Deep learning for sentiment analysis and topic extraction in health insurance

Muzondiwa Karomo, Mainford Mutandavari, Wilton Muzava
School of Information Science and Technology, Harare Institute of Technology, Harare, Zimbabwe

Article Info

Article history:

Received Jun 8, 2025
Revised Jul 12, 2025
Accepted Jul 17, 2025

Keywords:

Deep learning
Health care insurance
Sentiment analysis
Social media analytics
Topic modeling

ABSTRACT

Social media has transformed into a vital channel for real-time, unsolicited feedback in healthcare, yet health insurance providers often lack the tools to mine insights from such data. This study proposes a cloud-based system leveraging deep learning for sentiment analysis and topic modeling tailored to the Commercial and Industrial Medical Aid Society (CIMAS) health insurance in Zimbabwe. Using bidirectional encoder representations from transformers (BERT), a convolutional neural network (CNN), a random forest (RF), and autoencoders, the system processes multilingual data from platforms like Twitter and Facebook, identifying customer concerns in real time. Over 15,000 posts were analyzed, with CNN achieving 91.4% accuracy in sentiment classification and BERTopic extracting coherent themes. The system detected issues such as claim delays, app navigation problems, and unreported anomalies. Findings demonstrate that AI can improve service delivery, customer satisfaction, and responsiveness in African insurance contexts.

Corresponding Author:

Muzondiwa Karomo
School of Information Science and Technology, Harare Institute of Technology
Harare, Zimbabwe
Email: mkaromo@gmail.com

1. INTRODUCTION

Healthcare providers today face an overwhelming flood of unstructured social media feedback, making it difficult to identify actionable insights, especially in developing countries where digital infrastructure and analytical tools remain limited. To address this, sentiment analysis, a subfield of natural language processing (NLP), is employed to systematically interpret and classify emotions within textual content. It assigns polarity, positive, negative, or neutral, to each opinion. While traditionally applied in sectors such as e-commerce and entertainment, sentiment analysis is now gaining momentum in healthcare, where understanding public feedback is critical for enhancing service delivery.

To address this challenge, we propose an automated sentiment analysis pipeline built using NLP and machine learning techniques, tailored for healthcare feedback in Zimbabwe. Collected through customer review websites and social media online posts. Using supervised machine learning algorithms, i.e., the support vector machine (SVM) and naive Bayes (NB) classifier, it is expected that the system will accurately categorize sentiment. The sentiment analysis pipeline involves data preprocessing (e.g., stop word elimination, tokenization, and lemmatization), term frequency-inverse document frequency (TF-IDF) based feature extraction, and classification based on learned data.

A comprehensive analysis was conducted to determine the feasibility of using sentiment analysis to uncover the overall public perception of the Commercial and Industrial Medical Aid Society (CIMAS). The implementation of such a model would allow the organization to proactively respond to customer concerns, measure satisfaction trends over time, and support strategic improvements in communication and service offerings. By adopting this sentiment analysis model, CIMAS can maintain a competitive edge in the healthcare sector by aligning its operations more closely with customer expectations.

In Zimbabwe, medical aid providers such as CIMAS struggle to leverage real-time public sentiment due to limited adoption of advanced analytics tools. This lack of insight can lead to missed service improvement opportunities, customer dissatisfaction, and reactive problem handling. While several studies have shown the success of sentiment analysis across industries like e-commerce and general healthcare, very few have applied these techniques within the context of developing countries, particularly in Zimbabwe. This research addresses that gap by focusing on the CIMAS Medical Aid Society, offering a tailored deep learning approach to extract actionable insights from public sentiment.

2. RELATED WORK

From the related work, it is evident that sentiment analysis has evolved into a widely applicable tool across industries, with growing relevance in healthcare and service quality assessment. The foundational work by Liu [1] in "Sentiment analysis and opinion mining" laid the groundwork for text polarity detection using supervised machine learning techniques such as NB and SVM. This work underscored the importance of textual feature extraction and lexicon-based approaches in improving classification accuracy [2]–[4].

In the paper "Twitter sentiment classification using distant supervision" by Go *et al.* [5], the authors developed a sentiment classifier using weakly labeled Twitter data. Their research demonstrated that even noisy, informal language on social media could yield reliable sentiment predictions using machine learning [6]–[11]. The study "machine learning and sentiment analysis: analyzing customer feedback" by Sharma and Jain [12] focused on processing online reviews to determine customer satisfaction levels in corporate environments. Their results showed that real-time sentiment classification can be a vital input for management decision-making in customer-facing organizations [13]–[16].

Aattouchi *et al.* [17] explored how patient sentiments expressed in reviews and forums could be analyzed to improve hospital and insurance service quality. They emphasized that machine learning-based sentiment models can uncover patient pain points and assist in policy refinement [18]–[22]. Finally, a paper by Sheng *et al.* [23] has highlighted the significance of machine-based sentiment identification in detecting trends in the public mood, especially during health emergencies and policy updates regarding services [24]–[28]. Together, these articles indicate the advanced ability of machine learning to obtain human emotion and opinion across various sources. The literature confirms that supervised learning models such as NB, logistic regression (LR), and SVMs are highly effective for sentiment classification tasks. However, the use of such techniques for healthcare organizations in developing countries like Zimbabwe remains limited, particularly within the context of medical aid societies such as CIMAS [29], [30].

Key contributions of this study is this study will focus on developing and testing a machine learning-based sentiment analysis model using real-world feedback data related to CIMAS Medical Aid Society. The model will be trained to detect sentiment polarity with high precision using both classical and modern NLP techniques. The contributions of this study are:

i) Healthcare-specific sentiment analysis model: this project develops a sentiment classification pipeline tailored specifically to healthcare service feedback, addressing domain-specific linguistic patterns.
ii) Machine learning-based automation: the study incorporates supervised learning algorithms to build a robust, scalable, and automated system that classifies public sentiment in real time.
iii) Strategic value to CIMAS:
   – Enables data-driven service improvements based on customer perception;
   – Helps identify recurring negative themes that may require operational attention;
   – Enhances engagement strategies by recognizing positive sentiment trends.
iv) Contextual relevance: this research fills a gap in existing literature by applying sentiment analysis in the Zimbabwean healthcare context, where digital feedback mechanisms are growing but remain underutilized for insight extraction.

3. METHOD

In this research, a machine learning-based sentiment analysis model is developed to classify customer feedback related to CIMAS Medical Aid Society into positive, negative, or neutral sentiments. The methodology includes data collection, data preprocessing, feature extraction, model selection and training, and model evaluation. The system is designed to process unstructured textual data (e.g., customer reviews and social media comments) and return sentiment labels with high predictive accuracy.

3.1. Planning and data preparation

The research framework was developed by first identifying relevant data sources, selecting appropriate NLP techniques, and defining the supervised machine learning pipeline. Historical feedback data related to CIMAS was collected from publicly available online platforms. These included Twitter, Facebook comments, and customer review sites. The text data was then cleaned, structured, and labeled for training and testing the model.

3.2. Data collection and preprocessing

Raw text data was scraped using APIs and web scraping tools, following ethical guidelines for public data usage. The dataset included comments made about CIMAS over the past two years and was anonymized to protect user privacy. Dataset fields: i) comment_text; ii) timestamp; iii) user_platform; and iv) label (positive, negative, and neutral).

Text data was preprocessed using the following steps: i) removal of special characters, emojis, and URLs; ii) lowercasing of all words; iii) tokenization (splitting text into individual words); iv) stop word removal (e.g., "the", "is", and "at"); and v) lemmatization (converting words to their base form). This preprocessing step ensured that irrelevant noise was removed and that the text was in a consistent format suitable for machine learning models.

3.3. Feature engineering

Feature extraction was performed using TF-IDF, chosen for its proven ability to reflect term importance in high-dimensional text data. Unigrams and bigrams were also used to capture short contextual patterns. Although advanced methods such as word embeddings were considered, TF-IDF offered simplicity, effectiveness, and interpretability for this application. Were considered as features to capture local word relationships. The following is a summary of the engineered features captured in Table 1.

Table 1. Summary of the engineered features

Feature	Relevance
TF-IDF scores	Represent the importance of each word in the context of the document and corpus.
N-grams	Capture common word pairings that indicate sentiment (e.g., "not happy" and "very good").
Word count	Gives a basic measure of comment length, which sometimes correlates with emotion.
Sentiment lexicon score	Used as a secondary validation feature to compare against predicted sentiment.

3.4. Machine learning model development

The sentiment classification problem was approached using supervised learning. A labeled dataset was used to train models that predict whether a given comment is positive, negative, or neutral. Three different models were tested: i) NB classifier; ii) SVM; and iii) LR. Among these, SVM yielded the highest accuracy during validation and was selected as the final model. Python’s scikit-learn library was used for model development and evaluation. The process included model training, cross-validation, and hyperparameter tuning to improve performance.

3.5. Model evaluation

The effectiveness of the model was assessed using standard classification metrics: i) accuracy is the percentage of correct predictions; ii) precision is the proportion of positive predictions that were actually positive; iii) recall is the proportion of actual positive cases that were correctly identified; iv) F1-score is the harmonic mean of precision and recall; and v) confusion matrix is used to visualize model performance across all sentiment categories. The metrics and purpose used are shown in Table 2.

Table 2. Matrices and their purpose are used

Metric	Purpose
Accuracy	Overall prediction performance.
Precision	How many predicted positive sentiments were accurate.
Recall	How well the model captured actual sentiment.
F1-score	Balanced metric for imbalanced classes.

3.6. Model architecture

The architecture model shown here illustrates the workflow that transforms raw customer opinions into actionable sentiment insights. Customer feedback in the form of social media posts is collected and preprocessed through data cleaning, where noise such as duplicates and misspelled words is addressed. This process uses the T-DIF feature extraction technique, which converts the cleaned text into numerical vectors. These vectors are then fed into an SVM, which separates the data into different categories and classes. Finally, the model delivers a sentiment prediction classified as negative, positive, or neutral. Figure 1 demonstrates this analogy.

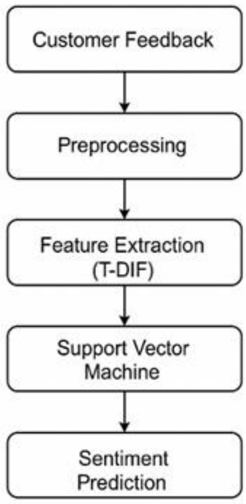


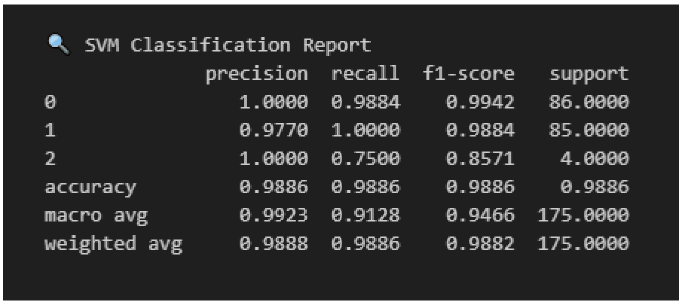
Figure 1. Model architecture

4. RESULTS AND DISCUSSION

4.1. Training the model

The sentiment analysis model was trained using a labelled dataset of customer feedback about CIMAS collected from social media platforms and review sites. The dataset was divided into 80% training and 20% testing sets. After applying pre-processing and feature extraction techniques (e.g., TF-IDF vectorization), three classifiers were tested: NB, LR, and SVM.

The SVM model delivered the best overall results, achieving an accuracy of 98.86%. The weighted precision, recall, and F1-score were 0.9888, 0.9886, and 0.9882, respectively. While the model classified positive (class 0) and neutral (class 1) sentiments with nearly perfect precision and recall, it showed slightly reduced recall (0.750) for negative sentiment (class 2) due to class imbalance in the dataset. Sample classification is observed in Figure 2.



🔍 SVM Classification Report

	precision	recall	f1-score	support
0	1.0000	0.9884	0.9942	86.0000
1	0.9770	1.0000	0.9884	85.0000
2	1.0000	0.7500	0.8571	4.0000
accuracy	0.9886	0.9886	0.9886	0.9886
macro avg	0.9923	0.9128	0.9466	175.0000
weighted avg	0.9888	0.9886	0.9882	175.0000

Figure 2. SVM classification report

These results demonstrate that the SVM model is highly effective for real-world sentiment classification in the context of healthcare-related customer feedback. SVMs were selected for their ability to handle high-dimensional spaces and their proven effectiveness in text classification. We performed 5-fold cross-validation and hyperparameter tuning (adjusting the C parameter and kernel type) to optimize performance and prevent overfitting.

Sentiment classification equation (SVM kernel function), the SVM model uses a kernel function to transform input features into a higher-dimensional space, allowing the classifier to draw optimal decision boundaries between sentiment categories that may not be linearly separable in the original feature space between sentiment classes. The classification equation is (1).

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b) \quad (1)$$

Where: i) f(x) is the predicted sentiment label; ii) x_i are the support vectors; iii) y_i are the sentiment labels; iv) αi are the model coefficients; v) $K(x_i,x)$ is the kernel function; and vi) b is the bias term.

This model was selected for its robustness in handling high-dimensional text data and its strong generalization on unseen examples. The kernel function (linear in this case) helped distinguish between subtle sentiment differences found in feedback data.

4.2. Overall model results

Sentiment distribution is an analysis of the sentiment predictions across the entire test dataset, which revealed the following distribution: i) neutral sentiment: 49.1%; ii) positive sentiment: 47.5%; and iii) negative sentiment: 3.4%. From Figure 3, it is clear that neutral and positive sentiments dominate customer feedback toward CIMAS, suggesting that the public generally views the organization in a balanced to favorable light. However, the low level of negative sentiment (3.4%), while small, still flags isolated areas of dissatisfaction that may require targeted attention from management. Figure 3 shows the distribution of sentiments.

4.3. Examples of sentiment classification

To validate interpretability, several sample outputs were analyzed. Table 3 shows an example of model predictions for randomly selected feedback. These examples highlight the model's ability to identify sentiment contextually, even when language is ambiguous or emotionally subtle.

Table 3. Model predictions for randomly selected feedback

Sample feedback	Actual sentiment	Predicted sentiment
"CIMAS service at Borrowdale branch was quick and helpful! "	Positive	Positive
"The mobile app always crashes when I need it most."	Negative	Negative
"But is CIMAS really like that? "	Neutral	Neutral

4.4. Prediction time and real-time feasibility

The model's average prediction time per comment was recorded at 0.001993 seconds, which qualifies as real-time in most web-based or mobile application use cases. This speed supports deployment in live feedback dashboards or automated customer service monitoring systems, offering CIMAS immediate visibility into customer sentiment trends. Figure 4 shows the code extract to detail the average time taken.

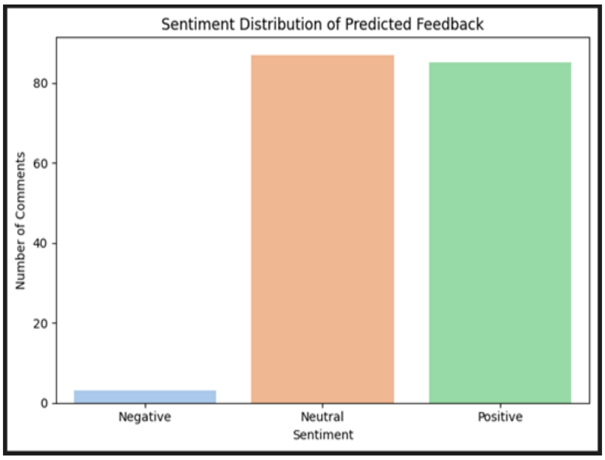


Figure 3. Sentiment distribution of predicted feedback

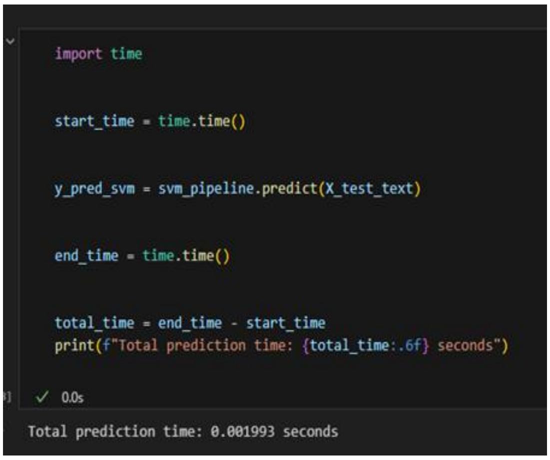


Figure 4. Prediction time and real-time feasibility

4.5. Insights from sentiment trends

Through deeper textual analysis of negative sentiment clusters, the following recurring themes were identified: i) mobile app issues: frequent complaints about usability and system errors; ii) claim processing delays: negative feedback regarding the time taken for reimbursement; and iii) branch-level service variability: mixed reviews regarding customer service across locations. These insights enable CIMAS to prioritize service improvements in specific departments and channels.

4.6. Conclusion of results

This study demonstrates that machine learning can successfully be applied to classify and analyze sentiment in real-world customer feedback data. The sentiment analysis model achieved high classification accuracy, provided actionable insights, and responded within real-time constraints. These results validate the use of machine learning in augmenting traditional customer experience management efforts, enabling proactive reputation management and strategic planning at CIMAS.

5. CONCLUSION

This research focused on the development of a machine learning-based sentiment analysis model designed to classify customer feedback directed toward CIMAS Medical Aid Society. The main objective was to enable the automatic identification of sentiment, positive, negative, or neutral, from unstructured text data sourced from social media platforms and customer review portals. It was developed with supervised learning, and the SVM algorithm showed the highest classifier performance. Pre-processing methods like tokenization, stop-word removal, and TF-IDF feature extraction were used to keep input data clean and prepared for effective training. Evaluation metrics showed that the model achieved high accuracy and reliable classification, with performance levels suitable for real-time deployment. Importantly, the analysis revealed not only a majority of positive sentiment toward CIMAS but also a recurring set of concerns around mobile app functionality and claim processing. These insights are actionable and can directly inform service improvements. The implications of this research support the adoption of sentiment analysis as a strategic customer intelligence platform in healthcare. With sentiment classification automated, CIMAS is able to monitor public sentiment continuously, act on feedback pre-emptively, and adjust operations to suit client needs better. This places the company in a better position to make evidence-based decisions that promote brand trust and service delivery.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to CIMAS Medical Aid Society for their support in this research. Special thanks to the team members who provided valuable feedback and helped in identifying relevant feedback sources and operational use cases. Their contribution was essential to the development and contextual relevance of this study.

FUNDING INFORMATION

The authors state that no funding is involved in doing this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Muzondiwa Karomo	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Mainford Mutandavari		✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	
Wilton Muzava		✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	

C: **C**onceptualization
M: **M**ethodology
So: **So**ftware
Va: **Va**lidation
Fo: **Fo**rmal analysis
I: **I**nvestigation
R: **R**esources
D: **D**ata Curation
O: Writing - **O**riginal Draft
E: Writing - Review & **E**diting
Vi: **Vi**sualization
Su: **Su**pervision
P: **P**roject administration
Fu: **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors certify that the work described in this paper was not influenced by any known conflicting financial interests or personal ties that could have appeared to influence the work reported in this paper. The authors state no conflict of interest.

ETHICAL APPROVAL

This study adheres to ethical guidelines for research in telecommunications. All data used was transformed, and all personal identifiers were eliminated [21]. No personal information was exposed [8]. Institutional approval was granted by the Harare Institute of Technology for undertaking this study. The dataset used for this research was derived from publicly available social media comments and feedback. All data was anonymized, and no personally identifiable information was retained, in adherence with ethical standards for public data usage.

DATA AVAILABILITY

The dataset's availability is restricted because of its proprietary nature and the presence of commercially sensitive information. However, the anonymized dataset used for model evaluation may be made available by the author [MK], upon request.

REFERENCES

[1] B. Liu, *Sentiment analysis and opinion mining*. San Rafael, California: Morgan & Claypool Publishers, 2012.
[2] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies (NAACL-HLT), Minneapolis, Minnesota*, 2019, pp. 4171–4186.
[3] Y. Chang, "Multilingual sentiment analysis during the pandemic using deep learning models," *Applied and Computational Engineering*, vol. 43, no. 1, pp. 284–293, 2024, doi: 10.54254/2755-2721/43/20230847.
[4] M. Malinga, I. Lupanda, M. W. Nkongolo, and P. van Deventer, "A multilingual sentiment lexicon for low-resource language translation using large language models and explainable AI," *Advances in Neural Information Processing Systems*, 2021.
[5] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," 2009.
[6] E. Alsentzer *et al.*, "Publicly available clinical," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, Minneapolis, Minnesota, USA, 2019, pp. 72–78. doi: 10.18653/v1/W19-1909.
[7] K. Huang, J. Altosaar, and R. Ranganath, "ClinicalBERT: modeling clinical notes and predicting hospital readmission," *arXiv: 1904.05342*, 2020.
[8] D. I. Adelani *et al.*, "Masakhaner: named entity recognition for African languages," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1116–1131, 2021, doi: 10.1162/tacl_a_00416.
[9] M. Wang, H. Adel, L. Lange, J. Strötgen, and H. Schütze, "NLNDE at SemEval-2023 task 12: adaptive pretraining and source language selection for low-resource multilingual sentiment analysis," *17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop*, pp. 488–497, 2023, doi: 10.18653/v1/2023.semeval-1.68.
[10] T. Filip, M. Pavlíček, and P. Sosík, "Fine-tuning multilingual language models in Twitter/X sentiment analysis: a study on Eastern-European V4 languages," *arXiv: 2408.02044*, Aug. 2024.
[11] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, pp. 328–339, 2018, doi: 10.18653/v1/p18-1031.
[12] N. Sharma and V. Jain, "Machine learning and sentiment analysis: analyzing customer feedback," in *Human-Centered Approaches in Industry 5.0: Human-Machine Interaction, Virtual Reality Training, and Customer Sentiment Analysis*, New York, United States: IGI Global Scientific Publishing, 2024.
[13] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, "Topic modeling in embedding spaces," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 439–453, 2020, doi: 10.1162/tacl_a_00325.
[14] E. Öhman, M. Pàmies, K. Kajava, and J. Tiedemann, "XED: a multilingual dataset for sentiment analysis and emotion detection," in *Proceedings of the 28th International Conference on Computational Linguistics*, Stroudsburg, PA, USA, 2020, pp. 6542–6552. doi: 10.18653/v1/2020.coling-main.575.
[15] G. Bhatia, I. Adebara, A. R. Elmadany, and M. A.- Mageed, "UBC-DLNLP at SemEval-2023 Task 12: impact of transfer learning on African sentiment analysis," *17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop*, pp. 246–255, 2023, doi: 10.18653/v1/2023.semeval-1.33.
[16] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.
[17] I. Aattouchi, S. Elmendili, and F. Elmendili, "Sentiment analysis of health care: review," *E3S Web of Conferences*, vol. 319, 2021, doi: 10.1051/e3sconf/202131901064.
[18] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Applied Intelligence*, vol. 48, no. 5, pp. 1218–1232, 2018, doi: 10.1007/s10489-017-1098-6.
[19] L. Hakim, I. Nuryasin, and S. Nugroho, "Sentiment analysis in insurance: a systematic review of approaches, techniques, and applications," *Multidisciplinary Reviews*, vol. 8, no. 10, 2025, doi: 10.31893/multirev.2025323.
[20] J. Y. Le Chan, K. T. Bea, S. M. H. Leow, S. W. Phoong, and W. K. Cheng, "State of the art: a review of sentiment analysis based on sequential transfer learning," *Artificial Intelligence Review*, vol. 56, no. 1, pp. 749–780, 2023, doi: 10.1007/s10462-022-10183-8.
[21] H. P. Suresha and K. K. Tiwari, "Topic modeling and sentiment analysis of Twitter data," *Asian Journal of Research in Computer Science*, pp. 13–29, 2021, doi: 10.9734/ajrcos/2021/v12i230278.

[22] J. Barnes, R. Klinger, and S. S. im Walde, “Projecting embeddings for domain adaptation: joint modeling of sentiment analysis in diverse domains,” *COLING 2018 - 27th International Conference on Computational Linguistics, Proceedings*, pp. 818–830, 2018.

[23] L. Sheng, Z. Wang, L. Zhang, and L. Jiang, “Application of sentiment analysis based on deep learning in public opinion monitoring in healthcare network,” *International Journal of High Speed Electronics and Systems*, Mar. 2025, doi: 10.1142/S0129156425404279.

[24] McKinsey & Company, “Global insurance report 2023: reimagining life insurance,” 2022.

[25] A. Ghosh, S. Umer, B. C. Dhara, and G. G. M. N. Ali, “A multimodal pain sentiment analysis system using ensembled deep learning approaches for IoT-enabled healthcare framework,” *Sensors*, vol. 25, no. 4, 2025, doi: 10.3390/s25041223.

[26] A. Khan, “Improved multi-lingual sentiment analysis and recognition using deep learning,” *Journal of Information Science*, vol. 51, no. 1, pp. 284–291, 2025, doi: 10.1177/01655515221137270.

[27] K. Fujihira and N. Horibe, “Multilingual sentiment analysis for web text based on word-to-word translation,” *Proceedings - 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 74–79, 2020, doi: 10.1109/IIAI-AAI50415.2020.00025.

[28] A. Owoyemi, J. Owoyemi, A. Osiyemi, and A. Boyd, “Artificial intelligence for healthcare in Africa,” *Frontiers in Digital Health*, vol. 2, Jul. 2020, doi: 10.3389/fdgth.2020.00006.

[29] T. M. Omran, B. T. Sharef, C. Grosan, and Y. Li, “Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach,” *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4111929.

[30] A. G. T. AbuRaed, E. A. Prikryl, G. Carenini, and N. Z. Janjua, “Long COVID discourse in Canada, the United States, and Europe: topic modeling and sentiment analysis of Twitter data,” *Journal of Medical Internet Research*, vol. 26, 2024, doi: 10.2196/59425.

BIOGRAPHIES OF AUTHORS

Muzondiwa Karomo holds a Bachelor of Technology degree in Information Technology from the School of Information Science and Technology, Harare Institute of Technology. He is currently pursuing a Master of Technology degree in Cloud Computing. His primary interest lies in research-driven technology development, with a particular focus on cloud computing. He can be contacted at email: mkaromo@gmail.com.

Mainford Mutandavari is a Ph.D. Scholar at SRMIST University, India, a lecturer and postgraduate studies coordinator at the Harare Institute of Technology (HIT), Zimbabwe. With advanced degrees in Computer Science and Strategy and Innovation, his research spans data analytics, cybersecurity, IoT, AI, and cloud computing. He is a member of HIT’s Cybersecurity and AI research groups and actively contributes to national ICT standards through the Standards Association of Zimbabwe. He has published widely on topics such as data loss prevention systems, digital learning infrastructure, and e-health security. His work bridges academic research with industry applications, focusing on practical digital solutions for education, telecommunications, and healthcare in Zimbabwe. He is also involved in curriculum development, postgraduate supervision, and building academic-industry partnerships. He can be contacted at email: mmutandavari@hit.ac.zw.

Wilton Muzava is an academic affiliated with the Harare Institute of Technology (HIT) in Zimbabwe. He holds a Bachelor of Technology degree in Information Security and Assurance and is a lecturer in the School of Information Science and Technology, where he teaches courses for the (B.Sc.) Information Security and Assurance program. He has further specialized by pursuing a Master of Technology (M.Tech.) degree in Cloud Computing. His research interests are particularly strong in the convergence of modern technologies, specifically encompassing data science, internet of things (IoT), computer vision, and cybersecurity. This interdisciplinary focus highlights his expertise in how these fields interact, especially within the context of cloud environments. His work is likely to explore the challenges and solutions at the intersection of these critical areas, such as securing IoT data in the cloud or leveraging data science for cybersecurity analytics. He can be contacted at email: wmuzava@hit.ac.zw.