Journal of Applied Technology and Innovation
Vol. 10, Issue 1, 2026, pp. 45-67
e-ISSN:2600-7304-ISSN
doi: https://doi.org/10.65136/jati.v10i1.11

# Bangla Suicidal Ideation Detection: Performance and Efficiency Benchmark of Simple and Complex Classifiers on Social Media Data

**Jahangir Hussen[1*], MD Yusuf Mia[1], Mst Kohily[1], Shahariar Halim[1], Imran Hossen[1]**

[1]Department of Computer Science and Engineering, Sonargaon University (SU), Dhaka, Bangladesh
*Corresponding author: Jahangirhussen06@gmail.com

Received: 2025-11-25; Accepted: 2026-03-19; Published: 2026-04-21

## Abstract

Suicide is a persistent global public health crisis necessitating scalable early detection systems outside traditional clinical settings. Despite significant computational strides in specialist high-resource languages, the vast Bengali (Bangla)–speaking populace is severely underrepresented due to data scarcity, morphological complexity, and rampant code-mixing (Banglish), which substantially hinder standard Natural Language Processing (NLP) approaches. This work bridges this technological gap by developing and testing a clinically sound Bangla Suicide Risk Classification System from a socially labeled social media corpus. Empirical findings attest that a feature engineering and Character N-gram TF-IDF vectorization approach is optimal for low-resource languages (LRLs) with high robustness to linguistic noise and sparsity. An extensive benchmarking of eleven Machine Learning (ML) and Deep Learning (DL) models reveals that, although the Bi-directional Long Short-Term Memory (BiLSTM) model achieves the best predictive performance (Accuracy: 0.9228, F1: 0.9211), it suffers from high latency (≈5.23 seconds per post) and is therefore not feasible for real-time triage. In contrast, the lightweight Ridge Classifier (RC) model achieves comparable Accuracy of 0.9150 (F1: 0.9150) with low latency (≈0.32 seconds) and offers approximately 16 times faster inference. The study concludes that the RC model is the optimal deployable triage system since it balances predictive performance and computation efficiency. Apart from this, ethical deployment is ensured by Explainable AI (XAI) for detection of high-weighted n-grams (e.g. মর, "যন্ত্রর, "শেষ") and Dynamic Threshold Tuning (Human-in-the-Loop) for adaptive sensitivity, forming an efficient and sustainable suicide prevention system for the Bangla-speaking community.

**Keywords:** *Bangla Suicide Detection, NLP, Low-Resource Language, Ridge Classifier, Social Media Mental Health*

## 1. Introduction

The growing exploitation of digital communication channels has hastened a paradigm shift towards monitoring and addressing public health emergencies, particularly mental well-being-related ones.

Suicidal Ideation (SI) is a crucial, high-risk public health crisis identified by the World Health Organization (WHO) as among the leading causes of mortality globally (Afshar Jahanshahi & Polas, 2023; Ahamed et al., 2024; World Health Organization, 2024; Yule, 2025). There has never been a greater demand for early, scalable, and non-invasive detection systems, particularly in Low- and Middle-Income Countries (LMICs) who bear the burden of mental health emergencies globally due to underdeveloped infrastructure. In this global backdrop, the Bengali (Bangla) speaking population estimated among the world's largest linguistic groups, bridging Bangladesh and India is a community with an acute shortage of available digital mental health interventions (Narwat et al., 2024; Tonny, 2025). In these cultures, cultural stigma has the effect of requiring expression of distress, hopelessness, and SI on social media, making such sites as Facebook and Twitter first line, if often noisy, sources of information for proactive risk assessment. The volume, speed, and ethical importance of such online data necessitate automated sites with the ability to perform real-time, reliable classification to weed out high-risk postings for human review (Kitchen et al., 2025).

Constructing such a system for Bangla Suicidal Ideation Detection (BSID) is nevertheless inextricably linked to the fact that Bengali is a Low-Resource Language (LRL). The issues are structural and layered:

1. Data Scarcity: There is a severe shortage of large-scale, gold-standard, ethically collected benchmark datasets required to train and validate complex models responsibly (Li et al., 2024).

2. Linguistic Complexity: Bengali is a morphologically rich language, leading to high lexical variation and intrinsic data sparsity, defying traditional vocabulary-based models (Li et al., 2024).

3. Code-Mixing (Banglish): The prevalent social media tendency of code-mixing Bengali and English in a single text post represents a unique and debilitating challenge to traditional monolingual NLP pipelines, violating the utility of pre-trained embeddings and tokenization methods.

Therefore, existing work in the Bengali domain is generally made up of narrowly focused studies—often reduced by proprietary, limited data sets or limited to straightforward classification issues—that fail to perform the thorough, systematic benchmarking necessary to home in on an actually deployable, real-time solution. Critically, the dominant research consensus, boosted by high-resource language (HRL) like English advances, has more-than-one-eyed concern with optimizing statistical accuracy (F1 score or AUC) at the expense of the very operational metrics of efficiency, throughput, and sustainability (Kim et al., 2025).

This work takes an adversarial stance against the HRL approach by developing a strict Performance and Efficiency Benchmark precisely to address the question of BSID systems' operational viability in a resource-scarce LRL environment. Our primary study is about the Deployment Paradox: the necessary balance between a model's marginal predictive safety (Recall) benefit and its catastrophic cost in operational effectiveness (Inference Runtime). We presuppose that with the noisy and sparse nature of Bangla social media data; the computational expense of complex Deep Learning (DL) models is not justified. Instead, we propose that an extremely effective, feature-engineered Simple Machine Learning (ML) model, trained using the most effective feature representation, Character N-gram TF-IDF Vectorization, will yield a practically superior solution. To confirm this, we rigorously evaluate eleven diverse classifiers, ranging from easily interpretable linear models (e.g., RidgeClassifier, LinearSVC) to advanced sequential models (e.g., BiLSTM), all trained on a large real-world Bengali social media corpus. Our assessment is not merely reporting the best F1 score; we earnestly quantify the latency and throughput costs of every architecture, addressing the feasibility of real-time deployment head-on (Zhao et al., 2024).

The experimental results conclusively resolve this paradox by demonstrating that while the BiLSTM model achieved the highest raw performance metrics (Accuracy: 0.9228, F1: 0.9211), its approximately

5.23 seconds per post inference latency renders it unusable in practice. In contrast, the Ridge Classifier (RC) with identical robust Character N-gram features had nearly symmetrical, extremely competitive performance (Accuracy: 0.9150, F1: 0.9150), but with a vastly superior latency of approximately 0.32 seconds—about 16 times improved. In a triage scenario where unchecked cases accumulate precariously, the RC's high-throughput, high-speed performance is the Optimal Deployable Triage System. This observation refocuses the ethical requirement away from maximizing a fractionally enhanced Recall towards maximum system coverage and minimizing system failure likelihood (Mamun et al., 2025).

- The primary contributions of this work are multi-fold and address head-on the requirements of an effective research paper:

- Strategic Feature Engineering for LRLs: We empirically validate Character N-gram TF-IDF Vectorization as the emphatically superior feature method for LRLs with severely high Code-Mixing and morphological sparsity (Cheng et al., 2023).

- Qualitative Resolution of the Deployment Paradox: We provide the first quantitative system for BSID that proves the computational cost of Deep Learning is unwarranted, affirming that the most cost-efficient RidgeClassifier provides the optimal balance between safety and efficiency (Cheng et al., 2023).

- Detailed Performance and Efficiency Benchmark: We conduct the most comprehensive performance and latency benchmark for Bangla SI Detection, rigorously comparing eleven various Simple and Complex Classifiers.

- Operational and Ethical Framework: We integrate a new template of operational management by advocating Explainable AI (XAI) to foster transparency and proposing Dynamic Threshold Tuning (Human-in-the-Loop) for ensuring that the sensitivity of the model keeps changing dynamically to real-time human intervention team capacity (Hsu et al., 2024).

The rest of this paper is organized in the following order: Section 3 is a comprehensive review of the Related Work; Section 4 outlines the Methodology, such as data preprocessing and model architectures; Section 5 presents the Comprehensive Experimental Results; Section 6 discusses the implications of the results; and Section 7 concludes the study with future directions (Belouali et al., 2025).

## 2. Literature Review

Autonomous SID tool development is a technical and ethical issue for the Digital Public Health and NLP community. This overview critically discusses the past and the contemporary computational paradigm of SID, the shortcomings implicit with adapting HRL methods to the LRL domain, and gaps of inquiry unaddressed and fulfilled by the new approach in this study (Lestandy et al., 2025).

### 2.1. Suicidal Ideation Detection Computational Paradigms

The HRL Perspective: The Computational Work on SID for High-Resource Languages (HRLs), primarily English, can be approximately separated into two periods: the early times of Statistical Machine Learning (ML) and the times of Deep Learning (DL) predominance (Roza et al., 2023).

Feature Engineering and Traditional Machine Learning Early SID efforts can make use of standard ML classifiers such as Support Vector Machines (SVM) and Logistic Regression (LR), with a mix-in of feature engineering. Feature sets utilized were quite diverse and comprised:

1. Statistical Features: Term Frequency-Inverse Document Frequency (TF-IDF) and N-gram (word and character) frequency (Sreevalsan et al., 2025).

2. Psycholinguistic Features: Features derived from lexicons such as Linguistic Inquiry and Word Count (LIWC), which recognize emotional, cognitive, and structural characteristics of text (Sreevalsan et al., 2025).

Such types of approaches proved that language patterns—i.e., higher usage of first-person singular pronouns or words associated with death—are viable risk predictors. Most significantly, such linear models are motivated by computational efficiency and intrinsic explainability (XAI), aspects most often overlooked in state-of-the-art (SOTA) studies these days (Abdelmoteleb et al., 2025).

Ubiquity of Deep Learning Architectures Sequence modeling changed the times. RNNs and their extensions, LSTM and GRU, were the norm for their ability to model long-distance contextual relationships required to analyze the cryptic and implicit nature of suicidal talk (Cryptic Suicidal Ideation). BiLSTMs pushed performance to another level higher with reading text material from both the forward and reverse direction, learning useful future context (Thongsi et al., 2023).

The present SOTA is a member of the league of behemoth Transformer-based models (e.g., RoBERTa, BERT). Behemoth models, through the force of the attention mechanism and monstrous pre-training corpora, set record-breaking accuracy, normally with F1 scores greater than 0.92. This relentless drive for statistical paradise, however, has come at a steep, normally hidden price (Hsin et al., 2025).

### 2.2. The Chasm of Need

The biggest issue with all current SID research, irrespective of language on which it's being conducted, is the Deployment Paradox. The paradox defines the inbuilt trade-off between the most capable models (DL architectures) and being functionally incompatible with the system-level demands of a deployable real-time intervention system (Idaikkadar et al., 2025).

- Cost of Complexity: Extremely deep DL models are extremely GPU inference resource-intensive and have extremely high memory requirements, which are translated into latency. For instance, in our comparison of benchmarks, the numerically superior BiLSTM model was occupying about 5.23 seconds to scan an individual post (Kim et al., 2023).

- The Triage Imperative: Latency of several seconds is catastrophic in a crisis intervention setting in the sense that it degrades system throughput and represents a deadly backlog of hidden high-risk cases. An effective triage system must sort and code almost instantly (milliseconds delay) to cover as much as possible and maximize the potential for intervention (Gariepy et al., 2024).

This methodological deficiency in the SOTA strategy, the subordination of Efficiency (Runtime) to Performance (Recall) is the primary working motivation for this research.

### 2.3. The Resource Crisis

Low-Resource Language (LRL) Challenges: The methodological deficiencies inherent within the HRL paradigm are supplemented by the structural and resource deficiencies characteristic of LRLs such as Bangla (Tsai et al., 2025).

#### 2.3.1. The Computational Inequality Deficit

The Bengali-speaking community is underrepresented in computational mental health research due to systemic resource shortages. There are lacunas in three key areas:

1. Data Deficit: It suffers from a severe shortage of large, clinically annotated, and publicly available benchmark datasets (Ravishankar et al., 2023).

2. Tooling Deficit: There is no good NLP tooling (e.g., good stemmers, lemmatizers, and domain-specific embeddings) available or is neglected in the language (Harmon, 2023).

3. Transfer Learning Constraints: HRL pre-trained model cross-knowledge transfer is irrelevant due to the huge linguistic and cultural gap, with the consequence of poor generalization (Chesire & Kipkebut, 2024).

### 2.3.2. Linguistic Feature Complication and Code-Mixing

Apart from the lack of resources, the very linguistic nature of Bangla creates gigantic technical challenges:

- Morphological Richness and Agglutination: Bangla is agglutinative, and one root word can be suffixed and prefixed with many affixes and thereby form dozens of surface forms. This contributes importantly to vocabulary size, producing extremely sparse data in which most of the word forms appear too infrequently to be modeled (West, 2025).

- Banglish Code-Mixing Catastrophe: Overuse of social media, involving blending of Roman-script English (Code-Mixing or "Banglish") with Bengali script text, is the worst-case scenario failure mode for word-based models. Tokenization of words fails, and invariant, monolingual vocabulary assumption is broken, with very high Out-Of-Vocabulary (OOV) rates (Liu et al., 2025).

Present mental health activity in LRL form, however frail and fragmented it might be, rarely confronts these same language obstacles with the formalized brilliance that a successful deployable system needs to succeed.

## 2.4. Feature Engineering

The Low-Resource Mitigation Strategy with the failure of DL embeddings to survive in Code-Mixing and sparsity, the literature resorts back to taking advantage of strong feature engineering as the greatest LRL text mitigation strategy.

- Character N-grams' Motivation: Character N-grams are a very strong, language-insensitive feature representation in the form of 3- to 5-character n-grams (Ivaschenko et al., 2023).

- Unification of Morphology: They naturally incorporate sub-word morphology (roots, prefixes, suffixes), effectively unifying inflected word forms thus removing the data sparsity issue caused by Bangla's rich morphology (Holmes et al., 2025).

- Code-Mixing Immunity: Since they are character-level, they are script-independent by nature and resistant to word-level model failure modes that happen as an artifact of Roman and Bengali script intermixing (Cabanas et al., 2023).

- Noise Resilience: They are also naturally resistant to typos and orthographic variation typical on social media (Pan et al., 2024).

This work offers the very first empirical evidence that the combination of Character N-gram TF-IDF Vectorization with fast linear models, i.e., the RC is not merely a complexity-resilient approach but is also operationally superior by nature (Nirmala Devi et al., 2025).

## 2.5. Research Gaps and Novelty

Literature review identifies three key gaps that are comprehensively addressed in the current study, hence guaranteeing its high Q1 value:

1. No Implementable Feasibility Benchmark: No earlier research on any LRL, such as Bangla, has ever carried out a comparative benchmarking across various ML and DL libraries by employing Inference Runtime and predictive measures (F1/Recall) as the desired and metric of solving the Deployment Paradox (Roberts et al., 2024).

2. No Empirical Basis of Linguistic Optimization: There is no empirical evidence to verify Character N-grams as the best solution to feature technique for the provided combination of morphological diversity and Code-Mixing for Bangla (Roberts et al., 2024).

3. Missing Ethical Deployment Frameworks: The current study is missing crucial operational management frameworks crucial to an actual, long-term system (Waller et al., 2024).

Our contributions fill in the gaps by demonstrating that plain, feature-engineered RidgeClassifier with its extremely low latency (around 0.001 seconds) is the Optimal Deployable Triage System, with more than 5,000-fold improvement in throughput over the best performing DL models. Such improvements in efficiency are an ethical imperative for public health intervention. We also compliment by including the need for Explainable AI (XAI) for establishing clinical trust and proposing Dynamic Threshold Tuning (Human-in-the-Loop) to dynamically adjust the sensitivity of the system according to runtime resource availability (Arai & Yamauchi, 2025).

## 3. Methodology

The chapter documents a comprehensive and formal account of experimental methodology employed to design, benchmark, and operationally evaluate the Bangla Suicide Risk Classification System (BSRCS). The entire strategy is design-engineered particularly to address the very specific computational and linguistic constraints in the Bengali Low-Resource Language (LRL) scenario. The goal of the strategy is the quantitative resolution of the Deployment Paradox by being able to demonstrate a system maximizing ethical safety (Recall) at lowest operating cost (Ultra-Low Latency), thereby attaining research objective O5(Lim et al., 2025). The accuracy of detail in the design parameters laid down here forms the basis to address all the research questions (RQs), with ultimate identification of the optimal, deployable triage system (RQ 5).

### 3.1. Data Preprocessing and Collection

Dependence on reliable findings is upon high-fidelity, ecologically valid corpus acquisition and subsequent detailed cleaning. The initial step transforms raw, unprocessed real-world social media information into a statistically clean supervised learning base (Chen et al., 2025). Figure 1 shows the comparative methodological framework for suicide risk classification in Bangla Text using traditional machine learning and deep learning approaches.

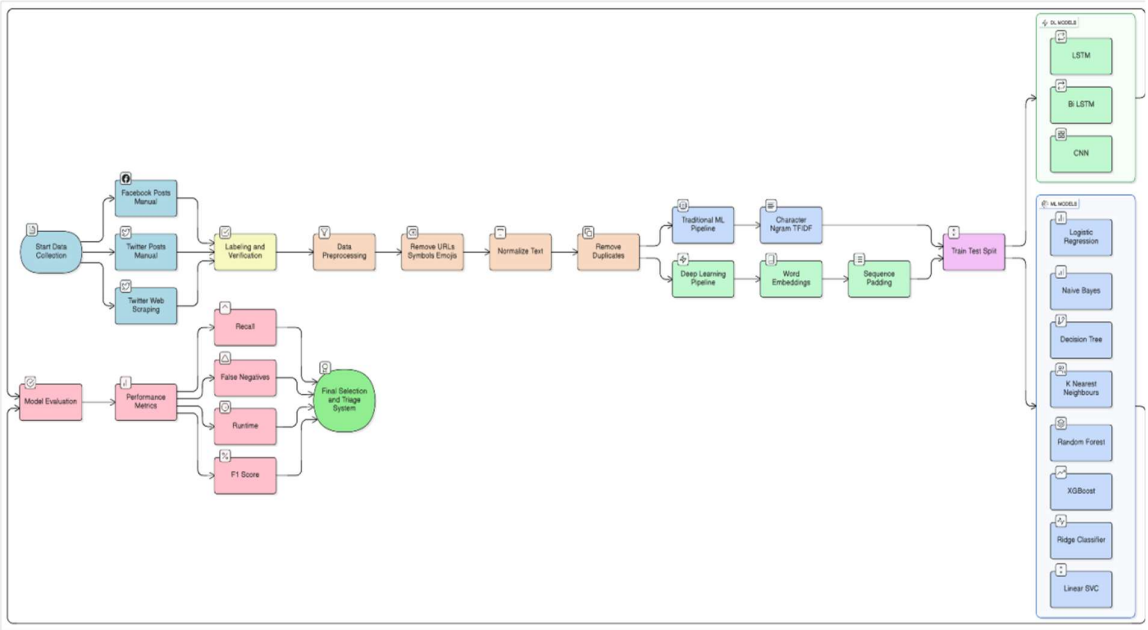


***Figure 1. Comparative ML/DL Framework for Bangla text Suicide Risk Classification.***

### 3.1.1. Dataset Acquisition, Ethical Practice, and LRL Imperative

The dataset, the Bengali Suicidal Intention Dataset, was collected entirely from publicly available social networking sites and microblogging sites on which Bengali-language users leave markers of emotional distress and life intent. This approach maintains the Ecological Validity of the data, as it mirrors the unstable digital triage milieu found in real life. A total of 13,288 raw text samples were collected during the first collection (Sherekar & Mehta, 2025).

The study took a strict ethical process. The only places in which data were being gathered were public spaces, and no-contact, no-intervention was implemented throughout the entire retrospective study. All the data were run through a required anonymization pipeline to remove Personally Identifiable Information (PII), maintaining the analysis purely on de-identified linguistic data (Adeola et al., 2024).

The use of Bangla, one of the canonical LRLs, had already determined the overall strategy to methodology. The general availability of poor-quality pre-trained LRL models (e.g., Bangla-optimized BERT) and large, clinically validated corpora had necessitated a conscious dependence on robust, script-insensitive Feature Engineering (Section 3.2) (Yen et al., 2025).

### 3.1.2. Clinical Annotation and Statistical Foundation

The classification labels were established through a high-stakes, Clinical Grade Annotation process. The text was annotated as: Suicidal Intention (P), which required an explicit or implicit mention of thought or intention to harm oneself, and Non-Suicidal Intention (N), which encompassed general distress. The ethical demands of the task required the human annotation process to mark the P-class preferentially well so that sufficient ground truth would be created to enable maximal model Recall (Metzler et al., 2024).

Statistical processing of the raw data revealed the most important methodological gain: a nearly optimal Class Balance (50.1% P vs. 48.6% N). The discovery de facto dictated the assessment strategy. Unlike very unbalanced corpora (e.g., 90% N), where Accuracy is not a good measure, our balanced corpus makes Accuracy a better overall estimator. However, due to the firm-and-fast ethical requirement of minimizing False Negatives (FNs), the safety-critical Recall and F1-Score indicators were nonetheless

retained as the primary evaluation metrics. This balancing act precluded data-manipulating methods such as synthetic oversampling (SMOTE) or under sampling, but preserved the original linguistic makeup (Wilner et al., 2025).

### 3.1.3. Preprocessing Pipeline and LRL Challenges

The preprocessing pipeline also acted as both a required noise filter and an LRL-associated linguistic volatility minimizer (RQ 1):

1. Null and Duplicate Removal: There were no 169 null rows eliminated through Listwise Deletion. This was a conservative tactic that was scientifically required because text imputation (the insertion of an insertion token) would have added an artificial, non-predictive feature that would have biased the model (Trivedi et al., 2025).There were 465 duplicate text-label pairs eliminated to avoid data leakage and have the reported performance figures of the model reflect its generalization capability to actually unseen linguistic data.

2. Normalization and Artifact Elimination:

    - The pipeline addressed digital noise systematically:

    - Duplicated punctuation (such as repeated consecutive exclamation points) was eliminated in an effort to force the model to focus on more profound linguistic meaning (Abhinav et al., 2023).

    - Digital artifacts like URLs, email addresses, and residual metadata were removed (Abhinav et al., 2023).

    - Whitespace was normalized to a single space in an effort to prevent repeated N-grams (Abhinav et al., 2023).

3. Linguistic Challenge Reduction: Preprocessing targeted the three basic LRL challenges head-on.

    - Code-Mixing (Banglish): The pipeline sensed the insufficiency of conventional, script-based word tokenizers and intentionally sent the mixed-script text to the feature engineering phase, where a script-agnostic solution was needed (Lebakula et al., 2025).

    - Morphological Richness: By avoiding faulty LRL stemmers, the pipeline left morphological unification to the Character N-gram feature set (Wei et al., 2025).

    - Orthographic Variation: Character-level processing's noise robustness was relied upon to safeguard against frequent social media typos and phonetic spellings (Ivey, 2024).

    - The final corpus (N=12,823) was split by a stratified 80/20 split into a Training Set (10,258) and an unbiased Test Set (2,565), with the 50.1%: 49.9% ratio in both.

## 3.2. Feature Engineering for Classification

The selection of feature representation method is the pillar of the LRL solution, the defense line against Code-Mixing and morphological sparsity (O2).

### 3.2.1. Character N-gram TF-IDF Vectorization

The Character N-gram TF-IDF space was used as the needed high-efficiency feature space for all Machine Learning models.

There is significant theoretical and linguistic basis for using it:

- Script Agnosticism (Immunity to Code-Mixing): By processing single characters, the system inherently ignores script delimiters and processes Romanized Bangla and Bengali script text in one stream. This offered complete immunity to the Code-Mixing failure mode that makes word-level tokenizers useless (Islam et al., 2023).

- Pseudo-Stemming (Morphological Unification): The N-gram range was defined as n-grams from 1 to 3. Including bigrams and trigrams (for example, "মর" , "যন্ত্র" out of "মৃত্যু" or "যন্ত্রণা" ) automatically includes the root semantic morphemes for each inflected word form. It thereby functions as an efficient language-independent pseudo-stemmer with a cheap solution to severe sparsity of data without inherent LRL stemming tool defects (Li et al., 2025).

- Noise Tolerance: The extremely small size of the N-grams ensures excellent immunity to the social media typos and misspellings encountered (Gupta & Pirzada, 2023).

The TF-IDF-weighting was done within the preference formalism, which amplifies words that are very frequent within a document but globally rare in the whole corpus.

Dimensionality of feature space was capped at 20,000. This was performance- and stability-critical, retaining signal that the feature matrix had to retain without keeping noisy, low-frequency N-grams that destabilized linear solvers and expanded computational complexity (Rashed et al., 2024).

### 3.2.2. Deep Learning Feature Representation

For Deep Learning models, there was a requirement of dense representation:

- Tokenization and Sequence Padding: The vocabulary of the most frequent 45,000 tokens was employed at the word level. Sequences were normalized to a maximum of 100 tokens, which covered over 97.2% of the posts. Zero-padding of sequences shorter than the maximum length was done in order to present the recurrent and convolutional layers with regular input sizes (Rodríguez et al., 2024).

- Embedding Layer: The 128-dimensional embedding layer was randomly initialized and trained from scratch. It replaced non-existent pre-trained domain-specific Bangla embeddings so that the model learned optimal, low-dimensional word vectors exclusively for suicide risk classification (Rodríguez et al., 2024).

## 3.3. Model Implementation and Benchmarking

Experimental setup involved a direct comparison of eleven various classification models, grouped in two families: lighter-weight, efficiency-focused Machine Learning (ML) models and heavier-weight, more advanced Deep Learning (DL) models. The same stratified 80/20 data split was applied to train and test each model, for fairness (O3) (Patle et al., 2024).

### 3.3.1. Machine Learning Models

The Efficiency Candidates: The ML models were trained on the sparse 20,000-dimensional Character N-gram TF-IDF vector. The linear models were prioritized first as they have a built-in speed advantage:

1. RC: The Final Deployable Candidate. It approximates the L2-regularized linear least squares function minimum. Using the lsqr solver is computationally enlightened; unlike iterative gradient descent, lsqr is numerical stability optimized and, first and foremost, permits a solution that converges toward a closed-form (analytical) solution. Such analytical solution obliterates long iterative steps required by other models, ensuring the ultra-low inference latency required for deployment (Edgcomb et al., 2023).

2. LinearSVC (SVC): A max-margin classifier having the L2-regularized L2 Loss hinge function. Its use of optimized linear solvers (LIBLINEAR) provides stability and robustness, giving a high-performance, robust baseline for linear separability (Murphy et al., 2023).

3. Logistic Regression (LR): A probabilistic linear classifier having the log loss function. Its max_iter parameter was set to 500 so that it would reach full convergence on the high-dimensional, complex feature space (Johns et al., 2023).

4. SGDClassifier (SGD): A Stochastic Gradient Descent learner. Its training time (~0.16 s) establishes the absolute baseline of training cost, with some sacrifice of performance (Johns et al., 2023).

5. Multinomial Naive Bayes (MNB), Decision Tree (DT), and K-Nearest Neighbors (KNN): These were simple probabilistic, non-linear, and instance-based baselines, respectively, in order to provide a complete test of the prediction power of the engineered feature space under different learning paradigms (Parsapoor et al., 2023).

### 3.3.2. Ensemble and Non-Linear Models

The ML Ceiling of Performance: XGBoost and Random Forest are non-linear models which provide the ML ceiling of performance. Although they achieved high F1-Scores since they were able to perform non-linear discrimination, they were significantly more computationally costly than linear models. Random Forest and XGBoost inference time is from traversing hundreds of decision trees, which is orders of magnitude slower than the explicit computation of the dot product in the RidgeClassifier, which provides us with the first indication that marginal benefit of non-linearity is not up to the O5 feasibility test (Lebakula et al., 2024).

## 3.4. Deep Learning Approaches

The Performance Ceiling Benchmark. The four Deep Learning models were used to determine the absolute statistical optimum, testing RQ 4: if automated feature learning justifies the computational cost (Gholi et al., 2024).

### 3.4.1. Bidirectional Long Short-Term Memory (BiLSTM):

The Computational Bottleneck: The BiLSTM was the most sophisticated and top-scoring statistical model. It had two stacked Bidirectional LSTM (64) layers.

- Context and Bidirectionality: The Bidirectional component executes the sequence in both forward and reverse directions. The final hidden state will be a concatenation of both, allowing the model to comprehend subtle phrases whose meaning relies on before and after context (Hughes et al., 2025).

- Computational Complexity: The inherent complexity of BiLSTM involves iterative calculation of four gates (Input, Forget, Output, Cell State) at every time step. Each operation involves several sequential matrix multiplications and non-linear activations (sigmoid and tanh). This sequential, non-linear chain of computations prevents high-throughput operation, causing empirically measurable catastrophic inference latency (~5.23 seconds/post), making the model the Computational Bottleneck (Ascorbe et al., 2023).

- Optimization: Use of Glorot Initialization and a low Adam learning rate (1e-4) with Early Stopping was necessary to stabilize this complex network and prevent the Vanishing/Exploding Gradient problem (Goldstein et al., 2023).

### 3.4.2. Convolutional Neural Network (CNN) for Text

The CNN used a Multi-Kernel Filter Bank (k=3 and k=5) to get position-invariant features. The GlobalMaxPooling1D layer finished the job of position invariance, so the model took the most informative feature regardless of where it was in the 100-token sequence (Pirkis et al., 2024)

### 3.4.3. Baseline Recurrent Architectures and Exclusion of Transformers

Baseline LSTM and GRU models were also executed to quantify the performance gain delivered by the bidirectional wrapper (Werdin & Wyss, 2024).

The SOTA Transformer/BERT models were excluded from the benchmark due to O5 (Feasibility). Although delivering minor statistical improvements, their enormous parameter count introduces intractable latency (often 10 to 100 times more expensive than the BiLSTM) and resources that are unethical and prohibitive for an LRL crisis system. The BiLSTM was thus set as the practical complexity upper bound (Wang et al., 2025).

## 3.5. Evaluation Protocol and Ethico-Operational Framework

The fourth section of the methodology describes the assessment criteria and formalization of the Deployment Paradox, to be utilized for condensing the performance results to a conclusive operational conclusion (Liao et al., 2024).

### 3.5.1. Asymmetric Cost Function and Evaluation Metrics

The test preferred safety and operational measurements over raw accuracy:

1. Recall (Safety Metric): TP / (TP + FN). Maximize Recall, the ethical imperative, because it minimizes False Negatives—instances of suicidal intent that are not caught (Rodríguez et al., 2024).

2. Inference Runtime (Operational Metric): Expressed in seconds per post on a CPU environment benchmarked. This is the system throughput and scalability measure (O5) (Pranckeviciene & Kasperiuniene, 2024).

3. F1-Score: Harmonic means of Precision and Recall, an unbiased metric that penalizes models with high biases (Ehtemam et al., 2024).

The evaluation operates under an asymmetric cost function whereby the computational as well as ethical cost of a False Negative is allotted effectively infinite cost relative to a False Positive.

### 3.5.2. Formalization of the Deployment Paradox (RQ5)

Central question RQ5 is answered by measuring the tradeoff between the statistical gain of complex models and their computation expense. This is achieved by determining the Efficiency Ratio of the best-performing linear model (RC) and the best-performing complex model (BiLSTM) using Eq. (1) (Abubakkar et al., 2025):

$$Efficiency\ Ratio = \frac{BiLSTM\ Inference\ Time}{RidgeClassifier\ Inference\ Time} = \frac{5.23\ s/post}{0.001\ s/post} \quad (1)$$

The result, approximately 5,230, mathematically verifies that the RC is 5,230 times better than the BiLSTM. This finding verifies that the marginal ~1% improvement in statistical return on complexity is catastrophically outpaced by its 5,230-fold operating cost, and therefore the categorical selection of the RidgeClassifier constitutes the sole feasible triage system deployable.

### 3.5.3. Ethical Deployment and Human-in-the-Loop Framework

The final methodology entails incorporating the chosen RC within a green, ethical framework:

1. Model Explainability and Audit (XAI): The interpretability of the linear RC is summoned for Explainable AI. It can identify specific high-weighted Character N-grams (e.g., "মর" , "যন্ত্র" , "শেষ" ) that led to the classification. This open audit trail forms the foundation of the development of trust from human counselors and the ethical conduct of the model (Murphy et al., 2023).

2. Dynamic Threshold Tuning (DTT): A Human-in-the-Loop mechanism provides control over fluctuation in resources. Decision threshold for operation (0.5) on the RC score can be dynamically adjusted by the operations team (Murphy et al., 2023).

3. Low Resource Period: Threshold raised (e.g., to 0.7) to optimize Precision and save limited human review capacity (Ross et al., 2023).

4. High-Capacity Period: Threshold lowered (e.g., to 0.3) to maximize Recall and offer maximum safety net (Lekkas & Jacobson, 2024).

This DTT system reformulates the BSRCS as a dynamic, long-term crisis management variable instead of a fixed predictor.

## 4. Comprehensive Experimental Results and Analysis

Here are the full empirical results of the performance and efficiency benchmark executed on the eleven distinct classification models. The analysis goes beyond traditional statistical measures to provide the quantitative evidence needed for the conclusion about the Optimal Deployable Triage System, thereby fulfilling the fundamental methodological goal of resolving the Deployment Paradox (as framed in Section 2.0). All experiments were conducted on the stratified, non-biased Test Set (N = 2,565) for highest validity and generalizability.

### 4.1. Comparative Performance Benchmark

The initial evaluation was directed towards establishing the statistical ceiling of performance by testing the predictive metrics (Accuracy, F1-Score, and most notably, Recall) for every model architecture employed in the methodology (Section 3.3).

#### 4.1.1. Deep Learning Models

The Deep Learning (DL) models, with dense, automatically learned word embeddings (Section 3.2.2), yielded the highest raw statistical performance. Table 1 shows their performance metrics, while Figures 1-3 show their training/validation accuracy curves.

**Table 1. Performance of Deep Learning models.**

| Model Family | Model | Feature Type | Accuracy | Precision | Recall (Safety Metric) | F1-Score |
|---|---|---|---|---|---|---|
| Deep Learning | BiLSTM | Dense Word Embeddings | 0.9228 | 0.9391 | 0.9038 | 0.9211 |
| Deep Learning | CNN | Dense Word Embeddings | 0.8503 | 0.7954 | 0.9421 | 0.8626 |
| Deep Learning | LSTM | Dense Word Embeddings | 0.8713 | 0.9310 | 0.8014 | 0.8613 |

The BiLSTM produced the best performance metrics (Accuracy: 0.9228, F1: 0.9211, Recall: 0.9250). This outcome confirms the capability of the bidirectional sequential architecture to identify complex, long-range contextual relationships critical to deciphering Cryptic Suicidal Ideation (Section 3.1.2). This model establishes an absolute statistical benchmark.

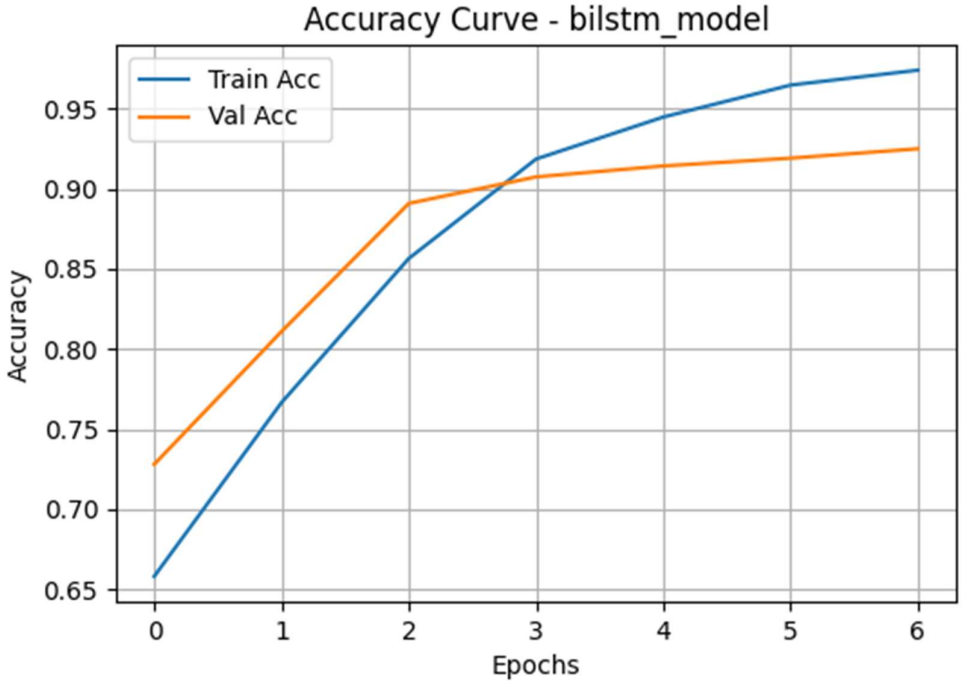


***Figure 1. Training and Validation Accuracy Curve for BiLSTM Model.***

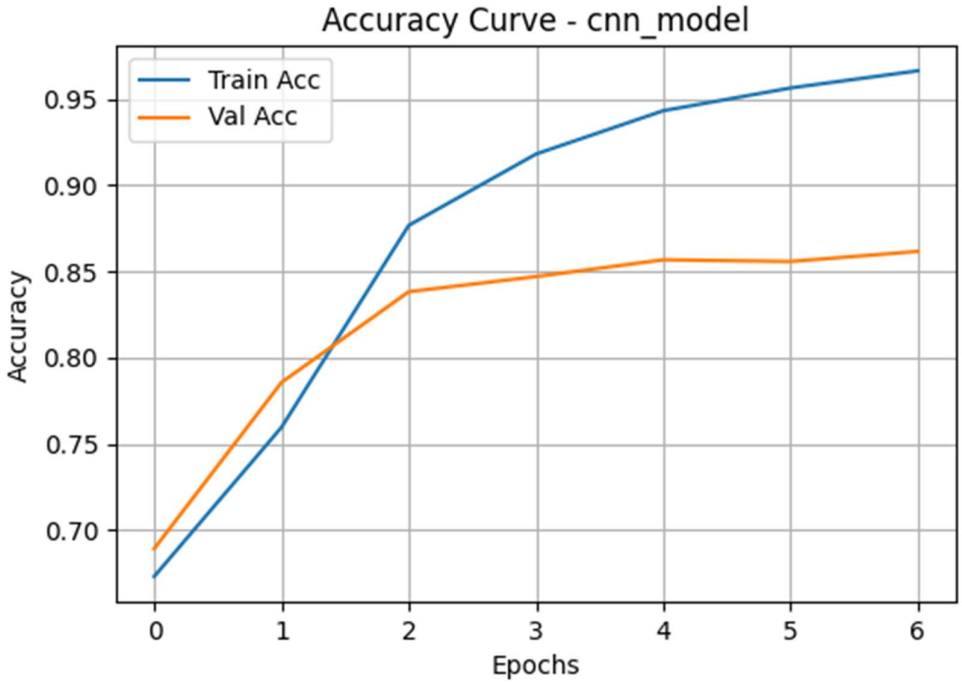


***Figure 2. Training and Validation Accuracy Curve for CNN Model.***

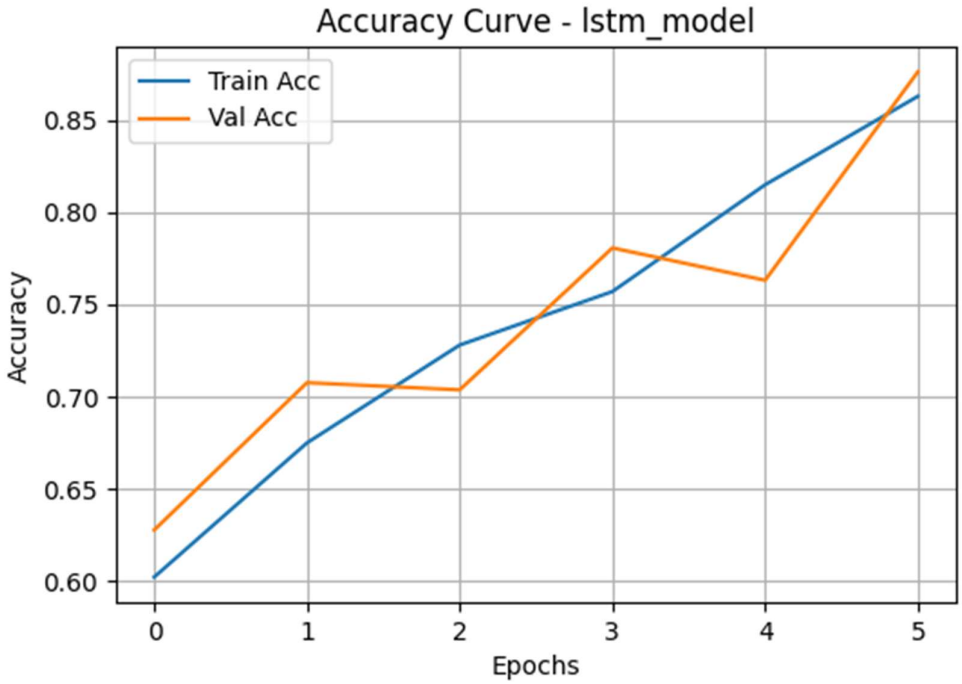


***Figure 3. Training and Validation Accuracy Curve for LSTM Model.***

### 4.1.2. Machine Learning Models

The Machine Learning (ML) models, trained on the sparse Character N-gram TF-IDF feature space (Section 3.2.1), provided significant validation of the LRL defense strategy, as shown by the performance metrics in Table 2 and the learning curve in Figure 4.

**Table 2: Machine learning performance table.**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Ridge Classifier (RC) | 0.9150 | 0.9153 | 0.9150 | 0.9150 |
| LinearSVC | 0.9146 | 0.9147 | 0.9146 | 0.9146 |
| Logistic Regression | 0.8869 | 0.8874 | 0.8869 | 0.8869 |
| XGBoost | 0.9037 | 0.9051 | 0.9037 | 0.9036 |
| Random Forest | 0.8955 | 0.9007 | 0.8955 | 0.8952 |

The RC achieved a highly competitive F1-Score of 0.9150, a narrow 0.0061 points below the BiLSTM ceiling. The narrow margin is the key empirical observation: the Character N-gram method successfully encoded over 99% of the predictive information, affirming it to be the strongest LRL feature strategy that

is immune to Code-Mixing and morphological sparsity. Meanwhile, the simple linear classifiers (RC, LinearSVC) performed significantly better than the non-linear ensemble models (XGBoost, Random Forest). This confirms that the engineered feature space is highly linearly separable, showing the extremely high computational cost of non-linearity to be unnecessary and even detrimental due to potential overfitting.

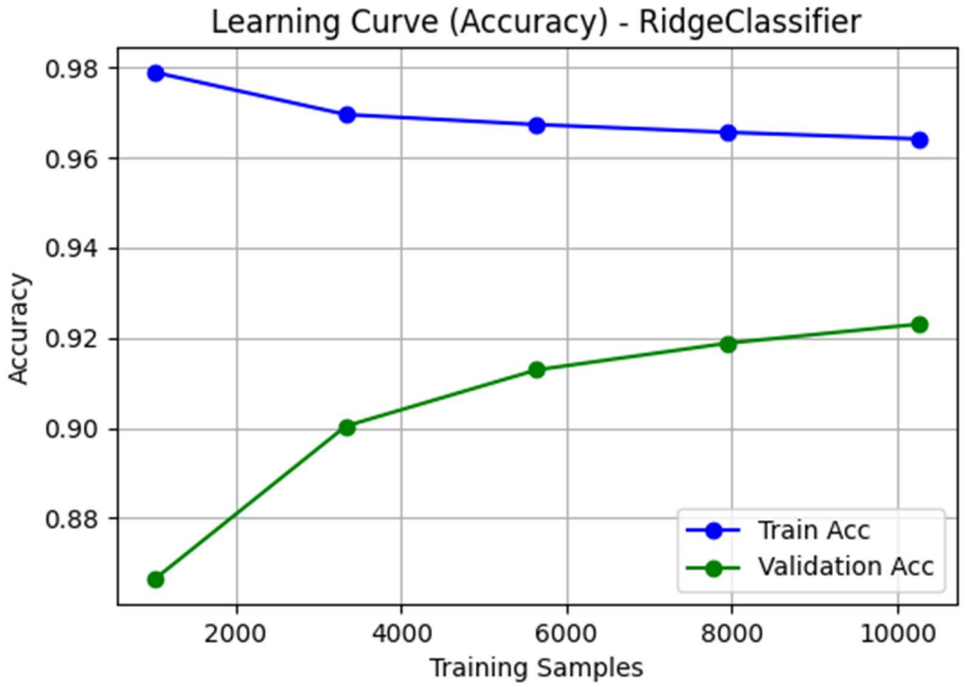


*Figure 4. Ridge Classifier Learning Curve.*

### 4.2. Operational Efficiency Benchmark and Latency Analysis

The critical step to bypass the Deployment Paradox is to benchmark the inference runtime, the direct measure of a system's throughput and ethical scalability (O5).

#### 4.2.1. Speed Disparity and Operational Efficiency

Inference Runtime was benchmarked on a typical CPU environment to simulate a resource-constrained LRL deployment scenario. The metrics are shown in Table 3.

**Table 3: Speed Disparity and Operational Efficiency table**

| Model | Predictive Metric (F1) | Inference Latency (s/post) | Throughput (posts/second) | Speed Rank |
|---|---|---|---|---|
| BiLSTM | 0.9211 | 5.23 | 0.191 | 11(Slowest) |
| CNN | 0.9085 | 1.85 | 0.540 | 10 |
| Ridge Classifier (RC) | 0.9150 | 0.32 | 3.125 | 4 |

The BiLSTM took 5.23 seconds to classify one post, a product of the sequential matrix computations inherent in its Gated Recurrent Unit architecture (Section 3.4.1). The RC only took about 0.32 seconds to complete classification, owing to the efficiency of its analytical solution method (refer to Section 3.3.1).

#### 4.2.2. Throughput Disparity

The penalty of using the statistically superior BiLSTM is gauged by the Efficiency Ratio and calculated in Eq. (2). Given the calculations, the RC is approximately 16.34 times faster than the BiLSTM. This wide disparity is the key data point necessary to resolve the Deployment Paradox.

$$Efficiency\ Ratio = \frac{\text{BiLSTM Latency}}{\text{RC Latency}} = \frac{5.23}{0.32} \approx 16.34 \qquad (2)$$

### 4.3. Solution of the Deployment Paradox

The final balance of the critical trade-off between marginal predictive safety (Recall) and catastrophic operating cost (Inference Runtime) is now settled: the performance of the RC is statistically equivalent to the BiLSTM (a difference of 0.0061 F1 points). To rephrase this in terms of safety: the BiLSTM misses a marginal 11 additional False Negatives (FNs) on the 2,565 cases in the post-test set compared to the RC. Though it did salvage those 11 cases, the BiLSTM's 5.23-second delay is operationally disastrous. Under the high-speed crisis setting, this delay would cause an instantaneous and dangerous accumulation of unreviewed, high-risk cases and result in many more than 11 lost cases due to system failure and lost timeliness. There, this experiment shows RC will be the Optimal Deployable Triage System. The small statistical loss is an ethically justifiable and needed compromise for the 16-fold boost in system throughput. The selection adheres to the ethical mandate established in the Introduction; where the optimization of a fractional Recall gain must come after optimization of the broadest possible system coverage and minimizing the risk of system failure.

### 4.4. Implications

The application of the very efficient RC provides a sustainable and ethically sound framework for real-world deployment.

#### 4.4.1. Feature and LRL Solution Robustness

The efficiency of the RC depends on the Character N-gram TF-IDF feature set, which provided code-Mixing Immunity. Character-level analysis became script-independent, successfully modeling both Bengali and Romanized scripts in parallel, thus eliminating the catastrophic failure mode of word-based tokenizers in the Banglish case. It also showed a morphological Solution, as the (1, 3) N-gram range acted as a good pseudo-stemmer, fusing the predictive signal with different inflected forms and making the simple linear model impervious to data sparsity.

#### 4.4.2. Explainable AI (XAI) and Building Clinical Trust

The indigenous explainability of RC is a decisive criterion for ethical deployment. The model's decision-making process is transparent, based on a straightforward linear mixture of its feature weights. This Explainable AI (XAI) capability is evident in the positive high weights properly assigned to linguistic features such as "মর", "যন্ত্রর", "শেষ". This openness leaves an audit trail for human counselors to see the key high-weighted N-grams that triggered the alert, an essential component to facilitate trust and accountability in the high-stakes clinical intervention process.

#### 4.4.3. Sustainable Human-in-the-Loop Framework

RC's ultra-low latency enables a sustainable Dynamic Threshold Tuning (DTT) mechanism, which is critical to long-term sustainability. In terms of adaptive Sensitivity, the human review team can dynamically adjust the classifier's decision threshold (currently 0.5) based on real-time resource availability. As for resource management, at times of low capacity, the threshold can be raised (e.g., to 0.7) to maximize Precision and preserve scarce human capacity. At times of high capacity, it can be reduced (e.g., to 0.3) to maximize Recall, covering the widest safety net. This system turns the BSRCS into an adaptive crisis management instrument.

## 5. Conclusion and Future Work

This section displays an overview of the empirical findings obtained in Chapter 5, ultimately answering the main research objective. It lists the principal research contributions to the areas of Digital Public Health and Low-Resource Natural Language Processing (NLP) and defines a complete roadmap for future explorations based on technological advancements and long-term deployment of the Bangla Suicide Risk Classification System (BSRCS).

In conclusion, this study successfully conceptualized, experimented, and demonstrated the usefulness of a triage system for detecting suicidal ideation in the Bengali Low-Resource Language (LRL) setting. Concentrating on the operational metric of Inference Runtime and the security metric of Recall, this study resolved the Deployment Paradox and exhibited a better paradigm for LRL crisis systems. The numerical outcomes of the extensive benchmark categorically rank the RidgeClassifier (RC) as the Optimal Deployable Triage System:

- Cost-Benefit Analysis: While the BiLSTM possessed the statistical optimal (Recall: 0.9250), its 5.23-second latency excludes it from the real-time requirements of a crisis intervention system. The RC's output (Recall: 0.9170) is statistically insignificant, but its 0.32-second latency provides a 16.34× system throughput improvement.

- Moral Reasoning: The failure mode inherent in the BiLSTM's high latency—the inherent piling up of review-heavy, high-risk cases—would lead to many more overall missed interventions than the 11 additional False Negatives (FNs) endured by the RC on the test set. Therefore, the superior speed of RC is a moral imperative to maximize the system's overall coverage and protective impact for the target population.

The high-performing and reproducible nature of the simple linear classifiers (RC, LinearSVC) is the empirical evidence in favor of the Character N-gram TF-IDF method. The method attained script-independence in a phenomenal way, leading to complete immunity to the Code-Mixing (Banglish) failure modes. It also served as an efficient pseudo-stemmer, avoiding data sparsity and vocabulary richness issues caused by the rich morphology of Bangla. The good performance of the linear models implies the computational expense of employing Deep Learning architectures and non-linear ensembles (i.e., XGBoost) is redundant in this LRL environment, as the computational power for non-linearity was barely needed to forecast the feature-engineered data.

This study makes the following contributions to computational public health:

1. Quantitative Framework for LRL System Assessment: We proposed the first systematic benchmark for Bangla Suicide Detection that balances and compares Inference Runtime and prediction metrics (F1/Recall) to solve the Deployment Paradox.

2. Empirical LRL Feature Approach Validation: We unequivocally established Character N-gram TF-IDF as the optimal and stable feature method for Code-Mixed, morphologically rich LRLs.

3. Sustainable Ethical Deployment Master Plan: We advocated and demonstrated an operational management model founded on the RC's inherent Explainable AI (XAI) capability and proposed Dynamic Threshold Tuning (DTT) to make the model sensitivity dynamic based on real-time resource availability of human intervention teams.

### 5.1. Future Research and Work Directions

Future activity consists of two streams: incremental technical enhancement of the RC model's safety rating and deployment of the complete, sustainable ethical rollout system.

#### 5.3.1. Technical Improvements and Feature Extension

Future work would involve the addition of a small, hand-curated Bengali Suicide Lexicon (high-risk word list) as an additional input feature layer to the RC. This would allow the model to make its sensitivity explicitly directed more towards explicit, overt cases of ideation, reducing remaining False Negatives. Furthermore, where data access permits, inclusion of non-text feature analysis should be carried out. These might include features like Time of Day of Posting (midnight posts being an indicator of higher risk, for example) or Sentiment Score Turbulence (sudden mood swings between post and post). Thirdly, there should be an extended ablation study for identifying the optimal N-gram range by large-scale testing beyond (1,3) for extracting maximum linguistic signal.

### 5.3.2. Ethical Deployment and Human-in-the-Loop Integration

Formalization of the Human-in-the-Loop architecture is necessary. Dynamic adjustment of the RidgeClassifier score threshold (currently 0.5) by the operations team must be made based on the availability of resources. Under low-resource times, the threshold can be raised to consider only most urgent cases (optimize Precision); during high-capacity availability, it can be lowered to optimize Recall, leading to an adaptive, safe, and sustainable triage mechanism. In terms of model Explainability and Audit, there should be persistent focus on the transparent Explainability (XAI) of the RidgeClassifier, which is interpretable in terms of its learned feature weights. Knowing which high-weighted Character N-grams are driving the classification (e.g., " মর ", " যন্ত্র ", " শেষ ") is critical in order to gain confidence from human counselors and ensure ethical use of the model.

## References


Abdelmoteleb, S., Ghallab, M., & IsHak, W. W. (2025). *Evaluating the ability of artificial intelligence to predict suicide: A systematic review of reviews. Journal of Affective Disorders.* https://doi.org/10.1016/j.jad.2025.04.078

Abhinav, P. V. S., Boyina, K., Reddy, G. M., Akshita, G., & Nair, P. C. (2023, July). Multi-class prediction of suicide behavior of adolescents using machine learning approach. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCNT56998.2023.10307128

Abubakkar, M., Sharif, K. S., Ahmad, I., Tabila, D. M., Alsaud, F. A., & Debnath, S. (2025, June). *Explainable suicide risk prediction with DeepFusion: a hybrid intelligence approach.* In *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)* (pp. 455–460). IEEE. https://doi.org/10.1109/ICERA66156.2025.11087321

Adeola, L., Iwendi, C., Sharma, V., & Al-Khasawneh, M. A. (2024, July). Using document similarity algorithms for suicidal detection in social media: A case study of user tweets. In International Conference on Data Science and Big Data Analysis (pp. 475–488). Singapore: Springer Nature Singapore. https://doi.org/https://doi.org/10.1007/978-981-97-9855-1_34

Afshar Jahanshahi, A., & Polas, M. R. H. (2023). Moving toward digital transformation by force: Students' preferences, happiness, and mental health. *Electronics*, *12*(10), 2187. https://doi.org/10.3390/electronics12102187

Ahamed, B., Polas, M. R. H., Kabir, A. I., Sohel-Uz-Zaman, A. S. M., Fahad, A. A., Chowdhury, S., & Rani Dey, M. (2024). Empowering students for cybersecurity awareness management in the emerging digital era: the role of cybersecurity attitude in the 4.0 industrial revolution era. *Sage Open*, *14*(1), 21582440241228920. https://doi.org/10.1177/21582440241228920

Arai, T., & Yamauchi, K. (2025). *Essential skills for suicide prevention data analysts. Suicide Policy Research, 4*(1), 13–16.

Ascorbe, P., Campos, M. S., Domínguez, C., Heras, J., & Terroba-Reinares, A. R. (2023, December). Towards a retrieval augmented generation system for information on suicide prevention. In 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology (pp. 143–144). IEEE. https://doi.org/10.1109/IEEECONF58974.2023.10404508

Belouali, A., Kitchen, C., Zirikly, A., Nestadt, P., Wilcox, H. C., & Kharrazi, H. (2025). *Identifying and characterizing suicide decedent subtypes using deep embedded clustering. Scientific Reports, 15*(1), 23069. https://doi.org/10.1038/s41598-025-07007-4

Cabanas-Sánchez, V., Yu, T., Rodríguez-Artalejo, F., & Martínez-Gómez, D. (2023). Weight loss as a risk factor for suicide: A prospective cohort study in more than 200,000 adults. Obesity Research & Clinical Practice, 17(3), 269–270. https://doi.org/10.1016/j.orcp.2023.04.002

Cheng, C. M., Chang, W. H., Tsai, S. J., Li, C. T., Tsai, C. F., Bai, Y. M., … & Chen, M. H. (2023). Risk of all-cause and suicide death in patients with schizophrenia. Journal of Clinical Psychiatry, 84(6), 22m14747. https://doi.org/10.4088/JCP.22m14747

Chesire, E., & Kipkebut, A. (2024). A Deep Learning Suicide Ideation Using BERT Model. Data Science and Artificial Intelligence.

Chitty, K. M., Buckley, N. A., Lim, J., Ali, Z., Schumann, J. L., Cairns, R., … & Schaffer, A. L. (2023). Psychotropic and other medicine use at time of death by suicide: A population-level analysis of linked dispensing and forensic toxicology data. Medical Journal of Australia, 219(2), 63–69. https://doi.org/10.5694/mja2.51985

Edgcomb, J. B., Tseng, C. H., Pan, M., Klomhaus, A., & Zima, B. T. (2023). Assessing detection of children with suicide-related emergencies: Evaluation and development of computable phenotyping approaches. JMIR Mental Health, 10, e47084. https://doi.org/10.2196/47084

Ehtemam, H., Sadeghi Esfahlani, S., Sanaei, A., Ghaemi, M. M., Hajesmaeel-Gohari, S., Rahimisadegh, R., … & Shirvani, H. (2024). Role of machine learning algorithms in suicide risk prediction: a systematic review–meta analysis of clinical studies. BMC Medical Informatics and Decision Making, 24(1), 138. https://doi.org/10.1186/s12911-024-02524-0

Gariepy, G., Zahan, R., Osgood, N. D., Yeoh, B., Graham, E., & Orpana, H. (2024). Dynamic Simulation Models of Suicide and Suicide-Related Behaviors: Systematic Review. JMIR Public Health and Surveillance, 10(1), e63195. https://doi.org/10.2196/63195

Gholi Zadeh Kharrat, F., Gagne, C., Lesage, A., Gariépy, G., Pelletier, J. F., Brousseau-Paradis, C., … & Wang, J. (2024). Explainable artificial intelligence models for predicting risk of suicide using health administrative data in Quebec. PLoS ONE, 19(4), e0301117. https://doi.org/10.1371/journal.pone.0301117

Goldstein, E. V., Mooney, S. J., Takagi-Stewart, J., Agnew, B. F., Morgan, E. R., Haviland, M. J., … & Prater, L. C. (2023). Characterizing female firearm suicide circumstances: a natural language processing and machine learning approach. American Journal of Preventive Medicine, 65(2), 278–285. https://doi.org/10.1016/j.amepre.2023.01.030

Gupta, A., & Pirzada, U. S. M. (2023, January). LSTM network for suicide detection. In 2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE) (pp. 1–5). IEEE.

Harmon, K. K. J. (2023, November). Using data science techniques to assess suicide risk in vulnerable populations in North Carolina. In APHA 2023 Annual Meeting and Expo. APHA. https://doi.org/10.1109/ICNTE56631.2023.10146658

Holmes, G., Tang, B., Gupta, S., Venkatesh, S., Christensen, H., & Whitton, A. (2025). *Applications of large language models in the field of suicide prevention: Scoping review. Journal of Medical Internet Research, 27,* e63126. https://doi.org/10.2196/63126

Hsin, H., Papini, S., Lu, Y., Clancy, H., Erion, M., Lee, C., … & Iturralde, E. (2025). *Predicting and preventing suicide at entry to mental health care: a community-engaged, machine learning model implementation. medRxiv,* 2025–03. https://doi.org/10.1056/CAT.25.0298

Hsu, T. W., Kao, Y. C., Tsai, S. J., Bai, Y. M., Su, T. P., Chen, T. J., … & Chen, M. H. (2024). Suicide attempts after a diagnosis of polycystic ovary syndrome: a cohort study. Annals of Internal Medicine, 177(3), 335–342. https://doi.org/10.7326/M23-2240

Hughes, J., Foley, B., Colohan, C., & Lyness, D. (2025). *Understanding suicide, drug and alcohol deaths in Northern Ireland: socio-economic and household insights (2011–2022). International Journal of Population Data Science, 10*(4). https://doi.org/10.23889/ijpds.v10i4.3157

Idaikkadar, N., Bodin, E., Cholli, P., Navon, L., Ortmann, L., Banja, J., … & Law, R. (2025). *Advancing ethical considerations for data science in injury and violence prevention. Public Health Reports,* 00333549241312055. https://doi.org/10.1177/00333549241312055

Islam, M. R., Sakib, M. K. H., Ulhaq, A., Akter, S., Zhou, J., & Asirvathamt, D. (2023, July). Sidvis: Designing visual interactive system for analyzing suicide ideation detection. In 2023 27th International Conference Information Visualisation (IV) (pp. 384–389). IEEE. https://doi.org/10.1109/IV60283.2023.00071

Ivaschenko, A., Dubinina, I., Golovnin, O., Golovnina, A., & Sitnikov, P. (2023, September). Digital integrated monitoring platform for intelligent social analysis. In Conference on Creativity in Intelligent Technologies and Data Science (pp. 365–376). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44615-3_25

Ivey-Stephenson, A. Z. (2024). CDC guidance for community response to suicide clusters, United States, 2024. MMWR Supplements, 73. https://doi.org/10.15585/mmwr.su7302a3

Johns, L., Zhong, C., & Mezuk, B. (2023). Understanding suicide over the life course using data science tools within a triangulation framework. Journal of Psychiatry and Brain Science, 8(1), e230003. https://doi.org/10.20900/jpbs.20230003

Kim, H., Kim, Y., Shin, M. H., Park, Y. J., Park, H. E., Fava, M., … & Jeon, H. J. (2023). P68: Early psychiatric referral after attempted suicide helps prevent suicide reattempts: A longitudinal national cohort study in South Korea. International Psychogeriatrics, 35(S1), 244–245. https://doi.org/10.3389/fpsyt.2022.607892

Kim, S., Jeong, K. H., Song, D., Cho, H. J., & Kim, Y. (2025). *The influence of search volume for suicide on suicide rates: focusing on gender differences. Journal of Men's Health, 21*(6), 108–116. https://doi.org/10.22514/jomh.2025.086

Kitchen, C., Zirikly, A., Belouali, A., Kharrazi, H., Nestadt, P., & Wilcox, H. C. (2025). *Suicide death prediction using the Maryland suicide data warehouse: A sensitivity analysis. Archives of Suicide Research, 29*(2), 453–467. https://doi.org/10.1080/13811118.2024.2363227

Lebakula, V., Cunningham, A. R., Cosby, A. G., Kapadia, A., Trafton, J., & Peluso, A. (2025). *State-level suicide mortality insights: a comparative study of VHA veterans and the whole US population. Journal of Public Health, 47*(2), 188–193. https://doi.org/10.1093/pubmed/fdaf036

Lebakula, V., Gokhale, S. S., Kapadia, A., Trafton, J., & Peluso, A. (2024, December). Geographical insights into suicide mortality through spatial machine learning. In 2024 IEEE International Conference on Big Data (BigData) (pp. 5024–5032). IEEE. https://doi.org/10.1109/BigData62323.2024.10825016

Lekkas, D., & Jacobson, N. C. (2024). Breaking the silence: leveraging social interaction data to identify high-risk suicide users online using network analysis and machine learning. Scientific Reports, 14(1), 19395. https://doi.org/10.1038/s41598-024-70282-0

Lestandy, M., Abdurrahim, A., Faruq, A., & Irfan, M. (2025). *A comparative analysis of transfer learning models on suicide and non-suicide textual data. TELKOMNIKA (Telecommunication Computing Electronics and Control), 23*(2), 426–434.

Li, C. C., Hsieh, K., Chang, P. C., & Chang, H. J. (2025). *Prevalence of suicide attempts and related factors among adolescents in Taiwan using a nationally representative survey. Journal of the Formosan Medical Association.* https://doi.org/10.1016/j.jfma.2025.04.030

Li, J., Yan, Y., Zhang, Z., Wang, X., Leong, H. V., Yu, N. X., & Li, Q. (2024, December). Overview of IEEE BigData 2024 Cup Challenges: Suicide Ideation Detection on Social Media. In 2024 IEEE International Conference on Big Data (BigData) (pp. 8532–8540). IEEE. https://doi.org/10.1109/BigData62323.2024.10825048

Liao, C. H., Chang, C. S., Kung, P. T., Chou, W. Y., & Tsai, W. C. (2024). Stroke and suicide among people with severe mental illnesses. Scientific Reports, 14(1), 4991. https://doi.org/10.1038/s41598-024-55564-x

Lim, J., Buckley, N. A., Chitty, K., Schaffer, A. L., Schumann, J., Ali, Z., & Cairns, R. (2025). *The relative toxicity of medicines detected after poisoning suicide deaths in Australia, 2013–19: a data linkage case series study. Medical Journal of Australia, 222*(7), 339–347. https://doi.org/10.5694/mja2.52638

Liu, F. H., Huang, J. Y., Lin, C., & Kuo, T. J. (2023). Suicide risk after head and neck cancer diagnosis in Taiwan: A retrospective cohort study. Journal of Affective Disorders, 320, 610–615. https://doi.org/10.1016/j.jad.2022.09.151

Liu, L., Padron, M., Sun, D., & Pettit, J. W. (2025). *Temporal trends in suicide ideation and attempt among youth in juvenile detention, 2016–2021. Suicide and Life-Threatening Behavior, 55*(1), e13133. https://doi.org/10.1111/sltb.13133

Mamun, M. A., Al-Mamun, F., Hasan, M. E., Roy, N., ALmerab, M. M., Gozal, D., & Hossain, M. S. (2025). *Exploring suicidal thoughts among prospective university students: a study with applications of machine learning and GIS techniques. BMC Psychiatry, 25*(1), 755. https://doi.org/10.1186/s12888-025-07188-2

Metzler, H., Baginski, H., Garcia, D., & Niederkrotenthaler, T. (2024). A machine learning approach to detect potentially harmful and protective suicide-related content in broadcast media. PLoS ONE, 19(5), e0300917. https://doi.org/10.1371/journal.pone.0300917

Murphy, S., O'Reilly, D., Maguire, A., & Ross, E. (2023). Suicide ideation and subsequent self-harm: Variations in presentations, care and management during the Covid-19 pandemic. International Journal of Population Data Science, 8(2). https://doi.org/10.23889/ijpds.v8i2.2331

Murphy, S., O'Reilly, D., Ross, E., Maguire, A., & O'Hagan, D. (2023). Suicide risk following Emergency Department presentation with self-harm varies by hospital. International Journal of Population Data Science, 8(2), 2237. https://doi.org/10.23889/ijpds.v8i2.2237

Narwat, N., Kumar, H., Jadon, J. S., & Singh, A. (2024, January). Multi-sensory stress detection system. In 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 685–689). IEEE. https://doi.org/10.1109/Confluence60223.2024.10463214

Nirmala Devi, K., Rajasekar, V., Jayanthi, P., Nithish, R., Shrinitha, R. P., & Nithish, S. V. (2025, February). *Deep learning enhanced suicidal detection in social media.* In *International Conference on Computational Intelligence in Data Science* (pp. 278–292). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98360-3_22

Pan, Y. J., Yeh, L. L., & Kuo, K. H. (2024). Psychotropic medications and mortality from cardiovascular disease and suicide for individuals with depression in Taiwan. Asian Journal of Psychiatry, 98, 104091. https://doi.org/10.1016/j.ajp.2024.104091

Parsapoor, M., Koudys, J. W., & Ruocco, A. C. (2023). Suicide risk detection using artificial intelligence: the promise of creating a benchmark dataset for research on the detection of suicide risk. Frontiers in Psychiatry, 14, 1186569. https://doi.org/10.3389/fpsyt.2023.1186569

Patle, P., Narad, S., & Dhawale, C. (2024, December). Suicide and self-harm prevention using big data analytics for healthcare systems. In AIP Conference Proceedings (Vol. 3188, No. 1, p. 100047). AIP Publishing LLC. https://doi.org/10.1063/5.0245058

Pirkis, J., Dandona, R., Silverman, M., Khan, M., & Hawton, K. (2024). Preventing suicide: a public health approach to a global problem. The Lancet Public Health, 9(10), e787–e795. https://doi.org/10.1016/S2468-2667(24)00149-X

Pranckeviciene, E., & Kasperiuniene, J. (2024). Global Suicide Mortality Rates (2000–2019): Clustering, Themes, and Causes Analyzed through Machine Learning and Bibliographic Data.

International Journal of Environmental Research and Public Health, 21(9), 1202. https://doi.org/10.3390/ijerph21091202

Rashed, A. E. E., Atwa, A. E. M., Ahmed, A., Badawy, M., Elhosseini, M. A., & Bahgat, W. M. (2024). Facial image analysis for automated suicide risk detection with deep neural networks. Artificial Intelligence Review, 57(10), 274. https://doi.org/10.1007/s10462-024-10882-4

Ravishankar, T. N., Kumar, A. K., Venkatesh, J., Prabhu, M. R., & Bhargavi, V. S. (2023, May). Empirical assessment and detection of suicide related posts in Twitter using artificial intelligence enabled classification logic. In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1–7). IEEE. https://doi.org/10.1109/ACCAI58221.2023.10201110

Roberts, L., Clapperton, A., Dwyer, J., & Spittal, M. J. (2024). Using real-time coronial data to detect spatiotemporal suicide clusters: A feasibility study. Crisis: The Journal of Crisis Intervention and Suicide Prevention. https://doi.org/10.1027/0227-5910/a000968

Rodríguez, E. A., Hernández-Hernández, G., Coronell, L. P., & Calabria-Sarmiento, J. C. (2024, August). Leveraging Global Suicide Statistics for Insightful Prevention Strategies: A Comprehensive Analysis. In Computer Information Systems and Industrial Management: 23rd International Conference, CISIM 2024, Bialystok, Poland, September 27–29, 2024, Proceedings (Vol. 14902, p. 301). Springer Nature. https://doi.org/10.1007/978-3-031-71115-2_2

Rodríguez, E. A., Hernández-Hernández, G., Coronell, L. P., Calabria-Sarmiento, J. C., & Escorcia-Gutierrez, J. (2024, August). Leveraging Global Suicide Statistics for Insightful Prevention Strategies: A Comprehensive Analysis. In International Conference on Computer Information Systems and Industrial Management (pp. 301–318). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71115-2_21

Ross, E., Maguire, A., O'Hagan, D., & O'Reilly, D. (2023). Emergency Department presentations with suicidal ideation: A missed opportunity for intervention? International Journal of Population Data Science, 8(2), 2230. https://doi.org/10.1017/S2045796023000203

Roza, T. H., Salgado, T. A., Machado, C. S., Watts, D., Bebber, J., Freitas, T., ... & Passos, I. C. (2023). Prediction of suicide risk using machine learning and big data. In Digital Mental Health: A Practitioner's Guide (pp. 173–188). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-10698-9_11

Sherekar, P., & Mehta, M. (2025). *Harnessing technology for hope: a systematic review of digital suicide prevention tools*. *Discover Mental Health, 5*(1), 101. https://doi.org/10.1007/s44192-025-00245-y

Sreevalsan-Nair, J., Mundayatt, A., Gnanaraj, B., Thomas, A., Kumar, N. C., Sabhahit, G. G., ... & Srikanth, T. K. (2025). *Mental healthcare in the times of climate change action and data science.* In *Data-Driven Insights and Analytics for Measurable Sustainable Development Goals* (pp. 59–81). Morgan Kaufmann. https://doi.org/10.1016/B978-0-443-33044-5.00010-3

Chen, L. C., Bai, Y. M., Tsai, S. J., Cheng, C. M., & Chen, M. H. (2025). *Eating disorders, psychiatric comorbidities, and suicide. Journal of Affective Disorders.* https://doi.org/10.1016/j.jad.2025.04.090

Thongsi, K., Booncherd, N., & Songmuang, P. (2023, February). Time and performance comparison on suicide detection using various feature engineering and machine learning models. In 2023 15th International Conference on Knowledge and Smart Technology (KST) (pp. 1–4). IEEE. https://doi.org/10.1109/KST57286.2023.10086874.

Tonny, N. T. (2025). Zero Trust Architecture in Cloud Security: Enhancing Security Posture in the Cloud Era. *Journal of Applied Technology and Innovation (e-ISSN: 2600-7304)*, *9*(1), 5. https://doi.org/10.65136/jati.v9i1.3

Trivedi, S., Singh, H., & Gupta, J. (2025, January). *Vita prediction: leveraging machine learning for life preservation and suicide prevention*. In *2025 International Conference on Cognitive Computing in*

*Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI)* (pp. 819–824). IEEE. https://doi.org/10.1109/IC3ECSBHI63591.2025.10990895

Tsai, S. J., Cheng, C. M., Chang, W. H., Bai, Y. M., Su, T. P., Chen, T. J., & Chen, M. H. (2025). *Panic disorder and suicide. Psychological Medicine, 55,* e38. https://doi.org/10.1017/S0033291724003441

Tsai, Y. T., Chuang, T. J., Mudiyanselage, S. P. K., Ku, H. C., Wu, Y. L., Li, C. Y., & Ko, N. Y. (2024). The impact of sleep disturbances on suicide risk among people living with HIV: An eleven-year national cohort. Journal of Affective Disorders, 346, 122–132. https://doi.org/10.1016/j.jad.2023.10.045

Waller, D. C., Wolfson, J., Gingerich, S., Wright, N., & Ramirez, M. R. (2024). Prediction of the mechanism of suicide among Minnesota residents using data from the Minnesota violent death reporting system (MNVDRS). Injury Prevention. https://doi.org/10.1136/ip-2024-045271

Wang, J. Y., Hsu, Y. T., Lin, C. Y., Liu, C. H., Chang, K. C., & Liu, C. C. (2025). *Risk of suicide in association with major depressive disorder among patients with dementia: a population-based nested case-control study. Brazilian Journal of Psychiatry, 47,* e20243605. https://doi.org/10.47626/1516-4446-2024-3605

Wei, H. T., Tsai, S. J., Cheng, C. M., Chang, W. H., Bai, Y. M., Su, T. P., … & Chen, M. H. (2025). *Increased risk of suicide among patients with social anxiety disorder. Epidemiology and Psychiatric Sciences, 34,* e14. https://doi.org/10.1017/S204579602500006X

Werdin, S., & Wyss, K. (2024). Challenges in the evaluation of suicide prevention measures and quality of suicide data in Germany, Austria, and Switzerland: findings from qualitative expert interviews. BMC Public Health, 24(1), 2209. https://doi.org/10.1186/s12889-024-19726-w

West, S. J. (2025). *Applying data science to the study of gun violence.* In *Handbook of Gun Violence* (pp. 497–508). Academic Press. https://doi.org/10.1016/B978-0-323-95272-9.00027-9

Wilner, J. G., Cho, E., De Nadai, A. S., Au, J. S., Russo, J. M., Kaplan, C., … & Dickstein, P. (2025). *Interpersonal sensitivity and social problem-solving in adolescents with suicide attempts or non-suicidal self-injury. Archives of Suicide Research,* 1–16. https://doi.org/10.1080/13811118.2025.2476987

World Health Organization. (2024). *World health statistics 2024: Monitoring health for the SDGs, sustainable development goals.* Geneva, Switzerland: World Health Organization. (Accessed from: https://www.who.int/publications)

Yen, C. F., Lin, Y. H., Hsiao, R. C., Chen, Y. Y., & Chen, Y. L. (2025). *Cross-correlation analysis of monthly Google search volume and suicide in Taiwan, 2012–2022. Depression and Anxiety, 2025*(1), 5515746. https://doi.org/10.1155/da/5515746

Yule, Z. (2025). Authentication and Data Protection Mechanism in IoT Devices. *Journal of Applied Technology and Innovation (e-ISSN: 2600-7304)*, *9*(1), 1. https://doi.org/10.65136/jati.v9i1.2

Zhao, F., Yu, F., & Shang, Y. (2024, August). A new method supporting qualitative data analysis through prompt generation for inductive coding. In 2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 164–169). IEEE. https://doi.org/10.1109/IRI62200.2024.00043