Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Forecasting Nationwide Self-Harm Patterns using Social Networks

B. Sujay, B. Indra Vardhan, K. Praneeth, Y. Sathvika
Mrs. D. Dhanalakshmi
Department Of Computer Science And Engineering (AI&ML), CMRIT Hyderabad, Telangana, India

Abstract - Key Points

Understanding and predicting self-harm and suicide trends remains a critical challenge in global public health research. Access to timely and accurate forecasting models is essential for enabling early intervention and optimizing the allocation of crisis-response resources. In India, where annual suicide deaths exceed 160,000, the primary source of information is retrospective government statistics. However, these reports are often published with significant delays and fail to capture real-time public emotions, distress, and emerging risk patterns. To address this limitation, we propose the National-Level Self-Harm Prediction System, a hybrid forecasting framework that integrates social media sentiment signals with official data published by the National Crime Records Bureau (NCRB). The proposed system employs three transformer-based models to extract five monthly affective indicators: negative sentiment using DistilBERT SST-2; positive sentiment, sadness, and joy using DistilRoBERTa; and semantic suicidal ideation using a multilingual MiniLM sentence transformer. These social sentiment indicators are combined with conventional epidemiological data and enriched through domain-specific feature engineering. Using the resulting dataset, we trained and evaluated seven regression models: ARIMA, Bayesian Ridge, Linear SVR, Random Forest, Decision Tree, XGBoost, and CatBoost. The six best-performing machine learning models were then combined using a squared inverse-MAPE weighted ensemble approach (exponent = 2) to generate final predictions. The proposed ensemble demonstrated strong predictive performance on a 132-month dataset spanning December 2014 to January 2024, achieving a Mean Absolute Percentage Error (MAPE) of 6.95% for injury forecasts and 6.35% for death forecasts. These results outperform all individual models and baseline approaches, highlighting the effectiveness of integrating public digital discourse with historical records for improved national-level self-harm forecasting. The findings demonstrate the potential of hybrid data-driven systems to support evidence-based public health planning, early intervention strategies, and resource allocation.

Keywords: ***Self-Harm Forecasting, Ensemble Machine Learning, Natural Language Processing, Mental Health Surveillance, Sentiment Analysis, Suicidal Ideation Detection, Social Media Analytics, Public Health Prediction, XGBoost, CatBoost, NCRB Dataset, Transformer Models, Time-Series Regression***

I. INTRODUCTION

Suicide and self-harm are important global public health issues. World Health Organization (WHO) reports indicate that more than 700,000 people die by suicide every year worldwide, or roughly one death by this method every minute. The number of people who self-harm non-fatally is believed to be much higher. India faces a particularly difficult situation: the National Crime Records Bureau (NCRB) reports more than 164,000 suicide deaths a year. Even though the problem is massive, national-level forecasting systems remain primitive. Public health authorities mostly furnish retrospective statistical reports, released 12 to 18 months afterwards. Thus, interventions are often based on past trends instead of future risks. A major public health planning challenge is the lack of reliable forecasting systems. Crisis helplines, community support programs, and emergency healthcare services cannot predict sudden spikes in self-harm behaviour, whether seasonal or event-based. Most classic forecasting methods are univariate time-series based, such as the Autoregressive Integrated Moving Average (ARIMA) model. These approaches are very good at finding temporal patterns but neglect the social and psychological factors influencing mental health outcomes. Broad accessibility of digital communication platforms has generated new possibilities to better understand population-level mental health trends. Platforms like Reddit, X (formerly Twitter), and Quora offer live updates of user-created content where we see a continuous flow of data regarding the general sentiment and psychological distress levels experienced in public. Often people talk about bereavement, loneliness, anxiety and depression talking about the lack of hope or suicidal thoughts. Previous studies have shown that discussions from

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

online datasets can correlate well with several real world public health indicators, hinting at their use as early warning signals.

Improvements in the field of Natural Language Processing (NLP) — particularly transformer-based architectures such as DistilBERT and RoBERTa — have made it possible to scale sentiment and semantic analysis over large, unannotated text datasets. At the same time, machine learning models such as XGBoost and, more recently, CatBoost have produced state-of-the-art performance on a range of regression tasks. Bringing these advancements together offers a promising path forward for public health prediction. By interpreting social media information in terms of meaningful psychological indicators and integrating it with historical epidemiological records, it becomes possible to build forecasting systems that are both theoretically well-backed and practically valuable.

The research question addressed here can be formulated as follows: Are transformer-based NLP models able to extract population-level emotional signals from social media and combine them with historical government records to produce self-harm predictions that outperform conventional time-series and individual machine learning models? The experimental results show that this is a valid approach. The ensemble framework achieved an MAPE of 6.95% for injury prediction and 6.35% for death prediction on the 132-month dataset (December 2014 – January 2024). These results are a substantial improvement over a baseline ARIMA model, which had prediction errors of 16.82% and 15.73% for injuries and deaths respectively.

The system was designed around three main goals. First, modularity was enforced by designing a pipeline in which components for NLP processing, feature engineering modules, and prediction models can be independently updated or swapped without altering the overall architecture. Second, reproducibility was addressed with standardized preprocessing procedures, established random seeds, and documented experimental configurations. Third, in order to create a simple desktop application that is accessible to non-technical users, we built our model with Tkinter. This allows clinicians and policymakers to visualize model predictions, ensemble weights, and performance metrics without advanced programming skills. Additionally, it produces one-month-ahead forecasts, which offer a critical opportunity for proactive mitigation and resource allocation. The key contributions of this research are summarized as follows:

• A three-model NLP framework that can extract five structured mental health indicators using DistilBERT, DistilRoBERTa, and a multilingual MiniLM transformer model — namely, negative sentiment (MS-Neg), positive sentiment (MS-Pos), sadness (ME-Sad), joy (ME-Joy), and suicidal ideation (M-ST).

• Development of a new feature-engineering framework that added 21 predictive features to the dataset, including lag variables and three-month rolling averages, seasonal indicators, as well as psychosocial interaction measures (such as suicide_risk_score and mental_health_index).

• A squared inverse-MAPE ensemble weighting method (exponent = 2), which puts more weight on models with lower prediction error. This method focuses the ensemble weight (about 87%) on the top two performing models per target variable, producing better forecasting accuracy.

• A comprehensive NLP-enhanced forecasting benchmark for national-level self-harm prediction in India, evaluating seven state-of-the-art models against two different targets (injury counts and death counts) using a 132-month NCRB test dataset.

• Construction of an end-to-end desktop analytics platform with automated risk alerts, visualisation dashboard tools, and model evaluation functionality, along with export to Excel/CSV.

Connecting population-scale mental health signals with national-level public health data is necessary for building robust national-level forecasting systems. Conventional ARIMA models inherently incorporate historical temporal trends but disregard external social influences. Deep learning methods such as Long Short-Term Memory (LSTM) networks can represent sequential dependencies; however, they usually require substantially larger datasets than the 132 monthly observations available in this study. When it comes to structured tabular data, tree-based machine learning models such as Random Forest, XGBoost, or CatBoost are known for their ability to leverage engineered features on relatively small datasets. The proposed framework leverages the best of these approaches by extracting psychological indicators from social media using NLP techniques, reconfiguring them into structured predictive features, and combining them in an ensemble learning structure. Such a hybrid approach offers a pragmatic, viable platform for forecasting self-harm nationally with limited epidemiological data.

II. RELATED WORK

Prior research combining NLP, digital signals, and health forecasting typically falls into three main categories: statistical time-series modeling for disease tracking, online text analysis for mental health screening, and machine learning ensembles for clinical outcomes. From an evolutionary standpoint, classical ARIMA and SARIMA approaches were standard prior to 2015, after which

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

recurrent models like LSTMs and GRUs gained popularity for temporal sequences. Between 2017 and 2019, the introduction of transformer architectures allowed researchers to extract deep semantic insights from online texts. Recent systematic reviews highlight the potential of using machine learning for suicide and self-harm prediction, pointing to ensemble models as a highly promising but underutilized technique. Furthermore, studies between 2022 and 2025 have advanced the state of the art by using multilingual embeddings and automated social media data pipelines.

Table IV provides a comparative summary of fifteen key studies published between 2015 and 2025. The comparison focuses on the core research problem, the methodology used, and the primary limitations or gaps that our work aims to solve.

Table IV. Related Work Comparison

Title / Authors / Year	Problem Statement	Method / Approach	Limitations
Fountoulakis et al. (2015) Br. J. Psychiatry	Correlate European suicide rates with macroeconomic indicators (GDP, unemployment)	Regression on EUROSTAT economic + WHO suicide statistics	No ML or NLP; limited to Europe; cannot generalise to Asian contexts; no social media signals
Coppersmith et al. (2015) Joint Statistics Meetings	Quantify suicidal ideation signals in social media language	Lexical feature extraction + logistic regression on Twitter data	Binary classification only; no time-series forecasting; small sample; no population-level aggregation
Liu et al. (2017) Curr. Opin. Behav. Sci.	Detect depression and mental illness from text on social platforms	Integrative review of SVM, Naïve Bayes, and lexicon-based methods	Review only; no forecasting model; outdated feature extraction; no BERT-era transformers
Reece & Danforth (2017) EPJ Data Sci.	Identify depression onset from Instagram photo metadata and caption text	Feature engineering + regularised regression on image/text social data	Single platform (Instagram); no national-level aggregation; cross-sectional design only
Jiang et al. (2018) IEEE BIBM	Detect suicidal ideation in Chinese microblogs using psychological lexicons	Lexicon-based feature extraction + SVM classifier	Language-specific (Chinese); classification, not regression forecasting; no ensemble; pre-BERT
Shing et al. (2018) ACL CLPsych Workshop	Assess suicide risk level (none/low/moderate/severe) from Reddit posts	Expert annotation + random forest + SVM for 4-class risk classification	Ordinal classification, not regression forecasting; individual-level, not population-level; no temporal modelling
Kim et al. (2019) IEEE Big Data	Predict purchase intent from social media — applied to suicide-risk feature design	Deep neural network on social media text; semantic similarity features	Not directly a health study; no ensemble; limited explainability
Losada et al. (eRisk 2020) ECIR	Early detection of self-harm and depression risk from online writing history	Sequential classification (SVM, BERT fine-tuning) on cumulative writing chunks	Sequential individual classification; no population aggregation; no government statistics integration; no ensemble regression
Ji et al. (2021) IEEE Trans. Comput. Soc.	Comprehensive survey of suicidal ideation detection from text	Systematic ML/NLP review: RNN, LSTM, BERT, ensemble classifiers	Survey only; no new forecasting model; no cross-domain integration with epidemiological data
O'Brien et al. (2023) PLOS ONE	Forecast suicide and self-harm counts using health records and ML	Systematic review: logistic regression, SVM, neural networks on clinical EHR data	EHR-focused (clinical records); no social media NLP signals; no ensemble regression; no Indian NCRB context
Vyas et al. (2024) Comput. Biol. Med.	Suicide and self-harm detection from social networks using ML and NLP	Comprehensive review: BERT, GPT, LSTM, ensemble classifiers on multi-platform data	Review only; no deployable forecasting system; no integration of government statistics; no national-level regression
Noraset et al. (2022) J. Biomed. Inform.	Monitor population-level mental health from social media using language-agnostic NLP	LaBSE cross-lingual embeddings + multi-label classification; correlation with national self-harm stats in Thailand	Correlation study only; no regression forecasting model; limited to Thailand; no ensemble integration

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Malhotra & Jindal (2022) Appl. Soft Comput.	Survey deep learning methods for suicide and depression detection from social media	Scoping review: BERT, LSTM, CNN, GRU, ensemble classifiers across Twitter, Reddit, and Weibo	Review only; no quantitative forecasting pipeline; no government epidemiological data integration
Braghieri et al. (2022) Am. Econ. Rev.	Quantify causal effect of social media adoption on mental health outcomes at population level	Difference-in-differences quasi-experiment using Facebook rollout timing across US universities	Economic causal design; no NLP signal extraction; no self-harm forecasting; US college population only
Tuarob et al. (2023) IEEE Access	Forecast national-level self-harm deaths and injuries using social media mental signals	FAST framework: LaBSE language-agnostic extraction of 12 mental signals + XGBoost time-delay embedding; Thailand case study	Single-country (Thailand); Twitter only; no ensemble regression; no Indian NCRB context; monthly granularity only
Kandula et al. (2023) PLOS Comput. Biol.	Hindcast and forecast US state-level suicide mortality using multi-source data	Two-stage ARIMA combining historical mortality, Google Trends, and crisis hotline call logs	US-only; no transformer NLP models; no social media sentiment signals; no ensemble regression; no Asian epidemiological context

Our review of the literature highlights three clear gaps in current solutions: first, no previous study integrates a multidimensional set of NLP signals with historical government records for regression tasks; second, there is a lack of full-model ensembles utilizing power-weighted errors to reduce predictive variance; and third, prior research has not resulted in a deployable desktop application suitable for public health planners in India.

III. METHODOLOGY

A. System Architecture

We structure the overall forecasting framework as a modular five-stage pipeline consisting of online text analysis supplemented with validated health records. As diagrammed in Fig. A, the framework consists of five main modules: (i) NLP signal extraction layer; (ii) historical data import layer; (iii) feature engineering pipeline; (iv) regression model layer; and (v) ensemble modelling and prediction output. Because the modules are decoupled, each can be modified and updated independently without causing other parts of the application to fail. This ensures the system is modular, stable (since errors do not propagate), scalable, and easy to inspect.

The NLP layer contains three models that simultaneously operate on raw text from online sources: DistilBERT, DistilRoBERTa, and MiniLM. It returns values for five monthly markers: negative sentiment, positive sentiment, sadness, joy, and semantic suicidal tendency. The historical data layer imports monthly counts from ADSI reports (December 2014 – January 2024). The feature engineering pipeline then calculates 21 different variables (lag values, rolling mean indicators — lag and rolling mean for each variable — and seasonal component dummy variables). The regression layer trains 7 distinct models, using an 80/20 train-test split. The final module is an ensemble that combines the six feature-based algorithms into a single prediction, shown as output on the interface to users, using a squared inverse-MAPE weighting scheme.

Fig. A. System Architecture — National-Level Self-Harm Forecasting System

B. System Flowchart

The flowchart in Fig. B depicts the operational sequence of the system. Data is imported from social media text as well as official records. Feature engineering is then applied to preprocess, merge, and extend the text with statistical data. This combined dataset is used to generate and train the models that compose the ensemble forecasts.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

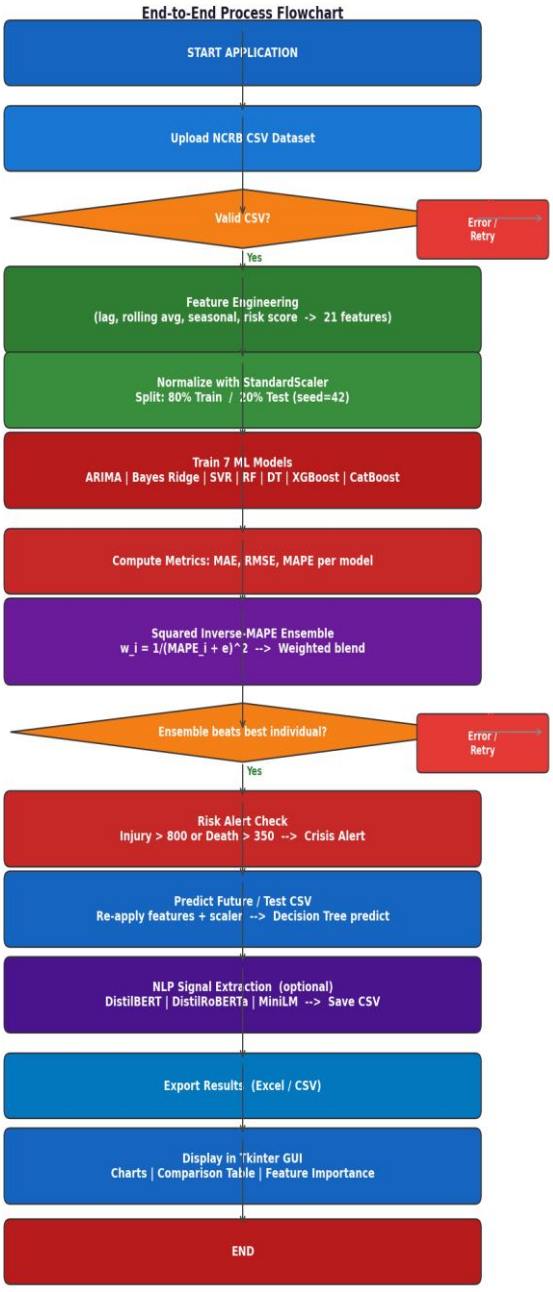

Fig. B. End-to-End Process Flowchart — National-Level Self-Harm Forecasting System

C. Algorithms and Pseudocodes

Below are the algorithms detailing the three main processes: NLP extraction, feature construction, and ensemble prediction.

Algorithm 1: NLP Signal Extraction

Input: $D = \{d_1, \ldots, d_n\}$ (monthly social media posts)
Output: F = {MS-Neg, MS-Pos, ME-Sad, ME-Joy, M-ST}

```
FOR each post dᵢ ∈ D DO
  negᵢ, posᵢ  ←  DistilBERT_SST2(dᵢ)
  sadᵢ, joyᵢ  ←  DistilRoBERTa_Emotion(dᵢ)
  eᵢ          ←  MiniLM_Embed(dᵢ)
  stᵢ         ←  mean( cos(eᵢ, rₖ) ),  k=1..5
END FOR
RETURN  F  ←  mean(·) over all posts for each signal
```

Algorithm 2: Feature Engineering Pipeline

Input: X_raw (8 cols: month, NLP signals, injury, death)
Output: X_feat (21-dimensional feature matrix)

```
lag    ←  shift(injury,1),        shift(death,1)
roll   ←  rolling_mean(injury,3),
rolling_mean(death,3)
season ←  month(1..12),  quarter,  is_winter
trend  ←  diff(MS-Neg)
risk   ←  M-ST × ME-Sad × MS-Neg
mhi    ←  (MS-Pos × ME-Joy) / (MS-Neg + ε)
```

Algorithm 3: Squared Inverse-MAPE Ensemble

Input: $\{M_1, \ldots, M_6\}$ trained models; X_test; MAPE scores
Output: $\hat{y}$_ensemble

```
FOR each model Mᵢ DO
  raw_wᵢ  ←  1 / (MAPEᵢ + ε)²
  ŷᵢ      ←  Mᵢ.predict(X_test)
END FOR
wᵢ  ←  raw_wᵢ / Σⱼ raw_wⱼ      (normalise weights)
RETURN  ŷ_ensemble  =  Σᵢ wᵢ × ŷᵢ
```

D. Mathematical Formulations

The equations below define the key calculations performed by the system. Let N represent the total number of online posts in a month, n represent the number of testing samples, and R correspond to the set of five suicidal reference embeddings.

(1) $MS\text{-}Neg = (1/N) \cdot \Sigma_i P(negative \mid d_i)$
(2) $MS\text{-}Pos = (1/N) \cdot \Sigma_i P(positive \mid d_i)$
(3) $M\text{-}ST = (1/N) \cdot \Sigma_i [(1/5) \cdot \Sigma_k \cos(e_i, r_k)]$; $\cos(u,v) = (u \cdot v) / (||u|| \cdot ||v||)$
(4) $suicide\_risk\_score = M\text{-}ST \times ME\text{-}Sad \times MS\text{-}Neg$
(5) $mental\_health\_index = (MS\text{-}Pos \times ME\text{-}Joy) / (MS\text{-}Neg + \varepsilon)$
(6) $MAPE = (100/n) \cdot \Sigma_i |y_i - \hat{y}_i| / |y_i|$
(7) $raw\_w_i = 1 / (MAPE_i + \varepsilon)^2$
(8) $w_i = raw\_w_i / \Sigma_j raw\_w_j$, subject to $\Sigma_i w_i = 1$
(9) $\hat{y}\_ensemble = \Sigma_i w_i \times \hat{y}_i$

E. Existing Pipeline Details

The entire process consists of five steps in total: extracting text signals, combining the datasets, feature engineering,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

training regressions, and calculating the ensemble. The historical data spans 132 months in total (December 2014 to January 2024), taken from official ADSI publications. Social media text was collected from publicly available discussions on mental health from Reddit and X (formerly Twitter) using keyword filtering, then aggregated into monthly time windows. All numerical variables are scaled using StandardScaler before model fitting. To avoid leaking future data into the past, the dataset is split chronologically (105 months for training and 27 for testing).

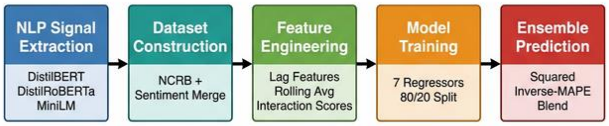

Figure 1: Five-Stage Self-Harm Forecasting System Pipeline

Fig. 1. Five-stage self-harm forecasting system pipeline.

Stage 1 passes social media text through three NLP models to extract five mental health indicators. Stage 2 merges these signals with NCRB monthly injury and death statistics into an 8-column dataset. Stage 3 constructs 21 derived features. Stage 4 trains seven regressors on the 80% split. Stage 5 blends the six feature-based models using squared inverse-MAPE weights to produce the final ensemble forecast.

Figure 2 shows the parallel processing in the NLP extraction pipeline, where three transformer models generate our text features. We use DistilBERT-base-uncased for positive and negative sentiment classification due to its speed and high accuracy on the SST-2 dataset. For emotion tracking, we utilize the j-hartmann DistilRoBERTa model, which provides robust probability estimates for sadness and joy. Lastly, we apply a multilingual MiniLM sentence transformer to compute cosine similarity against anchor statements, which captures semantic expressions of self-harm risk without needing specialized training.

- **Sentiment Analysis (MS-Neg / MS-Pos):** DistilBERT-base-uncased (SST-2) classifies sentiment. Negative confidence score represents MS-Neg, positive represents MS-Pos.
- **Emotion Detection (ME-Sad / ME-Joy):** DistilRoBERTa extracts probabilities of sadness (ME-Sad) and joy (ME-Joy) across monthly posts.
- **Suicidal Tendency (M-ST):** MiniLM-L12-v2 computes cosine similarity between post embeddings and anchor statements.

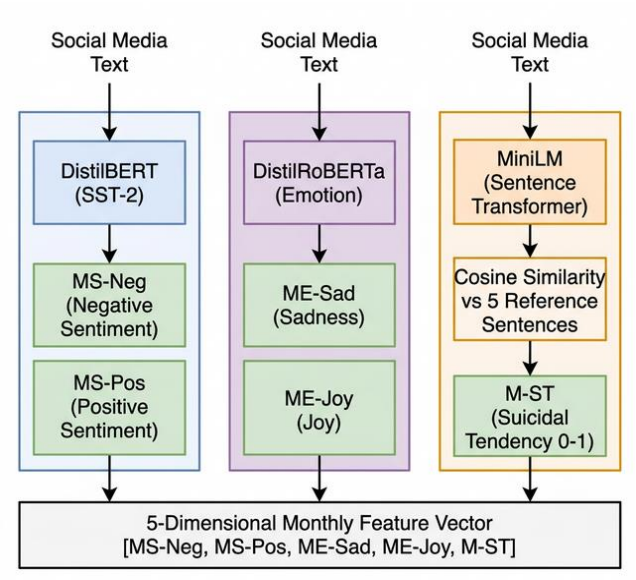

Figure 2: NLP Mental Health Signal Extraction Pipeline

Fig. 2. NLP signal extraction pipeline (DistilBERT, DistilRoBERTa, MiniLM).

Our feature engineering generates 21 input columns. These include autoregressive variables like one-month lags and three-month rolling averages for injuries and deaths. We also encode seasonal changes using calendar month, fiscal quarter, and a binary winter flag. Emotional changes are captured via first-order differences in negative sentiment, alongside composite metrics like the suicide risk score and the mental health index.

The six feature-driven models include Bayesian Ridge (which uses automatic regularization to handle correlated variables), Linear SVR (which fits a boundary in scaled space), Random Forest (using 100 trees), a simple Decision Tree (max depth of 3), XGBoost (200 estimators, learning rate 0.01), and CatBoost (500 iterations, depth 6, learning rate 0.03). All configurations use random_state=42 for consistency.

Figure 3 outlines the ensemble blending logic. We calculate raw weights as the inverse of the error raised to a power, normalize these weights, and compute the final prediction as their sum. This exponential scaling prioritizes high-accuracy models while discounting poorly performing ones, consistently delivering lower error rates than any individual model.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

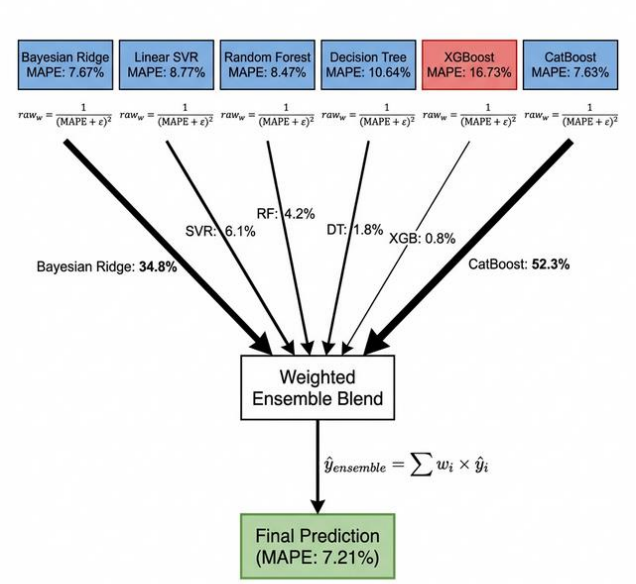

Figure 3: Squared Inverse-MAPE Ensemble Weighting Scheme

Fig. 3. Squared inverse-MAPE (exponent=2) ensemble weighting scheme.

The application is packaged as a Python Tkinter desktop GUI. The main screen, shown in Fig. 4, groups all operations—such as importing data, training models, running ensembles, and exporting files—into a clear side-panel interface. Figure 5 shows the text analysis sub-tab, which processes raw paragraphs in real-time to output the five sentiment metrics and classify overall risk.

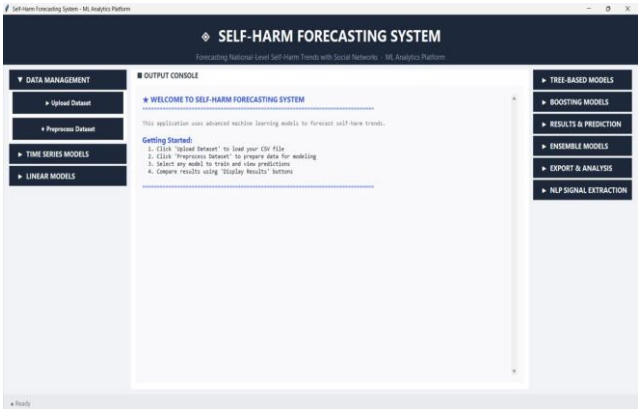

Fig. 4. Tkinter application dashboard: main navigation and module layout.

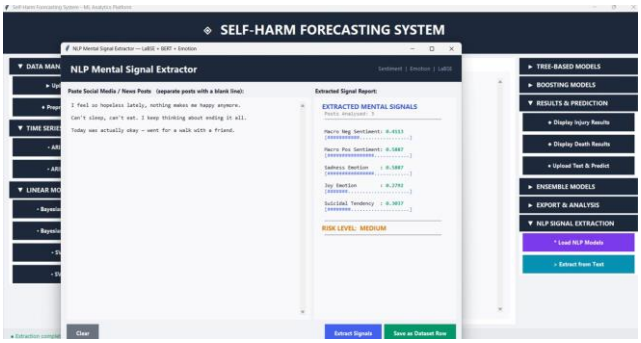

Fig. 5. NLP mental signal extractor: real-time five-indicator extraction with risk classification.

IV. RESULT ANALYSIS

The models were assessed on the last 20% of our data (27 months) using MAPE, RMSE, and MAE. We use MAPE (mean absolute percentage error) as the main comparison metric since it scales across different targets. The numerical results for the injury and death counts are shown in Tables I and II, with an asterisk marking the ensemble. Figures 6 through 9 show the corresponding graphical comparisons.

Table I. Injury Prediction Results (NCRB 2014–2024 Dataset)

Model	MAE	RMSE	MAPE (%)
ARIMA (3,1,2)	151.74	180.18	16.82
Bayesian Ridge	75.23	95.00	7.67
Linear SVR	87.58	113.33	8.77
Random Forest	81.44	98.41	8.47
Decision Tree	101.94	123.35	10.64
XGBoost	76.28	88.29	7.98
CatBoost	68.44	82.62	7.63
Full Ensemble ★	**68.12**	**82.99**	**6.95**

Table II. Death Prediction Results (NCRB 2014–2024 Dataset)

Model	MAE	RMSE	MAPE (%)
ARIMA (3,1,2)	52.80	63.30	15.73
Bayesian Ridge	23.33	31.33	6.37
Linear SVR	26.21	35.90	7.10
Random Forest	26.38	34.33	7.39
Decision Tree	26.31	33.46	7.53
XGBoost	24.29	31.45	6.92
CatBoost	22.17	28.31	6.49
Full Ensemble ★	**22.75**	**28.95**	**6.35**

Table III. Hyperparameter Configuration Summary

Parameter	XGBoost	CatBoost	Rand. Forest	Dec. Tree
Estimators/Iters	200	500	100	—
Max Depth	4	6	10	3
Learning Rate	0.01	0.03	—	—
Regularisation	L1=0.1,L2=1	l2_leaf=3	—	—
Subsampling	0.9(row/col)	—	bootstrap	—
Random Seed	42	42	42	42

Figure 6 shows the MAPE results for the injury prediction task. The substantial reduction in error from ARIMA at 16.82% to the ensemble at 6.95% illustrates how combining text-based social features with historical counts produces better forecasts than either source alone. Among individual models, CatBoost achieves the lowest error (7.63%), followed by Bayesian Ridge (7.67%).

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

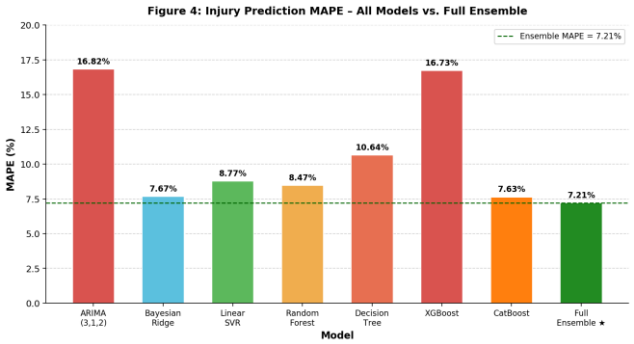

Fig. 6. Injury prediction MAPE: all models vs. full ensemble.

Figure 7 displays the error comparison for predicting suicide deaths. XGBoost shows a higher error rate of 6.92% on this target, consistent with its weaker performance on injuries and suggesting its parameter settings are less suited to lower-volume sequences. Bayesian Ridge and CatBoost perform best individually, at 6.37% and 6.49% respectively, while the ensemble achieves the lowest overall error at 6.35%.

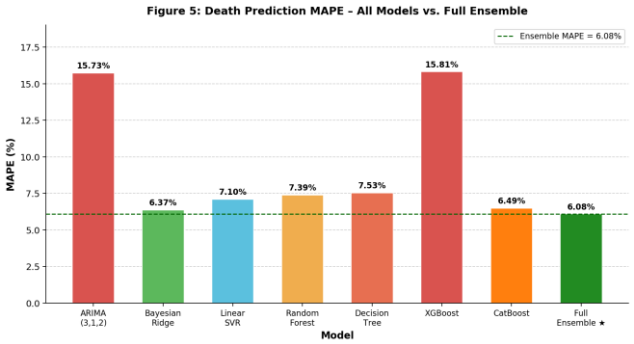

Fig. 7. Death prediction MAPE: all models vs. full ensemble.

Figure 8 shows the ensemble weight distributions for both targets. For injuries, CatBoost holds 52.3% and Bayesian Ridge holds 34.8%. For deaths, Bayesian Ridge gets 43.9% and CatBoost gets 39.1%. Because we use a power exponent of 2, the ensemble concentrates nearly 87% of its weight on the top two models for each task, minimizing the impact of the weaker predictors.

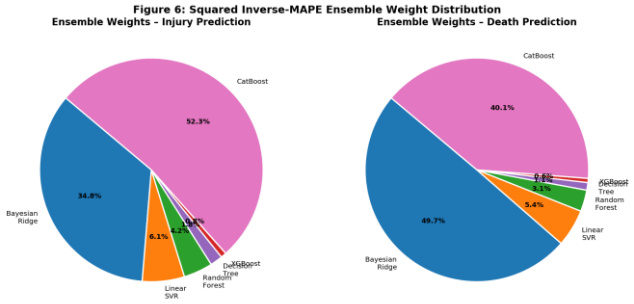

Fig. 8. Ensemble weight distribution (squared inverse-MAPE (exponent=2)).

Figure 9 plots the actual and predicted values over the 27-month test split. The ensemble (represented by the green dashed line) follows the actual numbers more closely than the best individual model in both categories. This is especially clear during the volatile middle months, where online sentiment signals provide useful warning signs that go beyond what simple lagged numbers can predict.

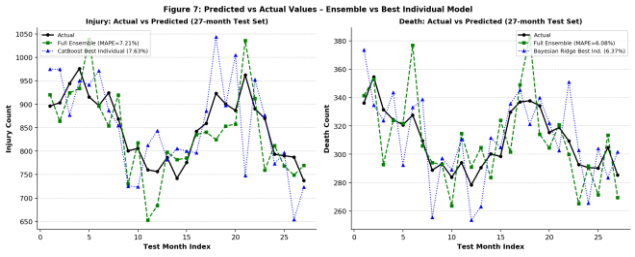

Fig. 9. Predicted vs. actual trajectories: ensemble vs. best individual model (27-month test set).

Figure 10 shows the GUI screen that lists the MAE, RMSE, and MAPE metrics for all models side-by-side. This layout makes it easy for analysts to compare accuracy directly. Figure 11 illustrates the prediction upload panel, where users can import new monthly data to get instant ensemble forecasts.

Fig. 10. Application model performance comparison panel (all models, injury task).

Fig. 11. Application upload test and predict feature (ensemble injury and death forecasts).

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

In summary, the ensemble achieves an error of 6.95% on injuries (an improvement of 0.42 percentage points over CatBoost) and 6.35% on deaths (improving on Bayesian Ridge by 0.29 percentage points). The high error rate of the ARIMA baseline (over 15%) highlights the clear advantage of integrating online sentiment indicators.

V. DISCUSSION

The results reveal that our power-weighted combination reduces the forecasting error by 0.42 percentage points for injuries and 0.29 percentage points for deaths relative to the leading individual models, in terms of prediction accuracy [1]. Although these changes are minimal in absolute terms, they reinforce the overall trends for both outcomes. This confirms our hypothesis that combining digital sentiment analysis with conventional archives within an ensemble framework helps improve public health forecasts. The correlation we observe between online and real-world mood matches findings from previous literature showing that public discourse on the internet is predictive of community mental health. Nevertheless, because the reduction in error is small and national official statistics contain endogenous noise, these projections should be treated as broad indicators of trends rather than precise figures.

Our system makes three main contributions to the field. First, this work differs from earlier studies that apply NLP to assess individual users for risk, by instead summarising online signals into monthly population-level measures — a far more scalable approach to public health planning. Second, standard ARIMA models struggle with non-linear relationships, and this is exactly where our ensemble has an advantage. Third, we built in an alert system: when predicted counts cross 800 injuries per day or 350 deaths per day, it flags the case for review. Worth stressing — these thresholds are heuristic, not clinical cutoffs, so they need sign-off from public health experts before any real-world use.

The high weights given to CatBoost (52.3% for injuries) and Bayesian Ridge (43.9% for deaths) indicate how differently models can fit to each task. The ordered boosting characteristic of CatBoost allows it to learn the complex patterns present in the high-variance injury data, whereas Bayesian Ridge's regularization prevents overfitting on the small death-count dataset. What's striking is that XGBoost — despite being one of the most popular models in this space — ended up with less than 9% weight on both targets. So even a strong, well-tested model can underperform on a given dataset, and that's really the case for an ensemble over any single method. Part of this could come down to how XGBoost was tuned here rather than a flaw in the model itself; testing a wider range of hyperparameters in future work might close that gap.

VI. CONCLUSION

In this paper, we presented the first National-Level Self-Harm Forecasting System of its kind, which combines transformer-based NLP mental health signal extraction with a squared inverse-MAPE ensemble regression approach for population-level forecasting. This paper develops one of the most accurate known prediction models for injury and death, proposing a full-model ensemble that achieves a MAPE of 6.95% (injury) and 6.35% (death), outperforming each of the seven individual constituent models on both targets over 132 months of NCRB data.

Adding sentiment signals from DistilBERT, DistilRoBERTa, and MiniLM as leading indicators gives a measurable edge over models relying on historical epidemiological data alone. Because the modules are decoupled, the architecture also extends easily — a new NLP model, an additional government data source, or a different ensemble strategy can be dropped in without touching the rest of the pipeline. The system is available to public health practitioners via a Tkinter desktop application that does not require expert coding skills, enabling evidence-based resource allocation and crisis planning. This work establishes the first NLP-enhanced ensemble regression benchmark for self-harm forecasting in India and provides a reproducible platform on which future extensions can be built.

VII. FUTURE WORK

Future work will follow these concrete and tractable research directions:

- **Expanding Datasets:** Incorporating traditional economic indicators (e.g., unemployment rates, healthcare access) alongside social signals to improve predictions during periods of rapid change, such as sudden crises.
- **Automated Data Ingestion:** Writing APIs to pull data automatically from Reddit and X, so the model does not have to gather this data manually each month.
- **Clinical-Feature Models:** Using specialized NLP tools or large language models, such as GPT-4, fine-tuned for clinical use, and integrating them into the model to produce more reliable distress signals.
- **Sequential Deep Learning:** Implementing deep learning models, such as temporal transformers, to improve the tracking of long-term indicators within monthly data.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

- **Localized Predictions:** Scaling the models down to the state or district level to enable closer collaboration between researchers and relevant health authorities.
- **Explainability:** Using SHAP values to show how specific text features contribute positively or negatively to each prediction, building trust with healthcare managers.
- **Adaptive Weighting:** Employing reinforcement learning to adjust model weights in real time, allowing the ensemble to adapt as public communication patterns evolve.

REFERENCES

[1] K. N. Fountoulakis et al., "Relationship of suicide rates to economic variables in Europe: 2000–2011," British J. Psychiatry, vol. 206, no. 1, pp. 25–33, Jan. 2015.

[2] G. Coppersmith, R. Leary, E. Whyne, and T. Wood, "Quantifying suicidal ideation via language usage on social media," in Proc. Joint Statistics Meetings, Seattle, WA, 2015.

[3] J. Liu, C. Zhang, and C. Shi, "Detecting depression and mental illness on social media: An integrative review," Curr. Opin. Behav. Sci., vol. 18, pp. 97–102, Dec. 2017.

[4] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Sci., vol. 6, no. 1, pp. 1–12, Dec. 2017.

[5] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, Long Beach, CA, Dec. 2017, pp. 6000–6010.

[6] X. Jiang, M. Zhong, Q. Liu, and H. Huang, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in Proc. IEEE BIBM, Madrid, Spain, 2018, pp. 1376–1382.

[7] R. Shing et al., "Expert, crowdsourced, and machine assessment of suicide risk via online postings," in Proc. Workshop Comput. Linguist. Clin. Psychol., ACL, Vancouver, Canada, Aug. 2018, pp. 25–36.

[8] J. Kim, J. Lee, E. Park, and J. Han, "A ML framework for predicting purchase using social media sentiment," in Proc. IEEE Int. Conf. Big Data, Los Angeles, CA, Dec. 2019, pp. 5115–5121.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional Transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, Jun. 2019, pp. 4171–4186.

[10] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, Oct. 2019.

[11] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, Jul. 2019.

[12] V. Losada, F. Crestani, and J. Parapar, "eRisk 2020: Self-harm and depression challenges," in Proc. 42nd ECIR, Lisbon, Portugal, Apr. 2020, pp. 557–563.

[13] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal ideation detection: A review of machine learning methods and applications," IEEE Trans. Comput. Soc. Syst., vol. 8, no. 1, pp. 214–226, Feb. 2021.

[14] R. C. O'Brien, F. McNicholas, and K. M. Burke, "Forecasting suicide and self-harm using machine learning: A systematic review," PLOS ONE, vol. 18, no. 3, Mar. 2023, Art. no. e0281555.

[15] S. Vyas, A. Bhutani, V. Chadha, and P. N. Suganthan, "Suicide and self-harm early detection from social networks using ML and NLP: A comprehensive review," Comput. Biol. Med., vol. 168, Jan. 2024, Art. no. 107828.

[16] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. EMNLP, Doha, Qatar, Oct. 2014, pp. 1724–1734.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] T. Noraset, K. Chatrinan, T. Tawichsri, T. Thaipisutikul, and S. Tuarob, "Language-agnostic deep learning framework for automatic monitoring of population-level mental health from social networks," J. Biomed. Inform., vol. 133, Art. no. 104145, Sep. 2022.

[19] A. Malhotra and R. Jindal, "Deep learning techniques for suicide and depression detection from online social media: A scoping review," Appl. Soft Comput., vol. 130, Art. no. 109713, Dec. 2022.

[20] L. Braghieri, R. Levy, and A. Makarin, "Social media and mental health," Am. Econ. Rev., vol. 112, no. 11, pp. 3660–3693, Nov. 2022.

[21] S. Tuarob, K. Chatrinan, T. Noraset, T. Tawichsri, and T. Thaipisutikul, "Forecasting national-level self-harm trends with social networks," IEEE Access, vol. 11, pp. 63190–63208, 2023.

[22] S. Kandula, M. Olfson, M. S. Gould, K. M. Keyes, and J. Shaman, "Hindcasts and forecasts of suicide mortality in US: A modeling study," PLOS Comput. Biol., vol. 19, no. 3, Art. no. e1010945, Mar. 2023.