Final Report – SentiFM: Epidemic Forecasting with Text Sentiment Analyses

Yule Wang
Georgia Tech
Atlanta, USA
yulewang@gatech.edu

Huanzhong Jia
Georgia Tech
Atlanta, USA
hjia43@gatech.edu

Junyang He
Georgia Tech
Atlanta, USA
jhe382@gatech.edu

Abstract

Traditional epidemiological forecasting models, such as compartmental (SIR/SEIR) and time-series regressions, heavily rely on lagging information indicators (e.g., confirmed cases, hospitalizations). This causes a critical blind spot, as infectious disease spread is fundamentally a behavior-mediated process shaped by public perception and emotional response. To fill this research gap, we propose a framework named, SentiFM, a sentiment-aware forecasting framework that incorporates both advanced Large Language Models (LLMs) *and* classical sequence models (e.g., LSTMs), as well as multimodal fusion techniques, to enhance epidemic prediction. We integrate rich, unstructured text data from social media as a real-time reflection of public mood. The algorithm involves using an LLM and classical sequence models to extract fine-grained sentiment signals and align them temporally with comprehensive epidemiological indicators. We will evaluate its performance on the WHO and a COVID-19 Twitter dataset, which consists of various sentiments and epidemiological features. To sum up, this work fills the research gap between advanced LLM capabilities and sentiment-aware epidemic forecasting. Code is available at https://github.com/JunyangHe/epidemic-sentiment-forecasting.git.

CCS Concepts

• Epidemiology → Time-series Forecasting ; **• Machine Learning → LLMs**.

ACM Reference Format:
Yule Wang, Huanzhong Jia, and Junyang He. 2018. Final Report – SentiFM: Epidemic Forecasting with Text Sentiment Analyses. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Traditional epidemiological forecasting frameworks—such as compartmental models (SIR/SEIR) and time-series regressions—rely predominantly on *lagging indicators* (e.g., confirmed cases, hospitalizations, deaths). While these signals capture the biological progression of an outbreak, they reveal transmission dynamics only after the fact, often with a delay of one to two weeks. Consequently, these models struggle to anticipate sudden behavioral shifts in the population that drive infection surges or declines.

However, infectious disease spread is fundamentally a **behavior-mediated** process. Public compliance with interventions—such as mask mandates, vaccination uptake, and mobility reduction—is directly shaped by collective perception and emotional response to unfolding events. For instance, fear and anxiety can suppress mobility and contact rates, while fatigue, skepticism, or optimism can rapidly reverse these behavioral gains. As such, understanding and predicting epidemic dynamics without quantifying public sentiment leaves a critical blind spot.

Social media platforms provide real-time reflections of public mood, offering rich, unstructured text data that capture evolving emotions, beliefs, and attitudes. Our project leverages **Large Language Models (LLMs)** to extract fine-grained sentiment signals from these social media streams and align them temporally with epidemiological indicators. We then integrate these sentiment-derived features as *exogenous variables* into modern forecasting architectures such as Chronos-2 [2]. By fusing behavioral sentiment with epidemiological time-series, we aim to enhance prediction accuracy and improve model responsiveness to shifts in public behavior.

2 Related Works

For the current LLM integration frameworks, specifically, EpiLLM [4] achieves in-depth and precision forecasting capacity by employing a dual-branch architecture that combines Graph Neural Networks (GNNs) with an MLP to simultaneously capture spatial and temporal patterns. This makes it powerful for complex, multi-source datasets requiring structural modeling. CovidLLM [14], on the other hand, focuses on flexibility and practicality. It converts time-series data into structured text prompts, utilizing a lightweight LLM for efficient prediction. Its advantage draws in its high training efficiency, making it suitable for resource-constrained environments. Nonetheless, neither of them have employed the power of LLMs for sentiment analyses in the scenarios of pandemic forecasting.

By comparison, most recently, **LLMs** purpose-built for public-health have been proposed to support multilingual monitoring across a variety of tasks, moving beyond generic models toward domain specialization with curated training corpora and public-health specific training sets [6, 13]. This work establishes a state-of-the-art capability by supporting up to $\sim$ 100 distinct infoveillance tasks, directly validating the paradigm of using domain-specific LLMs for comprehensive public attitude monitoring. Wang et al. [11] presents an LLM agentic framework for integrating non-numerical, unstructured news events into time series prediction. A key contribution of

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

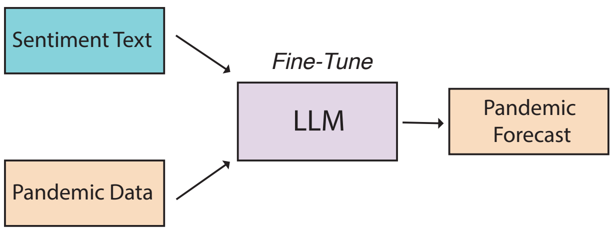

Figure 1: Framework Illustration of our SentiFM-LLM framework, which conducts epidemic forecasting with text sentiment analyses using LLMs project.

this framework is the proposed reflection mechanism, which allows the model to iteratively refine its understanding and conversion of complex external textual information into effective predictive signals. This offers a highly relevant blueprint for integrating social media sentiment in epidemic forecasting.

Regarding **sentiment analysis** in epidemic data, prior works [7, 8] have established social media sentiment as a critical feature to enhance the accuracy of time-series pandemic prediction models. However, their conventional sequential modeling frameworks, such as a stacking of LSTM/RNN layers, generally do not perform as well as LLMs in this complex language task. Beyond the standard structured models, **social media** remains a vital information source for public health surveillance. Early studies differentiated between indicator-based and event-based monitoring, focusing on data collection and geotagging. During the COVID-19 pandemic, social platforms [9, 10] proved instrumental in sensing public attitudes, monitoring infodemics, and tracking government responses.

To sum up, this project fills the research gap between advanced LLM capabilities and epidemic forecasting by proposing a sentiment-aware holistic framework. We demonstrate that this novel coupling of language advanced modeling with spatio-temporal dynamics extraction can significantly enhances forecasting accuracy and provides crucial, interpretable insights into the public health factors driving disease spread.

3 Methodology

3.1 Timeseries Forecasting

We perform forecasting with three main families of time series forecasting models: a classical statistical model (SARIMAX), a neural sequence model (LSTM), and a time–series foundation model (Chronos-2). As a traditional baseline, we use Seasonal Autoregressive Integrated Moving Average with exogenous regressors (SARIMAX), which captures linear auto-regressive and seasonal structure in the epidemiological series while being capable of incorporating sentiment as external covariates. We then employ a vanilla LSTM network constructed with PyTorch as a nonlinear multivariate forecaster, and the Chronos-2 model [2] proposed by Amazon as the more recent foundation model approach. Chronos-2 enables covariate input as exogenous information to the Chronos time series foundation model [3]. Across all models, we use a lookback window of 14 days and a forecast horizon of one day.

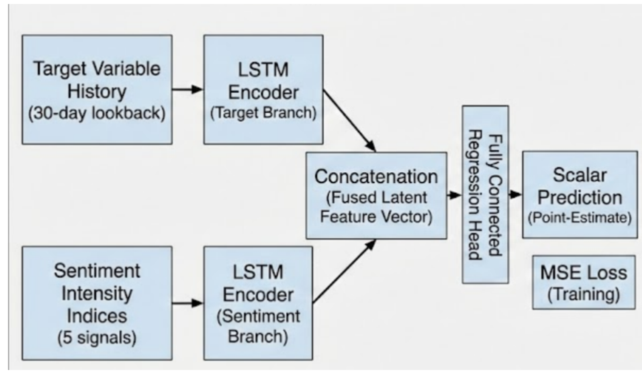

Figure 2: Framework Illustration of our SentiFM-LSTM framework.

3.2 Sentiment Integration

We incorporate sentiment analysis results as **exogenous variables** to enhance COVID-19 forecasting. All sentiment signals are aggregated to the same temporal granularity as the epidemiological data, forming a parallel time series $\mathbf{s}_t$ aligned with $\mathbf{y}_t$, the COVID target (e.g., new cases).

We evaluate three integration strategies with LSTM, Chronos-2, and LLM:

- **1. Naive Concatenation.** Aggregated sentiment features (emotion intensities, sentiment ratios, and tweet volume) are directly concatenated with lagged epidemiological inputs, producing a joint feature vector $\mathbf{x}_t = [\mathbf{y}_{t-\delta t:t}; \mathbf{s}_t]$. LSTM then uses this augmented series as multivariate input to predict $\mathbf{y}_{t+1}$. This baseline measures whether sentiment trends alone provide additive predictive value.
- **2. In-Context Learning.** We leverage the native In-Context Learning (ICL) capability of the Chronos-2 model pipeline, which is engineered to handle covariate information. Instead of fine-tuning the time series foundation model, the sentiment time series $\mathbf{s}_t$ is treated as exogenous covariates alongside the target series $\mathbf{y}_t$. Internally, Chronos-2's Group Attention and Time Attention layers dynamically learn the correlations between sentiment trends and future COVID-19 streams from the provided context.
- **3. Delayed Alignment (Performative Shift, τ=4).** Inspired by the performative forecasting framework [12], we account for behavioral delays between public sentiment and observable epidemic outcomes. Using a lead–lag analysis based on cosine similarity, we estimate an optimal delay $\tau*$ =4 days between composite sentiment changes and subsequent case trends. To approximate this performative feedback, we shift the sentiment series forward by $\tau*$—using $S_{t+\tau*}$ instead of $\mathbf{s}_t$—before feeding it into Chronos-2 as exogenous input. This *delayed alignment* captures the causal latency of public response (e.g., fear or optimism affecting mobility and exposure several days later) without introducing new model parameters.
- **4. Multimodal Fusion (Dual-Branch).** To mitigate noise interference between smooth epidemiological trends and

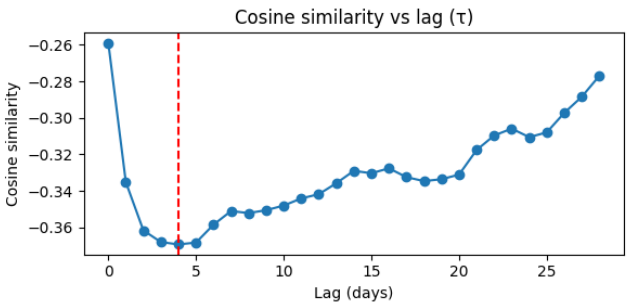

Figure 3: The best shift days (selected).

volatile social sentiment, we implemented a *Late Fusion* architecture. Instead of early concatenation, we employ two distinct LSTM encoders: a *Temporal Encoder* processes the target history $\mathbf{y}_{t-\delta t:t}$ to capture autoregressive dynamics, while a separate *Sentiment Encoder* processes the multivariate sentiment stream $\mathbf{s}_t$. The final hidden states from both branches are concatenated and passed through a Multi-Layer Perceptron (MLP) fusion head to generate the prediction. This approach allows the model to learn distinct latent representations for biological momentum and behavioral signals before integration.

3.3 Generative AI Forecasting (Llama-3)

To evaluate the efficacy of general-purpose reasoning versus specialized time-series architectures, we employed **Llama-3-8B-Instruct**. Unlike Chronos, which tokenizes values into a fixed vocabulary, Llama-3 processes time series as natural language sequences. We explored three distinct integration strategies:

- **Zero-Shot Prompting (Base).** Following the *LLMTime* methodology, we serialize the numerical history of confirmed cases into a comma-separated string representation. We construct a system prompt instructing the model to act as a pattern completion engine. No gradient updates are applied; the model relies entirely on its pre-trained next-token prediction capabilities to extrapolate the trend. This serves as our generative baseline.
- **Univariate Instruction Tuning.** To adapt the model to the specific magnitude and volatility of epidemiological data, we performed Supervised Fine-Tuning (SFT) using **QLoRA** (Quantized Low-Rank Adaptation). We froze the 8B parameters in 4-bit precision and trained low-rank adapters on the target dataset. The instruction format was strictly univariate:
  "Instruction: Predict the next COVID case count based on the history. Input: [History of Cases]. Response: [Target]."
  This allows the model to learn the specific "scale" of the pandemic without exogenous variables.

3.4 Algorithm In-depth Statistical Analysis

Given that our methodology employs deep neural network-based architectures like LLMs and Chronos-2, a complete theoretical analysis of this framework faces challenges inherent to the "black-box" properties on them. Thus, we focus instead on the core algorithmic implementation components and modeling objective of the SentiFM framework: In this section, we provide a detailed analysis of the modeling pipeline to better understand the theoretical assumptions and limitations of the SentiFM architecture.

Tokenization and Quantization. To process the continuous epidemiological time series multi-dimensional variables, we first convert them into a discrete token vocabulary through scaling and quantization. This procedure allows the data to be fit into standard LLM architectures, which are built to process sequences of discrete tokens (words). Through this approach, we can use the LLM to process numerical time series data.

To enable the use of Transformer Layers, which are designed for processing discrete language tokens, we apply a quantization procedure to the time-series inputs. The epidemiological values and sentiment signals are scaled and discretized into a fixed vocabulary of tokens. Notably, this introduces a quantization error term ϵ:

$$\tilde{\mathbf{x}}_t = \text{Quantize}(\mathbf{x}_t) = \mathbf{x}_t + \epsilon_q.$$

Modeling Objective. Given the historical preceding time-series point and the sentiment text tokens, our overall objective is to learn the conditional distribution of the next time-series value $\mathbf{y}_{t+1}$ given historical epidemiological observations and aligned sentiment signals. Formally, the model aims to estimate:

$$p\left(\mathbf{y}_{t+1} \mid \mathbf{y}_{t-\delta t:t}, \mathbf{s}_t; \Theta\right) = p(\mathbf{y}_{t+1} \mid \mathbf{x}_t; \Theta), \tag{1}$$

in which $\mathbf{x}_t = [\mathbf{y}_{t-\delta t:t}; \mathbf{s}_t]$ includes the preceding time-series features and the sentiment text features, Θ denotes all the trained parameters in the overall framework. In LLMs, this estimation is implicitly learned via next-token prediction objectives on tokenized input sequences, whereas in Chronos-2, it is implemented through attention-based temporal modeling of continuous-valued features and sentiment covariates.

With the incorporation of sentiment features as covariates, instead of a purely autoregressive process, here we now model:

$$\mathbf{y}_{t+1} \sim f(\mathbf{y}_{t-\delta t:t}, \mathbf{s}_t), \tag{2}$$

This modeling framework allows the model to potentially capture latent behavioral signals (e.g., fear, happiness) that modulate the epidemic dynamics. With the delayed alignment τ of the sentiment, we further have:

$$\mathbf{y}_{t+1} \sim f(\mathbf{y}_{t-\delta t:t}, \mathbf{s}_{t+\tau}). \tag{3}$$

Our framework can be interpreted through a probabilistic machine learning perspective, since that Chronos-2 produces probabilistic forecasts through quantile regression heads. It estimates conditional quantiles $\hat{q}_\alpha(\mathbf{y}_{t+1})$ for various levels $\alpha \in (0, 1)$. This procedure allows for interval forecasts and uncertainty quantification. We have the following equation:

$$p(\mathbf{y}_{t+1}) \approx \frac{1}{K} \sum_{k=1}^{K} \delta(\mathbf{y}_{t+1}^{(k)}), \tag{4}$$

where δ denotes a point mass and $\mathbf{y}_{t+1}^{(k)}$ is the k-th head sample from the foundation model's output.

Limitations. Due to their black-box nature, large-scale deep models such as Chronos-2 lack explicit interpretability compared to classical statistical models like SARIMAX. Furthermore, they may overfit to unreal (spurious) correlations between sentiment and case trends, especially when sentiment data is with low signal-to-noise ratio or misaligned. To mitigate this, we apply temporal regularization (e.g., delayed alignment) and architectural strategies (e.g., dual-branch fusion) to enforce the correct structured learning.

To sum up, while SentiFM leverages expressive models, its predictive performance relies heavily on careful preprocessing, causal assumptions about sentiment influence, and the statistical compatibility of multi-modal signals.

4 Experiment

4.1 Dataset

Covid time series: We use the official World Health Organization (WHO) COVID-19 global dataset [1], which provides daily country-level surveillance data reported by national authorities. The WHO repository aggregates standardized indicators across multiple thematic tables (epidemiological, testing, clinical, and vaccination), making it suitable for multivariate time series forecasting.

For this study, we integrated four components: (i) daily epidemiological data (confirmed cases, deaths, cumulative counts), (ii) testing activity (number of tests performed, test positivity), (iii) hospitalization (current admissions, ICU occupancy), and (iv) vaccination progress (doses administered, people partially/fully vaccinated). After alignment and merging on the date and location keys, the unified table contained 35 observable properties. To reduce inconsistency arising from cross-country reporting practices, we restricted the analysis to the United States.

Covid sentiment: We use the `COVID19_twitter_full_dataset` [5] as our primary source of social sentiment information. This dataset contains tweet-level records with fields such as *tweet_timestamp*, *region*, *sentiment label* (positive, neutral, negative), and *emotion intensities* including valence, fear, anger, happiness, and sadness. Each of the 6 features represents an individual tweet associated with COVID-related discussions. We aggregate tweets by date to compute daily averages of emotion scores, sentiment proportions, and tweet volume, producing a sentiment time series S_t that captures collective emotional trends aligned with the epidemiological data.

Data aggregation: We combine the WHO covid time series and the sentiment time series by shared time interval to get a dataset with 40 time series properties covering 9 full months from 2020/12/31 to 2021/9/1. We will finetune LLM with the 6 sentiment features in this dataset and use the finetuned model for forecasting on select covid time series properties. We will also try multivariate forecasting on covid time series data with and without sentiment time series to measure the effect of sentiment data.

4.2 Dataset Visualizations

Figures 4, 5, 6, 7 show selected timeseries streams in each of the four main categories of WHO covid-19 data. Figure 8 shows the correlation between the sentiment features and the WHO covid-19 features.

4.3 Naive Concatenation with LSTM

In this experiment, we study the effect of fear intensity index as sentiment exogenous information on four important epidemiological features. We only leverage fear intensity index out of all sentiment time series streams as it shares the highest correlation with our target COVID features, demonstrated in Figure 8, and is more

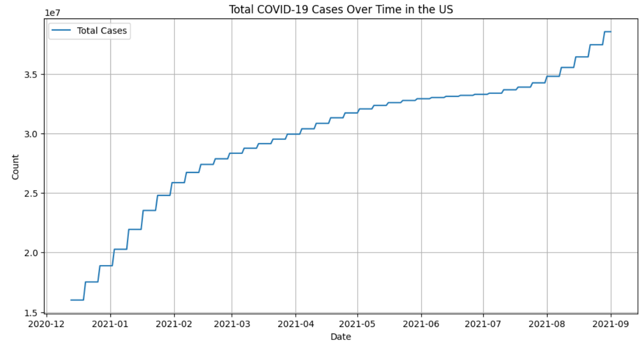

Figure 4: Total cases.

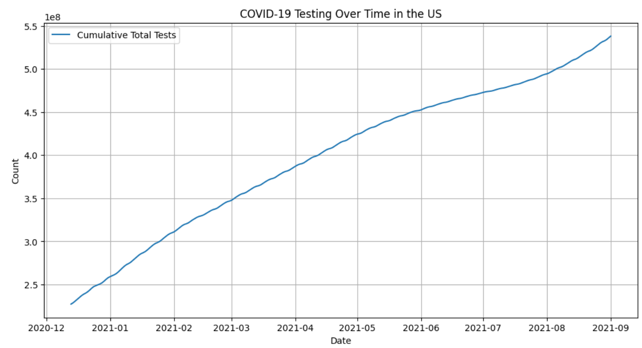

Figure 5: Testing data (selected).

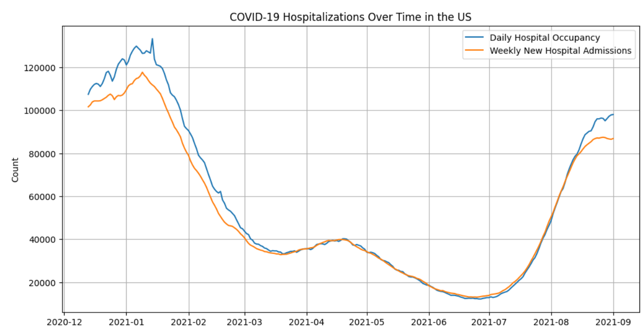

Figure 6: Hospitalization data (selected).

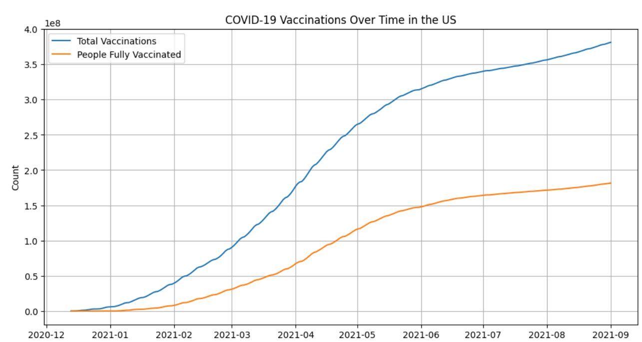

Figure 7: Vaccination data (selected).

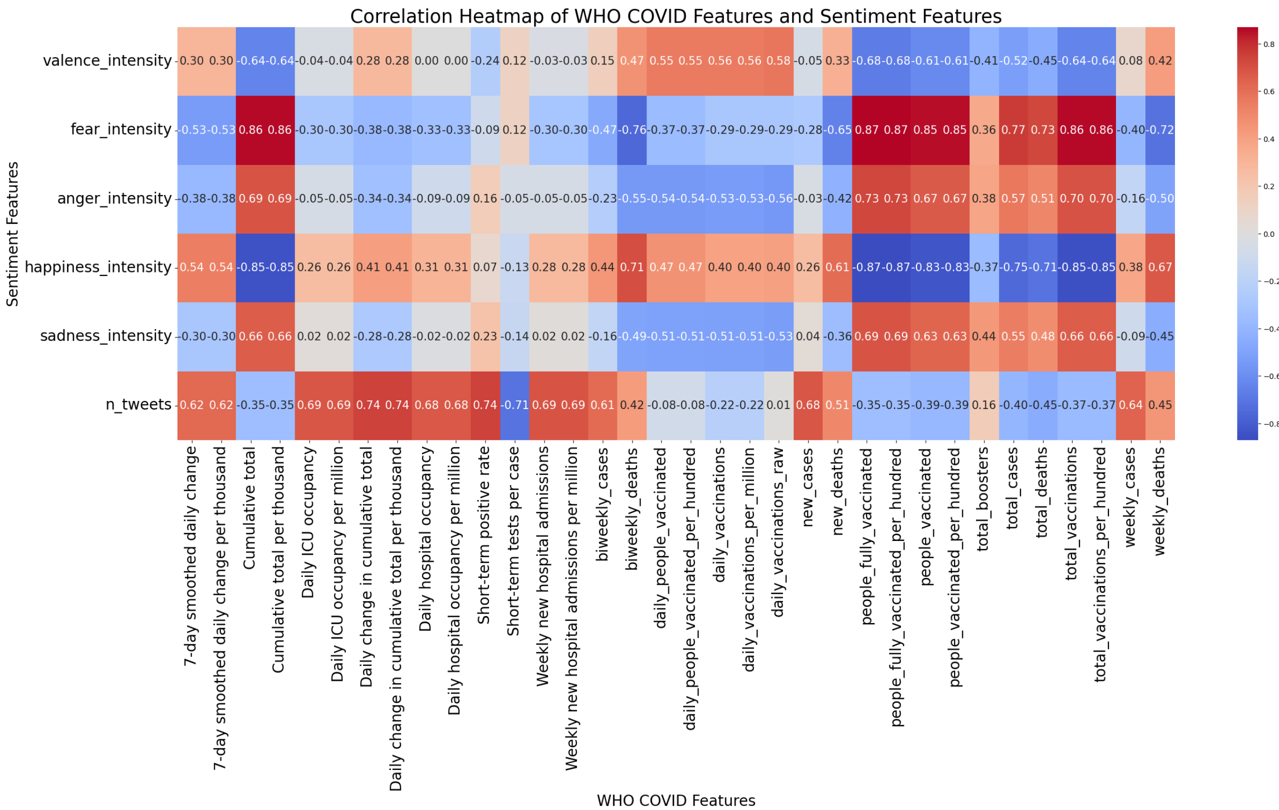

Figure 8: Correlation heatmap of covid and sentiment features.

Table 1: Effect of naively concatenated sentiment intensity features on epidemiological forecast with LSTM (shown in MSE)

COVID feature	Without sentiment	With sentiment
People vaccinated	0.0254	0.0211
Total tests	0.3379	0.3218
Total cases	0.2930	0.2919
Total deaths	0.489	0.151

suitable for the small network used in this experiment. A naive concatenation of epidemiological and sentiment data is chosen. We construct a simple one-layer Long Short-Term Memory network with a lookback window of 14 days and a forecast horizon of one day. Due to the limited size of the data, we choose a batch size of eight and hidden size of eight for a network with two LSTM layers and one linear layer. We train and evaluate the model under two configurations: 1) univariate training and forecasting on the COVID feature, 2) multivariate training with fear intensity index and univariate forecasting on the COVID feature. Results, shown in Table 1, are obtained from an average of 10 runs, each trained over 50 epochs with a learning rate of 0.001. Results show the potential of adding sentiment data to aid forecasting of highly correlated features through naive concatenation.

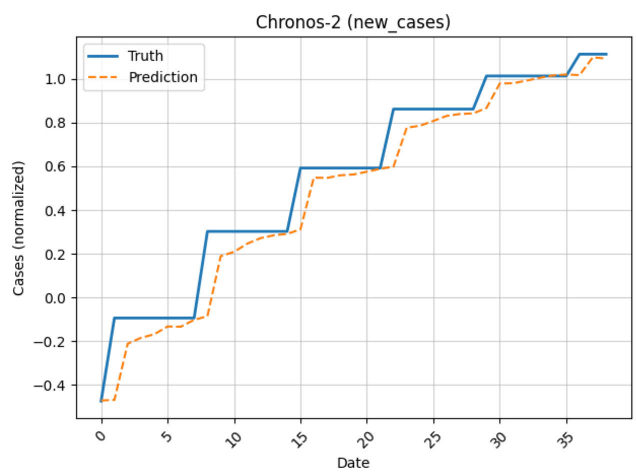

Figure 9: Chronos-2 total cases (ICL with sentiment data).

4.4 In-Context Learning with Chronos-2

In this experiment, we study the effect of including all five intensity indices as sentiment exogenous information on the same set of COVID-19 epidemiological features (cases, deaths, vaccination, and testing). Consistent with previous analyses, a lookback window of 14 days and a forecast horizon of one day were maintained. The

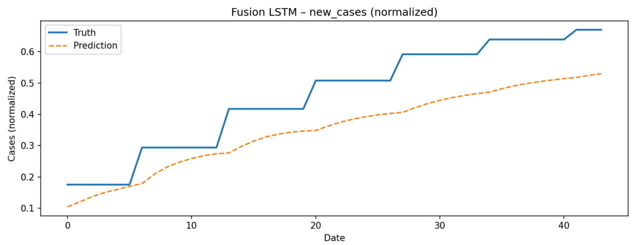

Figure 10: Multimodal Fusion total cases.

Table 2: Effect of naively concatenated sentiment intensity features on epidemiological forecast with Chronos-2 (shown in RMSE)

COVID feature	Without sentiment	With sentiment
People vaccinated	0.00421	0.00240
Total tests	0.00898	0.00628
Total cases	0.0677	0.0702
Total deaths	0.0286	0.0297

Chronos-2 model was evaluated across two inference configurations: 1) a univariate baseline utilizing only the historical COVID feature as input, and 2) an ICL configuration where the five sentiment streams were provided as explicit covariates via the Group Attention mechanism. As shown in Table 2, the inclusion of sentiment data yielded significant improvements in forecasting accuracy for vaccination and testing features. However, the model exhibited a marginal degradation in performance for total cases and deaths. This counter-intuitive result suggests that the latent sentiment signals relevant to total cases and deaths may already be implicitly embedded within the historical epidemiological patterns, making sentiment covariate input redundant or potentially noisy for these targets. The truth vs. the predicted result total cases result is shown in Figure 10.

4.5 Generative Forecasting with Llama-3 and QLoRA

In this experimental phase, we pivot to a generative paradigm, evaluating the zero-shot and fine-tuned capabilities of Llama-3-8B on the epidemiological time series. Unlike the specialized architectures discussed previously, Llama-3 is a general-purpose text-completion model, presenting a unique "modality mismatch" when applied to strict numerical regression. The fundamental challenge lies in the model's pre-training on vast textual corpora; directly generating precise floating-point values requires the model to override its inherent bias toward linguistic token sequences. To mitigate this, we employed rigorous prompt engineering—enforcing a strict System-User-Assistant schema—to constrain the output space to numerical tokens and reduce hallucinations.

Furthermore, the "Big Model, Small Data" paradox becomes evident here. The COVID-19 dataset, while temporally dense, is insufficient in volume to support full parameter tuning of an 8-billion parameter model without immediate overfitting. To address this, we utilized Quantized Low-Rank Adapters (QLoRA). By freezing the pre-trained backbone and injecting trainable low-rank matrices into the query and value projections, we enabled the model to adapt to the specific distribution of epidemiological data. As shown in the implementation below, this approach yielded a notable reduction in format errors and improved trend adherence, although the stochastic nature of token sampling remains a limiting factor compared to deterministic regression heads.

4.6 Deterministic Forecasting with Dual-Branch Multimodal Fusion

In this experiment, we explore the predictive capabilities of a **deterministic Dual-Branch Fusion LSTM**, explicitly integrating the five sentiment intensity indices as exogenous information. Consistent with the previous setup, the model utilizes a lookback window of 30 days to generate a forecast horizon of one day. We utilized a **Dual-Branch Conditioning mechanism**, where the target variable history and the sentiment streams are processed by parallel LSTM encoders.

As defined in the architecture, the final hidden states from both the target and sentiment branches are concatenated to create a combined latent feature vector. This fused representation is passed through a fully connected regression head to output a scalar prediction. Unlike stochastic generation models, this architecture functions as a **point-estimate regressor**, utilizing Mean Squared Error (MSE) loss to minimize the deviation between the predicted values and the actual epidemiological figures. This approach allows the model to learn the direct mapping between multimodal historical patterns and future target values, assuming a deterministic relationship between the sentiment signals and the target variables.

4.7 Scenario and Emotion-Specific Analysis

To further examine performative effects, we evaluate model behavior under different sentiment regimes and emotional components. We test the effect of each sentiment intensity index separately to examine whether fear-dominated sentiments or optimism-dominated sentiments are more beneficial to epidemiological forecasts. The results in Table 4 highlight which public emotions exert the strongest influence on epidemic trajectory predictions. Surprisingly, we observe that happiness intensity index leads to the highest accuracy increase for vaccination and testing forecasts. For cases and deaths, the highest accuracy is achieved without any sentiment information, but including happiness index still outperformed other intensity indices. This could be explained as happiness intensity carries the most "interesting" signals that are not entirely inherent in the target time series. We hypothesize that happiness intensity correlates with pro-social compliance behaviors as well as social interactions. This not only reflects public willingness to engage in testing or vaccination, but also signals increased human interactions as people engage in more gatherings and events.

5 Conclusion and Future Work

In this work we introduced **SentiFM**, a sentiment-aware framework that augments epidemic forecasting with social-media derived emotional signals. Across classical machine learning (SARIMAX), neural network (LSTM), foundation (Chronos-2), and generative (Llama-3) models, we systematically evaluated how Twitter-based

Table 3: Cross-Architecture Performance Evaluation (shown in MSE). Comparison across three epidemiological targets.

Model Architecture	New Cases	New Deaths	Vaccinations
Classical Baselines			
SARIMAX (Univariate)	0.1795	0.0508	0.0031
SARIMAX (Raw Sentiment)	0.1673	0.0492	**0.0003**
SARIMAX (Senti-Shift $\tau = 14$)	0.1583	0.0445	0.0004
Generative AI			
Llama-3 (Zero-Shot Base)	0.0754	0.0435	0.0018
Llama-3 (QLoRA Fine-Tuned)	0.0713	0.0410	0.0015
Proposed Method			
LSTM (Multimodal Fusion)	**0.0067**	**0.0125**	**0.0003**

Table 4: Effect of each sentiment intensity feature on epidemiological forecast with Chronos-2 (shown in RMSE, best configuration for each row highlighted in red)

COVID feature	Without sentiment	+ Valence intensity	+ Fear intensity	+ Anger intensity	+ Happiness intensity	+ Sadness intensity
People vaccinated	0.00421	0.00216	0.00210	0.00213	0.00206	0.00209
Total tests	0.00898	0.00457	0.00458	0.00506	0.00418	0.00467
Total cases	0.0677	0.0681	0.0690	0.0693	0.0681	0.0691
Total deaths	0.0286	0.0288	0.0293	0.0293	0.0288	0.0294

sentiment covariates influence COVID-19 prediction. Our results show that even simple naive concatenation can reduce error for highly correlated targets such as vaccinations and testing, while a dual-branch LSTM with multimodal fusion consistently achieves the lowest scaled MSE across all architectures. Chronos-2 in-context learning benefits vaccination and testing but yields mixed gains for cases and deaths, suggesting that some behavioral information is already implicit in the epidemiological history. Emotion-specific analysis further reveals that happiness intensity, rather than fear, provides the most useful exogenous signal, likely reflecting both pro-social compliance and increased social activity.

Future work will include expanding the time interval of forecasts to multiple years to allow for analysis on each sentiment index's effect during each sub-period of the pandemic. In addition, we plan to replace standard sentiment time series data with public-health–tuned LLM classifiers for richer event representations[**?**]. This enables the use of a wider range of infoveillance information in real-time forecasting.

References

[1] [n. d.]. Global COVID-19 Overview | WHO COVID-19 dashboard — data.who.int. https://data.who.int/dashboards/covid19. [Accessed 04-11-2025].

[2] Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. 2025. Chronos-2: From Univariate to Universal Forecasting. arXiv:2510.15821 [cs.LG] https://arxiv.org/abs/2510.15821

[3] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. 2024. Chronos: Learning the Language of Time Series. arXiv:2403.07815 [cs.LG] https://arxiv.org/abs/2403.07815

[4] Chenghua Gong, Rui Sun, Yuhao Zheng, Juyuan Zhang, Tianjun Gu, Liming Pan, and Linyuan Lv. 2025. EpiLLM: Unlocking the Potential of Large Language Models in Epidemic Forecasting. arXiv:2505.12738 [cs.LG] https://arxiv.org/abs/2505.12738

[5] Raj Gupta, Ajay Vishwanath, and Yinping Yang. 2022. COVID-19 Twitter Dataset with Latent Topics, Sentiments and Emotions Attributes. https://doi.org/10.3886/E120321V12. doi:10.3886/E120321V12 Version 12, distributed 2022-06-25.

[6] Mohammad Hossein Samaei, Faryad Darabi Sahneh, Lee W Cohnstaedt, and Caterina Scoglio. 2025. EpidemIQs: Prompt-to-Paper LLM Agents for Epidemic Modeling and Analysis. *arXiv preprint arXiv:2510.00024* (2025).

[7] Youngchul Song and Byungun Yoon. 2024. Prediction of infectious diseases using sentiment analysis on social media data. *Plos one* 19, 9 (2024), e0309842.

[8] Christos Stefanis, Elpida Giorgi, Konstantinos Kalentzis, Athanasios Tselemponis, Evangelia Nena, Christina Tsigalou, Christos Kontogiorgis, Yiannis Kourkoutas, Ekaterini Chatzak, Ioannis Dokas, et al. 2023. Sentiment analysis of epidemiological surveillance reports on COVID-19 in Greece using machine learning models. *Frontiers in Public Health* 11 (2023), 1191730.

[9] Shu-Feng Tsao, Helen Chen, Therese Tisseverasinghe, Yang Yang, Lianghua Li, and Zahid A Butt. 2021. What social media told us in the time of COVID-19: a scoping review. *Lancet Digit. Health* 3, 3 (March 2021), e175–e194.

[10] A Vijayaraj, K Bhavana, S SreeDurga, and S Lokesh Naik. 2022. Twitter based sentimental analysis of Covid-19 observations. *Mater. Today* 64 (May 2022), 713–719.

[11] Xinlei Wang, Maike Feng, Jing Qiu, Jinjin Gu, and Junhua Zhao. 2024. From news to forecast: Integrating event analysis in llm-based time series forecasting with reflection. *Advances in Neural Information Processing Systems* 37 (2024), 58118–58153.

[12] Zhiyuan Zhao, Haoxin Liu, Alexander Rodríguez, and B Aditya Prakash. 2025. Performative time-series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*. 3968–3979.

[13] Xinyu Zhou, Jiaqi Zhou, Chiyu Wang, Qianqian Xie, Kaize Ding, Chengsheng Mao, Yuntian Liu, Zhiyuan Cao, Huangrui Chu, Xi Chen, Hua Xu, Heidi J Larson, and Yuan Luo. 2025. PH-LLM: Public Health Large Language Models for infoveillance. (Feb. 2025).

[14] Shengjun Zhu, Siyu Liu, Yang Li, Qing Lei, Hongyan Hou, Hewei Jiang, Shujuan Guo, Feng Wang, Rongshang Chen, Xionglin Fan, et al. 2024. CovidLLM: a robust large language model with missing value adaptation and multi-objective learning strategy for predicting disease severity and clinical outcomes in COVID-19 patients. *Current Proteomics* 21, 6 (2024), 591–605.