Computers in Human Behavior Reports 19 (2025) 100733

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Review

Categorizing and assessing aspects of suicidal ideation detection approaches: A systematic review

Golnaz Nikmehr [a,*], Aritz Bilbao-Jayo [b], Aitor Almeida [b]

[a] Deustotech - University of Deusto, Bilbao, Spain
[b] University of Deusto, Bilbao, Spain

A R T I C L E I N F O

Keywords:
Suicidal ideation detection
Social networks
Systematic review

A B S T R A C T

Suicide remains a critical global issue and one of the leading causes of death worldwide. As this problem grows, the need for effective prevention strategies becomes increasingly urgent. Social networks and online platforms, such as Twitter, have emerged as spaces where people openly share their thoughts and emotions, including negative feelings, reflections on life, and even suicidal thoughts. This makes social media data an important resource for efforts to detect and reduce the risk of suicide.

This systematic review examines 92 studies published between 2018 and 2024 on the detection of suicidal ideation. The studies are categorized using a multidimensional framework that considers three key aspects: the platforms used for data collection, the analytical techniques applied, and the specific features employed to identify suicidal ideation.

By exploring these dimensions, the review highlights existing gaps and limitations in current methods, offering insights to guide future research and improve strategies for suicide prevention.

1. Introduction

According to the World Health Organization (WHO), approximately 800,000 people die by suicide annually, making it one of the leading global causes of death (World Health Organization, 2021). Among individuals aged 5–29 years, three of the top five causes of death — road traffic injuries, homicide, and suicide — are injury-related (World Health Organization, 2022). Suicide rates also vary significantly across nations and regions, underscoring the importance of considering diverse social and cultural factors. Understanding suicide, therefore requires a comprehensive approach that accounts for mental health, age, gender, and regional influences.

The risk factors for suicide are complex and shaped by a wide range of variables. For instance, Richardson, Robb, McManus, and O'Connor (2023) observed gender differences, noting that males are less likely than females to report suicidal thoughts or attempts. Key determinants include sociodemographic factors such as gender, age, and ethnicity, as well as life experiences like early adversity and trauma. Both physical and mental health play critical roles in influencing suicidal ideation.

Supporting this multifactorial view, Baldini, Gnazzo, Maragno et al. (2025) examined suicide risk among adolescent psychiatric inpatients, identifying key contributors such as insomnia, depression, and social-personal factors like bullying and physical inactivity. The study found that adolescents who attempted suicide were more likely to be female, experience depression, and suffer from sleep disturbances—particularly insomnia, which was closely linked to emotional dysregulation. These findings underscore the need for comprehensive intervention strategies addressing both psychological and social vulnerabilities.

Expanding on the role of sleep, Baldini, Gnazzo, Santangelo et al. (2025) also explored how sleep disturbances increase suicide risk among individuals experiencing first-episode psychosis (FEP). High rates of insomnia, circadian rhythm disruptions, and poor sleep quality were observed, with severe sleep issues associated with a 2.5-fold increase in suicidal ideation. These connections appear mediated by cognitive impairments and neurobiological changes, highlighting sleep as a key target for suicide prevention across high-risk groups.

The WHO and the American Association of Suicidology (AAS) define suicidal ideation as "thinking about, considering, or planning for suicide" (American Association of Suicidology, 2023). This includes passive ideation — thoughts of death without specific plans — and active ideation, which involves concrete plans and intent to act. Suicidal ideation arises from a complex mix of mental health conditions, traumatic experiences, and environmental stressors, reflecting its multifaceted nature.

According to Ilic and Ilic (2022), approximately 759,000 suicides occurred worldwide in 2019, with a global age-standardized rate (ASR) of 9.0 per 100,000 population, higher in males (12.6) than in females

* Corresponding author.
E-mail addresses: golnaz.nikmehr@deusto.es (G. Nikmehr), aritzbilbao@deusto.es (A. Bilbao-Jayo), aitor.almeida@deusto.es (A. Almeida).

https://doi.org/10.1016/j.chbr.2025.100733
Received 5 February 2025; Received in revised form 28 May 2025; Accepted 22 June 2025
Available online 9 July 2025

(5.4). Suicide rates increased with age, particularly among those aged 70 and older, and remained consistently higher in males across all age groups. These trends vary widely by gender and country, often driven by a complex interplay of socio-economic challenges and untreated mental health issues. As a result, implementing targeted, evidence-based prevention strategies is not just beneficial but critical. Ultimately, comprehensive and proactive prevention efforts are essential to reduce suicide rates, improve mental well-being, and advance global public health objectives.

Oexle, Waldmann, Staiger, Xu, and Rüsch (2018) examine the impact of mental illness stigma on suicidal ideation. Societal perceptions of mental health are often distorted by stigma and misunderstanding, significantly affecting individuals struggling with mental illness. Persistent beliefs equating mental health issues with weakness or moral failing foster discrimination and social isolation. This stigma discourages help-seeking and perpetuates silence and shame, making it harder for those in need to access support and resources.

Addressing these societal perceptions is essential for effective suicide prevention. Education and awareness campaigns can play a pivotal role in reshaping public attitudes, fostering empathy, and promoting open conversations about mental health. By encouraging understanding and acceptance, communities can create supportive environments that empower individuals to seek help and reduce the risk of suicidal thoughts and behaviors.

The COVID-19 pandemic has significantly exacerbated mental health challenges among young people, with studies indicating a marked rise in conditions such as depression and anxiety. Surveys conducted during the pandemic revealed that approximately 48% of youth experienced clinical depression, and 51% reported symptoms of anxiety, highlighting an urgent need for mental health support (Bell et al., 2023). Social isolation, disruptions to education and employment, and heightened loneliness contributed to this decline, particularly among those with pre-existing conditions.

Additionally, the pandemic disrupted key developmental milestones, increasing feelings of hopelessness and uncertainty about the future. Many young individuals reported that their symptoms worsened, with some experiencing severe outcomes such as self-harm and suicidal ideation. This alarming trend underscores the necessity for targeted mental health interventions and robust support systems tailored to youth in the post-pandemic context.

Cultural, societal, and socioeconomic factors significantly influence the development and detection of suicidal ideation. Often rooted in complex interactions between biological predispositions and life experiences, suicide risk is shaped by cultural norms, social expectations, and economic stress. Marginalized individuals — such as members of the LGBTQ community or those living in poverty — face elevated risks due to persistent stigma, discrimination, and limited access to mental health care. In many societies, cultural stigma surrounding mental illness suppresses open dialogue, leading to underreporting of suicidal ideation and misclassification of suicide deaths as accidental. Furthermore, societal norms equating vulnerability with weakness deter help-seeking, especially among high-risk groups such as men and veterans.

Detection and intervention efforts must also account for these broader contexts. While standardized tools like the PHQ-9 and the Columbia-Suicide Severity Rating Scale are widely used, they often lack cultural sensitivity and are less effective in diverse or underserved populations. Evidence shows these tools are insufficiently predictive in primary care settings, frequently missing high-risk individuals. Public health strategies such as school-based and community programs hold promise, but their effectiveness is often hampered by inconsistent implementation and limited evidence, especially in low-resource settings. Barriers such as poor healthcare access, a shortage of culturally competent providers, and economic hardship disproportionately affect vulnerable groups. Improving detection and response requires systemic change, including stigma reduction through education, culturally informed healthcare training, and policies addressing the social determinants of mental health (Bahamón et al., 2025).

Table 1
Research questions and their motivations.

Question	Motivation
What are the existing approaches in suicidal ideation detection?	To categorize them and have a clear view of different aspects of suicidal ideation detection.
What are the main platforms that are used in suicidal ideation detection area?	To identify the existing platforms such as social networks, interviews, questionnaires, and electronic health records and also the limitations of them.
What are the most popular methods in suicidal ideation detection?	To find the gaps and areas that may be interesting to explore for reducing the risk of suicide.
What are the best features to consider in suicidal ideation detection?	To have a better understanding of useful features to generalize the solution of the detection.

Social media has transformed how mental health and suicidal ideation are experienced and expressed, particularly among adolescents and young adults. On the positive side, these platforms can raise awareness, disseminate mental health information, and foster support networks that help youth recognize symptoms and seek assistance. However, the intense and immediate nature of online interactions can also influence emotional development in complex ways, potentially reinforcing both positive and negative mental health trajectories. These dual effects underscore the importance of further research and clinical strategies that leverage digital platforms while mitigating associated risks (Khalaf, Alubied, Khalaf, & Rifaey, 2023).

The increasing use of platforms like Twitter and Reddit has also opened new avenues for identifying suicidal ideation through online expression. Burnap, Colombo, Amery, Hodorog, and Scourfield (2017), for example, analyzed tweets using lexical, structural, emotive, and psychological features to classify indicators of suicidal ideation. This research highlights the potential of social media as a tool for early detection and intervention in suicide prevention efforts.

Given the growing statistics on suicide and the significant influence of social media on mental health, the urgency and importance of research in this area are evident. In response, our review focuses on recent studies that apply machine learning and deep learning techniques, methods that have become increasingly prominent and impactful in identifying patterns, predicting risk, and ultimately contributing to the prevention of suicidal behavior. The goal of this work is to systematically review, categorize, and analyze these methods to highlight their contributions and potential. In the following sections, we provide a detailed examination of these methodologies and their classification. The specific research objectives and motivations guiding our review are summarized in Table 1.

This paper is structured as follows: Section 2 describes the search strategy and quality metrics used in evaluating the studies reviewed. Section 3 presents the outcomes of the study selection process and provides a summary of the included studies. In Section 4, we introduce a novel framework for categorizing the studies reviewed. Section 5 offers a comprehensive summary of the systematic review, including a detailed table of relevant papers. Finally, Section 6 concludes the paper with key findings and recommendations.

2. Methods

This section outlines the methodology used to conduct the systematic review. Following the established guidelines provided by Kitchenham and Charters (2007), we adopted a systematic approach to ensure the reliability of our findings. Our process began with the development of comprehensive search strategies to identify relevant scholarly works. We then focused on addressing the predefined research questions and conducted a detailed analysis of the collected data.

Table 2
Search queries samples.

Query1	"Suicidal Ideation" AND "Detection" AND "Social Networks" AND "Machine Learning"
Query2	Suicidal Ideation Prediction in Social Networks using Machine Learning

2.1. Search strategy

The search process began with the careful selection of relevant terms to align with our research objectives and ensure optimal results. Key search terms included *Suicidal Ideation*, *Detection*, *Prediction*, *Social Networks*, *Machine Learning,* and *Deep Learning.* The search prioritized precise matches in the titles, abstracts, and keywords of relevant articles. To ensure a thorough review, we leveraged the advanced search capabilities of reputable academic databases, specifically Scopus and ACM. Furthermore, the search was refined to include publications from 2018 to 2024, enabling a focus on the most recent advancements in the field.

Table 2 provides examples of the search queries employed in this study. As shown, we used a combination of keywords such as "Suicidal Ideation", "Detection", "Social Networks", "Machine Learning", and "Deep Learning" connected by the Boolean operator AND. These queries were limited to publications from 2018 to 2024, aligning with the scope of our review, and were restricted to English-language sources. Additionally, we included a more specific query using the complete phrase "Suicidal Ideation Prediction in Social Networks using Machine Learning".

To enhance the quality and thoroughness of our review process, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page, McKenzie et al., 2021). PRISMA is a widely recognized guideline that promotes transparency, standardization, and comprehensive reporting in systematic reviews and meta-analyses. By following PRISMA, we ensured a methodologically robust and precise approach. The following sections provide a detailed account of our review methodology, emphasizing how the application of PRISMA contributed to the reliability and depth of our systematic review and meta-analysis.

2.2. Quality metrics

In the paper selection process, we utilized a set of quality metrics derived from reputable sources, including PRISMA (Page, McKenzie et al., 2021; Page, Moher et al., 2021) and other established methodologies (Castillo-Sánchez et al., 2020; Yeskuatov, Chua, & Foo, 2022), alongside our expertise in the field. In the following, we established specific selection criteria to ensure that only related and high-quality studies were included in our review:

- Publications from 2018 to 2024, ensuring the inclusion of the most recent and relevant research findings.
- Studies published in English to ensure accessibility and ease of comprehension.
- Focus on studies that introduce innovative and significant contributions to the existing body of knowledge.
- Prioritization of works that specifically address the detection or prediction of suicidal ideation, closely aligning with the objectives of our research.
- Inclusion of studies that utilize Machine Learning or Deep Learning methods, recognizing the importance of advanced computational techniques in effectively tackling the research question.

By following these strict selection criteria, we ensure that the studies included in our review are both relevant and of high quality, thereby reinforcing the reliability and robustness of our research findings.

2.3. Related work

Detecting suicidal thoughts is a critical aspect of preventing self-harm and suicide. In recent years, numerous studies have focused on this pressing issue, employing various methods to find the most effective solutions. These studies often combine machine learning techniques with mental health analysis to identify signs of suicidal ideation. For example, in Castillo-Sánchez et al. (2020), the authors conducted a systematic review following the PRISMA protocol, selecting sixteen studies from respected databases such as PubMed, Science Direct, IEEE Xplore, and Web of Science. These studies explored the integration of machine learning and social networks in suicide prevention research.

Additionally, in Arowosegbe and Oyelade (2023) Arowosegbe et al. explored the utilization of NLP for the detection and prevention of suicidal ideation. Their review involved the compilation of twenty papers from sources including PubMed, EMBASE, MEDLINE, PsycINFO, and Global Health databases. A significant portion of the research they reviewed focused on improving the ability to detect or predict suicidal thoughts on social networks. They mainly used advanced techniques like deep learning and state-of-the-art language models such as BERT. Also, a review (Schafer, Kennedy, Gallyer, & Resnik, 2021) conducted a systematic survey of studies covering a wide range of suicide-related outcomes, including suicidal ideation, suicide attempts, and actual suicide deaths.

In the review by Ji et al. (2020), two main categories for suicidal ideation detection are identified: clinical approaches, which involve direct interactions between social workers or experts and individuals, and machine learning techniques, which include feature engineering and deep learning methods for automatic detection based on online social content. The review also explores domain-specific applications of suicide ideation detection, utilizing various data sources such as questionnaires, EHRs, suicide notes, and user-generated content from online platforms.

Furthermore, a comprehensive review (Lopez-Castroman et al., 2019) highlights several projects aimed at preventing suicide. A prominent initiative mentioned is the 'Durkheim Project' funded by the U.S. Government, which seeks to improve suicide risk detection through data mining techniques applied to text and images from social media networks. Another important contribution is a systematic review and meta-analysis by Kusuma et al. (2022), which evaluates the performance of 54 machine learning models across 35 studies. This analysis reveals that the models achieved an AUC of 0.86, sensitivity of 0.66, and specificity of 0.87.

For an overview of the reviews in the field of suicidal ideation detection and their objectives, refer to Table 3. Our review distinguishes itself by introducing a novel categorization of existing approaches across multiple dimensions. This structure allows us to highlight the limitations within each category, thereby identifying gaps that can guide future research more effectively. As discussed, prior reviews often concentrate narrowly, some focusing just on social media platforms, specific techniques, or detailed methodological aspects. In contrast, our categorization provides a broader and more insightful perspective, which we elaborate on in the following sections.

3. Results

In this section, we systematically analyze and synthesize the results of the review to provide comprehensive answers to the research questions, which were designed to explore the methodologies used for detecting suicidal ideation on social media platforms. This section outlines the studies selected for inclusion, with Section 3.2 detailing the methodology used for retrieving and selecting the manuscripts. Section 3.3 then provides a high-level overview of the chosen studies.

Table 3
Related works conducted in the area covered by this research.

Paper	Time frame	Type	Objective
Lopez-Castroman et al. (2019)	2000–2017	Scoping review	Overview of different aspects connecting social media and suicidal behavior
Yeskuatov et al. (2022)	2018–2022	Literature review	Investigate the methods employed to detect suicidal ideations on the Reddit forum.
Arowosegbe and Oyelade (2023)	2016–2022	Systematic review	Use of NLP in detecting and preventing suicide ideation
Castillo-Sánchez et al. (2020)	2015–2019	Scoping review	Analyze the state-of-the-art on the use of machine learning methods for suicide detection
Schafer et al. (2021)	1984–2020	Meta-analysis	Compare the predictive accuracy of machine learning models in predicting suicide ideation, attempts, and death
Lasri et al. (2022)	2016–2022	Short literature review	Overview of different methods, including machine learning and deep learning, for suicide ideation detection
Abdulsalam and Alhothali (2024)	2014–2020	Literature review	Summary of current research efforts to detect suicidal ideation using machine learning algorithms on social media
Kusuma et al. (2022)	2002–2021	Systematic review/ Meta-analysis	Evaluate the performance of machine learning models in predicting suicide ideation, attempt, and death
Bhardwaj, Gupta, Goyal, Nagpal, and Jha (2022)	2013–2020	Literature review	Examine the existing studies and research in the detection of suicidal thoughts using machine learning and deep learning.
Ji et al. (2020)	2007–2020	Literature review	Introduce and discuss methods for suicidal ideation detection, with a focus on early detection and prevention of suicide attempts

3.1. Inclusion and exclusion criteria

During the screening phase, we applied specific inclusion and exclusion criteria to ensure the relevance and quality of the selected studies.

Inclusion Criteria

- Sufficient citation count indicating scholarly impact.
- Published in reputable journals or conferences with rigorous peer review.
- Clear and informative title and abstract conveying the study's scope and objectives.
- Focus on the use of Natural Language Processing (NLP) or machine learning methods.
- Studies conducted between 2018 and 2024.
- Studies written in English.

Exclusion Criteria

- Low citation count or published in low-quality venues.
- Inadequate or unclear title and abstract.
- Studies that focused solely on psychiatric or clinical aspects without computational analysis.
- Limited or no emphasis on NLP techniques or social media data.
- Duplicated or inaccessible full texts.

3.2. PRISMA based selection

Fig. 1 provides a clear representation of the PRISMA (Page, McKenzie et al., 2021; Page, Moher et al., 2021) flowchart, a widely recognized and standardized framework for conducting systematic reviews and meta-analyses. The flowchart outlines the step-by-step process of the review, beginning with the initial literature search and screening, followed by the assessment of study eligibility. Studies are then included or excluded based on predefined criteria. After the relevant studies are selected, data extraction and analysis are performed, leading to the synthesis of results and the creation of the final report. The flowchart concludes with the completion of the review process.

As illustrated in Fig. 2, the study selection process followed a structured PRISMA framework encompassing four main phases: identification, screening, eligibility assessment, and inclusion. This systematic approach ensured a comprehensive and focused review of relevant literature on suicidal ideation detection using NLP and social media data.

The process began with the identification of relevant studies through extensive database searches. Articles were retrieved from four major sources: Google Scholar (n = 17,200), ACM Digital Library (n = 23,789), Science Direct (n = 213), and Scopus (n = 31), resulting in an initial corpus of 41,233 records. Before formal screening, a set of preliminary exclusion criteria was applied. This included the removal of 21,145 duplicate records and 536 articles deemed irrelevant due to unrelated keywords or ambiguous publisher designations. After these initial exclusions, 19,552 unique articles proceeded to the screening stage.

During the screening phase, articles were evaluated based on their titles and abstracts. A total of 16,745 articles were excluded at the title screening stage due to lack of relevance or topic inconsistency. The abstract-level screening further excluded 2348 papers that did not meet the study's scope or focus. As a result, 459 articles were shortlisted for full-text retrieval and detailed evaluation.

In the eligibility phase, 10 reports could not be retrieved, leaving 449 papers for full-text assessment. Of these, 351 articles were excluded for focusing primarily on psychiatric aspects rather than computational methods. An additional 6 papers were excluded for addressing NLP without a direct link to suicidal ideation or social media analysis.

Finally, a total of 92 studies were included in the review. These studies were determined to be the most relevant and met all predefined inclusion and exclusion criteria (as outlined in Section 3.1). Together, they form the basis of our in-depth investigation into current methodologies and advancements in the use of NLP for detecting suicidal ideation on social media platforms.

3.3. Publishing patterns and research growth in the field

This section provides a detailed examination of the selected research papers, including an overview of their distribution across conferences

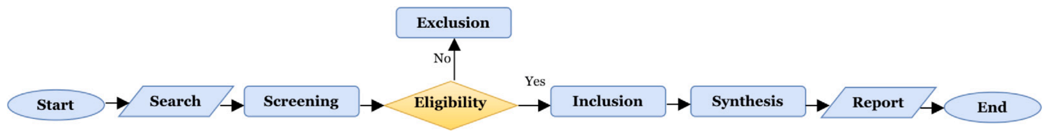


Fig. 1. PRISMA flow chart.

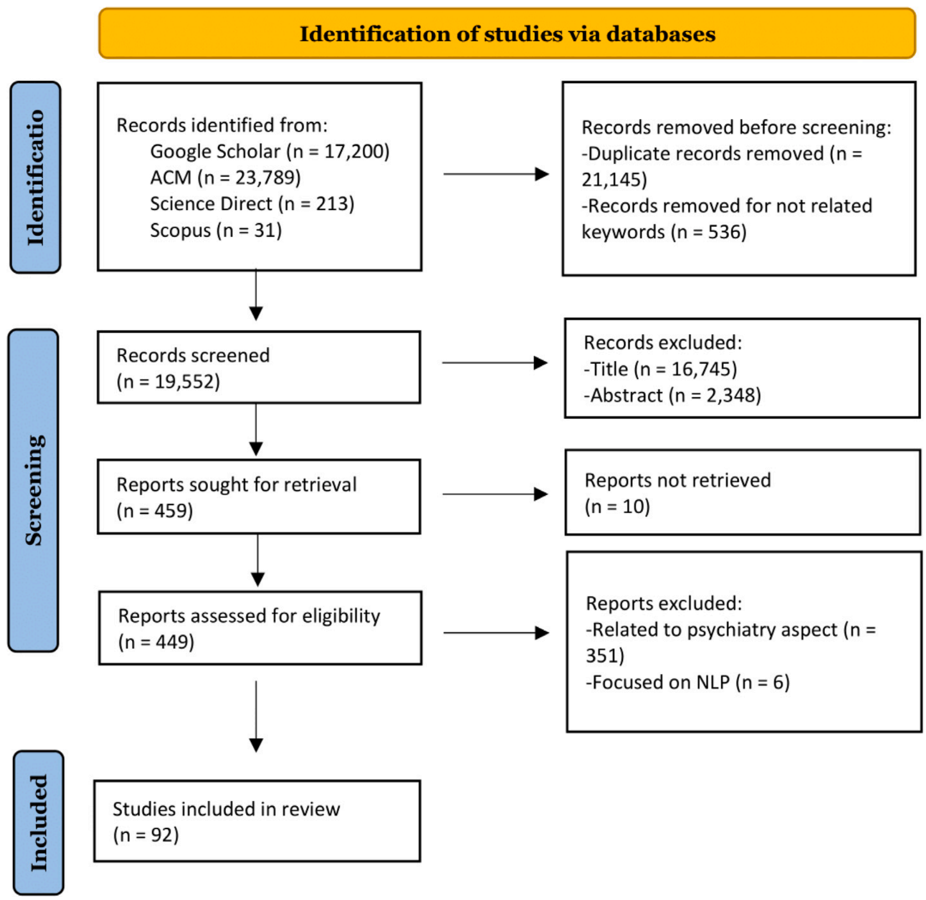
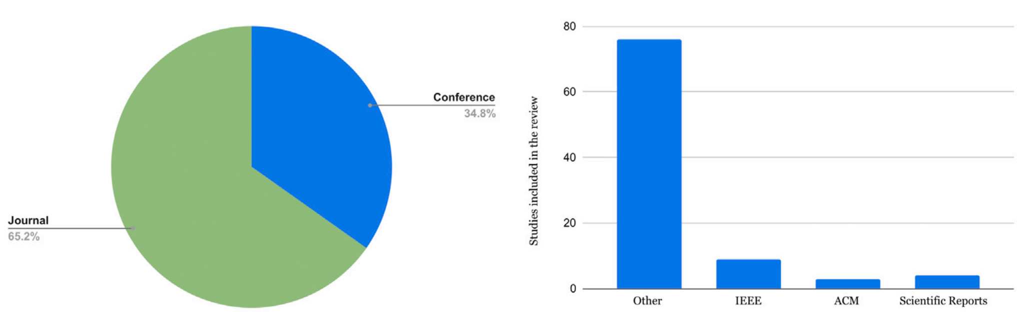


Fig. 2. Illustration of the systematic review process, including database search procedures.



(a) Type of Publications (b) Number of papers by publisher

Fig. 3. An overview of the publishers featured in the review.

and journals, as well as the identification of key publishers. As shown in Fig. 3, the majority of the studies included in this review were published in journals, accounting for 65.2% of the total, while 34.8% were conference papers. The publishers of these papers include IEEE, ACM, Scientific Papers, and a category labeled "Other", which includes respected sources such as the International Journal of Environmental Research and Public Health, MDPI, and Mathematics.

In this study, we conducted a comprehensive review of scholarly papers published between 2018 and 2024, analyzing their distribution over this period. The analysis highlights a notable trend starting in 2020, with increasing interest in utilizing machine learning methods for suicide ideation detection, suicide prevention, and mental health research.

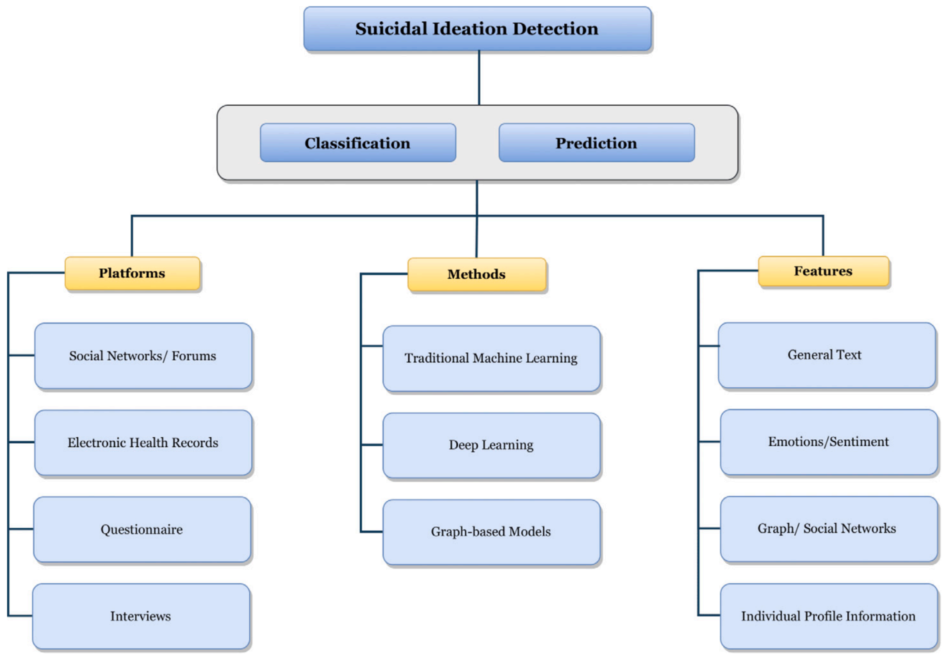


Fig. 4. The fundamental groupings found in examined papers.

4. Discussion

This section serves as the core of our discussion, focusing on the studies included in this review. As depicted in Fig. 4, it provides an overview of the categories considered in this analysis. To facilitate a thorough understanding, we have organized the studies into distinct segments, which will be explored in detail in the following sections.

Fig. 4 offers a foundational reference, systematically categorizing the task of detecting suicidal ideation into two primary domains: *Classification* and *Prediction*, which are further discussed in Section 2. The *Classification* domain involves categorizing instances into specific classes, while the *Prediction* domain incorporates temporal factors to forecast the classification of instances in future time frames.

Next, we analyze the relevant literature through three additional key categories: *Platforms*, *Methods*, and *Features*. Each of these categories is explored in detail in the following subsections.

The *Platforms* category is divided into four subcategories: “Social Networks”, “Questionnaires”, “Interviews”, and “Electronic Health Records (EHR)”. These subcategories are thoroughly explained in Section 4.1. In the *Methods* category, discussed in Section 4.2, we examine three subcategories: “Traditional Machine Learning”, “Deep Learning”, and “Graph-based Methods”. Finally, the *Features* category, detailed in Section 4.3, focuses on various features used in the literature, categorized into “Text”, “Emotional”, “Social Network Features”, and “Individual Profile Information”. To conclude, Section 5 provides a comprehensive table summarizing the papers analyzed within this categorical framework.

4.1. Platforms

As illustrated in Fig. 4, the studies reviewed in this paper leverage various platforms to detect suicidal ideation. The findings in Fig. 5 present majority of studies, representing 79.3% of the total, utilize Social Networks and Forums as their primary platform. Questionnaires emerge as the second most common platform, used in 12.2% of the studies. Interviews account for 6.1% of the research, while Electronic Health Records (EHR) are the least utilized platform, featured in only 2.4% of the studies reviewed.

4.1.1. Social networks/forums

Research focuses on this category, including social networks such as Twitter, Reddit, and Facebook and Forums like Sina Weibo and SanctionedSuicide. In the case of Twitter, the included papers use annotated tweets where users express their sentiments and thoughts related to suicide. Some of these papers apply specific lexicons containing phrases or keywords indicative of suicidal ideation, including expressions such as ‘Kill myself’, ‘Don’t want to exist’, ‘Want to be dead’, ‘Ready to die’, and ‘Hate my life’ (Valeriano Valdez, Condori-Larico, & Sulla-Torres, 2020). This approach introduces limitations and potential biases into the dataset.

Furthermore, certain studies involve tweets in languages other than the target language and address this by attempting translations. For instance, in one study (Valeriano Valdez et al., 2020), tweets are translated into Spanish, while another work focuses on translating English to Portuguese (de Carvalho, Giacon, Nascimento, & Nogueira, 2020). This language-based approach adds both limitations and specifics to the research, requiring specialized handling and consideration.

Reddit is among the most widely used social networking platforms, with a dedicated subreddit called *r/SuicideWatch* that focuses on suicide-related discussions. Some studies, such as Choi and Yang (2024), have utilized this subreddit to annotate posts as indicative of suicidal ideation. Data collection from Reddit can be facilitated through the Public Reddit Application Programming Interface (API). For instance, Chadha and Kaushik (2022) employed Reddit as a primary data source, where the annotation process was conducted using TextBlob, a lexicon-based sentiment analysis tool, and validated by a psychiatric expert. TextBlob uses predefined rules and a word-weight dictionary to calculate sentiment polarity. However, relying solely on text polarity or a subreddit to identify suicidal ideation has its limitations. This method may lead to data unreliability, as it oversimplifies the complex emotions and important details associated with suicidal ideation.

A widely utilized platform for detecting suicidal ideation is Sina Weibo, a Chinese microblogging site. Many posts on this forum convey users’ profound experiences, feelings of despair, and occasionally explicit expressions of self-harm or suicidal intentions. This platform’s dataset has been employed in studies such as Cao, Zhang, and Feng (2022) data was annotated based on users who explicitly mentioned

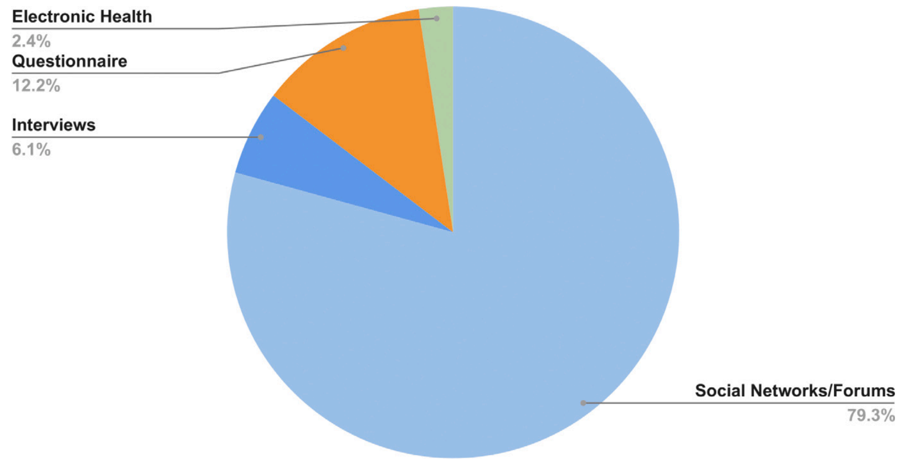


Fig. 5. The allocation of platforms employed in the studies under review.

suicidal thoughts, plans, or self-harm tendencies on at least five separate occasions over different days. For example, users express statements like, "At this moment, I especially want to die. I feel very tired. I really want to be free", "Even Weibo can't give me a reason to keep on going", or "I don't want to do anything for the last five days of my life" were categorized as being at risk of suicide. This manual annotation process is recognized as a more robust and reliable method for data labeling.

In Ophir, Tikochinski, Asterhan, Sisso, and Reichart (2020), the authors utilized Facebook as a social network platform for their research. They collected user-generated social media texts, collecting a dataset from 1002 Facebook users who willingly engaged in the completion of a clinically validated screening tool for assessing suicide risk. Additionally, these participants voluntarily disclosed a year's worth of their Facebook activity, contributing to a dataset that comprised 83,292 postings. It is important to note that using Facebook to detect signs of suicidal thoughts is not common in research. This is due to factors like limited data access and the platform's lower tendency for open expression of suicide-related feelings and thoughts by users.

Furthermore, Sakthi, Chen, and Sathiyanarayanan (2023) concentrate on sentiment messages from college students across various social media platforms like Twitter, Instagram, Facebook, and YouTube. Users on different social media platforms may express different sentiments, and the combination of users' posts may not be suitable for suicide ideation detection based on users.

4.1.2. Questionnaire

Another commonly used approach in the reviewed studies is the questionnaire-based platform. This method offers specificity by collecting data from a small group of willing participants, requiring careful planning to ensure data quality and reliability. However, it is more time-intensive compared to other data collection methods. For instance, the authors of Macalli et al. (2021) employed questionnaires in their study through the "Internet-based Students' Health Research Enterprise (i-Share)" project. This prospective, population-based study focused on student health at French universities in 2013.

Students were informed about the project via flyers, classroom announcements, and social media. Eligible participants included university students who were at least 18 years old and fluent in French. Online questionnaires gathered comprehensive data on demographics, health, personal history, and lifestyle. Examples of questionnaire variables included *Accommodation type*, *Opinion on resources*, *Year of study*, *Perceived parental support in childhood*, *Parental divorce*, and *Paternal depression history*.

The monitoring period spanned from February 2013 to September 2019, with a one-year follow-up to track suicidal thoughts and attempts. Participants reporting such experiences were categorized accordingly. While this approach enables detailed data collection, it has limitations, particularly in scaling up to include larger participant groups.

Kim, Gwak, Kim, and Gang (2022) employed self-report questionnaires to gather data on demographics, suicidal thoughts, depression, socioeconomic status, and general health information. The data collection took place between September and October 2020, focusing on adults aged 65 and older who lived independently and were registered with the Okcheon Public Health Department. A total of 650 participants were selected through a convenience sampling method. The main question asked was, "Have you contemplated terminating your life through any means within the past year?" Responses to this question were then analyzed to assess suicidal ideation.

It is important to note that questionnaire-based studies have inherent limitations. One key limitation is participants' hesitation to disclose negative or sensitive emotions, which can restrict the number of valid responses. Additionally, the accuracy of the data relies on whether participants are honest in their answers, which poses another significant challenge to data reliability.

4.1.3. Interviews

There are papers that use interviews for data collection, but like questionnaires, interviews may not always be the best choice for detecting suicidal ideation. This is because interviews have limitations related to the number of participants that can be effectively engaged and the complexities of data collection. In our observation, interviews are used less frequently than questionnaires. One reason for this is that interviews require face-to-face interactions, which can discourage people from participating in the study.

For instance, in Mens et al. (2020) conducted by Mens et al. a population-based investigation was carried out to document suicidal thoughts and behaviors among young adults aged 18 to 34 in Scotland. Participants took part in hour-long, in-person interviews at their homes and received a £25 incentive for their participation. In the follow-up phase of the ongoing study, participants who consented to further involvement were contacted 12 months later to complete an additional questionnaire. This follow-up survey was administered via email, mail, or telephone, and participants received additional compensation for their time.

4.1.4. Electronic Health Records

We studied some manuscripts that utilized Electronic Health Records (EHR) platforms in the detection of suicidal ideation. EHRs represent a digital transformation of traditional patient paper charts, offering real-time, patient-centric records that provide instant and secure access to authorized users. Beyond the standard clinical data found in a provider's office, EHR systems include a comprehensive array of patient information, including demographic details and diagnostic history, such as admissions and emergency visits.

In Zhu et al. (2020), collected Electronic Medical Records (EMRs) from psychiatric inpatients at a top Chinese psychiatric hospital. The data covers the period from January 1, 2010, to December 31, 2017, and includes information on individual inpatients, each identified by a unique hospital-specific ID. The dataset includes various data points like sociodemographic details (age, gender, nationality, place of origin, race, marital status), clinical diagnoses at discharge, chief complaints, and medical histories.

Zheng et al. (2020) used EHRs from patients who received medical care at three Berkshire Health System hospitals between January 1, 2015, and December 31, 2017, to identify patients at high risk of suicide attempts. While the EHR platform is valuable, it has some limitations, such as data availability and possible gaps in capturing the data. One significant limitation is the possibility of unreported suicide attempts. This highlights the need to train hospital staff to recognize warning signs and conduct follow-up inquiries to identify potential unreported suicide attempts.

4.1.5. Datasets

In this section, we provide an overview of the datasets utilized in research on suicidal ideation. Table 4 categorizes these datasets by their source, size, annotation methods, and public accessibility. Social network datasets, such as those from Sina Weibo, Twitter, and Reddit, typically require manual annotations. In contrast, datasets like the Patient Health Questionnaire-9, Student Interviews, and EHR employ automatic annotation during data collection. Some studies identify suicidal posts through keyword-based searches. For example, in Aldhyani and Alshebami (2022), posts containing terms like 'suicide', 'kill myself', and 'end my life' were flagged. Other studies may combine various types of datasets to improve their analysis. These variations emphasize the diversity of dataset characteristics in detecting suicidal ideation, as summarized in Table 4. Notably, the most commonly used datasets in the reviewed studies are the *Reddit CLPsycho 2019 Shared Task* and the *UMD Reddit Suicidality Dataset*.

As discussed, datasets for suicidal ideation detection can originate from a variety of sources, including interviews, questionnaires, EHRs, and social media platforms. However, traditional sources such as interviews, questionnaires, and EHRs often present challenges, such as limited accessibility, privacy concerns, and insufficient data volume, which require large-scale research efforts. In contrast, recent technological advancements, particularly the availability of social media data through public APIs, have significantly improved data accessibility. These platforms offer vast, real-time data streams that are well-suited for machine learning and deep learning applications, which require large and diverse datasets for effective model training. This shift marks a transformative step forward in enabling more scalable and data-driven approaches to mental health research (Chancellor & De Choudhury, 2020).

4.1.6. Limitations

As discussed in the previous subsections and confirmed through a thorough review of the literature across multiple platforms, it is evident that each method comes with its own set of limitations. These limitations include factors such as the linguistic nature of the data, age-specific parameters, and the difficulty of generalizing findings to broader contexts in suicidal ideation detection. These factors reduce the universal applicability and reliability of the methods. For example, the study in Abdulsalam and Alhothali (2024) provides a comparative analysis of research conducted in different languages, such as Russian and Japanese.

Specifically, when using interview-based platforms, the study often focuses on a select group of participants, such as adolescents on Instagram who have previously experienced suicidal thoughts (Lekkas, Klein, & Jacobson, 2021). However, this limited demographic may not adequately represent the general population. The small sample size further prevents the ability to generalize findings, and the lack of diversity among interviewees complicates the development of robust models, making them prone to overfitting.

When utilizing social network data such as Twitter, inherent biases and constraints tied to the data source can exclude a comprehensive understanding of the complexity and diversity of individuals' emotions and thoughts. Privacy policies on social media platforms further complicate efforts to identify individuals at risk of suicide. Limitations in the annotation process and issues with inter-rater agreement can introduce additional challenges.

Focusing on a single social network, such as Reddit or a specific subreddit, also risks reducing the generalizability of findings to broader online mental health contexts. For instance, in Yao et al. (2020), Yao et al. examined suicidality predictions in opioid-related subreddits and observed factors that could skew results. Extreme withdrawal symptoms, expressions of anger, or periods of sobriety were sometimes misinterpreted as indicators of suicidal intent. Conversely, when predicting opioid use within the r/suicidewatch subreddit, the emotional diversity of discussions — ranging from anger and despair to moments of happiness — posed challenges to maintaining predictive accuracy.

Similarly, Sawhney et al. (2021) highlighted that Reddit data could be influenced by demographic disparities, biases from expert annotators, and the platform's unique characteristics. These intricacies may introduce hidden challenges, especially when translating findings to real-world applications, underscoring the complexity of leveraging social network data for mental health research.

In another study (Jacobucci et al., 2021), the authors examined the use of traditional measurement methods, such as Likert-type items and rating scales, in the 'Amazon Mechanical Turk questionnaire'. They highlighted potential limitations of questionnaire methods in assessing and predicting suicide risk, such as standardized response options may fail to capture critical information necessary for accurate risk prediction. In contrast, text-based responses could provide insights into underlying processes or emotions that individuals may be unable or unwilling to articulate explicitly through fixed response formats.

Roy et al. (2020) have underscored that the expression of suicidal ideation on social media platforms may not necessarily correspond to actual suicidal behavior. Furthermore, the dataset employed in the study consisted exclusively of Twitter users who had previously posted about suicide or self-harm, potentially introducing a lack of representativeness for the broader population. Additionally, social media data is frequently characterized by noise, incompleteness, and inconsistencies, all of which can impact data quality.

In terms of questionnaires or interviews, the process involves careful participant selection and relies on the truthfulness of their answers, which can take a significant amount of time. Protecting data privacy is a top concern, as any lapses in this area can lead to complications. Participants may also be hesitant to openly share their true thoughts about suicide, which is a major challenge. Additionally, questionnaires are sometimes filled out at different times, which can affect the quality of predictions. In Van Vuuren et al. (2021), the identification of suicidal behavior was based on just four questions, which could limit the accuracy of the results. Moreover, it relied on self-reported data, which may be subject to biases and misinterpretation by respondents.

The limited use of EHR platforms can be attributed to data availability issues. EHR data requires specific infrastructure and equipment, which may not be readily accessible.

Class imbalance poses a significant challenge across all platforms, as the majority class (individuals without suicidal ideation) often dominates, introducing biases into predictive models. Research aimed at addressing this imbalance remains limited. Using binary classifiers to identify suicide-related posts can also lead to biases and inaccuracies, warranting cautious interpretation of results. Additionally, language translation further complicates the process, as it can lead to misinterpretations and a loss of the original meaning.

Table 4
Details of datasets utilized in the reviewed studies.

Dataset	Details	Annotation methodology	Accessibility
Sina Weibo microblog (Ma, 2021)	3652 users with suicide risk 3677 users without suicide risk	NA	–
Twitter (Yatapala & Kumara, 2021)	9217 Tweets 4062 Tweets suicidal 5144 Tweets non-suicidal	Manually checked	Yes (AminuIsrael, 2020)
Reddit (Aladağ, Muderrisoglu, Akbas, Zahmacioglu, & Bingol, 2018)	508,398 Reddit posts 10,785 posts randomly selected	Manually by psychiatrists and a computer scientist.	–
Reddit C-SSRS Suicide Dataset (Gaur et al., 2021; Naseem, Kim, Khushi, & Dunn, 2023)	448 users 7327 posts	Experts	Yes (Gaur et al., 2019b)
UMD Reddit Suicidality dataset (Ji, Li, Huang, & Cambria, 2022; Shing et al., 2018)	11,129 users (r/SuicideWatch) 1,556,194 posts	Crowdsourcing, Experts	Yes (Zirikly, Resnik, Uzuner, & Hollingshead, 2019)
Reddit (Tadesse, Lin, Xu, & Yang, 2019)	3549 suicide-indicative posts 3652 non-suicidal posts	Manually	–
Patient Health Questionnaire-9(PHQ-9) (Kim & Lee, 2022)	7994 participants 7214 no suicidal 780 suicidal	–	–
Interview (Cohen et al., 2020)	267 interviews from 60 students in eight schools by ten therapists. 29 indicated suicide or self-harm risk.	–	–
Electronic Medical Records (Zhu et al., 2020)	3600 psychiatric inpatients 1800 with suicidal behaviors 1800 without suicidal behaviors	–	
Reddit-Twitter (Ji et al., 2018)	3549 Reddit posts 10,288 tweets	Manually annotate	–
Twitter (Sawhney, Manchanda, Mathur, Shah, & Singh, 2018)	5213 texts (use generated lexicon and Twitter REST API)	Human annotators	–
Twitter (Mishra et al., 2019)	34,306 tweets	Two student annota-tors with expertise in clinical psychology and social media	–
Twitter (Chadha & Kaushik, 2021)	14,202 posts (keyword-based)	Human annotators	–
Reddit CLPsycho 2019 Shared Task (Allen, Bagroy, Davis, & Krishnamurti, 2019; Ambalavanan, Jagtap, Adhya, & Devarakonda, 2019; Bitew et al., 2019a; Chen, Aldayel, Bogoychev, & Gong, 2019; Matero et al., 2019; Morales, Dey, Theisen, Belitz, & Chernova, 2019)	Task A, Task B: 496 users in train, 128 users in test Task C: 993 users in train, 248 users in test	Human annotators	–
Twitter (Sawhney et al., 2022; Sawhney, Joshi, Gandhi, & Shah, 2020)	34,306 tweets, 32,558 unique users based on the lexicon of 143 suicidal phrases	Two annotators(students of Clinical Psychology)	–
Sina Weibo (Liu et al., 2020)	65,352 posts (Chinese posts)	Manually annotate	–
Twitter (Kumar, Rao, Nayak, & Chandra, 2020)	60188 tweets (collected based on specific keywords)	Manually based on the annotation rule	–
Scottish Wellbeing Study (Mens et al., 2020)	3508 young adults	Interview based	–
Sina Weibo (Cao et al., 2022)	Crawled 7329 users (each user follows 5 neighbor users (as friends) on average) 3652 users with suicidal ideation 3677 users without suicidal ideation	NA	Yes (bryant03, 2021)
Twitter (Roy et al., 2020)	512,526 tweets from 283 SI cases 3,518,494 tweets from 2655 controls	NA	–
Facebook (Ophir et al., 2020)	83,292 posts, 1002 users (include psychosocial information about the users)	Psycho-diagnostic measures and validation checks, ensuring the quality of the self-reported user responses	–
Twitter (Valeriano Valdez et al., 2020)	2068 posts (Spanish language)	Human annotators	–
Twitter (de Carvalho et al., 2020)	15,000 tweets (Brazilian Portuguese), 2446 labeled tweets	Expert	Yes (viniciosfaustino, 2020)
Berkshire Health System dataset (Zheng et al., 2020)	118,252 patients, 245 of whom attempted suicide in the year 2016. 118,095 patients and 203 cases with a suicide attempt in the year 2017	EHR based	–

(*continued on next page*)

Table 4 (*continued*).

Reddit (Haque, Un Nur, Jahan, Mahmud, & Shah, 2020)	3549 suicidal suggestive texts and several non-suicidal texts	Manually annotate	–
Psychiatry Department of Hospital (Toledo-Acosta et al., 2020)	20,000 adult outpatients	Doctors, clinical psychologists, and nurses	–
Reddit (Shah et al., 2020)	7098 data with user-post column (r/SuicideWatch)	Manually annotate	–
Twitter (Ramírez-Cifuentes et al., 2020)	1,214,474 tweets, 305,637 images, 252 users	Clinicians	–
Prison population (Horvath, Dras, Lai, & Boag, 2021)	353 participants including borderline personality disorder (BPD) and antisocial personality disorder (APD)	Interview based	–
Opioid Users on Reddit (Yao et al., 2020)	500 r/Opiates posts 500, r/SuicideWatch posts	Amazon Mechanical Turk	–
Twitter (Rabani, Khan, & Khanday, 2020)	18 756 tweets	Manually annotate	–
Sina Weibo (Ma & Cao, 2020)	2500 users with suicide risk 2500 users without suicide risk	NA	–
Hualien Armed Forces General Hospital (Lin et al., 2020)	3546 military men and women aged 18–50 years	Questionnaire based on Brief Symptom Rating Scale (BSRS-5)	–
Reddit (Sawhney, Joshi, Gandhi, & Shah, 2021)	270,000 users, 2181 potential suicidal users	Clinical experts	–
Questionnaires (Choi et al., 2021)	17,482 for training, 14,238 for testing	Questionnaire based	–
Instagram (Lekkas et al., 2021)	52 participants (adolescents on Instagram with a prior history of lifetime suicidal ideation)	Interview based	–
French i-Share cohort (Macalli et al., 2021)	5066 college students	Questionnaire based	–
Reddit C-SSRS (Ashok Kumar, Trueman, & Abinesh, 2021)	500 users, 15,755 posts	Experts	–
General population of Amsterdam students (Van Vuuren et al., 2021)	8998 unique students, 732 students reported suicide ideation	Questionnaire based	–
Amazon Mechanical Turk questionare (Jacobucci, Ammerman, & Tyler Wilcox, 2021)	947 online participants	Questionnaire based	–
Minnesota Multiphasic Personality Inventory-2 (Kim, Lee, & Lee, 2021)	7824 college students, 3685 males, 4139 females	Questionnaire based	–
Reddit (Aldhyani & Alshebami, 2022; Nikhileswar, Vishal, Sphoorthi, & Fathimabi, 2021)	2,32,074 posts (SuicideWatch, Teenagers subreddits) 116,037 suicidal posts 116,037 non-suicidal posts	Based on subreddit	Yes (Suicide, 2021)
Twitter-Reddit (Rabani, Khan, & Khanday, 2021)	7582 suicidal tweets + 2676 reddit posts	Psychiatrists and psychologists	–
CLPsych 2021 Shared Tasks (Wang et al., 2021)	Twitter and questionnaire Subtask1: 114 train users/test users22, Subtask2: 164 train users/30 test users	NA	–
Reddit (Renjith, Abraham, Jyothi, Chandran, & Thomson, 2022)	69,600 posts	Crowdsourcing, Experts	Yes (Zirikly et al., 2019)
Suicide notes (Zhang, Schoene, & Ananiadou, 2021)	659 suicide notes, 431 last statements, and 2000 neutral posts from r/fitness, r/parenting, r/teaching, r/relationships	–	Yes (Suicide, 2020)
Reddit (Chadha & Kaushik, 2022)	10,000 suicidal and 10,000 non-suicidal posts	TextBlob + psychiatric expert	–
Twitter (Metzler, Baginski, Niederkrotenthaler, & Garcia, 2022)	3202 tweets retrieved via the data reseller Crimson Hexagon	Manually annotate	–
Twitter (Haque, Islam, Islam, & Ahsan, 2022)	49,178 tweets based on suicide-related phrases	VADER and TextBlob, manually reviewed	–
Twitter (Diniz et al., 2022)	5699 tweets	3 psychologists different psychological approaches: cognitive behavioral theory, psychoanalytic theory, and humanistic theory	–
Korean National Health and Nutrition Survey (Park & Lee, 2022)	12,816 workers	Questionnaire based	–

(continued on next page)

Table 4 (*continued*).

Sina Weibo (Liu, Shi, & Jiang, 2022)	40,222 posts 2272 with suicidal ideation, 37,950 without suicidal ideation	Manually annotate	–
Behavioral Health Licensed Clinicians (Cohen et al., 2022)	70 participant	Interview based	–
Twitter (Ananthakrishnan et al., 2022)	9119 tweets intention 5121 and no intention 3998	NA	–
Twitter (Chatterjee, Samanta, Kumar, & Sarkar, 2022)	188 704 tweets, 1169 users	Manually annotate	–
Self-report questionnaires (Kim et al., 2022)	650 Adults aged over 65 living in rural areas of South Korea	Questionnaire based	–
High school and college students (Sakthi et al., 2023)	Twitter, Instagram, Facebook and Youtube of each student	NA	–

Annotations play a vital role in analysis but present their own challenges. When restricted to specific phrases and keywords, they risk overlooking important subtleties associated with suicidal ideation, potentially introducing bias. Misclassifications are particularly problematic due to the rarity of suicidal thoughts, emphasizing the need for careful consideration and refinement in data handling and model design.

4.2. Methods

The evolution of statistical analysis has progressed significantly, transitioning from basic descriptive techniques to advanced inferential and computational methods. Early approaches primarily involved summarizing data using measures such as the mean, median, and mode, accompanied by simple visualizations like histograms and pie charts. Over time, more sophisticated techniques, such as regression analysis, meta-analysis, and methods for detecting publication bias, have been developed, allowing researchers to uncover deeper patterns, assess complex relationships, and address sources of bias, thus improving the reliability and validity of statistical conclusions (Huang, Ribeiro, Musacchio, & Franklin, 2017). In recent years, particularly since 2018, there has been a notable shift toward the adoption of machine learning and deep learning techniques. These approaches are a central focus of this review.

The techniques used in the reviewed studies fall into two main categories, as discussed in Section 4: Classification and Prediction. Within each category, we provide further details about subcategories, which include ‘Traditional Machine Learning’, ‘Deep Learning’, and ‘Graph-based Methods’. A visual representation of the distribution of these categorizations is depicted in Fig. 6.

The category of “Traditional Machine Learning” refers to established and widely used techniques, including methods like Support Vector Machines (SVM) (Cortes & Vapnik, 1995), Random Forest (Breiman, 2001), and K-Nearest Neighbors (KNN) (Cover & Hart, 1967), among others.

In contrast, “Deep Learning” encompasses a broader range of advanced approaches that leverage multi-layered neural networks. This includes techniques such as Convolutional Neural Networks (CNN) (LeCun, Bottou, Bengio, & Haffner, 1998), Long Short-Term Memory networks (LSTM) (Hochreiter & Schmidhuber, 1997), transformer-based models like BERT (Devlin, Chang, Lee, & Toutanova, 2018), and hybrid models that integrate various deep learning methodologies.

4.2.1. Classification

From our perspective, classification involves assigning input data points to predefined categories or classes with precision. Specifically, in suicidal ideation detection, the input data is textual and can either reflect suicidal thoughts or not, categorized into discrete class labels: *suicidal* and *non-suicidal*. Within the scope of this review, the majority of the studies fall under the “Classification” category, accounting for approximately 87% of the total (Fig. 7).

Deep Learning. We delve into research studies that employ deep learning as their primary approach. One notable study (Sawhney et al., 2020) proposes a classification method using a time-aware transformer-based model for the initial evaluation of suicidal risk on social media platforms. What sets this work apart is its innovative incorporation of Emotional Historic Context (EHC), which analyzes users’ emotional states over an extended period to assess their mental well-being, as highlighted in Benton, Coppersmith, and Dredze (2017). The dataset for this study comprises Twitter posts that include a lexicon of 143 distinct suicidal phrases (Mishra et al., 2019). However, it is important to note that relying on such lexicons may introduce biases into the data. For the data, the authors utilized SentenceBERT (Reimers & Gurevych, 2019), an adaptation of the BERT model designed for generating contextual word embeddings. This approach enhanced the model’s ability to interpret the textual content better. For evaluation, the study employed the macro F1 score and recall as performance metrics. The authors reported a recall score of 0.8, demonstrating the effectiveness of their method compared to alternative approaches.

In Ma (2021), the authors presented a novel classification approach aimed at the detection of suicide risk. Their significant contribution to the field involved the introduction of a dual attention mechanism, designed to capture implicit correlations between text and images within the same social media posts. To evaluate the effectiveness of their method, the authors utilized a dataset sourced from the Sina Weibo microblogging platform, as previously discussed in Section 4.1. The methodology employed in their research involved the utilization of a Gated Recurrent Unit (GRU) layer (Cho et al., 2014) for the extraction of textual information from the textual content, while the Residual Network (ResNet) (He, Zhang, Ren, & Sun, 2016) architecture was applied to analyze and process the accompanying images. The fusion of these two outputs was facilitated through the novel dual attention mechanism. In comparative analyses with alternative approaches, the authors reported that their proposed method exhibited an impressive accuracy rate, approaching approximately 90% overall and demonstrating even higher performance, reaching approximately 91% when specifically considering image-related aspects.

In the research conducted by Haque et al. (2020), a Transformer-based methodology was introduced for the detection of suicidal ideation within social media posts, leveraging the capabilities of pre-trained language models. The study used a substantial dataset containing 3549 texts indicating suicidal thoughts, collected from the Reddit social media platform. It was compared to a set of non-suicidal texts. The model had three main parts: a data layer, an embedding layer, and a classification layer. In the data layer, preprocessing was applied to the suicidal texts, such as expanding abbreviations and removing URLs. Then, the embedding layer used pre-trained language models to convert these processed texts into numerical vectors. Finally, the classification layer employed a neural network to classify the texts as either suicidal or non-suicidal. The results presented in the study revealed the superior performance of Transformer-based models when compared to other deep neural network models, as evidenced by higher

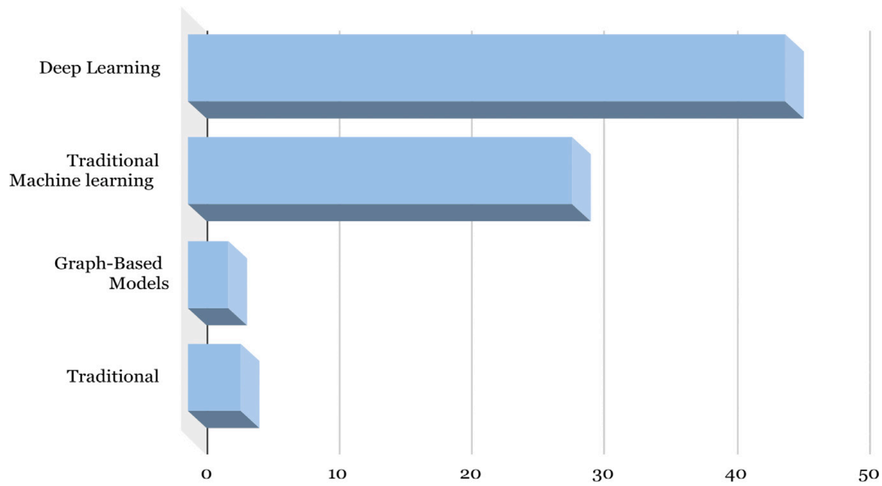


Fig. 6. Distribution of techniques utilized in the reviewed studies.

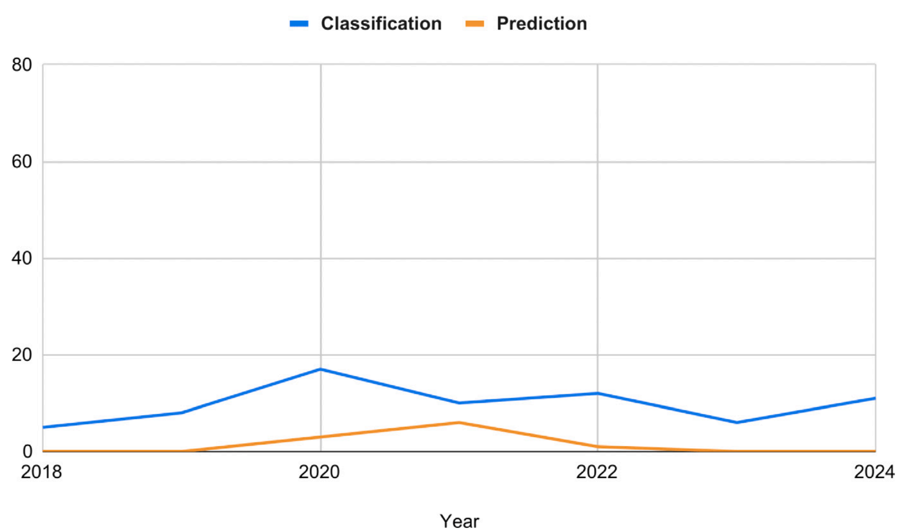


Fig. 7. Manuscripts techniques distributed in Classification and Prediction categories.

accuracy, recall, precision, and F1-score metrics. Notably, the ROBERTa model (Liu et al., 2019), which is a BERT pre-trained model depending on the dataset size, emerged as the top-performing model across all evaluated metrics. The study concluded that the Transformer-based approach, augmented by pre-trained language models, serves as an effective and promising tool for the detection of suicidal ideation within social media posts.

Aldhyani and Alshebami (2022) conducted an investigation into the application of deep learning techniques for the classification of suicidal ideation expressed on social media platforms. Their approach involved the development of a hybrid deep learning model, which integrated CNN with Bidirectional Long Short-Term Memory (Bi-LSTM) architecture (Author & OtherAuthor, 2019). The primary aim was to recognize and analyze suicidal tendencies within textual data. To assess the efficiency of their model, the researchers performed a comparative evaluation against a baseline model, XGBoost, which is a machine learning classifier. In these experiments, they used both textual features and features generated by the Linguistic Inquiry and Word Count-22 (LIWC-22) tool (Author & OtherAuthor, 2020) to classify the posts into two distinct categories: suicidal and non-suicidal.

In Ji et al. (2022) introduce a methodology aimed at detecting suicidal ideation and mental disorders within the content of social media. This approach leverages lexicon-based sentiment scores and relation networks, incorporating attention mechanisms to enhance text representation and classification accuracy. It is a multi-class classification task, necessitating a comprehensive understanding of user-level and post-level information. To validate their methodology, the authors conducted experiments using publicly available datasets. Additionally, the method utilized relational reasoning, incorporating relation networks (RNs) to integrate multichannel data and detailed textual representations. Their results indicated the superiority of their approach compared to existing methods. However, it is important to acknowledge that the proposed model exhibited challenges in predicting low-risk suicidal ideation. In Table 5, we reference additional research papers that address suicidal ideation classification within a multi-class dataset.

On the other hand, some works have recently tried to focus on LLMs and their applications such as Ghanadian, Nejadgholi, and Al Osman (2024) that Ghanaian et al. introduce a novel approach to synthetic data generation to detect suicidal ideation by using large language models (LLMs) and integrating relevant social factors from psychology literature. It focuses on creating targeted synthetic datasets that distinguish between varying levels of suicidal risk, thereby enhancing the model's predictive capabilities. By crafting specific prompts that reflect psychological themes, the authors improve the quality and diversity of the generated data, addressing the challenges of data scarcity and sensitivity in real-world contexts. The study demonstrates that classifiers fine-tuned on these topic-oriented synthetic datasets significantly outperform those trained on less relevant data, showcasing the potential

Table 5
Studies employ a multiclass approach when detecting suicidal ideation.

Study title	Classes
Suicidal ideation and mental disorder detection with attentive relation networks (Ji et al., 2022)	Depression, Suicide, Anxiety, Bipolar, PTSD
Suicidal risk identification in social media (Ashok Kumar et al., 2021)	Indicator, ideation, behavior, attempt, supportive
Detecting potentially harmful and protective suicide-related content on twitter: Machine learning approach (Metzler et al., 2022)	Coping, Awareness, Prevention, Suicide cases, Irrelevant
An investigation of deep learning systems for suicide risk assessment (Morales et al., 2019)	No risk, Low risk, Moderate risk, and Severe risk.
Towards ordinal suicide ideation detection on social media (Sawhney et al., 2021)	Support, Indirect suicidal ideation, Suicidal ideation, Plan and preparation, Suicide attempt.
Machine learning for suicidal ideation identification on twitter for the Portuguese language (de Carvalho et al., 2020)	Strongly concerning, Possibly concerning, Safe to ignore.

of this approach to enhance mental health interventions through more effective data-driven insights.

Traditional Machine Learning. In this part, we aim to highlight several studies that have employed traditional machine learning methodologies in the context of identifying suicidal ideation. Kumar et al. (2020), present RF as the chosen machine learning algorithm for a classification task focused on the identification of tweets indicative of suicidal ideation. The authors initiated data collection on the Twitter platform by filtering keywords such as ‘die’, ‘to die’, ‘suicide’, ‘kill myself’, and ‘end my life’, among others. Additionally, they incorporated n-grams as search key phrases to further refine the identification of tweets containing expressions of suicidal thoughts. Consequently, tweets featuring any of these suicidal keywords were categorized as instances of suicidal ideation. A critical component of applying machine learning techniques is the selection of features for model training. This study utilized a diverse array of features, including Term Frequency–Inverse Document Frequency (TF–IDF) (Salton & McGill, 1992), topic-based features, LIWC attributes, and various statistical properties. Among the models evaluated, the RF model outperformed others, such as Logistic Regression, achieving an impressive accuracy rate of 0.996. This high performance highlights the significant role of the carefully chosen keywords used for data collection and keyword matching in the study.

Kim and Lee (2022) used machine learning techniques, including KNN, RF, and Neural Network Classification, to predict suicidal thoughts based on depressive symptoms and social isolation. Data from community residents, including age, the Patient Health Questionnaire-9 (PHQ-9), and the Lubben Social Network Scale, were used as predictors of suicidal thoughts. Two prediction models were used: Model 1, which considered only depression as a predictor, and Model 2, which included social isolation in addition to depression. These models were trained, validated, and tested. The effectiveness of the models was evaluated using metrics such as the area under the curve, specificity, and accuracy. Incorporating social isolation as a feature significantly improved the accuracy of predicting suicidal ideation, with the Random Forest (RF) model proving to be the most effective method.

In Liu et al. (2022), aimed to enhance the detection of suicidal posts on social media by integrating single and multidimensional features. Their approach identified the best classification models, achieving an accuracy of 80.61% and an F1-score of 79.20%. The study highlights the importance of feature fusion, which combines diverse feature types to improve the identification of suicidal posts. While alternative text representation methods and demographic factors were acknowledged as potentially significant predictors, they were not explored in this research. Word embedding techniques, such as Word2vec, were employed to represent text features and capture contextual word relationships, demonstrating practical applications in suicide risk assessment. Additionally, the K-means algorithm was used to cluster suicide-related keywords based on Word2vec vectors, further enhancing the analysis.

Goel et al. in Goel and Digalwar (2024) employed an ensemble learning technique known as the Max Voting Ensemble classifier. This method combines the predictions of multiple individual machine learning models, such as SVM, LR, RF, and others, to enhance the overall accuracy and robustness of the classification task. By leveraging the strengths of various classifiers, the Max Voting Ensemble effectively captures subtle linguistic patterns and emotional cues indicative of suicidal ideation, thereby improving precision and accuracy in identifying at-risk individuals. The study demonstrates that this ensemble approach significantly outperforms individual models, showcasing its potential for timely intervention and support in mental health contexts, particularly in the analysis of social media content related to suicidal thoughts.

Graph-Based Models. Among the research papers employing graph-based methodologies, this approach remains relatively underutilized, as shown in Fig. 6. One major study (Mishra et al., 2019) highlights the potential of leveraging social engagement, ego networks, and user attributes to enhance the detection of suicidal ideation. The researchers compiled a large dataset of manually annotated tweets and introduced a deep-learning model called SNAP-BATNET. This model integrates text-based features with information from historical author profiling and graph embeddings to improve classification performance.

Social graphs were constructed to support author profiling to capture demographic features and enhance classifier performance. Four types of weighted, undirected graphs were developed: the *Follower Graph*, *Mentions Graph*, *RepliedTo Graph*, and *Quotes Graph*. Graph embeddings were generated for both weighted and unweighted graphs and individually analyzed for the classification task. Through extensive quantitative analyses and comparisons with baseline models, the study demonstrated that SNAP-BATNET outperformed alternative approaches, showcasing the effectiveness of combining social network insights with author profiling in suicidal ideation detection.

The study presented in Sawhney et al. (2022) introduces an innovative approach called Hyperbolic Conversation Network (HCN) for detecting suicidal ideation in online social media conversations. The primary focus of HCN is to model the intricate dynamics of conversation structures, including the interplay between comments and replies originating from an initial post.

HCN leverages the scale-free properties of conversation trees, where a small number of comments often attract a disproportionate number of responses. To accurately represent the exponential growth and hierarchical patterns of these structures, the model utilizes hyperbolic space, enabling a more effective representation of the diverse dynamics found in discussions about suicidal ideation.

Using real-world Twitter data, the study compared HCN with state-of-the-art methods. The results showed that HCN not only outperformed existing approaches in detecting suicidal ideation but also required significantly less user-specific data. Additionally, the model was more environmentally sustainable due to its smaller size. A key finding of the research was that comments posted within the first 30 min of a conversation play a critical role in identifying individuals at risk.

4.2.2. Prediction

In our objective within the Prediction category, we have various classes, with the key distinction being the consideration of future events. In suicidal ideation detection, the collection of data is considered in the context of time and the target is to predict the future risk of suicide. A smaller portion of the papers in this review, about 13%, falls into the "Prediction" category (Fig. 7). Next, we will explore the prediction studies using different techniques.

Deep Learning. In Sawhney et al. (2021) Sawhney et al. introduce an innovative framework termed 'SISMO,' an acronym for 'Suicide Ideation Screening using Multi-Objective optimization.' SISMO, designed as a hierarchical attention model, takes into account the graded nature of increasing suicide risk levels and factors in the temporal aspect of user posts. By considering the historical posts of a user and their contextual representation, the model aims to capture the changes in the user's mental state over time and make predictions about their suicide risk levels. In other words, it is formulated as an ordinal regression problem and multi-class classification in the field of suicide risk assessment on social media. For this purpose the data used for training and evaluation is based on the Columbia-Suicide Severity Scale (C-SSRS) (Gaur et al., 2019a) categories: *Support*, *Indicator*, *Low Risk*, *Moderate Risk*, and *High Risk* which is labeled by clinical experts.

In a related study (Zheng et al., 2020), the authors developed an Early-Warning System (EWS) to predict high-risk suicide attempts using deep learning and electronic health records (EHRs). Their model employed advanced machine learning algorithms and deep neural networks to create a risk stratification tool that assigns a risk score to individuals, reflecting the likelihood of a suicide attempt within the next year. The evaluation demonstrated strong performance, achieving an area under the curve (AUC) of 0.792.

Notably, the proposed model integrates seamlessly into routine medical care without requiring additional data collection, making it a cost-effective solution. This approach has significant potential to prevent suicide attempts and save lives by identifying high-risk individuals on a large scale. It also enables the creation of personalized intervention plans, further contributing to the reduction of suicide attempt rates.

Gaur et al. (2021) developed deep learning algorithms to assess suicide risk on Reddit, focusing on the severity and timing of risk using the C-SSRS. They applied two distinct deep learning approaches: time-variant modeling (TvarM) and time-invariant modeling (TinvM).

TvarM leverages sequential models such as recurrent neural networks (RNNs) and LSTMs to capture contextual relationships within posts, while also incorporating the temporal aspect of user interactions. In contrast, TinvM aggregates all user posts, regardless of timing, to generate user-level predictions about suicidality. This approach classifies users' risk levels by utilizing key features across all posts. The performance of both models was evaluated using the AUC metric. The time-variant model demonstrated superior performance in identifying suicide-related ideations and supportive behaviors, while the time-invariant model proved more effective in predicting suicide-related behaviors and attempts.

Traditional Machine Learning. In Roy et al. (2020), Roy et al. present a machine learning framework designed to predict future occurrences of suicidal ideation by analyzing social media data, specifically from Twitter. They developed an algorithm called Suicide Artificial Intelligence Prediction Heuristic (SAIPH), which can predict whether an individual is expressing suicidal thoughts and is notable for its ability to make forecasts up to three weeks in advance. The study utilized a large dataset of over 3000 Twitter users who had previously posted content related to suicide or self-harm. Regular searches were conducted using keywords such as "I suicide thinking OR planning" to identify users expressing suicidal thoughts. The study also considered the time aspect, focusing on instances of Suicidal Ideation (SI) that lasted at least 120 days. Various time windows (4, 7, 14, and 21 days) were analyzed to predict the future risk of suicidal thoughts.

To predict future risks, the authors combined neural networks and Random Forest models to process and analyze Twitter data. The SAIPH algorithm demonstrated a high level of accuracy in forecasting future suicidal ideation. The authors believe this approach could be valuable in identifying individuals at risk and facilitating personalized interventions. However, the algorithm is designed for Twitter data, and further research is required to determine its applicability to other social media platforms and its ability to track changes in mood over time.

In Mens et al. (2020), Mens et al. analyzed a population-based dataset to predict suicidal thoughts and suicide attempts over a one-year period. The study compared the performance of various machine learning algorithms to assess their predictive capabilities. While the primary focus was on examining baseline risk factors and psychological metrics as predictors of future suicidal behaviors, less attention was given to how these behaviors evolve over time. To predict future suicidal thoughts, the key predictive factors were internal entrapment and perceived burden. In contrast, when predicting suicide attempts during the follow-up period, significant predictors included feelings of defeat, optimism, internal entrapment, and depressive symptoms.

Graph-Based Models. In a different study (Choi et al., 2021), a Deep Graph Neural Network model (GNN) was used to predict suicide risk in young adults. The model chosen for this task was the Graph Isomorphism Network (GIN), a variant of GNN known for its ability to effectively handle graph-structured data, similar to the Weisfeiler–Lehman (WL) test. GIN is particularly adept at classifying various types of graphs and has gained recognition for its outstanding performance in graph classification, achieving a state-of-the-art status.

In this model, the key components were the inclusion of multidimensional questionnaires and the prediction of suicidal thoughts within a 2-week timeframe. It involved deep features related to factors like depression, anxiety, resilience, self-esteem, and clinical-demographic information. The results of this model are promising, especially in the context of remote suicide risk assessment for young adults, which has become increasingly important due to challenges like the COVID-19 pandemic.

4.2.3. Limitations

In the previous sections, we explored various methodologies used to detect suicidal ideation, with a particular focus on Deep Learning techniques. A key factor influencing the effectiveness of these methods is the quality of data used for training the models. However, data availability can present significant limitations. Most studies in this field emphasize the combination of different models, such as LSTM, CNN, and Transformers, to extract more comprehensive features from the data. It is important to note that employing hybrid deep learning models comes with a substantial computational cost and requires access to large datasets for training. Additionally, generalizing results across different datasets or social media platforms can be challenging due to variations in language use and cultural differences.

Haque et al. (2022) highlight limitations in using word embeddings within deep learning models. Their approach may fail to fully capture the semantic meanings of words and phrases, as it represents words as compact vectors without considering their grammatical relationships. Additionally, the method discussed in Ji et al. (2022) relies on sentiment lexicons and topic modeling to extract sentiment- and topic-related risk indicators from relational text encoding. This dependency on external resources can introduce the risk of error propagation.

Moreover, Sawhney et al. (2021) acknowledges a tradeoff between inherent selection bias in the analyzed data and the informed consent of users. This study also recognizes the subjectivity in the interpretation of results, which may vary among individuals, and highlights the potential lack of interpretability of the model to non-experts. Furthermore, it is important to emphasize that the classification of suicidal ideation is a complex task that necessitates precise consideration of ethical and legal implications. Automated systems should be employed in continuity with human experts to safeguard the well-being of individuals at risk of suicide.

Some approaches based on lexicon or semantics may not be accurate enough to understand the subjective nature of personal distress and suicidal risk. These methods might miss important, detailed risk factors that make a person vulnerable to suicide. Additionally, many studies fail to report standard metrics for information extraction and text data mining, like recall, precision, and F-measure. This lack of reporting makes it difficult to assess the quality and generalizability of results, especially when the conclusions are based on a limited sample of manually selected examples.

Traditional Machine Learning methods face several significant limitations. A key challenge is the insufficient exploration of optimal feature selection for each classifier, which can directly impact algorithm performance. Additionally, the high dimensionality of features — stemming from the numerous unique words in a document — can degrade classifier performance. Noise and redundant words further complicate this issue by confusing classifiers and reducing accuracy.

Model evaluation often focuses on specific demographic groups, predominantly young adults, limiting the generalizability of findings to other populations or age groups. Moreover, critical external factors, such as cultural differences and socioeconomic status, are often overlooked, despite their potential influence on suicide risk prediction (Choi et al., 2021).

The authors of Roy et al. (2020) also note limitations in their algorithm's ability to predict suicide risk for individuals experiencing suicidal thoughts for the first time. These cases are rare, posing challenges for neural network-based models due to data imbalance. Furthermore, the algorithm struggles to account for external environmental factors that may alter data patterns over time, further complicating accurate risk prediction.

In the graph-based techniques, we encounter data limitations such as sparse metadata within social networks which constrain the model for constructing graphs and utilizing for suicidal ideation detection. In Sawhney et al. (2022) the focus on analyzing conversation trees from online communities may not fully capture a user's mental state due to the diverse range of comments and their impact on distressed users. The method's emphasis on capturing detailed timing patterns in comment releases that may be challenging for real-time implementation.

In conclusion, the detection of suicidal ideation is a complex task with various challenges and limitations in different research approaches. It is crucial to recognize that the effectiveness of these methods can depend on the specific social media platform and user characteristics. Understanding and addressing these limitations is vital for the progress of this field and the development of more reliable and ethical solutions for identifying individuals at risk of suicide.

4.3. Features

This section provides an overview of the common features employed in studies reviewed for the detection of suicidal ideation. Among the features utilized, text-based attributes play a central role. These include well-established techniques, including Term Frequency–Inverse Document Frequency (TF–IDF), n-grams, LIWC, Bag of Words(BOW), and word embeddings. These textual features serve as foundational elements in the analysis of textual data, aiding in the identification of linguistic patterns associated with suicidal thoughts.

In addition, many studies include emotional and sentiment-related features. One important tool in this category is 'EMPATH', which is designed to help measure and understand the emotional and thematic aspects of text. For example, in the research presented in Chen et al. (2019), sentiment features were used to detect similarities between users' posts. A specific feature called "sentiment profile" was utilized to assess the likeness between a user's current posts and their past ones. This set of features results in a vector of sentiment values related to a user's previous posts. Using the Levenshtein Distance metric, the study calculated the similarity between these vectors, highlighting the usefulness of sentiment features in identifying similarities between users' posts and their potential connection to suicide risk. In a different study (Bitew et al., 2019b), the authors employed a pre-trained model known as DeepMoji (Felbo, Mislove, Søgaard, Rahwan, & Lehmann, 2017) to extract the emotional content embedded within user-generated posts.

Another emerging category of features in this domain focuses on graph-based or social network attributes, as highlighted in Table 5. These features effectively capture the intricate relationships and interactions within social networks, offering valuable insights into the dynamics of detecting suicidal ideation. In the context of individual profile information, which is user's details including age, demographic characteristics, and personal attributes, such as health status. This category finds utilization in research studies that employ questionnaire-based or interview-based platforms to gather user data. Conversely, in social network platforms, access to such comprehensive user information is usually unavailable. For instance, in the research conducted in Van Vuuren et al. (2021), which focuses on the detection of suicidal ideation using a questionnaire-based platform, various features are employed. Some features include the Ask Suicide-Screening-Questionnaire—Revised (ASQ)[1] risk baseline, depression score, frequency of vegetable consumption (1–2 days a week), gender, frequency of hash or weed usage (less than monthly), and the consumption of five or more glasses of alcohol on a daily or near-daily basis.

Ji et al. in Ji et al. (2018) aimed to utilize text features, specifically focusing on what they refer to as the 'topic feature'. This feature is derived from the latent Dirichlet allocation (LDA) algorithm, as introduced by Blei et al. in Blei, Ng, and Jordan (2003). LDA is an unsupervised machine learning algorithm known for its capability to identify underlying topics in a collection of documents. This study used LDA to uncover hidden categories in user-generated posts by analyzing word distributions. Each post was represented as a combination of categories, with probabilities indicating its association with each. By extracting topic features through LDA, researchers captured the themes in both suicidal and non-suicidal posts, incorporating this information into a supervised learning model to enhance classification.

The authors in Shing et al. (2018) used various characteristics to assess the risk of suicide through online posts. These features include:

- Emotion features (NRC): The count of emotion-tokenized lemmas occurring in the post based on the NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2013). The emotions included are anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.
- Mental disease lexicon (mentalDisLex): The maximum count of the post's tokens or lemmas that match entries in the mental disease lexicon introduced (Zirikly, Kumar, & Resnik, 2016).
- Syntactic features: These include the proportion of transitive verbs (out of all verbs), the proportion of active verbs, the proportion of passive verbs, the proportion of active verbs with "I" as subject, the proportion of passive verbs with "I" as subject, and proportion of transitive verbs with "me" or "myself" as object.
- Topic model posteriors: The researchers used Latent Dirichlet Allocation (LDA) to infer a 20-topic model on the training set using each post as a document, in order to use the set of topic posteriors as features.
- Word embeddings: The researchers calculated 300-dimensional embeddings for the entire Reddit corpus using a word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013).

[1] The ASQ is a four-question screening instrument with good content validity. The four questions together assessed major facets of established suicide risk factors.

The feature vector for each user is the average of the feature vectors from the relevant set of user's posts, which differs depending on the task.

Haque et al. (2022) utilized the BOW technique for feature extraction, converting textual data into matrix or vector formats for machine learning applications. This approach aimed to enhance the performance of models in identifying suicidal ideation from Twitter data. BOW has proven to be a highly effective and widely used feature extraction technique in machine learning, especially for text classification tasks. It simplifies high-dimensional textual data by mapping it into lower-dimensional feature sets, reducing computational complexity while enhancing the efficiency of machine learning models. The preprocessing step played a pivotal role in enabling more accurate text analysis for identifying suicidal ideation.

In Liu et al. (2020) features used in the Conditional Random Field (CRF) model for suicidal ideation cause extraction (SICE) from social texts include:

- Word features: The experiment shows that word features worked best for SICE, indicating the importance of analyzing the words used in social texts.
- Part of speech (POS) and dependence relationship (DP) features: These features can be covered by word features to some extent, suggesting that analyzing the linguistic structure of social texts can contribute to SICE.
- Suicidal psychology, emotion, and language feature: These features improve the effectiveness of SICE, highlighting the significance of considering psychological, emotional, and linguistic aspects in understanding suicidal ideation causes.
- Character embeddings: Adding character embeddings to the CRF model significantly improves the extraction using Char-BiLSTM-CRF, indicating the importance of character-level information in SICE.
- Three-word embeddings: Word2vec, ELMo, and BERT were compared as three-word embeddings.

The CRF model used a feature selection strategy based on the greedy method, selecting the best single-type features and examining the combination of all six types of features.

In their study, Gaur et al. (2021) proposed an innovative approach to represent user-generated posts by leveraging ConceptNet—a semantic network known for encoding general knowledge and common sense concepts as interconnected nodes. ConceptNet, detailed in works by Speer, Chin, and Havasi (2017) and Speer and Lowry-Duda (2017), provided the foundational framework for generating vector representations of textual content. These vector representations facilitated the measurement of semantic proximity between n-gram phrases and concepts from medical lexicons. The process accounted for both syntactic properties and contextual usage, offering an understanding of the relationships between textual elements and medical concepts.

4.3.1. Limitations

In suicidal ideation detection, beyond data collection, the primary focus lies in feature engineering, a process intended to improve detection accuracy by creating more meaningful variables for machine learning models. However, this process can be computationally intensive, especially when dealing with a large number of variables. Interestingly, a recent study (Horvath et al., 2021) found that just 29 predictors were sufficient for effective detection, highlighting the efficiency that can be achieved with a carefully selected feature set. This finding suggests that further research into feature engineering could refine its scope and enhance model performance, making it a promising direction for future advancements.

It is crucial to recognize the limitations of linguistic features in detecting suicidal ideation, as they do not account for other significant factors such as demographic details and mental health history. These aspects often remain inaccessible in the context of social networks. Additionally, the effectiveness of LIWC features is constrained by their reliance on a predefined dictionary, which may fail to capture the full range of linguistic differences. For example, in Aldhyani and Alshebami (2022), the LIWC-22 tool used for extracting linguistic features was noted to potentially miss critical aspects of suicidal ideation, such as metaphors or sarcasm. Furthermore, relying solely on textual features may not represent the complex and multifaceted nature of suicidal ideation expressed on social media platforms. It is also worth noting that LIWC-22 is not freely available.

Moreover, some studies have focused exclusively on textual data, neglecting other modalities like images or videos that could provide valuable insights into the detection of suicidal ideation (Haque et al., 2020). These limitations underscore the need for more holistic and inclusive approaches to better understand and identify suicidal behaviors online.

Text-based feature extraction methods like Word2Vec, Doc2Vec, and TF–IDF each have their own limitations. For example, Word2Vec and Doc2Vec are based on the distributional hypothesis, which suggests that words with similar contexts share similar meanings. While effective, this assumption can lead to issues such as polysemy, where words with multiple meanings may be inadequately represented. These models can also struggle when the meaning of a word is highly dependent on the speaker's intent or context.

Similarly, TF–IDF, though widely used, may not fully capture the semantic depth of text. By focusing primarily on word frequency within a corpus, it may miss nuanced meanings. Studies that rely on frequency analysis can limit themselves to counting terms, neglecting the complex insights that can be gained from syntactic and semantic approaches. Additionally, TF–IDF is sensitive to document length and lacks a deeper understanding of semantics. For instance, in Haque et al. (2022), the BOW technique was employed as a feature extraction method. While this improved accuracy, it still could not fully capture the meaning of words, as it mainly focused on how often words appeared in the text.

In Sawhney et al. (2021), recognize that the characteristics they examined may not completely capture the complexity of suicidal thoughts. They suggest adding more features like time-based and network-related attributes to improve the model's performance. They also consider using mixed features that combine word, part-of-speech (POS), and dependency information. However, it is important to note that this approach may not solve the problem of word categorization ambiguity, which could affect the model's accuracy in identifying specific types of words related to suicidal thoughts. In another study, Ramírez-Cifuentes et al. (2020), the authors mention that the lexicon extracted from Reddit may introduce vocabulary biases, especially when translating to Spanish, even though they tried to validate the terms with specialized clinicians.

Feature selection techniques like linear forward selection (LFS) and genetic algorithms (GA) are commonly used to identify relevant features. However, it is important to recognize that these methods do not always guarantee the selection of the optimal features, as pointed out in Shah et al. (2020). As such, carefully selecting the right features is crucial when aiming to identify the most impactful indicators for detecting suicidal thoughts.

5. Summary

This review provides a comprehensive summary of the papers examined, as detailed in Table 6. The studies are organized chronologically, from the earliest to the most recent. In Fig. 4, we have categorized these papers based on their contributions to the detection of suicidal ideation. Some papers focus on predicting suicidal thoughts, while others employ a classification approach. Notably, these studies span a variety of platforms and methodologies, with each carefully considering its data sources and feature sets.

Recent research has made significant strides in identifying gaps and enhancing efforts to reduce suicide risk. However, it is essential to recognize that finding an optimal solution to this complex and multifaceted issue remains an ongoing challenge.

Table 6
An overview of studies on the detection of suicidal ideation across various categories.

Title	Prediction	Classification	Technique	Evaluation metrics	Text features	Emotion features	Social networks features	Individual profile information features	Platform	Year
Ji et al. (2018)	No	Yes	Traditional machine learning	Accuracy, Recall, Precision, F1, AUC	Yes	Yes	No	No	Social Networks/Forums	2018
Shing et al. (2018)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall, F1	Yes	Yes	No	No	Social Networks/Forums	2018
Aladağ et al. (2018)	No	Yes	Traditional machine learning	Accuracy, F1, Precision, Recall	Yes	Yes	No	No	Social Networks/Forums	2018
Sawhney et al. (2018)	No	Yes	Deep learning	Accuracy, Recall, Precision, F1	Yes	No	No	No	Social Networks/Forums	2018
Mishra et al. (2019)	No	Yes	Graph-based models	Micro F1, Precision, Recall	Yes	Yes	Yes	Yes	Social Networks/Forums	2019
Chen et al. (2019)	No	Yes	Traditional machine learning	Recall, Precision, F1	Yes	Yes	No	No	Social Networks/Forums	2019
Morales et al. (2019)	No	Yes	Deep learning	Macro F1	Yes	No	No	Yes	Social Networks/Forums	2019
Chadha and Kaushik (2021)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall	NA	NA	NA	NA	Social Networks/Forums	2019
Matero et al. (2019)	No	Yes	Deep learning	Macro F1, Accuracy	Yes	No	No	Yes	Social Networks/Forums	2019
Ambalavanan et al. (2019)	No	Yes	Deep learning	Macro F1, Accuracy	Yes	No	No	No	Social Networks/Forums	2019
Bitew et al. (2019a)	No	Yes	Traditional machine learning	Precision, Recall, F1	Yes	Yes	No	No	Social Networks/Forums	2019
Allen et al. (2019)	No	Yes	Deep learning	Precision, Recall, F1	Yes	Yes	No	No	Social Networks/Forums	2019
Sawhney et al. (2020)	No	Yes	Deep learning	Macro F1, Recall	Yes	Yes	No	No	Social Networks/Forums	2020
Liu et al. (2020)	No	Yes	Deep learning	Precision, Recall, F1	Yes	Yes	No	No	Social Networks/Forums	2020
Kumar et al. (2020)	No	Yes	Traditional machine learning	Precision, Recall, F1, Accuracy	Yes	No	No	No	Social Networks/Forums	2020
Mens et al. (2020)	Yes	No	Traditional machine learning	AUC, Sensitivity, Specificity, Positive Predictive Value, Balanced Accuracy	Yes	No	No	No	Interviews	2020
Cao et al. (2022)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1, MacroF1	Yes	No	Yes	Yes	Social Networks/Forums	2020
Roy et al. (2020)	Yes	No	Traditional machine learning	AUC	Yes	Yes	No	Yes	Social Networks/Forums	2020
Ophir et al. (2020)	No	Yes	Deep learning	AUC	Yes	No	No	Yes	Social Networks/Forums	2020
Valeriano Valdez et al. (2020)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall, F1	Yes	No	No	No	Social Networks/Forums	2020
Zhu et al. (2020)	No	Yes	Traditional machine learning	Precision, Recall, F1, Accuracy	Yes	No	No	No	Electronic Health Records	2020
de Carvalho et al. (2020)	No	Yes	Deep learning	Precision, Recall, F1, Accuracy	Yes	No	No	No	Social Networks/Forums	2020
Zheng et al. (2020)	Yes	No	Deep learning	AUC, ROC	Yes	No	No	Yes	Electronic Health Records	2020
Haque et al. (2020)	No	Yes	Deep learning	Accuracy, Recall, Precision, F1	Yes	Yes	No	No	Social Networks/Forums	2020
Toledo-Acosta et al. (2020)	No	No	NLP	NA	Yes	Yes	No	No	Questionnaire	2020
Shah et al. (2020)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall, F1, AUC	Yes	No	No	No	Social Networks/Forums	2020
Ramírez-Cifuentes et al. (2020)	No	Yes	Traditional machine learning	Precision, Recall, F1, Accuracy, AUC	Yes	Yes	Yes	No	Social Networks/Forums	2020
Horvath et al. (2021)	No	Yes	Traditional machine learning	AUC, F1	No	No	No	Yes	Interviews	2020
Yao et al. (2020)	No	Yes	Deep learning	Accuracy, Recall, Precision, F1	Yes	No	No	No	Social Networks/Forums	2020
Cohen et al. (2020)	No	Yes	Traditional machine learning	AUC	Yes	No	No	No	Interviews	2020
Tadesse et al. (2019)	No	Yes	Deep learning	Accuracy, F1, Precision, Recall	Yes	Yes	No	No	Social Networks/Forums	2020
Rabani et al. (2020)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall	Yes	No	No	No	Social Networks/Forums	2020
Ma and Cao (2020)	No	Yes	Deep learning	Accuracy, F1	Yes	No	No	No	Social Networks/Forums	2020
Luo, Du, Tao, Xu, and Zhang (2020)	No	No	Deep learning	Jaccard Distance	Yes	No	No	Yes	Social Networks/Forums	2020
Lin et al. (2020)	No	Yes	Traditional machine learning	Accuracy, Sensitivity, Recall, Specificity, Precision, F1, AUC, ROC	No	No	No	Yes	Questionnaire	2020
Sawhney et al. (2021)	Yes	No	Deep learning	FN, FP, Graded Precision, Graded Recall	Yes	Yes	No	No	Social Networks/Forums	2021
Choi et al. (2021)	Yes	No	Graph-based models	Sensitivity, Specificity, Accuracy, AUC	Yes	No	Yes	Yes	Questionnaire	2021
Lekkas et al. (2021)	No	Yes	Traditional machine learning	Sensitivity, Specificity, AUC	Yes	Yes	Yes	No	Interviews	2021
Macalli et al. (2021)	Yes	No	Traditional machine learning	Out-of-Bag error, AUC, Positive Predicted Value, Sensitivity	Yes	No	No	Yes	Questionnaire	2021
Ma (2021)	No	Yes	Deep learning	Accuracy, F1	Yes	Yes	Yes	No	Social Networks/Forums	2021
Ashok Kumar et al. (2021)	No	Yes	Deep learning	F1, AUC-ROC	Yes	No	No	No	Social Networks/Forums	2021
Gaur et al. (2021)	Yes	No	Deep learning	AUC	Yes	No	No	Yes	Social Networks/Forums	2021
Van Vuuren et al. (2021)	Yes	No	Traditional machine learning	AUC, Sensitivity, Specificity, Positive predictive value, Negative predictive value, Balanced accuracy	No	No	No	Yes	Questionnaire	2021
Jacobucci et al. (2021)	No	Yes	Deep learning	AUC	Yes	Yes	No	No	Questionnaire	2021
Yatapala and Kumara (2021)	No	Yes	Deep learning	Accuracy	Yes	No	No	No	Social Networks/Forums	2021
Kim et al. (2021)	No	Yes	Traditional machine learning	Precision, Recall, F1, AUC	No	No	No	Yes	Questionnaire	2021
Nikhileswar et al. (2021)	No	Yes	Deep learning	Accuracy, Recall, Precision, F1	Yes	No	No	No	Social Networks/Forums	2021

(*continued on next page*)

Table 6 (*continued*).

Rabani et al. (2021)	No	Yes	Traditional machine learning	F-Measure	Yes	No	No	No	Social Networks/Forums	2021
Wang et al. (2021)	Yes	No	Deep learning	F1 scores, F2 scores	Yes	Yes	No	No	Social Networks/Forums	2021
Renjith et al. (2022)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1	Yes	No	No	No	Social Networks/Forums	2021
Zhang et al. (2021)	No	Yes	Deep learning	Precision, Recall, F1	Yes	Yes	No	No	Social Networks/Forums	2021
Chadha and Kaushik (2022)	No	Yes	Deep learning	Accuracy, Precision, Recall, Specificity, F1	Yes	No	No	No	Social Networks/Forums	2022
Aldhyani and Alshebami (2022)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1, specificity	Yes	Yes	No	No	Social Networks/Forums	2022
Metzler et al. (2022)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1	Yes	No	No	No	Social Networks/Forums	2022
Sawhney et al. (2022)	No	Yes	Graph-based models	Macro F1, Recall	Yes	No	No	No	Social Networks/Forums	2022
Ji et al. (2022)	No	Yes	Deep learning	Accuracy, F1, and AUC	Yes	No	No	No	Social Networks/Forums	2022
Haque et al. (2022)	No	Yes	Deep learning	Accuracy, F1	Yes	No	No	No	Social Networks/Forums	2022
Kim and Lee (2022)	No	Yes	Traditional machine learning	Precision, recall, F1-value, Accuracy - Sensitivity, Specificity, Positive Predictive value, Negative predictive value, AUC	Yes	No	No	Yes	Questionnaire	2022
Diniz et al. (2022)	No	Yes	Deep learning	Accuracy, Precision, F1, AUC	Yes	No	No	No	Social Networks/Forums	2022
Park and Lee (2022)	No	Yes	Traditional machine learning	Sensitivity, Specificity, Accuracy, F1, AUC, Positive Predictive Value, Negative Predictive Value	Yes	No	No	Yes	Questionnaire	2022
Liu et al. (2022)	No	Yes	Traditional machine learning	Accuracy, F1	Yes	No	No	Yes	Social Networks/Forums	2022
Cohen et al. (2022)	Yes	No	Traditional machine learning	AUC, Brier scores	No	No	No	Yes	Interviews	2022
Ananthakrishnan et al. (2022)	No	Yes	Deep learning	Precision, Recall, F1	Yes	No	No	No	Social Networks/Forums	2022
Chatterjee et al. (2022)	No	Yes	Traditional machine learning	Accuracy, Recall, Precision, F1	Yes	Yes	No	No	Social Networks/Forums	2022
Kim et al. (2022)	No	Yes	Traditional machine learning	AUC, ROC, Sensitivity, Specificity, Accuracy	No	No	No	Yes	Questionnaire	2023
Sakthi et al. (2023)	No	Yes	Deep learning	Sensitivity, Specificity, Precision, Recall, F-score, and Accuracy	Yes	Yes	No	No	Social Networks/Forums	2023
Naseem et al. (2023)	No	Yes	Deep learning	Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2023
Islam et al. (2023)	No	Yes	Deep learning	Accuracy, Precision, Recall, AUC	Yes	No	No	No	Social Networks/Forums	2023
Priyamvada et al. (2023)	No	Yes	Deep learning	Accuracy, Precision, Recall	Yes	No	No	No	Social Networks/Forums	2023
Buddhitha and Inkpen (2023)	No	Yes	Deep learning	Precision, Recall, F1, AUC-ROC	Yes	No	No	No	Social Networks/Forums	2023
Setiawan, Kimberly, Suharjo, and Harefa (2024)	No	Yes	Deep learning	Precision, Recall, F1, AUC-ROC	Yes	No	No	No	Social Networks/Forums	2024
Singh, Dewan, Mittal, Kumar, and Mishra (2024)	No	Yes	Deep learning	Support, Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Kancharapu and Ayyagari (2024)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1-Score	Yes	No	Yes	No	Social Networks/Forums	2024
Ghanadian et al. (2024)	No	Yes	Deep learning	Accuracy, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Mishra, Sucharitha, Siddhartha, Raju, and Mohanty (2024)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Boonyarat, Liew, and Chang (2024)	No	Yes	Deep learning	Precision, Recall, F1-Score	Yes	Yes	No	No	Social Networks/Forums	2024
Choi and Yang (2024)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Herath and Wijayasiriwardhane (2024)	No	Yes	Traditional machine learning	Accuracy, Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Mirtaheri, Greco, and Shahbazian (2024)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1-Score	Yes	No	No	No	Social Networks/Forums	2024
Gorai and Shaw (2024)	No	Yes	Deep learning	Accuracy, Precision, Recall, F1-Score	Yes	Yes	No	No	Social Networks/Forums	2024

5.0.1. Limitations

The reviews discussed in Section 2.3 highlight several limitations. One key limitation is the absence of a generic approach in their methodologies, along with an overemphasis on specific details, such as the explanation of TFIDF. Furthermore, some reviews fail to incorporate the PRISMA framework, which is essential for conducting a thorough and comprehensive literature review. Another limitation is that some reviews analyze only a limited number of papers, which may not provide a complete picture of the existing approaches. Additionally, many of the reviews overlook the specific limitations and gaps within the various approaches.

Given the critical importance of data in detecting suicidal ideation, it is essential to investigate the datasets used in these studies, as well as the methods employed for data collection and annotation. Equally important is the need to focus on the prediction of suicidal ideation, a key component of suicide prevention efforts. Therefore, a comprehensive evaluation of the techniques used in this field is crucial for advancing the understanding and effectiveness of these approaches.

6. Conclusion

In this systematic review, we thoroughly examined the literature on the identification of suicidal thoughts from 2018 to 2024, analyzing 92 papers—a notably large sample compared to other reviews. We developed a comprehensive search strategy based on the PRISMA framework and applied a hierarchical taxonomy for a structured analysis of the relevant studies. The papers were categorized into two primary domains: classification and prediction, providing a distinctive framework for our review.

Our analysis also explored the platforms, methods, and features discussed within each category, revealing significant limitations and highlighting key opportunities to address research gaps in the field. We classified platforms into four main categories: social networks, Electronic Health Records (EHR), questionnaires, and interviews. To provide a deeper understanding, we conducted an in-depth review of the datasets associated with these platforms, offering detailed insights into their size and annotation methods. This approach not only enables comparability across methods but also helps to pinpoint gaps and limitations, making a substantial contribution to this review.

In summary, our findings show that the majority of studies relied heavily on social network data sources due to their widespread availability and ease of access. In contrast, EHRs, questionnaires, and interviews were less accessible and yielded smaller datasets, which posed limitations for methodologies that require large-scale data. This underscores one of the key strengths of our review: the ability to highlight platform-related disparities and their impact on research in this area.

We categorized the methodologies in the reviewed papers into three main groups: traditional machine learning methods (e.g., Random Forest), deep learning techniques, and graph-based approaches. Each of these methods was evaluated separately within both classification and prediction contexts. Our analysis revealed that deep learning methods were the most prominent in the literature, largely due to their superior performance compared to traditional machine learning approaches, although they require larger datasets.

In our analysis of the features essential for effective suicide detection, we categorized them into four primary groups: text-based features, emotional features, social network-derived features, and individual profile information. Our findings showed a strong preference for text-based features, owing to their computational feasibility and the abundance of textual data in most of the studies we reviewed. Emotional features also received significant attention, as they are closely linked to suicidal ideation, reflecting expressions of sentiment and emotion.

We also identified the use of social network features, such as follower counts on platforms like Twitter, emphasizing the role of social dynamics in suicide detection. On the other hand, demographic information such as age and gender played a minimal role, largely due to their limited availability or inadequacy in the datasets, especially in questionnaire or interview-based platforms.

6.1. Future works

Based on our analysis, the following represent key directions for future research that we have identified in this review. Recognizing the importance of timely detection in preventing suicidal ideation, we emphasize the need for accurately predicting future occurrences. Advancing predictive methodologies is, therefore, a key priority for future research and represents one of the central contributions of this review. Within the Method category, our analysis revealed that graph-based techniques remain underutilized, highlighting a promising and underexplored direction for further investigation. Furthermore, in our examination of the Platform category, we underscore the importance of thoughtfully selecting and integrating diverse feature sets to enhance the effectiveness of detection models. By drawing attention to feature preferences and their practical implications, our review offers a comprehensive perspective that can inform and guide future developments in the field.

CRediT authorship contribution statement

Golnaz Nikmehr: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis. **Aritz Bilbao-Jayo:** Supervision. **Aitor Almeida:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We gratefully acknowledge the support of the Ministry of Economy, Industry, and Competitiveness of Spain under Grant No.: INCEPTION(PID2021-128969OB-I00).

Data availability

No data was used for the research described in the article.

References

Abdulsalam, A., & Alhothali, A. (2024). Suicidal ideation detection on social media: a review of machine learning methods. *Social Network Analysis and Mining*, *14*(1), 188. http://dx.doi.org/10.1007/s13278-024-01348-0.

Aladağ, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., & Bingol, H. O. (2018). Detecting suicidal ideation on forums: proof-of-concept study. *Journal of Medical Internet Research*, *20*(6), Article e9840.

Aldhyani, T., & Alshebami, A. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *International Journal of Environmental Research and Public Health*, *2022*, 1–16. http://dx.doi.org/10.3390/ijerph191912635.

Allen, K., Bagroy, S., Davis, A., & Krishnamurti, T. (2019). ConvSent at CLPsych 2019 task a: using post-level sentiment features for suicide risk prediction on reddit. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 182–187).

Ambalavanan, A. K., Jagtap, P. D., Adhya, S., & Devarakonda, M. (2019). Using contextual representations for suicide risk assessment from internet forums. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 172–176).

American Association of Suicidology (2023). Know the signs: How to tell if someone might be suicidal. https://suicidology.org/2023/06/01/know-the-signs-how-to-tell-if-someone-might-be-suicidal/.

AminuIsrael (2020). Predicting-suicide-ideation. URL https://github.com/AminuIsrael/Predicting-Suicide-Ideation.

Ananthakrishnan, G., Jayaraman, A. K., Trueman, T. E., Mitra, S., Abinesh, A., & Murugappan, A. (2022). Suicidal intention detection in tweets using BERT-based transformers. In *2022 international conference on computing, communication, and intelligent systems* (pp. 322–327). IEEE.

Arowosegbe, A., & Oyelade, T. (2023). 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).

Ashok Kumar, J., Trueman, T. E., & Abinesh, A. K. (2021). Suicidal risk identification in social media. *Procedia Computer Science*, *189*, 368–373. http://dx.doi.org/10.1016/j.procs.2021.05.106, URL https://www.sciencedirect.com/science/article/pii/S187705092101231X. AI in Computational Linguistics.

Author, J., & OtherAuthor, J. (2019). Bidirectional long short-term memory networks for natural language processing. *Journal of Artificial Intelligence Research*, *30*, 123–145.

Author, J., & OtherAuthor, J. (2020). Introducing LIWC-22: An updated linguistic inquiry and word count dictionary for psychological text analysis. *Journal of Psycholinguistic Research*, *49*(4), 405–419. http://dx.doi.org/10.1007/s10936-019-09686-1.

Bahamón, M. J., Javela, J. J., Ortega-Bechara, A., Matar-Khalil, S., Ocampo-Flórez, E., Uribe-Alvarado, J. I., et al. (2025). Social determinants and developmental factors influencing suicide risk and self-injury in healthcare contexts. *International Journal of Environmental Research and Public Health*, *22*(3), http://dx.doi.org/10.3390/ijerph22030411, URL https://www.mdpi.com/1660-4601/22/3/411.

Baldini, V., Gnazzo, M., Maragno, M., Biagetti, R., Stefanini, C., Canulli, F., et al. (2025). Suicidal risk among adolescent psychiatric inpatients: the role of insomnia, depression, and social-personal factors. *European Psychiatry*, *68*(1), Article e42. http://dx.doi.org/10.1192/j.eurpsy.2025.29.

Baldini, V., Gnazzo, M., Santangelo, G., D'Agostino, A., Varallo, G., Scorza, M., et al. (2025). Are sleep disturbances a risk factor for suicidal behavior in the first episode of psychosis? Evidence from a systematic review. *Journal of Psychiatric Research*, *185*, 186–193. http://dx.doi.org/10.1016/j.jpsychires.2025.03.053, URL https://www.sciencedirect.com/science/article/pii/S0022395625002079.

Bell, I. H., Nicholas, J., Broomhall, A., Bailey, E., Bendall, S., Boland, A., et al. (2023). The impact of COVID-19 on youth mental health: A mixed methods survey. *Psychiatry Research*, *321*, Article 115082. http://dx.doi.org/10.1016/j.psychres.2023.115082.

Benton, A., Coppersmith, G., & Dredze, M. (2017). Ethical research protocols for social media health research. In *Proceedings of the first ACL workshop on ethics in natural language processing* (pp. 94–102). Valencia, Spain: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W17-1612, URL https://aclanthology.org/W17-1612.

Bhardwaj, T., Gupta, P., Goyal, A., Nagpal, A., & Jha, V. (2022). A review on suicidal ideation detection based on machine learning and deep learning techniques. In *2022 IEEE world AI IoT congress* (pp. 027–031). IEEE.

Bitew, S. K., Bekoulis, G., Deleu, J., Sterckx, L., Zaporojets, K., Demeester, T., et al. (2019a). Predicting suicide risk from online postings in reddit the ugent-idlab submission to the clpysch 2019 shared task a. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 158–161).

Bitew, S. K., Bekoulis, G., Deleu, J., Sterckx, L., Zaporojets, K., Demeester, T., et al. (2019b). Predicting suicide risk from online postings in reddit the UGent-IDLab submission to the CLPysch 2019 shared task A. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 158–161). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W19-3019, URL https://aclanthology.org/W19-3019.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Boonyarat, P., Liew, D., & Chang, Y.-C. (2024). Leveraging enhanced BERT models for detecting suicidal ideation in thai social media content amidst COVID-19. *Information Processing and Management*, *61*(4), Article 103706, URL https://www.sciencedirect.com/science/article/pii/S0306457324000669.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

bryant03 (2021). Sina-weibo-dataset. URL https://github.com/bryant03/Sina-Weibo-Dataset.

Buddhitha, P., & Inkpen, D. (2023). Multi-task learning to detect suicide ideation and mental disorders among social media users. *Frontiers in Research Metrics and Analytics*, *8*, Article 1152535. http://dx.doi.org/10.3389/frma.2023.1152535, Copyright© 2023 Buddhitha and Inkpen.

Burnap, P., Colombo, G., Amery, R., Hodorog, A., & Scourfield, J. (2017). Multi-class machine classification of suicide-related communication on Twitter. *Online Social Networks and Media*, *2*, 32–44. http://dx.doi.org/10.1016/j.osnem.2017.08.001, URL https://www.sciencedirect.com/science/article/pii/S2468696417300605.

Cao, L., Zhang, H., & Feng, L. (2022). Building and using personal knowledge graph to improve suicidal ideation detection on social media. *IEEE Transactions on Multimedia*, *24*, 87–102. http://dx.doi.org/10.1109/TMM.2020.3046867.

Castillo-Sánchez, G., Marques, G., Dorronzoro, E., Rivera-Romero, O., Franco-Martín, M., & De la Torre-Díez, I. (2020). Suicide risk assessment using machine learning and social networks: a scoping review. *Journal of Medical Systems*, *44*(12), 205.

Chadha, A., & Kaushik, B. (2021). A survey on prediction of suicidal ideation using machine and ensemble learning. *The Computer Journal*, *64*(11), 1617–1632.

Chadha, A., & Kaushik, B. (2022). A hybrid deep learning model using grid search and cross-validation for effective classification and prediction of suicidal ideation from social network data. *New Gen. Comput.*, *40*(4), 889–914. http://dx.doi.org/10.1007/s00354-022-00191-1.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Medicine*, *3*(1), 1–11.

Chatterjee, M., Samanta, P., Kumar, P., & Sarkar, D. (2022). Suicide ideation detection using multiple feature analysis from Twitter data. In *2022 IEEE Delhi section conference* (pp. 1–6). IEEE.

Chen, L., Aldayel, A., Bogoychev, N., & Gong, T. (2019). Similar minds post alike: Assessment of suicide risk using a hybrid model. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 152–157). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W19-3018, URL https://aclanthology.org/W19-3018.

Cho, K., Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. http://dx.doi.org/10.3115/v1/D14-1179.

Choi, K., Kim, S., Kim, B.-H., Jeon, H., Kim, J.-H., Jang, J., et al. (2021). Deep graph neural network-based prediction of acute suicidal ideation in young adults. *Scientific Reports*, *11*, http://dx.doi.org/10.1038/s41598-021-95102-7.

Choi, H.-S., & Yang, J. (2024). Innovative use of self-attention-based ensemble deep learning for suicide risk detection in social media posts. *Applied Sciences*, *14*(2), http://dx.doi.org/10.3390/app14020893, URL https://www.mdpi.com/2076-3417/14/2/893.

Cohen, J., Wright-Berryman, J., Rohlfs, L., Trocinski, D., Daniel, L., & Klatt, T. W. (2022). Integration and validation of a natural language processing machine learning suicide risk prediction model based on open-ended interview language in the emergency department. *Frontiers in Digital Health*, *4*, Article 818705.

Cohen, J., Wright-Berryman, J., Rohlfs, L., Wright, D., Campbell, M., Gingrich, D., et al. (2020). A feasibility study using a machine learning suicide risk prediction model based on open-ended interview language in adolescent therapy sessions. *International Journal of Environmental Research and Public Health*, *17*(21), 8187.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, *13*(1), 21–27.

de Carvalho, V. F., Giacon, B., Nascimento, C., & Nogueira, B. M. (2020). Machine learning for suicidal ideation identification on Twitter for the portuguese language. In *Intelligent systems: 9th Brazilian conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, proceedings, part i* (pp. 536–550). Berlin, Heidelberg: Springer-Verlag, http://dx.doi.org/10.1007/978-3-030-61377-8_37.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. CoRR abs/1810.04805. arXiv:1810.04805. URL http://arxiv.org/abs/1810.04805.

Diniz, E. J., Fontenele, J. E., de Oliveira, A. C., Bastos, V. H., Teixeira, S., Rabêlo, R. L., et al. (2022). Boamente: A natural language processing-based digital phenotyping tool for smart monitoring of suicidal ideation. *vol. 10*, In *Healthcare* (p. 698). MDPI, 4.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1615–1625). Copenhagen, Denmark: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D17-1169, URL https://aclanthology.org/D17-1169.

Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., et al. (2019a). Knowledge-aware assessment of severity of suicide risk for early intervention. In *The world wide web conference* (pp. 514–525). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3308558.3313698.

Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., et al. (2019b). Reddit C-SSRS suicide dataset. URL https://zenodo.org/records/2667859.

Gaur, M., Aribandi, V., Alambo, A., Kursuncu, U., Thirunarayan, K., Beich, J., et al. (2021). Characterization of time-variant and time-invariant assessment of suicidality on reddit using C-SSRS. *PLoS One*, *16*(5), Article e0250448.

Ghanadian, H., Nejadgholi, I., & Al Osman, H. (2024). Socially aware synthetic data generation for suicidal ideation detection using large language models. *IEEE Access*, *PP*, http://dx.doi.org/10.1109/ACCESS.2024.3358206.

Goel, R., & Digalwar, M. (2024). Suicidal thought detection using max voting ensemble technique. *Procedia Computer Science*, *235*, 2587–2598. http://dx.doi.org/10.1016/j.procs.2024.04.244, URL https://www.sciencedirect.com/science/article/pii/S1877050924009207. International Conference on Machine Learning and Data Engineering (ICMLDE 2023).

Gorai, J., & Shaw, D. K. (2024). A BERT-encoded ensembled CNN model for suicide risk identification in social media posts. *Neural Computing and Applications*, *36*(18), 10955–10970. http://dx.doi.org/10.1007/s00521-024-09642-w, Published June 1, 2024.

Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. *Technologies*, *10*(3), http://dx.doi.org/10.3390/technologies10030057, URL https://www.mdpi.com/2227-7080/10/3/57.

Haque, F., Un Nur, R., Jahan, S., Mahmud, Z., & Shah, F. (2020). A transformer based approach to detect suicidal ideation using pre-trained language models. http://dx.doi.org/10.1109/ICCIT51783.2020.9392692.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Herath, S., & Wijayasiriwardhane, T. (2024). A social media intelligence approach to predict suicidal ideation from sinhala facebook posts. (pp. 1–6). http://dx.doi.org/10.1109/SCSE61872.2024.10550899.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Horvath, A., Dras, M., Lai, C. C., & Boag, S. (2021). Predicting suicidal behavior without asking about suicidal ideation: machine learning and the role of borderline personality disorder criteria. *Suicide and Life-Threatening Behavior*, *51*(3), 455–466.

Huang, X., Ribeiro, J. D., Musacchio, K. M., & Franklin, J. C. (2017). Demographics as predictors of suicidal thoughts and behaviors: A meta-analysis. *PLoS One*, *12*(7), Article e0180793.

Ilic, M., & Ilic, I. (2022). Worldwide suicide mortality trends (2000–2019): A joinpoint regression analysis. *World Journal of Psychiatry*, *12*(8), 1044–1060. http://dx.doi.org/10.5498/wjp.v12.i8.1044, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9476842/. Epub ahead of print, August 19, 2022.

Islam, M. R., Kowsar Hossain Sakib, M., Prome, S. A., Wang, X., Ulhaq, A., Sanin, C., et al. (2023). Machine learning with explainability for suicide ideation detection from social media data. In *2023 10th international conference on behavioural and social computing* (pp. 1–6). http://dx.doi.org/10.1109/BESC59560.2023.10386773.

Jacobucci, R., Ammerman, B. A., & Tyler Wilcox, K. (2021). The use of text-based responses to improve our understanding and prediction of suicide risk. *Suicide and Life-Threatening Behavior*, *51*(1), 55–64.

Ji, S., Li, X., Huang, Z., & Cambria, E. (2022). Suicidal ideation and mental disorder detection with attentive relation networks. *Neural Computing and Applications*, *34*(13), 10309–10319. http://dx.doi.org/10.1007/s00521-021-06208-y.

Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, *8*(1), 214–226.

Ji, S., Yu, C. P., Fung, S.-f., Pan, S., Long, G., & Cong, G. (2018). Supervised learning for suicidal ideation detection in online user content. *Complex*, *2018*, http://dx.doi.org/10.1155/2018/6157249.

Kancharapu, R., & Ayyagari, S. N. (2024). Suicidal ideation prediction based on social media posts using a GAN-infused deep learning framework with genetic optimization and word embedding fusion. *International Journal of Information Technology*, *16*(4), 2577–2593. http://dx.doi.org/10.1007/s41870-023-01725-6, Published April 1, 2024.

Khalaf, A. M., Alubied, A. A., Khalaf, A. M., & Rifaey, A. A. (2023). The impact of social media on the mental health of adolescents and Young adults: A systematic review. *Cureus*, *15*(8), Article e42990. http://dx.doi.org/10.7759/cureus.42990, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10476631/. Epub August 5, 2023.

Kim, J., Gwak, D., Kim, S., & Gang, M. (2022). Identifying the suicidal ideation risk group among older adults in rural areas: Developing a predictive model using machine learning methods. *Journal of Advanced Nursing*, *79*, n/a–n/a. http://dx.doi.org/10.1111/jan.15549.

Kim, S., & Lee, K. (2022). The effectiveness of predicting suicidal ideation through depressive symptoms and social isolation using machine learning techniques. *Journal of Personalized Medicine*, *12*(4), 516.

Kim, S., Lee, H.-K., & Lee, K. (2021). Detecting suicidal risk using MMPI-2 based on machine learning algorithm. *Scientific Reports*, *11*(1), 15310.

Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*: *Tech. rep. EBSE 2007-001*, Keele University and Durham University Joint Report.

Kumar, R., Rao, K., Nayak, S., & Chandra, R. (2020). Suicidal ideation prediction in twitter data using machine learning techniques. *Journal of Interdisciplinary Mathematics*, *23*, 117–125. http://dx.doi.org/10.1080/09720502.2020.1721674.

Kusuma, K., Larsen, M., Quiroz, J. C., Gillies, M., Burnett, A., Qian, J., et al. (2022). The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: A meta-analysis and systematic review. *Journal of Psychiatric Research*.

Lasri, S., et al. (2022). Suicide ideation detection on social networks: Short literature review. *Procedia Computer Science*, *215*, 713–721.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lekkas, D., Klein, R. J., & Jacobson, N. C. (2021). Predicting acute suicidal ideation on instagram using ensemble machine learning models. *Internet Interventions*, *25*, Article 100424.

Lin, G.-M., Nagamine, M., Yang, S.-N., Tai, Y.-M., Lin, C., & Sato, H. (2020). Machine learning based suicide ideation prediction for military personnel. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 1907–1916.

Liu, D., Fu, Q., Wan, C., Liu, X., Jiang, T., Liao, G., et al. (2020). Suicidal ideation cause extraction from social texts. *IEEE Access*, *8*, 169333–169351.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach.

Liu, J., Shi, M., & Jiang, H. (2022). Detecting suicidal ideation in social media: An ensemble method based on feature fusion. *International Journal of Environmental Research and Public Health*, *19*(13), 8197.

Lopez-Castroman, J., Moulahi, B., Azé, J., Bringay, S., Deninotti, J., Guillaume, S., et al. (2019). Mining social networks to improve suicide prevention: A scoping review. *Journal of Neuroscience Research*, *98*, http://dx.doi.org/10.1002/jnr.24404.

Luo, J., Du, J., Tao, C., Xu, H., & Zhang, Y. (2020). Exploring temporal suicidal behavior patterns on social media: Insight from Twitter analytics. *Health Informatics Journal*, *26*(2), 738–752.

Ma, Y. (2021). Social media-based suicide risk detection via social interaction and posted content. In *ICAIIS 2021*, *2021 2nd international conference on artificial intelligence and information systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3469213.3470345.

Ma, Y., & Cao, Y. (2020). Dual attention based suicide risk detection on social media. In *2020 IEEE international conference on artificial intelligence and computer applications* (pp. 637–640). IEEE.

Macalli, M., Navarro, M., Orri, M., Tournier, M., Thiébaut, R., Côté, S., et al. (2021). A machine learning approach for predicting suicidal thoughts and behaviours among college students. *Scientific Reports*, *11*, http://dx.doi.org/10.1038/s41598-021-90728-z.

Matero, M., Idnani, A., Son, Y., Giorgi, S., Vu, H., Zamani, M., et al. (2019). Suicide risk assessment with multi-level dual-context language and BERT. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 39–44).

Mens, K., Schepper, C., Wijnen, B., Koldijk, S., Schnack, H., de Looff, P., et al. (2020). Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study. *Journal of Affective Disorders*, *271*, http://dx.doi.org/10.1016/j.jad.2020.03.081.

Metzler, H., Baginski, H., Niederkrotenthaler, T., & Garcia, D. (2022). Detecting potentially harmful and protective suicide-related content on Twitter: Machine learning approach. *Journal of Medical Internet Research*, *24*, Article e34705. http://dx.doi.org/10.2196/34705.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Mirtaheri, S. L., Greco, S., & Shahbazian, R. (2024). A self-attention TCN-based model for suicidal ideation detection from social media posts. *Expert Systems with Applications*, *255*, Article 124855. http://dx.doi.org/10.1016/j.eswa.2024.124855, URL https://www.sciencedirect.com/science/article/pii/S0957417424017226.

Mishra, R., Prakhar Sinha, P., Sawhney, R., Mahata, D., Mathur, P., & Ratn Shah, R. (2019). SNAP-BATNET: Cascading author profiling and social network graphs for suicide ideation detection on social media. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: student research workshop* (pp. 147–156). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/N19-3019, URL https://aclanthology.org/N19-3019.

Mishra, T. K., Sucharitha, G., Siddhartha, N., Raju, B., & Mohanty, S. N. (2024). Prediction of suicidal behaviour among the users on social media using NLP and ML. In *2024 international conference on emerging systems and intelligent computing* (pp. 74–81). http://dx.doi.org/10.1109/ESIC60604.2024.10481588.

Mohammad, S., & Turney, P. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*, http://dx.doi.org/10.1111/j.1467-8640.2012.00460.x.

Morales, M., Dey, P., Theisen, T., Belitz, D., & Chernova, N. (2019). An investigation of deep learning systems for suicide risk assessment. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 177–181). Minneapolis, Minnesota: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W19-3023, URL https://aclanthology.org/W19-3023.

Naseem, U., Kim, J., Khushi, M., & Dunn, A. (2023). Graph-based hierarchical attention network for suicide risk detection on social media. In *WWW '23 companion, Companion proceedings of the ACM web conference 2023* (pp. 995–1003). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3543873.3587587.

Nikhileswar, K., Vishal, D., Sphoorthi, L., & Fathimabi, S. (2021). Suicide ideation detection in social media forums. In *2021 2nd international conference on smart electronics and communication* (pp. 1741–1747). IEEE.

Oexle, N., Waldmann, T., Staiger, T., Xu, Z., & Rüsch, N. (2018). Mental illness stigma and suicidality: the role of public and individual stigma. *Epidemiology and Psychiatric Sciences*, *27*(2), 169–175. http://dx.doi.org/10.1017/S2045796016000949.

Ophir, Y., Tikochinski, R., Asterhan, C., Sisso, I., & Reichart, R. (2020). Deep neural networks detect suicide risk from textual facebook posts. *Scientific Reports*, *10*, http://dx.doi.org/10.1038/s41598-020-73917-0.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, Article 105906. http://dx.doi.org/10.1016/j.ijsu.2021.105906, URL https://www.sciencedirect.com/science/article/pii/S1743919121000406.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, http://dx.doi.org/10.1136/bmj.n160, URL https://www.bmj.com/content/372/bmj.n160. arXiv:https://www.bmj.com/content/372/bmj.n160.full.pdf.

Park, H., & Lee, K. (2022). A machine learning approach for predicting wage workers' suicidal ideation. *Journal of Personalized Medicine*, *12*(6), 945.

Priyamvada, B., Singhal, S., Nayyar, A., Jain, R., Goel, P., Rani, M., et al. (2023). Stacked CNN - LSTM approach for prediction of suicidal ideation on social media. *Multimedia Tools and Applications*, *82*(18), 27883–27904. http://dx.doi.org/10.1007/s11042-023-14431-z.

Rabani, S. T., Khan, Q. R., & Khanday, A. M. U. D. (2020). Detection of suicidal ideation on Twitter using machine learning & ensemble approaches. *Baghdad Science Journal*, *17*(4), 1328–1328.

Rabani, S. T., Khan, Q. R., & Khanday, A. (2021). A novel approach to predict the level of suicidal ideation on social networks using machine and ensemble learning. *ICTACT Journal of Soft Computing*, *11*(2), 2288–2293.

Ramírez-Cifuentes, D., Freire, A., Baeza-Yates, R., Puntí, J., Medina-Bravo, P., Velazquez, D. A., et al. (2020). Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *Journal of Medical Internet Research*, *22*(7), Article e17758.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. CoRR abs/1908.10084. arXiv:1908.10084. URL http://arxiv.org/abs/1908.10084.

Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Journal of King Saud University-Computer and Information Sciences*, *34*(10), 9564–9575.

Richardson, C., Robb, K. A., McManus, S., & O'Connor, R. C. (2023). Psychosocial factors that distinguish between men and women who have suicidal thoughts and attempt suicide: findings from a national probability sample of adults. *Psychological Medicine*, *53*(7), 3133–3141.

Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., & Kaminsky, Z. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digital Medicine*, *3*, http://dx.doi.org/10.1038/s41746-020-0287-6.

Sakthi, U., Chen, T. M., & Sathiyanarayanan, M. (2023). CyberHelp: Sentiment analysis on social media data using deep belief network to predict suicidal ideation of students. In *2023 international conference on intelligent data communication technologies and internet of things* (pp. 206–211). http://dx.doi.org/10.1109/IDCIoT56793.2023.10053425.

Salton, G., & McGill, M. J. (1992). TF-IDF: Term frequency-inverse document frequency. In *Proceedings of the 15th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 21–30). ACM.

Sawhney, R., Agarwal, S., Neerkaje, A. T., Aletras, N., Nakov, P., & Flek, L. (2022). Towards suicide ideation detection through online conversational context. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1716–1727). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3477495.3532068.

Sawhney, R., Joshi, H., Gandhi, S., & Shah, R. R. (2020). A time-aware transformer based model for suicide ideation detection on social media. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 7685–7697). Online: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.emnlp-main.619, URL https://aclanthology.org/2020.emnlp-main.619.

Sawhney, R., Joshi, H., Gandhi, S., & Shah, R. R. (2021). Towards ordinal suicide ideation detection on social media. In *Proceedings of the 14th ACM international conference on web search and data mining* (pp. 22–30). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3437963.3441805.

Sawhney, R., Manchanda, P., Mathur, P., Shah, R., & Singh, R. (2018). Exploring and learning suicidal ideation connotations on social media with deep learning. In *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 167–175).

Schafer, K. M., Kennedy, G., Gallyer, A., & Resnik, P. (2021). A direct comparison of theory-driven and machine learning prediction of suicide: A meta-analysis. *PLoS One*, *16*(4), 1–23. http://dx.doi.org/10.1371/journal.pone.0249833.

Setiawan, M. D., Kimberly, S., Suharjo, F. A., & Harefa, J. (2024). Transfer learning using indobert with long short-term memory classifier for detection of suicide ideation themed Indonesian Twitter posts. In *2024 4th international conference of science and information technology in smart administration* (pp. 276–281). http://dx.doi.org/10.1109/ICSINTESA62455.2024.10747816.

Shah, F., Jahan, S., Haque, F., Mahmud, Z., Ragib, U., & Un Nur, R. (2020). A hybridized feature extraction approach to suicidal ideation detection from social media post. http://dx.doi.org/10.1109/TENSYMP50017.2020.9230733.

Shing, H.-C., Nair, S., Zirikly, A., Friedenberg, M., Daumé, H., III, & Resnik, P. (2018). Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic* (pp. 25–36). New Orleans, LA: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W18-0603, URL https://aclanthology.org/W18-0603.

Singh, M., Dewan, P., Mittal, M., Kumar, A., & Mishra, A. (2024). Suicide detection using Twitter sentiment analysis. In *2024 11th international conference on computing for sustainable global development* (pp. 808–813). http://dx.doi.org/10.23919/INDIACom61295.2024.10499165.

Speer, R., Chin, J., & Havasi, C. (2017). Conceptnet 5.5: An open multilingual graph of general knowledge. *vol. 31*, In *Proceedings of the AAAI conference on artificial intelligence*. 1.

Speer, R., & Lowry-Duda, J. (2017). Conceptnet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge. arXiv preprint arXiv:1704.03560.

Suicide notes. (2020). URL https://www.kaggle.com/datasets/mohanedmashaly/suicide-notes.

Suicide and depression detection. (2021). URL https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. *Algorithms*, *13*(1), 7.

Toledo-Acosta, M., Barreiro, T., Reig-Alamillo, A., Müller, M., Aroca Bisquert, F., Barrigon, M. L., et al. (2020). Cognitive emotional embedded representations of text to predict suicidal ideation and psychiatric symptoms. *Mathematics*, *8*(11), 2088.

Valeriano Valdez, K., Condori-Larico, A., & Sulla-Torres, J. (2020). Detection of suicidal intent in spanish language social networks using machine learning. *International Journal of Advanced Computer Science and Applications*, *11*, http://dx.doi.org/10.14569/IJACSA.2020.0110489.

Van Vuuren, C., van Mens, K., de Beurs, D., Lokkerbol, J., van der Wal, M., Cuijpers, P., et al. (2021). Comparing machine learning to a rule-based approach for predicting suicidal behavior among adolescents: Results from a longitudinal population-based survey. *Journal of Affective Disorders*, *295*, 1415–1420.

viniciosfaustino (2020). tweet-collector. URL https://github.com/viniciosfaustino/tweet-collector.

Wang, N., Luo, F., Shivtare, Y., Badal, V. D., Subbalakshmi, K., Chandramouli, R., et al. (2021). Learning models for suicide prediction from social media posts. arXiv preprint arXiv:2105.03315.

World Health Organization (2021). *Global health estimates*: *Tech. rep.*, World Health Organization, URL https://www.who.int/data/global-health-estimates.

World Health Organization (2022). *WHO urges more effective prevention of injuries and violence causing 1 in 12 deaths worldwide*: *Tech. rep.*, World Health Organization, URL https://www.who.int/news/item/29-11-2022-who-urges-more-effective-prevention-of-injuries-and-violence–causing-1-in-12-deaths-worldwide.

Yao, H., Rashidian, S., Dong, X., Duanmu, H., Rosenthal, R. N., & Wang, F. (2020). Detection of suicidality among opioid users on reddit: machine learning–based approach. *Journal of Medical Internet Research*, *22*(11), Article e15293.

Yatapala, K., & Kumara, B. (2021). Detection of suicide ideation in Twitter using ANN. In *2021 6th international conference on information technology research* (pp. 1–5). IEEE.

Yeskuatov, E., Chua, S.-L., & Foo, L. K. (2022). Leveraging reddit for suicidal ideation detection: A review of machine learning and natural language processing techniques. *International Journal of Environmental Research and Public Health*, *19*(16), http://dx.doi.org/10.3390/ijerph191610347, URL https://www.mdpi.com/1660-4601/19/16/10347.

Zhang, T., Schoene, A. M., & Ananiadou, S. (2021). Automatic identification of suicide notes with a transformer-based deep learning model. *Internet Interventions*, *25*, Article 100422.

Zheng, L., Wang, O., Hao, S., Ye, C., Liu, M., Xia, M., et al. (2020). Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Translational Psychiatry*, *10*, http://dx.doi.org/10.1038/s41398-020-0684-2.

Zhu, H., Xia, X., Yao, J., Fan, H., Wang, Q., & Gao, Q. (2020). Comparisons of different classification algorithms while using text mining to screen psychiatric inpatients with suicidal behaviors. *Journal of Psychiatric Research*, *124*, http://dx.doi.org/10.1016/j.jpsychires.2020.02.019.

Zirikly, A., Kumar, V., & Resnik, P. (2016). The GW/UMD clpsych 2016 shared task system. (pp. 166–170). http://dx.doi.org/10.18653/v1/W16-0321.

Zirikly, A., Resnik, P., Uzuner, o., & Hollingshead, K. (2019). UMD reddit suicidality dataset. URL https://users.umiacs.umd.edu/~resnik/umd_reddit_suicidality_dataset.html.