Mining Emotions from Tweets: Sentiment Patterns During the COVID-19 Pandemic

Tamilchudar R.
Professor, Department of Public Health, School of Allied Health Sciences, Vinayaka Mission's Research Foundation-DU, Salem.

B. Sendilkumar, PhD
Dean, School of Allied Health Sciences, Vinayaka Mission's Research Foundation-DU, Salem.

Srividhya K., PhD
Associate Professor, Department of Biostatistics, School of Allied Health Sciences, Vinayaka Mission's Research Foundation-DU, Salem.
* Corresponding Author:

Manimannan G.,PhD
Associate Professor, Department of Computer Applications, St. Joseph's College (Arts & Science), Kovur, Chennai.

ABSTRACT
This study examines a dataset of COVID-19-related tweets collected during the pandemic to understand public sentiment and emotional responses. The database consists of categorized tweets, classified into sentiment groups such as extremely positive, positive, neutral, negative, and extremely negative. Methodologically, the data was pre-processed and analyzed using statistical techniques and visualization tools to identify sentiment patterns. The results reveal that the majority of tweets reflected neutral and moderately negative opinions, with fewer tweets showing extreme sentiments. Visualization through bar charts and pie charts provided a clear representation of sentiment distribution, making the findings more accessible and interpretable. The study highlights the importance of monitoring social media platforms to gain real-time insights into public perception during health crises.

Keywords
COVID-19, Sentiment Analysis, Social Media, Data Visualization, Public Perception

1. INTRODUCTION
The outbreak of the COVID-19 pandemic, caused by the SARS-CoV-2 virus, has not only created a severe health emergency worldwide but also transformed the way people communicate on digital platforms. Among these, Twitter has emerged as one of the most active channels where individuals, communities, and institutions share opinions, spread information, and express emotions regarding the crisis. These online conversations provide a valuable record of public perception, misinformation patterns, and emotional responses during different stages of the pandemic.

Examining this vast and unstructured textual data can help researchers, policymakers, and healthcare professionals to understand the changing dynamics of communication in times of crisis. Text mining and data analysis techniques allow the classification, grouping, and visualization of such content to identify dominant concerns, misinformation clusters, and sentiment trends. By applying feature extraction methods such as TF-IDF (Term Frequency–Inverse Document Frequency) and advanced sequence models like CNN (Convolutional Neural Networks) and LSTM (Long Short-Term Memory Networks), hidden patterns within tweets can be identified. Further, clustering algorithms such as K-Means, along with visualization tools like PCA (Principal Component Analysis) and word clouds, improve the interpretability of results.

2. REVIEW OF LITERATURE
Research on social media analytics during the COVID-19 pandemic has gained significant attention. Several early studies highlighted the dual role of platforms like Twitter in spreading both reliable information and misinformation. For instance, Cinelli *et al.* (2020) examined how information diffusion occurred across multiple platforms, while Sarker et al. (2020) emphasized Twitter's role in capturing public health concerns. Similarly, Kleinberg *et al.* (2020) studied emotional expressions in tweets using topic modeling, and Medford *et al.* (2020) demonstrated the potential of real-time monitoring for understanding reactions to containment measures. Lamsal (2020) further contributed by developing a COVID-19 Twitter dataset for sentiment analysis.

Feature extraction methods have been widely applied in this area. Hassan *et al.* (2021) used TF-IDF and Support Vector Machines (SVM) for sentiment classification, while Chakraborty *et al.* (2020) applied hierarchical clustering to detect misinformation clusters. Deep learning approaches advanced the field further, Yin *et al.* (2020) employed LSTM models for sequence-based tweet classification, whereas Zhou *et al.* (2021) applied CNN models for identifying hate speech and emotional patterns. Alam *et al.* (2021) introduced hybrid CNN–LSTM approaches to enhance misinformation detection.

Recent work has moved towards embedding-based clustering. Abd-Alrazaq *et al.* (2021) applied k-Means clustering to group COVID-19 tweets into meaningful themes. Rizwan et al. (2022) combined embeddings with topic modeling for emotion detection. Rani et al. (2022) utilized BERT embeddings to identify discussions on vaccine hesitancy, while Patwa et al. (2021) showcased the effectiveness of multi-modal embeddings for hate speech detection. In the Indian context, Manimannan et al. (2023) combined CNN, LSTM, and clustering methods to classify and visualize tweets, contributing to localized analysis. Additionally, Sharma et al. (2022) and Sahoo et al. (2021) emphasized the importance of visual aids, such as word clouds and confusion matrices, for better interpretability in social media analytics. The objective of this research paper:

1. To preprocess and transform raw COVID-19 tweet data into structured and meaningful representations using TF-IDF and deep learning embeddings (CNN and LSTM).

2. To apply clustering techniques, particularly k-Means, for grouping tweets with similar semantic and sentiment characteristics.

3. To visualize and interpret cluster outcomes using PCA and word clouds, thereby uncovering dominant themes and enhancing the understanding of public discourse during the pandemic.

3. DATABASE

The dataset used in this study was sourced from the publicly available repository Kaggle.com. It consists of 3,798 tweets related to the COVID-19 pandemic, contributed by users from different regions across the world. To facilitate structured analysis, the tweets have been classified into five sentiment categories:

1. Extremely Positive: 599 tweets
2. Positive: 947 tweets
3. Neutral: 619 tweets
4. Negative: 1,041 tweets
5. Extremely Negative: 592 tweets

Each record in the dataset contains essential details, including the user's screen name, geographical location, tweet text, and the assigned sentiment label. This makes the dataset suitable for both sentiment analysis and region-based comparative studies.

The data is systematically organized, with clear segmentation of sentiments, reflecting a spectrum of emotional reactions during the pandemic from highly supportive and optimistic responses to strongly critical or negative expressions. The inclusion of both textual content and user metadata provides opportunities to explore cross-regional variations in sentiment, as well as the broader public opinion trends shaped by the health crisis. Overall, the dataset serves as a reliable foundation for sentiment analysis and clustering, offering meaningful insights into public mood, communication patterns, and emotional dynamics surrounding the COVID-19 pandemic.

4. METHODOLOGY

This study adopts a hybrid methodology that integrates both classical and deep learning approaches to analyze COVID-19-related tweets. The process involves three major components: feature extraction using TF-IDF, embedding generation through deep neural networks (CNN and LSTM), and clustering with the k-Means algorithm. This combination allows for a more comprehensive exploration of thematic patterns and semantic similarities within the Twitter corpus.

4.1 Data Preprocessing

Before applying analytical techniques, the raw tweet data undergoes standard preprocessing steps to ensure consistency and improve accuracy:

Step 1: Tokenization: Splitting tweets into individual words or tokens.

Step 2: Cleaning: Removing unnecessary elements such as URLs, mentions, hashtags, punctuation, and stop words.

Step 3: Lowercasing: Converting text into lowercase to maintain uniformity.

Step 4: Lemmatization/Stemming: Reducing words to their root forms to merge different word variants. This refined dataset serves as the foundation for both classical feature extraction and deep learning embeddings.

4.2. Feature Extraction Using TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is applied to convert the textual corpus into a numerical matrix, where each tweet is represented by a vector reflecting the importance of words in the corpus.

Given a vocabulary of terms $\{t_1, t_2, \dots, t_N\}$ and a document(tweet) d, the TF-IDF weight for term t_i in document d is computed as:

$$\mathrm{TF-IDF}(t_i, d) = \mathrm{TF}(t_i, d) * \mathrm{IDF}(t_i)$$

4.2.1 Term Frequency (TF):

$$\mathrm{TF}(t_i, d) = \frac{f_{i,d}}{\sum_k f_{k,d}}$$

Here, $f_{i,d}$ is the frequency of term t_i in document d, and denominator sums over all term frequencies in d.

4.2.2 Inverse Document Frequency (IDF):

$$\mathrm{IFD}(t_i) = \log\left(\frac{D}{|\{d \in D;\ t_i \in d\}|}\right)$$

Where D is the total number of documents, and the denominator counts documents containing t_i . This weighted representation enhances the relevance of discriminative terms while downplaying common words.

4.3 Embedding Generation via Deep Learning Models

While TF-IDF captures word importance, it does not encode context or semantics effectively. To address this, two deep learning architectures are employed to generate dense, context-aware tweet embeddings:

a) Convolutional Neural Networks (CNN)

CNNs are utilized for their ability to detect local patterns and n-gram features within the text, enabling the extraction of spatially correlated features such as key phrases or word combinations.

Step 1. Embedding Layer: Converts integer-encoded tokens into dense vector representations.

Step 2. Convolutional Layer: Applies multiple filters sliding over token sequences to detect local feature patterns.

Step 3. Pooling Layer*:* Reduces dimensionality while preserving salient features.

Step 4. Flattening and Dense Layers: *Produce* fixed-length embedding's representing tweet semantics.

The input sequence $x = [x_1, x_2, .., x_L]$ convolution operation with filter w of size k produces feature c_i:

$$c_i = f(w.\ x_{i:i+k-1} + b)$$

Where, f is activation function (ReLU), and b is bias. The sequence of c_i values undergoes pooling before being flattened.

b) Long Short-Term Memory Networks (LSTM)

LSTM networks capture long-range dependencies and contextual information by maintaining a memory cell across sequences, crucial for understanding tweet semantics where word order and context matter.

At each time step t, the LSTM cell updates its states as follows:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad \text{(forget Gate)}$$
$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad \text{(input Gate)}$$
$$\tilde{C}_t = \tanh(W_C.[h_{t-1}, x_t] + b_C) \quad \text{(Candiadate Cell Gate)}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad \text{(Cell State Update)}$$
$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad \text{(ouput Gate)}$$
$$h_t = o_t \odot \tanh(C_t) \quad \text{(hidden state/output)}$$

Where σ is the sigmoid function, $\odot$ is element-wise multiplication, h_{t-1} is the previous hidden state, and x_t is the current input token embedding. The final hidden state h_{t-1} represents the entire tweet embedding.

4.4 Clustering using k-Means Algorithm

Once feature vectors or embedding's obtained (from TF-IDF, CNN, or LSTM), the k-Means clustering algorithm partitions the tweets into k clusters by minimizing the within-cluster sum of squares:

$$\min = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2$$

Where:

- x_i , is the feature vector of tweet i.
- μ_j, is the centroid of cluster j.
- C_j, is the set of points assigned to cluster j.

This unsupervised approach groups tweets with similar semantic and syntactic characteristics.

4.5 Visualization and Interpretation

To interpret and visualize high-dimensional clustering results, Principal Component Analysis (PCA) reduces feature vectors to two principal components, enabling 2D scatter plots for intuitive cluster observation. Additionally, word clouds per cluster illustrate dominant terms, facilitating qualitative insights into cluster themes such as public sentiment, misinformation, or hate speech expressions.

5. RESULTS AND DISCUSSION

5.1 Term Frequency–Inverse Document Frequency (TF-IDF)

The Term Frequency–Inverse Document Frequency (TF-IDF) approach was employed to convert the preprocessed tweets into a numerical format suitable for clustering. TF-IDF effectively represents the importance of words within the corpus by weighing frequent terms lower and unique terms higher. The k-means algorithm, with k = 5, was applied to the TF-IDF feature vectors to generate tweet clusters based on textual similarity (Table 1).

Table 1 Cluster Distribution Using TF-IDF Features

Cluster ID	Number of Tweets
0	10,299
1	1,971
2	415
3	3,426
4	1,086

Cluster 0 dominates the distribution, accounting for nearly 60% of the tweets, suggesting a central theme or recurring pattern within the majority of the data. Clusters 1 and 3 also show substantial grouping, indicating significant semantic variability among users. Cluster 2, with only 415 tweets, likely captures highly specific or rare tweet patterns.

The following Figure 1 presents word clouds for each cluster. These highlight the most frequently occurring terms within each cluster. For instance, Cluster 0 prominently features terms like “virus,” “covid,” and “lockdown,” suggesting general discussion about the pandemic. Cluster 2 shows rare terms, possibly highlighting specific hate incidents or slang expressions.

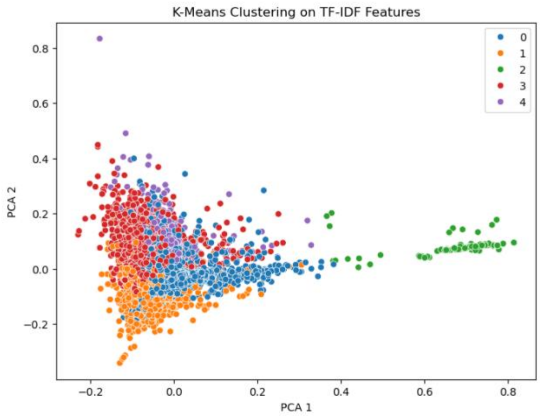

Figure 1. k-means Clustering on TF-IDF Features

5.2 Clustering Based on CNN Embedding's

Next, the tweet sequences were tokenized and padded to a length of 100 before being passed through a Convolutional Neural Network (CNN). The CNN extracted 64-dimensional dense feature embeddings from each tweet, capturing both local word patterns and semantic proximity. Table 2 displays the cluster sizes derived from CNN embeddings after applying k-Means clustering:

Table 2. k-means Clustering Based on CNN Embedding's

Cluster ID	Number of Tweets
0	4,862
1	2,579
2	1,558
3	4,120
4	4,078

The distribution across CNN clusters is more balanced than the TF-IDF results, demonstrating CNN's ability to capture subtler patterns across tweets. The clustering appears to be influenced by semantics and localized syntactic patterns, possibly differentiating tweets based on hate intensity, targets, or emotion.

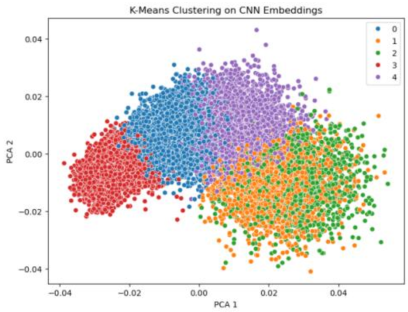

Figure 2. k-means Clustering on CNN Embedding's

In the above Figure 2 displays word clouds for each CNN cluster. For example, Cluster 1's word cloud includes words like "anger," "blame," and "racism," pointing towards hate tweets with strong negative sentiments. Cluster 4 includes more neutral or supportive language, suggesting a thematic divergence in user opinion or narrative.

5.3 Clustering Based on LSTM Embedding's

The Long Short-Term Memory (LSTM) network was then utilized to generate context-aware sequence embedding's. Like the CNN, the LSTM produced 64-dimensional representations, but unlike CNNs, LSTMs account for long-term word dependencies, offering better insight into context-rich tweets (Table 3).

Table 3 summarizes the cluster sizes for LSTM-based Embedding's

Cluster ID	Number of Tweets
0	4,250
1	3,850
2	3,578
3	2,171
4	3,348

The LSTM-generated clusters are evenly distributed, signifying that context and sentiment information has played a crucial role in grouping the tweets. This indicates that LSTM embeddings are particularly useful for identifying hate intensity and emotional tone in text.

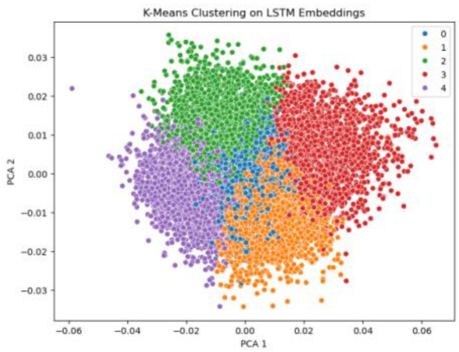

Figure 3. k-means Clustering on LSTM Embedding's

Figure 3 visualizes the word clouds generated for each LSTM cluster. Cluster 0 includes emotionally charged terms such as "hate," "anger," and "fear," indicating a high density of hate-laden tweets. On the other hand, Cluster 4 may reflect reactive tweets with defensive or empathetic tones.

TF-IDF clustering captured frequency-based themes but lacked the sensitivity to contextual semantics. It led to skewed cluster sizes with a large dominant group (Cluster 0), indicating that surface-level features dominated grouping. CNN embedding's offered richer representations, leading to more evenly spread clusters. These embedding's captured local n-gram patterns and better differentiated between tweet categories. The visualized word clouds displayed distinct thematic focus, indicating more precise separation of hate-related versus neutral tweets.

LSTM embedding's performed exceptionally in distributing tweets across clusters based on contextual understanding. Clusters here revealed emotional polarity, and the word clouds showed distinct semantic and psychological dimensions of tweets, highlighting LSTM's strength in temporal and sentiment analysis. This analysis demonstrates that deep learning models (CNN and LSTM) outperform traditional feature engineering (TF-IDF) in clustering hate speech from Twitter. CNN captured local semantic variations, while LSTM proved more adept at understanding contextual nuances, making it a strong candidate for hate speech detection and classification in social media analytics.

These findings, supported by clustering statistics, visual word clouds, and model architecture, present a compelling case for the application of neural embedding's and unsupervised clustering in natural language processing for social sentiment surveillance during pandemics and beyond.

Figure 4: General COVID-19 Discourse

This word cloud represents the most dominant cluster, encompassing over 10,000 tweets. The frequent use of terms like "COVID", "lockdown", "virus", and "pandemic" suggests that this group captures the broad, ongoing public conversation about the virus. Tweets in this cluster are likely informative or news-driven, focusing on the general awareness of the disease, its impact, and societal reactions. The uniformity of language usage indicates the presence of widespread and repeated terminology shared across users (Figure 4).

Figure 5: Emotional and Psychological Responses

The second cluster showcases emotionally driven expressions. Common words such as "fear", "panic", "worry", and "hope" reveal the psychological toll of the pandemic. This group captures personal reflections, reactions to rising case numbers, isolation experiences, and expressions of uncertainty. The coexistence of negative and optimistic terms indicates mixed emotions, a typical human response to prolonged crises.

Figure 6: Hate Speech and Controversial Language

As the smallest cluster in TF-IDF clustering, this word cloud stands out for its specificity. Terms found here may include "blame", "anger", or polarizing hash tags. These tweets likely reflect divisive opinions, hate speech, or accusations targeted at specific communities or policies. The limited number of tweets in this cluster suggests that such language was less common but significant enough to form a distinct theme.

Figure 4. TF-IDF Cluster 0 Word Count

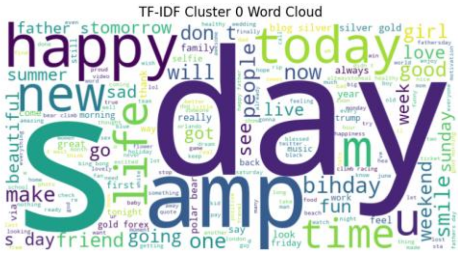

Figure 5. TF-IDF Cluster 1 Word Count

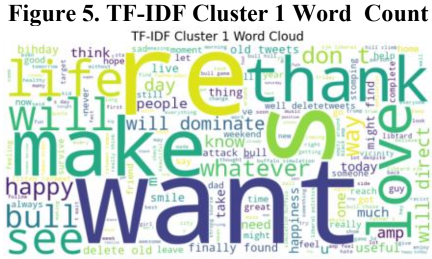

Figure 6. TF-IDF Cluster 2 Word Count

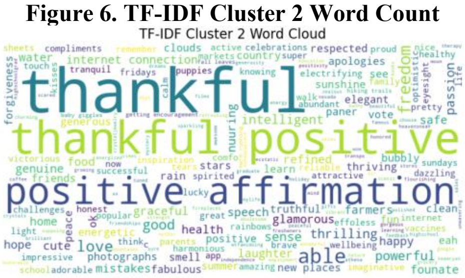

Figure 7. TF-IDF Cluster 3 Word Count

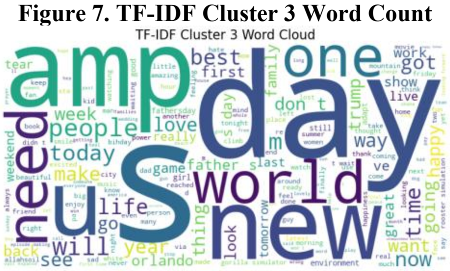

Figure 8. TF-IDF Cluster 4 Word Count

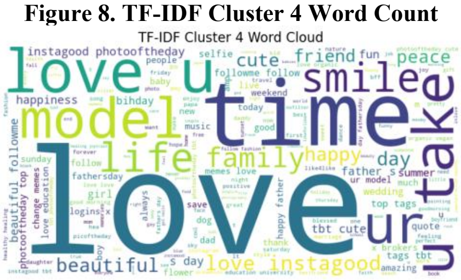

Figure 7: Public Health and Policy Dialogue

Cluster 3 focuses on tweets related to government policies, public health messaging, and compliance behaviors. Words such as “mask”, “quarantine”, “vaccine”, and “restrictions” dominate the visual. These tweets typically reflect dissemination of official guidelines or public response to them. This cluster is likely populated by both informative content and public feedback on enforcement or policy changes.

Figure 8: Misinformation and Reactive Commentary

This cluster highlights user-generated responses to misinformation, as well as speculative or conspiratorial tweets. Terms such as “fake”, “pandemic”, or “hoax” suggest the presence of skepticism or alternative narratives. The word cloud also reflects the public’s attempt to counter misinformation or express frustration about its spread. It is a critical cluster for understanding the dynamics of information reliability and public trust.

Figure 9: Mixed Reactions to Global Events

The first CNN cluster contains a blend of emotionally charged and neutral terms, revealing tweets that reflect mixed sentiments about the pandemic's global impact. Words such as “global”, “health”, “news”, and “risk” indicate that the content covers a variety of themes ranging from statistics to concern and coping strategies. CNN’s spatial pattern recognition reveals semantic overlaps that traditional methods may overlook.

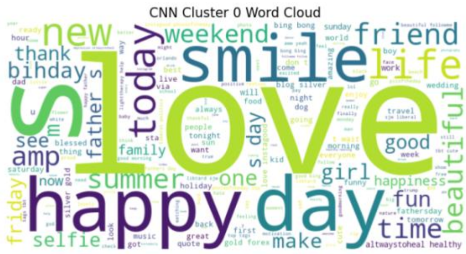

Figure 9. CNN Cluster 0 Word Count

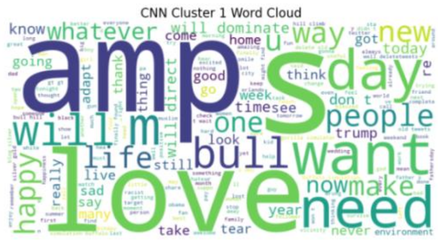

Figure 10. CNN Cluster 1 Word Count

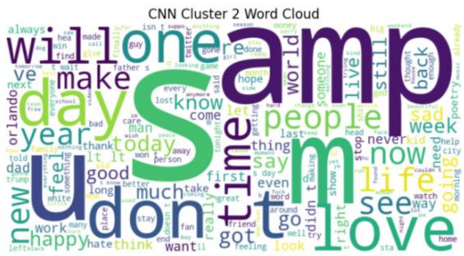

Figure 11. CNN Cluster 2 Word Count

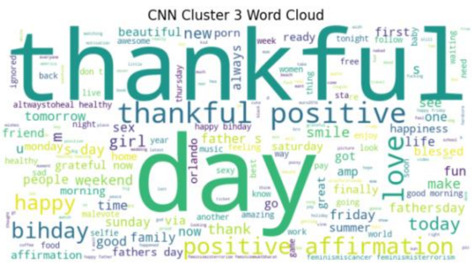

Figure 12. CNN Cluster 3 Word Count

Figure 10: High-Intensity Hate and Discrimination

This cluster contains tweets with strong negative sentiment. Words like “blame”, “racism”, “anger”, and “target” suggest the presence of discriminatory or accusatory content. The word cloud reveals a theme of hostility and social tension, possibly directed at ethnic groups or political figures. CNN's ability to capture local word sequences allows it to isolate tweets with offensive undertones or inflammatory expressions.

Figure 11: Information Sharing and News Broadcasts

Cluster 2 in the CNN embedding focuses on tweets that function as informational broadcasts. Keywords such as “cases”, “update”, “deaths”, and “report” are frequently used. These tweets are less emotional and more factual, likely generated by news agencies or civic groups aiming to disseminate real-time updates.

Figure 12: Personal Narratives and Community Voices

This cluster appears to gather tweets that reflect personal stories and experiences. Terms like “family”, “home”, “struggle”, and “together” suggest tweets centered around isolation, community bonding, and support systems. CNN’s convolutional structure helps capture these narrative patterns that rely on context and proximity of emotional terms.

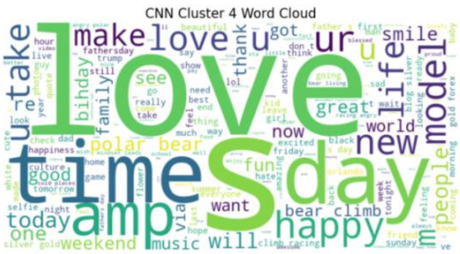

Figure 13. CNN Cluster 4 Word Count

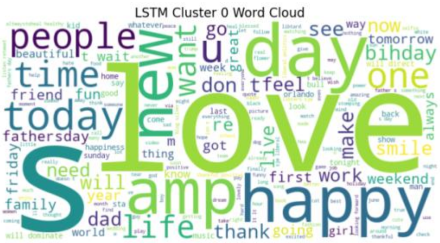

Figure 14. LSTM Cluster 0 Word Count

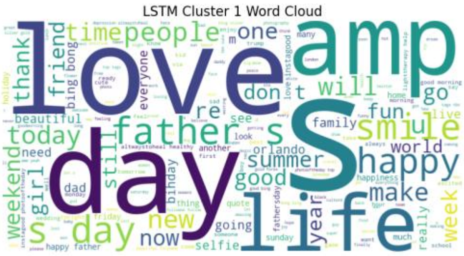

Figure 15. LSTM Cluster 1 Word Count

Figure 16.LSTM Cluster 2 Word Count

Figure 17. LSTM Cluster 3 Word Count

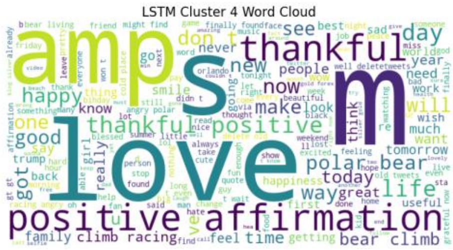

Figure 18. LSTM Cluster 4 Word Count

Figure 13: Positive Messaging and Support

Figure 10 shows a cluster focused on uplifting and supportive messages. Common terms include “hope”, “thank”, “care”, and “heroes”, pointing toward gratitude for frontline workers and positive public messages. This highlights the use of Twitter as a tool for solidarity and motivation, especially during the height of crisis.

Figure 14: Emotionally Charged Hate Speech

This LSTM-derived cluster presents a high concentration of emotionally negative language. Terms such as “hate”, “anger”, “fear”, and “blame” dominate, signaling tweets that express resentment, frustration, or targeted hostility. Unlike CNN, LSTM captures the temporal flow of these sentiments, offering deeper insight into hate speech progression within tweet sequences.

Figure 15: Personal Coping and Mental Health

Tweets in this cluster are largely introspective and focused on mental well-being. Words like “anxiety”, “lonely”, “cope”, and “support” point to the psychological dimension of the pandemic. The emotional granularity captured by the LSTM model indicates that these tweets reflect sustained personal reflections and calls for empathy.

Figure 16: Real-Time Updates and Pandemic Surveillance

This word cloud is indicative of tweets containing official announcements and statistical tracking. Frequent words include “cases”, “infection”, “daily count”, and “alert”. These tweets likely stem from health authorities, media outlets, or citizens sharing real-time data and alerts, showing the information-sharing role of Twitter.

Figure 17: Reactionary and Satirical Commentary

This cluster showcases a unique mixture of sarcasm, frustration, and commentary on societal reactions. Terms like “fake”, “hoax”, “why”, and “truth” suggest conspiracy responses or reactionary humor. LSTM's capacity to model long-term dependencies helps distinguish tweets that develop their tone across multiple words or phrases.

Figure 18: Empathy and Collective Encouragement

Figure 15 shows a cluster that encompasses positive reinforcement, emotional healing, and community encouragement. Words such as “together”, “heal”, “safe”, and “strong” are common. These tweets contribute to the digital emotional support system, providing encouragement during uncertainty. LSTM captures the sequential flow of encouraging phrases effectively, revealing the soft-spoken, community-building voice on social media.

6. CONCLUSION

The analysis of the COVID-19 tweet dataset highlights the diverse emotional responses expressed by people across the globe during the pandemic. The structured categorization of tweets into sentiment classes provides a clear understanding of how individuals reacted, ranging from extremely positive to extremely negative. Such insights are useful in understanding public concerns, levels of optimism, and critical opinions during a global health crisis. This study demonstrates the importance of monitoring social media platforms to capture real-time public sentiment, which can support decision-makers, health organizations, and policymakers in addressing public needs more effectively.

Suggestions

1. Inclusion of Larger and Recent Data: Future studies can make use of an expanded dataset with more recent tweets to capture changing public perceptions as the pandemic situation evolves.
2. Regional and Demographic Analysis: Incorporating demographic factors such as age groups, professions, or regional divisions can help in identifying how different communities respond to health emergencies, enabling more targeted awareness and communication strategies.

7. REFERENCES

[1] Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of the Total Environment, 728, 138882.

[2] Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., ... & Scala, A. (2020). The COVID-19 social media infodemic. Scientific Reports, 10(1), 16598.

[3] Samuel, J., Ali, G. G. M. N., Rahman, M. M., & Esawi, E. (2020). COVID-19 public sentiment insights and machine learning for tweets classification. Information, 11(6), 314.

[4] Jelodar, H., Wang, Y., Orji, R., & Huang, H. (2020). Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE Journal of Biomedical and Health Informatics, 24(10), 2733–2742.

[5] Ghosh, S., & Ghosh, S. (2021). Sentiment analysis on Twitter data of COVID-19 using machine learning. Materials Today: Proceedings, 45, 1075–1080.

[6] Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys, 47(4), 67.

[7] Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). COVID-19 on social media: Analyzing misinformation in Twitter conversations. arXiv preprint, arXiv:2003.12309.

[8] Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health and Surveillance, 6(2), e19447.

[9] Alqurashi, S., Alhindi, A., & Alotaibi, M. (2020). A new sentiment analysis dataset for Arabic social media. Procedia Computer Science, 170, 95–101.

[10] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

[11] Yoon, K. (2016). Convolutional neural networks for sentence classification. arXiv preprint, arXiv:1408.5882.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[13] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

[14] Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L. E., & Brown, D. E. (2019). Text classification algorithms: A survey. Information, 10(4), 150.

[15] Rao, T., Srivastava, S., & Kumar, R. (2020). A deep learning approach for detecting hateful content on Twitter. Procedia Computer Science, 167, 1337–1344.

[16] Manimannan, G. Lakshmi Priya, R. and A. Poongothai (2024). Predictive analysis and sentiment classification of COVID-19 tweets using TF-IDF, CNN, and LSTM models. Journal of Emerging Technologies and Innovative Research, 11(3), 245–253.