

Transformer-Based Sentiment Analysis of Monkeypox (Mpox) Discourse on Twitter for Digital Epidemiology

Hazel Chamboko[#], Kudakwashe Maguraushe*, Belinda Ndlovu[#]

[#]Department of Informatics and Analytics, National University of Science and Technology, Bulawayo, Zimbabwe

* School of Computing, University of South Africa, Pretoria, South Africa

E-mail: [hazelchamboko\[at\]gmail.com](mailto:hazelchamboko[at]gmail.com), [magark\[at\]unisa.ac.za](mailto:magark[at]unisa.ac.za), [belinda.ndlovu\[at\]nust.ac.zw](mailto:belinda.ndlovu[at]nust.ac.zw)

ABSTRACTS

The global Monkeypox (Mpox) outbreak has generated substantial public discourse on social media platforms such as Twitter (X), where users express emotions ranging from fear and scepticism to reassurance and support. Understanding these sentiments is increasingly important for public health surveillance, digital epidemiology, and crisis communication, as they provide timely insights into public perceptions and behavioural responses during infectious disease outbreaks. However, conventional sentiment analysis approaches, including lexicon-based and traditional machine learning methods, often struggle to capture the contextual complexity, ambiguity, and informal linguistic patterns characteristic of social media discourse. This study evaluates the effectiveness of transformer-based transfer learning for sentiment analysis of Mpox-related Twitter posts. A total of 6,377 tweets were collected using Apify web-scraping tools, of which 5,142 English-language tweets were retained after filtering duplicates and non-English content. Initial weak-supervision labels were generated using VADER to bootstrap training, after which a transformer-based model was fine-tuned to learn contextual representations beyond lexicon-based rules. The model was evaluated using an 80:20 train-test split, achieving an overall accuracy of 83.92%, with macro-averaged precision, recall, and F1-score of 79.78%, 79.67%, and 79.57%, respectively. The results indicate that the transformer-based model outperformed the VADER baseline in capturing nuanced sentiment patterns in Mpox-related discourse. To demonstrate practical applicability, the trained model was deployed through an interactive Streamlit dashboard that enables users to upload tweet datasets, generate sentiment predictions, and visualise sentiment trends. The findings highlight the potential of transformer-based sentiment analysis for large-scale social discourse monitoring and its applicability in digital epidemiology, infodemiology, and public health decision-support during infectious disease outbreaks.

Manuscript received Apr 19, 2026;
revised May 3, 2026, accepted Aug
7, 2026 Date of publication Sep
30, 2026. International Journal,
JITSI : Jurnal Ilmiah Teknologi
Sistem Informasi licensed under a
Creative Commons Attribution-
Share Alike 4.0 International
License



Keywords / Kata Kunci — *Digital Epidemiology, Mpox, Sentiment Analysis, Transfer Learning, Transformer Models*

CORRESPONDING AUTHOR

Kudakwashe Maguraushe
School of Computing, University of South Africa, Pretoria, South Africa
Email: [magark\[at\]unisa.ac.za](mailto:magark[at]unisa.ac.za)

1. INTRODUCTION

The Monkeypox virus causes monkeypox (Mpox), a zoonotic viral disease belonging to the Orthopoxvirus genus within the Poxviridae family. Following the eradication of smallpox in 1980, Mpox emerged as the most significant Orthopoxvirus infection affecting humans [1]. Historically endemic to parts of Central and West Africa, Mpox has recently evolved into a global public health concern. Between June 2022 and June 2024, 116

countries reported Mpox cases, with a cumulative total of 99,176 confirmed cases worldwide [2]. The rapid international spread of the disease, coupled with uncertainties surrounding transmission, vaccination, and containment strategies, has intensified public discourse across digital platforms. Consequently, understanding public perceptions, emotional responses, and behavioural reactions to Mpox has become increasingly important for policymakers and public health authorities seeking to design effective interventions [3], [4].

In recent years, natural language processing (NLP) has become an essential tool for analysing large-scale textual data generated through social media and online platforms [3]. Within NLP, sentiment analysis, also referred to as opinion mining or text sentiment analysis, focuses on identifying and classifying emotional or attitudinal expressions in text [5]. During pandemics and other public health emergencies, sentiment analysis enables researchers and decision-makers to understand the affective states expressed by the public regarding disease outbreaks, health policies, and interventions [6]. Public sentiment, widely expressed on social media, provides valuable insights that can support healthcare communication, policy formulation, and epidemic response strategies [7].

Social media platforms such as Twitter have become central to the dissemination and amplification of public opinion during disease outbreaks. The first publicly available dataset for Mpox analysis was introduced by [8], who collected approximately 255,000 Twitter posts related to the outbreak from 7 May 2022 onwards. Similarly, [9] combined Twitter and web news data to predict the spread of the Mpox virus by analysing trends and extracting statistically relevant country-specific insights. More broadly, the mining and analysis of Twitter conversations have attracted significant attention across healthcare, epidemiology, big data, and computer science, as demonstrated by studies examining diseases such as Ebola, E. coli, Dengue, and Human Papillomavirus (HPV) [10]. These studies highlight the growing importance of social media as a real-time indicator of public sentiment and behavioural trends during infectious disease outbreaks.

The recent Mpox outbreak, particularly in regions such as the Democratic Republic of Congo (DRC), has exposed critical public health challenges and reinforced the importance of understanding public perceptions of disease risk, prevention, and response [4]. Social media often disseminates information faster than traditional news media, serving as both an information source and a channel for misinformation, stigma, and public panic [11]. Consequently, real-time sentiment monitoring has become increasingly important for governments and public health institutions seeking to formulate timely mitigation strategies.

Despite this growing interest, accurately analysing sentiment in outbreak-related social media discourse remains challenging. Social media text is typically short, noisy, informal, and highly context-dependent, often containing slang, sarcasm, abbreviations, emojis, and rapidly evolving linguistic expressions. Traditional lexicon-based approaches, such as the National Research Council (NRC) lexicon, can identify emotion-bearing words effectively but often fail to capture contextual nuances, such as satire, slang, or implicit sentiment [12]. Similarly, [13] applied classical machine learning algorithms, including Support Vector Machine, Naïve Bayes, Logistic Regression, and Random Forest, using Bag-of-Words (BoW) and TF-IDF feature extraction. Although these methods achieved moderate performance, they struggled to capture nuanced emotions such as resilience, sarcasm, and contextual ambiguity.

Lexicon-based tools such as VADER have also been widely adopted due to their simplicity and computational efficiency [14]. However, VADER cannot reliably distinguish bot-generated tweets or adapt effectively to domain-specific linguistic variations [10]. In contrast, [15] demonstrated that combining TextBlob and Flair improved stigmatisation and disinformation detection, though their approach required substantial data volumes to achieve high accuracy. Similarly, [16] employed comparable techniques but failed to detect automated tweets and recommended broader multilingual studies using more advanced models. While the Twitter API remains one of the most commonly used tools for data collection, it imposes significant restrictions on access volume and on the retrieval of historical data, thereby constraining large-scale or longitudinal analyses [17].

Recent advances in transformer-based NLP offer a promising alternative to these limitations [18]. Transformer architectures such as BERT and RoBERTa use self-attention mechanisms to capture long-range contextual dependencies across text sequences [19], enabling improved interpretation of contextual polarity, sarcasm, negation, and semantic nuance. These models have achieved state-of-the-art performance across a range of NLP tasks due to their ability to learn contextual embeddings from large-scale corpora. In outbreak-related social media analysis, where discourse is highly dynamic and emotionally complex, such contextual modelling is particularly valuable.

A key advantage of transformer-based models is their compatibility with transfer learning. Transfer learning is the process of leveraging knowledge from a source task or dataset to improve performance on a related target task [20]. This is especially beneficial in settings where labelled data are limited. For years, the scarcity of sufficiently large labelled datasets has hindered the training of deep neural networks from scratch [21]. Studies have shown that transfer learning approaches often outperform traditional models, particularly when domain-specific datasets are relatively small [22]. Furthermore, transfer learning reduces training time and computational requirements while improving memory efficiency and predictive performance [23], [24].

Despite these advancements, few studies have systematically investigated transformer-based transfer learning for Mpox-related sentiment analysis while simultaneously integrating such models into practical monitoring systems. Existing work such as [25] has largely focused on descriptive analytics, lexicon-based methods, or traditional machine learning pipelines without fully exploiting contextual language modelling or demonstrating deployable real-time analytics tools. This creates both a methodological and practical gap in the literature.

To address this gap, this study investigates the use of a transformer-based transfer learning model, specifically ELECTRA, for sentiment classification of Mpox-related Twitter discourse. The study further integrates the model into an interactive dashboard for real-time or near real-time sentiment monitoring. Specifically, this paper makes three contributions. First, it evaluates the effectiveness of transformer-based transfer learning for classifying Mpox-related tweets into positive, negative, and neutral sentiment categories. Second, it compares the proposed model against lexicon-based and traditional machine learning approaches. Third, it demonstrates the model's practical applicability by developing a deployable analytics platform to support digital epidemiology and public health communication monitoring.

Accordingly, this study seeks to answer the following research questions:

RQ1: To what extent can transformer-based transfer learning models accurately classify sentiment in Mpox-related Twitter discourse?

RQ2: How does the proposed model compare with traditional machine learning and lexicon-based sentiment analysis approaches?

RQ3: Can the resulting sentiment classification model be integrated into a practical analytics tool for monitoring public discourse during infectious disease outbreaks?

By answering these questions, this study contributes to the intersection of machine learning, digital epidemiology, and public health informatics, while offering methodological and practical insights for future epidemic intelligence systems.

2. RESEARCH METHODOLOGY

The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework was used to guide the development, training, and deployment of the Mpox sentiment analysis model [26]. CRISP-DM provides a structured framework for developing and deploying data mining and machine learning solutions [26], [27]. The model consists of the business understanding, data understanding, data preparation, modelling, evaluation, and deployment phases. The methodology adopted in this study implemented these phases to ensure transparency and reproducibility of the modelling process.

2.1. Business Understanding

In the initial phase, research objectives were defined, and the relevance of analysing information disseminated to the public about infectious disease outbreaks was identified.

Social media platforms such as Twitter (X) serve as a major channel of communication during public health crises, with users commenting on emerging diseases and expressing discontent. Understanding these sentiments may thus help public health authorities monitor misinformation, assess public perception, and improve communication strategies [28].

This research was supposed to develop a sentiment classification system that automatically categorises Tweets related to Mpox into positive, negative, and neutral classes. Several visualisation interfaces were to be developed to present sentiment patterns in the datasets.

2.2. Data Collection and Understanding

The data collection and understanding phase aimed to acquire a relevant and representative corpus of Mpox-related social media discourse for sentiment analysis and subsequent model training. To achieve this, publicly available Twitter (X) posts were collected using Apify, a web-scraping platform that retrieves Twitter content by interacting with publicly accessible endpoints and reverse-engineering Twitter's unofficial API. Apify has been widely used in social media analytics research due to its ability to collect large-scale publicly available content beyond the restrictions imposed by the official Twitter API [15].

A keyword-based retrieval strategy was employed to capture tweets related to Mpox. The search terms were designed to reflect common references to the disease, its symptoms, outbreak discussions, prevention measures, and public reactions. These included terms associated with the disease name, symptom descriptions, vaccine discussions, public health interventions, and related debates. Using this retrieval strategy, 6,377 tweets were initially collected. Each record in the dataset contained key metadata attributes necessary for downstream analysis, including:

- Tweet text: the primary unstructured textual content used for sentiment analysis;
- Date and time of posting: used for temporal trend analysis and sentiment evolution monitoring;
- Author identifier: used for traceability and duplicate detection during pre-processing.

The collected dataset was exported in CSV format to facilitate integration with Python-based data analysis tools. Data exploration and pre-processing were conducted using libraries such as Pandas for structured data manipulation, NLTK for text processing, and regular expressions (Regex) for pattern-based cleaning and filtering.

An initial exploratory review of the dataset revealed several characteristics typical of social media discourse. These included informal language, inconsistent punctuation, hyperlinks, hashtags, mentions, emojis, abbreviations, and other non-standard textual elements. Such characteristics can negatively affect sentiment analysis performance if not addressed appropriately. Consequently, pre-processing was required to normalise the textual content before sentiment annotation and model training. This involved cleaning noisy text, tokenisation, and standardisation of lexical forms to improve the quality of downstream machine learning inputs.

To improve consistency in sentiment annotation, a language filtering procedure was applied. Since initial weak supervision labels were generated using the VADER sentiment analysis tool, which performs optimally on English-language text, non-English tweets were excluded. This process resulted in the removal of 1,235 tweets, reducing the dataset to 5,142 English-language tweets. Duplicate tweets and retweets with identical textual content were also removed to minimise bias in sentiment distribution and ensure that each record represented a unique opinion or information-sharing instance.

The final dataset therefore, consisted of 5,142 unique Mpox-related English-language tweets, each containing cleaned textual content and associated metadata. This corpus served as the basis for subsequent pre-processing, weakly supervised sentiment annotation, transformer-based model fine-tuning, and temporal sentiment trend analysis.

2.3. Text Pre-processing

The collected tweets in their raw form contained substantial noise and irregularities commonly associated with social media text. These included web links, hashtags, user mentions, emojis, inconsistent punctuation, and other non-standard textual elements that could negatively affect sentiment classification performance. To improve data quality and ensure consistency for downstream analysis, a structured text pre-processing pipeline was implemented to clean and normalise the textual content prior to sentiment annotation and model training. The pre-processing pipeline consisted of the following steps:

- Lowercasing: All text was converted to lowercase to ensure lexical consistency and reduce feature sparsity caused by case variations.
- URL Removal: Hyperlinks and embedded URLs were removed using regular expression (Regex) patterns, as they generally do not contribute meaningful semantic information to sentiment analysis.
- User Mention and Hashtag Removal: Twitter-specific tokens such as user mentions (e.g., @username) and hashtags were removed or normalised to reduce noise and prevent overfitting to platform-specific artefacts.
- Punctuation, Emoji, and Special Character Removal: Non-alphanumeric symbols, emojis, and unnecessary punctuation marks were removed to simplify textual representations and improve tokenisation consistency.
- Whitespace Normalisation: Excessive or irregular whitespace was standardised to ensure clean token separation and consistent formatting.

These pre-processing operations were implemented using Python libraries, including re for pattern matching, NLTK for tokenisation and text handling, and Pandas for structured data manipulation.

Figure 1 illustrates the overall pre-processing workflow applied to the raw Twitter dataset before sentiment annotation and transformer-based model fine-tuning.

FIGURE 1. Text Pre-processing Pipeline

Textual Features, The different procedures followed were:

- Extracting the word tokens from the tweet using regular expressions.
- A word counter that indicates the number of words in the tweet
- A sentence length counter that indicates the number of characters.
- A spelling error count was identified as such through a spell-checking algorithm.

Such features saved the users of the tweets and gave room for further exploratory analysis on linguistic patterns.

Temporal features, The original data contained 'date' and 'time', so the two were merged into a single feature called 'datetime'. This allowed the extraction of further temporal features:

- Hour of posting

After the pre-processing step, duplicate tweets and retweets were removed to avoid a biased sentiment distribution. Thus, pre-processing steps produced clean text suitable for feature extraction and sentiment analysis.

2.4. Feature Engineering and Sentiment Labelling

To make the dataset more useful for model building and exploratory analysis, additional features were added to the existing set of cleaned tweets.

- Day of posting
- Month and the year

The previously mentioned attributes further enabled analysis of minute-by-minute activity trends within the dataset.

Sentiment labelling. The Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiments were employed to derive the labels of sentiments. The VADER sentiment analyser calculates the compound polarity score that ranges from -1 (most negative) to +1 (most positive). Tweets were classified into the three standard VADER sentiment classes using the following thresholds:

- Positive: compound ≥ 0.05
- Neutral: $-0.05 < \text{compound} < 0.05$
- Negative: compound < -0.05

The labeled dataset created was used as input for the sentiment classification model based on a transformer architecture. To evaluate the dataset's balance, this class distribution was compared with the sentiment annotations. Neutral tweets accounted for a slightly higher percentage than the other two tweet types, echoing informational posts and news-related tweets typically seen on social media, while positive and negative tweets were minor but similar in number.

It was important to emphasise that VADER's main functionality for sentiment annotation was driven by the weak-supervision mechanism and the generation of initial polarity labels. The main goal of the study was not to capitalise on VADER predictions but to examine the role of a contextual transformer model in learning richer representations of sentiment patterns across the dataset. The comparison in Results indicates the marked superiority of our ELECTRA model over the lexicon baseline, suggesting that the transformer architecture captures contextual relationships beyond the fixed sentiment rules employed by VADER.

2.5. Model Architecture

It used the ELECTRA transformer model for sentiment analysis. The ELECTRA transformer (Efficiently Learning Inference Kernel Rethinking Architectures) boosts training efficiency by detecting replaced tokens. Essentially, unlike BERT, which operates on a tradition of masked language modelling, the ELECTRA models learn a generator-discriminator model to make the novel token, where the generator replaces the missing token in the sentence, while the discriminator learns to detect if the token was present in the original form or was modified.

By allowing the use of the entire input sequence during training, this paper's model learns contextual embeddings efficiently. Figure 2 shows the ELECTRA architecture, as used in the experiment.

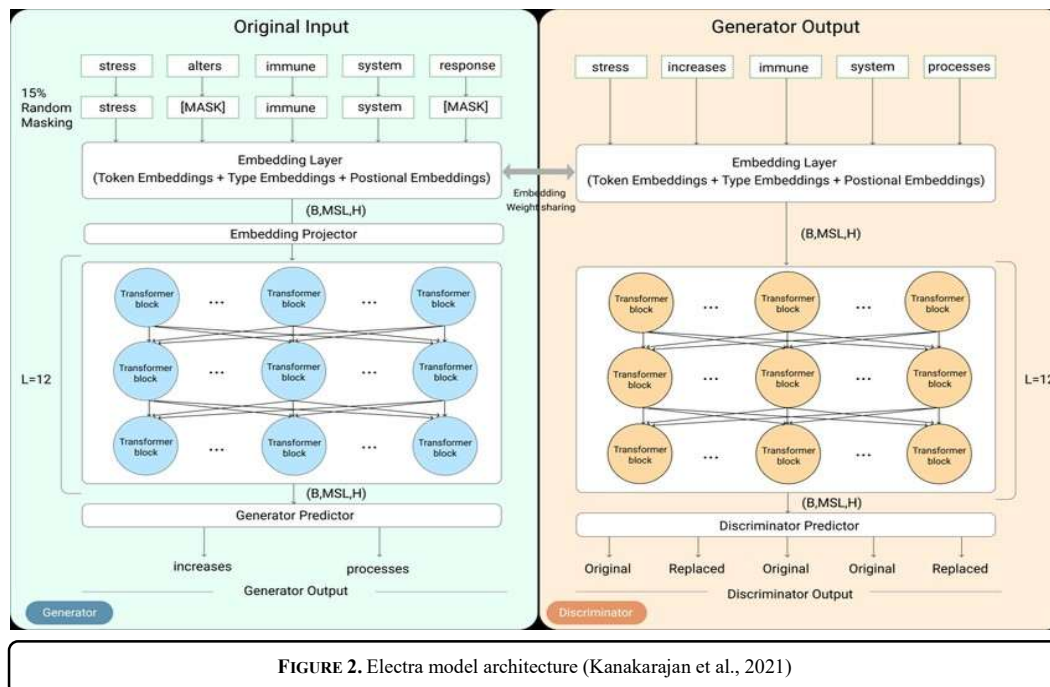


FIGURE 2. Electra model architecture (Kanakarajan et al., 2021)

2.6. Model Training

The study ran about 20 epochs with MultiModalPredictor using the default transformer settings from the Hugging Face library, including calibration, adalearning rate-rate scheduling, batch size-size adjustments based on available the GPU resource in Google Colab. In our work, the pre-trained Google/electra-base-discriminator model was fine-tuned from the labelled tweet dataset.

Training was done according to these parameters:

- 80:20 split for evaluation
- 20 epochs
- Accuracy evaluation metric
- Learning time: approximately 1 hour on Google Colab GPU resources.

AutoGluon automation includes pre-processing, feature extraction, hyperparameter optimisation, and model selection. Even though the dataset used in the study was relatively small, given the massive scale of large NLP benchmarks, transformer-based transfer learning is inherently well-suited to such scenarios. Electra and similar pretrained transformer models inherit linguistic knowledge from their pre-training corpora; they can be fine-tuned on smaller datasets and still achieve high accuracy [19]. This is a major advantage in transfer learning, given that a small dataset may be the only source of learning in specialised applications, such as analysing social media discourse during an emergent public health event, where vast, annotated datasets do not always exist.

The fine-tuning was already done on a pretrained transformer, so explicit data augmentation techniques were eschewed. Language-based transfer learning, in principle, would render further dataset augmentation unnecessary.

2.7. Model Evaluation Strategy

Model quality was assessed on a hold-out of the test set. The following were metrics for model evaluation:

- Accuracy
- Precision (macro-average)
- Recall (macro-average)
- F1-Score (macro-average)

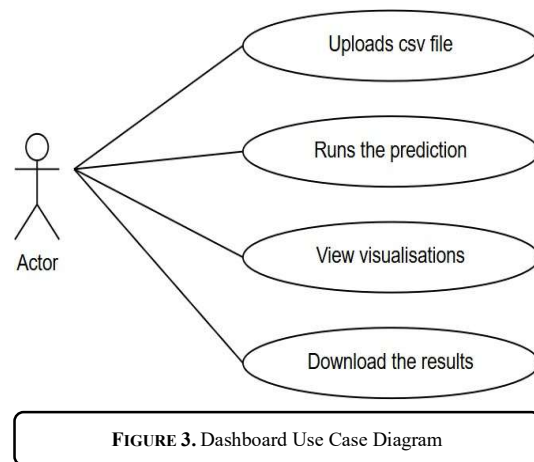
The macro-average was used to give equal weight to evaluations across sentiment classes, which is vital, especially when working with unbalanced datasets. A confusion matrix is used to compare classification models across sentiment classes. In analysing the human-understandable internal representation of tweet sentiment, t-distributed stochastic neighbour embedding (t-SNE) was applied to the learned text embeddings.

2.8. Model Deployment

Upon completion of model evaluation, the trained classifier was deployed through an interactive dashboard built with Streamlit. The dashboard was built in Python using Visual Studio Code and allowed users to perform sentiment analysis on new tweet datasets. The following options are available to users:

- Users can upload CSV files containing tweets.
- Automatic generation of sentiment predictions
- Users can visualise sentiment distributions and word clouds.
- Users get to download the classified dataset.

Figure 3 illustrates the user interaction flow on the deployed dashboard.



3. RESULTS

3.1. Model Performance

The Transformer model for sentiment classification was applied and compared against standard classification metrics (accuracy, precision, recall, and F1-score) on common classifiers. Separation of training and testing was done at an 80:20 ratio, and the model was trained on the training data; truly held-out data was not used during development.

In terms of metrics, the ELECTRA-based model achieves the best accuracy of 83.92% and surpasses the macro precision, recall, and F1-score values of 79.78%, 79.67%, and 79.57%, respectively. These macro-averaged metrics ensure fairness in evaluation, so that the smallest sentiment class does not dominate the model's evaluation. The above methods establish that the transformer-based model effectively detected contextual sentiment signals in Mpox-related social media discourse; nonetheless, this suggests that social media discourse on Twitter is often noisy and linguistically diverse. Social media datasets have highly skewed sentiment category distributions; hence, a macro-metrics-based evaluation was preferred to ensure balanced performance across good, neutral, and bad classes rather than simply following the majority class.

To further assess the effectiveness of the proposed approach, ELECTRA's performance was compared against a baseline lexicon-based sentiment analysis using VADER. The sentiment is mostly expressed in social media

data, with VADER used to detect sentiment in emoticons, slang, and punctuation. However, unlike transformer-based models, VADER relies heavily on predefined sentiment dictionaries and does not capture deep contextual relationships among words. Table 1 presents a comparison between the lexicon-based baseline and the proposed transformer-based model.

TABLE 1. Performance comparison between lexicon-based and transformer-based sentiment models

Model	Approach	Accuracy	Precision (Macro)	Recall (Macro)	F1 Score (Macro)
VADER	Lexicon-based sentiment analysis	68.41%	65.32%	64.87%	64.95%
ELECTRA (Proposed)	Transformer-based transfer learning	83.92%	79.78%	79.67%	79.57%

Specifically, the VADER baseline achieved an overall accuracy of 68.41%, whereas the ELECTRA-based model achieved 83.92%, representing an absolute improvement of 15.51 percentage points. Similar improvements were observed across the macro-averaged precision, recall, and F1-score metrics, indicating more balanced predictive performance across all sentiment classes. This is particularly important in multi-class sentiment classification, where imbalanced class performance may undermine a model's practical reliability in real-world monitoring applications.

The improved performance of ELECTRA can be attributed to its ability to model contextual relationships among words within a sentence. Unlike lexicon-based methods that infer sentiment using predefined polarity scores assigned to individual words or phrases, transformer-based architectures leverage self-attention mechanisms and large-scale pre-training to capture richer semantic and contextual dependencies. As a result, the model can better interpret complex linguistic patterns commonly found in social media text, including abbreviations, informal language, implicit sentiment, and context-dependent polarity shifts.

These findings reinforce the value of transfer learning for sentiment analysis of outbreak-related social media discourse. By leveraging pre-trained contextual representations and fine-tuning them on domain-specific data, transformer-based models can outperform traditional lexicon-based approaches in handling large volumes of unstructured and linguistically diverse content during public health events. This suggests that contextual language models may provide a more reliable foundation for real-time sentiment monitoring and digital epidemiology applications.

3.2. Confusion Matrix Analysis

To examine classification behaviour in greater detail, a normalised confusion matrix was generated as shown in Figure 4.

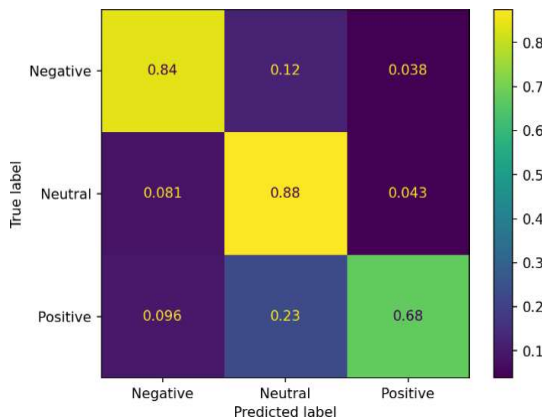


FIGURE 4. Confusion Matrix of the ELECTRA sentiment classifier

VADER, a widely used lexicon-based sentiment analysis tool, remains popular in social media analytics due to its simplicity, interpretability, and low computational requirements. However, the comparative results in Table 1 indicate that Mpox-related tweet classification benefits substantially from context-aware transformer-based models. In this study, the ELECTRA model achieved superior performance across all evaluation metrics, suggesting that contextual language representations provide a more robust mechanism for interpreting informal, nuanced, and context-dependent social media expressions.

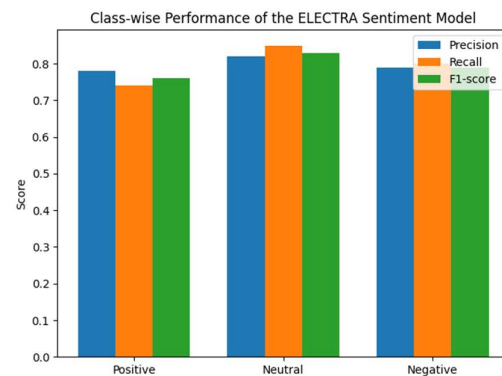


FIGURE 5. Class-wise precision, recall, and F1-score for the ELECTRA sentiment classification model across positive, neutral, and negative tweet categories

The confusion matrix indicates that 84% of negative tweets and 88% of neutral tweets were correctly classified. The positive class achieved a lower recall, with some instances misclassified as neutral. This suggests that distinguishing between positive and neutral sentiment remains a challenge due to subtle linguistic

differences. Surprisingly, major misclassification occurred between positive and neutral texts. In other words, positive and neutral sentiment types exhibit a somewhat similar undertone.

Overall, the confusion matrix indicates that, while the model still needs improvement, it performs well in recognising tweets with these specific polarities.

For example, this overlap between positive and neutral classifications could also involve subtler aspects of linguistic expression, such as irony, humour, and indirect sentiment, making them even harder for an automatic sentiment classifier to detect.

3.3. Class-wise Performance Analysis

To focus on the classifier's performance across various sentiment classes, class-wise precision, recall, and F1-score are analysed. As shown in Figure 5, the ELECTRA model's performance across the positive, neutral, and negative sentiment categories is presented. Class-based precision, recall, and F1 were computed for the sentiment classification model based on ELECTRA across the positive, neutral, and negative tweet classes. The model's overall performance varies across the three sentiment classes. The neutral class is systematically more accurate.

3.4. Embedding Space Visualisation

t-SNE was used to further examine how the model captured tweet sentiment. The t-SNE maps shown in this section were generated as a graphical depiction to give us, though still very preliminary, an understanding of the structure of the learned embedding space. Thus, the classifier assessment was the key. More evaluation can be done using the confusion matrix. t-SNE helps us understand how sentiment classification patterns are internal to the model. Figure 6 shows the two-dimensional projection of the training dataset's embeddings.

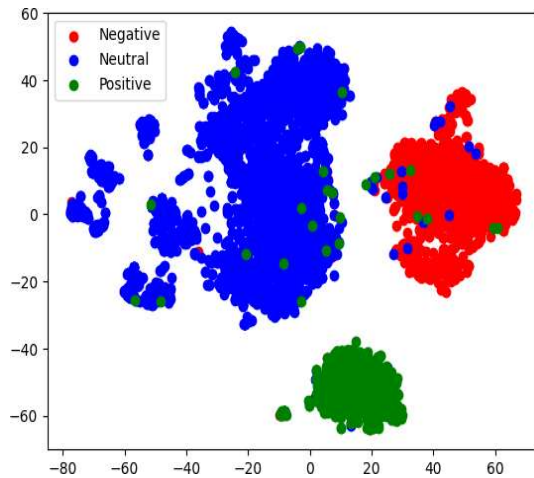


FIGURE 6. t-SNE Visualisation of training embeddings

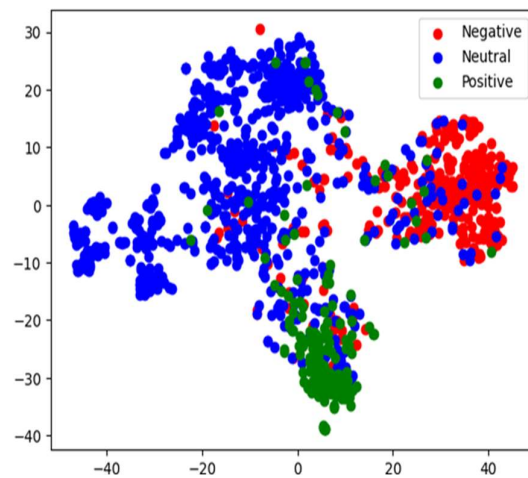


FIGURE 7. t-SNE Visualisation of test embeddings

The visualisation reveals that distinct clusters correspond to sentiment classes. The negative and neutral tweets have their own relatively well-defined clusters, while affirmative tweets only overlap slightly with neutral tweets. A t-SNE visualisation of the test dataset is shown in Figure 7.

Similar clustering patterns suggest that the model may have learned transferable semantic structures rather than solely memorising training patterns. These visualisations should be interpreted as a qualitative representation of the learned embedding structure and should not be used to evaluate the model's performance. These results are included in earlier analysis alongside clearly displayed classification figures.

3.5. Dashboard Output

Post-evaluation, the trained model was integrated into the Streamlit dashboard to demonstrate the practical applicability of the sentiment analysis system. Users can upload Tweet datasets and view sentiment predictions, with visualisations such as sentiment distribution charts and word clouds. The dashboard interface used for analysing the Tweets is shown in Figure 8.

The deployment of the machine learning models is used to build visualisation tools that track public discourse during health crises.



FIGURE 8. Dashboard landing page

4. DISCUSSION

The findings of this study provide strong empirical evidence that transformer-based transfer learning architectures are effective for modelling the linguistic complexity of Mpox-related social media discourse. In contrast to traditional sentiment analysis methods that rely on static lexicons or sparse vector representations, contextual transformer models capture semantic dependencies across words, phrases, and sentence structures, thereby enabling more nuanced interpretation of informal and highly dynamic online communication. This is particularly relevant in public health crises, where discourse is often characterised by slang, abbreviations, sarcasm, emotionally mixed expressions, and rapidly evolving narratives. The strong performance of the ELECTRA-based classifier therefore, reinforces growing evidence that contextual language models are becoming increasingly central to digital epidemiology and infodemiology applications [10], [15].

The model achieved balanced performance across the three sentiment classes—positive, negative, and neutral—as reflected by the macro-averaged precision, recall, and F1-score values. This is an important finding because social media datasets are often inherently imbalanced, with informational or neutral content dominating strongly polarised expressions. By assigning equal importance to each class, macro-averaged metrics provide a more reliable measure of robustness under class imbalance. The relatively stable class-wise performance suggests that the model generalises effectively across diverse sentiment categories, reducing the risk of majority-class bias and increasing confidence in its applicability to real-world public discourse monitoring.

The confusion matrix provides additional insight into the model's behaviour. The classifier performed particularly well on negative and neutral tweets, achieving class-specific accuracies of 84% and 88%, respectively. Negative tweets often contain explicit expressions of fear, criticism, or frustration, while neutral tweets are more likely to reflect factual updates or formal information-sharing. These categories may therefore exhibit clearer lexical and semantic markers. By contrast, the lower performance observed for the positive class highlights the inherent ambiguity of positive sentiment during public health crises. Positive discourse in outbreak settings is often subtle and may manifest as reassurance, solidarity, gratitude, or restrained optimism rather than overtly positive emotional language [12], [13]. This semantic overlap with neutral discourse is consistent with prior findings in crisis informatics and health communication research.

The comparative analysis further demonstrates the limitations of conventional sentiment analysis methods in outbreak-related discourse. Previous studies on Mpox-related Twitter data have predominantly employed classical machine learning approaches, such as Support Vector Machines, Naïve Bayes, and Logistic Regression, with Bag-of-Words or TF-IDF representations [15]. While these approaches can perform adequately in structured domains, they often struggle to capture contextual semantics, sarcasm, implicit sentiment, and evolving linguistic trends on social media. Similarly, lexicon-based tools such as VADER and TextBlob rely on predefined sentiment dictionaries, which limit adaptability in domain-specific or rapidly changing contexts [10], [13], [29]. In contrast, transformer-based models such as ELECTRA leverage large-scale pre-training to learn richer contextual representations, enabling more robust task-specific fine-tuning and stronger downstream performance [10].

The t-SNE visualisations provide qualitative support for these quantitative findings. Distinct clustering patterns observed in both training and test embedding spaces suggest that the model learned semantically meaningful representations rather than relying solely on superficial lexical cues. Tweets with similar emotional polarity tended to cluster together, indicating that the model captured deeper contextual relationships among words and phrases. Although overlap remained between positive and neutral clusters, this likely reflects the inherent ambiguity of natural language and the contextual fluidity of sentiment expression in crisis communication environments, rather than model inadequacy. From a practical perspective, integrating the transformer model into a deployable analytics platform is an important contribution. By combining AutoGluon for model optimisation and Streamlit for interactive visualisation, the study moves beyond algorithmic experimentation toward a usable decision-support artefact. This aligns with broader efforts in digital epidemiology to transform social media analytics into actionable surveillance tools for public health agencies. Automated sentiment-tracking systems may support early detection of misinformation, public panic, vaccine hesitancy, and communication gaps during emerging outbreaks. As social media increasingly functions as a real-time public opinion sensor, such systems can enhance situational awareness and inform responsive public health messaging strategies.

The implications of this study are threefold. Theoretically, it contributes to the growing body of evidence supporting transformer-based contextual modelling for health-related social media analytics. Methodologically, it demonstrates the utility of weak supervision combined with transformer fine-tuning in domains where manually annotated datasets are scarce. In practice, it offers a scalable prototype capable of supporting health communication monitoring and crisis response planning. These contributions position the study at the intersection of machine learning, public health informatics, and digital surveillance. Nevertheless, several limitations should be acknowledged. First, the dataset was primarily restricted to English-language tweets, limiting its representativeness of global discourse. Since Mpox outbreaks have affected multilingual populations, future work should explore multilingual transformer architectures such as XLM-RoBERTa and mBERT to improve linguistic inclusivity. Second, although the dataset size was sufficient for training and evaluation, it remains smaller than

that used in large-scale social media surveillance systems. Continuous data collection pipelines could improve temporal robustness and enable longitudinal analysis of sentiment evolution. Third, sentiment labels were generated using weak supervision through VADER, which may introduce label dependency and bias. Future research should incorporate manually annotated corpora and human-in-the-loop validation to improve robustness and external validity. Finally, future studies could extend beyond sentiment classification to include misinformation detection, topic evolution analysis, and behavioural intent modelling, thereby broadening the system's applicability in epidemic intelligence. Overall, this study demonstrates that transformer-based sentiment analysis offers significant promise for enhancing real-time understanding of public discourse during infectious disease outbreaks. By integrating contextual language modelling with deployable analytics interfaces, the proposed framework contributes toward more intelligent, scalable, and actionable tools for digital epidemiology, infodemiology, and public health crisis communication

5. LIMITATIONS

Although the findings demonstrate the potential of transformer-based transfer learning for analysing public sentiment during infectious disease outbreaks, several limitations should be acknowledged. First, the dataset was restricted to English-language tweets, which improved consistency during sentiment annotation and model training but limited the generalisability of the findings. Mpox outbreaks affected multilingual regions where languages such as French, Portuguese, and various indigenous African languages are widely spoken. Consequently, the analysed corpus may not fully represent the diversity of global public discourse surrounding the disease. Future studies should explore multilingual transformer architectures, such as XLM-RoBERTa or mBERT, to improve linguistic inclusivity and cross-cultural generalisability in epidemic-related sentiment analysis. Second, the initial sentiment labels were generated using the VADER sentiment analysis tool as a weak supervision mechanism prior to transformer fine-tuning. VADER is widely used in social media analytics for its ability to handle informal language and explicit emotional expressions. However, as a lexicon-based approach, it may not adequately capture sarcasm, irony, slang, or subtle shifts in contextual sentiment within complex tweet structures. Since the quality of weakly supervised labels directly influences downstream learning, any bias or misclassification introduced at this stage may have affected model performance. Future research should incorporate manually annotated datasets or human-in-the-loop validation to reduce label dependency and improve annotation reliability. Third, the dataset may contain automated or bot-generated content, which can distort observed sentiment distributions by artificially amplifying specific narratives or emotional tones. Social media platforms are known to host automated accounts that influence public discourse during health crises. Although duplicate tweets and retweets were removed during pre-processing, explicit bot detection was beyond the scope of this study. Future studies could integrate bot-detection frameworks or account-level behavioural analysis to improve the reliability of social media sentiment datasets. Fourth, although the final dataset of 5,142 tweets was sufficient for model training and evaluation, larger-scale datasets could improve robustness and external validity. Sentiment patterns during health crises are highly dynamic and may evolve rapidly in response to policy changes, media coverage, and emerging epidemiological developments. The dataset used in this study, therefore, represents a temporal snapshot of public sentiment during the selected collection period. Future research could implement continuous or longitudinal data collection pipelines to capture sentiment evolution over time and improve temporal generalisability. Fifth, the study employed a single hold-out validation strategy using an 80:20 train-test split. While this approach provides a practical estimate of model performance, it may not fully capture variability across different data partitions. Future studies could employ k-fold cross-validation or repeated stratified validation to provide more robust estimates of generalisation performance and reduce the risk of partition-specific bias. Finally, this study focused primarily on sentiment polarity classification rather than deeper semantic or discourse-level analysis. Public discussions surrounding infectious disease outbreaks are often multidimensional, involving misinformation, stigma, fear, vaccine hesitancy, trust, and policy debates. Restricting analysis to positive, negative, and neutral sentiment may oversimplify these complex narratives. Future work could extend this framework to incorporate misinformation detection, stance detection, topic modelling, or behavioural intent analysis to provide richer insights into the societal impact of infectious disease outbreaks and the role of social media in shaping public understanding of health risks. Despite these limitations, the study provides a useful methodological and practical foundation for applying transformer-based sentiment analysis in digital epidemiology, infodemiology, and public health surveillance contexts

6. FUTURE WORK

The findings of this study, together with its identified limitations, provide several opportunities for future research. First, future studies should explore multilingual sentiment analysis to capture public discourse across diverse linguistic communities better. Since Mpox outbreaks have affected multilingual regions, restricting analysis to English-language tweets may overlook important cultural and linguistic variations in public sentiment. Transformer architectures such as multilingual BERT and XLM-RoBERTa have demonstrated strong

performance in cross-lingual text classification and offer promising potential for analysing Mpox-related discourse across multiple languages. Such approaches may improve the representativeness of sentiment analysis and enable comparative analyses of public perceptions across geographic and linguistic contexts.

Second, improving the quality of sentiment annotation remains an important direction for future work. This study relied on weak supervision through automated sentiment scoring to generate initial labels. Although effective for bootstrapping model training, automated annotation may introduce systematic bias or fail to capture subtle emotional expressions. Future research could adopt hybrid annotation strategies that combine automated labelling with human-in-the-loop validation to improve label reliability and reduce annotation noise. Such approaches may be particularly useful for identifying more nuanced emotional states, such as fear, resilience, uncertainty, or scepticism, which extend beyond simple sentiment polarity.

Third, future research could implement continuous, real-time data-collection pipelines to monitor sentiment evolution during ongoing health crises dynamically. Integrating automated scraping tools such as Apify with batch-processing or streaming architectures could enable longitudinal sentiment analysis and real-time surveillance of public reactions to emerging outbreaks, policy changes, vaccination campaigns, or misinformation events. Such systems would further strengthen applications in digital epidemiology, infodemiology, and public health surveillance.

Fourth, further experimentation with advanced transformer architectures and ensemble learning approaches may improve classification robustness and predictive performance. Recent advances in models such as RoBERTa, DeBERTa, and domain-adapted transformer variants have demonstrated strong performance across a wide range of NLP tasks. Ensemble modelling, which combines the strengths of multiple models, may improve resilience to noisy and ambiguous social media text, enhance the detection of subtle sentiment patterns, and improve model stability.

Finally, future work should extend beyond polarity-based sentiment classification toward richer semantic and discourse-level analysis. Social media discussions during health crises often involve multidimensional themes such as misinformation, stigma, fear, vaccine hesitancy, and policy debates, all of which may influence behavioural responses and policy compliance [4], [10]. Integrating sentiment analysis with complementary techniques such as topic modelling, stance detection, misinformation detection, and behavioural intent analysis may provide deeper insights into how public perceptions of emerging infectious diseases evolve and how social media shapes public understanding of health risks.

Collectively, these future directions may contribute toward the development of more scalable, multilingual, and intelligent epidemic intelligence systems capable of supporting evidence-based public health communication and crisis response.

7. CONCLUSION

This study evaluated the effectiveness of transformer-based transfer learning for sentiment analysis of Mpox-related Twitter discourse. Using the ELECTRA architecture within the AutoGluon framework, the proposed system classified tweets into positive, neutral, and negative sentiment categories, achieving an overall accuracy of 83.92% and balanced macro-averaged precision, recall, and F1-score. These findings suggest that transformer-based contextual language models are effective in modelling the linguistic variability, contextual ambiguity, and informal characteristics of social media text in outbreak-related discourse.

The comparative analysis indicates that contextual transformer models provide a more robust mechanism for sentiment classification than traditional lexicon-based and classical machine learning approaches, particularly in contexts characterised by informal language, implicit sentiment, and evolving discourse patterns. By leveraging contextual embeddings and transfer learning, the proposed approach demonstrates the practical value of modern NLP techniques in analysing large-scale unstructured social media data during public health events.

Beyond predictive performance, this study contributes a practical decision-support artefact by developing an interactive Streamlit dashboard for sentiment monitoring and trend visualisation. This demonstrates the feasibility of translating transformer-based sentiment models into usable tools for researchers, policymakers, and public health institutions seeking timely insights into public perceptions, communication challenges, and behavioural responses during infectious disease outbreaks.

More broadly, this study contributes to the growing fields of digital epidemiology, infodemiology, and public health informatics by illustrating how social media analytics can support outbreak surveillance and public health communication. As online platforms continue to shape the spread of information, misinformation, and public sentiment during health crises, analysing these digital traces may provide valuable situational awareness for evidence-based decision-making.

This study contributes to the growing fields of **digital epidemiology**, **infodemiology**, and **public health informatics** by demonstrating how transformer-based sentiment analysis can support outbreak surveillance and public health communication. As social media platforms continue to shape the spread of information, misinformation, and public sentiment during health crises, analysing these digital traces may provide valuable situational awareness for evidence-based decision-making. The findings of this study provide a practical and

methodological foundation upon which more scalable, multilingual, and real-time epidemic intelligence systems can be developed.

ACKNOWLEDGMENTS

We would like to acknowledge the reviewers for their time in reviewing this manuscript.

Ethics Considerations

The study analysed publicly available Twitter data collected through automated scraping tools. Although the data are publicly available, ethical considerations regarding user privacy and contextual integrity were taken into account by avoiding the reporting of identifiable user information. The study therefore, complied with ethical guidelines for social media research that permit the analysis of publicly accessible user-generated content for academic purposes. Nevertheless, the findings should be interpreted with awareness of potential biases inherent in social media data, including demographic representation and platform-specific communication patterns.

Data Availability

The dataset used in this study was collected using Apify and is available upon reasonable request. In addition to providing tweet identifiers, the authors can, upon request, provide the pre-processing scripts, training configuration parameters, and model evaluation procedures to support the reproducibility of the experimental workflow.

Funding

No funding received

REFERENSI

- [1] C. Van Dijk *et al.*, “Emergence of mpox in the post-smallpox era—a narrative review on mpox epidemiology,” *Clin. Microbiol. Infect.*, vol. 29, no. 12, pp. 1487–1492, 2023, doi: 10.1016/j.cmi.2023.08.008.
- [2] A. Ganesan *et al.*, “Mpox treatment evolution: past milestones, present advances, and future directions,” *Naunyn. Schmiedeberg's Arch. Pharmacol.*, vol. 398, no. 2, pp. 1057–1080, 2025, doi: 10.1007/s00210-024-03385-0.
- [3] K. K. Mohbey, G. Meena, S. Kumar, and K. Lokesh, “A CNN-LSTM-Based Hybrid Deep Learning Approach for Sentiment Analysis on Monkeypox Tweets,” *New Gener. Comput.*, vol. 42, no. 1, pp. 89–107, 2024, doi: 10.1007/s00354-023-00227-0.
- [4] R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, “Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach,” *IEEE Access*, vol. 11, no. May, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.
- [5] T. Alqurashi, “Arabic Sentiment Analysis for Twitter Data: A Systematic Literature Review,” *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 2, pp. 10292–10300, 2023, doi: 10.48084/etasr.5662.
- [6] Y. Wang, “Sentiment Analysis in Analyzing Monkeypox-Related Tweets Based on Deep Learning,” in *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence*, 2024, no. Emti, pp. 215–220, doi: 10.5220/0012923200004508.
- [7] A. Yazdani, M. Shamlou, M. Khaki, and A. Nahvijou, “Use of sentiment analysis for capturing hospitalized cancer patients’ experience from free-text comments in the Persian language,” *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, pp. 1–14, 2023, doi: 10.1186/s12911-023-02358-2.
- [8] N. Thakur, “MonkeyPox2022Tweets: A Large-Scale Twitter Dataset on the 2022 Monkeypox Outbreak, Findings from Analysis of Tweets, and Open Research Questions,” *Infect. Dis. Rep.*, vol. 14, no. 6, pp. 855–883, 2022, doi: 10.3390/idr14060087.
- [9] K. Jahanbin, M. Jokar, and V. Rahmadian, “Using twitter and web news mining to predict the monkeypox outbreak,” *Asian Pac. J. Trop. Med.*, vol. 15, no. 5, pp. 236–238, 2022, doi: 10.4103/1995-7645.346083.
- [10] N. Thakur, “Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox,” *Big Data Cogn. Comput.*, vol. 7, no. 2, 2023, doi: 10.3390/bdcc7020116.

- [11] F. M. Dandure and B. Ndlovu, "Federated Learning for Privacy-Preserving Sentiment Analysis in Distributed Electronic Health Record Environments : A Systematic Literature Review," vol. 8, no. 2, pp. 1812–1842, 2026, doi: 10.63158/journalisi.v8i2.1546.
- [12] J. Mantik et al., "Analysis of twitter user sentiment on the monkeypox virus issue using the nrc lexicon," J. Mantik, vol. 6, no. 4, pp. 2685–4236, 2023.
- [13] C. Mhlanga and B. Ndlovu, "Sentiment Analysis in Electronic Health Records for Patient-Centric Care : A Systematic Literature Review of Methods , Applications , and Challenges," vol. 8, no. 2, pp. 1740–1775, 2026, doi: 10.63158/journalisi.v8i2.1554.
- [14] H. Dia and N. Pettersson, "Evaluating the Accuracy of Sentiment Analysis Models when Applied to Social Media Texts," vol. TRITA-EECS, no. 2024:0000. KTH Royal Institute of Technology, 2024.
- [15] P. Rajkhowa et al., "Factors Influencing Monkeypox Vaccination: A Cue to Policy Implementation," J. Epidemiol. Glob. Health, vol. 13, no. 2, pp. 226–238, 2023, doi: 10.1007/s44197-023-00100-9.
- [16] K. K. Agustiningsih, E. Utami, and H. Al Fatta, "Sentiment Analysis of COVID-19 Vaccine on Twitter Social Media: Systematic Literature Review," Proc. - 2021 IEEE 5th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. Appl. Data Sci. Artif. Intell. Technol. Glob. Challenges Dur. Pandemic Era, ICITISEE 2021, pp. 121–126, 2021, doi: 10.1109/ICITISEE53823.2021.9655960.
- [17] J. Machemedze and B. Ndlovu, "Transformer-Based Models for Electronic Health Records and Omics in Healthcare : A Systematic Literature Review," J. Appl. Informatics Comput., vol. 10, no. 1, pp. 90–105, 2026, doi: 10.30871/jaic.v10i1.11893.
- [18] O. T. Chikumo and B. Ndlovu, "Transformer-based Models for Cardiovascular Disease Predictions from Electronic Health Records : A Systematic Review," J. Appl. Informatics Comput., vol. 10, no. 1, 2026, doi: 10.30871/jaic.v10i1.11899.
- [19] J. Na et al., "Sentiment analysis of online reviews of energy-saving products based on transfer learning and LBBA model," J. Environ. Manage., vol. 360, no. May, p. 121083, 2024, doi: 10.1016/j.jenvman.2024.121083.
- [20] A. Ghorbanali, M. K. Sohrabi, and F. Yaghmaee, "Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks," Inf. Process. Manag., vol. 59, no. 3, p. 102929, 2022, doi: 10.1016/j.ipm.2022.102929.
- [21] A. G. A. and V. V., "Sentiment analysis on a low-resource language dataset using multimodal representation learning and cross-lingual transfer learning," Appl. Soft Comput., vol. 157, no. April 2023, p. 111553, 2024, doi: 10.1016/j.asoc.2024.111553.
- [22] A. De Arriba, M. Oriol, and X. Franch, "Applying Transfer Learning to Sentiment Analysis in Social Media," Proc. IEEE Int. Conf. Requir. Eng., vol. 2021-September, pp. 342–348, 2021, doi: 10.1109/REW53955.2021.00060.
- [23] H. Chamboko, B. Ndlovu, and K. Maguraushe, "Twitter (X) Sentiment Analysis on Monkeypox : A Systematic Literature Review," Int. J. Informatics Dev., vol. 14, no. 2, pp. 629–639, 2025, doi: 10.14421/ijid.2025.5196.
- [24] Z. Revesai, K. Maguraushe, B. Ndlovu, and S. Dube, "Predictive Modeling and Risk Assessment of Monkeypox Transmission Using Bayesian Networks : An Interpretable Machine Learning Approach," Bhateja, V., Biju, S.M., Udgata, S.K. Inf. Syst. Des. AI ML Appl. ISDIA 2025. Lect. Notes Networks Syst., vol. 1565, pp. 281–296, 2026, doi: 10.1007/978-981-95-0375-9_23.
- [25] A. Khumaidi, "Data Mining for Predicting the Amount of Coffee Production Using Crisp-Dm Method," J. Techno Nusa Mandiri, vol. 17, no. 1, pp. 1–8, 2020, doi: 10.33480/techno.v17i1.1240.
- [26] Y. A. Singgalen, "Analyzing an Interest in GPT 4o through Sentiment Analysis using CRISP-DM," J. Inf. Syst. Informatics, vol. 6, no. 2, pp. 882–898, 2024, doi: 10.51519/journalisi.v6i2.740.
- [27] N. Omojunikanbi, "Public relations and effective communication during a global health crisis: Combating disinformation, misinformation, and fake news on Covid-19," J. Commun. Media Res., vol. 14, no. 1, pp. 64–71, 2022.