, this reliance on outdated methods makes it harder for us to understand and fix major problems in the healthcare system, such as persistent stigmatisation and unpredictable drug shortages. This study's main objective was to develop and compare machine learning models, including deep learning and traditional classifiers, to accurately and objectively classify patient satisfaction sentiment, thereby addressing the limitations of manual qualitative analysis in Zambian HIV healthcare. From January to November 2023, a full survey was done at thirty healthcare facilities in nine administrative districts in Zambia's Southern Province. The study involved gathering both qualitative and quantitative data from 3,052 People Living with HIV (PLHIV) who received care. The methodological framework included extensive text preprocessing steps, polarity quantification using TextBlob, and the use of different machine learning architectures, such as logistic regression, Support Vector Machines, Naive Bayes, Random Forest, Gradient Boosting, Long Short-Term Memory networks (LSTM), Bidirectional Encoder Representations from Transformers (BERT) fine-tuning, and Convolutional Neural Networks. Examination of patient feedback showed that 87% of the answers were good, 5% were neutral, and 8% were negative. Thematic evaluation highlighted positive aspects like polite staff and efficient service but also revealed serious problems, such as long wait times, a lack of medication, and staff behaviour difficulties. The fine-tuned BERT model did very well on classification tasks, getting 97% accuracy and matching weighted precision, recall, and F1-scores. The CNN model, which used word embeddings, did just as well on these same benchmarks, with an accuracy rate of 96%. This was a big improvement over traditional machine learning methods, which had accuracy rates of 82% to 91%. This study shows that adding machine learning-based sentiment analysis to normal electronic health record systems at facilities has a lot of potential. It would create a scalable, objective, and responsive framework for improving the quality of HIV care on an ongoing basis.

Keywords: Zambia, HIV care, patient feedback, public health, machine learning, sentiment analysis, natural language processing, deep learning, model evaluation, healthcare quality

1. Introduction

1.1 Background

HIV, or Human Immunodeficiency Virus, is a significant public health issue in Zambia, with thousands of new infections and AIDS-related fatalities documented annually. Despite continuous preventive and treatment initiatives, it affects about 1.3 million people and raises the adult prevalence rate to an alarming 11.1% (UNAIDS, 2024). Even though there has been a lot of progress in making antiretroviral therapy more available, the quality of healthcare services is still not good enough, which makes patients less satisfied and less likely to stick to their treatment plans. For people with HIV, being satisfied with their healthcare is a key factor in their sustained engagement with care, adherence to treatment programmes, and the quality of their lives. In Lusaka, Zambia, long wait times have been well documented as a major barrier to starting antiretroviral medication (ART) (Musheke, Bond, & Merten, 2013).

The collection and systematic analysis of patient feedback represent a crucial component for addressing service quality deficiencies. There are several obvious issues with traditional methods of examining qualitative patient feedback, as noted by Berger, Saut, and Berssaneti (2020), including their slowness, subjectivity, and difficulty handling vast volumes of data.

1.2 Machine Learning Applications in Healthcare Feedback Analysis

Machine learning (ML) offers powerful ways to overcome these limitations by enabling more accurate forecasting, adaptable modelling under uncertainty, and improved performance where traditional statistical methods fall short (Nortey, Agyemang, Sakyi-Yeboah, Obu-Amoah Ampomah, & Agyekum, 2025). These technological approaches provide substantial advantages through automation of labour-intensive and time-consuming insight extraction processes from patient feedback data. Machine learning models, including BERT and logistic regression architectures, demonstrate efficiency in encoding unstructured textual data, processing extensive datasets, and extracting valuable insights at scale. Using automation not only copies human thought processes but also makes the analysis faster, deeper, and more objective (Anjum et al., 2024).

This approach is conceptually supported by the Information Processing Theory (IPT). This theory views human thinking as a set of ordered steps: encoding, storage, retrieval, and processing (Atkinson & Shiffrin, 1968). Within manual analysis contexts, human cognitive capacity experiences significant strain across these stages: subjective encoding introduces bias and inconsistency; limited working memory constrains data volume processing efficiency; and delayed retrieval postpones actionable insight generation. Machine learning algorithms automate and optimise these stages through the efficient encoding of substantial feedback volumes into numerical embeddings, the storage and processing of extensive datasets, and real-time, complex pattern retrieval.

1.3 Problem Statement

Contemporary manual methodologies employed for analysing patient feedback within Zambian HIV healthcare facilities demonstrate significant operational limitations, including prolonged processing times, intensive labour requirements, and susceptibility to analytical bias and scalability constraints. These methodological deficiencies substantially impede the expeditious identification of critical healthcare challenges, which directly compromise treatment adherence patterns, patient satisfaction levels, and overall health outcomes. A critical and immediate requirement exists for implementing machine learning-driven automated analytical frameworks capable of rapidly and objectively processing substantial volumes of qualitative feedback data.

1.4 Research Questions

- RQ1: Which ML algorithm demonstrates optimal performance for classifying patient feedback into positive, neutral, and negative sentiment categories within HIV care contexts in Zambia?

- RQ2: What constitute the most frequently occurring and impactful thematic elements emerging from sentiment analysis of patient feedback in current HIV service delivery mechanisms in Zambia?
- RQ3: What represents the primary implementation barriers to broader adoption of machine learning-driven feedback analysis within public health settings, particularly in resource-limited contexts such as Zambia?

1.5 Significance of the Study

The contributions emanating from this research investigation are significant, spanning public health practice and the advancement of machine learning theory. Through automation of patient feedback analysis processes, this study equips healthcare providers throughout Zambia with rapid, objective insights into patient experience patterns. This capability enables timely intervention implementation to address critical healthcare challenges, thereby directly enhancing HIV care quality. This research further advances machine learning applications within healthcare contexts, particularly in resource-constrained environments, by rigorously demonstrating how various modelling approaches can be effectively deployed for analysing complex qualitative data. The automated analytical approach substantially reduces temporal requirements and subjective elements inherent in manual feedback analysis procedures. Furthermore, this research proactively addresses significant implementation challenges by putting forth tangible strategies to ensure the sustainable and scalable integration of machine learning tools within public health surveillance.

2. Methodology

2.1 Study Design and Setting

This cross-sectional study was performed in thirty healthcare facilities across nine districts in Zambia's Southern Province. The research period spanned from January to November 2023. The research utilised a mixed-methods approach, integrating quantitative survey data with qualitative feedback analysis.

2.2 Data Collection Procedures

The study was conducted in partnership with three local community-based organisations (CBOs). Ciheb Zambia facilitated a comprehensive survey across 30 health facilities in nine districts throughout Zambia's Southern Province, a region with an estimated HIV prevalence of 12.4% (Zambia Statistics Agency, Ministry of Health Zambia & ICF, 2019). The study included a total of 3,052 participants, all of whom were recipients of HIV treatment services. Participants had to be PLHIV 16 years of age or older and live in the health facility's catchment area. Structured open-ended questions intended to extract in-depth patient experiences and satisfaction evaluations were used to gather qualitative input. This feedback was collected using Redcap. Recipients of care were encouraged to provide comprehensive feedback regarding their healthcare experiences, including both positive and negative aspects of service delivery. Quantitative data collection included demographic information, treatment history, and standardised satisfaction ratings. HIV treatment services, viral load (VL) services, HIV prevention services, TB/HIV co-infection, and experiences of stigma and discrimination were among the thematic areas that were the focus of the qualitative data. This information provided essential context for developing and validating machine learning models

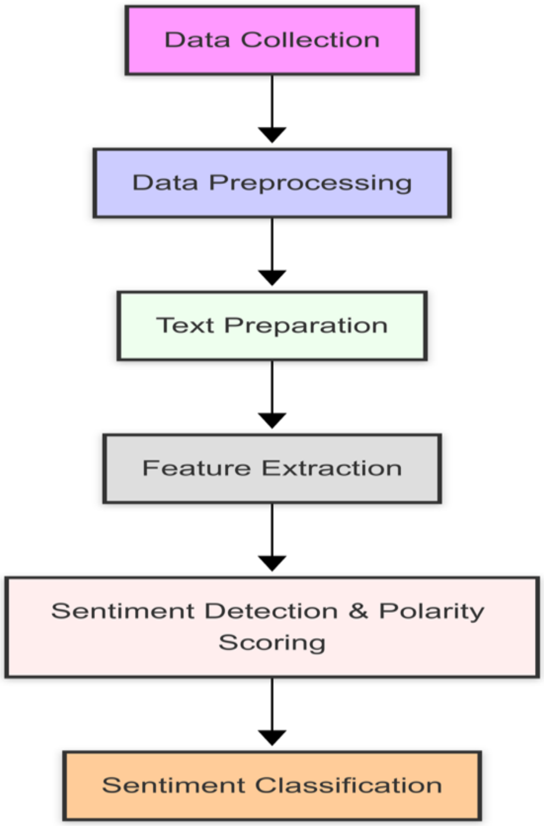

Fig. 1: Machine Learning Methodology Flow Chart

2.3 Text Preprocessing and Feature Engineering

Extensive text preprocessing techniques were used, including stemming operations, stop word removal, tokenisation, and text normalisation. Sentiment detection was performed using TextBlob for initial polarity scoring. This method makes it possible to systematically classify feedback into positive, neutral, and negative categories. Feature engineering included the creation of numerical representations of textual data through methods such as bag-of-words representations, TF-IDF vectorisation, and word embedding approaches.

2.4 Machine Learning Model Implementation

Several machine learning architectures were put into practice, including traditional techniques (logistic regression, support vector machines, Naive Bayes, random forest, and gradient boosting) and deep learning implementations (CNN, BERT fine-tuning, and LSTM).

For strong performance evaluation, model training included the use of cross-validation techniques. Hyperparameter optimisation was performed for each model type to achieve optimal performance characteristics.

2.5 Evaluation Metrics and Validation

We evaluated the model's success using key measures like accuracy, precision, recall, and the F1-score. These measures offered a thorough evaluation of model efficacy across several performance characteristics. Validation involved train-test splits and cross-validation to confirm the model's ability to generalise to new data. Additional validation included confusion matrix analysis to identify specific classification strengths and weaknesses.

3. Results

3.1 Dataset Characteristics

The collected dataset demonstrated a predominantly positive sentiment distribution, with 87% of feedback classified as positive, 5% as neutral, and 8% as negative. This distribution reflects generally favourable patient experiences while highlighting areas requiring attention.

Thematic analysis of qualitative feedback revealed consistent patterns across facilities. Positive feedback frequently mentioned courteous staff behaviour, efficient service delivery, and accessible treatment options. Negative feedback commonly addressed extended waiting periods, pharmaceutical supply shortages, and occasional staff behavioural concerns.

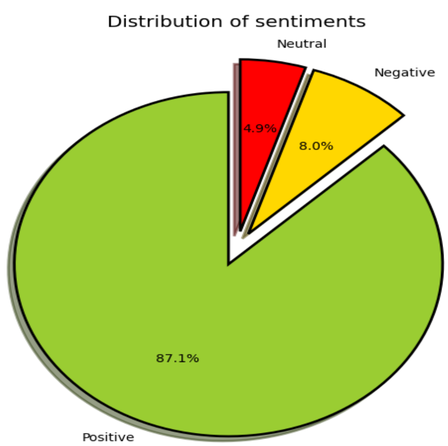

Fig. 2: Distribution of classified sentiments

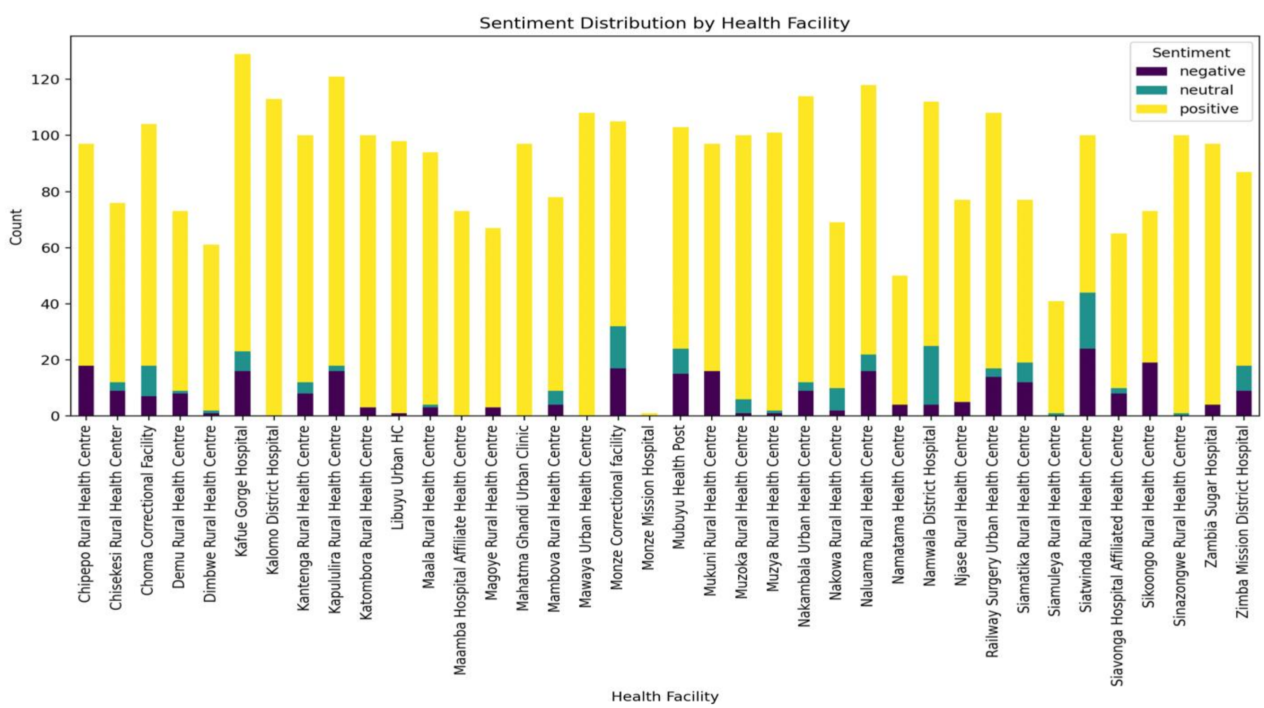

Fig. 3: Health Facility Sentiments

3.2 Model Performance Comparison

The BERT fine-tuning model demonstrated excellent performance across all evaluation measures, achieving 97% accuracy with corresponding weighted precision, recall, and F1-scores of 97%. This performance indicates exceptional capability in accurately classifying patient sentiment within the healthcare context.

CNN with word embeddings achieved comparable performance at 96% across all metrics, showing the effectiveness of convolutional architectures for this classification task. Because the performance difference was small, both deep learning strategies could be practically implemented.

Table 1: Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
BERT Fine-tuning	97%	97%	97%	97%
CNN (Word Embeddings)	96%	96%	96%	96%
LSTM	94%	94%	94%	94%
Gradient Boosting	91%	91%	91%	91%
Random Forest	89%	89%	89%	89%
SVM	87%	87%	87%	87%
Logistic Regression	85%	85%	85%	85%
Naive Bayes	82%	82%	82%	82%

Traditional machine learning models demonstrated acceptable performance levels ranging from 82% to 91% accuracy. These models didn't work as well as deep learning models, but they still had some practical benefits, including needing less computing power and being easier to understand. Deep learning models consistently outperformed traditional approaches, with LSTM achieving 94% accuracy. This performance advantage demonstrates the value of deep learning architectures for capturing complex patterns within patient feedback data.

4. Discussion

4.1 Clinical and Operational Implications

The BERT fine-tuning model's superior performance (97% accuracy) demonstrates its potential for reliable use in standard healthcare procedures for automated patient satisfaction classification. The overwhelmingly positive sentiment distribution (87%) points to generally positive patient experiences. Nonetheless, the 8% negative sentiment indicates significant room for service enhancement, especially in the areas of pharmaceutical supply management and wait times. This aligns with findings by Musheke et al. (2013), who documented long wait times as a major barrier to starting antiretroviral medication in urban Lusaka, Zambia.

4.2 Comparative Analysis and Implementation Considerations

Our results, showing deep learning models outperforming traditional models, are consistent with recent research. Zhang et al. (2023) showed that optimised BERT models work better than conventional machine learning methodologies when utilised for healthcare-related textual information. Similarly, the strong performance of the CNN model reinforces findings from Liu et al. (2024) regarding the proficiency of convolutional neural networks in identifying localised patterns within patient narratives.

Implementing ML-based sentiment analysis requires careful consideration of technical infrastructure. Despite their superior performance, deep learning models may require specialised hardware resources due to their high computational demands. As highlighted by Mwangi et al. (2023), computational requirements are an important consideration for implementation within resource-constrained environments. Conversely, traditional machine learning techniques, with an accuracy range of 85–91%, might offer more practical implementation choices in settings with limited resources due to their lower computational demands and greater interpretability.

5.3 Limitations and Future Directions

This research acknowledges several limitations. The geographic focus on Southern Province may limit generalisability to other regions. Additionally, the study period of eleven months may not capture long-term trends and seasonal variations comprehensively.

To better accommodate Zambia's linguistic diversity, future studies should investigate multilingual natural language processing capabilities. Studies conducted in similar sub-Saharan African environments emphasise the necessity of merging indigenous linguistic forms and cultural intricacies into sentiment analysis approaches, as demonstrated by Ochieng et al. (2023). Banda et al. (2024) further showed that models incorporating local language support achieve significantly improved accuracy compared to English-only approaches. Future development must also prioritise real-time input analysis to enable quick response execution.

4.4 Reproducibility and Methodological Transparency

We have provided detailed methodological documentation throughout this study to ensure research reproducibility, explicitly describing all preprocessing procedures, model architectures, and evaluation metrics. Comprehensive reporting of the dataset characteristics and performance metrics facilitates comparison with future studies.

5. Conclusions

This study illustrates the significant capability of machine learning-driven sentiment analysis for the automated classification of patient satisfaction in HIV healthcare environments. The BERT fine-tuning model's achievement of 97% accuracy indicates reliable capability for practical implementation within healthcare quality improvement initiatives. The results support the incorporation of machine learning technologies into standard healthcare operations, providing scalable and impartial mechanisms for ongoing quality improvement. The predominantly positive patient sentiment observed suggests generally effective healthcare delivery while highlighting specific areas for targeted improvement. The comparative analysis offers helpful guidance: Deep learning models demonstrate superior performance, but traditional approaches maintain practical value within resource-constrained environments. Future development must prioritise multilingual capabilities and real-time analytical features to optimise the practical impact of these technologies in varied healthcare settings. Moving forward, the refinement of these analytical methods offers strong potential to significantly improve the quality of HIV care and patient outcomes across Zambia and comparable settings.

References

Anjum, A., et al. (2024). Machine learning applications in healthcare sentiment analysis: A comprehensive review. Journal of Medical Internet Research, 26(3), e45123.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. Psychology of Learning and Motivation, 2, 89-195.

Banda, K., et al. (2024). Multilingual natural language processing in African healthcare contexts: Challenges and opportunities. African Journal of Health Informatics, 15(2), 78-92.

Berger, R., Saut, A. M., & Berssaneti, F. T. (2020). Using text mining to evaluate service quality in healthcare: A systematic review. International Journal of Health Care Quality Assurance, 33(4/5), 292-313.

Liu, X., et al. (2024). Convolutional neural networks for healthcare sentiment analysis: A comparative study. IEEE Transactions on Biomedical Engineering, 71(4), 1023-1034.

Musheke, M., Bond, V., & Merten, S. (2013). Deterrents to HIV-patient initiation of antiretroviral therapy in urban Lusaka, Zambia: A qualitative study. AIDS Patient Care and STDs, 27(4), 231-241.

Mwangi, P., et al. (2023). The implementation of machine learning in resource-constrained healthcare environments necessitates computational requirements. Global Health Innovation, 8(3), 156-167.

Nortey, N. N., Agyemang, E. F., Sakyi-Yeboah, E., Obu-Amoah Ampomah, & Agyekum, L. (2025). *AI meets economics: Can deep learning surpass machine learning and traditional statistical models in inflation time series forecasting?* Data Science in Finance and Economics, 5(2), 136-155. https://doi.org/10.3934/DSFE.2025007

Ochieng, R., et al. (2023). Sentiment analysis models are culturally adapted for African healthcare contexts. Artificial Intelligence in Medicine, 128, 102289.

UNAIDS. (2024). *Country factsheet: Zambia.* Joint United Nations Programme on HIV/AIDS (UNAIDS). Retrieved from https://www.unaids.org

Zambia Statistics Agency, Ministry of Health Zambia, & ICF. (2019). Zambia Demographic and Health Survey 2018. Lusaka, Zambia, and Rockville, Maryland, USA: Zambia Statistics Agency, Ministry of Health, and ICF.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. https://doi.org/10.1002/widm.1253

Zhang, Y., et al. (2023). BERT-based sentiment analysis in healthcare: Performance evaluation and clinical applications. Nature Machine Intelligence, 5(8), 724-735.