

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452

NAAS Rating (2026): 4.49

Maths 2026; 11(8): 160-168

© 2026 Stats & Maths

<https://www.mathsjournal.com>

Received: 13-06-2026

Accepted: 15-07-2026

Diana Atieno OpiyoDepartment of Mathematics and
Physics, Technical University of
Mombasa, P.O. Box 90420-
80100, Mombasa, Kenya**Cyril Wandera**Department of Mathematics,
Statistics and Computing, Rongo
University, P.O. Box 103-40404,
Rongo, Kenya**Dorca Nyamusi Stephen**Department of Mathematics and
Physics, Technical University of
Mombasa, P.O. Box 90420-
80100, Mombasa, Kenya

Deep learning model on suicidal ideation detection using social media text data

Diana Atieno Opiyo, Cyril Wandera and Dorca Nyamusi StephenDOI: <https://www.doi.org/10.22271/maths.2026.v11.i8b.2677>

Abstract

The Coronavirus Disease 2019 pandemic caused widespread psychological distress and changed the way people express that distress online. This study examined suicidal ideation among Kenyan users of the social media platform X, formerly Twitter, during and after the pandemic, and developed a deep learning model to detect such expressions in real time. A complete enumeration of publicly available posts that contained any of twenty distress-related keywords, originated from users located in Kenya, and were published between 1 January 2020 and 31 July 2021 returned 6,651 posts. After cleaning and polarity-based labelling, 6,265 posts were retained and split into 5,012 training and 1,253 test posts. Weekly post counts were tested for a monotonic trend using the Mann-Kendall test. A Long Short-Term Memory network was trained to classify the posts and compared with logistic regression, a naive Bayes classifier and a random forest, using accuracy, recall and AUC-ROC. Differences between classifiers were assessed using pairwise McNemar tests. Posting volume rose sharply between late February and April 2020, reaching a maximum of 165 posts in the week beginning 6 April 2020, and a second peak of 127 posts followed the renewed restrictions of March 2021. Across the whole period the Mann-Kendall test confirmed a significant declining trend ($Z = -5.64$, $\tau = -0.44$, slope = -0.75 posts per week, $p < 0.001$). On the held-out test set, LSTM achieved an accuracy of 92.1%, an F1-score of 94.4% and an area under the curve of 94.5%. It achieved the highest accuracy and F1-score of the four classifiers, compared to traditional models. The area under the curve was highest for the random forest at 96.4%, followed by logistic regression at 95.7%. The pairwise tests confirmed that every difference between models was significant ($p < 0.001$). The model was deployed as a web application that screens several platforms at once and returns a single risk profile. The findings show that social media text carries measurable signals of population mental health, and that deep learning can support, though not replace, professional clinical judgement.

Keywords: Suicidal ideation, deep learning, long short-term memory, text classification, Mann-Kendall trend test, social media analytics

1. Introduction

Suicidal ideation refers to thoughts of ending one's own life, ranging from simple thoughts to detailed plans. The Coronavirus Disease 2019 (COVID-19) pandemic raised the burden of such thoughts worldwide. In the United States, 10.7% of adults reported having seriously considered suicide in the preceding 30 days, rising to 25.5% among those aged 18 to 24 ^[1]. A household survey in Eswatini during the first wave found moderate to severe psychological distress in almost half of respondents ^[2], and a study of Chinese university students linked pandemic stress and social isolation to suicidal ideation³. In Kenya the burden was already high before the pandemic: a survey of 9,742 students reported a 22.6% prevalence of suicidal ideation, with depression, stress and economic hardship emerging as risk factors⁴. Clinical evidence points the same way, with emergency-room presentations for suicide attempts becoming both more frequent and more severe during the pandemic ^[5].

Social media platforms have become places where distress is expressed openly, which makes them a useful source of data for mental health research. Several studies have applied machine learning to this problem. Sawhney *et al.* showed that recurrent architectures capture the connotations of suicidal language⁶. Metzler *et al.* hand-labelled 3,202 posts into twelve content

Corresponding Author:**Diana Atieno Opiyo**Department of Mathematics and
Physics, Technical University of
Mombasa, P.O. Box 90420-
80100, Mombasa, Kenya

categories and compared word-frequency and deep learning classifiers⁷. Haque *et al.* collected 49,178 posts using eighteen keywords and compared Long Short-Term Memory (LSTM) and related architectures against conventional classifiers^[8]. Priyamvada *et al.* reported 93.92% accuracy from a stacked convolutional and LSTM model^[9], and Benjachairat *et al.* built a Thai-language corpus labelled by counselling psychologists, showing that the approach transfers beyond English^[10]. Related work has applied transformer models to suicide notes^[11] and deep learning to pandemic-era sentiment on the same platform^[12].

However, almost all of this work is set in high-income or Asian contexts. The African, and specifically the Kenyan, setting is largely absent, even though language use, cultural reference and the shape of the pandemic response all differ. Additionally, most published models stop at the evaluation table and are never made usable by the practitioners who would benefit from them. Existing deployed tools are also narrow: a chatbot-based detector requires the user to converse with a machine^[13], and a smartphone application designed for transgender users targets transgender population^[14].

This study therefore had four objectives: to determine the temporal pattern of suicidal ideation expressed by Kenyan users of X during and after the pandemic; to develop an LSTM classifier for such posts; to evaluate that classifier against three conventional machine learning baselines; and to deploy it as a working web application. Two hypotheses were tested. The first was that there is no significant monotonic trend in the weekly volume of suicidal ideation posts over the study period. The second was that there is no significant difference between the classification performance of any two of the models compared.

2. Materials and Methods

2.1. Study Design

A quantitative design was adopted because the objectives required systematic measurement of a large body of text, statistical testing of temporal trends, and the training of models whose performance could be judged against numerical criteria. The work followed a five-stage pipeline: data collection, preprocessing, model development, model evaluation and deployment. Structuring the study in this way allowed each stage to be validated on its own and supported reproducibility.

2.2. Data Collection and Study Population

Posts were retrieved through the X Developer Application Programming Interface (API) using the full-archive search endpoint under Academic Research credentials. Twenty keywords semantically linked to anxiety, depression and suicidal ideation were used: *depression, hate myself, want to die, stressed out, struggle, suffering, depressed, kill myself, suicide cases, end my life, hopeless, worthless, can't go on, better off dead, no reason to live, self harm, overwhelmed, empty, alone and giving up*. These capture both explicit references to self-harm and broader distress.

The study did not draw a sample. A census, or complete enumeration, was used, in which every post belonging to the target population was studied. Where those limits do not apply, and the population is finite, clearly defined and fully accessible, the complete set of units can be observed directly and no inference from a part to the whole is required^[15, 16]. The target population was defined by three conditions that a post had to meet at the same time: it had to contain at least one target keyword; it had to originate from a user located in

Kenya, as determined by the geographic place metadata attached to the post and the country recorded in the profile location field; and it had to have been published between 1 January 2020 and 31 July 2021. Applying these conditions returned 6,651 posts, which were retained for further pre-processing. Completeness is therefore achieved with respect to what the API could return, since deleted, protected and platform-removed posts were not retrievable.

2.3. Ethical Considerations

The study was approved by the Technical University of Mombasa Scientific Ethics Review Committee under approval number TUM SERC MSC/028/2025A, and licensed by the National Commission for Science, Technology and Innovation under licence number NACOSTI/P/26/4186508. No human participants were contacted, and only posts from open, unrestricted accounts were collected. Individual informed consent was not feasible for secondary data of this scale; the requirement was met through safeguards. Handles were removed during collection, posts were paraphrased in all reporting so that they cannot be traced by verbatim search, and data were held on encrypted storage accessible only to the research team. The study followed the ethical guidelines of the Association of Internet Researchers and the platform's terms of service.

2.4. Data Preprocessing and Labelling

Social media text is informal and noisy, so the raw posts were normalised before modelling. URLs, hashtags, user mentions, emojis, reserved words, isolated single letters, punctuation, non-alphabetic characters and redundant whitespace were removed. All text was lowercased and tokenised. Three linguistic signals known from earlier work to accompany distress were retained: explicit self-harm terms, negative-emotion words, and first-person singular pronouns, the last of which has been linked to self-focused rumination and depressive states^[17]. Each remaining post was labelled as suicidal or non-suicidal based on its polarity score, computed with a lexicon-based sentiment library, TextBlob library. Of the 6,651 posts collected, 386 (5.8%) returned a neutral polarity score and could not be assigned a binary label; these were excluded, leaving 6,265 labelled posts. Latent Dirichlet Allocation was applied to describe the dominant themes in the corpus.

2.5. Trend Analysis

Posts were aggregated into weekly counts. Weekly aggregation gives evenly spaced observations and smooths the volatility of daily counts while preserving genuine temporal signal. The Mann-Kendall test was used to assess whether a monotonic trend was present. It is non-parametric and assumes neither normality nor linearity, which suits text data in this study. The test statistics compare the direction of change between every pair of time-ordered observations, as given in Equation (1).

$$S = \sum_{i=1}^{n-1} \text{sgn}(x_j - x_i) \quad (1)$$

Here n is the number of weekly observations, and x_i and x_j are the counts at times i and j respectively, with j greater than i . The sign function returns +1 when the later count is larger, 0 when the two are equal, and -1 when the later count is smaller. A positive S indicates an increasing trend and a negative S a decreasing one. For samples of this size the statistic is standardised to a Z value, as given in Equation (2).

$$Z = \frac{S - \text{sgn}(S)}{\sqrt{\text{Var}(S)}} \quad (2)$$

The strength of the association was quantified using Kendall's tau, a rank correlation coefficient bounded by -1 and $+1$, where values near the extremes indicate a strong monotonic relationship. The test gives statistical validation that the trend is not due to random noise.

2.6. Long Short-Term Memory Network

An LSTM network formed the core model. An LSTM is a recurrent neural network built to process ordered sequences,

such as the words of a post, while retaining information across long stretches of that sequence. Ordinary recurrent networks tend to lose earlier inputs as a sequence grows, a problem known as the vanishing gradient, which weakens their ability to link words that are far apart. The LSTM avoids this through a memory cell and a set of gates that regulate what information is kept and what is discarded. This matters for suicidal ideation, where meaning often depends on the combination and order of words rather than on isolated terms. The structure of a single cell is shown in Figure 2.1

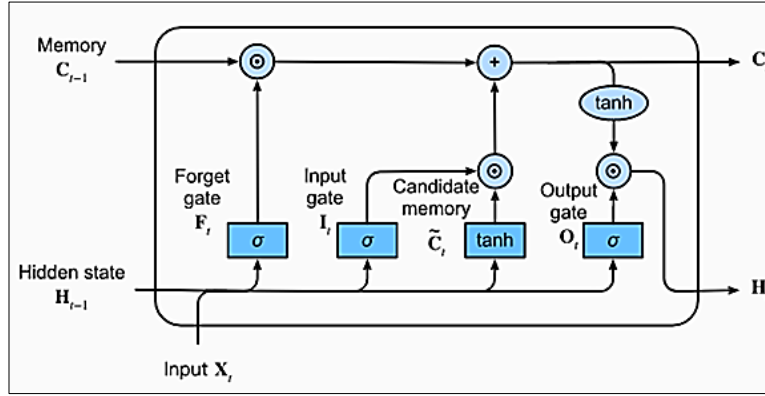


Fig 2.1: Structure of a Long Short-Term Memory Cell

At each time step t the cell receives the current input vector X_t , the previous hidden state H_{t-1} and the previous memory state C_{t-1} and returns an updated memory state C_t and hidden state H_t . Three gates, each a sigmoid transformation whose output lies between 0 and 1, control the flow of information; a value near 0 blocks information and a value near 1 lets it pass. The forget gate decides what proportion of the old memory to erase, as given in Equation (3).

$$F_t = \sigma(W_F X_t + U_F H_{t-1} + b_F) \quad (3)$$

The input gate decides what proportion of the new information to admit, as given in Equation (4).

$$I_t = \sigma(W_I X_t + U_I H_{t-1} + b_I) \quad (4)$$

The candidate memory is the new information proposed for storage, bounded by the hyperbolic tangent function, as given in Equation (5).

$$\tilde{C}_t = \tanh(W_C X_t + U_C H_{t-1} + b_C) \quad (5)$$

The memory state is updated by combining the gated old memory with the gated candidate memory, as given in Equation (6). This step is the heart of the LSTM, since it fixes how much of the past is retained and how much new information is added.

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (6)$$

The output gate decides what part of the memory is exposed, as given in Equation (7), and the hidden state is that exposed portion, as given in Equation (8).

$$O_t = \sigma(W_O X_t + U_O H_{t-1} + b_O) \quad (7)$$

$$H_t = O_t \odot \tanh(C_t) \quad (8)$$

In Equations (3) to (8), $\odot$ denotes element-wise multiplication, σ is the sigmoid function and $\tanh$ bounds values between -1 and 1 . The matrices W are applied to the input, the matrices U to the previous hidden state, and the vectors b are biases; all are estimated during training. Each LSTM layer holds $4(dh + h^2 + h)$ trainable parameters, where d is the embedding dimension and h the number of hidden units.

The outcome $Y \in \{0, 1\}$ was predicted from the final hidden state H_T , which summarises the whole post, through a single-neuron dense layer with sigmoid activation, as given in Equation (9).

$$\hat{y} = \sigma(w_y \cdot H_T + b_y) \quad (9)$$

Parameters were estimated by minimising the binary cross-entropy loss, given in Equation (10), which measures the distance between the predicted probabilities and the true labels and heavily penalises confident but wrong predictions.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (10)$$

In Equation (10), N is the number of training posts, y_i is the true label of post i , taking the value 1 when the post is suicidal and 0 otherwise, and $\hat{y}_i$ is the predicted probability that the post is suicidal. The logarithms make the penalty grow sharply when a high probability is assigned to the wrong class, and the leading negative sign makes the loss a positive quantity to be minimized.

2.6.1. Model Configuration

The 6,265 labelled posts were split 80 to 20 into training and test sets using stratified allocation, giving 5,012 training and 1,253 test posts and preserving the class proportions in both subsets. The full configuration is given in Table 2.1

Table 2.1: Architecture and Training Configuration of the Long Short-Term Memory Model

Component	Setting	Purpose
Vocabulary	10,000 most frequent tokens, plus an out-of-vocabulary token	Limits sparsity while retaining the working vocabulary of the corpus
Sequence length	100 tokens, post-padded	Covers the length of almost every post without excessive padding
Embedding layer	10,000 × 128	Maps each token to a dense 128-dimensional vector learned during training
LSTM layer 1	64 units, full sequence returned	Learns word-level sequential structure and passes it forward
LSTM layer 2	32 units, final state returned	Compresses the sequence into one summary vector for classification
Dropout	0.5 and 0.4	Randomly disables units during training to limit overfitting
Output layer	1 neuron, sigmoid activation	Returns the probability that a post expresses suicidal ideation
Optimiser	Adam, learning rate 0.0003	Adapts the step size per parameter for stable convergence
Batch size / epochs	16, up to 15 with early stopping	Training halted at epoch 5 and the best weights were restored

2.7. Baseline Classifiers

Three conventional classifiers were trained on the same data as baselines, and all four models were evaluated on the same held-out test set of 1,253 posts. 5-fold cross-validation was applied to the three baselines during development.

2.7.1. Logistic Regression

Logistic regression models the probability that a post belongs to the suicidal class by passing a linear combination of the features through the logistic function, as given in Equation (11).

$$P(y = 1 | x) = \frac{1}{1 + \exp(-(\beta_0 + \sum_{i=1}^p \beta_i x_i))} \quad (11)$$

Here x_i are the p input features, β_i the estimated coefficients and β_0 the intercept. The method is interpretable, because each coefficient shows the direction and size of a feature's effect, but it assumes an approximately linear relationship with the log-odds of the outcome.

2.7.2. Naive Bayes

The naive Bayes classifier applies Bayes' theorem and assigns each post to the class with the highest posterior probability, assuming that features are conditionally independent given the class, as given in Equation (12).

$$\hat{y} = \arg \max_{c \in \{0,1\}} P(c) \prod_{i=1}^p P(x_i | c) \quad (12)$$

Here $P(c)$ is the prior probability of class c and $P(x_i | c)$ the likelihood of feature i given that class. The independence assumption rarely holds in natural language, since words in a post are related.

2.7.3. Random Forest

The random forest builds many decision trees on random subsets of the data and features and combines them by majority vote. Splits within each tree are chosen to reduce Gini impurity, which measures how mixed the classes are at a node, as given in Equation (13).

$$G(t) = 1 - \sum_{k=1}^K p_k^2 \quad (13)$$

Here p_k is the proportion of observations at node t belonging to class k , and K is the number of classes. A pure node returns zero. The classifier was implemented with term frequency-inverse document frequency features (TF-IDF), 200 trees and a maximum depth of 30.

2.8. Evaluation Metrics

Models were assessed using accuracy, F1-score and area under the receiver operating characteristic curve¹⁸. Suicidal

was treated as the positive class throughout. Accuracy is the proportion of all posts classified correctly. Because accuracy alone is misleading when classes are unbalanced, the F1-score, the harmonic mean of precision and sensitivity given in Equation (14), was used as balanced measure.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (14)$$

2.9. Statistical Comparison of Models

Differences in performance were tested using the pairwise McNemar test, a non-parametric test for two classifiers evaluated on the same data. Rather than comparing overall accuracy, it examines only the cases on which the two models disagree¹⁹. The statistic is given in Equation (15).

$$\chi^2 = \frac{(b - c)^2}{b + c} \quad (15)$$

Here b is the number of posts classified correctly by the first model but not the second, and c the number classified correctly by the second but not the first. A significant result indicates that the disagreement is unlikely to be due to chance.

3. Results

3.1. Temporal Pattern of Suicidal Ideation

Weekly volume rose sharply between late February and April 2020, reaching a maximum of 165 posts in the week beginning 6 April 2020, immediately after Kenya's first confirmed case on 13 March 2020, the curfew of 27 March 2020 and the movement restrictions of 5 April 2020. Volume fell after April but remained high through 2020, with a secondary spike of 141 posts in early September 2020. A further peak of 127 posts occurred in late March 2021, coinciding with the renewed restrictions announced that month. From May to July 2021 the trend was at lower levels. The pattern is shown in Figure 2.2.

Because a visual trend can be due to noise or random variation, the Mann-Kendall test was computed to statistically validate the trend. The results are given in Table 3.1.

Table 3.1: Mann-Kendall Trend Test on Weekly Counts of Suicidal Ideation Posts

Metric	Value
Trend	Decreasing
h (Reject H ₀)	True
p-value	1.66 × 10 ⁻⁸
Z-statistic	-5.64
Kendall's Tau	-0.44
Slope	-0.75 (tweets per week)
Intercept	104.5

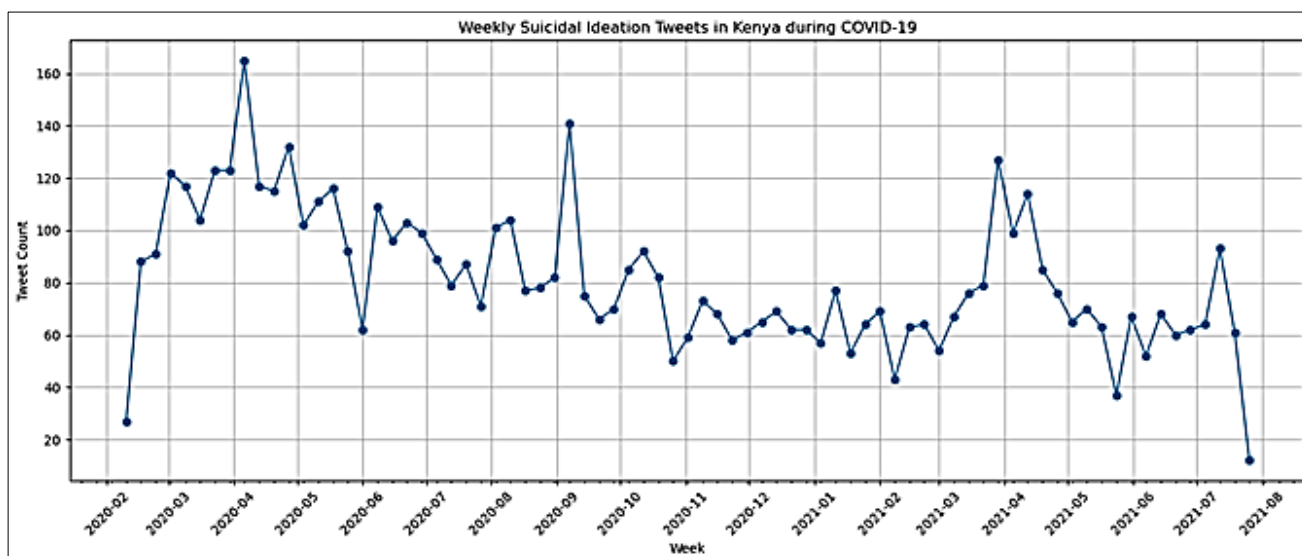


Fig 2.2: Weekly Count of Suicidal Ideation Posts

The first null hypothesis, that no monotonic trend is present, was rejected ($p < 0.001$). Kendall's tau of -0.44 indicates a moderate inverse association between time and volume, and the slope estimate implies a fall of about 0.75 posts per week across the period. The initial surge during the most disruptive phase of the pandemic was therefore followed by a sustained decline, a pattern consistent with gradual psychological adaptation, the easing of restrictions and the resumption of economic activity.

3.2. Training and Performance of the Developed Model

Training accuracy rose from 0.681 in the first epoch to 0.989 by the fifth, while training loss fell from 0.622 to 0.042.

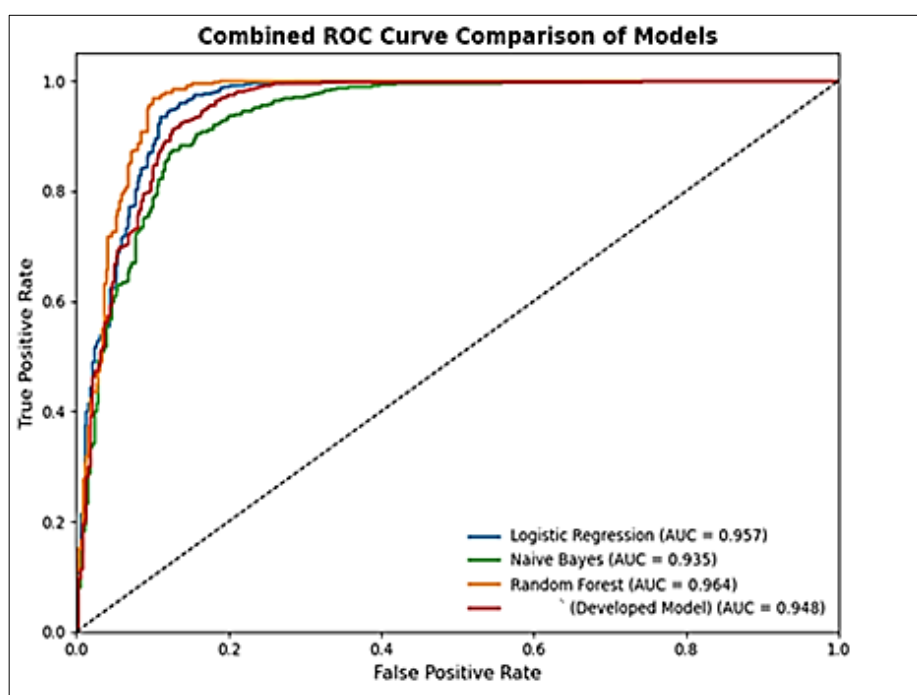
Validation accuracy stabilised near 0.921, and validation loss began to rise after the second epoch, indicating that further training would overfit. Early stopping therefore achieved at epoch 5 during training and restored the best weights. On the held-out test set the model returned an accuracy of 0.921, an F1-score of 0.944 and an area under the curve of 0.945.

3.3. Comparative Performance

All four models were evaluated on the same 1,253 held-out data. The results are shown in Table 3.2, and the receiver operating characteristic curves in Figure 3.1.

Table 3.2: Comparative Performance of the Four Classifiers on the Held-Out Test Set

Model	Accuracy	F1-score	AUC
Long Short-Term Memory	0.921	0.944	0.945
Logistic regression	0.886	0.924	0.957
Naive Bayes	0.811	0.879	0.935
Random Forest	0.831	0.891	0.964



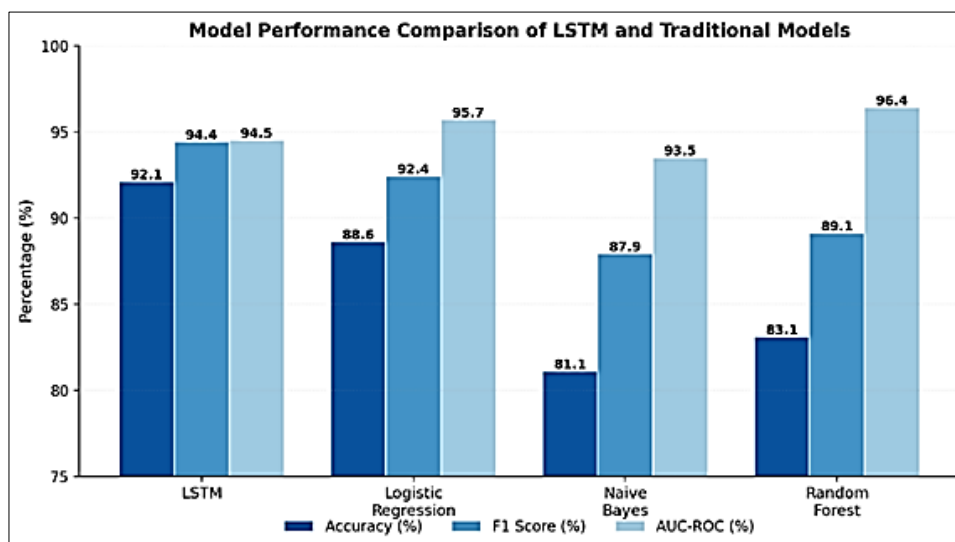


Fig 3.1: Receiver Operating Characteristic Curves (Left) and Bar Graph for the Four Classifiers (Right)

The LSTM gave the highest accuracy at 0.921, followed by logistic regression (0.886), the random forest (0.831) and naive Bayes (0.811). Also, LSTM achieved the highest F1-score. In area under the curve, the random forest scored highest at 0.964 and logistic regression second at 0.957. The area under the curve measures how well a model ranks posts across all possible thresholds. All four curves lie well above

the diagonal, as shown in Figure 3.1 (left) so every model achieved meaningful discrimination.

3.4. Error Analysis

Summary metrics do not reveal which kind of error a model makes, so the confusion matrix was computed as shown in Figure 3.2

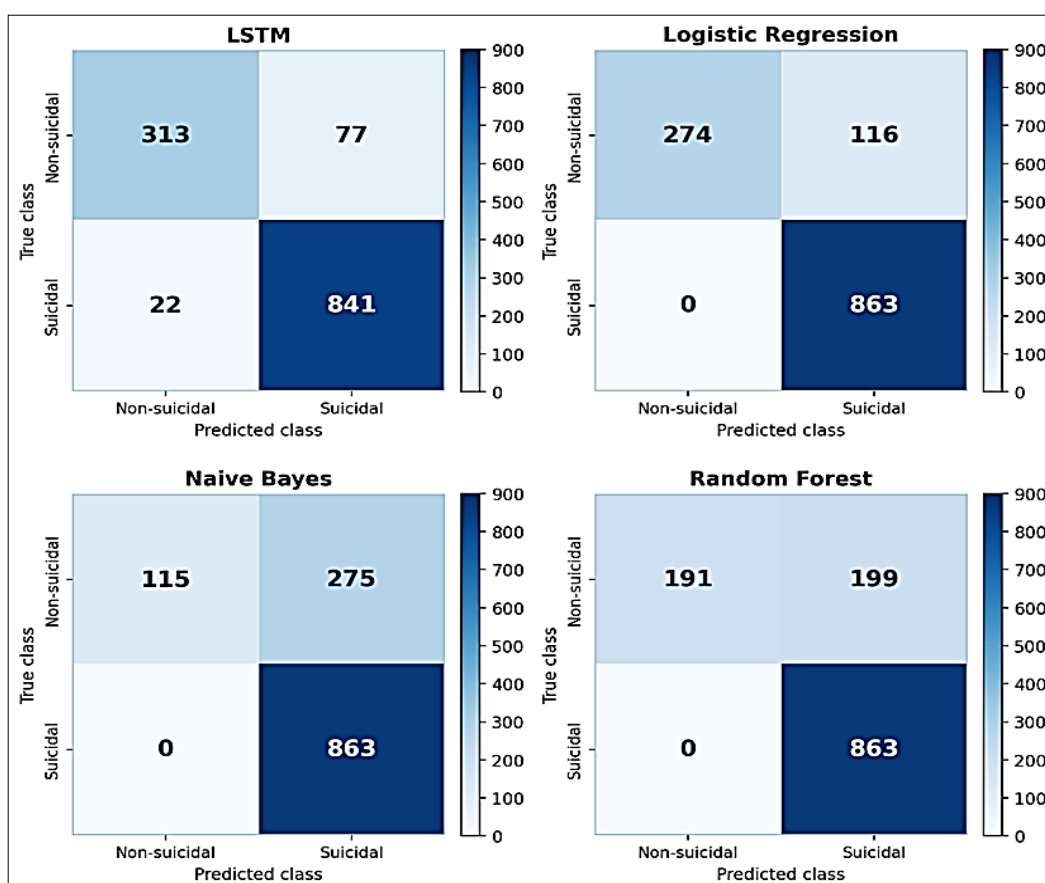


Fig 3.2: Confusion Matrices for the Four Classifiers on the Held-Out Test Set

The three baselines produced no false negatives, meaning that no genuine suicidal post was overlooked. Naive Bayes wrongly flagged 275 of the 390 non-suicidal posts, the

random forest 199, and logistic regression 116. LSTM had 22 false negatives and only 77 false alarms, correctly recognizing 313 of the 390 non-suicidal posts as illustrated in Figure 3.2.

3.5. Pairwise McNemar Statistical Test

Pairwise McNemar tests were applied to the paired predictions of each model. The results are given in Table 3.3.

Table 3.3: Pairwise McNemar Statistical Test Between Classifiers

Model comparison	b	c	p-value
LSTM vs Logistic Regression	105	24	< 0.001
LSTM vs Naive Bayes	208	44	< 0.001
LSTM vs Random Forest	132	56	< 0.001
Logistic Regression vs Naive Bayes	161	15	< 0.001
Logistic Regression vs Random Forest	85	45	< 0.001
Naive Bayes vs Random Forest	27	103	< 0.001

The second null hypothesis was rejected for every pair of models ($p < 0.001$). In all three comparisons involving the LSTM, b greatly exceeded c , meaning that the deep learning model corrected far more of the baselines' errors. Among the baselines, logistic regression significantly outperformed both naive Bayes and the random forest. These results confirm that the observed differences in classification performance were not due to chance.

3.6. Deployment as a Screening Tool: The trained model was deployed as a web application. A counsellor enters an

individual's handles or profile links in one form, and the system retrieves posts from the preceding six months across Reddit, Bluesky, Mastodon, YouTube, Facebook and X in a single run. Six months was chosen because suicidal ideation typically recurs, so one post is a weak signal, while older material may reflect distress since resolved. Video audio is transcribed automatically using Whisper Package and passed through the same trained weights. Each item receives a risk score, and scores are aggregated into a unified profile, as shown in Figure 3.3

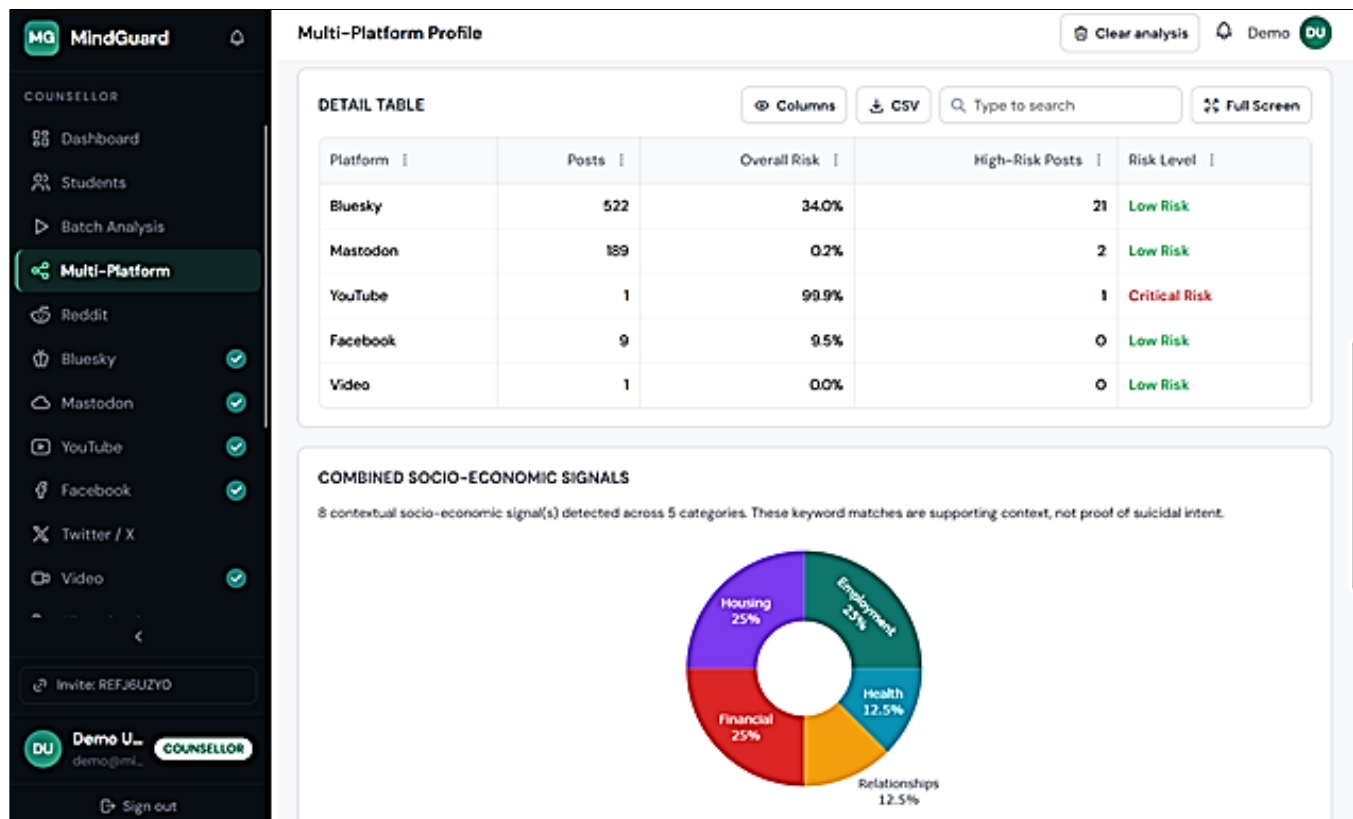


Fig 3.3: Unified Multi-Platform Output of the Deployed App for Demonstration Account

In the demonstration run shown in Figure 3.3, five platforms were analyzed. The aggregate risk was 34.0% for Bluesky across 522 posts, 0.2% for Mastodon across 189 posts and 9.5% for Facebook across 9 posts, all rated low risk, while a single YouTube item scored 99.9% and was rated critical. This illustrates the practical value of item-level scoring: the aggregate profile was low, yet the one critical item remained visible. The application also reports the proportion of socio-economic signals detected covering employment, housing, financial, health and relationship categories, as supporting context and not as evidence of intent.

4. Discussion

4.1. Temporal Patterns

The sharp rise in suicidal ideation posts between late February and April 2020 coincided with Kenya's first confirmed case and the containment measures that followed. This agrees with the wider pandemic literature. Killgore *et al.* found that suicidal ideation among people under lockdown rose from 17.6% to 30.7%²⁰, and Schluter *et al.* reported a cross-country increase from 24.2% to 27.5%²¹. The present study supports that association and extends it to a Kenyan social media setting that has been largely absent from previous work.

Where this study differs from previous studies is in its observation window. The previous studies were largely cross-sectional and captured only the rising phase. The Mann-Kendall test applied here covered 19 months and revealed a significant long-term decline, suggesting that the initial surge was followed by gradual adaptation. The secondary peak in March 2021 further supports the link between isolation, quarantine, economic strain and online expressions of hopelessness. Because the data are real-time posts and not questionnaire responses, the study also avoids the recall limitations of survey-based designs.

4.2. Model Performance

The finding that the LSTM achieved the highest accuracy and F1-score is consistent with Sawhney *et al.*, who showed that recurrent architectures are appropriate for modelling sequential data and context of language⁶. This also explains why LSTM was appropriate for this type of dataset. A model may rank well and still be unusable if its threshold flags most of the corpus; for a deployed tool, performance at the operating point is very significant.

4.3. Deployment and Practical Implications

Deployment as a full-stack web application confirms the feasibility of technology-assisted screening reported by Bruen *et al.*²², while addressing a limitation of their study that the screening tool was confined to an inpatient setting and a single device type, the present application is device-agnostic and reachable through any browser. The demonstration run also shows why item-level scoring is important. A single critical item among low-risk posts would disappear in any aggregate measure, yet it is exactly the item a counsellor needs to see.

4.4. Limitations

Several limitations should be noted. Keyword retrieval misses ideation expressed indirectly, sarcastically or through metaphor. Kenyan slang and English-Kiswahili code-switching are imperfectly captured. Additionally, 386 posts with neutral polarity were excluded, which may have removed genuinely ambiguous cases. Also, the study did not consider posts that were written by bot, which should be considered for future research.

5. Conclusion

This study examined suicidal ideation expressed by Kenyan users during and after the COVID-19 pandemic and developed a deep learning model to detect it. Posting volume rose sharply during the first lockdown of March and April 2020 and again during the renewed restrictions of March 2021, and the Mann-Kendall test confirmed a statistically significant long-term decline across the whole period. Among the four classifiers evaluated on the same held-out data, the Long Short-Term Memory network achieved the highest accuracy and F1-score and was the only model that separated the two classes. Deployed as a multi-platform web application, the model worked in practice as a screening tool and does not replace clinical diagnosis. The study makes two recommendations follow. National emergency response plans should include explicit mental health provision, since online distress is related closely with restriction announcements. Also, automated screening developed in this study should be adopted only as a decision-support tool, with a qualified professional retaining responsibility for every clinical judgement. Future work should test transformer architectures,

add bot detection, and extend the model to Kiswahili and code-switched text.

Acknowledgement

The authors thank Dr. Kevin Tole from Technical University of Mombasa for his insights on model development, and the Department of Mathematics and Physics, Technical University of Mombasa, for institutional support. The authors declare no conflict of interest and received no external funding for this work.

References

1. Czeisler MÉ, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, *et al.* Mental health, substance use, and suicidal ideation during the COVID-19 pandemic. *MMWR Morb Mortal Wkly Rep.* 2020;69(32):1049-1057.
2. Shongwe MC, Huang SL. Suicidal ideation and predictors of psychological distress during the COVID-19 pandemic in Eswatini: a population-based household telephone survey. *Int J Environ Res Public Health.* 2021;18(13):6700.
3. Zhou SJ, Wang LL, Qi M, Yang XJ, Gao L, Zhang SY, *et al.* Depression, anxiety, and suicidal ideation in Chinese university students during the COVID-19 pandemic. *Front Psychol.* 2021;12:669833.
4. Ndeti DM, Mutiso VN, Weisz JR, Okoth CA, Musyimi C, Muia EN, *et al.* Socio-demographic, economic and mental health problems were risk factors for suicidal ideation among Kenyan students aged 15 plus. *J Affect Disord.* 2022;302:74-82.
5. Kang JH, Lee SW, Ji JG, Yu JK, Jang YD, Kim SJ, *et al.* Changes in the pattern of suicide attempters visiting the emergency room after COVID-19 pandemic: an observational cross sectional study. *BMC Psychiatry.* 2021;21(1):571.
6. Sawhney R, Manchanda P, Mathur P, Shah R, Singh R. Exploring and learning suicidal ideation connotations on social media with deep learning. *Proc Workshop Comput Approaches Subjectivity Sentiment Soc Media Anal.* 2018;9:167-175.
7. Metzler H, Baginski H, Niederkroenthaler T, Garcia D. Detecting potentially harmful and protective suicide-related content on Twitter: machine learning approach. *J Med Internet Res.* 2022;24(8):e34705.
8. Haque R, Islam N, Islam M, Ahsan MM. A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. *Technologies.* 2022;10(3):57.
9. Priyamvada B, Singhal S, Nayyar A, Jain R, Goel P, Rani M, *et al.* Stacked CNN-LSTM approach for prediction of suicidal ideation on social media. *Multimed Tools Appl.* 2023;82(18):27883-27904.
10. Benjachairat P, Senivongse T, Taephant N, Puvapaisankit J, Maturosamnan C, Kultananawat T. Classification of suicidal ideation severity from Twitter messages using machine learning. *Int J Inf Manag Data Insights.* 2024;4(2):100280.
11. Zhang T, Schoene AM, Ananiadou S. Automatic identification of suicide notes with a transformer-based deep learning model. *Internet Interv.* 2021;25(1):100422.
12. Chintalapudi N, Battineni G, Amenta F. Sentimental analysis of COVID-19 tweets using deep learning models. *Infect Dis Rep.* 2021;13(2):329-339.

13. Elsayed N, ElSayed Z, Ozer M. CautionSuicide: a deep learning based approach for detecting suicidal ideation in real time chatbot conversation. arXiv [Preprint]. 2024;2401.01023:1-6.
14. Dubov A, Fraenkel L, Goldstein Z, Arroyo H, McKellar D, Shoptaw S. Development of a smartphone app to predict and improve the rates of suicidal ideation among transgender persons (TransLife): qualitative study. J Med Internet Res. 2021;23(3):e24023.
15. Ahmed SK. How to choose a sampling technique and determine sample size for research: a simplified guide for researchers. Oral Oncol Rep. 2024;12:100662.
16. Lohr SL. Sampling: design and analysis. 3rd ed. Boca Raton: Chapman and Hall/CRC; 2021. p. 1-52.
17. Zimmermann J, Brockmeyer T, Hunn M, Schauenburg H, Wolf M. First-person pronoun use in spoken language as a predictor of future depressive symptoms: preliminary evidence from a clinical sample of depressed patients. Clin Psychol Psychother. 2016;24(2):384-391.
18. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022;75(1):25-36.
19. Sundjaja JH, Shrestha R, Krishan K. McNemar and Mann-Whitney U tests. StatPearls [Internet]. Treasure Island: StatPearls Publishing; 2021. p. 1-8.
20. Killgore WDS, Cloonan SA, Taylor EC, Allbright MC, Dailey NS. Trends in suicidal ideation over the first three months of COVID-19 lockdowns. Psychiatry Res. 2020;293:113390.
21. Schluter PJ, Génèreux M, Hung KK, Landaverde E, Law RP, Mok CPY, *et al.* Patterns of suicide ideation across eight countries in four continents during the COVID-19 pandemic era: repeated cross-sectional study. JMIR Public Health Surveill. 2022;8(1):e32140.
22. Bruen AJ, Wall A, Haines-Delmont A, Perkins E. Exploring suicidal ideation using an innovative mobile app-strength within me: the usability and acceptability of setting up a trial involving mobile technology and mental health service users. JMIR Ment Health. 2020;7(9):e18407.