

From Fear to Forecast: Machine Learning Solutions for Mass Psychosis in Crisis Scenarios

Raahil Sheikh¹, Lavleen Sharma², Disha Singh³, Pragyan Priyadarshini Rout², Shreya Das², Tejaswini Yadav⁴, 

1. *Space Science, Beyond Space, Mumbai, IND*
2. *Space Science, Beyond Space, Chandigarh, IND*
3. *Computer Science Engineering, Beyond Space, Mohali, IND*
4. *Aerospace Engineering, Beyond Space, Kanpur, IND*

Received: March 13, 2026 | Review began: April 07, 2026 | Review ended: May 13, 2026 | Published: May 27, 2026

© **Copyright** 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Mass hysteria, also known as mass psychogenic illness is referred to as a quick outbreak of fear, anxiety, or unreasonable behavior in many people at once, especially during an epidemic. The COVID-19 pandemic triggered a global climate of fear, uncertainty, exacerbated by a significant surge in misinformation. Forecasting and controlling such circumstances help reduce the effects and enhance public health measures. This research focuses on using machine learning (ML) to predict and understand mass hysteria in pandemic scenarios that may occur in future. The researchers analyze and observe large datasets, for example, social media posts, search engine data, health statistics, and news trends to identify patterns that lead to collective panic with techniques like natural language processing (NLP), time-series forecasting, and sentiment analysis which are used to spot early warning signs. It also monitors how hysteria develops over time to control it in the future. NLP techniques can analyze emotions and concerns expressed in text. The study also examines how misinformation, social factors, and demographic variables contribute to mass hysteria. This work combines ML models with behavioral and public health research to provide insights that can help policymakers, health organizations, and media platforms take action to reduce panic and build trust during future crises. This research thus underlines the potential of ML frameworks in solving the psychological and social challenges of pandemics, but it also puts across that preparation for, and management of a global emergency are very much needed in working across disciplines.

Categories: AI applications, Machine Learning (ML), Medical Expert systems

Keywords: covid-19, machine learning, mass hysteria, natural language processing, pandemic

Introduction

Mass hysteria, often termed as mass psychogenic illness (MPI), is characterized as the rapid spread of anxiety-driven symptoms, fear, or irrational behavior among groups of individuals in the absence of identifiable organic causes. The COVID-19 pandemic demonstrated how unprecedented public health crises can trigger collective emotional responses amplified by global communication networks. For instance, by 25 March 2020, India declared a nationwide lockdown spanning 21 days to limit transmission across its 1.4 billion population [1].

How to cite this article:

Sheikh R, Sharma L, Singh D, et al. (May 27, 2026) From Fear to Forecast: Machine Learning Solutions for Mass Psychosis in Crisis Scenarios. *Cureus J Comput Sci* 3 : es44389-026-00085-z. DOI <https://doi.org/10.7759/s44389-026-00085-z>

During the subsequent months, citizens faced restricted mobility, financial uncertainty, widespread isolation, and disruption of social systems, leading to heightened levels of panic, anxiety, loneliness, and depression [2]. Media coverage displaying continuous case counts and mortality figures further amplified public distress, while social media emerged as a significant vector for misinformation. The resultant exposure to sensationalized content and alarming imagery contributed to behavioral responses such as panic buying, stockpiling, and hoarding of medical supplies, placing strain on essential resource chains and aggravating societal tensions. Parallely, certain communities exhibited protests and unrest against preventive measures including mask mandates and vaccination campaigns, reflecting how collective stress can escalate into disruptive socio-political action during crises [3].

This scenario illustrates a classic nocebo-driven psychosocial cascade, wherein heightened fear triggers psychogenic symptoms and behavioral disturbances at scale. Research confirms that MPI commonly spreads through verbal or visual transmission, inducing symptoms in the absence of biomedical pathology. In modern societies, digital social platforms have replaced localized verbal contagion with high-velocity information propagation, enabling MPI-related content ranging from rumors to visibly distressing imagery to transcend geographical boundaries instantaneously. Despite its psychological roots, MPI lacks standardized medical treatment and is instead understood as a psychosocial response emerging from perceived threat environments combined with stress, uncertainty, and social vulnerability [4,5].

How to cite this article:

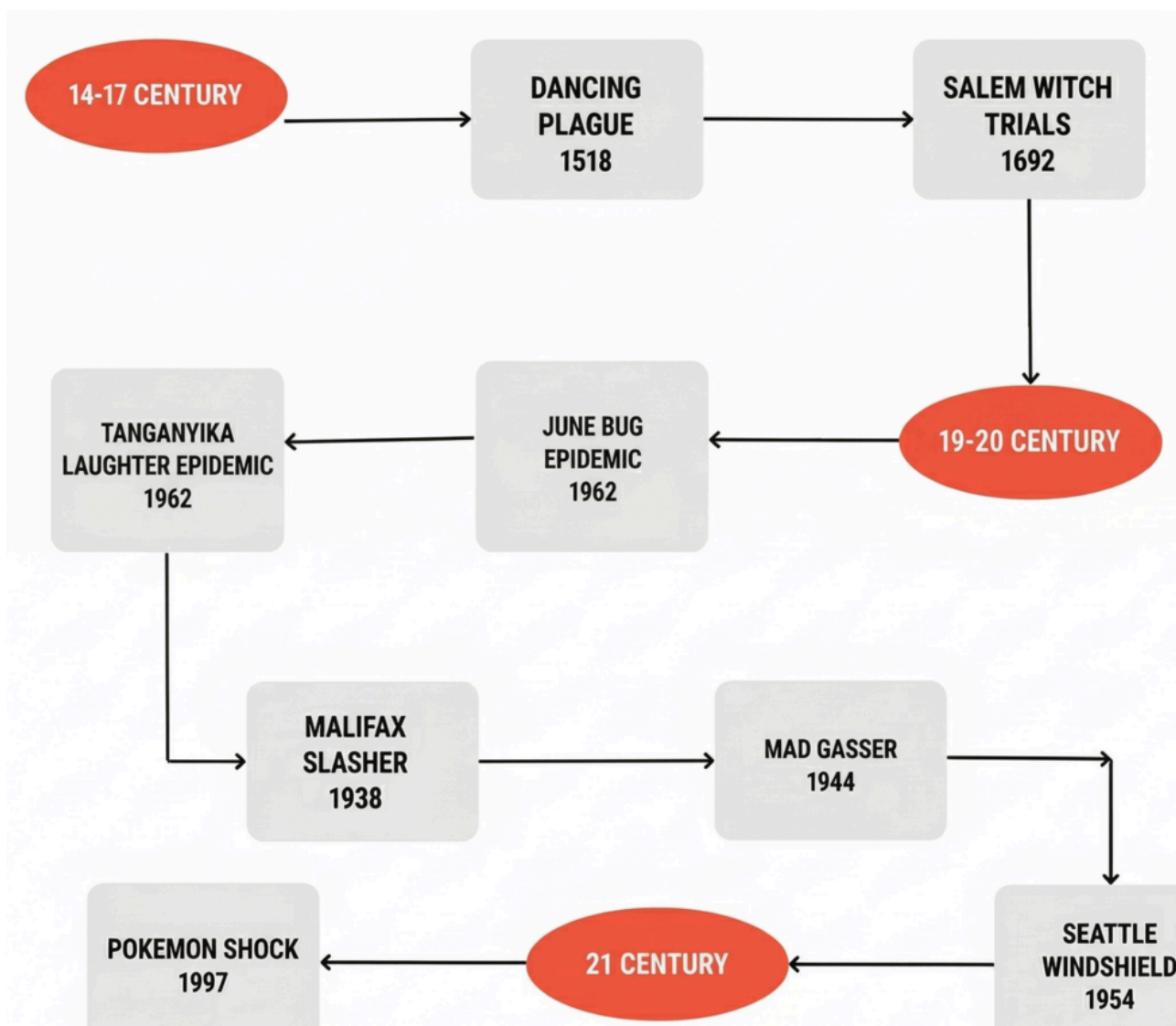


FIGURE 1: Timeline of previous notable instances of mass hysteria.

As illustrated in Figure 7, the notable incidents of mass hysteria throughout history have aggravated panic and widespread anxiety amongst the people. MPI was first documented in the 15th century when mass motor outbreaks were often misattributed to demonic possessions, i.e., the Salem Witch Trials being the classic example of this [6,7,8]. Modern literature now categorizes MPI into physical, psychological, and behavioral. Physical symptoms include fainting and tremors, emerging due to high stress and line-of-sight transmissions. Psychological symptoms include anxiety and paranoia, often propagated through direct observations. Behavioral symptoms such as the social contagion of the 1518 Dancing Plague often serve as a catalyst for escalation [9,10,11].

In the digital era, platforms such as X, formerly known as Twitter, accelerate the spread of fear and misinformation during crises, making MPI harder to detect and manage through traditional public health frameworks. The key problem addressed in this study is how to identify and predict mass hysteria in real time using machine learning (ML) and natural language processing (NLP) applied to social media data [12]. Recent longitudinal studies, such as that of the Digital Mass Hysteria framework, where information overload and the shock of a new disease created a vacuum of panic, which was filled by the social media users' panic [9].

How to cite this article:

Another study focused on the cascading-resonance model which highlighted that the negative emotion transfer operates through a mimicry and transition process. Fear has been the long-term driver of collective experiences, whilst anger accelerates and triggers immediate disruptive behavior [13]. ML models have been implemented for the diagnosis of mental health by integrating acoustic or physiological data with social media texts to increase model accuracy. Convolutional Neural Networks (CNN), Random Forest (RF), Support Vector Machine (SVM), Deep Neural Networks, and Extreme Learning Machine (ELM) are prominent models for predicting mental health conditions [14].

While prior work has examined pandemic-related sentiment trends, misinformation dynamics, and psychological effects, most studies are retrospective, lack predictive modeling, and do not integrate behavioral theory with temporal forecasting or practical decision-support mechanisms [15].

This study specifically examines social media data, with a focus on X, where public sentiment and reactions to crises unfold in real-time. Through sentiment analysis, trend monitoring, and behavioral data evaluation, we seek to identify early warning signs of mass hysteria and anticipate its escalation. The primary objectives of this study include:

- Developing an ML-based model to detect early signs of mass hysteria.
- Examining the impact of misinformation, fear, and anxiety on the spread of hysteria through social media.
- Utilizing NLP techniques to assess public sentiment and predict hysteria outbreaks.
- Providing actionable insights to support effective crisis management and public health communication.

Materials And Methods

The methodological framework of this research integrates NLP, ML, and time-series analysis to study patterns of emotional contagion, misinformation, and panic-driven discourse on social media platforms during crisis situations. The approach is structured into six interdependent components, namely data collection, data preprocessing, feature engineering, model selection, experimental setup, and performance evaluation.

Data collection

Previous studies have been able to identify emotion labels from social texts and link verbal and non-verbal representations through referential activity. Due to the current ethical considerations and user privacy concerns, X was selected as the primary data source because of its role as a high-velocity information channel during emergencies, where users actively share opinions, emotions, and experiences in real time.

Data were retrieved through the CML-COVID dataset, an X dataset of roughly 20 million tweets from roughly 6 lakh users. used primarily for Sentiment Analysis and NLP Applications. These were collected during the months of March 2020 and July 2020 with the help of trigger words relevant to the pandemic. Query terms included crisis-related expressions linked to pandemics, supply shortages, fear-driven behaviors, and misinformation narratives [16].

For each tweet, metadata fields such as timestamps, engagement counts (likes, retweets, and replies), and user information were collected to enable temporal and diffusion-based modelling. Strict ethical considerations were maintained throughout the data collection process by limiting extraction to publicly available content and removing personally identifiable user attributes not required for research.

Data preprocessing

Since raw X data are highly noisy and unstructured, a multi-stage preprocessing procedure was implemented to transform the corpus into standardized machine-interpretable text. The tweets were first segmented into lexical units through tokenization, enabling linguistic and syntactic analysis at a granular level. Stop-words, including frequently occurring function words that do not convey semantic meaning, were removed to enhance the discriminative power of textual features.

How to cite this article:

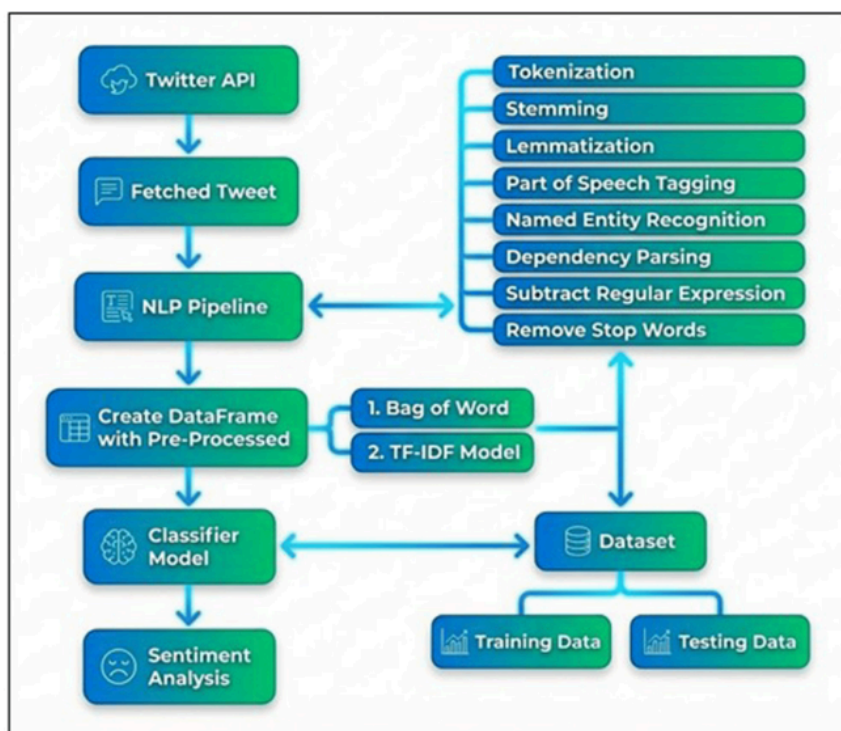


FIGURE 2: Sentimental analysis pipeline for X API.

TF-IDF, Term frequency inverse document frequency; API, application programming interface.

This part will provide emotive and textual analysis of tweets relating to COVID-19, as well as ML models, NLP, and X. To manage X data effectively, there are a few major problems that are explicitly illustrated in Figure 2. A multi-stage NLP pipeline designed to transform unstructured social tweets into a standardized ML-interpretable format. The pipeline begins with data acquisition from X API, followed by preprocessing steps of tokenization, stop-word removal and lemmatization, which was applied to normalize morphological variations by converting words to their canonical base forms, thereby reducing sparsity in the textual representation. To preserve semantic integrity, the pipeline integrates Named Entity Recognition and Part-of-Speech tagging for crisis-specific context. Finally, the refined text is converted into vectors using (TF-IDF) for high-fidelity features for the downstream sentimental analysis and ML models tasked with detecting markers of mass psychogenic distress.

Feature engineering

Following preprocessing, feature engineering was undertaken to capture semantic, temporal, and behavioral attributes associated with mass hysteria expressions. Semantic attributes were captured by transforming cleaned text into high-dimensional vectors using vectorization schemes such as TF-IDF and distributed word embeddings. To capture deeper context, Word2Vec embeddings were utilized to map semantic relationships between synonymous anxiety-driven terms. These representations allow ML models to interpret context, similarity, and emphasis within crisis-related language.

In addition to textual semantics, network diffusion attributes such as retweets, likes, and replies were incorporated to contextualize the amplification effects of emotionally charged or misleading content. The retweet-to-time ratio was calculated to achieve rapid distribution within the first hour of posting. An engagement intensity was calculated using the interaction index by combining the likes, retweets and replies (replies had more weight).

This was done to identify high-resonance content that signals the onset of emotional contagion. Such engagement metrics often reflect the contagiousness of panic narratives and the extent of public resonance.

How to cite this article:

Temporal attributes derived from tweet timestamps were also included as a parameter to enable time-series modeling of sentiment evolution across critical phases of a crisis. The integration of textual, network-based, and temporal features provides a holistic representation of how panic-influencing content is generated, propagated, and transformed on social media.

Model selection

Multiple ML and deep learning models were employed to evaluate their ability to classify sentiment and detect panic-related communication patterns with high fidelity. Traditional models such as Logistic Regression (LR) and RF were selected due to their robustness, interpretability, and strong baseline performance on sparse feature spaces generated by TF-IDF vectors.

LR was employed as a baseline linear classifier. It was used to establish the probabilistic relationship between specific TF-IDF weighted tokens (e.g., "shortage," "panic") and the target sentiment classes. Given its higher interpretability, LR allowed for the identification of the most significant linguistic predictors of distress. RF was implemented to handle non-linear relationships within the sparse feature space. By utilizing an ensemble of 100 decision trees with Gini impurity as the split criterion, the RF model effectively captured complex interactions between network engagement metrics (retweets/likes) and textual sentiment that a linear model might overlook.

To model sequential dependencies and contextual variations inherent in crisis communication, the long short-term memory (LSTM) networks were implemented, allowing the system to capture temporal relationships within tweet sequences. Additionally, transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT) were incorporated due to their capacity to process bidirectional linguistic context and handle semantic ambiguity typical of social media discourse.

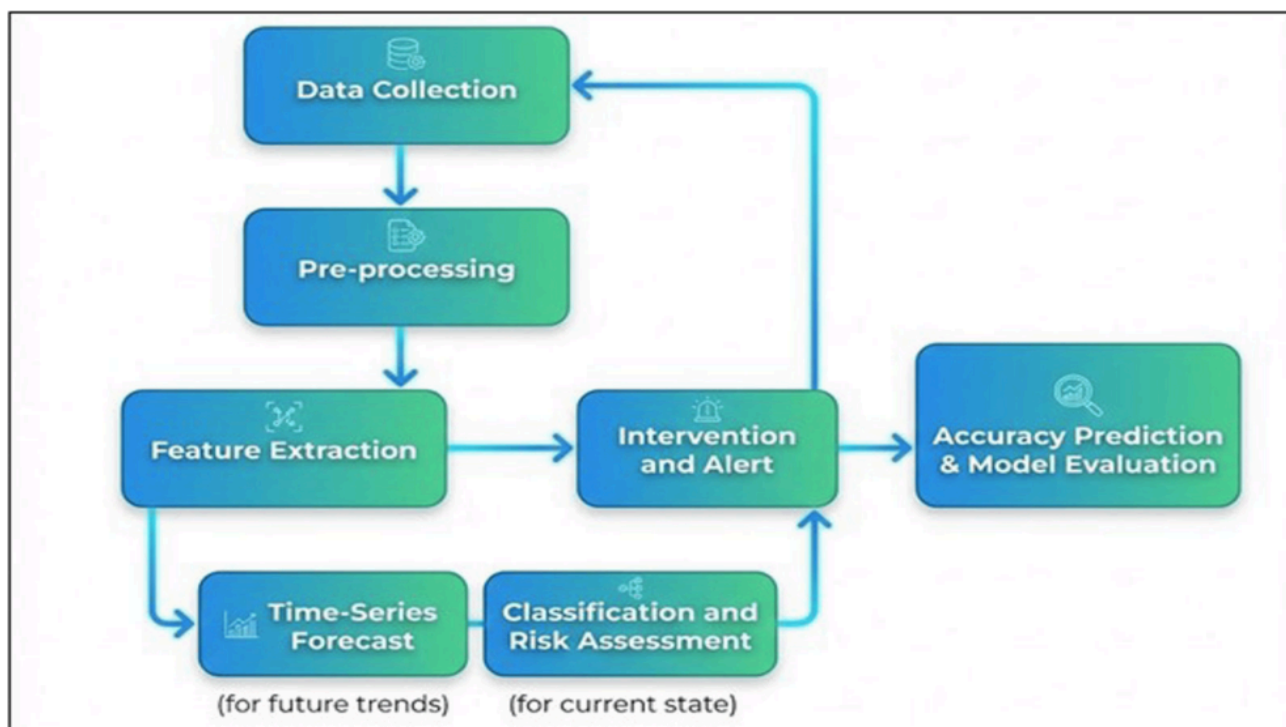


FIGURE 3: Flowchart for machine learning pipeline for mass hysteria prediction.

How to cite this article:

Figure 3 represents the ML pipeline opted for in the Mass Hysteria Prediction. With the help of real-time data from social media, it is further preprocessed by cleaning the text data and converting it into numerical features. The use of models spanning classical statistical learning, recurrent neural networks, and transformer architectures enabled a comprehensive comparative assessment of algorithmic suitability for crisis informatics applications.

Using fundamental sentiment analysis tools like VADER, the emotional intensity is predicted by extracting features like social network patterns. With the help of Time-Series Forecasting, the trends are analyzed in sentiment fluctuations and LSTM to predict panic escalation. With the aid of this pipeline, the risk levels can be classified, and the output scores shall indicate the likelihood of mass hysteria events. If risk crosses a threshold, the authorities are alerted to suppress and mitigate the panic.

Experimental setup

Our workflow was structured within Google Colab, using Python 3.10 (GPU: NVIDIA T4, 16GB RAM), providing an efficient cloud-based environment for data processing, model training, and evaluation. To visualize trends and insights effectively, we employed Matplotlib, enabling us to generate detailed graphical representations of sentiment shifts and engagement levels over time.

For statistical analysis and data preprocessing, Scipy played a crucial role in handling numerical computations, filtering noise, and optimizing model parameters. Additionally, Risk Prognosis Assessment and Prediction Evaluation (RPAPE) was integral to our model selection process, allowing us to systematically evaluate various machine learning architectures to determine the most effective model for predicting mass psychosis trends. RPAPE focuses on the prognostic reliability of the system, focusing on its ability to differentiate between transient sentiment fluctuations and sustained risk signals to precede MPI events. It assessed the time interval between an ML-detected sentiment spike and the actual surge in misinformation.

By comparing multiple algorithms, fine-tuning hyperparameters, and assessing performance metrics, we ensured the robustness and reliability of our approach. SMOTE oversampling was included in the pipeline to ensure minority class balance and better generalization over all the classes. Unlike standard oversampling, which indicates minority instances and risk of overfitting, SMOTE generates synthetic examples by identifying the k-nearest neighbors for a minority class point. By increasing the representation of high-risk psychogenic labels to match the majority of neutral classes, SMOTE prevents in developing a major-class bias.

In the optimized pipeline, the dataset was partitioned into an 80/20 train-test split, with SMOTE oversampling applied to the training set to ensure a balanced representation of acute panic labels. The deep learning architecture comprised an LSTM network with a 64-dimensional embedding layer, trained over 10 epochs with a batch size of 64 to capture long-range temporal dependencies. For linguistic nuance, BERT was fine-tuned over 3 epochs to adapt its bidirectional encoders to pandemic-specific semantics. These were integrated with a (2,1,2) ARIMA model to stabilize the time-series forecasting of sentiment fluctuations.

Evaluation metrics

To ensure a rigorous assessment of the classification and prediction capabilities of the selected models, multiple performance metrics were utilized rather than relying solely on overall accuracy. Accuracy provided a general measure of the proportion of correctly identified labels across the dataset, while precision measured the reliability of positive predictions and therefore the system's ability to minimize false alarms.

Recall quantified the model's capability to detect true positive signals, a critical factor when attempting to capture panic-inducing content that may appear less frequently yet holds greater importance in crisis monitoring. The F1-score, defined as the harmonic mean of precision and recall, offered a balanced evaluation particularly suited for imbalanced datasets.

How to cite this article:

Parameter	Value
Train/Test split	80%/20%
SMOTE oversampling	Yes (minority class balanced)
LSTM embedding dim	64
LSTM epochs/batch	10/64
BERT fine-tuning epochs	3
ARIMA order	(2,1,2)
Evaluation metrics	Accuracy, Precision, Recall, F1-Score

TABLE 1: Model hyperparameters and experimental configuration.

Furthermore, the area under the receiver operating characteristic curve metric was used to evaluate the trade-off between true positive and false positive rates over varying decision thresholds, providing deeper insight into the classifier’s discriminative power. Given the massive scale of the CML-COVID dataset (~20 million tweets), an 80/20 hold-out validation split was utilized.

This approach provides a sufficiently large and diverse testing set to ensure the model’s performance is representative of real-world variance without the prohibitive computational cost of iterative cross-validation. Table 1 illustrates cumulative model hyperparameters and the evaluation criteria that ensure a robust and multidimensional assessment of model performance within the context of crisis-related sentiment analysis.

Results And Discussion

The results of this study demonstrate the viability of using natural language processing and ML techniques to detect, analyze, and forecast mass hysteria phenomena by identifying distinct phases of MPI escalation and characterizing the life cycle of digital mass hysteria. The sentiment analysis revealed significant temporal fluctuations in public mood over the course of the pandemic-related crisis period, illustrating how collective emotions evolved in response to uncertainty, misinformation, and subsequent stabilization phases.

How to cite this article:

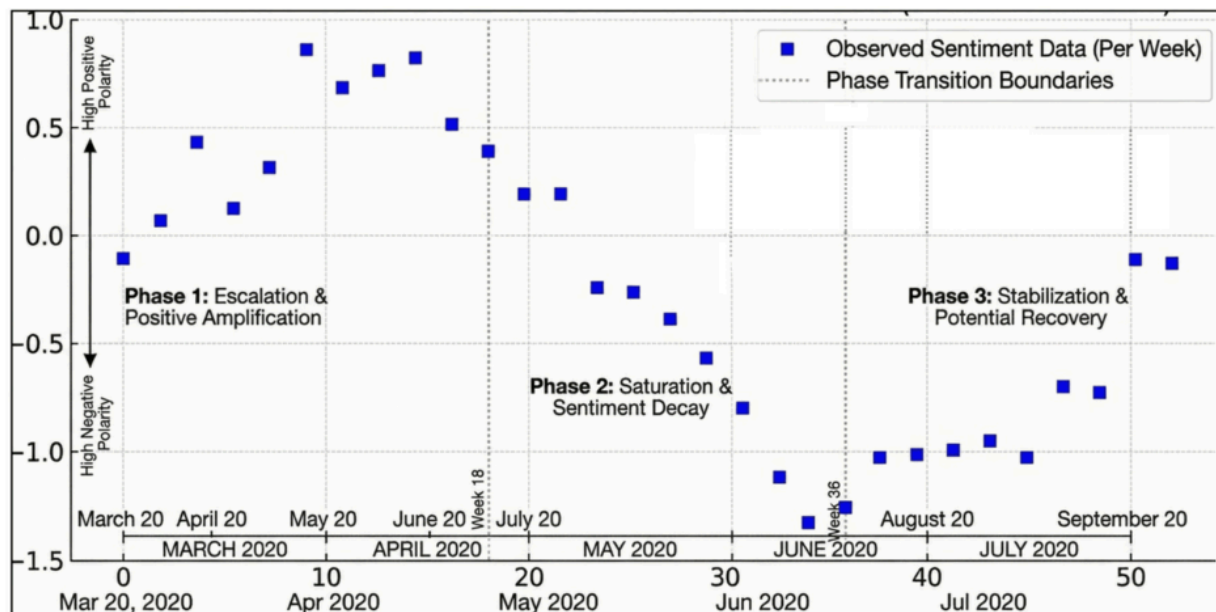


FIGURE 4: Sentiment analysis result.

Figure 4 represents the sentiment score, which is a reflection of public mood or emotions that range from negative (fear, anxiety) to positive sentiment (hope, relief). The positive scores will imply calmer or more optimistic feelings, and negative scores would be of fear, panic, or distress.

The sentiment trajectory is divided into three distinct phases, where Phase 1 reflects an early surge of negative sentiment driven by fear and initial panic, Phase 2 shows a steep decline into intense negativity likely induced by misinformation surges or escalating crisis conditions, and Phase 3 illustrates stabilization as positive and negative sentiment oscillations reduce over time. This temporal progression highlights the cyclical dynamics of public emotion during crises and underscores the role of social media in both amplifying and attenuating hysteria-driven narratives. Further analysis using ML-based forecasting models showed that predicted sentiment values closely aligned with observed sentiment trends, indicating strong time-series predictability of emotional responses during crisis scenarios.

How to cite this article:

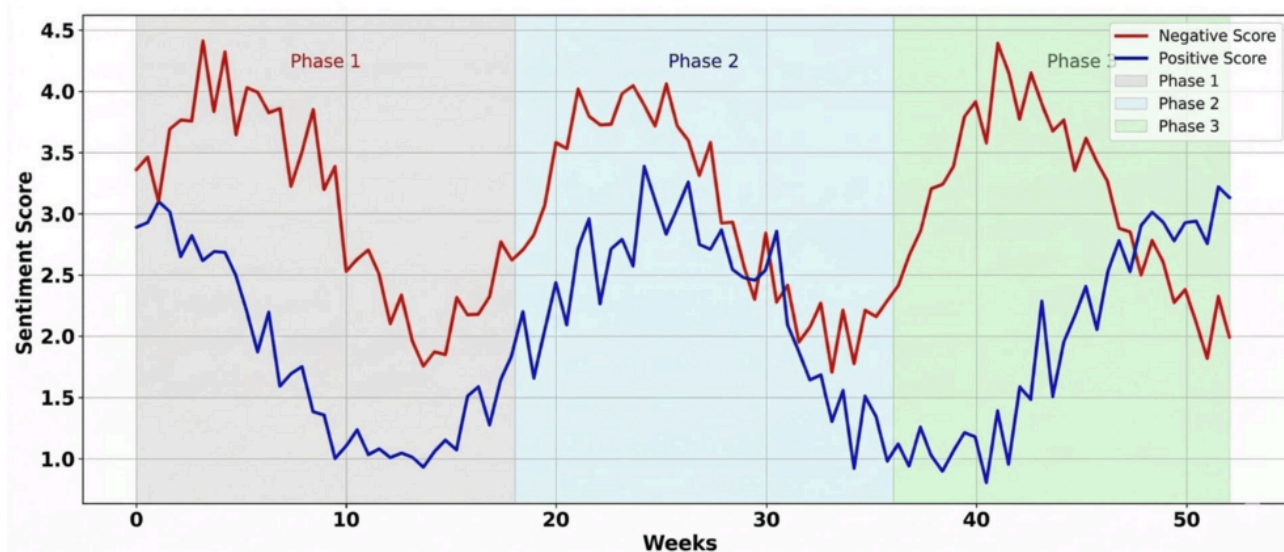


FIGURE 5: Sentimental score over 50 weeks.

Figure 5 gives a nuanced split of the dynamics of sentiment over time into three distinct phases (Phase 1, Phase 2, and Phase 3), with the blue line representing positive sentiment and the red line representing negative sentiment explicitly on the plot. The predicted versus actual sentiment curves revealed accurate identification of sentiment inflexion points, especially the sharp negative dip in Phase 2 and the gradual sentimental recovery in Phase 3. The sharp downward trajectory in Phase 2 serves as the primary diagnostic marker for imminent mass hysteria, where the high velocity of sentiment decay indicates a shift from individual anxiety to collective emotional contagion.

The accuracy of these curves was validated against a ground-truth baseline established through temporal alignment with documented pandemic milestones and manual expert annotation of the corpus, ensuring that the model's predictions reflect historical reality. These predictive insights are valuable for policymakers and public health authorities, as they enable early detection of emotional volatility and offer timing windows for strategic interventions such as fact-checking campaigns or psychological support efforts.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	78%	0.75	0.72	0.73
Random Forest	85%	0.83	0.79	0.81
LSTM (Deep Learning)	89%	0.88	0.85	0.86
BERT (NLP-based)	92%	0.91	0.89	0.90

TABLE 2: Model performance for detecting mass hysteria through Twitter words.

How to cite this article:

Model performance evaluation, as illustrated in Table 2, confirmed the superiority of transformer-based models for sentiment classification. Table 2 summarizes the comparative results: LR yielded an accuracy of 78% (Precision 0.75, Recall 0.72, F1-score 0.73), Rf achieved 85% (Precision 0.83, Recall 0.79, F1-score 0.81), LSTM achieved 89% (Precision 0.88, Recall 0.85, F1-score 0.86), and BERT achieved the highest accuracy of 92% (Precision 0.91, Recall 0.89, F1-score 0.90). This performance hierarchy confirms that models capable of capturing contextual linguistic relationships, such as LSTM and BERT, are more effective in crisis informatics tasks where semantics, sentiment polarity, and contextual cues are strongly interdependent.

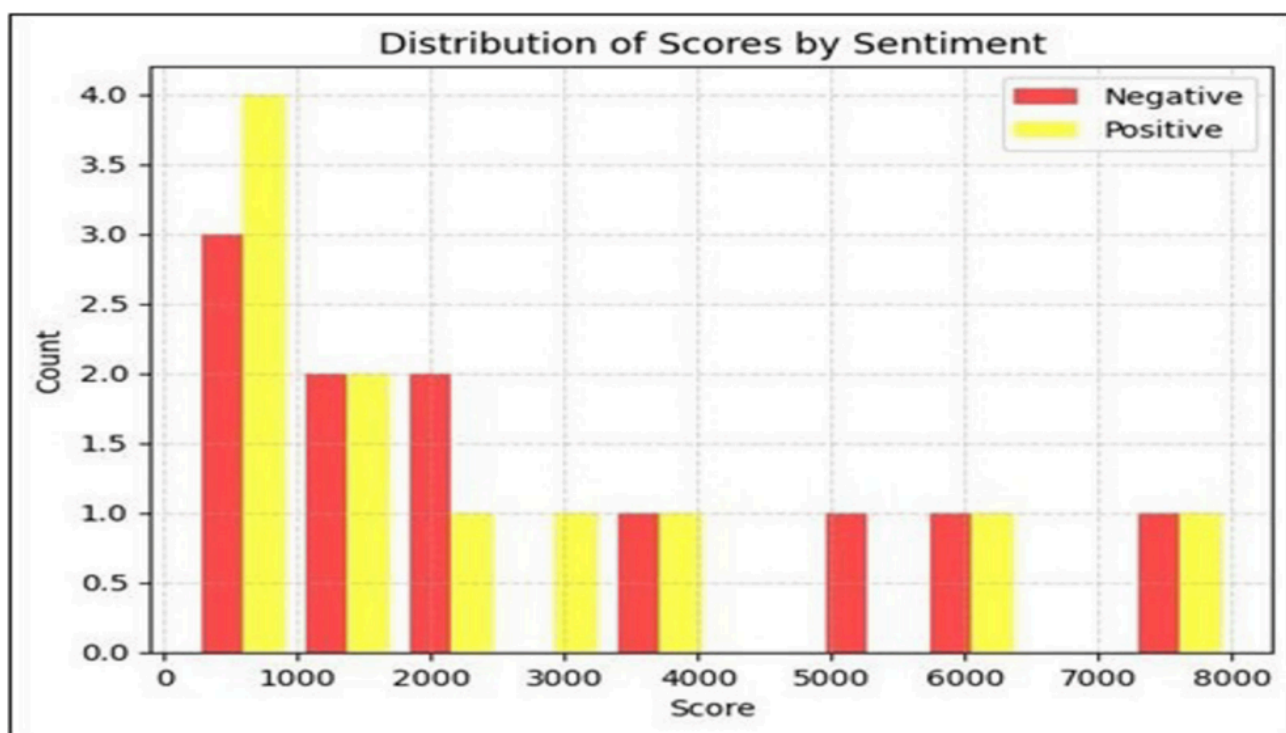


FIGURE 6: The graph portrays the distribution of affect scores for both positive and negative sentiments.

Figure 6 represents the sentiment scores clustered predominantly in the lower positive and negative ranges, indicating that mild-to-moderate emotional responses were more common than extreme reactions. This score was calculated by aggregating individual tweet polarities into discrete intensity bins. The sentiment valence, ranging from -1 (maximal distress) to +1 (maximal resilience) was calculated by summing the valence of each word in the tweet, for amplifiers and negations. The system then counts how many tweets fall in the positive and negative frequency to depict the sentiment score.

Positive sentiment frequencies exceeded negative frequencies in the lower-scoring band, suggesting a counterbalancing emotional resilience within the population. Meanwhile, the scarcity of extreme negative sentiment responses indicates that although fear and anxiety were present, they did not uniformly escalate into severe collective hysteria. This aligns with historically documented behavioral patterns of mass psychogenic illness, where only certain demographic clusters undergo acute psychological response escalation while others maintain emotional stability.

While polarity provides a broad view, these frequency distributions allow for a more fine-grained observation of emotional intensity; specifically, the concentration of scores in the high-intensity negative tail corresponds to the fear-driven signatures associated with psychogenic events

How to cite this article:

These results collectively demonstrate that ML models can effectively detect emotional contagion patterns, that crises manifest in temporally segmented sentiment phases, and that predictive modeling can assist in anticipating phase transitions associated with panic escalation. The integration of sentiment dynamics, predictive curves, and performance metrics underscores the potential for computational systems to support crisis management, misinformation mitigation, and mental health intervention strategies during future pandemics or socio-digital emergencies.

Conclusions

This study demonstrates the efficacy of leveraging ML and NLP techniques to detect, analyze, and predict mass hysteria phenomena during crisis events using social media data, with Twitter serving as the primary data source. The sentiment analysis and time-series modeling revealed distinct phases of emotional response, with Phase 1 exhibiting early negative sentiment reflecting initial panic, Phase 2 showing intensified negative sentiment corresponding to misinformation propagation and heightened uncertainty, and Phase 3 indicating stabilization as public emotions gradually balanced between positive and negative sentiment. Predictive modeling using LSTM and transformer-based architectures such as BERT achieved superior performance, with BERT attaining 92% accuracy, 0.91 precision, 0.89 recall, and 0.90 F1-score, outperforming classical models such as LR and RF. These results confirm that contextualized language models are highly effective in capturing nuanced sentiment dynamics, including fear, anxiety, and misinformation-induced panic, which are critical indicators of potential mass psychogenic illness. Analysis of sentiment score distributions further revealed that moderate emotional responses were predominant, whereas extreme negative sentiment was comparatively rare, suggesting a natural resilience in the population against full-scale hysteria. The limitation or future scope of this study would be the real-time implementation of this hysteria pipeline integrated with either social media applications or media coverages. Additional limitation is as followed; due to the Twitter API rate limit may underrepresent rural or low-internet population. Model was also only trained on COVID-19 panic and does not generalize over different type of crisis.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Acquisition, analysis, or interpretation of data: Tejaswini Yadav, Disha Singh, Shreya Das

Critical review of the manuscript for important intellectual content: Tejaswini Yadav, Disha Singh, Shreya Das

Concept and design: Raahil Sheikh, Lavleen Sharma, Pragyan Priyadarshini Rout

Drafting of the manuscript: Raahil Sheikh, Lavleen Sharma, Pragyan Priyadarshini Rout

Supervision: Lavleen Sharma, Shreya Das

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

How to cite this article:

References

1. Bagus P, Peña-Ramos JA, Sánchez-Bayón A: [COVID-19 and the political economy of mass hysteria](#). International Journal of Environmental Research and Public Health. 2021, 18:1376. [10.3390/ijerph18041376](#)
2. Sahu AA: [COVID-19 pandemic: inducing pan-hysteria and panic disorder; an overview](#). Indian Journal of Forensic Medicine & Toxicology. 2020, 14:6600-6604. [10.37506/ijfmt.v14i4.12645](#)
3. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E: [Misinformation of COVID-19 on the internet: infodemiology study](#). JMIR Public Health and Surveillance. 2020, 6:e18444. [10.2196/18444](#)
4. Ali-Gombe A, Guthrie E, McDermott N: [Mass hysteria: one syndrome or two?](#). The British Journal of Psychiatry. 1996, 168:633-635. [10.1192/bjp.168.5.633](#)
5. Donaldson LJ, Cavanagh J, Rankin J: [The dancing plague: a public health conundrum](#). Public Health. 1997, 111:201-204. [10.1038/sj.ph.1900371](#)
6. Islam MN, Inan TT, Rafi S, Akter SS, Sarker IH, Islam AKM Najmul: [A systematic review on the use of AI and ML for fighting the COVID-19 pandemic](#). IEEE Transactions on Artificial Intelligence. 2021, 1:258-270. [10.1109/tai.2021.3062771](#)
7. Shorten C, Khoshgoftaar TM, Furht B: [Deep learning applications for COVID-19](#). Journal of Big Data. 2021, 8:1-54. [10.1186/s40537-020-00392-9](#)
8. Mergenthaler E, Bucci W: [Linking verbal and non-verbal representations: computer analysis of referential activity](#). British Journal of Medical Psychology. 1999, 72:339-354. [10.1348/000711299160040](#)
9. Jeong D, Hanson-DeFusco J, Kim D, Lee CK: [Digital mass hysteria during pandemic? A study of Twitter communication patterns in the US during the stages of COVID-19 vaccination](#). Behavioral Sciences. 2024, 14:389. [10.3390/bs14050389](#)
10. Hasan MR, Maliha M, Arifuzzaman M: [Sentiment analysis with NLP on Twitter data](#). 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2). 2019, 1-4. [10.1109/IC4ME247184.2019.9036670](#)
11. Hochreiter S, Schmidhuber J: [Long short-term memory](#). Neural Computation. 1997, 9:1735-1780. [10.1162/neco.1997.9.8.1735](#)
12. Devlin J, Chang MW, Lee K, Toutanova K: [BERT: pre-training of deep bidirectional transformers for language understanding](#). In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019, 1:4171-4186. [10.18653/v1/N19-1423](#)
13. Wilkie DCH, Lipnickas G, Pham NTA: [Emotional contagion on social media: pathways, effects, and insights for marketers](#). Journal of Marketing Management. 2025, 42:57-90. [10.1080/0267257x.2025.2570739](#)
14. Madububambachu U, Ukpebor A, Ihezue U: [Machine learning techniques to predict mental health diagnoses: a systematic literature review](#). Clinical Practice and Epidemiology in Mental Health. 2024, 20:1-16. [10.2174/0117450179315688240607052117](#)
15. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP: [SMOTE: synthetic minority over-sampling technique](#). Journal of Artificial Intelligence Research. 2002, 16:321-357. [10.1613/jair.953](#)
16. Dashtian H, Murthy D: [CML-COVID: a large-scale COVID-19 Twitter dataset with latent topics, sentiment and location information](#). ArXiv. 2021, 1-6. [10.48550/arXiv.2101.12202](#)

How to cite this article: