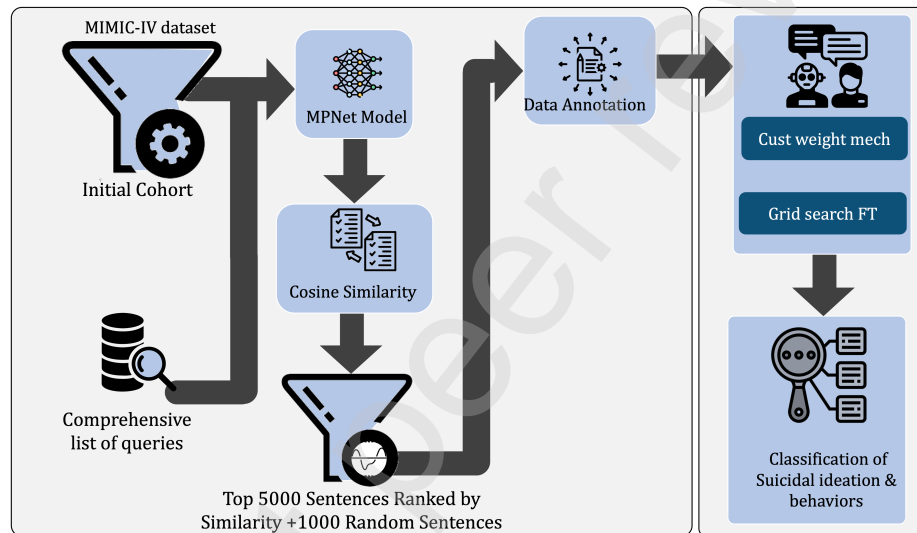


Graphical Abstract

Enhancing Suicidal Behavior Detection in EHRs: A Multi-Label NLP Framework with Transformer Models and Semantic Retrieval-Based Annotation

Kimia Zandbiglari, MS, Shobhan Kumar, PhD, Muhammad Bilal, PhD,
Amie Goodin, PhD, MPP, Masoud Rouhizadeh, PhD, MS, MA



Enhancing Suicidal Behavior Detection in EHRs: A Multi-Label NLP Framework with Transformer Models and Semantic Retrieval-Based Annotation

Kimia Zandbiglari, MS^a, Shobhan Kumar, PhD^a, Muhammad Bilal, PhD^a,
Amie Goodin, PhD, MPP^a, Masoud Rouhizadeh, PhD, MS, MA^{a,b}

^a*Department of Pharmaceutical Outcomes & Policy, University of
Florida, Gainesville, FL, USA*

^b*Division of Biomedical Informatics & Data Science, Johns Hopkins University School of
Medicine, Baltimore, MD, USA*

Abstract

Background: Suicide is a leading cause of death worldwide, making early identification of suicidal behaviors crucial for clinicians. Current Natural Language Processing (NLP) approaches for identifying suicidal behaviors in Electronic Health Records (EHRs) rely on keyword searches, rule-based methods, and binary classification, which may not fully capture the complexity and spectrum of suicidal behaviors. This study aims to create a multi-class labeled dataset with annotation guidelines and develop a novel NLP approach for fine-grained, multi-label classification of suicidal behaviors, improving the efficiency of the annotation process and accuracy of the NLP methods.

Methods: We develop a multi-class labeling system based on guidelines from FDA, CDC, and WHO, distinguishing between six categories of suicidal behaviors and allowing for multiple labels per data sample. To efficiently create an annotated dataset, we use an MPNet-based semantic retrieval framework to extract relevant sentences from a large EHR dataset, reducing annotation space while capturing diverse expressions. Experts annotate the extracted sentences using the multi-class system. We then formulate the task as a multi-label classification problem and fine-tune transformer-based models on the curated dataset to accurately classify suicidal behaviors in EHRs.

Results: Lexical analysis revealed key themes in assessing suicide risk, considering an individual's history, mental health, substance use, and family background. Fine-tuned transformer-based models effectively identified sui-

cidal behaviors from EHRs, with Bio_ClinicalBERT achieving the highest accuracy (0.82) and F1 scores (0.82), outperforming general domain transformers. Bio_ClinicalBERT showed strong performance in identifying Family History and Non-Suicidal behaviors, with some challenges in distinguishing between Active and Passive Suicidal Ideation. The proposed approach, using task-specific NLP models and a multi-label classification system, captures the complexity of suicidal behaviors more effectively than traditional binary classification. However, direct comparisons with existing studies are difficult due to varying metrics and label definitions.

Conclusion: This study presents a robust NLP framework for detecting suicidal behaviors in EHRs, leveraging task-specific fine-tuning of transformer-based models and a semi-automated pipeline. Despite limitations, the approach demonstrates the potential of advanced NLP techniques in enhancing the identification of suicidal behaviors. Future work should focus on model expansion and integration to further improve patient care and clinical decision-making.

Keywords: Suicidal behaviors, Natural Language Processing (NLP), Transformer-based language models, Electronic Health Records (EHRs), Multi-Label classification, Mental health informatics, Deep learning

1. Introduction & Background

Suicide is a major public health issue and the 11th leading cause of death in the United States [1, 2, 3]. Despite the critical need to identify individuals at risk, accurately assessing suicidal behaviors remains a significant challenge [4, 5]. The stigma surrounding mental health likely contributes to the under-reporting of the nearly one million suicide and suicidal behavior events officially reported each year [4, 5]. These behaviors are often documented in unstructured clinical notes within Electronic Health Records (EHRs), rendering diagnostic codes alone inadequate for capturing the full scope of suicidal behaviors [6]. Identifying individuals at risk of suicide in clinical settings, such as primary care or emergency departments (ED), has been associated with better linkage to treatment, which can result in reduced suicide morbidity and mortality in some populations. Despite these benefits, current evidence suggests that EHR identification remains understudied as a potential screening replacement [7]. Manual review of EHRs is time-consuming, error-prone, irreproducible, and impractical for large datasets, and is susceptible to bias due to varying reviewer expertise and inconsistent adherence to guidelines. Natural Language Processing (NLP) techniques offer a robust and reproducible solution for extracting information on suicidality from clinical notes, standardizing the identification of suicidal behaviors, reducing bias, and enhancing the accuracy of suicide risk assessment in EHRs [8, 9]. However, existing NLP approaches heavily rely on keyword searches and rule-based methods, which have limitations in capturing complex contextual patterns. Moreover, the resource-intensive nature of creating large-scale annotated training data for suicidality detection poses challenges for developing comprehensive NLP models. In this work, we propose a semi-automatic approach that leverages transformer-based language models to address these limitations, while introducing a multi-class labeling system for identifying a spectrum of suicidal behaviors. We utilize the publicly available “Medical Information Mart for Intensive Care (MIMIC)” dataset to foster reproducibility and transparency in our research.

1.1. Moving beyond keyword search, while reducing manual annotation effort

To accurately identify and extract suicidal behaviors from EHRs, NLP methods need to go beyond the keyword and phrase searches that rule-based approaches in existing studies heavily rely on. While these methods are

effective in finding exact matches, they have significant limitations in capturing complex contextual patterns not explicitly stated using those keywords, and this limitation can lead to critical information being overlooked [6, 10, 11, 12, 13, 14].

Suicidal behaviors, particularly suicidal ideation, are often not explicitly documented using standard terminologies [15]. Instead, they can be expressed in diverse ways within the free text, without specific keywords or phrases [16]. For instance, patients may exhibit suicidality by discussing feelings of hopelessness, worthlessness, or a desire to “disappear forever” without directly mentioning suicide. Examples from EHRs include: “Patient clarifies that because of the suffering in the world, perhaps life is not worth living” or “Patient finds life hard every day and at times prefers not to be alive”. Keyword search methods likely miss these key instances of ideation, failing to capture implicit concepts. Consequently, keyword searches may have high precision but low recall, missing these implicit references. [11] study resulted in an 88.5% sensitivity (recall) and a 100.0% Positive Predictive Value (PPV, AKA. Precision) for current suicidal ideation/attempts, and a 100.0% sensitivity and PPV for historical suicidal ideation/attempts.

On the other hand, the limited availability of extensive training samples poses a challenge for complex semantic-based NLP methods in detecting suicidal behaviors. Creating large-scale annotated training data is resource-intensive, as suicidal behavior documentation is rare in patient records, making it impractical to annotate thousands of notes. For instance, Fernandes et al. (2018) [10] manually annotated 500 documents from a psychiatric research database [17] to identify suicide ideation and attempts. Similarly, Zhong et al. (2019) [18] thoroughly reviewed and annotated 200 patients’ medical records, including 40 notes with a suicide-related diagnostic code and 160 without, resulting in 9,341 annotations to train a model for identifying suicidal behaviors in pregnant women. Approximately 73% of the 40 notes and 28% of the 160 notes were confirmed to have suicidal behaviors. This study could achieve a sensitivity of 58%, Predictive Positive Value (PPV) of 63%, Negative Predictive Value (NPV) of 88%, and AUC of 83% [19]. These examples highlight the extensive annotation required to develop effective models for recognizing such sensitive behaviors, given the limited training data. Therefore, balancing model complexity and available data is crucial for effective model development, especially when dealing with the diverse and often implicit expressions of suicidal behaviors mentioned earlier.

To address the limitations of term searches and rule-based approaches in

identifying suicidal behaviors in EHRs, as well as the infeasibility of large-scale annotation work, we propose a semi-automatic approach that leverages transformer-based language models and their ability to capture semantic relevance and contextual information. This mechanism aims to enhance the accuracy and comprehensiveness of behavioral pattern detection while reducing the annotation effort required for training NLP models. In contrast to conventional term searches, the proposed model reveals hidden clinical insights on suicidal behaviors by selectively extracting clinically rich and semantically relevant excerpts, providing a deep understanding without the need for extensive note review and labeling. The extracted sentences assist experts in annotating the data more effectively with minimal effort, thereby accelerating data curation and building the NLP model more efficiently than traditional rule-based methods.

1.2. Enhancing suicidal behavior classification with multi-class labeling

Existing NLP models for identifying suicidal behaviors often rely on simple, binary classification labels, which may fail to capture the intricacies and details of suicidal thoughts and behaviors in the text [6, 10, 11, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. Rawat et al. [31] reported Precision, Recall, and F1-Score of 87%, 89%, and 88%, respectively, for predicting any evidence of suicidal ideation or behaviors in the text. For predicting suicide attempts and suicidal ideation specifically, they achieved a Precision of 62% and 61%, a Recall of 64% and 69%, and an F1-Score of 63% and 64%, respectively. To address these limitations, our research introduces a multi-class labeling system that identifies a spectrum of suicidal behaviors. This system adopts a granular approach to categorization, distinguishing between six well-defined categories of suicidal thoughts and behaviors as defined by the Food and Drug Administration (FDA) [32], Centers for Disease Control and Prevention (CDC) [33], and World Health Organization (WHO) [34]. These categories include: “Suicide Attempt”, “Passive Suicidal Ideation”, “Active Suicidal Ideation”, “non-suicidal self-injurious behaviors”, “Family History”, and “Non-suicidal”, along with a category for marking sentences as “irrelevant”. Table 2 presents the definitions and illustrative examples for each label, as detailed in Section 2.2.2. Curating our annotated datasets using well-established guidelines and labels from trusted sources enhances their reliability, quality, and analytical depth within our NLP framework.

Recognizing the dynamic nature of suicidal behaviors is crucial. Individuals who have attempted suicide may shift away from a suicidal state over

time, while patients with strong suicidal thoughts might later exhibit milder ideation or non-suicidal tendencies (e.g., “Patient states he has attempted to take his life in the past but is currently without plan or intent”) [35, 36]. A single patient’s record can contain documentation of various stages of suicidality, including past suicide attempts, passive or active suicidal thoughts, non-suicidal periods, and other relevant information, making it complex to accurately capture suicide risk in EHR clinical notes. Our work employs a “Sentence-Level Multi-Label Classification (MLC)” approach that diverges from conventional single-label classifications for EHRs [4, 8, 9, 10, 31]. This approach adopts a “sigmoid” loss function that treats each class independently and allows the assignment of multiple, non-mutually exclusive labels to each data sample (Table 1). The MLC approach enhances the accuracy and comprehensiveness of capturing the nuances of suicidal behavior characteristics expressed across EHRs [37], addressing the complexities of analyzing real-world documentation where a single sample may exhibit multiple facets of suicidality.

1.3. Fostering public data for reproducibility and transparency

In this work, we utilize the publicly available MIMIC dataset. Unlike studies reliant on smaller datasets [38] or private data sources [27, 39, 40], this rich dataset provides valuable medical information to propel research forward. Open data availability offers numerous advantages, including enabling transparency, reproducibility, and validation of findings [41], enhancing community engagement and collaboration, improving efficiency, and building trust [42, 43]. The annotated dataset and trained models will be available on PhysioNet, subject to required data usage training and agreements. The pre-processing scripts, annotation schema, and NLP codes will be publicly shared on GitHub.

1.4. Statement of Significance

1.4.1. Problem

Accurate identification of suicidal behaviors in Electronic Health Records (EHRs) is crucial for timely intervention, but the complex nature of suicidal expressions in clinical notes poses significant challenges.

1.4.2. What is Already Known

Traditional approaches, such as keyword searches and binary classifiers, often fail to capture the semantic nuances and spectrum of suicidal behaviors

in EHRs, leading to an oversimplified representation of the problem. Creating large-scale annotated training data for complex semantic-based NLP methods is resource-intensive, as suicidal behavior documentation is rare in patient records. Recent advancements in NLP, particularly transformer-based models, have shown promise in improving detection, but annotation effort remains a significant challenge.

1.4.3. What This Paper Adds:

This study presents a novel framework that leverages transformer-based models to selectively extract clinically relevant sentences from EHRs, optimizing the annotation process and enabling the development of a multi-label classification system. Our approach, guided by expert-curated guidelines from the FDA, CDC, and WHO, allows for a more granular understanding of suicidal behaviors, moving beyond the limitations of binary labels. By addressing the challenges of existing methods, this framework demonstrates significant improvements in both the efficiency of the annotation process and the accuracy of suicidal behavior detection, ultimately facilitating more targeted interventions and enhanced patient care.

Table 1: Label definitions, the source of definitions, and illustrative examples

Label	Definition	Source of definition	Examples
Suicidal attempt	A potentially self-harming act, indicating at least some intention to cause death consequently. The evidence of the individual's intent to end their own life, to some extent, can be either explicit or deduced from their behavior or situation. A suicide attempt may or may not lead to actual harm. [32, 33]	FDA and CDC [32, 33]	"Patient reports one prior suicide attempt by overdosing in ____," "Reports multiple suicide attempts, by wrist cutting and intentional OD on medications."
Passive suicidal ideation	This refers to a desire to be dead. It involves thoughts of not wanting to be alive anymore or wishing to fall asleep and not wake up. [32]	FDA [32]	"She has significant passive suicidal ideation, though no active plan." "Patient was without plans to hurt himself or others, but verbalized the idea that it might be of benefit that himself and others were dead."
Active suicidal ideation	Active suicidal ideation refers to thoughts of ending one's life that vary in intensity: from nonspecific thoughts without method, intent, or plan, to specific thoughts that include method, intent, and detailed plans. [10]	FDA [32]	"She reports that she had been planning to kill herself via overdose on Klonopin for days." "Thoughts of jumping off a bridge or cutting self with a knife"
Non-suicidal	In the absence of suicidal ideation or self-injurious behavior, it can be inferred that the patient is non-suicidal. To annotate a sentence as non-suicidal, we looked for a strong indication of negation of suicidal ideation and self-injurious behavior [32].	FDA [32]	"Psychiatric: Patient was not having suicidal thoughts on admission." "She also denies suicidal ideation and says that she wants to live but repeatedly says that she feels depressed."
Non-suicidal Self Injurious behaviors (NSSI)	Self-injurious behavior without any intention to cause death is performed for reasons other than ending one's life. This behavior serves different purposes, such as relieving distress (referred to as self-mutilation) through actions like superficial cuts, scratches, hitting, banging, or burns. Additionally, it may also be aimed at influencing others or the environment to bring about a change. [32]	FDA [32]	"The patient reports feeling overwhelmed and stressed for the past several months. They describe using physical pain as a method to cope with emotional distress."
Family History	A family history of suicide refers to the occurrence of suicide or suicidal behaviors among biological relatives of an individual. This includes completed suicides, suicide attempts, or any documented behaviors indicating suicidal thoughts within the family. [44] This history is considered a significant risk factor for suicide, as genetic and environmental factors that contribute to suicidal behaviors can be familial. [45]	[44]	"The patient reported one brother who committed suicide by jumping in front of a bus x years ago."
Irrelevant	The "Irrelevant" category in our annotation refers to clinical notes that are unrelated to suicide or any defined categories. These notes cover different medical topics, treatments, or procedures, not relevant to the study's focus. By excluding irrelevant notes, we maintain specificity and accuracy in analyzing suicidal behavior, mental health, and related aspects of interest. Annotators ensured consistent labeling to ensure the quality of data used for training and evaluation		"Clozapine 25 mg PO QAM" "She was given ASA 324mg, SL NTG and morphine 2mg IV x3 for pain, ativan 0.5mg IV x1."

2. Materials & Methodology

2.1. Dataset Creation: Initial Cohort Sampling and Preprocessing of EHRs

We utilized the MIMIC-IV critical care database [46], a large, open-access resource containing de-identified health data from over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2008-2019 [47]. The dataset includes a wide range of information such as demographics, vital signs, laboratory tests, medications, and clinical notes. This study was approved by the University of Florida Institutional Review Board (IRB20220219) to ensure ethical data use. We used the International Classification of Diseases (ICD) 9 and 10 codes to extract notes from 100 patients, comprising 80 patients with suicide-related ICD codes as cases and 20 patients without such codes as the control group. The notes were segmented into sentences using PySBD [48], a clinical text segmentation tool known for its accuracy and efficiency, resulting in a total of 48,935 sentences.

2.1.1. Similarity scoring approach for efficient annotation sample identification

As discussed in subsection 1.1, manual annotation of clinical notes for suicidal behaviors is resource-intensive due to the sparse documentation in EHRs [49]. To expedite annotation, we propose a semi-systematic approach that identifies and extracts sentences from MIMIC-IV notes that are semantically similar to a small, comprehensive list of relevant phrases. This approach recognizes that the documentation of suicidal behavior and ideation may vary across patients, but assumes that the semantic meaning remains consistent despite differences in phrasing. By precisely selecting relevant samples from a large number of EHR notes, our method significantly simplifies annotation efforts.

We developed six distinct sets of comprehensive semantic search patterns, each corresponding to one of the six relevant categories of suicidal thoughts and behaviors. Each set contains 10-20 carefully selected words or phrases that encapsulate the diverse range of expressions within its respective category. These search patterns ensure that we capture relevant information from the clinical notes in our suicide dataset. To ensure the reliability of our search patterns, we used validated instruments such as the “Columbia-Suicide Severity Rating Scale” (C-SSRS) [50], the “Suicide Ideation Questionnaire” (SIQ) [51], and the “Beck Scale for Suicide Ideation” (BSSI) [52]. These

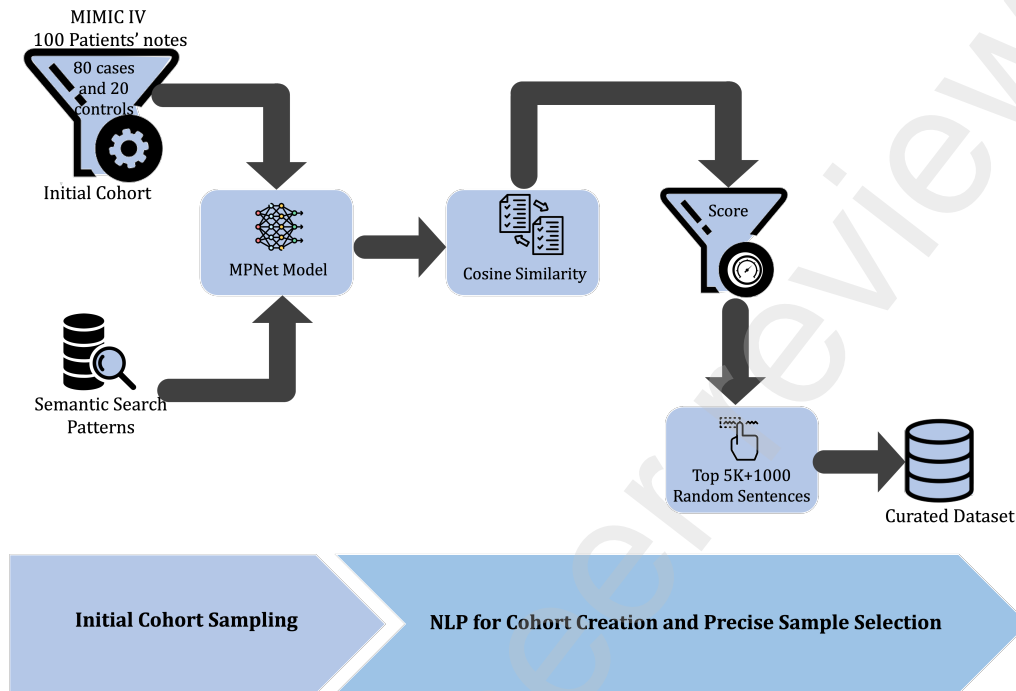


Figure 1: Graphical outline of the cohort creation and precise sample selection

rigorously designed and validated instruments provide a robust foundation for generating search patterns that accurately reflect clinical realities.

We then developed a semantic retrieval framework using the MPNet (Masked and Permuted Pre-training for Language Understanding) model [53] to efficiently extract relevant sentences from EHRs within our suicide dataset and create a customized dataset. MPNet, a mini-BERT model, plays a pivotal role by encoding both search patterns and sentences into dense vector representations. By employing MPNet embeddings, we can accurately identify sentences in the notes that exhibit semantic similarity to our search patterns. We quantify semantic similarity using the cosine similarity metric, comparing vector pairs across embeddings (search patterns versus sentences), ranking them based on the similarity score, and keeping only the top 5000 sentences. This approach facilitates the extraction of sentences pertinent to each of the six relevant categories, while sentences not belonging to the top 5000 sentences are considered irrelevant. To ensure a comprehensive dataset, we supplemented the extracted sentences with an additional 1,000 randomly selected sentences from our suicide dataset, enhancing the sensitivity and

specificity of our framework. We also ensured that the extracted sentences were unique per patient, maintaining the integrity and diversity of the data.

2.2. Manual Annotation and Gold-Standard Annotated Dataset Creation

2.2.1. Beyond Binary Labels: A Granular Labeling Approach for Suicidal Behaviors

Suicidal ideation and behaviors are not simply binary distinctions between suicide attempts and ideation; they are much more intricate. These behaviors are influenced by various factors. Recognizing the limitations of categorizing them with just a few labels, and realizing that this approach can result in a loss of valuable information about suicidal behavior, we have introduced a more nuanced system as mentioned in subsection 1.2. This system defines six distinct relevant categories and one irrelevant category: “Suicide attempt”, “Active suicidal ideation”, “Passive suicidal ideation”, “Non-suicidal”, “Non-suicidal Self Injurious behaviors (NSSI)”, “Family history”, and “Irrelevant”.

As discussed earlier, we utilized a sentence-level “Multi-Label Classification” (MLC) approach to address the presence of different aspects of risk factors within a single piece of information. MLC allows the assignment of multiple, non-mutually exclusive labels to each data sample, enabling a deeper understanding of complex suicidal ideation and behaviors. Table 2 shows some examples of annotated sentences with multiple labels.

2.2.2. The Annotation Schema for Labeling Extracted Sentences

To establish a reliable and precise annotation schema for labeling the sentences extracted from our suicide dataset using the semantic retrieval framework, we conducted an extensive review of reputable references and guidelines on suicidal ideation and behaviors. We relied on resources from well-known organizations including: (“World Health Organization” (WHO)[54], “Centers for Disease Control and Prevention” (CDC)[55], “United States Food and Drug Administration” (FDA) [56], as well as studies published in reputable journals. These sources provided a robust foundation for developing our annotation schema, ensuring accuracy and consistency in our work. The written schema, relevant examples, and expert training were offered to annotators to ensure consistent and accurate annotations. This schema employs clearly defined labels to provide reliable and high-quality annotations for the extracted sentences. Table 1 describes the labels, corresponding definitions derived from source materials, and illustrative examples per category. Our annotation strategy, which incorporates domain guidelines and structured

Table 2: Examples of Annotated Sentences with Multiple Labels

Sentence	Labels
On presentation, the patient was evaluated and felt to be at an increased risk of harm to himself based upon thoughts of suicide.	Active suicidal ideation Self-harm risk assessment
Psychiatric: The patient denied SI, denied intent or plan for self-harm, but admitted that during stressful periods she wondered “what it would be like to not exist.” She notes recent improvements in her life, such as working again and having a stable living situation.	Non-suicidal Self-harm risk assessment
He reportedly expressed suicidal ideation with a plan to overdose on Valium and Percocet and also stated that he had committed suicide in the past.	Active suicidal ideation Suicide attempt

training, facilitated precise categorization of the extracted sentences related to suicidal behavior.

2.2.3. Manual Sentence Annotation Process

targeting the top 5,000 sentences and an additional 1,000 random sentences, totaling 6,000 sentences, extracted from 260 notes of 100 patients within our suicide dataset using the semantic retrieval framework ranked by cosine similarity involved two rounds of manual labeling by two trained annotators per round. Given the complex and multifaceted nature of suicide, individual sentences may encompass multiple aspects. Our trained annotators assigned all relevant labels to capture the full context of each extracted sentence accurately, ensuring complete categorization even when various factors intertwine to contribute to suicidal ideation and behaviors.

In the initial round, we randomly selected 150 sentences from the over 6,000 extracted sentences for double review, ensuring each received two independent assessments. Analysis of Inter-Rater Reliability (IRR) [57] showed an overall agreement of 93.3%. Differences in assessments were resolved

Table 3: The frequency and percentage of occurrence of each label in relevant sentences

Label	# of occurrence	% of occurrence
Suicide attempt	163	13.63
Active suicidal ideation	481	40.22
Passive suicidal ideation	108	9.03
Non-suicidal	341	28.51
NSSI	155	12.96
Family history	39	0.33

through discussions, emphasizing rationales and adherence to guidelines, which resulted in mutually agreed-upon labels. This collaborative approach resulted in 100% agreement on the first 150 extracted sentences and allowed us to refine the annotation schema.

In the second annotation round, 350 new sentences randomly selected from the remaining 5,850 extracted sentences underwent double review, bringing the total to 500 sentences across both rounds. The IRR analysis for this round revealed an even higher agreement level of 97.71%, demonstrating improved consistency through refined guidelines and effective resolution of initial discrepancies. As with the first round, any remaining differences were addressed through discussion, ultimately achieving 100% agreement. After the double review process, annotators gained a shared understanding of labels and definitions. We then proceeded to annotate the remaining extracted sentences.

Out of 6,000 annotated sentences, we had a mix of relevant and irrelevant sentences. Specifically, there were 1258 sentences marked as relevant and 4742 sentences marked as irrelevant. Among the relevant sentences, 1056 had only one label, 177 had two labels, and 34 had three or more labels. The following table shows the distribution of 8 distinct labels in relevant sentences:

2.3. Identifying Suicidal Behaviors in EHRs using Transformer-based Models

This study aims to demonstrate the effectiveness of using transformer-based models to identify suicidal behaviors in clinical notes from EHRs. We developed a two-step classification pipeline leveraging an expert-annotated dataset created with a novel MPNet-based approach. The pipeline first distinguishes between suicide-relevant and irrelevant sentences, followed by a

detailed classification of suicidal behavior within the relevant sentences.

2.3.1. Step 1: Classifying Suicide-Relevant and Irrelevant Sentences

In the first step, we employed the Bio_ClinicalBERT model [58] to classify sentences as either suicide-relevant or irrelevant. Bio_ClinicalBERT is a transformer-based model specifically designed for medical text, making it well-suited for this task. The model was trained on 80% of the expert-annotated dataset and tested on the remaining 20%, using basic hyperparameter tuning. This approach proved highly effective, achieving a precision of 0.99, recall of 0.98, and an F1-score of 0.985 in distinguishing between relevant and irrelevant sentences.

2.3.2. Step 2: Detailed Classification of Suicidal Behavior

The second step focuses on the detailed classification of suicidal behavior within the correctly identified relevant sentences from Step 1. We explored five transformer-based models: BERT [59], RoBERTa [60], XLNet [61], Bio_ClinicalBERT [58], and BioBERT [62]. These models offer a deep understanding of language and can be fine-tuned for specific tasks, such as suicidal behavior detection.

To adapt the transformer-based models for our task, we replaced the classification head with a new dense layer for multi-class classification. The models were then fine-tuned end-to-end on the relevant sentences using the AdamW optimizer and cross-entropy loss. During fine-tuning, both the pre-trained weights and the newly initialized classification layer were updated via backpropagation.

Algorithm 1 outlines the proposed approach for identifying detailed suicidal behaviors using transformer-based models.

2.3.3. Computational Environment

The transformer-based models in this work were trained on the University of Florida's HiPerGator supercomputer. HiPerGator is equipped with state-of-the-art processors and specialized nodes designed to handle-intensive deep learning tasks [63]. We used two NVIDIA Ampere A100 80GB GPUs, along with an AMD EPYC processor with 32 cores and 256GB of RAM.

Input : Expert-annotated dataset (extracted clinical notes)

Output: Transformer-based model for suicidal behavior detection in extracted clinical notes

Step 1: Classifying Suicide-Relevant and Irrelevant Sentences

Step 1.1: $Sent_list \leftarrow Input_data$

Step 1.2: $Preprocessed_data \leftarrow do_following$

for i **in** $Sent_list$ **do**

 Tokenize the text

 Convert tokens to token IDs

 Create attention masks

end

Step 1.3: Split Preprocessed data into training (80%) and test sets (20%)

Step 1.4: Initialize Bio_ClinicalBERT model with pre-trained weights

Step 1.5: Fine-tune Bio_ClinicalBERT model on the training set using basic hyperparameter tuning

Step 1.6: Evaluate the model on the test set using precision, recall, and F1-score

Step 1.7: Save the fine-tuned Bio_ClinicalBERT model for identifying suicide-relevant sentences

Step 2: Detailed Classification of Suicidal Behavior

Step 2.1: $Relevant_Sent_list \leftarrow OutputofStep1$

Step 2.2: $Preprocessed_data \leftarrow do_following$

for i **in** $Relevant_Sent_list$ **do**

 Tokenize the text

 Convert tokens to token IDs

 Create attention masks

end

Step 2.3: Split Preprocessed data into training and test sets

Step 2.4: **for** i **in** $Transformer_Models$ ($BERT$, $RoBERTa$, $XLNet$, $BioBERT$, $Bio_ClinicalBERT$) **do**

 Initialize model with pre-trained weights

 Replace the classification head with a new dense layer for multi-class classification

 Fine-tune the model on the training set using AdamW optimizer and cross-entropy loss

 Evaluate the model on the test set using evaluation metrics

 Save the best-performing model for the current transformer architecture

end

Step 2.5: Compare the performance of different transformer models

Step 2.6: Select the best-performing model among all architectures based on test set metrics

Step 2.7: Save the overall best-performing model for detailed suicidal behavior detection.

Algorithm 1: Transformer-based approach for identifying suicidal behaviors in clinical notes.

3. Results and Discussion

3.1. Lexical Analysis of Suicidal Behavior Categories

This lexical analysis examines the most frequent words associated with various suicidal behavior categories in our dataset to provide insights into the language used to describe and assess these behaviors. The word clouds and narratives offer an understanding of the key themes, patterns, and associations that emerge from the text data.

The analysis of suicide attempts (Figure 2-a) reveals a complex interplay of factors, with intentional self-directed violence at the core. Individuals' histories, indicated by words like "past," "prior," and "recent," suggest ongoing struggles. The prominence of "psychiatric," "patient," "hospital," and "discharge" highlights the crucial role of mental healthcare in assessing and treating those at risk, while "alcohol" emerges as a significant correlate. The concepts of "risk," "reports," and "history" emphasize the importance of comprehensive evaluation and understanding of each person's unique situation.

Active suicidal ideation (Figure 2-b) is characterized by the persistent nature of the suicidal mindset, with "suicidal," "ideation," and "thoughts" at the forefront. The presence of "plan" and "intent" suggests a higher level of risk, as individuals may have progressed to formulating a plan of action. The involvement of healthcare systems, including "hospital," "admission," "ed" (emergency department), and "discharge," is critical in managing and treating individuals in crisis. "Risk" and "safety" highlight the urgent need for assessment and intervention to prevent potential harm.

Passive suicidal ideation (Figure 2-c) is characterized by the presence of suicidal thoughts without active intent or planning. The terms "passive," "ideation," and "thoughts" are at the forefront, while "denies" suggests that individuals may initially minimize these thoughts when assessed. The importance of self-reporting and clinical evaluation in identifying passive ideation is indicated by "patient" and "states." While passive ideation lacks concrete plans, the presence of "plan" and "intent" suggests the need to assess the potential for escalation.

The non-suicidal category (Figure 2-d) is characterized by the explicit denial of suicidal thoughts or behaviors, with "denies," "denied," and "si" (suicidal ideation) being prominent. The high frequency of "patient" and "ideation" indicates that these assessments often occur in a clinical setting. While suicidal ideation is absent, the presence of "thoughts," "content," and

“mood” implies a discussion of the individual’s general thought content and emotional state. The inclusion of “harm,” “self,” and “plan” suggests that risk assessments are still conducted to ensure no potential for self-harm.

Non-suicidal self-injury (NSSI) (Figure 2-e) is characterized by deliberate, self-directed behavior involving self-inflicted injuries, often in the form of cutting. The prominence of “self,” “harm,” and “behavior” underscores the intentional nature of these acts, which are performed to cause personal harm without suicidal intent. The presence of “risk,” “assessment,” and “history” highlights the importance of evaluating an individual’s past experiences and current risk factors when assessing NSSI. The inclusion of “denies,” “thoughts,” and “suicide” suggests that while NSSI is distinct from suicidal behavior, there is still a need to assess for suicidal ideation and intent.

Finally, family history (Figure 2-f) plays a significant role in assessing an individual’s suicide risk, with “family,” “history,” and “suicide” being the most frequent words. The presence of “suicides” (plural) hints at the possibility of multiple family members having died by suicide, while “mother” and “past” indicate the importance of maternal history and past events. The relevance of a family history of mental health conditions and substance use disorders is suggested by “mental,” “illness,” “psychiatric,” “disorder,” “substance,” and “abuse.”

In conclusion, the lexical analysis of the most frequent words associated with each suicidal behavior category provides valuable insights into the complex nature of these behaviors, the factors that influence them, and the clinical contexts in which they are assessed and treated. The word clouds and narratives highlight the importance of considering an individual’s history, mental health status, substance use, and family background when evaluating suicide risk, while also underscoring the critical role of healthcare systems in managing and treating individuals experiencing suicidal thoughts or engaging in self-harm behaviors.

3.2. Comparative Analysis of Fine-tuned Transformer-based Models

The comparative analysis of fine-tuned transformer-based models on the test data, as presented in Table 4, demonstrates the strong performance of all models in identifying suicidal behaviors from electronic health records. The weighted average F1 scores ranged from 0.78 to 0.82, indicating the models’ effectiveness in this critical task. Notably, the models excelled in identifying and categorizing crucial suicidal behaviors and risk factors from clinical texts, particularly in the “Suicide Attempt”, “Active Suicidal Ideation”, and



“Family History” categories. Across all models, F1 scores for these categories surpassed 0.76, with Bio_ClinicalBERT achieving the highest F1 score of 0.88 in the “Suicide Attempt” category, outperforming its counterparts.

The models also demonstrated strong performance in the “Passive Suicidal Ideation” and “Non-Suicidal” categories, with F1 scores ranging from 0.67 to 0.84. Bio_ClinicalBERT and BioBERT stood out in these categories, showcasing their proficiency in handling clinical text data related to passive suicidal ideation and non-suicidal behaviors.

When comparing the performance of general domain transformers (BERT, RoBERTa, and XLNet) to biomedical and clinical domain models (BioBERT and Bio_ClinicalBERT), the domain-specific models generally outperformed their general domain counterparts. Bio_ClinicalBERT, in particular, achieved the highest overall accuracy of 0.82 and the highest macro and weighted average F1 scores of 0.82, underlining its superior performance in identifying suicidal behaviors from electronic health records.

However, the models faced some challenges in accurately identifying “Non-Suicidal Self-Injury” (NSSI), with F1 scores ranging from 0.72 to 0.76. This suggests that there is still room for improvement in the models’ ability to differentiate between suicidal and non-suicidal self-injury behaviors.

The fine-tuned transformer-based models demonstrated strong overall performance in identifying suicidal behaviors from electronic health records, with Bio_ClinicalBERT achieving the highest performance across various categories. The domain-specific models, BioBERT and Bio_ClinicalBERT, generally outperformed the general domain transformers, emphasizing the importance of domain adaptation in enhancing the models’ ability to process clinical text data related to suicidal behaviors and risk factors.

3.3. Performance Analysis and Error Patterns

To gain a deeper understanding of our best-performing model, we conducted a detailed performance analysis and error pattern examination of the Bio_ClinicalBERT classifier. The confusion matrix in Figure 3 reveals the classifier’s performance across nine categories related to suicidal behavior and risk assessment.

The model demonstrates strong performance in identifying instances related to Family History (100% accuracy) and Non-Suicidal behaviors (90% accuracy). This suggests that the model has learned to effectively recognize patterns and language specific to these categories, which may have more distinct features compared to the other categories.

Table 4: Comparative analysis of performance of suicidal behavior classification among fine-tuned transformer-based models.

Model	Precision	Recall	F1 Score	Support
BERT				
Suicide Attempt	0.88	0.72	0.79	32
Active Suicidal Ideation	0.83	0.77	0.80	94
Passive Suicidal Ideation	0.68	0.65	0.67	20
Non-Suicidal	0.76	0.86	0.81	63
NSSI	0.68	0.79	0.73	24
Family History	0.75	1.00	0.86	6
Accuracy			0.78	239
Macro Avg	0.76	0.80	0.77	239
Weighted Avg	0.79	0.78	0.78	239
RoBERTa				
Suicide Attempt	0.87	0.81	0.84	32
Active Suicidal Ideation	0.87	0.78	0.82	94
Passive Suicidal Ideation	0.74	0.70	0.72	20
Non-Suicidal	0.75	0.87	0.81	63
NSSI	0.69	0.75	0.72	24
Family History	0.86	1.00	0.92	6
Accuracy			0.80	239
Macro Avg	0.80	0.82	0.80	239
Weighted Avg	0.81	0.80	0.80	239
XLNet				
Suicide Attempt	0.85	0.69	0.76	32
Active Suicidal Ideation	0.82	0.82	0.82	94
Passive Suicidal Ideation	0.68	0.65	0.67	20
Non-Suicidal	0.74	0.84	0.79	63
NSSI	0.81	0.71	0.76	24
Family History	0.86	1.00	0.92	6
Accuracy			0.79	239
Macro Avg	0.79	0.78	0.78	239
Weighted Avg	0.79	0.79	0.79	239
BioBERT				
Suicide Attempt	0.89	0.78	0.83	32
Active Suicidal Ideation	0.83	0.78	0.80	94
Passive Suicidal Ideation	0.71	0.75	0.73	20
Non-Suicidal	0.74	0.83	0.78	63
NSSI	0.72	0.75	0.73	24
Family History	0.86	1.00	0.92	6
Accuracy			0.79	239
Macro Avg	0.79	0.81	0.80	239
Weighted Avg	0.80	0.79	0.79	239
Bio_ClinicalBERT				
Suicide Attempt	0.96	0.81	0.88	32
Active Suicidal Ideation	0.87	0.78	0.82	94
Passive Suicidal Ideation	0.68	0.75	0.71	20
Non-Suicidal	0.79	0.90	0.84	63
NSSI	0.70	0.79	0.75	24
Family History	0.86	1.00	0.92	6
Accuracy			0.82	239
Macro Avg	0.81	0.84	0.82	239
Weighted Avg	0.83	0.82	0.82	239

However, the model shows some difficulty in distinguishing between Active and Passive Suicidal Ideation, with 20% of Passive Suicidal Ideation instances misclassified as Active Suicidal Ideation. This confusion may stem from the conceptual similarity between these two categories, as they both involve thoughts related to suicide. The language used to express these thoughts may have overlapping features, making it challenging for the model to differentiate between them consistently.

Another area of potential confusion is between Suicidal Ideation and Non-Suicidal instances, with 11% of Suicidal Ideation instances misclassified as Non-Suicidal. This may be due to the presence of ambiguous or non-explicit language in some Suicidal Ideation instances, leading the model to interpret them as Non-Suicidal. Additionally, there is a small percentage of misclassification between Suicide Attempt and NSSI (3.1% in each direction), possibly due to shared language patterns related to self-harm behaviors.

To improve the model's performance, future work could focus on refining the training data to include more instances that highlight the subtle differences between easily confused categories, such as Active and Passive Suicidal Ideation. Furthermore, incorporating domain-specific knowledge and expert input in the data annotation process may help to create more robust and distinguishable features for the model to learn from. Despite these challenges, the Bio.ClinicalBERT classifier demonstrates promising results in automating the categorization of suicide-related content, which could support mental health professionals in risk assessment and intervention efforts.

3.4. Comparison to Existing Work

Comparing our results with existing studies on predicting suicidal ideation and behaviors is challenging due to the varying metrics reported and the differences in label definitions. While some studies focus on Precision, Recall, and F1-Score [31], others report Negative Predictive Value (NPV), Positive Predictive Value (PPV), and Area Under the Curve (AUC) [19]. Additionally, our study includes six labels for suicidal behavior and one irrelevant label, whereas most other studies typically use only two to four labels, making direct comparisons difficult.

Despite these challenges, our work demonstrates the effectiveness of applying NLP and ML methods to identify and categorize suicidal behaviors in EHRs. Our models achieved strong performance, with weighted average F1 scores ranging from 0.78 to 0.82. The ClinicalBERT model, which was trained specifically on medical texts, demonstrated superior performance

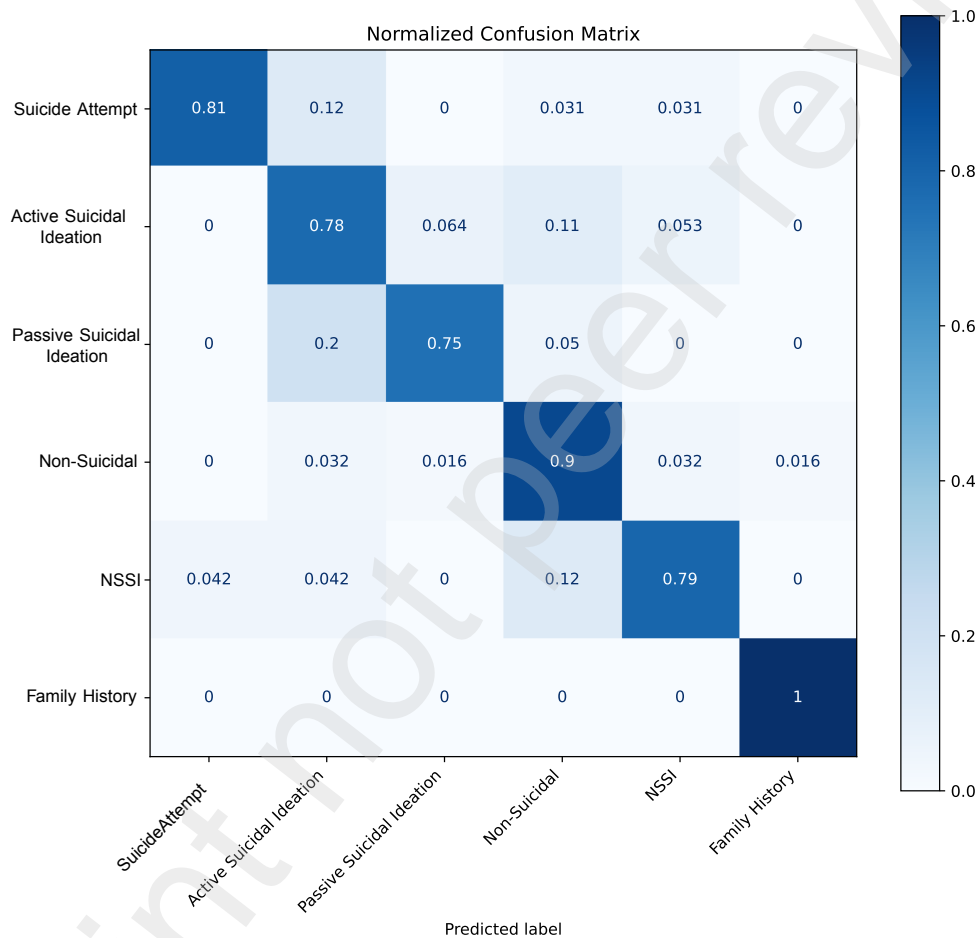


Figure 3: Confusion matrix for the Bio_ClinicalBERT classifier, illustrating the model's performance across six suicidal behavior categories.

compared to general domain transformers, highlighting the importance of domain-specific training in detecting suicidal behaviors from EHRs.

Our approach addresses the limitations of keyword search and rule-based methods [6, 10, 11, 12, 13, 14] for EHR analysis by proposing task-specific, fine-tuned NLP language models that excel at understanding context and language nuances. This is crucial when interpreting the everyday language used to express suicidal behaviors in EHRs. The creation of a semi-automated data pipeline that generates a unique dataset for suicidal behavior identification in EHRs is a key achievement of this work, as it enriches the quality of the dataset while reducing manual annotation efforts.

Moreover, our decision to use a multi-label classification system encompassing six distinct categories for suicidal behavior and one irrelevant label ensures a more precise understanding of suicidal behaviors compared to traditional binary classification methods. This system captures the detailed and varied nature of suicidal behaviors, allowing for more accurate identification and classification compared to existing studies [6, 10, 11, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. The use of these labels, based on trustworthy references and resources from the FDA, CDC, and WHO, enhances the reliability and quality of our annotated datasets, providing a deeper analytical depth within our NLP framework.

In summary, while direct comparisons with existing studies are challenging due to varying metrics and label definitions, our work demonstrates the effectiveness of applying domain-specific NLP models to detect and categorize suicidal behaviors in EHRs. The proposed multi-label classification system with six suicidal behavior labels and one irrelevant label, along with the semi-automated data pipeline, represent significant advancements in capturing the complex nature of suicidal behaviors, ultimately aiding in early intervention and patient support.

3.5. Clinical Implications

The tool developed in this study has valuable applications in clinical practice for assessing and managing patients with suicidal behavior. By automatically classifying clinical notes based on suicidal behavior categories, it identifies high-risk patients, expediting early interventions and saving time for clinicians and caregivers. Additionally, it provides insights into the complexity of suicidal behavior; however, this tool should be used to complement, not replace, clinical judgment. Integrating this tool could enhance suicide

risk assessment, utilizing natural language processing and machine learning for targeted interventions and improved patient care.

3.6. Limitations

Despite the advancements made in this study, there are several limitations to consider. First, the data used were from a single health system, which might limit the generalizability of our findings to other settings. Additionally, variations in clinical documentation by clinical setting and provider types could affect the model's performance. For instance, a visit to the emergency department for an unrelated injury or ailment is unlikely to contain information relevant to detecting suicidal behavior risk, whereas a visit to primary care for a specific mental health concern may have a wealth of useful information about potential suicidal behavior risk that can be flagged for other clinicians. Second, the lack of consideration for the temporal aspect of suicidal behavior is a notable limitation. We acknowledge that some patients may have had suicidal ideation in the past but were not suicidal at the time of note collection. Future research should consider the temporal context and incorporate period assertion to enhance the accuracy of identifying individuals with current suicidal behavior.

3.7. Future Directions

Moving forward, it is essential to continue exploring the integration of advanced NLP and machine learning techniques in clinical practice. Future works can focus on applying patient-level classification of suicidal behaviors, integrating structured data forms, and creating computable phenotypes to support reliable clinical research. Expanding our model to include additional datasets and healthcare settings will help validate and refine its performance, ensuring its applicability across diverse clinical environments.

This work creates opportunities for future advancements in several areas. First, testing the generalizability of the task-specific fine-tuning methodology on other clinical NLP tasks, such as diagnosis prediction and adverse event detection, could further establish its utility. Second, developing dynamic risk assessment models that consider changes in language patterns and behavioral indicators over time could enable the identification of evolving risk factors and facilitate proactive interventions to prevent escalation to critical levels. Finally, incorporating explainable AI techniques into NLP models can enhance interpretability and transparency, providing clinicians with insights into the model's decision-making process.

4. Conclusion

This work presents a robust NLP framework for detecting suicidal behaviors in Electronic Health Records (EHRs). Task-specific fine-tuning of NLP transformer-based language models has improved the accuracy and sensitivity of predicting suicidal behavior in EHRs, demonstrating the effectiveness of transfer learning and domain-specific fine-tuning in enhancing NLP model performance for identifying suicidal behaviors in clinical text data.

The proposed semi-automated NLP pipeline for sentence extraction from unstructured EHRs enables the creation of a high-quality dataset, facilitating a more granular understanding of suicidal behaviors within EHRs and the subsequent building of NLP models. This approach improves health-care decisions and emphasizes the importance of customized NLP models for valuable insights from EHRs.

Despite the limitations of this study, such as the use of data from a single health system and the lack of consideration for the temporal aspect of suicidal behavior, this work demonstrates the potential of advanced NLP techniques in enhancing the identification and understanding of suicidal behaviors in EHRs. Future works should focus on expanding the model to include additional datasets, integrating structured data forms, and developing dynamic risk assessment models.

In conclusion, this work highlights the potential of task-specific fine-tuning of transformer-based language models in improving the detection of suicidal behaviors in EHRs. With further advancements and validation, this approach can significantly contribute to enhancing patient care and supporting clinical decision-making in the context of suicide prevention and mental health management.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the University of Florida Department of Pharmaceutical Outcomes Policy, AI in the Health Sciences Initiative, College of Pharmacy. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST

None.

FUNDING SOURCES

This work was supported by the National Institute on Aging, a part of the National Institutes of Health, under award number P30AG066506, and by internal funding from the 1Florida Alzheimer's Disease Research Center.

DATA AVAILABILITY STATEMENT Although the dataset is publicly available, accessing it requires the completion of CITI Data or Specimens Only Research and consent to the Data User Agreement, as mandated by PhysioNet's data sharing rules. Should you need additional information or assistance with the dataset, please reach out to the corresponding authors of this paper.

References

- [1] American Foundation for Suicide Prevention, [Suicide statistics](#), accessed: 2024-05-06 (2024).
URL <https://afsp.org/suicide-statistics/>
- [2] National Institute of Mental Health, [Suicide](#), accessed: 2023-09-13 (2023).
URL <https://www.nimh.nih.gov/health/statistics/suicide>
- [3] Centers for Disease Control and Prevention, [Web-based injury statistics query and reporting system \(wisqars\)](#), accessed: 2023-09-13 (2023).
URL <https://www.cdc.gov/injury/wisqars/index.html>
- [4] A. Arowosegbe, T. Oyelade, Application of natural language processing (nlp) in detecting and preventing suicide ideation: A systematic review, *International Journal of Environmental Research and Public Health* 20 (2) (2023) 1514.
- [5] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, S. Y. Ahn, et al., Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010, *The lancet* 380 (9859) (2012) 2095–2128.
- [6] C. A. Bejan, M. Ripperger, D. Wilimitis, R. Ahmed, J. Kang, K. Robinson, T. J. Morley, D. M. Ruderfer, C. G. Walsh, Improving ascertainment of suicidal ideation and suicide attempt with natural language processing, *Scientific reports* 12 (1) (2022) 15146.
- [7] J. J. Mann, C. A. Michel, R. P. Auerbach, Improving suicide prevention through evidence-based strategies: a systematic review, *Focus* 21 (2) (2023) 182–196.
- [8] P. G. Shekelle, P. J. Pronovost, R. M. Wachter, S. L. Taylor, S. M. Dy, R. Foy, S. Hempel, K. M. McDonald, J. Ovretveit, L. V. Rubenstein, et al., Advancing the science of patient safety (2011).
- [9] M. Vassar, H. Matthew, The retrospective chart review: important methodological considerations, *Journal of educational evaluation for health professions* 10 (2013).

- [10] A. C. Fernandes, R. Dutta, S. Velupillai, J. Sanyal, R. Stewart, D. Chandran, Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing, *Scientific reports* 8 (1) (2018) 7426.
- [11] F. Xie, D. S. L. Grant, J. Chang, B. I. Amundsen, R. C. Hechter, Identifying suicidal ideation and attempt from clinical notes within a large integrated health care system, *The Permanente Journal* 26 (1) (2022) 85.
- [12] E. J. Diniz, J. E. Fontenele, A. C. de Oliveira, V. H. Bastos, S. Teixeira, R. L. Rabêlo, D. B. Calçada, R. M. Dos Santos, A. K. de Oliveira, A. S. Teles, Boamente: A natural language processing-based digital phenotyping tool for smart monitoring of suicidal ideation, in: *Healthcare*, Vol. 10, MDPI, 2022, p. 698.
- [13] N. J. Carson, B. Mullin, M. J. Sanchez, F. Lu, K. Yang, M. Menezes, B. L. Cook, Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records, *PloS one* 14 (2) (2019) e0211116.
- [14] C. Cliffe, A. Seyedsalehi, K. Vardavoulia, A. Bittar, S. Velupillai, H. Shetty, U. Schmidt, R. Dutta, Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders: a retrospective cohort study, *BMJ open* 11 (12) (2021) e053808.
- [15] A. J. Millner, T. M. Augenstein, K. H. Visser, K. Gallagher, G. A. Vergara, E. J. D'Angelo, M. K. Nock, Implicit cognitions as a behavioral marker of suicide attempts in adolescents, *Archives of Suicide Research* 23 (1) (2019) 47–63.
- [16] National Institute of Mental Health, [Suicide FAQ](https://www.nimh.nih.gov/health/publications/suicide-faq), accessed: September 4, 2023 (Year of publication, e.g., 2021).
URL <https://www.nimh.nih.gov/health/publications/suicide-faq>
- [17] Himmelfarb Health Sciences Library, [Psychiatry databases](https://guides.himmelfarb.gwu.edu/Psychiatry/databases), retrieved from <https://guides.himmelfarb.gwu.edu/Psychiatry/databases>

(n.d.).

URL <https://guides.himmelfarb.gwu.edu/Psychiatry/databases>

- [18] Q.-Y. Zhong, L. P. Mittal, M. D. Nathan, K. M. Brown, D. Knudson González, T. Cai, S. Finan, B. Gelaye, P. Avillach, J. W. Smoller, et al., Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem, *European journal of epidemiology* 34 (2019) 153–162.
- [19] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, S. Fidler, Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, in: *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27.
- [20] A. E. Aladağ, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu, H. O. Bingol, Detecting suicidal ideation on forums: proof-of-concept study, *Journal of medical Internet research* 20 (6) (2018) e9840.
- [21] S. Ji, C. P. Yu, S.-f. Fung, S. Pan, G. Long, Supervised learning for suicidal ideation detection in online user content, *Complexity* 2018 (2018).
- [22] K. Nikhileswar, D. Vishal, L. Sphoorthi, S. Fathimabi, Suicide ideation detection in social media forums, in: *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, 2021, pp. 1741–1747.
- [23] R. Haque, N. Islam, M. Islam, M. M. Ahsan, A comparative analysis on suicidal ideation detection using nlp, machine, and deep learning, *Technologies* 10 (3) (2022) 57.
- [24] M. Mulholland, J. Quinn, Suicidal tendencies: The automatic classification of suicidal and non-suicidal lyricists using nlp, in: *Proceedings of the sixth international joint conference on natural language processing*, 2013, pp. 680–684.
- [25] S. T. Rabani, Q. R. Khan, A. M. U. D. Khanday, Detection of suicidal ideation on twitter using machine learning & ensemble approaches, *Baghdad science journal* 17 (4) (2020) 1328–1328.

- [26] T. E. Workman, J. L. Goulet, C. A. Brandt, A. R. Warren, J. Eleazer, M. Skanderson, L. Lindemann, J. R. Blossnich, J. O. Leary, Q. Z. Treitler, Leveraging contextual relatedness to identify suicide documentation in clinical notes through zero shot learning, arXiv preprint arXiv:2301.03531 (2023).
- [27] K. Haerian, H. Salmasian, C. Friedman, Methods for identifying suicide or suicidal ideation in ehrs, in: AMIA annual symposium proceedings, Vol. 2012, American Medical Informatics Association, 2012, p. 1244.
- [28] A. Haque, V. Reddi, T. Giallanza, Deep learning for suicide and depression identification with unsupervised label correction, in: Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part V 30, Springer, 2021, pp. 436–447.
- [29] R. Rhenaldy, L. S. Karenza, H. Hidayaturrahman, M. K. Ario, Classification between suicidal ideation and depression through natural language processing using recurrent neural network, Indonesian Journal of Artificial Intelligence and Data Mining 5 (2) (2022) 76–82.
- [30] M. M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of suicide ideation in social media forums using deep learning, Algorithms 13 (1) (2019) 7.
- [31] B. P. S. Rawat, S. Kovaly, W. R. Pigeon, H. Yu, Scan: suicide attempt and ideation events dataset, in: Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting, Vol. 2022, NIH Public Access, 2022, p. 1029.
- [32] Guidance for industry: Suicidal ideation and behavior - prospective assessment of occurrence in clinical trials, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-suicidal-ideation-and-behavior-prospective-assessment-occurrence-clinical-trials>, accessed on 25th August 2023.
- [33] National hospital ambulatory medical care survey: 2018 emergency department summary, <https://www.cdc.gov/nchs/data/nhsr/nhsr108.pdf>, accessed on 25th August 2023.

- [34] Mental disorders, <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>, accessed on 22nd May 2023.
- [35] A. N. Stover, I. R. Rockett, G. S. Smith, T. LeMasters, V. G. Scott, K. M. Kelly, E. L. Winstanley, Distinguishing clinical factors associated with unintentional overdose, suicidal ideation, and attempted suicide among opioid use disorder in-patients, *Journal of psychiatric research* 153 (2022) 245–253.
- [36] J. J. Lastra-Díaz, J. Goikoetxea, M. A. H. Taieb, A. García-Serrano, M. B. Aouicha, E. Agirre, A reproducible survey on word embeddings and ontology-based methods for word similarity: Linear combinations outperform the state of the art, *Engineering Applications of Artificial Intelligence* 85 (2019) 645–665.
- [37] T. Gonçalves, P. Quaresma, A preliminary approach to the multilabel classification problem of portuguese juridical documents, in: *Portuguese Conference on Artificial Intelligence*, Springer, 2003, pp. 435–444.
- [38] M.-H. Metzger, N. Tvardik, Q. Gicquel, C. Bouvry, E. Poulet, V. Potinet-Pagliaroli, Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a french pilot study, *International journal of methods in psychiatric research* 26 (2) (2017) e1522.
- [39] H. S. Bhat, S. J. Goldman-Mellor, Predicting adolescent suicide attempts with neural networks, *arXiv preprint arXiv:1711.10057* (2017).
- [40] T. Tran, D. Phung, W. Luo, R. Harvey, M. Berk, S. Venkatesh, An integrated framework for suicide risk prediction, in: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1410–1418.
- [41] L. M. Federer, C. W. Belter, D. J. Joubert, A. Livinski, Y.-L. Lu, L. N. Snyders, H. Thompson, Data sharing in plos one: an analysis of data availability statements, *PloS one* 13 (5) (2018) e0194768.
- [42] N. Sarandria, A literature review in immuno-oncology: Pathophysiological and clinical features of colorectal cancer and the role of the doctor-patient interaction, *Journal of Cancer Therapy* 13 (12) (2022) 654–684.

- [43] physionet, [Physionet publishing guidelines](https://physionet.org/about/publish/#guidelines), accessed: 2023-08-20 (2023). URL <https://physionet.org/about/publish/#guidelines>
- [44] J. A. Bridge, T. R. Goldstein, D. A. Brent, [Adolescent Suicide and Suicidal Behavior](https://www.ncbi.nlm.nih.gov/books/NBK107191/), 2022, accessed: 2024-05-06. URL <https://www.ncbi.nlm.nih.gov/books/NBK107191/>
- [45] D. A. Brent, N. Melhem, Familial transmission of suicidal behavior, *Psychiatric Clinics of North America* 31 (2) (2008) 157–177.
- [46] A. E. W. Johnson, L. Bulgarelli, L.-F. Shen, et al., MIMIC-IV, a freely accessible electronic health record dataset, *Scientific Data* 10 (2023) 1. doi:10.1038/s41597-022-01899-x.
- [47] A. Johnson, L. Bulgarelli, T. Pollard, et al., MIMIC-IV. doi:10.13026/S6N6-XD98.
- [48] N. Sadvilkar, M. Neumann, [PySBD: Pragmatic sentence boundary disambiguation](https://www.aclweb.org/anthology/2020.nlpss-1.15), in: *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, Association for Computational Linguistics, Online, 2020, pp. 110–114. URL <https://www.aclweb.org/anthology/2020.nlpss-1.15>
- [49] H. D. Anderson, W. D. Pace, E. Brandt, R. D. Nielsen, R. R. Allen, A. M. Libby, D. R. West, R. J. Valuck, Monitoring suicidal patients in primary care using electronic health records, *The Journal of the American Board of Family Medicine* 28 (1) (2015) 65–71.
- [50] Columbia University, Columbia-Suicide Severity Rating Scale: Pediatric Screener Lifeline Version, https://cssrs.columbia.edu/wp-content/uploads/C-SSRS_Pediatric-SLC_11.14.16.pdf, accessed: 2024-05-06 (2016).
- [51] Psychological Assessment Resources, Suicide ideation questionnaire (siq), <https://www.parinc.com/Products/Pkey/413>, accessed: 2024-05-06 (2024).
- [52] Pearson Assessments, Beck scale for suicide ideation, <https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Beck-Sca>

- [le-for-Suicide-Ideation/p/100000157.html](#), accessed: 2024-05-06 (2024).
- [53] Mpnet, https://huggingface.co/docs/transformers/model_doc/mpnet (accessed 25 Jul 2023).
 - [54] World health organization.
URL <https://www.who.int>
 - [55] Centers for disease control and prevention.
URL <https://www.cdc.gov/index.htm>
 - [56] U.s. food and drug administration.
URL <https://www.fda.gov>
 - [57] G. K. Krippendorff, Package 'irr'.
URL <https://cran.r-project.org/web/packages/irr/irr.pdf>
 - [58] K. Huang, J. Altosaar, R. Ranganath, Clinicalbert: Modeling clinical notes and predicting hospital readmission, CoRR abs/1904.05342 (2019). [arXiv:1904.05342](https://arxiv.org/abs/1904.05342).
URL <http://arxiv.org/abs/1904.05342>
 - [59] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
 - [60] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pre-training approach, arXiv preprint arXiv:1907.11692 (2019).
 - [61] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, Advances in neural information processing systems 32 (2019).
 - [62] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, Biobert: a pre-trained biomedical language representation model for biomedical text mining, CoRR abs/1901.08746 (2019). [arXiv:1901.08746](https://arxiv.org/abs/1901.08746).
 - [63] University of Florida Research Computing, [HiperGator Supercomputer](#), accessed: September 4, 2023 (2023).
URL <https://www.rc.ufl.edu/about/hipergator/>