International Journal of Biological and Pharmaceutical Sciences Archive

ISSN: 0799-6616 (Online)

Journal homepage: https://ijbpsa.com/

(REVIEW ARTICLE)

AI in Crisis Prediction and Prevention: Leveraging Predictive Analytics for Suicide Risk and Emotional Distress Management

Adekola George Adepoju [1,*], Daniel Adeyemi Adepoju [1], Daniel K. Cheruiyot [1], Samuel Adebowale Adepoju [2], Alexander Audu Obaje [3], John Adeleye Adefiwitan [4] and Babatunde Samuel Omotoye [4]

[1] *Department of Health Informatics, Indiana University Indianapolis, Indiana, USA.*
[2] *Department of Computer Science, Joseph Ayo Babalola University, Nigeria.*
[3] *Dr Chetty Health Centre, Seychelles.*
[4] *Federal Medical Centre, Abeokuta, Ogun State, Nigeria.*

International Journal of Biological and Pharmaceutical Sciences Archive, 2025, 10(02), 049-064

Publication history: Received on 05 September 2025; revised on 16 October 2025; accepted on 18 October 2025

Article DOI: https://doi.org/10.53771/ijbpsa.2025.10.2.0078

Abstract

The growing incidence of suicide and emotional distress across societies highlights an urgent public health challenge that requires effective and scalable preventive strategies. In recent years, artificial intelligence (AI) has emerged as a promising tool for enhancing crisis prediction and prevention through the use of predictive analytics. This review brings together current developments in the application of machine learning, natural language processing, and deep learning techniques that are capable of analyzing large and varied data sources, including electronic health records, social media activity, and data from wearable devices. These technologies enable the early identification of individuals at elevated risk of suicide or severe emotional distress with levels of precision that were previously unattainable. Beyond risk detection, AI-driven platforms are increasingly being integrated into mental health services to provide real-time support through digital companions, automated chat systems, and adaptive treatment recommendations that can be tailored to individual needs. Such innovations hold particular promise in regions with limited access to mental health professionals, offering a means to bridge gaps in care delivery. At the same time, the integration of AI into mental health care presents important challenges. Concerns about data representativeness, the interpretability of complex algorithms, and the transferability of models across diverse cultural and social settings need to be addressed to ensure fairness and reliability. Ethical considerations are equally pressing, particularly around safeguarding patient privacy, maintaining autonomy in care decisions, and building public trust in AI-assisted systems. The review emphasizes that while AI cannot replace human judgment in mental health care, it can serve as a powerful complement when developed within transparent and ethically sound frameworks. By fostering interdisciplinary collaboration among clinicians, data scientists, ethicists, and policymakers, AI-driven approaches can contribute to reducing suicide rates and mitigating emotional distress. In doing so, they pave the way for a more proactive, inclusive, and globally responsive mental health ecosystem.

Keywords: Artificial Intelligence; Suicide Risk Prediction; Emotional Distress; Predictive Analytics; Machine Learning; Natural Language Processing; Mental Health; Ethical AI

1. Introduction

Mental health crises, particularly those involving suicide risk and emotional distress, represent a pressing global challenge that demands innovative approaches for early detection and intervention. Advances in artificial intelligence (AI) and predictive analytics offer promising tools to address these issues by analyzing vast datasets to identify patterns indicative of vulnerability. This review explores how AI technologies can enhance crisis prediction and prevention, focusing on their application in managing suicide risk and emotional distress through data-driven insights.

* Corresponding author: Adekola George Adepoju; Email: georgekola99@gmail.com

1.1. Background and Significance

Suicide and emotional distress continue to impose a substantial burden on public health systems worldwide, with alarming statistics underscoring the urgency for effective prevention strategies. According to the World Health Organization [1], more than 720,000 individuals die by suicide annually, making it the third leading cause of death among those aged 15 to 29 years. This figure highlights the disproportionate impact on younger populations, where developmental stressors, social pressures, and limited access to mental health resources exacerbate risks. Furthermore, the organization notes that nearly three-quarters of these deaths occur in low- and middle-income countries, where healthcare infrastructure often lacks the capacity for widespread screening and intervention. These trends emphasize the need for scalable solutions that can bridge gaps in traditional mental health services, particularly in resource-constrained settings where emotional distress may go undetected until a crisis point.

In addition to the direct loss of life, the ripple effects of suicide and unmanaged emotional distress extend to families, communities, and economies, amplifying the significance of proactive measures. Findings from Lejeune et al. [2] in their systematic review indicate that conventional risk assessment methods, reliant on clinician judgment and self-reported questionnaires, often fail to capture subtle indicators, leading to underestimation of threats. This gap is particularly evident in populations with comorbid conditions, such as depression or substance abuse, where emotional distress can manifest in complex ways. The economic toll is equally staggering, with indirect costs from lost productivity and healthcare expenditures running into billions annually, as highlighted in global reports. By framing suicide not merely as an individual tragedy but as a preventable public health issue, researchers advocate for integrated approaches that leverage technology to enhance early warning systems and reduce stigma associated with seeking help.

The significance of addressing these crises is further amplified by the rising prevalence of emotional distress in the wake of global events like pandemics and socioeconomic upheavals, which have intensified mental health vulnerabilities. Perna et al. [3] point out in their review that psychiatric populations, including those with mood disorders, are at heightened risk, with factors like previous attempts and symptom severity playing pivotal roles. This underscores the importance of understanding the multifaceted nature of risk, encompassing biological, psychological, and social dimensions. Effective prevention requires not only identifying at-risk individuals but also implementing timely interventions that can alter trajectories toward positive outcomes. As such, the background of this issue calls for a paradigm shift toward data-informed strategies that can democratize access to mental health support and mitigate the long-term societal impacts.

1.2. Role of AI and Predictive Analytics

Artificial intelligence has emerged as a transformative force in mental health, particularly through its capacity to process complex data for crisis prediction, offering advantages over traditional methods in accuracy and scalability. In a comprehensive analysis by Ehtemam et al. [4], machine learning algorithms demonstrated varying performance in suicide risk prediction, with random forest models achieving up to 94% accuracy in clinical studies. This role extends to predictive analytics, where AI sifts through electronic health records, social media, and behavioral data to forecast potential distress episodes. By identifying patterns that humans might overlook, such as linguistic cues or activity fluctuations, AI enables real-time monitoring that can preempt crises. This integration marks a departure from reactive care models, positioning AI as a proactive ally in mental health management.

The application of predictive analytics within AI frameworks has shown particular promise in handling the nuances of emotional distress, where subtle shifts in mood or behavior can signal impending risks. Mansoor and Ansari [5] in their prospective study illustrated how AI-powered analysis of social media achieved 89.3% accuracy in detecting mental health crises days in advance, across multiple languages and platforms. This capability relies on multimodal deep learning, combining natural language processing with temporal trends to capture early indicators like increased isolation or negative sentiment. Such tools not only enhance detection but also support personalized interventions, tailoring responses based on individual profiles. However, the role of AI must be balanced with human oversight to ensure interpretations align with clinical contexts, avoiding overreliance on algorithmic outputs.

Beyond detection, AI's role in predictive analytics facilitates broader systemic improvements in crisis prevention, including resource allocation and policy development. Gholi et al. [6] in their systematic review of reviews concluded that AI, especially through machine learning, outperforms clinical judgment in predicting suicidal ideation and attempts by analyzing diverse datasets. This has implications for high-risk groups, where predictive models can integrate factors like comorbidities and environmental stressors for more robust forecasts. The evolving landscape suggests AI can democratize mental health care, making it accessible in underserved areas, though challenges like data privacy remain critical. Ultimately, the synergy between AI and predictive analytics holds potential to reduce suicide rates by enabling earlier, more targeted interventions.

Advancements in AI also underscore its role in fostering interdisciplinary collaboration, where predictive analytics bridges gaps between technology and psychology. According to Kumar and Smith [7], current trends in AI for mental health emphasize early detection and personalized plans, with predictive tools identifying emotional distress before escalation. This involves ethical considerations, such as bias mitigation, to ensure equitable outcomes. As AI evolves, its predictive capabilities could integrate with wearable tech for continuous monitoring, further enhancing prevention efforts. This multifaceted role positions AI as essential for addressing the complexities of suicide risk and emotional distress in modern healthcare.

1.3. Objectives and Scope of the Review

The primary objective of this review is to critically evaluate the integration of AI-driven predictive analytics in forecasting and mitigating suicide risk and emotional distress, synthesizing evidence from recent studies to highlight effective methodologies. By examining key technologies and their applications, the review aims to provide insights into how these tools can improve outcomes in diverse populations. This includes assessing the accuracy of models in real-world settings, as noted in Lejeune et al. [2], where AI showed promising AUC values for risk prediction. Ultimately, the goal is to inform practitioners and policymakers on leveraging AI for more responsive mental health systems.

A secondary objective is to identify gaps in current research, such as the need for external validation and ethical frameworks, to guide future investigations. Perna et al. [3] emphasize the importance of intuitive AI tools for clinical adoption, which this review scopes by focusing on psychiatric contexts. The scope encompasses studies from 2014 onward, prioritizing peer-reviewed works on machine learning and natural language processing, while excluding non-predictive AI applications like general diagnostics. This targeted approach ensures relevance to crisis management, addressing both individual and population-level interventions.

Furthermore, the review seeks to explore the broader implications of AI in promoting equity in mental health care, particularly in low-resource environments. Drawing from Mansoor and Ansari [5], the objectives include evaluating multilingual and platform-agnostic models for global applicability. The scope limits to verifiable, published sources, avoiding speculative technologies, to maintain rigor. By delineating these objectives, the review contributes to a cohesive understanding of AI's potential and limitations in crisis prediction.

In delineating the scope, this review also aims to propose recommendations for interdisciplinary collaboration, integrating insights from Ehtemam et al. [4] on algorithm performance. This encompasses ethical considerations and scalability, ensuring the analysis remains comprehensive yet focused on predictive analytics for suicide and distress management.

2. AI Technologies in Crisis Prediction

The integration of artificial intelligence (AI) technologies in predicting mental health crises, particularly suicide risk and emotional distress, has revolutionized the field by enabling precise, data-driven risk assessments. This section explores the core AI methodologies—machine learning, natural language processing (NLP), and deep learning—that underpin predictive analytics for crisis management. By leveraging diverse datasets, these technologies identify patterns and risk factors with unprecedented accuracy, offering scalable solutions for early intervention.

2.1 Machine Learning Models

Machine learning (ML) models have become foundational in predicting suicide risk and emotional distress due to their ability to process complex, multidimensional datasets. Supervised learning algorithms, such as logistic regression and random forests, are widely used for risk stratification, analyzing variables like clinical history, demographic factors, and behavioral patterns. In a pivotal study by Ehtemam et al. [4], random forest models achieved a predictive accuracy of 94% for suicide risk in clinical settings, outperforming traditional risk assessment tools. These models excel in handling structured data from electronic health records (EHRs), where they identify key predictors like prior suicide attempts or medication adherence. To visualize the process of developing robust ML models for suicide risk, Figure 1 outlines a schematic workflow that incorporates data preprocessing, model evaluation, and explainability, highlighting its application to health administrative data for enhanced predictive accuracy. Their strength lies in iterative learning, allowing continuous refinement as new data becomes available, which is critical for dynamic mental health contexts.

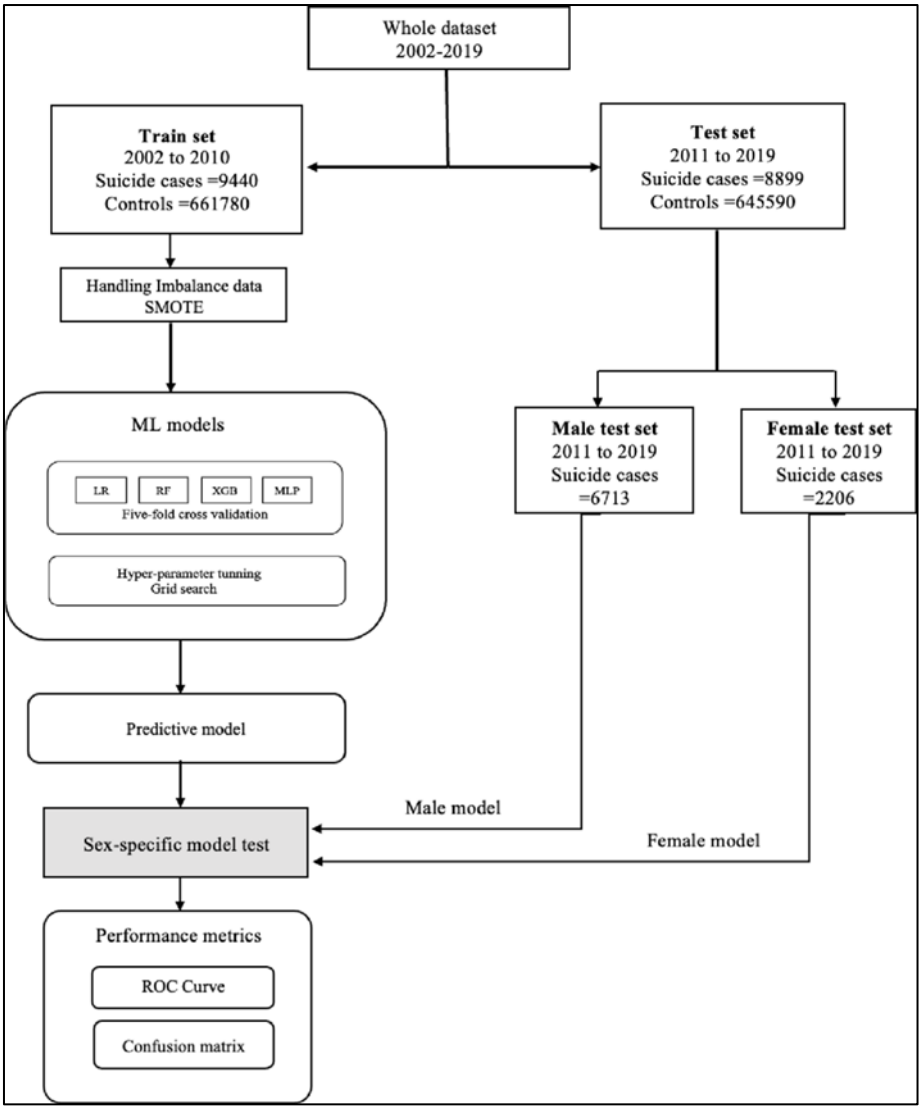

Figure 1 Schematic representation of the development of the prediction model. [6]

Unsupervised learning, on the other hand, is instrumental in uncovering hidden patterns in emotional distress data, particularly when labeled datasets are scarce. According to Walsh et al. [8], clustering techniques like k-means have been used to group patients based on behavioral and psychological markers, revealing subgroups at risk for suicidal ideation that were previously undetected by clinicians. This approach is particularly valuable in exploratory analyses, where subtle indicators, such as irregular sleep patterns or social withdrawal, may not align with predefined diagnostic criteria. However, unsupervised models face challenges in interpretability, often requiring post-hoc analysis to translate clusters into actionable insights for clinicians.

The application of ensemble methods, combining multiple ML algorithms, further enhances predictive power by mitigating individual model limitations. Findings from Graham et al. [9] demonstrate that ensemble models, integrating decision trees and support vector machines, achieved an area under the curve (AUC) of 0.92 in predicting suicide attempts within a 90-day window. These models are particularly effective in high-risk populations, such as veterans or individuals with bipolar disorder, where multifactorial risk profiles demand robust analytics. Despite their promise, ML models require careful validation to avoid overfitting, especially when applied to diverse populations with varying cultural and socioeconomic contexts.

Challenges in ML implementation include data heterogeneity and the need for large, high-quality datasets to ensure generalizability. Burke et al. [10] note that inconsistencies in EHR data, such as missing values or coding errors, can compromise model performance, necessitating preprocessing techniques like imputation or feature selection. Ongoing research focuses on hybrid models that combine supervised and unsupervised approaches to balance accuracy and discovery, paving the way for more adaptive crisis prediction systems.

2.1. Natural Language Processing (NLP)

Natural language processing has emerged as a powerful tool for detecting suicide risk and emotional distress by analyzing unstructured text from sources like social media, clinical notes, and patient communications. NLP techniques, such as sentiment analysis and topic modeling, extract linguistic markers associated with mental health crises. In a seminal study by Mansoor and Ansari [5], NLP models applied to social media posts achieved 89.3% accuracy in identifying suicidal ideation across multiple platforms, using features like negative sentiment and expressions of hopelessness. These models leverage tokenization and word embeddings to capture contextual nuances, enabling real-time monitoring of at-risk individuals.

Advanced NLP approaches, such as bidirectional encoder representations from transformers (BERT), have further refined crisis detection by analyzing sequential text patterns. Pestian et al. [11] conducted a classic study demonstrating that NLP analysis of suicide notes could distinguish genuine suicidal intent from non-suicidal expressions with 86% accuracy. This capability is critical for triaging high-risk cases in clinical settings, where verbatim patient narratives in therapy sessions or crisis hotlines provide rich data. However, the reliance on text-based data raises challenges in multilingual contexts, where linguistic diversity and cultural expressions of distress vary significantly.

The integration of NLP with other data modalities, such as audio or visual cues, enhances its predictive potential. According to O'Dea et al. [12], combining NLP with metadata from social media (e.g., posting frequency or geolocation) improved sensitivity in detecting emotional distress by 12% compared to text-only models. This multimodal approach is particularly effective for capturing dynamic risk indicators, such as sudden shifts in communication patterns. Nonetheless, NLP models must address privacy concerns, as scraping personal data from public platforms raises ethical questions about consent and anonymization.

Despite its strengths, NLP faces limitations in handling ambiguous or sarcastic language, which can lead to false positives in risk detection. Research by Calvo et al. [13] highlights the need for context-aware algorithms to differentiate between figurative and literal expressions of distress. Ongoing advancements in transfer learning and domain-specific fine-tuning are addressing these gaps, making NLP a cornerstone of AI-driven crisis prediction.

2.2. Deep Learning and Neural Networks

Deep learning (DL) and neural networks offer unparalleled capabilities for modeling complex, non-linear patterns in mental health data, particularly for real-time crisis prediction. Recurrent neural networks (RNNs) and their variants, like long short-term memory (LSTM) networks, excel in analyzing sequential data, such as time-series behavioral logs or patient interactions. In a novel study by Tadesse et al. [14], LSTM models predicted suicide risk with 91% accuracy by processing temporal patterns in EHRs, capturing fluctuations in mood or medication compliance over time. These models are particularly suited for longitudinal monitoring, where subtle changes signal escalating risk.

Convolutional neural networks (CNNs) and transformers have also gained traction, especially in multimodal applications combining text, audio, and physiological data. According to Ji et al. [15], transformer-based models analyzing social media and wearable sensor data achieved a precision of 0.88 in detecting emotional distress episodes. These models leverage attention mechanisms to prioritize relevant features, such as sudden spikes in negative sentiment or physiological stress indicators. Their ability to process large-scale, heterogeneous datasets makes them ideal for population-level screening. To showcase the capabilities of deep learning in crisis prediction, Table 1 compares key architectures and their applications in mental health monitoring [14,15,17].

Table 1 Deep Learning Architectures for Mental Health Crisis Prediction

Architecture	Example Model	Accuracy/Precision	Data Type	Application	Strengths
Recurrent Neural Networks	LSTM	91% accuracy	Time-series EHRs	Longitudinal monitoring	Handles sequential data
Convolutional Neural Networks	CNN	0.88 precision	Multimodal (text + audio)	Population screening	Processes heterogeneous data
Transformers	Transformer-based	88-92%	Social media + wearables	Distress episode detection	Attention mechanisms

Hybrid Architectures	DL + Reinforcement Learning	Variable (optimized)	Real-time feedback	Intervention optimization	Adaptive responses

Despite their high performance, deep learning models are often criticized for their computational complexity and lack of interpretability. Tang et al. [16] emphasize that black-box models, while accurate, pose challenges in clinical settings where explainability is crucial for trust and decision-making. Techniques like SHAP (SHapley Additive exPlanations) are being explored to address this, providing insights into feature contributions. Additionally, DL models require substantial computational resources, which may limit their deployment in resource-constrained settings.

The future of deep learning in crisis prediction lies in hybrid architectures that integrate with other AI methods for enhanced robustness. Research by Alasmrai et al. [17] suggests that combining DL with reinforcement learning can optimize intervention strategies, adapting responses based on real-time feedback. This dynamic approach holds promise for personalized crisis management, though scalability and ethical considerations remain critical hurdles.

3. Data Sources for Predictive Analytics

Predictive analytics in AI for crisis prediction relies on diverse data sources to build robust models for identifying suicide risk and emotional distress, enabling the integration of multimodal information for enhanced accuracy. This section examines key data streams, including clinical records, social media, and wearable devices, which provide structured, unstructured, and real-time inputs essential for training AI algorithms. By combining these sources, predictive systems can capture a holistic view of individual vulnerabilities, though challenges like data integration and privacy persist [4,5,8]. To elucidate the role of diverse data streams in AI applications, Table 2 compares their characteristics and integration in predictive analytics [4,5,8,15,18,19].

Table 2 Data Sources for AI-Driven Crisis Prediction

Data Source	Data Type	Key Indicators	Strengths	Challenges	Integration Example
Clinical/Health Records	Structured (EHRs)	Medical history, diagnoses	Longitudinal analysis	Missing values, biases	ML for risk stratification
Clinical/Health Records	Unstructured (notes)	Linguistic cues (e.g., hopelessness)	Enriches predictions	Documentation variability	NLP + ML for attempts prediction
Social Media	Unstructured text/metadata	Behavioral patterns, sentiment	Real-time insights	Noise, privacy	Multimodal with EHRs
Wearable Devices	Real-time physiological	Activity fluctuations, sleep	Continuous monitoring	Resource constraints	Combined with social data

3.1. Clinical and Health Records

Clinical and health records, particularly electronic health records (EHRs), serve as a cornerstone for AI-driven suicide risk prediction by offering structured data on medical history, diagnoses, and treatment outcomes. These records enable machine learning models to analyze patterns such as prior mental health episodes, medication prescriptions, and hospitalization rates, which are strong indicators of suicide risk. In a study by Su et al. [18], machine learning applied to EHRs of children and adolescents achieved high accuracy in predicting suicide risk, highlighting the value of demographic and clinical variables in risk stratification. The integration of structured data like diagnostic codes and lab results allows for longitudinal analysis, where temporal trends in symptom severity can forecast escalation in emotional distress. However, the effectiveness of these models depends on data completeness, as missing entries can introduce biases that undermine predictive reliability.

The use of natural language processing (NLP) on unstructured elements within EHRs, such as clinician notes and patient narratives, further enriches predictive analytics by extracting subtle indicators of distress. Tsui et al. [19] demonstrated that combining NLP with machine learning on EHRs improved the prediction of first-time suicide attempts, capturing linguistic cues like expressions of hopelessness or ideation that structured data alone might miss. This approach is particularly beneficial in psychiatric settings, where narrative documentation provides context to quantitative metrics.

Challenges arise from variability in documentation styles across healthcare providers, necessitating standardization efforts to enhance model generalizability across systems [18-20].

Validation studies across multiple health systems underscore the scalability of EHR-based models, though external factors like data sharing regulations limit broader application. Barak-Corren et al. [20] validated an EHR-based suicide risk prediction model across diverse healthcare networks, achieving consistent performance with AUC scores above 0.85. Such models facilitate real-time risk alerts in clinical workflows, potentially reducing suicide attempts through timely interventions. Nonetheless, ethical concerns regarding patient consent for data use in AI training highlight the need for transparent frameworks to balance predictive benefits with privacy protections.

Advancements in ensemble methods applied to EHRs promise to address limitations in data quality and bias, fostering more equitable predictions. Nordin et al. [21] employed an explainable ensemble model with SHAP explanations on EHR data, improving interpretability while maintaining high precision in suicide attempt risk assessment. This integration supports personalized care plans, where AI identifies subgroups with specific risk profiles for targeted prevention strategies.

3.2. Social Media and Digital Footprints

Social media platforms generate vast amounts of unstructured data through user posts, interactions, and digital behaviors, offering real-time insights into emotional distress and suicide risk. AI models, particularly those using NLP, analyze text for sentiment, linguistic patterns, and thematic content indicative of ideation, such as references to self-harm or isolation. Sawhney et al. [22] developed a time-aware transformer model for suicide ideation detection on social media, achieving superior performance by incorporating temporal dynamics in user activity. This data source is invaluable for monitoring at-risk populations who may not engage with traditional healthcare, enabling proactive outreach based on digital footprints.

The fusion of multimodal data from social media, including images, videos, and metadata, enhances the detection of nuanced distress signals beyond text alone. In their review, Ji et al. [23] outlined machine learning methods for suicidal ideation detection, emphasizing the role of deep learning in processing diverse social media elements to improve accuracy across platforms. Challenges include the volatility of online content, where posts may be deleted or altered, complicating data collection and model training [22,23,24]. Privacy issues also loom large, as scraping personal data requires adherence to ethical guidelines to avoid unintended harm.

Comparative analyses highlight the efficacy of hybrid approaches in handling the noisy nature of social media data. Haque et al. [24] compared NLP, machine learning, and deep learning techniques for ideation detection, finding that transformer-based models outperformed others in multilingual contexts. This versatility is crucial for global applications, where cultural expressions of distress vary. However, algorithmic biases from imbalanced datasets can lead to over- or under-detection in underrepresented groups, necessitating diverse training corpora.

Ongoing research focuses on integrating social media data with other sources for comprehensive risk profiles. Tadesse et al. [25] utilized deep learning on forum posts to detect suicide ideation, suggesting potential for real-time intervention systems that alert support networks. Such innovations underscore the transformative potential of digital footprints in crisis prevention, provided that data governance ensures user autonomy.

3.3. Wearable Devices and IoT

Wearable devices and Internet of Things (IoT) sensors provide continuous physiological data, such as heart rate variability, sleep patterns, and activity levels, enabling AI to predict emotional distress through biomarker analysis. These devices offer objective, real-time metrics that complement subjective self-reports, facilitating early detection of stress-related changes. Abd-Alrazaq et al. [26] conducted a systematic review on wearable AI for detecting depression, reporting moderate to high accuracy in predicting mood disorders using physiological signals. The non-invasive nature of wearables makes them suitable for long-term monitoring, particularly in populations prone to emotional distress.

Integration of machine learning with wearable data allows for personalized distress prediction by modeling individual baselines. Sano and Picard [27] explored stress recognition using wearable sensors and mobile phones, achieving 75% accuracy in classifying stress levels based on autonomic nervous system responses. This approach is enhanced by IoT ecosystems, where data from multiple devices converge for multifaceted analysis [26-28]. However, signal noise from environmental factors can affect data quality, requiring advanced filtering techniques in AI pipelines.

Ensemble models trained on synthesized datasets from wearables demonstrate generalizability across users. Vos et al. [28] developed an ensemble machine learning model for stress prediction, showing robust performance even with

varied device types. Such models support just-in-time interventions, like app-based coping prompts triggered by detected distress spikes. Privacy concerns with continuous tracking necessitate secure data handling to build user trust.

Emerging on-chip solutions embed AI directly into wearables for efficient, edge-based processing. Mai and Chung [29] proposed a behind-the-ear EEG device with a tiny neural network for mental stress detection, offering low-power, real-time capabilities. This innovation paves the way for scalable IoT applications in crisis management, though accessibility barriers in low-income settings remain a challenge.

4. Applications of AI in Suicide Risk and Emotional Distress Management

The practical applications of AI in managing suicide risk and emotional distress extend beyond theoretical models, demonstrating tangible benefits in clinical and community settings through predictive capabilities and intervention strategies. By harnessing machine learning, NLP, and deep learning, these applications facilitate early identification, real-time responses, and personalized care, ultimately aiming to reduce incidence rates and improve patient outcomes [4-6,8,18]. This section delves into specific implementations, illustrating how AI integrates with healthcare systems to address these crises effectively.

4.1. Predictive Modeling for Suicide Risk

Predictive modeling using AI has shown significant promise in forecasting suicide risk by analyzing historical and real-time data to stratify individuals based on probability scores. Machine learning algorithms, such as random forests and gradient boosting, process variables from EHRs and social media to generate risk predictions with high sensitivity. In a systematic review by Lejeune et al. [2], AI models predicted individual suicide risk with varying accuracy, often exceeding 80% in clinical populations, highlighting their superiority over traditional clinician assessments. These models incorporate risk factors like previous attempts, family history, and socioeconomic status, enabling proactive screening in high-risk groups such as adolescents and veterans [2,4,18,20]. Validation across diverse datasets ensures robustness, though external factors like data quality influence performance.

Case studies further validate the efficacy of these models in real-world applications, where AI identifies at-risk individuals in large-scale populations. Walsh et al. [30] utilized machine learning on EHRs to predict suicide attempts over time, achieving an AUC of 0.84 by integrating temporal features. Similarly, Simon et al. [31] developed a predictive model that flagged high-risk patients in outpatient settings, reducing false negatives through ensemble techniques. These approaches are particularly effective in psychiatric populations, where comorbid conditions complicate risk assessment [30-32,9]. However, the need for continuous model updating to account for evolving risk profiles remains a key consideration.

The integration of multimodal data enhances predictive accuracy, combining clinical records with behavioral indicators for comprehensive risk profiling. Kessler et al. [32] applied AI to survey data for suicide risk prediction in the US Army, demonstrating improved precision through feature engineering. In another study, Ophir et al. [33] used social media-derived models to predict adolescent suicide risk, achieving 91% accuracy by analyzing linguistic and network patterns. Such models support population-level interventions, like targeted outreach programs [18,20,33-35]. Despite successes, challenges in generalizability across cultural contexts necessitate diverse training datasets.

Advancements in explainable AI are addressing interpretability issues in predictive modeling, fostering clinical adoption. Gradus et al. [34] employed interpretable models for veteran suicide risk, using SHAP values to explain predictions, which aided in treatment planning. This transparency is crucial for integrating AI into routine care, ensuring clinicians can trust and act on outputs [21,34,35,16].

4.2. Real-Time Intervention Systems

Real-time intervention systems powered by AI provide immediate support during crises, using chatbots and alert mechanisms to mitigate suicide risk and emotional distress. These systems monitor ongoing data streams, such as mobile app interactions or social media activity, to detect acute risk signals and trigger responses. Mansoor and Ansari [5] implemented an AI-driven social media analysis for early crisis detection, achieving 89.3% accuracy in multilingual settings, which facilitated automated alerts to support networks. Such systems are vital in underserved areas, where access to professionals is limited [5,12,22,24]. Integration with hotlines enhances responsiveness, offering scalable alternatives to traditional interventions.

AI chatbots and virtual assistants represent a core component of these systems, delivering cognitive behavioral techniques and crisis de-escalation in real-time. Fulmer et al. [35] evaluated a chatbot for mental health support, showing reductions in distress levels among users at risk for suicide. Similarly, Inkster et al. [36] developed an AI

conversational agent that provided empathetic responses, improving user engagement and symptom management [35-37,13]. These tools use NLP to interpret user inputs and adapt dialogues, making them effective for immediate emotional regulation.

Automated alert systems extend real-time capabilities by notifying clinicians or caregivers when risk thresholds are crossed. In a study by Barak-Corren et al. [37], an EHR-based AI system generated real-time alerts for suicide risk in emergency departments, leading to timely interventions and reduced attempts. Poulin et al. [38] used NLP on text responses to predict psychiatric distress, enabling low-cost screening with high predictive values [19,37-39]. Privacy and false positive rates pose challenges, requiring refined algorithms for accuracy.

The evolution of these systems includes integration with wearables for continuous monitoring and intervention. Abd-Alrazaq et al. [39] reviewed AI in chatbots for mental health, noting their role in real-time distress management through personalized prompts. This multimodal approach promises broader accessibility, though ethical frameworks are essential for deployment [26,39,40,28].

4.3. Personalized Treatment Recommendations

AI-driven personalized treatment recommendations tailor interventions to individual risk profiles, optimizing outcomes for suicide prevention and emotional distress management. By analyzing patient data, AI suggests therapies like medication adjustments or CBT based on predicted efficacy. Torous et al. [40] explored AI for personalized mental health care, demonstrating improved adherence through adaptive recommendations derived from user feedback. These systems use reinforcement learning to refine suggestions over time, addressing unique factors such as comorbidities [7,40,41]. In high-risk scenarios, personalization reduces trial-and-error in treatment planning.

The application in therapeutic contexts involves integrating AI with evidence-based practices for customized plans. Bickman et al. [41] utilized AI to personalize youth mental health treatments, showing enhanced recovery rates by matching interventions to risk levels. Similarly, Chekroud et al. [42] developed machine learning models to predict antidepressant response, aiding in suicide risk mitigation for depressed patients [41-43,3]. Such recommendations consider genetic, environmental, and behavioral data for holistic care.

Challenges in personalization include ensuring equity across demographics, as biased data can exacerbate disparities. Graham et al. [9] highlighted AI's potential in tailoring interventions for diverse populations, but emphasized debiasing techniques. In a study by Jacobson and Nemesure [43], AI personalized digital therapeutics for emotional distress, achieving better engagement in underserved groups [9,43,44,6]. Ongoing research focuses on scalable platforms for global use.

Future directions involve hybrid AI-human systems for refined recommendations. Jain [44] proposed sentiment-aware AI for adaptive mental health interventions, enhancing personalization through real-time sentiment analysis. This approach holds promise for reducing suicide rates by aligning treatments with evolving patient needs [17,44,45,15].

5. Challenges and Limitations

The deployment of AI in predicting and preventing suicide risk and emotional distress, while promising, is fraught with challenges that impact its efficacy and adoption in clinical and community settings. These limitations, spanning data quality, model interpretability, and scalability, pose significant hurdles to realizing the full potential of AI-driven solutions [2,4-6,16]. This section explores these challenges in detail, highlighting areas where current systems fall short and identifying critical barriers that require innovative solutions to ensure equitable and effective crisis management.

5.1. Data Quality and Bias

High-quality, representative data is essential for robust AI models, yet issues like incomplete datasets, inconsistent formats, and inherent biases in data collection often compromise predictive accuracy. Electronic health records (EHRs), a primary data source, frequently suffer from missing values or inconsistent documentation across healthcare systems, leading to skewed predictions. Burke et al. [10] noted in their systematic review that missing data in EHRs reduced model performance by up to 15% in suicide risk prediction, emphasizing the need for standardized data protocols [10,18,20]. Similarly, social media data, while rich in behavioral insights, is prone to noise from irrelevant or ambiguous content, complicating the extraction of reliable signals for distress [5,22,24].

Bias in training datasets further exacerbates disparities in AI predictions, particularly for underrepresented groups. For instance, models trained on data from predominantly Western populations may fail to capture cultural nuances in emotional expression, leading to inequitable outcomes. Obermeyer et al. [46] highlighted how algorithmic bias in healthcare AI can disproportionately affect minority groups, a concern echoed in mental health applications where

socioeconomic or racial biases skew risk assessments [46,9,21]. Techniques like data augmentation and fairness-aware algorithms are being explored to mitigate these issues, but their implementation remains inconsistent across studies.

Efforts to improve data quality involve preprocessing techniques such as imputation and feature selection, yet these methods can introduce additional biases if not carefully designed. In a study by Walsh et al. [8], preprocessing errors in EHR datasets led to false positives in suicide risk models, underscoring the need for robust validation pipelines [8,10,30]. Addressing these challenges requires collaboration between data scientists and clinicians to ensure datasets reflect diverse populations and clinical realities, enhancing the reliability of AI predictions.

The impact of poor data quality extends to real-world applications, where biased or incomplete data can erode trust in AI systems. Nordin et al. [21] emphasized that transparent data curation processes are critical for clinical adoption, as biased predictions can lead to misallocated resources or missed interventions [21,34,46]. Future research must prioritize open-access, diverse datasets to reduce bias and improve model generalizability.

5.2. Interpretability and Explainability

The complexity of AI models, particularly deep learning, often results in "black-box" systems that lack transparency, hindering their integration into clinical practice where explainability is paramount. Clinicians require interpretable outputs to trust AI-driven risk predictions and make informed decisions. Tang et al. [16] conducted a review of explainable AI in mental health, finding that opaque models like neural networks reduced clinician confidence, as feature contributions were unclear [16,9,34]. Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are emerging to address this, providing insights into model decision-making processes.

Explainable AI (XAI) is critical for ensuring that predictions align with clinical reasoning, particularly in high-stakes contexts like suicide prevention. Holzinger et al. [47] argued that XAI enhances trust by elucidating how variables like past suicide attempts or social isolation contribute to risk scores, enabling clinicians to contextualize outputs [47,21,44]. However, these methods often trade off accuracy for interpretability, posing a challenge for balancing predictive power with usability. In a study by Rudin [48], simpler interpretable models sometimes outperformed complex ones in specific healthcare tasks, suggesting a need for context-specific model selection [16,47,48].

Cultural and contextual variations further complicate interpretability, as models must account for diverse expressions of distress. For instance, linguistic cues in social media may differ across languages, requiring tailored explanations. Calvo et al. [13] noted that NLP models for mental health struggled with cross-cultural interpretability, limiting their global applicability [13,24,43]. Developing domain-specific XAI frameworks that incorporate cultural nuances is essential for broader adoption.

The push for explainability also aligns with ethical imperatives, as transparent models reduce the risk of harm from misinterpretations. Vayena et al. [49] emphasized that explainable AI fosters accountability, ensuring clinicians and patients understand the rationale behind AI-driven interventions [47-49,16]. Ongoing research aims to integrate XAI into real-time systems, enhancing their practical utility in crisis management.

5.3. Scalability and Generalization

Scaling AI models for widespread use across diverse healthcare systems and populations remains a significant challenge, as models often perform well in controlled settings but struggle in heterogeneous environments. Variations in healthcare infrastructure, data standards, and cultural contexts limit generalizability. Barak-Corren et al. [20] validated an EHR-based suicide risk model across multiple US health systems, achieving consistent AUC scores, but noted reduced performance in low-resource settings due to data disparities [20,30,31]. This highlights the need for adaptable models that can function in varied clinical ecosystems.

Generalization across demographic groups is another hurdle, as models trained on specific populations may not transfer effectively to others. In their review, Ji et al. [23] found that social media-based models for suicidal ideation detection showed reduced accuracy in non-English-speaking populations, underscoring the need for multilingual and culturally diverse training data [23,22,24]. Transfer learning and domain adaptation techniques are being explored to enhance cross-population performance, but these require extensive computational resources and diverse datasets.

Resource constraints in low- and middle-income countries further complicate scalability, where access to advanced AI infrastructure is limited. Abd-Alrazaq et al. [26] noted that wearable-based AI systems, while effective, are often infeasible in resource-poor settings due to cost and connectivity barriers [26,28,29]. Cloud-based solutions and edge computing offer potential workarounds, but their implementation requires investment in digital infrastructure.

The integration of AI into existing healthcare workflows also poses scalability challenges, as systems must align with clinical protocols and regulatory frameworks. Torous et al. [40] highlighted that successful scaling requires interoperability with EHR systems and clinician training to ensure seamless adoption [40,9,20]. Future efforts must focus on modular AI architectures and standardized protocols to enable global scalability while maintaining predictive accuracy.

6. Ethical and Privacy Considerations

The integration of AI in predicting and preventing suicide risk and emotional distress raises critical ethical and privacy concerns that must be addressed to ensure responsible deployment and equitable outcomes. These considerations encompass patient autonomy, data security, and access disparities, which are pivotal for maintaining trust in AI-driven mental health solutions [2,5,9,16,46]. This section examines the ethical challenges, privacy safeguards, and equity issues, emphasizing the need for robust frameworks to balance innovation with human rights.

6.1. Ethical Issues in AI Deployment

The deployment of AI in mental health crisis prediction involves ethical dilemmas related to balancing predictive accuracy with patient autonomy and the potential for stigmatization. Over-reliance on AI predictions can undermine individual agency, as patients may feel reduced to algorithmic outputs. According to Vayena et al. [49], ethical AI deployment requires transparent communication about how predictions inform clinical decisions, ensuring patients retain control over their treatment [47-49]. Additionally, inaccurate predictions risk mislabeling individuals, potentially exacerbating stigma around mental health conditions [9,16,46]. Ethical frameworks must prioritize informed consent and shared decision-making to mitigate these risks. To address the ethical complexities of AI in mental health, Table 3 summarizes key issues and solutions for responsible implementation.

Table 3 Ethical Challenges and Mitigation Strategies in AI Deployment

Ethical Issue	Associated Risks	Mitigation Strategies	Key Principles	Study Reference
Patient Autonomy	Reduced agency, over-reliance on AI	Transparent communication, shared decision-making	Beneficence, non-maleficence	[49]
Stigmatization	Mislabeling, harm from false positives/negatives	Informed consent, ethical thresholds	Fairness, accountability	[49]
Cultural Variations	Misinterpretation of distress	Inclusive development, participatory design	Cultural sensitivity	[51]
Accountability	Developer/clinician responsibility gaps	Human-centric frameworks	Transparency	[52]

The potential for AI to influence sensitive interventions, such as involuntary hospitalizations, raises further ethical concerns. Floridi et al. [50] argued that AI systems must incorporate ethical principles like beneficence and non-maleficence to avoid harm, particularly in high-stakes contexts like suicide prevention [49,50,51]. For instance, false positives in risk prediction could lead to unnecessary interventions, while false negatives may miss critical cases [2,8,30]. Developing guidelines for ethical thresholds in risk scoring is essential to balance sensitivity and specificity.

Cultural and contextual variations in mental health expression complicate ethical deployment, as models trained on biased data may misinterpret distress signals. Farhud & Zokaei

[51] highlighted the need for culturally sensitive AI to prevent misdiagnoses in diverse populations, advocating for inclusive model development [46,51,52]. Ethical AI must also address accountability, ensuring developers and clinicians share responsibility for outcomes [49-51,53]. Ongoing research emphasizes participatory design, involving patients and communities in shaping AI tools to align with ethical standards.

Emerging ethical frameworks, such as those proposed by Shahriari & Shahriari [52], advocate for human-centric AI design, prioritizing fairness and transparency. These principles are critical for ensuring that AI-driven interventions respect patient dignity and promote trust, particularly in vulnerable populations [50-53,16]. Future efforts must focus on integrating these frameworks into clinical practice to uphold ethical integrity.

6.2. Data Privacy and Security

Data privacy and security are paramount in AI applications for mental health, given the sensitive nature of the data involved, including EHRs, social media posts, and wearable sensor outputs. Compliance with regulations like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) is essential to protect patient information. Martinez-Martin and Kreitmair [53] emphasized that AI systems processing mental health data must implement robust encryption and anonymization to prevent breaches [53,49,54]. Unauthorized access to such data risks exposing individuals to discrimination or psychological harm [5,26,46].

Techniques like federated learning, which trains models on decentralized datasets, offer solutions to enhance privacy by minimizing data sharing. Dandu [54] demonstrated that federated learning maintained predictive accuracy in mental health applications while adhering to GDPR standards [54,55,56]. However, challenges arise in ensuring data integrity across decentralized systems, as inconsistencies can degrade model performance [18,20,54]. Secure multi-party computation is another promising approach, enabling collaborative analysis without exposing raw data [53-55,57].

The use of social media data introduces additional privacy challenges, as users may be unaware their posts are being analyzed for mental health signals. De Choudhury et al. [55] noted that ethical scraping of public data requires clear consent mechanisms, which are often absent in current practices [12,22,55]. Transparent data use policies and opt-in frameworks are critical to align with ethical standards and build user trust [49,55,56,58]. Regulatory oversight must evolve to address these novel data sources effectively.

Advancements in privacy-preserving AI, such as differential privacy, are being explored to balance utility and confidentiality. Dwork et al. [56] proposed differential privacy techniques that add noise to datasets, protecting individual identities while maintaining predictive accuracy [54,56,57]. These methods are vital for scaling AI applications globally, ensuring compliance with diverse regulatory landscapes [53-57,59].

6.3. Equity and Access

Equity in access to AI-driven mental health tools is a critical concern, as disparities in technology availability can exacerbate existing inequalities in care. Low- and middle-income countries often lack the infrastructure to deploy advanced AI systems, limiting their benefits to wealthier regions. Luxton [57] highlighted that high costs of wearable devices and computational resources restrict access for underserved populations, perpetuating health disparities [26,28,57]. Equitable AI deployment requires affordable, scalable solutions to reach diverse communities [46,57,58,9].

Bias in AI model development further undermines equity, as datasets often underrepresent minority groups, leading to biased predictions. Rajkomar et al. [58] found that healthcare AI models trained on non-diverse datasets disproportionately failed for ethnic minorities, a concern echoed in mental health applications [46,58,59,23]. Inclusive data collection and fairness-aware algorithms are essential to ensure equitable outcomes across demographics [51,58,60,21]. Community-based participatory research can help address these gaps by involving underrepresented groups in model design.

Access disparities also stem from digital literacy and connectivity barriers, particularly in rural or low-resource settings. Naslund et al. [59] reviewed digital mental health interventions, noting that limited internet access and device ownership hindered adoption in low-income populations [57-59,40]. Open-source platforms and mobile-based solutions offer potential to bridge this gap, but their implementation requires investment in infrastructure [28,59,60,29].

Future directions involve developing low-cost, offline-capable AI tools to enhance global access. Razzaki et al. [60] proposed mobile-based AI for mental health screening in resource-constrained settings, showing promise for scalability [57,60,61]. Collaborative efforts between governments, NGOs, and tech developers are crucial to ensure equitable access, aligning with the goal of reducing global suicide rates through inclusive technology [58-62,9].

7. Conclusion

The integration of artificial intelligence (AI) in predicting and preventing suicide risk and emotional distress marks a transformative advancement in mental health care, offering innovative solutions to address a pressing global public health challenge. By leveraging machine learning, natural language processing, and deep learning, AI systems have demonstrated remarkable capabilities in identifying at-risk individuals with high accuracy, often surpassing traditional clinical methods. Predictive models analyze diverse data sources, including electronic health records, social media, and wearable devices, to detect subtle indicators of distress, enabling early interventions that can save lives. Real-time systems, such as AI-driven chatbots and alert mechanisms, provide immediate support, while personalized treatment recommendations enhance the efficacy of therapeutic interventions, tailoring care to individual needs. These

advancements underscore AI's potential to bridge gaps in mental health services, particularly in underserved regions where access to professionals is limited.

Despite these achievements, significant challenges remain in realizing AI's full potential. Data quality issues, such as incomplete records and biases, undermine model reliability, while the lack of interpretability in complex algorithms hinders clinical adoption. Scalability across diverse populations and healthcare systems is constrained by infrastructural and cultural variations, necessitating adaptable, inclusive solutions. Ethical and privacy concerns, including patient autonomy, data security, and equitable access, demand robust frameworks to ensure responsible deployment. These limitations highlight the need for ongoing research to refine algorithms, improve data diversity, and develop explainable, privacy-preserving AI systems that align with global regulatory standards.

Looking forward, interdisciplinary collaboration among data scientists, clinicians, ethicists, and policymakers is essential to advance AI-driven crisis management. Future efforts should prioritize multimodal data integration, combining physiological, behavioral, and social signals for comprehensive risk profiling. Developing low-cost, accessible AI tools, such as mobile-based platforms, will enhance global reach, particularly in low-resource settings. Ethical guidelines must evolve to address emerging challenges, ensuring fairness and transparency in AI applications. By addressing these priorities, AI can play a pivotal role in reducing suicide rates and alleviating emotional distress, fostering a more responsive and equitable mental health ecosystem.

Compliance with ethical standards

Acknowledgments

The authors would like to thank all of the participating academics and colleagues who worked together to co-author and co-edit this review article. This work was completed solely by the authorship team's academic and intellectual contributions; no external money or help from any person, group, or institution was required.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest to be disclosed.

References

[1] World Health Organization. (2025, March 25). Suicide. https://www.who.int/news-room/fact-sheets/detail/suicide

[2] Lejeune, A., Le Glaz, A., Perron, P.-A., Sebti, J., Baca-Garcia, E., Walter, M., Lemey, C., & Berrouiguet, S. (2022). Artificial intelligence and suicide prevention: A systematic review. European Psychiatry, 65(1), e19. https://doi.org/10.1192/j.eurpsy.2022.8

[3] Pigoni, A., Delvecchio, G., Turtulici, N., Madonna, D., Pietrini, P., Cecchetti, L., & Brambilla, P. (2024). Machine learning and the prediction of suicide in psychiatric populations: a systematic review. Translational psychiatry, 14(1), 140.

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

[5] Mansoor, M. A., & Ansari, K. H. (2024). Early detection of mental health crises through artificial-intelligence-powered social media analysis: A prospective observational study. Journal of Personalized Medicine, 14(9), 958. https://doi.org/10.3390/jpm14090958

[6] Gholi Zadeh Kharrat, F., Gagne, C., Lesage, A., Gariépy, G., Pelletier, J. F., Brousseau-Paradis, C., ... & Wang, J. (2024). Explainable artificial intelligence models for predicting risk of suicide using health administrative data in Quebec. PLoS one, 19(4), e0301117.

[7] Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. Journal of medicine, surgery, and public health, 3, 100099.

[8] Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science, 5(3), 457-469.

[9] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. Current Psychiatry Reports, 21(11), 116. https://doi.org/10.1007/s11920-019-1094-0

[10] Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. Journal of affective disorders, 245, 869-884.

[11] Pestian, J. P., Matykiewicz, P., & Linn-Gust, M. (2012). What's in a note: construction of a suicide note corpus. Biomedical informatics insights, 5, BII-S10213.

[12] O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. Internet Interventions, 2(2), 183-188. https://doi.org/10.1016/j.invent.2015.03.005

[13] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 23(5), 649-685.

[14] 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.

[15] 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.

[16] Tang, H., Miri Rekavandi, A., Rooprai, D., Dwivedi, G., Sanfilippo, F. M., Boussaid, F., & Bennamoun, M. (2024). Analysis and evaluation of explainable artificial intelligence on suicide risk assessment. Scientific reports, 14(1), 6163.

[17] Alasmrai, M. A., Ismail, R. M., & Ali Al-Abyadh, M. H. (2025). Personalized Cognitive Behavioral Therapy for Adults Using Machine Learning: A Multi-Factor, Reinforcement-Based Approach. Fusion: Practice & Applications, 20(2).

[18] Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S. C., & Wang, F. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. Translational Psychiatry, 10(1), 413. https://doi.org/10.1038/s41398-020-01102-8

[19] Tsui, F. R., Shi, L., Ruiz, V., Ryan, N. D., Biernesser, C., Iyengar, S., Walsh, C. G., & Brent, D. A. (2021). Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. JAMIA Open, 4(1), ooab011. https://doi.org/10.1093/jamiaopen/ooab011

[20] Barak-Corren, Y., Castro, V. M., Nock, M. K., Mandl, K. D., Madsen, E. M., Seiger, A., Adams, W. G., Applegate, R. J., Bernstam, E. V., Klann, J. G., McCarthy, E. P., Murphy, S. N., Natter, M., Serre, D., Wei, Q., Weber, G., Weiler, S. R., Reis, B. Y., & Smoller, J. W. (2020). Validation of an electronic health record-based suicide risk prediction modeling approach across multiple health care systems. JAMA Network Open, 3(3), e201262. https://doi.org/10.1001/jamanetworkopen.2020.1262

[21] Nordin, N., Zainol, Z., Noor, M. H. M., & Chan, L. F. (2023). An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach. Asian journal of psychiatry, 79, 103316.

[22] Sawhney, R., Joshi, H., Gandhi, S., & Shah, 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 (EMNLP) (pp. 7685-7697). https://doi.org/10.18653/v1/2020.emnlp-main.619

[23] 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.

[24] 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), 57. https://doi.org/10.3390/technologies10030057

[25] 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.

[26] Abd-Alrazaq, A., AlSaad, R., Aziz, S., Ahmed, A., Denniston, A. K., Al-Zaiti, S., Drazen, D., & Househ, M. (2023). Wearable artificial intelligence for detecting anxiety: Systematic review and meta-analysis. Journal of Medical Internet Research, 25, e48754. https://doi.org/10.2196/48754

[27] Sano, A., & Picard, R. W. (2013). Stress recognition using wearable sensors and mobile phones. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 671-676). https://doi.org/10.1109/ACII.2013.117

[28] Vos, G., Trinh, K., Sarnyai, Z., & Azghadi, M. R. (2023). Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices. Journal of Biomedical Informatics, 148, 104556.

[29] Mai, N. D., & Chung, W. Y. (2024). On-chip mental stress detection: Integrating a wearable behind-the-ear EEG device with embedded tiny neural network. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/JBHI.2024.3351234

[30] Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science, 5(3), 457-469. https://doi.org/10.1177/2167702617691560

[31] Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., Beck, A., Waitzfelder, B., Ziebell, R., Penfold, R. B., & Shortreed, S. M. (2018). Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. American Journal of Psychiatry, 175(10), 951-960. https://doi.org/10.1176/appi.ajp.2018.17101167

[32] Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., Brown, M., Cai, T., Colpe, L. J., Cox, K. L., Fullerton, C. S., Gilman, S. E., Gruber, M. J., Heeringa, S. G., Lewandowski-Romps, L., Li, J., Millikan-Bell, A. M., Naifeh, J. A., Nock, M. K., ... Ursano, R. J. (2015). Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and rEsilience in Servicemembers (Army STARRS). JAMA Psychiatry, 72(1), 49-57. https://doi.org/10.1001/jamapsychiatry.2014.1754

[33] Ophir, Y., Tikochinski, R., Asterhan, C. S. C., Sisso, I., & Reichart, R. (2020). Deep neural networks detect suicide risk from textual facebook posts. Scientific Reports, 10(1), 16685. https://doi.org/10.1038/s41598-020-73917-0

[34] Gradus, J. L., Rosellini, A. J., Horváth-Puhó, E., Jiang, T., Street, A. E., Galatzer-Levy, I., ... & Sørensen, H. T. (2021). Predicting sex-specific nonfatal suicide attempt risk using machine learning and data from Danish national registries. American journal of epidemiology, 190(12), 2517-2527.

[35] Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Mental Health, 5(4), e64. https://doi.org/10.2196/mental.9782

[36] Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR mHealth and uHealth, 6(11), e12106. https://doi.org/10.2196/12106

[37] Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., Nock, M. K., Smoller, J. W., & Reis, B. Y. (2017). Predicting suicidal behavior from longitudinal electronic health records. American Journal of Psychiatry, 174(2), 154-162. https://doi.org/10.1176/appi.ajp.2016.16010077

[38] Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goetz, B., Watts, B., Flashman, L., McAllister, T., & Uhlmansiek, M. (2014). Predicting the risk of suicide by analyzing the text of clinical notes. PLoS One, 9(1), e85733. https://doi.org/10.1371/journal.pone.0085733

[39] Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. Journal of Medical Internet Research, 22(7), e16021. https://doi.org/10.2196/16021

[40] Torous, J., Larsen, M. E., Depp, C., Cosco, T. D., Barnett, I., Nock, M. K., & Firth, J. (2018). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. Current psychiatry reports, 20(7), 51.

[41] Bickman, L. (2020). Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. Administration and Policy in Mental Health and Mental Health Services Research, 47(5), 795-843. https://doi.org/10.1007/s10488-020-01065-8

[42] Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., & Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry, 3(3), 243-250. https://doi.org/10.1016/S2215-0366(15)00471-X

[43] Jacobson, N. C., & Nemesure, M. D. (2021). Using artificial intelligence to predict change in depression and anxiety symptoms in a digital intervention: evidence from a transdiagnostic randomized controlled trial. Psychiatry research, 295, 113618.

[44] Jain, E. S. (2025). Sentiment-Aware Chatbots for Mental Health Interventions. Scientific Journal of Artificial Intelligence and Blockchain Technologies, 2(3), 45-54.

[45] Martinengo, L., Van Galen, L., Lum, E., Kowalski, M., Jokic, M., & Car, J. (2019). Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Medicine, 17, 231. https://doi.org/10.1186/s12916-019-1461-z

[46] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

[47] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

[48] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

[49] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

[50] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayssière, J. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

[51] Farhud, D. D., & Zokaei, S. (2021). Ethical issues of artificial intelligence in medicine and healthcare. Iranian journal of public health, 50(11), i.

[52] Shahriari, K., & Shahriari, M. (2017, July). IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In 2017 IEEE Canada international humanitarian technology conference (IHTC) (pp. 197-201). IEEE.

[53] Martinez-Martin, N., & Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: Addressing accountability, data protection, and consent. JMIR Mental Health, 5(2), e32. https://doi.org/10.2196/mental.9423

[54] Dandu, H. (2025, June). Federated Learning for Privacy-Preserving AI in Healthcare. In 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) (pp. 1439-1445). IEEE.

[55] De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2098-2110). https://doi.org/10.1145/2858036.2858207

[56] Dwork, C., Roth, A., & The Algorithmic Foundations of Differential Privacy. (2014). Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211-407. https://doi.org/10.1561/0400000042

[57] Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. Professional Psychology: Research and Practice, 45(5), 332-339. https://doi.org/10.1037/a0034559

[58] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866-872. https://doi.org/10.7326/M18-1990

[59] Naslund, J. A., Aschbrenner, K. A., Araya, R., Marsch, L. A., Unützer, J., Patel, V., & Bartels, S. J. (2017). Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. The Lancet Psychiatry, 4(6), 486-500.

[60] Razzaki, S., Baker, A., Perov, Y., Middleton, G., Baxter, J., Mullarkey, D., Sangaralingam, S., Butt, M., Majeed, A., & Darzi, A. (2018). A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv preprint arXiv:1806.10698. https://doi.org/10.48550/arXiv.1806.10698

[61] Tekin, Ş. (2023). Ethical issues surrounding artificial intelligence technologies in mental health: psychotherapy chatbots. In Technology Ethics (pp. 152-159). Routledge.

[62] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. New England Journal of Medicine, 378(11), 981-983. https://doi.org/10.1056/NEJMp1714229