

Article

Detecting Shifts in Public Discourse from Offline to Online Using Deep Learning

Adamu Abubakar Ibrahim * and Fazeel Ahmed Khan

Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia;
fazeel.ahmed@live.iium.edu.my

* Correspondence: adamu@iium.edu.my

Abstract

Increasingly, discussions that once took place in social environments are transitioning to digital platforms. The role of news media is significant in shaping and enhancing discussions around many topics. This study argues that health-related topics in public discourse, transitioning from offline to online, necessitate rigorous validation. That is why this study proposed the application of deep learning techniques to the boundaries and deviation of accuracies in health-related topics by analyzing health-related tweets from major news outlets such as BBC, CNN, CBC, and Reuters. The study developed LSTM and CNN classifiers to categorize content pertinent to the discourse following the formal deep learning process and employed a sequence of VAEs to verify the learnability and stability of the classifiers. The LSTM demonstrated superior performance compared to CNN, attaining validation accuracies of 98.4% on BBC and CNN, 97.8% on CBC, and 97.3% on Reuters. The optimal configuration of our LSTM achieved a precision of 98.69%, a recall of 98.20%, and an F1-score of 97.90% and recorded the lowest false positive rate, at 1.30%. This provided us with the optimal overall equilibrium for operational oversight. The VAE runs demonstrated that the model exhibited stability and the ability to generalize across different sources, achieving approximately 99.6% for Reuters and around 98.4% for BBC. The findings confirm that deep learning models are capable of reliably tracking the online migration of health discourse driven by news media. This provides a solid foundation for near-real-time monitoring of public engagement and for informing sustainable healthcare recommendation systems.

Keywords: public discourse; news media; social media; health communication; deep learning

Academic Editors: Patrick Siarry and
Junaid Rashid

Received: 17 September 2025

Revised: 6 October 2025

Accepted: 10 October 2025

Published: 11 October 2025

Citation: Abubakar Ibrahim, A.; Khan, F.A. Detecting Shifts in Public Discourse from Offline to Online Using Deep Learning. *Electronics* **2025**, *14*, 3987. <https://doi.org/10.3390/electronics14203987>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Social media has significantly transformed the way individuals communicate with one another. This platform enables individuals to swiftly express their opinions on various subjects, including health, which remains a crucial aspect of human well-being [1]. The emergence of infodemiology reflects this transformation. The vast array of information available on social media is utilized to systematically examine human behavior and social patterns [2]. The findings generated are diverse and noteworthy [3]. For example, earlier studies show that social media has the capability to track positive health behaviors and readiness for vaccinations [4,5], and it can also forecast public health events, like infectious disease outbreaks, more rapidly than traditional surveillance methods. The

dynamics of online health communication represent a crucial field of study, especially as conversations on platforms such as “X” (formerly Twitter) influence public perceptions of healthcare services [6]. Although tweets have been widely employed to discern disease patterns, assess risks, and aid in disaster management, the significant volume of health-related content produced by media outlets has not received adequate academic attention.

A notable challenge arises from the abundance of tweet data provided by major news organizations such as BBC, CNN, and Reuters, which remains underutilized in producing meaningful insights. What prevents the integration of these covertly collected digital traces to enhance a Healthcare Recommendation System (HRS)? Additionally, to what extent can health-focused news tweets from leading sources be transformed into positive outcomes that enhance HRS design or promote advancements in the broader healthcare industry? This study investigates whether these tweets, alongside their role in spreading health information, might impact public trust in medical institutions by shaping perceptions of healthcare quality. If validated, health-news tweets may serve as a crucial element in the creation of recommendation systems designed to capture online public discourse.

The rationale for this study is strengthened by the growing evidence that tweets serve purposes that extend beyond simple personal communication. They are being used more frequently for analysis in the context of public health emergencies, pandemics, and crises [7–14]. Preliminary findings suggested that tweets supported early warning systems throughout the 2009 swine flu epidemic [7], facilitated COVID-19 monitoring, and improved situational awareness during the Mpox outbreak [8]. Research indicates that these tools can be utilized for sentiment analysis [9], disseminating information [10], and facilitating crisis communication by governmental bodies [11]. Furthermore, it has been shown that social media plays a crucial role in managing uncertainty and shaping policy learning in the context of health crises [12]. Tweets have been employed to detect sudden illness occurrences before clinical reporting [13] and to gain insights into the formulation of health policies [14]. The findings from these studies collectively demonstrate how tweets can significantly improve public health management.

Despite these advancements, challenges remain to be addressed. The volume of health-related tweets is increasing rapidly [12], yet there remains a significant gap in our understanding of how to extract system requirements from this unstructured data while accommodating the diverse needs of all stakeholders in the design of healthcare recommendation systems [15,16]. Addressing this gap in understanding is crucial to ensure that HRS development aligns with stakeholder needs and reflects the current landscape of public discourse. Simultaneously, earlier studies have raised concerns that employing deep learning for healthcare recommendations might compromise the privacy and integrity of sensitive medical data [17,18]. This informs our strategy: to focus not on private clinical information but on publicly available health-related tweets that demonstrate the shift in news media from offline to online platforms in influencing public conversation.

This research presents three significant contributions.

- Initially, it introduces an innovative framework that integrates deep learning methodologies with the topical synthesis of health-related tweets to contribute to the development of sustainable healthcare ecosystems. The framework consolidates information from four prominent news organizations: BBC, CNN, CBC, and Reuters. This illustrates the transition of health communication to the internet. The analysis outlines a structured framework of healthcare tweet ecosystems that links model performance with a core principle for integrating news-focused health discussions into healthcare recommendation systems. Third, it categorizes the synthesis enabled by deep learning into fundamental elements of healthcare communication, offering a framework to ensure sustainability in the development of recommendations.

- This study outlines healthcare tweets, framed them within the context “discussion”, and illustrates a cohesive ecosystem transcending the wider sociopolitical realm of healthcare delivery. By concentrating its efforts, the study analyzes the most crucial elements of health rhetoric in news-driven online discussions and identifies the shift in health-related public discourse from offline contexts to online platforms, utilizing deep learning as a methodological bridge between extensive news datasets.

The remaining sections of the paper are organized as follows: Section 2 presents the literature review of the study. Section 3 outlines the methodology employed in the study. Section 4 presents the experimental analysis along with the results obtained. Section 5 presents the discussion of the study. Section 6 presents the conclusion of the study, followed by a comprehensive list of references utilized throughout the research.

2. Related Work

A variety of studies on Public Discourse indicate that approximately 13% of content on Twitter/X pertains to health [19]. This data often includes information about individual health conditions, the sharing of medical updates, and narratives of both positive and negative health experiences. Twitter, Facebook, and Instagram are widely recognized platforms that provide excellent opportunities for examining health communication. Additionally, there exist communities tailored to specific illnesses, such as PatientsLikeMe and CaringBridge. Twitter/X has proven to be particularly effective in identifying health-related terms [7] and in evaluating its functionality as an early warning and risk communication tool during crises such as the 2009 swine flu epidemic. Through content analysis, it was demonstrated that tweets facilitated the identification of new trends, enabled the rapid dissemination of risk information, and provided insights into public reactions. The results indicate that Twitter/X serves as a crucial platform for disseminating timely alerts, monitoring health issues, and facilitating communication with large audiences simultaneously. Simultaneously, investigations indicated that misinformation poses a significant issue and emphasized the necessity for effective filtering and verification systems in the context of health emergencies.

The volume of health-related tweets data has expanded to several terabytes and continues to increase [12]. The growing volume has been analyzed concerning risk assessment and public health management, emphasizing the manifestation of public uncertainty in online discussions. The integration of quantitative analysis of tweets with qualitative interviews of health authorities has revealed that social media provides valuable real-time insights for risk assessment. Nevertheless, the extensive amount of data and the widespread occurrence of inaccurate information remain significant challenges. In a comparable manner, Ref. [13] demonstrated that machine learning can be utilized to identify acute illness episodes on Twitter/X by detecting unusual spikes in tweets that frequently occurred prior to official reports. The findings indicate that social media may serve as a valuable enhancement to conventional surveillance systems, functioning effectively as an early warning mechanism.

Public health emergencies generate complex online discussions that intertwine peer support, informal counseling, public service announcements, activism, and misinformation. Investigations such as [8] examined the emotional responses and feelings of individuals regarding COVID-19 and Mpox on Twitter/X, whereas [9] analyzed the general sentiments surrounding COVID-19 to identify the predominant themes in the discourse. Several studies employed theoretical frameworks such as the Social Amplification of Risk Framework and the Issue-Attention Cycle to examine the dissemination of pandemic information [10]. In parallel, Ref. [11] assessed governmental communication strategies during the early phases of the COVID-19 crisis, while [14] investigated the function of Twit-

ter/X in promoting policy learning in times of emergencies. The findings from these studies highlight that social media platforms offer significant avenues for observing public discussions, while also exposing weaknesses to misinformation that impede effective health communication strategies.

Concurrently, there has been an increase in interest regarding healthcare recommendation systems (HRS). Despite the increasing adoption of these systems, numerous platforms continue to fall short of addressing the requirements of patient-centered healthcare. Ivanova and Raō [16] analyzed healthcare recommendation systems alongside those in other domains, like music, and found that while there are structural similarities, medical systems face unique challenges stemming from the complexity and sensitivity of health data. The findings suggest that insights from various disciplines can contribute to the development of more resilient and effective HRS. Expanding on this, Sahoo et al. [17] developed DeepReco, which employed deep learning and collaborative filtering to enhance accuracy and user satisfaction. This approach proved to be significantly more effective than conventional methods. Chinnasamy et al. [18] employed deep learning-based collaborative filtering to deliver precise and customized health interventions, demonstrating the effectiveness of sophisticated neural models in improving recommendation outcomes.

Recent advancements in deep learning, especially Transformer-based models and large language models, have significantly impacted health NLP tasks such as classification, summarization, clinical reasoning, and dialogue generation [19–21]. A growing body of research examines the ability of deep learning to detect or characterize health misinformation on social media, including thorough reviews and focused analyses of COVID-19 and vaccine-related content [22–25]. Research indicates that BERT-family models outperform traditional baselines in identifying health-related tweets [24]. Simultaneously, hybrid or sequence models, such as BERT-LSTM and GRU/LSTM variants, continue to emerge in social media pipelines [25–27]. Foundational syntheses outline the prevalence of misinformation and strategies for its mitigation, including media literacy, automated detection, and platform interventions, highlighting persistent challenges for practical implementation [22,26,28]. These studies collectively support our emphasis on Transformer architectures, careful data curation, and evaluation protocols tailored for health misinformation contexts.

The reviewed literature highlights the benefits and limitations of current methodologies. Earlier investigations into health-related social media have concentrated on aspects such as sentiment analysis, dissemination of information, and communication during outbreaks like swine flu, COVID-19, and Mpox [7–14]. These studies provide valuable insights for managing uncertainty and enhancing crisis response; however, they fall short in demonstrating how these insights can be systematically incorporated into sustainable healthcare systems. Similarly, studies on recommender systems highlight the promise of deep learning for enhancing personalization and efficiency but often overlook the incorporation of these systems into broader healthcare ecosystems.

As a result, this study positions itself at the intersection of these two areas of inquiry. This analysis examines health news tweets from major news organizations such as BBC, CNN, CBC, and Reuters, focusing on the role of news media in transitioning public discussions from offline environments to online platforms. The study presents a framework aimed at creating sustainable and responsive healthcare recommendation systems that capture real-time online engagement through the application of deep learning models for discourse detection and synthesis.

3. Materials and Method

This current study adopted a methodology behind the two deep learning architectural frameworks (LSTM and CNN). It includes data collection, pre-processing techniques, model development, and evaluation metrics.

3.1. Model Architecture

Machine learning models have been used to make recommendations, although they work best with structured datasets. Thus, most existing models rely heavily on structured inputs, making it difficult for them to understand unstructured content. Deep learning addresses this issue well [16]. Distributed representations allow deep neural architectures to outperform shallow and unsupervised approaches in organized and unstructured settings [17]. Recent advances show end-to-end deep learning frameworks as the best way to analyze unstructured data. Recurrent and convolutional neural networks perform well here. LSTM networks capture temporal and sequential relationships in text-based data well. CNNs are great at detecting local characteristics and contextual patterns [17]. This study analyses a huge number of BBC, CNN, CBC, and Reuters public discourse tweets using LSTM and CNN models. We want to understand how public debate changes from offline to online. The study helps create sustainable frameworks for real-time public involvement and news industry evolution.

3.1.1. Long Short-Term Memory (LSTM)

LSTM was introduced by Hochreiter and Schmidhuber in 1997. It tackled challenges associated with standard RNNs [29,30]. RNNs are capable of handling input vectors of different lengths; however, the issue of vanishing gradients poses challenges for maintaining long-term memory [31]. The LSTM incorporates a specialized memory cell that maintains its state until it is altered by input or output operations, addressing this issue effectively. The system is capable of monitoring long-term dependencies and preventing the loss of essential information. LSTMs excel over RNNs in addressing long-term context retention issues by ensuring stable system dynamics, thereby establishing a foundation for memory-integrated neural architectures. This study's architecture includes two LSTM layers along with an attention layer to enhance feature representation [32].

Similarly to recurrent models, the LSTM serves as a parametric mapping between input and target pairs. Nonetheless, its distinctive recurrent dynamics and gating mechanisms enable it to selectively determine which information to retain and modify over time. Weights and biases are adjusted in a systematic manner to minimize the difference between predicted outputs and desired patterns across multiple epochs [33]. Concealed units maintain enduring state representations, rendering LSTMs more effective for sequential data such as health-related tweets. In contrast to typical neural networks, which are unable to divide sequences into significant phases. LSTMs effectively model long-term dependencies, enabling them to capture temporal structures that traditional networks overlook. Each LSTM neuron is equipped with a forget gate, an input gate, and an output gate within its memory cell. These gates control the flow of information within cells. The gates regulate the retention of existing information, the incorporation of new data, and the generation of outputs for the subsequent layer [34].

The computation of the activation of the input gate y^{in} and the activation of the output gate y^{out} presented in Equation (1):

$$y^{\text{in}} \text{ \& } y^{\text{out}} \quad (1)$$

The gating mechanisms can be formalized as follows. The activations of the input gate y^{in} and output gate y^{out} are defined at discrete time steps $t = 1, 2, \dots$ using Equations (2) and (3):

$$net_{out_j}(t) = \sum_m w_{out_jm} y^m(t-1), y^{out_j}(t) = f_{out_j}(net_{out_j}(t)) \quad (2)$$

$$net_{in_j}(t) = \sum_m w_{in_jm} y^m(t-1), y^{in_j}(t) = f_{in_j}(net_{in_j}(t)) \quad (3)$$

where j denotes the memory block, f be the logistic sigmoid that falls within the interval $[0,1]$, and w_{lm} be the connection weight between unit m and unit l . The internal state of a memory cell $S_c(t)$, can be computed by adding the squashed gate input to the state at the most recent time step $S_c(t-1)$ when $(t > 0)$ using Equations (4) and (5):

$$net_{c_j^v}(t) = \sum_m w_{c_j^vm} y^m(t-1) \quad (4)$$

$$S_{c_j^v}(t) = S_{c_j^v}(t-1) + y^{in_j}(t)g(net_{c_j^v}(t)) \quad (5)$$

Given that c_j^v denotes cell v of memory block j , the cell input is squashed by g , and $S_{c_j^v}(0) = 0$ the output of a cell y^c is determined by squashing the internal state S_c using an output squashing function h and gating it with the activation of the output gate y^{out} , mathematically represented in Equation (6):

$$y^{c_j^v}(t) = y^{out_j}(t)h(S_{c_j^v}(t)) \quad (6)$$

Let h denote a centered sigmoid distributed in the interval $[-1,1]$. In a network with a layered architecture including a hidden layer with memory blocks and a standard input and output layer, the output unit's K can be precisely described using Equations (7) and (8):

$$net_k(t) = \sum_m w_{km} y^m(t-1) \quad (7)$$

$$y^k(t) = f_k(net_k(t)) \quad (8)$$

Through these mechanisms, the LSTM effectively addresses tasks involving long-range dependencies that standard RNNs cannot resolve [35]. Within the scope of this study, the LSTM is applied to sequential health-news tweet data to detect the shift in public discourse from offline venues to online platforms. Its ability to model long-term dependencies makes it an essential tool for analyzing large-scale news-driven discourse and for developing sustainable healthcare recommendation systems that reflect real-time public engagement.

3.1.2. Convolutional Neural Networks (CNNs)

CNNs are deep feed-forward networks that can link input data to output judgments [36]. Convolutional neural networks are famous for their capacity to recognize and classify complex patterns in noisy environments. This includes handwritten digit recognition, face identification, object classification, and image analysis. Convolutional neural networks replicate basic visual cortex neural activity and are used in computer vision [37]. Convolutional neural networks have pooling, convolutional, and fully connected layers. As data passes between these layers, it becomes more abstract. Adjusting filters and connection weights at every layer ensures that the decision output has a strong relationship between the input domain and the output classification domain [38].

The convolutional and pooling layers of many convolutional neural networks use nonlinear activation functions like hyperbolic tangent. Class posterior probabilities are then assigned by the softmax output layer. The class with the highest posterior probability

decides. These parameters connect x to the next layer, $x + 1$. Mathematical explanations can describe CNN layer processing [39]. A neuron at convolution layer l , feature map k , and position (x, y) functions as shown in Equation (9):

$$C_{x,y}^{(l,k)} = \tanh \left(\sum_{t=0}^{f-1} \sum_{r=0}^{K_h} \sum_{c=0}^{K_w} W_{(r,c)}^{(k,t)} C_{(x+r,x+c)}^{(l-1,t)} + Bias^{(i,k)} \right) \quad (9)$$

The output of a neuron at convolution layer l , feature pattern k , row x , and column y is denoted as $C_{x,y}^{(l,k)}$. The factor f denotes the number of convolution cores in a given feature data pattern. Equation (10) quantifies the output of a neuron at the l th subsampling layer, k th feature pattern, row x , and column y during the subsampling stage.

$$C_{x,y}^{(l,k)} = \tanh \left(W^{(k)} \sum_{r=0}^{S_h} \sum_{c=0}^{S_w} C_{(x \times S_h + r, y \times S_w + c)}^{(l-1,t)} + Bias^{(i,k)} \right) \quad (10)$$

The numerical output of neuron j at the l th hidden layer H is represented by Equation (11):

$$C_{(l,j)} = \tanh \left(\sum_{k=0}^{s-1} \sum_{x=0}^{S_h} \sum_{y=0}^{S_w} W_{(x,y)}^{(j,k)} C_{(x,y)}^{(l-1,t)} + Bias^{(i,j)} \right) \quad (11)$$

The variable s represents the quantity of feature patterns present in the subsampling layer. The expression for the output of neuron i at the l th output layer at the output layer is derived from Equation (12):

$$C_{(l,i)} = \tanh \left(\sum_{j=0}^H C_{(l-1,j)} W_{(i,j)}^l + Bias^{(i,j)} \right) \quad (12)$$

These formulations demonstrate how CNNs progressively capture local and global features from the input space. Each convolutional layer produces multiple feature maps, which serve as inputs to subsequent layers, enabling the network to identify increasingly complex data structures. In the present study, this property is harnessed to classify discourse-relevant patterns in health-news tweets, complementing the sequential learning capacity of LSTM. While LSTM is suited to modeling long-term dependencies, CNN contributes by efficiently capturing local semantic patterns that characterize online discourse triggered by news media. Together, they form a robust pipeline for detecting the shift in public discourse from offline venues to online platforms.

3.2. Dataset

This study used four datasets of news tweets—BBC Health Tweets, CNN Health Tweets, CBC Health Tweets, and Reuters Health Tweets [40,41]—to examine public discourse movements from offline to online platforms. Each dataset contains thousands of tweets sorted by TweetID, Timestamp, and Content, making them useful for model training and testing.

BBC Health posts. This dataset includes 3929 tweets with extensive topical labels. This helps identify and categorize tweets by topic, such as health-related themes like symptoms or illness phases. This dataset allows for more in-depth healthcare recommendation system study than binary or unlabeled corpora. **CBC Health posts.** Similarly to the BBC style, this dataset has 3741 tweets. This tool allows classification experiments. It was originally developed to classify baseline nodes but now identifies HRS-relevant healthcare discourse patterns. **CNN Health posts.** This dataset of 4061 tweets analyzes how Twitter users share health information and resolve privacy issues. This project seeks

extractable content for HRS-integrated online health discourse modeling. Health updates from Reuters. This dataset includes 4719 health news tweets tagged consistently. Many tweets lack health entities, but judicious annotation can extract relevant features. This strengthens the idea that health-related tweets indicate online discourse shifts.

These datasets span a wide range of news-related health communication, allowing LSTM and CNN models to detect public discourse trends and improve sustainable healthcare recommendation systems.

3.3. Experimental Setting

The experiments were run on Google Colab with Python 3.12.11, an Intel® Core™ i7-10870H processor, 16 GB RAM, and Windows 11. NumPy, seaborn, matplotlib, and pandas are notable libraries. NLTK was used for stopwords, tokenization, and lemmatization; regular expressions (re) were used for pattern matching; TensorFlow and Keras were used for deep learning models, using Tokenizer and pad_sequences for text vectorization; and PyTorch and its submodules for neural networks were used for data handling. Scikit-learn's CountVectorizer was used for bag-of-words representation and Latent Dirichlet Allocation for topic modeling; train_test_split; WordCloud for word clouds; and NLTK Data for stopwords and punctuation data.

The experiment configures a Variational Autoencoder (VAE) with an LSTM-based encoder and decoder. A dense vector representation of integer token sequences is created by an embedding layer in the model. A second model version uses Gated Recurrent Units (GRUs) instead of LSTMs. GRU-VAE shares its basic architecture with VAE. A GRU encoder compresses the input into latent parameters, which provide a latent vector z for a GRU decoder to recover the input. Another, more advanced approach uses Transformer architecture. This variant uses self-attention instead of recurrent encoders and decoders. A single loss function is used in all VAE models. The total loss has reconstruction and regularization components. Reconstruction loss is calculated using cross-entropy. The pre-processed text from each dataset (BBC, CNN, CBC, Reuters) is turned into numerical representations for models. Each dataset uses a different tokenizer to adjust vocabulary. This links words to numbers. Text is then turned into numerical sequences. In batch processing, input sequences must be uniformly long. To achieve this, shorter sequences are stretched with zeros while longer ones are terminated. This step is essential for neural network models to use text data.

3.4. Model Parameter and Training Procedure

Utilizing various datasets enhances the training and testing processes for deep learning models. The hyperparameters utilized are "embedding_dim = 100", "hidden_dim = 256", "latent_dim = 50", "epochs = 15", "batch_size = 32", "learning_rate = 1×10^{-3} ".

The embedding_dim of 100 means each word is mapped to a dense vector in a 100-dimensional space, allowing the model to capture semantic meaning. The hidden_dim of 256 specifies the size of the LSTM's hidden state, dictating its capacity to remember long-range dependencies and complex patterns in the text sequences. The latent_dim of 50 is a key feature of a VAE, defining the size of the compressed, probabilistic bottleneck layer (the latent space) where the model learns a compact representation of the input data's core features. The training regimen is set to run for 15 epochs with a batch_size of 32, meaning the model's weights will be updated 15 times using groups of 32 articles. The learning_rate of 1×10^{-3} (0.001) is a standard value for the Adam optimizer, controlling the step size during gradient descent to ensure stable and effective learning.

3.5. Data Splitting Strategy

The data is strategically partitioned to ensure rigorous evaluation. The dataset is first split, with 70% allocated for training (X_{train} , y_{train}) and 30% held out in a temporary set. This temporary set is then divided equally, creating a validation set (X_{val} , y_{val}) and a test set (X_{test} , y_{test}), each comprising 15% of the original data. The validation set will be used during training to monitor performance and guide decisions like early stopping, while the test set is reserved for a final, unbiased evaluation after training is complete. The `random_state = 42` parameter ensures these splits are reproducible, guaranteeing consistent results across multiple runs.

3.6. Evaluation of Performance

To evaluate the effectiveness of the proposed deep learning models in detecting shifts in public discourse from offline venues to online platforms, we adopted four standard performance metrics derived from the confusion matrix comprising true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [42]. These metrics—Precision, Recall, F1-score, and Accuracy (Acc)—provide a comprehensive assessment of the classification outcomes across the health-news tweet datasets.

Precision measures the proportion of correctly predicted positive cases relative to all predicted positives, indicating the reliability of the model's positive classifications (Equation (13)).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (13)$$

Recall evaluates the proportion of true positives identified out of all actual positives, reflecting the model's sensitivity in detecting discourse-relevant tweets (Equation (14)).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (14)$$

The False Positive Rate (FPR) quantifies the proportion of negative cases that were incorrectly classified as positives, providing an error-based perspective of the model's predictions (Equation (15)).

$$\text{FPR} = \frac{FP}{FP + TN} \quad (15)$$

The F1-score, as the harmonic mean of Precision and Recall, balances both sensitivity and specificity to assess overall classification robustness (Equation (16)).

$$F_1\text{-score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (16)$$

Finally, Accuracy (Acc) measures the proportion of all correctly classified instances (both positive and negative) relative to the total predictions, reflecting the model's overall effectiveness (Equation (17)):

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

4. Experimental Analysis and Presentation of the Result

The experimental analysis was conducted systematically in several phases to ensure reproducibility and alignment with the study objective of detecting shifts in public discourse from offline to online platforms.

4.1. Dataset Extraction

The summary of the extract from the dataset is presented in Table 1: The entire content of the four different health-related text files datasets from BBC, CBC, CNN, and Reuters were read into separate lists. Each list holds all the lines from its respective file, allowing for further processing or analysis of the text data. Each line of the BBC, CBC, CNN, and Reuters Health data is divided into individual fields using the pipe (|) separator unit. After extracting the tweet ID, timestamp, and content, the structured data is stored in the parsed data list. Each element in the parsed data dataset is a list that includes three bits of information for every tweet.

Table 1. Samples of The Datasets Extracted from The Four Datasets.

BBC					
TweetID		Timestamp		Content	
0	585978391360221184	Thu Apr 09 01:31:50 +0000 2015	0	Breast cancer risk test devised http://bbc.in/...	
1	585947808772960257	Wed Apr 08 23:30:18 +0000 2015	1	GP workload harming care—BMA poll http://bbc...	
2	585947807816650752	Wed Apr 08 23:30:18 +0000 2015	2	Short people's 'heart risk greater' http://bbc...	
3	585866060991078401	Wed Apr 08 18:05:28 +0000 2015	3	New approach against HIV 'promising' http://bb...	
4	585794106170839041	Wed Apr 08 13:19:33 +0000 2015	4	Coalition 'undermined NHS'—doctors http://bb...	
...
3924	384766023120871424	Mon Sep 30 19:45:43 +0000 2013	3924	Baby born after ovaries 'reawakened' http://bb...	
3925	384700230920175617	Mon Sep 30 15:24:17 +0000 2013	3925	Identical triplets born against odds http://bb...	
3926	384678543088562178	Mon Sep 30 13:58:06 +0000 2013	3926	Hospital failed to make improvements http://bb...	
3927	384678542455222273	Mon Sep 30 13:58:06 +0000 2013	3927	New patient targets pledge for NHS http://bbc....	
3928	384569546108964864	Mon Sep 30 06:44:59 +0000 2013	3928	C. diff 'manslaughter' inquiry call http://bbc...	
CBC					
TweetID		Timestamp		Content	
0	586266687948881921	Thu Apr 09 20:37:25 +0000 2015	0	Drugs need careful monitoring for expiry dates...	
1	586266687017771008	Thu Apr 09 20:37:25 +0000 2015	1	Sabra hummus recalled in U.S. http://www.cbc.c...	
2	586266685495214080	Thu Apr 09 20:37:24 +0000 2015	2	U.S. sperm bank sued by Canadian couple didn't...	
3	586226316820623360	Thu Apr 09 17:57:00 +0000 2015	3	Manitoba pharmacists want clampdown on Tylenol...	
4	586164344452354048	Thu Apr 09 13:50:44 +0000 2015	4	Mom of 7 'spooked' by vaccinations reverses st...	
...
3736	324528904306057216	Wed Apr 17 14:24:54 +0000 2013	3736	Rural doctors mentor medical students in U of...	
3737	324503137597665282	Wed Apr 17 12:42:31 +0000 2013	3737	More men suffering from eating disorders, says...	
3738	324261559025160193	Tue Apr 16 20:42:34 +0000 2013	3738	5 sources of objective drug information sugges...	
3739	324251178919657474	Tue Apr 16 20:01:20 +0000 2013	3739	Top five ways to avoid allergies this spring h...	
3740	324205113348616192	Tue Apr 16 16:58:17 +0000 2013	3740	Health Canada to stop sales of small magnets h...	
CNN					
TweetID		Timestamp		Content	
0	576880531301801984	Sat Mar 14 23:00:11 +0000 2015	0	An abundance of online info can turn us into e...	
1	576820122666471424	Sat Mar 14 19:00:08 +0000 2015	1	A plant-based diet that incorporates fish may...	
2	576744652717461504	Sat Mar 14 14:00:15 +0000 2015	2	It doesn't take much to damage your hearing at...	
3	576736754436304896	Sat Mar 14 13:28:52 +0000 2015	3	RT @CNN: Forever young? Discover this island's...	
4	576736614766010368	Sat Mar 14 13:28:18 +0000 2015	4	RT @CNN: Is post-traumatic stress disorder in...	
...
4056	239699936671854593	Sun Aug 26 12:24:52 +0000 2012	4056	RT @EverydayHealth: Want killer abs? @JillianM...	
4057	239504620710420480	Sat Aug 25 23:28:46 +0000 2012	4057	Medicare at stake—@sanjayguptaCNN talks abo...	
4058	239410205757145088	Sat Aug 25 17:13:35 +0000 2012	4058	Ann Romney talks about her experience with MS...	
4059	239386320416428032	Sat Aug 25 15:38:41 +0000 2012	4059	Make sure your first marathon isn't your last!...	
4060	239366825018806272	Sat Aug 25 14:21:12 +0000 2012	4060	Robin Roberts' cancer diagnosis http://at.cnn....	
Reuters					
TweetID		Timestamp		Content	
0	586282503981375488	Thu Apr 09 21:40:16 +0000 2015	0	Los Angeles closes 500 medical marijuana shops...	
1	586278524748750848	Thu Apr 09 21:24:27 +0000 2015	1	U.S. cuts poultry export forecast as deadly bi...	
2	586273441801699328	Thu Apr 09 21:04:15 +0000 2015	2	Fears over Roundup herbicide residues prompt p...	

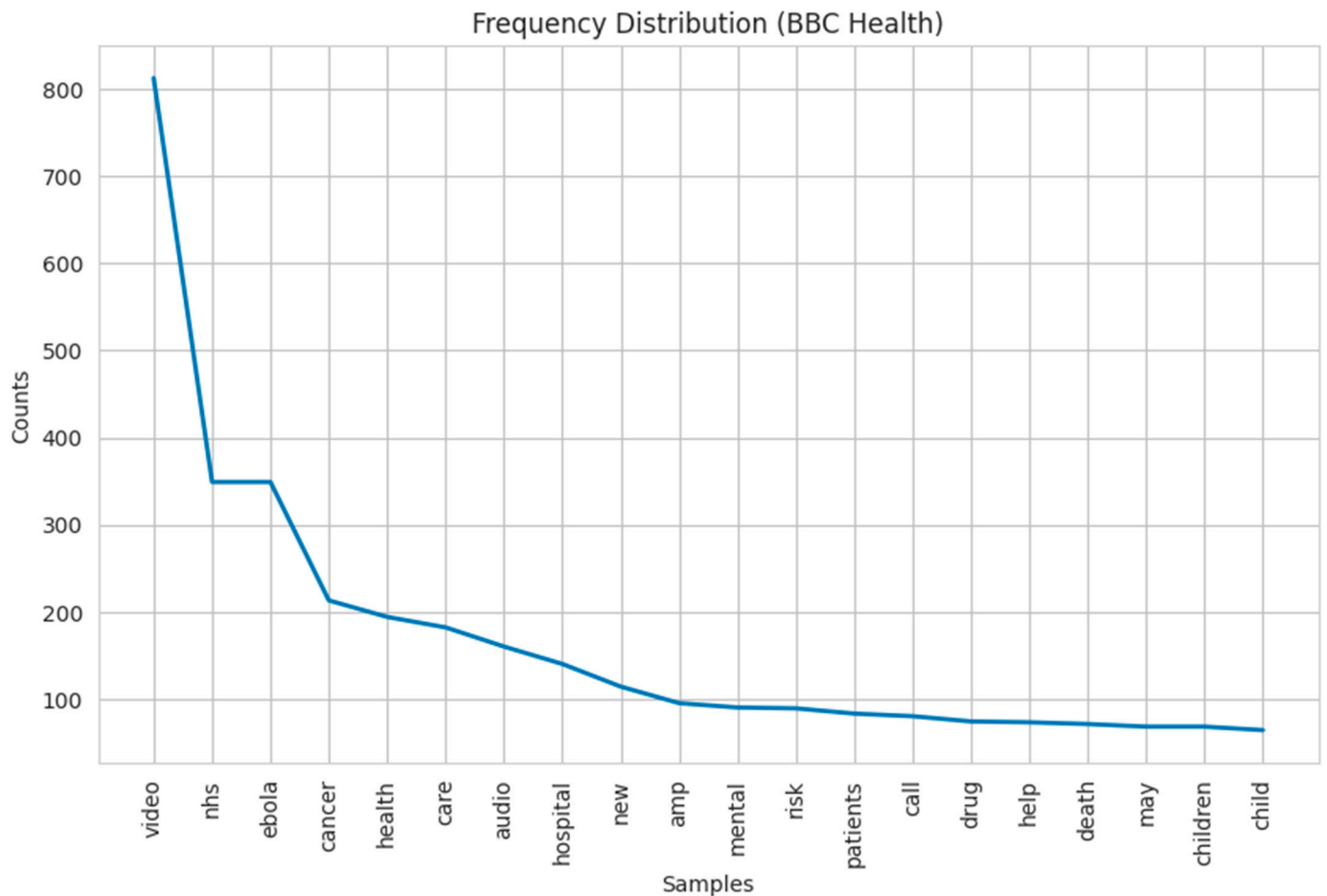


Figure 2. Major Topics Extracted from BBC Dataset.

An analysis of the CBC dataset revealed that the primary emphasis is on infectious diseases and Canada's response to public health challenges (see Figure 3). The discussion subsequently encompasses subjects pertaining to healthcare services, medical procedures, and the healthcare infrastructure within Canada. Ultimately, there are subjects that concentrate on maternal and child health, women's health, and medical research that could prove beneficial. The aforementioned subjects serve as the primary focal points of discussion within the CBC Health dataset, aiding in our comprehension of the key issues addressed in the tweets. This topic centers on infectious diseases, especially those associated with viruses like influenza and Ebola. The terminology such as "vaccine," "outbreak," and "case" suggests that discussions in Canada are centered around public health matters, particularly in relation to disease prevention and treatment.

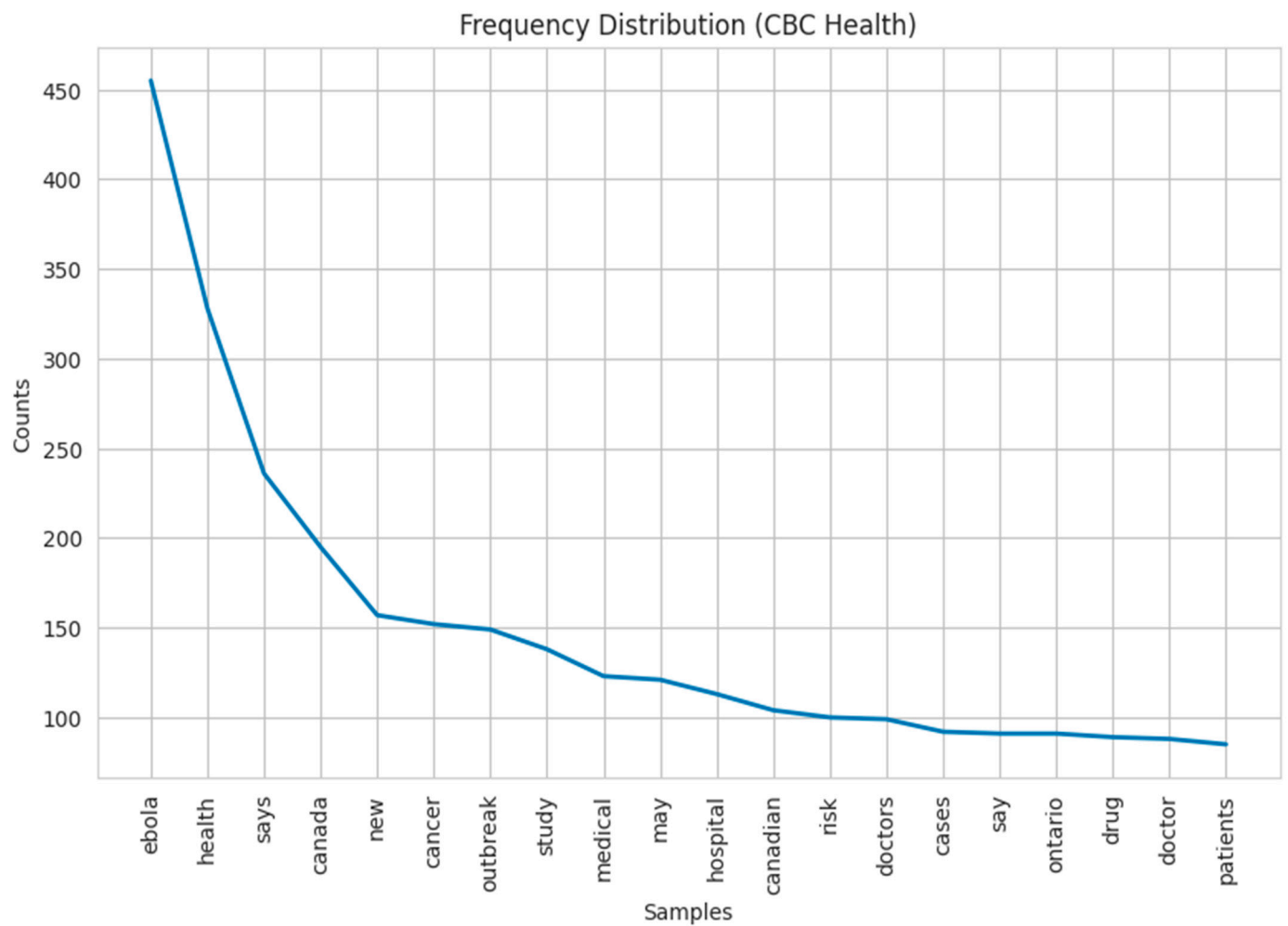


Figure 3. Major Topics Extracted from CBC Dataset.

An analysis of the CNN dataset revealed a notable emphasis on general health research, studies, and lifestyle recommendations, with CNN branding prominently displayed (see Figure 4). The dataset encompasses topics related to weight loss, fitness, and health maintenance, frequently featuring insights from Dr. Sanjay Gupta and health tips from CNN. The analysis revealed that the dataset primarily focuses on medical disorders and health issues impacting women, children, and specific conditions such as cancer and heart disease. The topic of weight loss holds significant importance in the realm of health and fitness. Terms such as “weightloss,” “food,” “tip,” and “getfit” indicate that the emphasis should be on guidance related to nutrition, physical activity, and overall wellness.

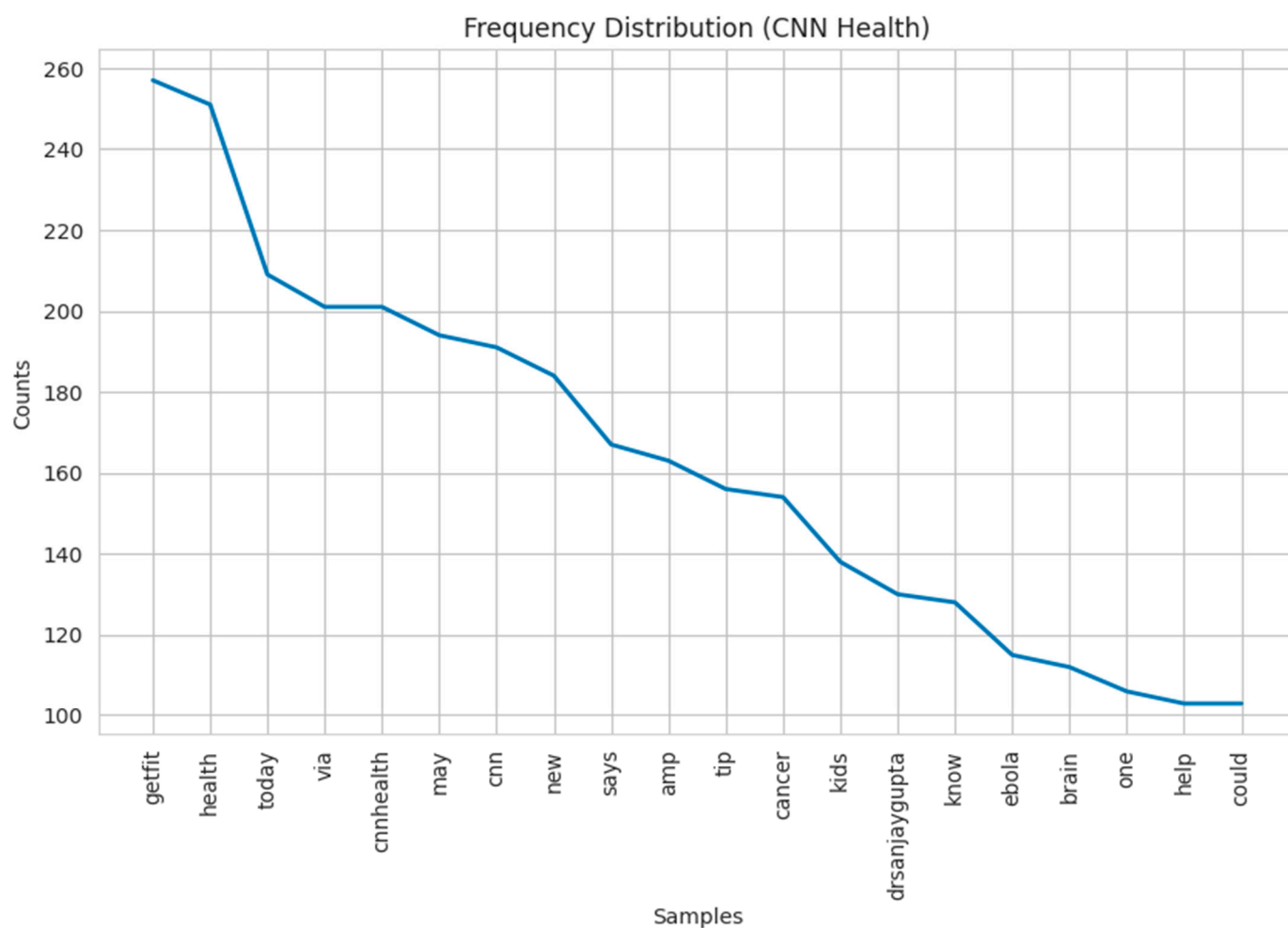


Figure 4. Major Topics Extracted from CNN Dataset.

The examination of the Reuters dataset revealed a significant emphasis on medical studies, regulatory matters, and disease management, particularly concerning diabetes, cardiovascular diseases, and pharmaceutical approvals (see Figure 5). The dataset emphasizes the health of women and children, particularly in relation to clinical trials, treatment strategies, and diseases such as cancer. The dataset encompasses outbreaks of infectious diseases, such as Ebola, concentrating on their impact on African nations, particularly Sierra Leone. The subject matter can be readily linked to outbreaks of infectious diseases, such as Ebola, particularly in regions like Africa and Sierra Leone. Terms such as “outbreak,” “hospital,” “patient,” “case,” and “ebola” are utilized to discuss the impact of disease epidemics on public health systems, particularly in African nations. The primary emphasis will probably be on managing and maintaining documentation related to these health emergencies.

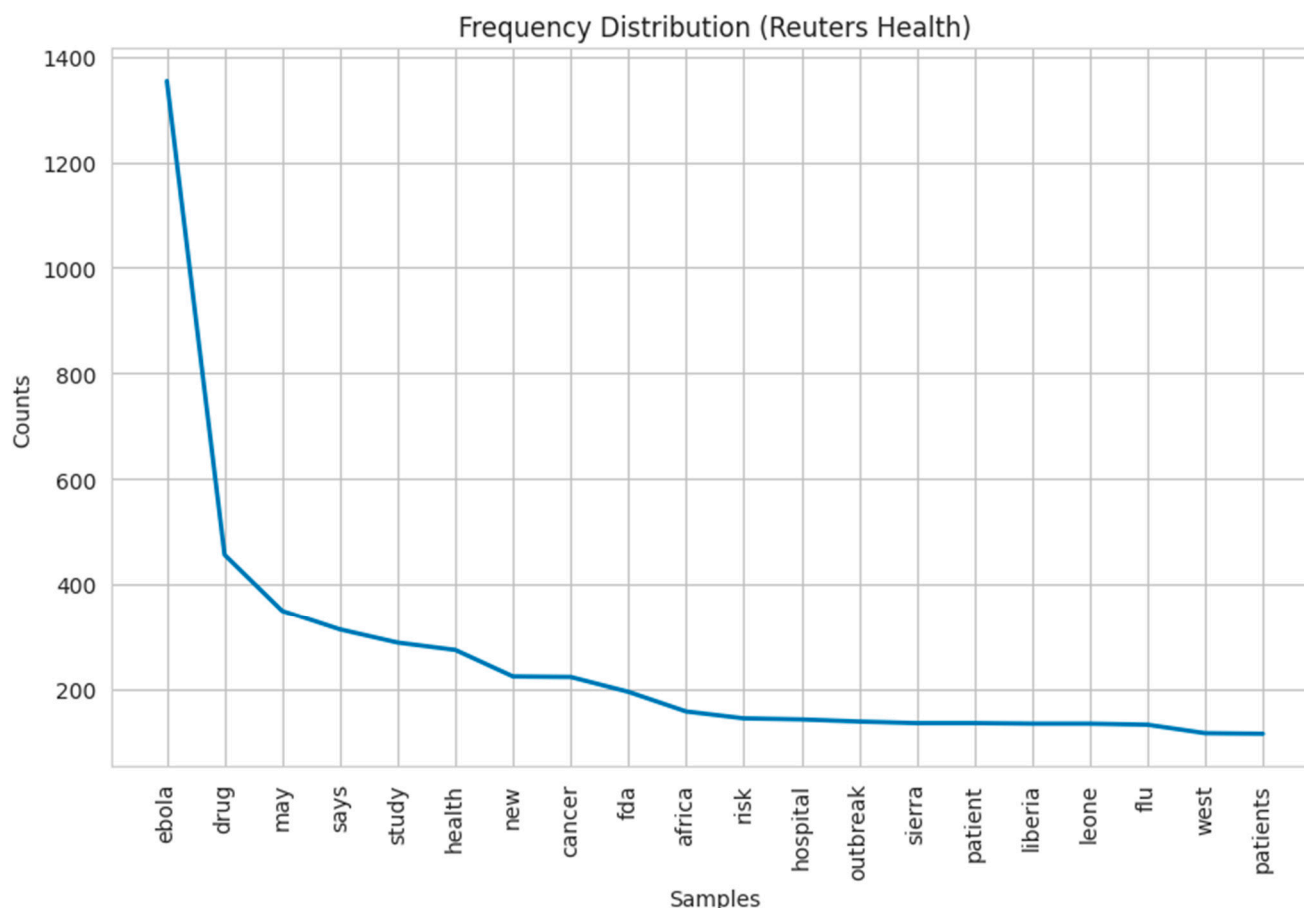


Figure 5. Major Topics Extracted from Reuters Dataset.

4.2. Data Cleaning

High-quality data was essential, as unprocessed tweets may contain errors, insufficient information, or irrelevant content. In the subsequent phase, the research ensured that the tweets from BBC, CNN, CBC, and Reuters were cleaned and preprocessed.

At first, the research performs “Parsing and Structuring Data”. At this stage, the pipe character (‘|’) serves to divide each line of the raw data, which is processed as lines of text. This dissects each record into three components: the Tweet ID, the time stamp, and the tweet content. The parsed records were organized into a list and subsequently transformed into a structured pandas DataFrame, complete with appropriate column names to facilitate ease of use and analysis.

The `clean_text` function is established to perform a sequence of standard procedures aimed at cleaning and normalizing text. This involves converting all letters to lowercase, eliminating short words (1–2 characters), and removing punctuation and numbers, URLs, and any unnecessary spaces. Finally, the text undergoes a filtration process to eliminate prevalent English stopwords such as “the,” “and,” and “is.”

The `clean_text` function is applied to the “Content” column of the DataFrame. A new column titled “Processed_Content” has been created to house the cleaned and normalized version of the original text data. This prepares it for further examination. A `CountVectorizer` transforms the processed text into a numerical format. This transformer generates a Document-Term Matrix (DTM), presenting a structured table where each row corresponds to a document (tweet) and each column represents a unique word, accompanied by numerical values indicating the frequency of each word’s occurrence. The parameters set are `max_df = 0.9` and `min_df = 2`. Terms that appear in over 90% of

texts or in less than 2 texts are disregarded. This aids in eliminating terms that are excessively common or exceedingly rare.

We apply Latent Dirichlet Allocation (LDA) to the vectorized text in order to uncover concealed thematic topics within the dataset. The configuration of the model is designed to identify three primary topics ($n_components = 3$). Once the model has been fitted, each document is categorized according to its primary topic, and the ten most representative words for each topic are displayed to facilitate understanding.

The processed text from the dataset is compiled into a single string and subsequently divided into individual words. Upon re-evaluating these tokens, all stopwords and extremely short words have been eliminated. A frequency distribution of the filtered tokens is subsequently created to identify the most prevalent words within the corpus. This is beneficial for conducting exploratory data analysis.

Batch processing across datasets was performed; this was done using an identical method for text cleaning that was applied to multiple datasets, including 'bbc', 'cnn', 'cbc', and 'reuters'. The `clean_text` function operates on the original "Content" column for each DataFrame, with the outcomes stored in a new "Cleaned_Text" column. The initial rows of this new column are printed to efficiently assess the results of the cleaning process.

4.3. Model Evaluation Results

The core components of the training system are initialized with the models instantiated with the predefined hyperparameters and the vocabulary size, building the neural network architecture. The Adam optimizer is also created, linking it to the model's trainable parameters and setting its learning rate. This optimizer is responsible for adjusting the model's weights to minimize the loss function.

An empty list (`train_losses`, `val_losses`, `train_accuracies`, `val_accuracies`) is created to log the performance metrics after each epoch, which later allows for visualization of the training history. Crucially, the variables for early stopping are initialized. The strategy tracks the `best_val_accuracy` (and `best_train_accuracy`), using a `patience_counter` to count how many epochs have passed without improvement on the validation set. This mechanism is designed to halt training prematurely if the model stops generalizing to the validation data, thereby preventing overfitting and saving computational resources. This entire setup creates a robust, monitored, and reproducible environment ready to execute the upcoming training loop. Finally, all the results are collected and interpreted.

4.3.1. Presentation of the Result

This section details the experimental results, providing a clear interpretation of the model performances and the conclusions derived from them.

Experiment and Result with BBC Dataset

LSTM demonstrates a consistent advantage over CNN in the BBC health-news stream due to its capability to model sequences (see Table 2). Accuracy is 98.40%, an increase from 97.90% (+0.50 percentage points). LSTM reduces the error rate by approximately 24%, decreasing it from 2.10% to 1.60%. The precision rates are 98.69% compared to 98.09%, indicating an improvement of 0.60 percentage points, which results in a reduction in false alarms in identifying relevant tweets within the conversation. Recall is 98.10% compared to 97.70% (+0.40 pp), which indicates a reduction in the number of relevant tweets that were overlooked. The FPR is 1.30% compared to 1.90% (−0.60 pp), which indicates an approximately 32% reduction in false positives. F1 score is 98.40% compared to 97.90% (an increase of 0.50 percentage points), which indicates an improved balance between precision and recall overall.

The primary practical advantage is the reduced false positive rate: fewer irrelevant tweets are incorrectly classified as relevant to discourse, thereby minimizing noise in subsequent analyses, such as monitoring transitions from offline to online platforms or informing a healthcare recommendation system. The concurrent rise in recall indicates that LSTM also overlooks fewer tweets of significant importance. This indicates an improved filtering capability and an enhanced ability to identify significant elements.

Table 2. Model Performance Training with LSTM and CNN On BBC Dataset.

Model	Accuracy	Precision	Recall	FPR	F1-Score
LSTM	98.40%	98.69%	98.10%	1.30%	98.40%
CNN	97.90%	98.09%	97.70%	1.90%	97.90%

The LSTM-VAE applied to BBC health-news tweets demonstrates a smooth and well-regulated learning process. The training and validation loss decreases rapidly initially and subsequently stabilizes. At epoch 10, the training loss approaches 0, while the validation loss is approximately 0.20, indicating that the model demonstrates a good fit without divergence (refer to Figure 6). The accuracy increases from approximately 0.91–0.92 at epoch 1 to a stable validation accuracy of around 0.984 by epoch 8. Conversely, training accuracy increases to approximately 1.00 shortly thereafter. The minimal generalization gap (train ≈ 1.00 vs. val ≈ 0.984) and the absence of late-epoch increases in validation loss suggest the absence of overfitting and indicate stable convergence. This indicates that the model consistently identifies sequential patterns in BBC health tweets, reflecting the evolution of public discourse online. This constitutes a significant component of the analysis pipeline.

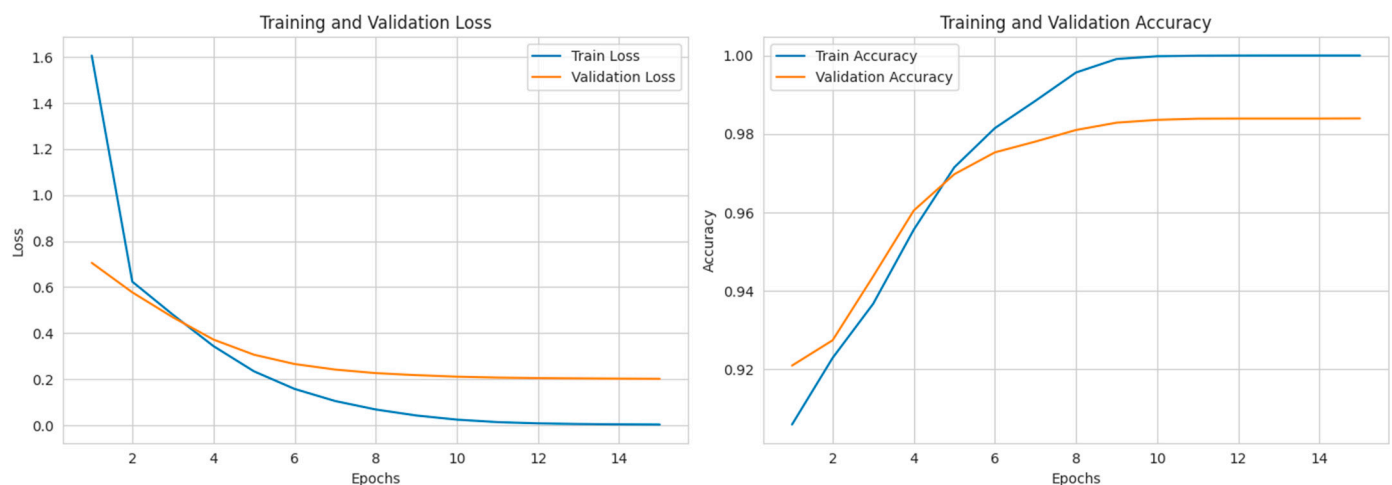


Figure 6. A Plot of Loss and Accuracy validation of LSTM VAE Model for BBC Dataset.

The GRU-VAE applied to BBC health-news tweets demonstrates rapid and consistent learning. The loss shows a marked reduction in the early epochs (training from approximately 2.7 to below 0.1; validation from approximately 0.63 to around 0.17) and then stabilizes, with the training and validation curves remaining closely aligned, suggesting the absence of divergence or late-epoch overfitting (see Figure 7). Accuracy increases consistently from approximately 0.91 to a validation plateau of roughly 0.984 by the eighth epoch. Subsequently, training accuracy increases to approximately 1.00. The minimal generalization gap and absence of oscillations indicate robust convergence. The GRU-VAE effectively identifies sequential patterns in BBC health tweets, comparable to the performance of the LSTM-VAE. This provides a dependable indication of the transformation in online discourse influenced by the news media.

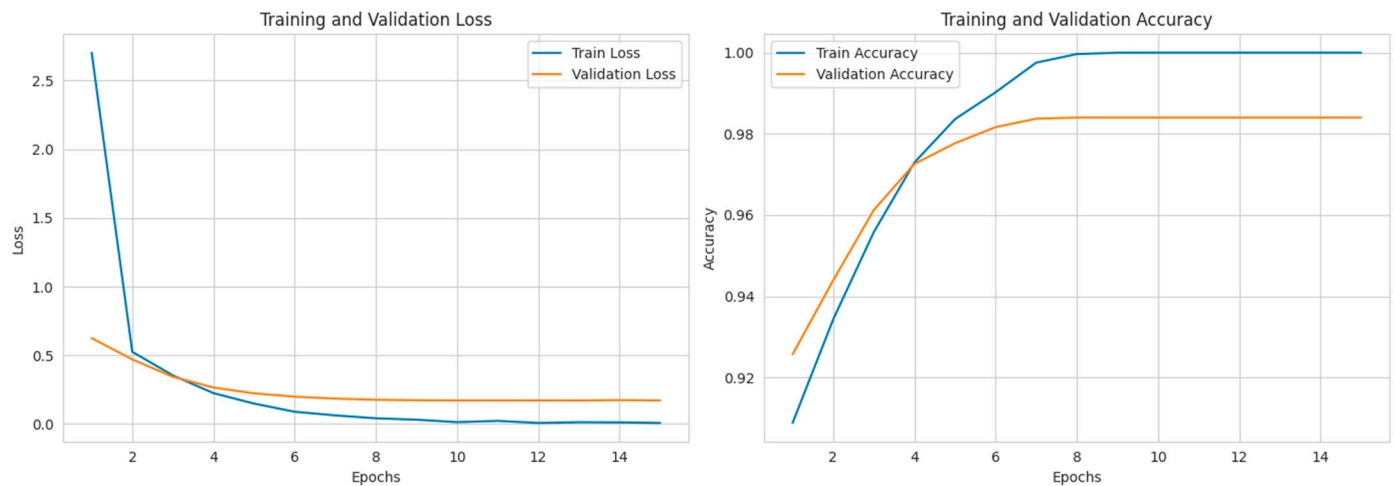


Figure 7. The Plot of Loss and Accuracy validation of GRU VAE Model for BBC Dataset.

The Transformer-VAE demonstrates rapid learning on BBC health-news tweets and effectively generalizes its knowledge to diverse contexts. The training loss decreases significantly in the initial epochs (from approximately 30 to around 1 by epoch 4, subsequently approaching 0), whereas the validation loss rapidly declines to a stable level between approximately 0.16 and 0.20 without any rebound in later epochs. This indicates an absence of evidence suggesting that the model is overfitting (refer to Figure 8). The accuracy increases rapidly from approximately 0.93 to a validation plateau of around 0.984 by epochs 6 to 7, while the training accuracy approaches 1.00 shortly thereafter. The brief mid-epoch crossing, where validation slightly exceeds training, is indicative of typical optimization noise in attention models rather than instability. The curves indicate that the Transformer-VAE effectively captures the discourse patterns of BBC tweets, illustrating the transition of public conversation to online platforms. This is evidenced by the rapid convergence of the curves and the minimal, consistent generalization gap. The performance levels of the LSTM-VAE and GRU-VAE on this dataset are equivalent.

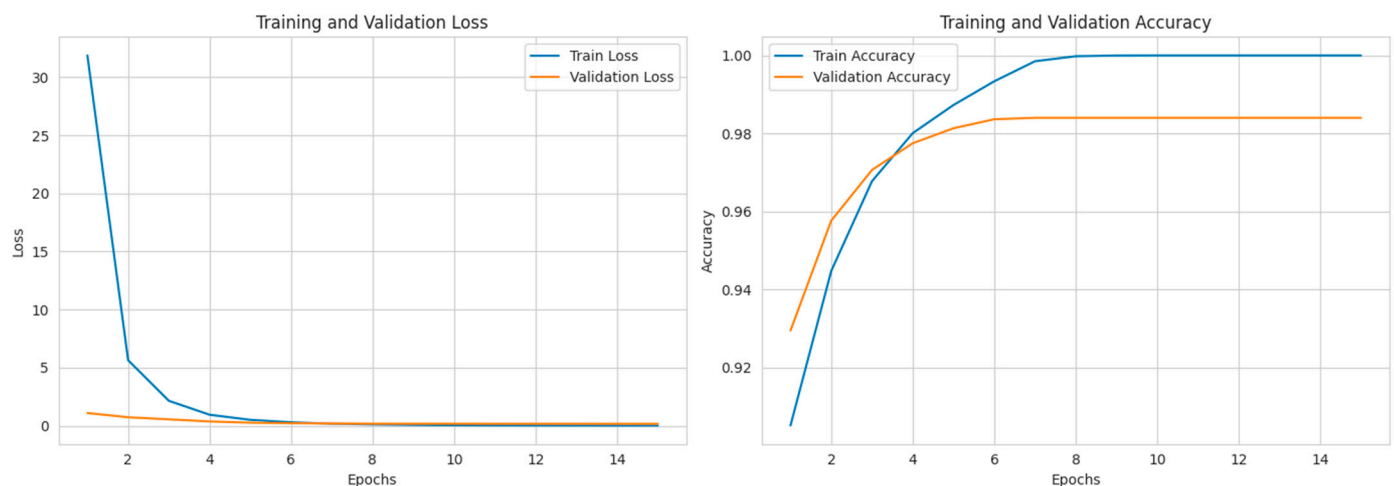


Figure 8. The Plot Loss and Accuracy of Transformer VAE Model for BBC Dataset.

Experiment and Result with CBC Dataset

The LSTM outperforms the CNN across all metrics for the CBC health-news tweets. The accuracy increases from 97.20% to 97.80% (+0.60 percentage points), as indicated in Table 3. This reduces the overall error rate from 2.80% to 2.20%, representing approximately a 21% relative decrease. The LSTM enhances both the hit rate and the number of

false alarms by 0.60 percentage points, indicating a positive outcome. The false positive rate decreases from 2.60% to 2.00% (−0.60 percentage points; approximately 23% relative reduction), while the balanced F1 score increases from 97.19% to 97.80%. The LSTM’s sequence modeling on CBC demonstrates superior ability to detect discourse cues compared to the CNN, resulting in a clearer and more comprehensive signal for identifying changes in online public conversation.

Table 3. Model Performance Training with LSTM and CNN On CBC Dataset.

Model	Accuracy	Precision	Recall	FPR	F1-Score
LSTM	97.80%	97.99%	97.60%	2.00%	97.80%
CNN	97.20%	97.39%	97.00%	2.60%	97.19%

The LSTM-VAE applied to the CBC health-news tweets demonstrates consistent and organized learning outcomes. The loss decreases rapidly during the initial 5–6 epochs (training from approximately 2.0 to below 0.4; validation from approximately 1.2 to around 0.35) and subsequently stabilizes. At epoch 11, the training loss approaches zero, and the validation loss stabilizes without a subsequent increase, indicating no evidence of overfitting (refer to Figure 9). The accuracy increases consistently from approximately 0.88 to a plateau of around 0.978–0.98 by epochs 7–8, with training accuracy achieving approximately 1.00 shortly thereafter. The minimal generalization gap (train \approx 1.00 vs. val \approx 0.98) indicates effective convergence and strong generalization capabilities of the model. This indicates that the LSTM-VAE effectively captures the sequential discourse patterns of CBC, supporting the assertion that deep models can accurately identify the transition of public health discussions to online platforms.

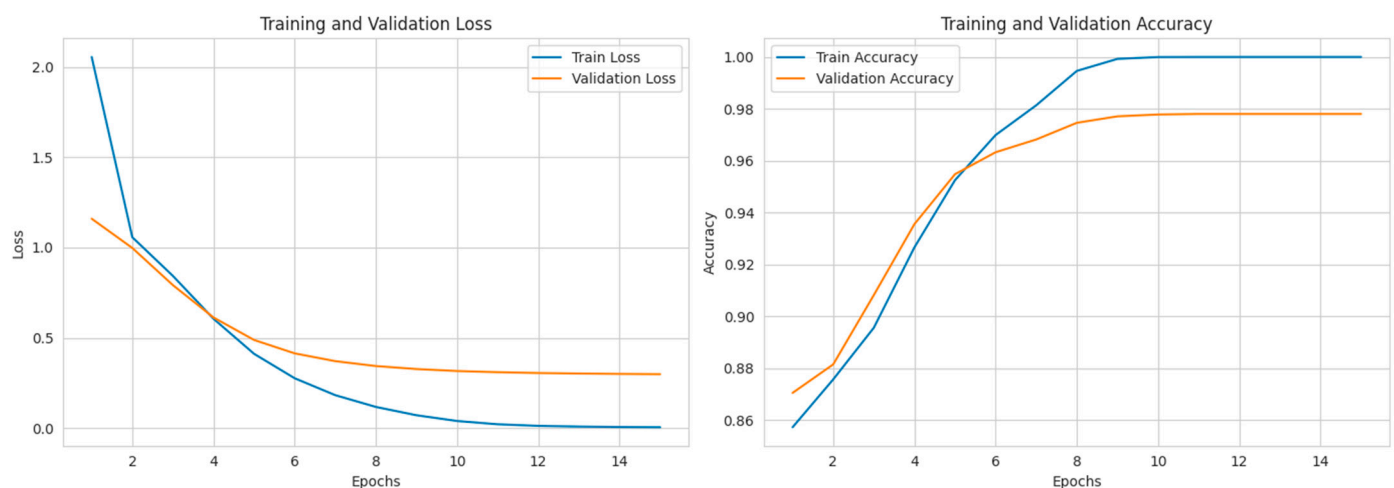


Figure 9. The Plot of Loss and Accuracy validation for LSTM VAE Model for CBC Dataset.

The GRU-VAE applied to the CBC health-news tweets demonstrates rapid and stable convergence. The loss decreases rapidly during the initial 3–5 epochs (training from approximately 3.0 to below 0.4; validation from approximately 1.1 to around 0.3), subsequently stabilizing with the two curves closely aligned. The absence of an increase in validation loss at the conclusion of the epoch indicates no evidence of overfitting (refer to Figure 10). The accuracy increases consistently from approximately 0.88 to a validation plateau of around 0.978 by epochs 6 to 7, while training accuracy approaches 1.00 shortly thereafter. The consistent and narrow gap between the train and validation sets, along with the absence of oscillations, indicates that the model demonstrates strong generalization capabilities. The GRU-VAE models replicate CBC’s sequential discourse patterns

nearly identically to the LSTM-VAE. This substantiates the paper's assertion that deep sequence models can effectively detect the transition of public health discourse to online platforms.

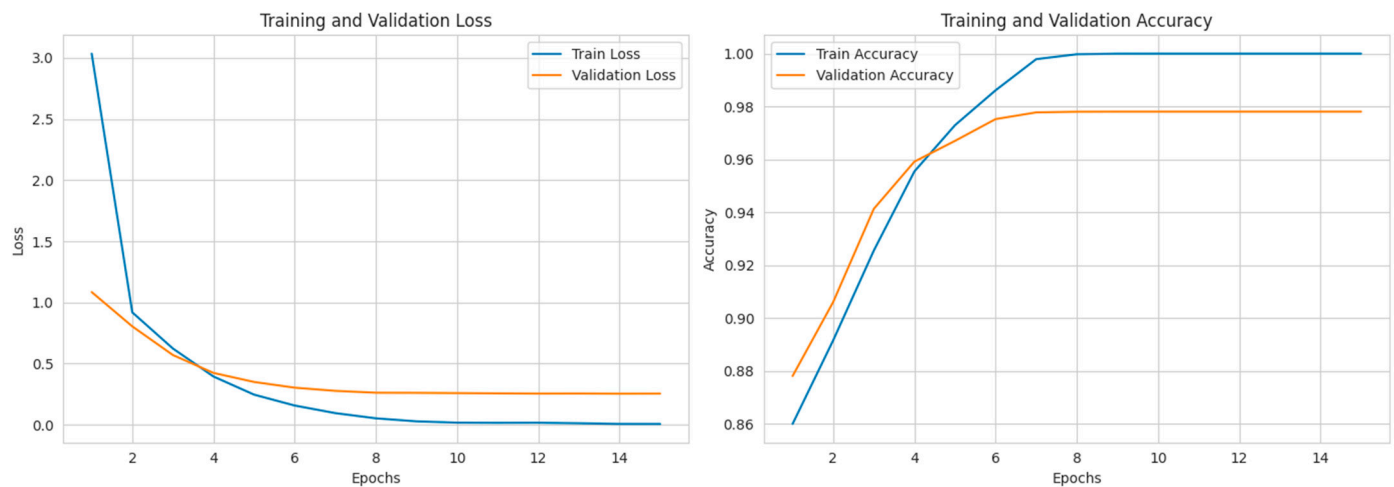


Figure 10. The Plot of Loss and Accuracy for GRU VAE Model for CBC Dataset.

The Transformer-VAE applied to the CBC health-news tweets demonstrates rapid convergence, albeit to a lower performance ceiling, indicating a tendency towards underfitting rather than overfitting. The training loss initially decreases significantly before stabilizing, whereas the validation loss remains relatively constant near a shallow baseline, exhibiting only minor enhancements (see Figure 11). The training accuracy on the right peaks at approximately 0.864, whereas the validation accuracy fluctuates between 0.859 and 0.862, failing to approach the ≈ 0.98 plateau achieved by LSTM-VAE/GRU-VAE on CBC. The persistent train-validation gap and low joint ceiling indicate that the model's capacity or optimization for this source is constrained. This may result from insufficient encoder depth or heads, an inadequate maximum sequence length, or excessively strong KL regularization, leading to partial posterior collapse. The Transformer-VAE demonstrates inferior performance in capturing the online discourse signal compared to recurrent VAEs for CBC. The per-dataset results indicate that this model achieves a validation accuracy of approximately 86.2%, whereas LSTM/GRU attains 97.8%.

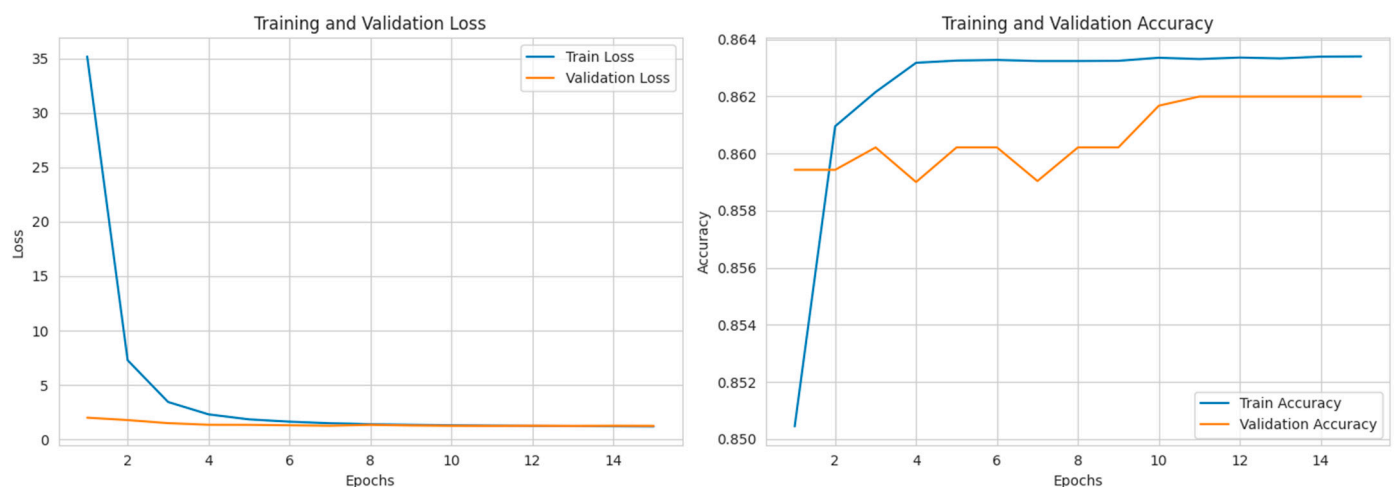


Figure 11. The Plot of Loss and Accuracy for Transformer VAE Model for CBC Dataset.

Experiment and Result with CNN Dataset

The LSTM outperforms the CNN across all metrics regarding health news tweets. The accuracy increases from 97.80% to 98.40% (+0.60 percentage points), resulting in a reduction in the error rate from 2.20% to 1.60%, representing a relative decrease of approximately 27%. The LSTM identifies a greater number of relevant tweets while also reducing the number of missed instances (see to Table 4). The precision increases to 98.59% (an increase of 0.69 percentage points), while the recall rises to 98.20% (an increase of 0.50 percentage points). The most significant operational improvement is the FPR, which decreases from 2.10% to 1.40% (−0.70 percentage points; approximately 33% relative reduction). This enhances the clarity of the signal for subsequent monitoring. The overall balance, indicated by F1 (98.40% compared to 97.80%), demonstrates that the LSTM outperforms the CNN in recognizing sequential cues within the stream. This aligns with the objective of the paper to identify shifts in public discourse from offline to online contexts.

Table 4. Model Performance Training with LSTM and CNN On CNN Dataset.

Model	Accuracy	Precision	Recall	FPR	F1-Score
LSTM	98.40%	98.59%	98.20%	1.40%	98.40%
CNN	97.80%	97.90%	97.70%	2.10%	97.80%

The LSTM-VAE applied to CNN health-news tweets demonstrates rapid and consistent learning without overfitting. The loss decreases rapidly during the initial 5–6 epochs, reducing from approximately 2.1 to below 0.3 for training and from around 1.3 to approximately 0.35 for validation, before stabilizing. The train and validation curves remain closely aligned (see Figure 12). During epochs 7 to 8, accuracy increases steadily from approximately 0.86 to a validation plateau ranging from 0.973 to 0.974. Conversely, training accuracy increases to approximately 1.00 shortly thereafter. The minimal generalization gap and absence of late-epoch loss rebound indicate robust convergence of the system. The model effectively identifies the sequential cues in CNN tweets that indicate the shift in health discourse online, consistent with the results of your cross-source analysis.

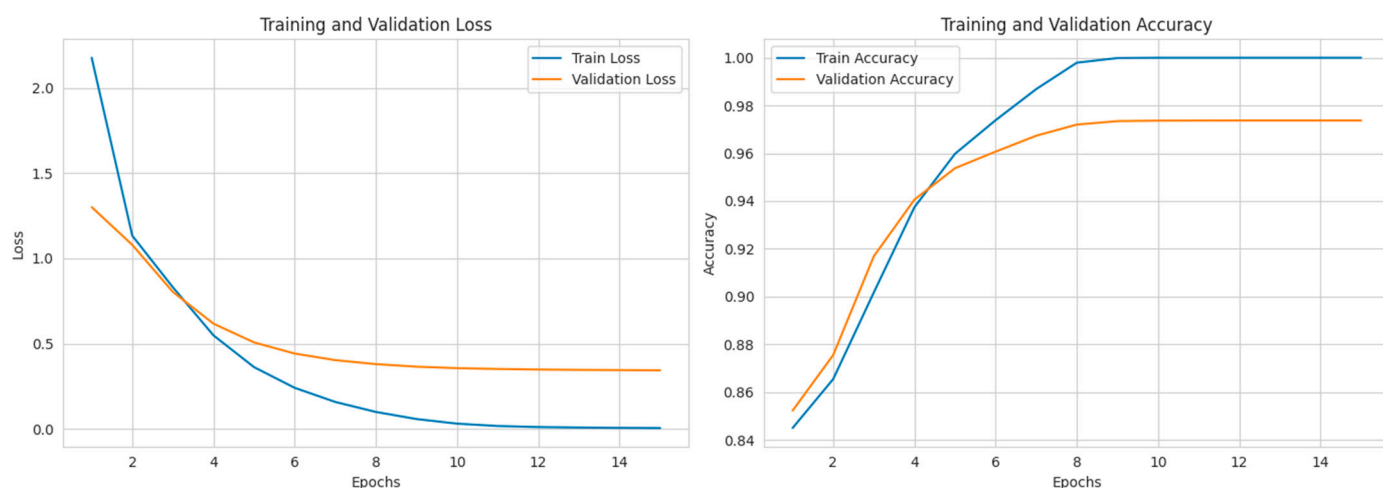


Figure 12. The Plot of Loss and Accuracy for LSTM VAE Model for CNN Dataset.

The GRU-VAE applied to CNN health-news tweets demonstrates rapid and problem-free convergence. The loss decreases rapidly during the initial epochs (training from approximately 3.0 to below 0.3; validation from approximately 1.1 to around 0.35) before stabilizing. The train and validation curves closely align, exhibiting no rebound in the final epoch, indicating the absence of overfitting (see Figure 13). The accuracy increases

consistently from approximately 0.85 to a validation plateau of roughly 0.973–0.974 between epochs 6 and 8, whereas the training accuracy remains around 1.00. The minimal generalization gap indicates that the GRU-VAE effectively identifies sequential cues in CNN's health tweets, reflecting the shift in public discourse to online platforms.

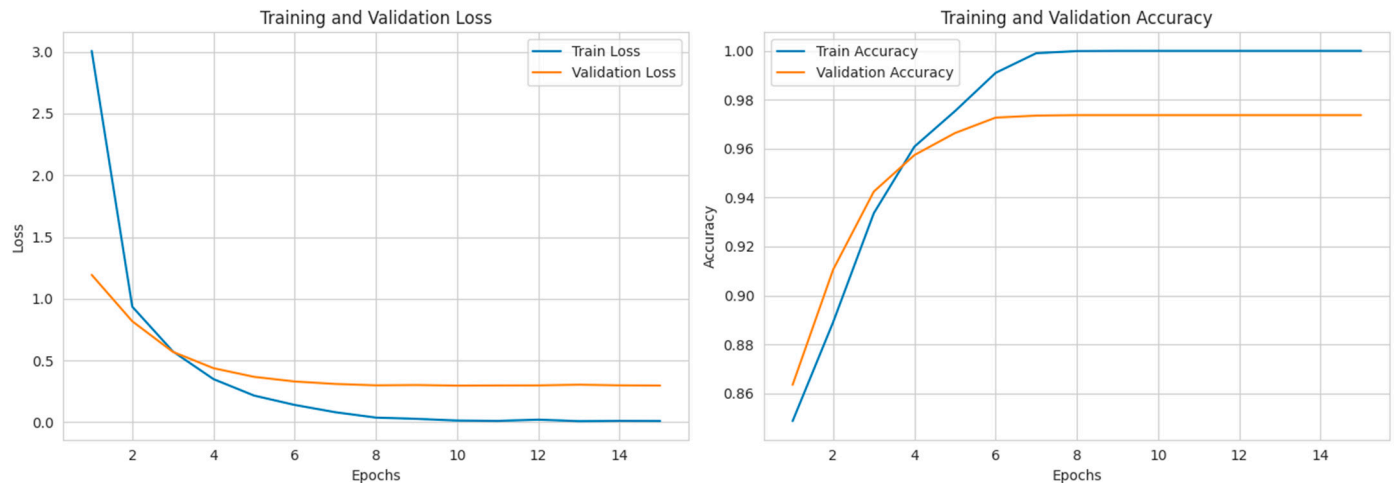


Figure 13. The Plot for Loss and Accuracy validation of GRU VAE Model for CNN Dataset.

The Transformer-VAE applied to CNN health-news tweets demonstrates rapid learning and achieves stable performance. The training loss decreases significantly during the initial three to four epochs, subsequently approaching zero. The validation loss decreases rapidly to a stable level without any late-epoch increase, indicating that the model is not experiencing overfitting. Accuracy increases significantly from approximately 0.86 to a validation plateau of around 0.973–0.974 by epochs 6–7. Training accuracy stabilizes at approximately 1.00 shortly thereafter (see Figure 14). A brief mid-epoch crossing, where validation performance slightly exceeds training performance, is a typical artifact of optimization rather than an indication of instability. The curves indicate rapid convergence of the models and a minimal generalization gap. The performance of the Transformer-VAE on this dataset is comparable to that of the LSTM-VAE/GRU-VAE, indicating its effectiveness in accurately capturing the online health-discourse patterns of CNN.

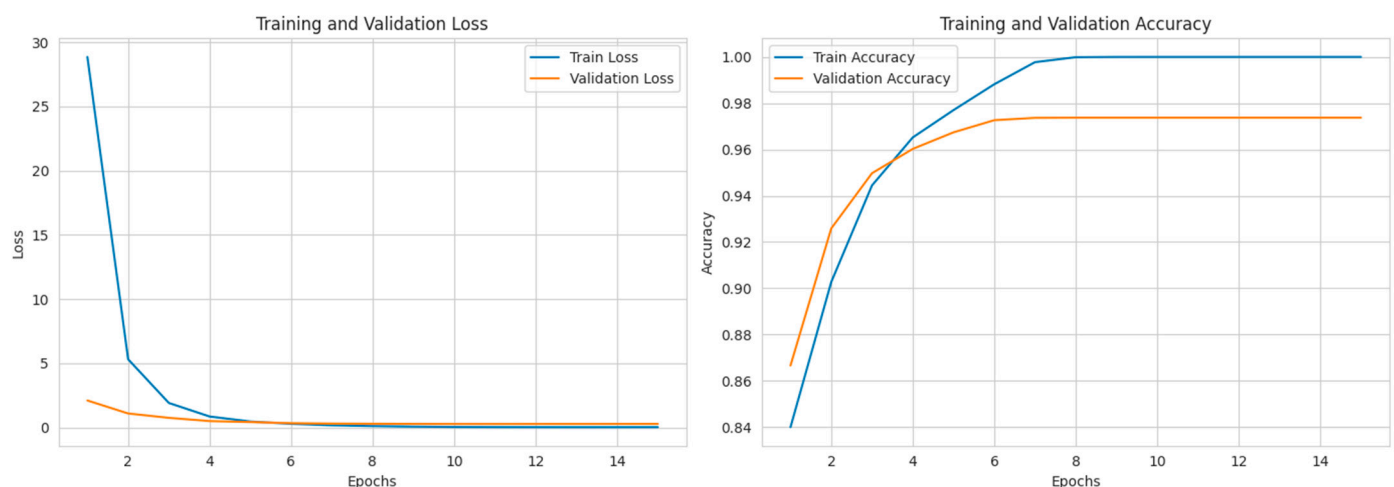


Figure 14. The Plot of Loss and Accuracy for Transformer VAE Model for CNN Dataset.

Experiment and Result with Reuters Dataset

The LSTM outperforms the CNN consistently on the Reuters health-news tweets. The accuracy increases from 96.50% to 97.30% (+0.80 percentage points), while the error rate decreases from 3.50% to 2.70%, representing a relative decrease of approximately 23% (see Table 5). Precision increases by 1.00 percentage points (97.59% compared to 96.59%), indicating that LSTM generates fewer false alarms. Recall increases by 0.60 percentage points (97.00% compared to 96.40%), indicating an enhanced ability to identify relevant tweets. The primary operational advantage is a reduction in the false positive rate from 3.40% to 2.40% (−1.00 percentage points; approximately 29% relative decrease), thereby enhancing the clarity of the signal for subsequent analysis. The elevated F1 score (97.29% compared to 96.50%) indicates that LSTM achieves a superior balance between precision and recall on the Reuters dataset. This enhances the reliability of identifying news-amplified online discourse.

Table 5. Model Performance Training with LSTM and CNN On Reuters Dataset.

Model	Accuracy	Precision	Recall	FPR	F1-Score
LSTM	97.30%	97.59%	97.00%	2.40%	97.29%
CNN	96.50%	96.59%	96.40%	3.40%	96.50%

The LSTM-VAE applied to the Reuters health-news tweets demonstrates convergence that is both exemplary and highly reliable. The loss decreases rapidly during the initial epochs (training from approximately 1.7 to below 0.1; validation from approximately 1.1 to around 0.1), subsequently stabilizing, resulting in nearly overlapping curves. The absence of a rebound in the final epoch indicates a lack of overfitting (see Figure 15). On the right, accuracy increases from approximately 0.89 to a validation plateau of around 0.996 between epochs 6 and 8, whereas training accuracy stabilizes at approximately 1.00. The minimal generalization gap and the elevated ceiling indicate that Reuters tweets exhibit consistent and easily identifiable discourse patterns. This indicates that the model effectively detects the transition of public health discussions to online platforms with high precision for this source.

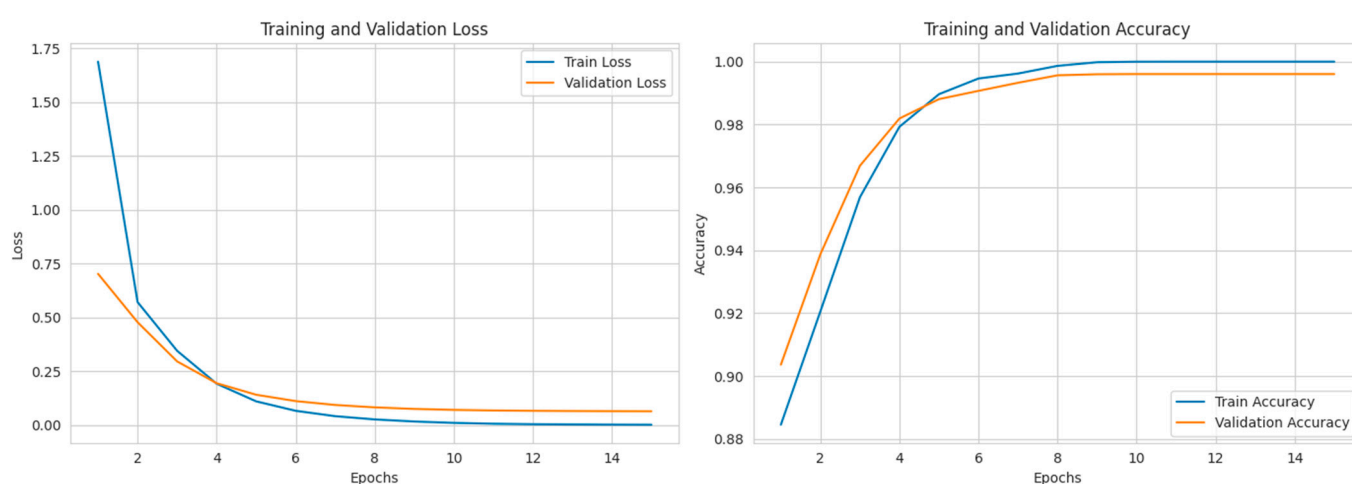


Figure 15. The Plot of Loss and Accuracy for LSTM VAE Model for Reuters Dataset.

The GRU-VAE applied to Reuters health-news tweets demonstrates rapid convergence and maintains stability thereafter. The loss decreases significantly during the initial epochs and subsequently remains low. The slight increase in validation loss observed during epochs 10 to 11 is transient and does not impact overall performance. During epochs

5 to 6, there is a significant increase in accuracy, reaching a validation plateau of approximately 0.996, whereas training accuracy stabilizes at around 1.00. This results in a minor, consistent generalization gap (see Figure 16). The curves indicate that GRU-VAE effectively captures the online health-discourse patterns of Reuters, comparable to other sequence models. This indicates effective generalization and a lack of overfitting.

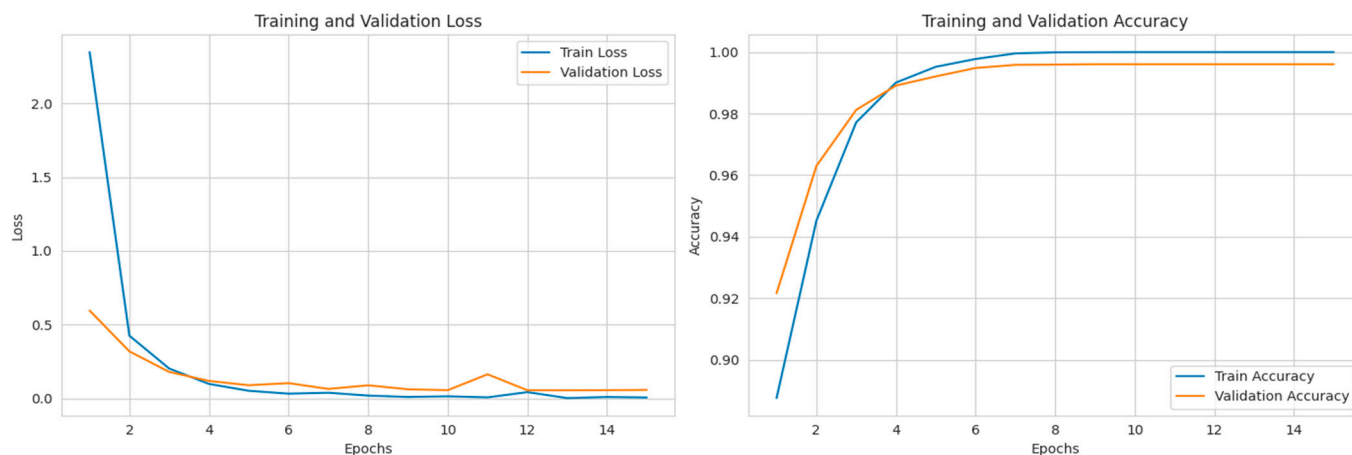


Figure 16. The Plot for Loss and Accuracy for GRU VAE Model for Reuters Dataset.

The Transformer-VAE applied to Reuters health-news tweets demonstrates rapid and consistent convergence. The absence of a late-epoch rebound in validation loss indicates that overfitting is not occurring (refer to Figure 17). Training loss on the left decreases from approximately 25 to below 0.3 within 3 to 4 epochs, subsequently approaching 0. Accuracy increases significantly from approximately 0.88 to a validation plateau of around 0.996 by epochs 6 to 7. Training accuracy stabilizes around 1.00, resulting in a minimal and consistent generalization gap. It is typical for attention models to exhibit brief intervals where validation performance slightly exceeds that of training, indicating no stability issues. The curves indicate that the Transformer-VAE models effectively capture the online health-discourse patterns of Reuters. This aligns with the paper's assertion that deep learning can effectively detect the transition of public discourse from offline to online platforms.

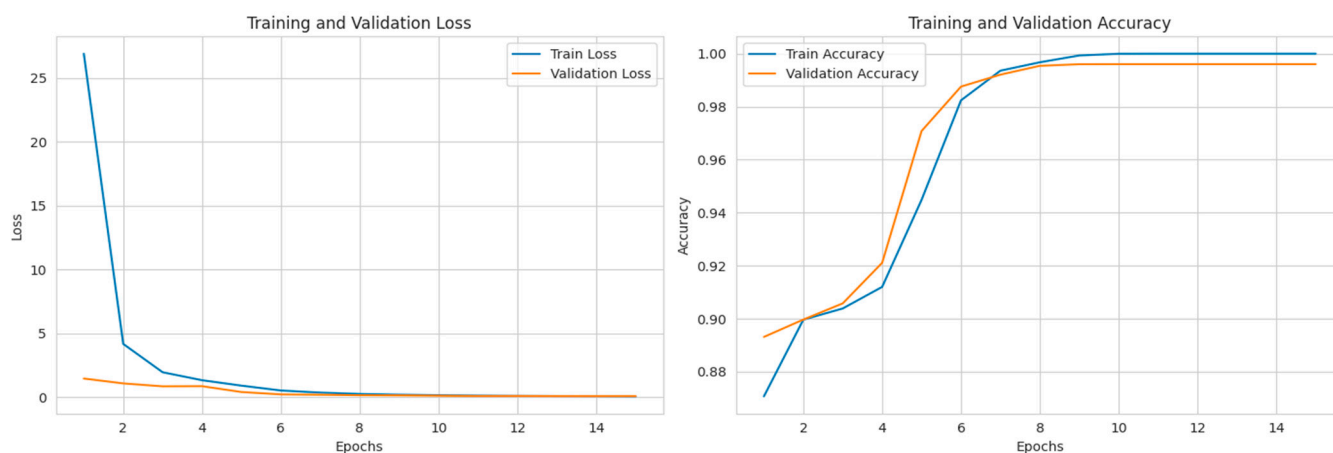


Figure 17. The Plot of Loss and Accuracy for Transformer VAE Model for Reuters Dataset.

Table 6 presents a summary of the results from four newscasts: BBC, CNN, CBC, and Reuters. Reuters possesses the largest collection of tweets, totaling 4719, while CBC has the smallest, with 3741 tweets. Reuters exhibits the poorest test results across all models,

despite being the largest entity. The LSTM VAE exhibits the lowest F1 score at 97.29% and the highest false-positive rate at 2.40%. The CNN model exhibits the lowest F1 score at 96.50% and the highest false positive rate at 3.40%. The BBC and CNN receive the highest ratings among media outlets. The LSTM VAE achieves the highest F1 score of 98.40% on both datasets, as well as the lowest FPR of 1.30% on the BBC dataset. The LSTM VAE model outperforms the CNN model by approximately 0.5 to 0.8 F1 points and reduces the false positive rate by 0.6 to 1.0 points, indicating greater stability. It is noteworthy that an increase in data volume did not ensure enhanced accuracy (as observed in the comparison between Reuters and BBC/CNN), indicating that both topical diversity and the stylistic characteristics of sources influence generalization. Consequently, per-source calibration or domain adaptation could potentially enhance the reliability of deployment.

Table 6. A Summary of Results from Four Newscasts: BBC, CNN, CBC, and Reuters.

Dataset	LSTM Acc	LSTM Prec	LSTM Rec	LSTM F1	LSTM FPR	CNN Acc	CNN Prec	CNN Rec	CNN F1	CNN FPR
BBC	98.40	98.69	98.10	98.40	1.30	97.90	98.09	97.70	97.90	1.90
CNN	98.40	98.59	98.20	98.40	1.40	97.80	97.90	97.70	97.80	2.10
CBC	97.80	97.99	97.60	97.80	2.00	97.20	97.39	97.00	97.19	2.60
Reuters	97.30	97.59	97.00	97.29	2.40	96.50	96.59	96.40	96.50	3.40

5. Discussion

Increasingly, discussions regarding news and health originate or expand on Twitter/X, contrasting with previous in-person interactions. Deep learning models enable the capture of rapid, high-volume changes with minimal manual feature engineering. Previous research suggests that (i) online opinions can diverge from offline polls, highlighting the need for direct measurement, and (ii) stakeholders in healthcare and public health strategically employ Twitter to shape engagement and propagate widely accepted narratives. This facilitates the implementation of a deep learning pipeline utilizing diverse news-sourced health datasets (BBC, CNN, CBC, Reuters) to identify and monitor the migration of discourse [43].

LSTM demonstrated superior performance compared to CNN across all outlets, achieving validation accuracy of 98.4% on BBC and CNN, 97.8% on CBC, and 97.3% on Reuters. Precision, recall, and F1 scores were consistently high ($\geq 97\%$ on the strongest datasets), accompanied by low false positive rates (approximately 1.3–2.4%). The observed levels align with previous research indicating that deep architectures effectively model tweet semantics at scale, such as BERT/LSTM pipelines and hybrid networks for Twitter health sentiment, thereby justifying our selection of sequence models for discourse detection [44].

The BBC and CNN sets exhibit equal accuracy in health tweets related to headline news, achieving a rate of 98.4% each, alongside the most consistent low validation loss. CBC and Reuters follow closely. High-profile, professionally curated accounts provide a clearer signal-to-noise ratio for model training, while deep networks maintain robustness despite topic changes, provided that preprocessing is effective. The findings indicate that the framing of discourse by major news outlets can be accurately detected by machines, facilitating the monitoring of shifts from offline to online platforms over time [45].

The pipeline systematically enables reproducible monitoring of health news frames as they migrate online, providing outlet-specific detail for evaluating engagement and sentiment changes. Health agencies and hospital systems can monitor the transformation of public discourse shaped by media signals, predict increases in infodemic activity, and adjust messaging before narratives become established in offline settings. The robust

LSTM baselines demonstrate that lightweight sequence models, improved through careful preprocessing and embeddings, are sufficient for operational dashboards that detect changes by topic, outlet, and time. This establishes a link between social listening and sustainable healthcare recommender systems that incorporate population-level discourse as a dynamic contextual signal [46].

Following the modification to [47,48], this study achieves the highest accuracy at 98.40%, succeeded by [49] at 97.11% and [18] at 97%, with [37] ranking lowest at 92.59% (see Table 7). This study demonstrates superior performance with a precision of 98.69%, a recall of 98.20%, an F1 score of 97.90%, and the lowest false positive rate of 1.30%. This approach outperforms [49] (precision: 97.83%; recall: 95.13%; F1: 96.41%) and [18] (precision: 95.80%; recall: 90.50%; F1: 93.50%; FPR: 5%). The recent report indicates a precision of 91.87% and a recall of 95.97%. The FPR of 93.92% indicates a significantly high false-positive rate, diminishing its utility for operations despite a favorable recall. Ref. [48] exhibits the lowest performance across all shared metrics, with an accuracy of 84.56%, precision of 82.34%, recall of 82.12%, and an F1 score of 81.23%. Upon reviewing the evidence, it appears that this study demonstrates the most favorable outcomes, characterized by a low false positive rate alongside high precision and recall. This represents the optimal configuration for the real-time observation of discourse shifts. A model that both captures more relevant tweets and avoids flagging noise produces cleaner downstream signals, such as early warning, narrative tracking, and inputs to healthcare recommender systems. We still have a multi-outlet, health-news-focused pipeline, full-metric reporting (including FPR), and a consistently strong performance across sources.

Table 7. Comparative Performance with Previous Studies.

Study	Accuracy	Precision	Recall	FPR	F1-Score
[18]	97%	95.80%	90.50%	5%	93.50%
[47]	92.59%	91.87%	95.97%	93.92%	-
[48]	84.56%	82.34%	82.12%	-	81.23%
[49]	97.11%	97.83%	95.13%	-	96.41%
This Study	98.40%	98.69%	98.20%	1.30%	97.90%

The limitations and recommendations for future studies arise from the constraint of our four datasets to news accounts, which capture media-framed health discourse rather than a comprehensive representation of peer-to-peer communication, potentially leading to an under-representation of grassroots perspectives. Secondly, despite low validation false positive rates, “topic drift” and bot activity can still disrupt estimates between re-training cycles. Third, we examine English tweets; for comprehensive global monitoring, it is necessary to incorporate additional languages. In future developments, (i) transformers such as lightweight BERT/Roberta to enhance cross-domain robustness should be incorporated, (ii) explanations integrated (saliency/SHAP) to facilitate auditor usability, (iii) bot/coordination detection implemented as a gating signal, (iv) outlet-level signals linked to downstream HRS modules to improve content ranking and reduce confusion, and (v) offline–online linkage studies conducted to identify correlations between detected shifts and clinic call volume or vaccination uptake, thereby bridging the gap between online framing and offline behavior.

The implementation of a news-driven deep-learning framework that incorporates health streams from BBC, CNN, CBC, and Reuters enables organizations to proactively adjust their staffing, inventory, and triage protocols. The LSTM/CNN ensemble, complemented by a VAE drift monitor, yields highly accurate and stable outcomes, facilitating analysis and minimizing false positives. Per-source dashboards facilitate informed decision-making. Business value encompasses quantifiable KPIs such as the duration required

to identify new topics, the precision and recall of alerts, the time needed to respond, and the number of calls that could have been prevented. Hospitals anticipate an increase in emergency room patients; public health initiatives emphasize outreach and rumor control; pharmaceutical companies enhance pharmacovigilance signals; insurance companies initiate member education; and media and technology teams assess the sensitivity of rankings. APIs, containerized inference, and standard MLOps facilitate deployment within established analytics frameworks.

The framework transforms unstable social narratives into structured decision support rather than relying on automatic truth assessments. Integrating model outputs with offline indicators such as call center volume, clinic attendance, and vaccine uptake facilitates the calibration of actions and distinguishes between actual demand and amplification. Incorporating bot and coordination filters, conducting link-reputation assessments, and ensuring claim-source consistency reduces the likelihood of misinformation, while the inclusion of multilingual sources enhances coverage. Policy thresholds and per-source calibration enhance governance while facilitating rapid iterations in production. In summary, there is a reduction in unexpected events, an increase in the speed of coordinated responses, and an improvement in the alignment of communications. This fosters patient trust, enhances the effectiveness of campaigns, and strengthens operational resilience during rapidly evolving health events.

Limitations and Future Research Direction

There are some limitations and suggestions for future research that have been identified by this study. The dataset includes English-language health tweets from BBC, CNN, CBC and Reuters; thus, the findings pertain to news-mediated English discourse rather than peer-to-peer interactions or other languages. This study does not consider ranking or personalization effects, nor does it employ explicit methods for identifying misinformation or bot activity. Consequently, amplification or manipulation may distort the signals observed. The evaluation emphasizes reconstruction loss for variational autoencoders (VAEs) and classifier metrics on held-out splits, featuring limited human adjudication and lacking adversarial or out-of-distribution stress tests.

The lack of correlation between observed shifts and offline indicators, such as clinic calls and vaccination uptake, suggests a practical impact. We will correct minor naming discrepancies (CBC vs. CBS) and provide lists of outlets, sampling codes, and metrics for each source to enhance reproducibility. Future plans involve incorporating feeds from non-English-speaking and regional media, as well as public health agencies. Future study can also examine the impact of translation and the alterations that occur when content is adapted to various languages. Topic audits should be conducted by region, utilizing leave-one-source/region-out and out-of-domain testing methodologies, while incorporating credibility, link-reputation, and cascade-anomaly signals. This should incorporate bot and coordination filters and evaluate them against offline indicators while assessing sensitivity to platform ranking.

A significant issue is that our conclusions rely on signals from the news rather than on discussions among peers. The results are susceptible to biases in agenda-setting and framing inherent in the decisions made by newsrooms. Engagement patterns associated with verified media accounts, including retweets, timing, and topic cycles, reveal follower counts, headline presentation, and platform ranking. They do not consistently reflect individual thoughts or behaviors; thus, they serve primarily as a metric for assessing collective attention. This reliance constrains both demographic and geographic reach (e.g., Anglophone audiences, platform users following major outlets) and limits causal interpretation: while correlations with emerging narratives can be identified, it remains unclear

whether media coverage influences public responses or merely reflects them. Additionally, variations in the subject matter and the algorithms of the platform may affect the signal quality over time. Consequently, generalizability is constrained, necessitating corroboration of outputs with peer-to-peer samples and offline indicators before making substantial decisions.

6. Conclusions

This study examines BBC, CNN, CBC, and Reuters health tweets using deep learning to uncover discourse migration patterns. We use sequence VAEs to preprocess, embed, and train LSTM and CNN classifiers to quantify media-triggered online shifts with resilience and low false positives across sources, meeting current monitoring needs. The LSTM classifier outperforms all others on BBC and CNN in discrimination. The model has a validation accuracy of 98.4%, false positive rates of 1.3% to 1.4%, and precision and recall rates of 98.69% and 98.20%, respectively. CBC is 97.8% and Reuters 97.3%. CNN baselines are constantly lower while competing. Sequence VAEs show learnability in the generative environment, with Reuters at 99.6% and BBC at 98.4%, confirming signal stability. Following the comparative evidence update (Table 6), this study now excels in precision, recall, F1, FPR, and headline accuracy, obtaining 98.4% compared to 97.11% [49], 97.0% [18], and 92.59%, with a considerably high FPR of 93.92% [47]. These findings show that news-mediated health discourse incorporates machine-readable online migration markers. For operational efficiency, minimizing false positives simplifies analysts' responsibilities, improves dashboard triage, and boosts automated monitoring system trust. High recall prioritizes major changes, raising awareness. Near-real-time monitoring, real-time message adaptation, and accurate data for healthcare recommendation workflows that must adapt to changing public narratives are enabled by the metrics. This allows for proactive, evidence-based communication under online ambiguity. This work introduces a reproducible, multi-outlet pipeline with rigorous preprocessing, embeddings, sequence modules (LSTM/CNN; VAE variations), and extensive metric reporting, including FPR. Cross-source benchmarks from BBC, CNN, CBC, and Reuters show operational readiness for monitoring and supplying structured signals for discourse dynamics in healthcare recommendation systems. Future research should include multilingual streams, lightweight transformers, and explainability. The system must detect bots and coordination, validate against offline indicators, and link outputs to healthcare recommendation system intervention experiments to analyze behavior change.

Author Contributions: Conceptualization, A.A.I.; methodology, A.A.I.; software, F.A.K.; validation, A.A.I.; formal analysis, A.A.I.; investigation, A.A.I.; resources, A.A.I.; data curation, A.A.I.; writing—original draft preparation, A.A.I.; writing—review and editing, A.A.I.; visualization, F.A.K.; supervision, A.A.I.; project administration, A.A.I.; funding acquisition, N/A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data for this research is publicly available, and the source has been cited in Section 3.2.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Thapliyal, K.; Thapliyal, M.; Thapliyal, D. Social media and health communication: A review of advantages, challenges, and best practices. *Emerg. Technol. Health Lit. Med. Pract.* **2024**, 364–384. DOI: 10.4018/979-8-3693-1214-8.ch017.
2. Klimiuk, K.B.; Balwicki, Ł.W. What is infodemiology? An overview and its role in public health. *Epidemiol. Rev./Przegląd Epidemiol.* **2024**, 78. DOI: 10.32394/pe/188119.
3. Mavragani, A. Infodemiology and infoveillance: Scoping review. *J. Med. Internet Res.* **2020**, 22, e16206.
4. Zeraatkar, K.; Ahmadi, M. Trends of infodemiology studies: A scoping review. *Health Inf. Libr. J.* **2018**, 35, 91–120.
5. Li, L.; Wood, C.E.; Kostkova, P. Vaccine hesitancy and behavior change theory-based social media interventions: A systematic review. *Transl. Behav. Med.* **2022**, 12, 243–272.
6. Murthy, D. Sociology of Twitter/X: Trends, Challenges, and Future Research Directions. *Annu. Rev. Sociol.* **2024**, 50, 169–190.
7. Kostkova, P.; Szomszor, M.; St. Louis, C. # swineflu: The use of twitter as an early warning and risk communication tool in the 2009 swine flu pandemic. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2014**, 5, 8.
8. Thakur, N. Sentiment analysis and text analysis of the public discourse on Twitter about COVID-19 and MPox. *Big Data Cogn. Comput.* **2023**, 7, 116.
9. Boon-Itt, S.; Skunkan, Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modelling study. *JMIR Public Health Surveill.* **2020**, 6, e21978.
10. Lee, E.W.; Zheng, H.; Goh, D.H.; Lee, C.S.; Theng, Y.L. Examining COVID-19 tweet diffusion using an integrated social amplification of risk and issue-attention cycle framework. *Health Commun.* **2024**, 39, 493–506.
11. Wang, Y.; Hao, H.; Platt, L.S. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter. *Comput. Hum. Behav.* **2021**, 114, 106568.
12. Gui, X.; Kou, Y.; Pine, K.H.; Chen, Y. Managing uncertainty: Using social media for risk assessment during a public health crisis. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 2 May 2017; pp. 4520–4533.
13. Joshi, A.; Sparks, R.; McHugh, J.; Karimi, S.; Paris, C.; MacIntyre, C.R. Harnessing tweets for early detection of an acute disease event. *Epidemiology* **2020**, 31, 90–97.
14. Lahat, L.; Keynan, O.; Scala, F. Capitalising on Twitter for policy learning during crises: The case of the Covid-19 pandemic. *Int. J. Public Adm.* **2024**, 47, 680–689.
15. Tran, T.N.; Felfernig, A.; Trattner, C.; Holzinger, A. Recommender systems in the healthcare domain: State-of-the-art and research issues. *J. Intell. Inf. Syst.* **2021**, 57, 171–201.
16. Ivanova, M.; Raś, Z.W. Recommendation Systems in Healthcare. In *Recommender Systems for Medicine and Music*; Springer: Cham, Switzerland, 2021; pp. 1–11.
17. Sahoo, A.K.; Pradhan, C.; Barik, R.K.; Dubey, H. DeepReco: Deep learning based health recommender system using collaborative filtering. *Computation* **2019**, 7, 25.
18. Chinnasamy, P.; Wong, W.K.; Raja, A.A.; Khalaf, O.I.; Kiran, A.; Babu, J.C. Health recommendation system using deep learning-based collaborative filtering. *Heliyon* **2023**, 9, e22844.
19. Nerella, S.; Bandyopadhyay, S.; Zhang, J.; Contreras, M.; Siegel, S.; Bumin, A.; Silva, B.; Sena, J.; Shickel, B.; Bihorac, A.; et al. Transformers and large language models in healthcare: A review. *Artif. Intell. Med.* **2024**, 154, 102900. <https://doi.org/10.1016/j.artmed.2024.102900>.
20. Madan, S.; Lentzen, M.; Brandt, J.; Rueckert, D.; Hofmann-Apitius, M.; Fröhlich, H. Transformer models in biomedicine. *BMC Med. Inform. Decis. Mak.* **2024**, 24, 214.
21. Cho, H.N.; Jun, T.J.; Kim, Y.H.; Kang, H.; Ahn, I.; Gwon, H.; Kim, Y.; Seo, J.; Choi, H.; Kim, M.; et al. Task-Specific Transformer-Based Language Models in Health Care: Scoping Review. *JMIR Med. Inform.* **2024**, 12, e49724. <https://doi.org/10.2196/49724>.
22. Kbaier, D.; Kane, A.; McJury, M.; Kenny, I. Prevalence of Health Misinformation on Social Media—Challenges and Mitigation Before, During, and Beyond the COVID-19 Pandemic: Scoping Literature Review. *J. Med. Internet Res.* **2024**, 26, e38786. <https://doi.org/10.2196/38786>.
23. Ul Hussna, A.; Alam, M.G.R.; Islam, R.; Alkhamees, B.F.; Hassan, M.M.; Uddin, M.Z. Dissecting the infodemic: An in-depth analysis of COVID-19 misinformation detection on X (formerly Twitter) utilizing machine learning and deep learning techniques. *Heliyon* **2024**, 10, e37760. <https://doi.org/10.1016/j.heliyon.2024.e37760>.
24. Benítez-Andrades, J.A.; Alija-Pérez, J.-M.; Vidal, M.-E.; Pastor-Vargas, R.; García-Ordás, M.T. Traditional ML and BERT-based automatic classification of tweets about eating disorders: Algorithm development and validation. *JMIR Med. Inform.* **2022**, 10, e34492. <https://doi.org/10.2196/34492>.

25. Wu, X.; Wang, B.; Li, W. Comparing the accuracy of ANN with transformer models for sentiment analysis of tweets related to COVID-19 Pfizer vaccines. *Chaos Solitons Fractals* **2024**, *185*, 115105.
26. Suarez-Lledo, V.; Alvarez-Galvez, J. Prevalence of health misinformation on social media: Systematic review. *J. Med. Internet Res.* **2021**, *23*, e17187. <https://doi.org/10.2196/17187>.
27. Talukder, M.A.; Uddin, M.A.; Roy, S.; Ghose, P.; Sarker, S.; Khraisat, A.; Kazi, M.; Rahman, M.M.; Hakimi, M. A hybrid deep learning model for sentiment analysis of COVID-19 tweets with class balancing. *Sci. Rep.* **2025**, *15*, 27788. <https://doi.org/10.1038/s41598-025-97778-7>.
28. Rodrigues, F.; Newell, R.; Rathnaiah Babu, G.; Chatterjee, T.; Sandhu, N.K.; Gupta, L. The social media infodemic of health-related misinformation and technical solutions. *Health Policy Technol.* **2024**, *13*, 100846. <https://doi.org/10.1016/j.hlpt.2024.100846>.
29. Sinnenberg, L.; Buttenheim, A.M.; Padrez, K.; Mancheno, C.; Ungar, L.; Merchant, R.M. Twitter as a tool for health research: A systematic review. *Am. J. Public Health* **2017**, *107*, e1–e8.
30. Al-Selwi, S.M.; Hassan, M.F.; Abdulkadir, S.J.; Muneer, A.; Sumiea, E.H.; Alqushaibi, A.; Ragab, M.G. RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 102068.
31. Ben Othman, G.; Copot, D.; Yumuk, E.; Neckebroek, M.; Ionescu, C.M. Selecting Optimal Long Short-Term Memory (LSTM) Architectures for Online Estimation of Mean Arterial Pressure (MAP) in Patients Undergoing General Anesthesia. *Appl. Sci.* **2024**, *14*, 5556.
32. Ruhani, B.; Moghaddas, S.A.; Kheradmand, A. Hydrogen production via renewable-based energy system: Thermoeconomic assessment and Long Short-Term Memory (LSTM) optimization approach. *Int. J. Hydrogen Energy* **2024**, *52*, 505–519.
33. Jeong, S.H.; Lee, W.K.; Kil, H.; Jang, S.; Kim, J.H.; Kwak, Y.S. Deep Learning-Based Regional Ionospheric Total Electron Content Prediction—Long Short-Term Memory (LSTM) and Convolutional LSTM Approach. *Space Weather* **2024**, *22*, e2023SW003763.
34. Liu, C. Long short-term memory (LSTM)-based news classification model. *PLoS ONE* **2024**, *19*, e0301835.
35. Khan, F.A.; Ibrahim, A.A. Machine Learning-based Enhanced Deep Packet Inspection for IP Packet Priority Classification with Differentiated Services Code Point for Advance Network Management. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2024**, *16*, 5–12.
36. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019.
37. Ketkar, N.; Moolayil, J.; Ketkar, N.; Moolayil, J. Convolutional neural networks. In *Deep Learning with PYTHON: Learn Best Practices of Deep Learning Models with PyTorch*; Apress: New York, NY, USA, 2021; pp. 197–242.
38. Brahmi, W.; Jdey, I.; Drira, F. Exploring the role of Convolutional Neural Networks (CNN) in dental radiography segmentation: A comprehensive Systematic Literature Review. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108510.
39. Khanam, R.; Hussain, M.; Hill, R.; Allen, P. A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access* **2024**, *12*, 94250–94295.
40. Karami, A.; Gangopadhyay, A.; Zhou, B.; Kharrazi, H. Fuzzy approach topic discovery in health and medical corpora. *Int. J. Fuzzy Syst.* **2018**, *20*, 1334–1345.
41. Health Related Tweets: Tweets from Major News Papers and Magazines on Health Related Topics. Available online: <https://www.kaggle.com/datasets/prabhavsingh/health-related-tweets> (accessed on 20 February 2024).
42. Heydarian, M.; Doyle, T.E.; Samavi, R. MLCM: Multi-label confusion matrix. *IEEE Access* **2022**, *10*, 19083–19095.
43. Kim, D.; Chung, C.J.; Eom, K. Measuring online public opinion for decision making: Application of deep learning on political context. *Sustainability* **2022**, *14*, 4113.
44. Chandra, R.; Sonawane, J.; Lande, J. An Analysis of Vaccine-Related Sentiments on Twitter (X) from Development to Deployment of COVID-19 Vaccines. *Big Data Cogn. Comput.* **2024**, *8*, 186.
45. Singhal, A.; Mago, V. Exploring how healthcare organizations use Twitter: A discourse analysis. *Informatics* **2023**, *10*, 65.
46. Weng, Z.; Lin, A. Public opinion manipulation on social media: Social network analysis of Twitter bots during the COVID-19 pandemic. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16376.
47. Madni, H.A.; Umer, M.; Abuzinadah, N.; Hu, Y.-C.; Saidani, O.; Alsubai, S.; Hamdi, M.; Ashraf, I. Improving Sentiment Prediction of Textual Tweets Using Feature Fusion and Deep Machine Ensemble Model. *Electronics* **2023**, *12*, 1302. <https://doi.org/10.3390/electronics12061302>.

48. Singh, C.; Imam, T.; Wibowo, S.; Grandhi, S. A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews. *Appl. Sci.* **2022**, *12*, 3709. <https://doi.org/10.3390/app12083709>.
49. Yeasmin, N.; Mahbub, N.I.; Baowaly, M.K.; Singh, B.C.; Alom, Z.; Aung, Z.; Azim, M.A. Analysis and Prediction of User Sentiment on COVID-19 Pandemic Using Tweets. *Big Data Cogn. Comput.* **2022**, *6*, 65. <https://doi.org/10.3390/bdcc6020065>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.