SAHAY: Multimodal, Privacy-Preserving AI for Suicide Risk Detection and Intervention in India

Salam Michael Singh[1] , **Manik Inder Singh Sethi**[3] ,
Suresh Bada Math[3] and **Tanmoy Chakraborty**[1,2]
[1]Department of Electrical Engineering, Indian Institute of Technology Delhi, India
[2]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, India
[3] Department of Psychiatry, National Institute of Mental Health and Neuro Sciences
{mike95iitd, tanchak}@iitd.ac.in, manikindersingh@gmail.com, sbm@nimhans.ac.in

Abstract

Suicide accounts for one of the leading causes of death in India, with over 164,033 deaths reported in 2021. Despite increased awareness, the gap between the need for consistent treatment and actual accessibility remains a challenge due to limited mental health infrastructure, the stigma surrounding mental illness in society, and the lack of real-time detection mechanisms. Traditional suicide risk assessments often miss early signs of distress, which rely heavily on clinical evaluations and self-reporting. Although AI-based monitoring seems promising, currently available models focus only on risk prediction without intervention and treatment, leaving a critical gap in tackling crisis management. In this proposal, we strive to design SAHAY, the first-of-its-kind AI-based, suicide prevention framework that seamlessly couples prediction with prevention and treatment access. Leveraging multimodal data, including the social media text and Electronic Health Records (EHR) and Ecological Momentary Assessments (EMA) such as wearable physiological data, SAHAY aims to assess suicide risk dynamically. Unlike existing models, SAHAY is culturally adaptive, multilingual and seamlessly integrates with India's TeleMANAS mental health support system to provide structured AI-human collaboration for long-term care and crisis interventions. It will be an adaptable, scalable, modular, and plug-and-play solution based on the Digital Public Infrastructure principle. Additionally, we intend to incorporate AI-driven geo-spatial crisis mapping to identify suicide hotspots in underserved regions. By combining real-time multimodal risk detection, professional mental health intervention, and geo-spatial outreach, SAHAY represents a scalable, adaptable, and end-to-end suicide prevention system. The design of SAHAY aligns with UN Sustainable Development Goals (SDGs) 3, 4, 5, 10, and 17, promoting inclusive, accessible, and data-driven mental healthcare.

1 Introduction

Suicide remains a significant global public health challenge, accounting for nearly 800,000 deaths annually, with India contributing over 164,033 deaths in 2021 alone.[1] Despite national efforts in suicide prevention, including the launch of TeleMANAS[2] for accessible mental health services, India continues to face critical barriers such as stigma, lack of trained professionals, and insufficient early detection systems [Garg *et al.*, 2019; Gururaj *et al.*, 2016]. Suicide disproportionately affects young adults, with over 60% of victims having no prior contact with mental health services [Chaudhary *et al.*, 2024]. The National Mental Health Survey (NMHS) reports a 75% treatment gap, highlighting the urgent need for early identification and intervention [Patel *et al.*, 2018]. With the rapid adoption of digital technologies, AI-driven suicide prevention systems, specifically generative AI, are promising in suicide prevention [Chakraborty *et al.*, 2025], and they have emerged as strong alternatives to traditional clinical assessments, enabling real-time risk detection and scalable intervention strategies. However, existing solutions suffer from limited multimodal data integration, a lack of regional adaptation, and weak AI-human collaboration frameworks, making them unsuitable for the diverse socio-cultural landscape of India.

Suicide: Indian Perspective and the Need for Early Detection India has one of the highest suicide rates globally, with suicide the leading cause of death among individuals aged 15–29. The suicide mortality rate stood at 16.5 per 100,000 in 2016, well above the global average of 10.5 per 100,000 [Dandona *et al.*, 2018]. Mental health stigma, inadequate psychiatric infrastructure, and economic distress further exacerbate the crisis [Gaiha *et al.*, 2020]. Despite the WHO's recommendation of 3 psychiatrists per 100,000 people, India has only 0.75 per 100,000, with rural areas facing severe shortages [Garg *et al.*, 2019]. Traditional suicide detection relies on clinical assessments, self-reporting, and periodic screenings, which are reactive rather than proactive

[1]https://ncrb.gov.in

[2]TeleMANAS (Tele Mental Health Assessment and Networking Across States) is India's national tele-mental health service, launched to provide free, 24/7 mental health support through toll-free helplines.

[Patel *et al.*, 2018]. Over 80% individuals with mental health conditions in India do not seek professional help within the first year of onset [Gururaj *et al.*, 2016]. These methods also fail to capture real-time distress signals, which are critical for prevention. AI-driven approaches using social media analysis, physiological markers (heart rate variability, sleep patterns), and real-time self-reports (Ecological Momentary Assessment (EMA) surveys) offer a scalable early intervention strategy. However, for effective deployment in India, such systems must be multilingual, culturally sensitive, and seamlessly integrated with human-led interventions.

Existing Studies and Research Gaps AI-driven suicide detection has leveraged Natural Language Processing (NLP), physiological monitoring, and EHR analytics for risk assessment [Pigoni *et al.*, 2024]. Social media-based NLP models detect ideation through text patterns and sentiment shifts [Chancellor and De Choudhury, 2020; Chatterjee *et al.*, 2022], while wearable data on heart rate and sleep patterns correlate with mental health decline [Torous *et al.*, 2018]. EHR-based models predict suicide attempts using historical markers [Zang *et al.*, 2024]. However, existing approaches are fragmented, focusing on either prediction or intervention without a comprehensive, closed-loop system. Most of them lack multilingual support, seamless AI-human collaboration, and privacy safeguards, making them unsuitable for India's diverse population [Murthy, 2017]. Additionally, a few models are not plug-and-play, limiting integration with NGOs and private mental health platforms. The absence of geo-spatial crisis mapping further restricts targeted interventions in suicide hotspots. To address these gaps, we propose SAHAY [3], a multimodal, privacy-preserving AI framework that fully integrates prediction, prevention, and treatment access. Unlike existing models, SAHAY closes the loop by combining social media analysis, wearable data, and EMA surveys while incorporating LLMs fine-tuned for Indian languages. Designed for scalability, it can be seamlessly integrated with TeleMANAS and other public, NGO, and private mental health platforms. In addition, AI-driven crisis mapping enables targeted outreach in high-risk areas. With end-to-end encryption and anonymised data handling, SAHAY ensures secure, adaptive, and ethical suicide prevention at scale.

Our Major Contributions

1. **A Multi-faceted Data Integration and Risk Detection System**: We propose a comprehensive suicide risk assessment system that integrates social media activity, wearable physiological data, EMA surveys, and EHR records, processed through LLMs for dynamic risk classification and geo-spatial hotspot identification.
2. **End-to-End Suicide Prevention Pipeline**: Unlike existing models that focus on either risk prediction or intervention, SAHAY closes the loop by ensuring seamless AI-human collaboration, transitioning high-risk individuals from AI-driven support to professional mental health treatment, while leveraging geo-spatial crisis mapping for targeted deployment in suicide hotspots via TeleMANAS.
3. **Privacy-First and Ethically Aligned AI**: SAHAY incorporates end-to-end encryption, anonymized data handling, and user consent mechanisms, ensuring compliance with GDPR, DPDP Act, and AI safety standards, while mitigating algorithmic bias through cross-cultural dataset training.
4. **Scalable and Adaptive Framework**: We intend SAHAY to be a modular, privacy-preserving (as per Digital Public Infrastructure (DPI)) that can be seamlessly integrated into India's mental health ecosystem, ensuring nationwide accessibility and interoperability and supports regional language adaptation, and geo-spatial crisis mapping.

[3]SAHAY is a Hindi word meaning *help* or *assistance*.

2 Goals

Through SAHAY, we aim to address the critical gaps in early suicide detection while ensuring sensitive care delivery. This section elaborates on the crisis and the key goals.

Limitations of Traditional Suicide Detection Methods Mental health assessment and suicide prevention in India face significant challenges due to their reliance on traditional clinical methods. Existing approaches primarily depend on in-person assessments, self-reporting, and periodic clinical evaluations, which often fail to capture the dynamic nature of mental health deterioration. The Lancet Commission on Global Mental Health highlights that these approaches capture only a fraction of at-risk individuals, with the treatment gap exceeding 75% in low and middle-income countries such as India [Patel *et al.*, 2018] where over 150,000 individuals die by suicide annually, with significant underreporting likely making this figure an underestimate [Armstrong and Vijayakumar, 2018].

The scarcity of mental health professionals in India, particularly in rural areas with only 0.75 psychiatrists per 100,000 population in contrast to the recommended 3 per 100,000 population by WHO [Garg *et al.*, 2019], makes traditional assessment methods inaccessible [Choudhary *et al.*, 2024]. This creates significant gaps in monitoring and intervention, especially in regions where mental health stigma is prevalent. Studies show that over 60% of individuals who died by suicide had no recorded prior contact with mental health services [Chaudhary *et al.*, 2024]. Clinical assessments, while valuable, provide only periodic snapshots of an individual's mental state, missing crucial behavioural changes between visits. Studies from the National Mental Health Survey of India reveal that over 80% of individuals with mental health conditions do not receive any form of professional help within the first year of onset [Gururaj *et al.*, 2016]. The traditional system's reliance on active help-seeking behaviour particularly fails to identify individuals from marginalised communities who face additional barriers to accessing mental health services [Singh *et al.*, 2016; Dandona *et al.*, 2018].

Towards Multimodal Data Integration for Enhanced Suicide Detection Recent advances in digital mental health technologies have demonstrated the potential of integrating multiple data sources for more effective suicide risk assessment [Torous *et al.*, 2018]. **(i) Social Media:** Digital footprints/phenotyping through social media activity patterns and online behaviour changes have emerged as valuable indicators of mental health status. NLP techniques applied to social media content have shown promising results in identify-

ing at-risk individuals [Chancellor and De Choudhury, 2020]. Studies indicate that changes in social media usage patterns can precede suicide attempts by several weeks [De Choudhury *et al.*, 2013]. **(ii) Wearable Devices:** This technology provides continuous physiological data modalities that objectively indicate mental state changes. Studies using platforms like MindLAMP demonstrate strong correlations between biomarkers such as sleep patterns, activity levels, heart rate variability, and mental health status [Torous *et al.*, 2018]. Integration with EHR provides crucial historical context, with machine learning models showing efficacy in predicting suicide attempts through EHR analysis [Zang *et al.*, 2024].

AI-Expert Collaboration Framework In SAHAY, we aim to propose a novel framework for effective collaboration between automated systems and mental health professionals, focusing on real-time risk assessment and expert validation systems. Khosravi *et al.* [Khosravi *et al.*, 2024] showed that AI systems can effectively augment clinical decision-making by continuously monitoring and analysing multimodal data sources. However, integrating expert knowledge remains crucial for system accuracy and cultural competency.

Longitudinal Care Management Arya [Arya, 2024] demonstrated that the risk of suicide attempts remains elevated for extended periods after initial crisis points, with up to 60% of attempts occurring within the first year of suicidal ideation. SAHAY's longitudinal care framework addresses this challenge through continuous monitoring and adaptive support strategies. Studies of long-term digital interventions have demonstrated significant reductions in suicide risk when combined with professional support [Yosep *et al.*, 2024]. Integrating passive monitoring with active intervention strategies has proven effective in maintaining engagement over extended periods [Mikus *et al.*, 2018]. By incorporating automated monitoring with regular professional assessment, SAHAY addresses a key limitation of current prevention strategies in India, where follow-up care is often inconsistent or unavailable [Omboni *et al.*, 2022].

3 Alignment with SDG

Our proposal is rightly aligned with the following SDGs:

1. **Good Health and Well-being (SDG 3)**: SAHAY tackles India's mental health crisis with a comprehensive suicide prevention system, integrating wearable data, social indicators, and professional care.
2. **Quality Education (SDG 4)**: SAHAY promotes mental health awareness and suicide prevention through culturally sensitive education and early warning recognition training.
3. **Gender Equality (SDG 5)**: The rate of suicide among Indian men is 2.5 times that among women, and there is hesitancy in seeking help among men [Yadav *et al.*, 2023]. SAHAY adopts gender-specific mental health approaches, addressing distinct suicide risk patterns with culturally appropriate support.
4. **Decent Work and Economic Growth (SDG 8)**: SAHAY fosters economic growth by supporting adolescents, and youth, who are pillars of a nation's growth and development.
5. **Reduced Inequalities (SDG 10)**: SAHAY bridges urban-rural mental healthcare disparities with AI-driven detection, multilingual support, and integration with TeleMANAS.
6. **Partnership for the Goals (SDG 17)**: Collaboration with NIMHANS (National Institute of Mental Health & Neuro Sciences, an Institute of National Importance under the Government of India), SAHAY combines AI-driven detection with expert-guided intervention and long-term care via TeleMANAS.

Apart from these SDG goals, SAHAY also strongly aligns with the Leave No One Behind (LNOB) principle by addressing critical mental health disparities in India's vulnerable populations. Through AI-driven geo-spatial analytics, SAHAY identifies high-risk areas and deploys community-based crisis teams alongside TeleMANAS support, bridging India's severe psychiatric care shortage (0.75 psychiatrists per 100,000 [Garg *et al.*, 2019]). The system intends to support various Indian languages, ensuring linguistic minorities aren't excluded from mental healthcare. SAHAY's privacy-preserving design aims to tackle the treatment gap driven by stigma while prioritising early intervention for adolescents, India's largest and most vulnerable demographic group.

4 SAHAY: System Description

In this section, we present the technical architecture and methodological framework of SAHAY, detailing our approach to multimodal data integration, LLM-based risk detection, human-expert intervention through TeleMANAS, geographical crisis hotspot mapping and long-term care management. Each component is designed for privacy preservation and cultural sensitivity, working together to create a comprehensive suicide prevention system. Figure 1 shows a schematic diagram of SAHAY.

4.1 Data Sources and Integration

We plan to aggregate and analyse diverse data modalities for early detection and intervention, ensuring privacy and cultural sensitivity.

Data Sources, Collection and Multi-Modal Integration We aim to integrate multiple validated data sources to comprehensively assess mental health risk factors. We plan to leverage existing datasets while also curating and annotating data in-house. Key sources include real-time EMA data from smartphone sensing and wearable devices such as MINDLAMP [Torous *et al.*, 2018], which continuously monitor activity patterns, sleep quality, and social interactions. Additionally, we will utilize the SHARP (Suicide and Health Assessment for Prevention) dataset, which provides clinically validated suicide risk progression patterns through longitudinal patient data. For social media analysis, ethical concerns necessitate a cautious approach. Therefore, we will begin with established datasets of mental health discourse [De Choudhury *et al.*, 2013]. The clinical assessment component incorporates validated psychometric instruments, including the Patient Health Questionnaire (PHQ-9) for depression screening, the Generalized Anxiety Disorder Assessment (GAD-7) for anxiety evaluation, and the Columbia-

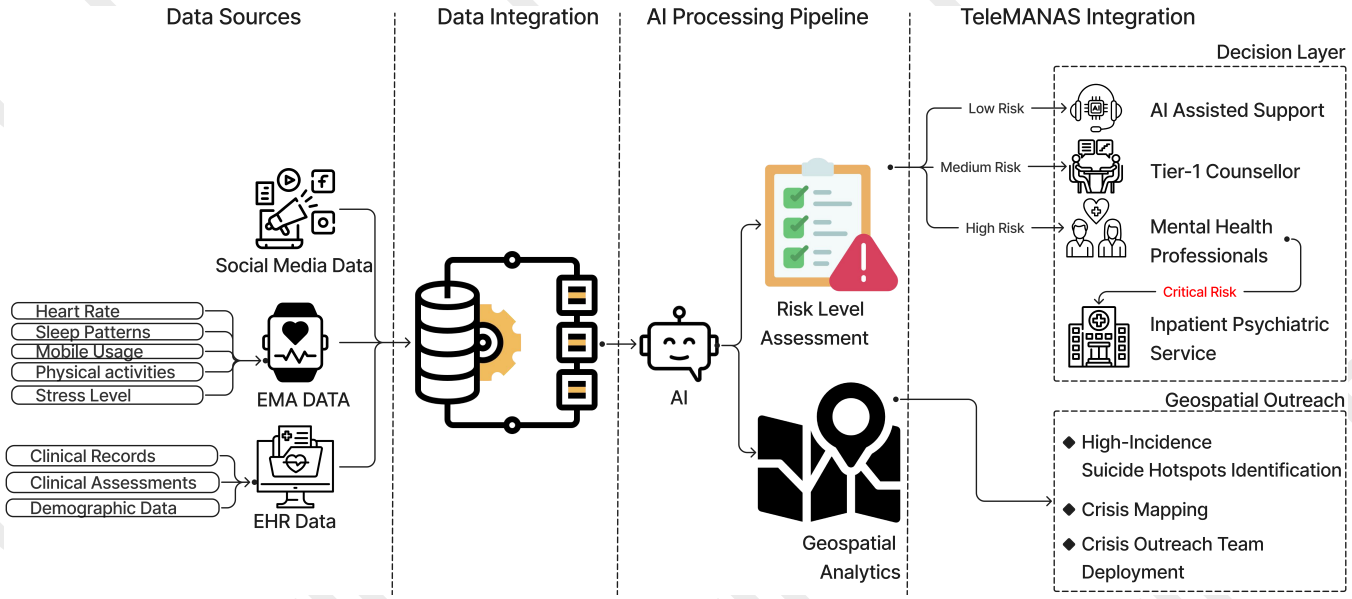

Figure 1: Our proposed schematic diagram of SAHAY outlines an AI-driven framework for suicide detection, prevention, and treatment access in India. SAHAY consists of four key components: data sources, multimodal data integration, an AI processing pipeline, and TeleMANAS integration. We will collect data from various sources, including social media, EHR, and EMA data, followed by a secure pipeline for multimodal aggregation. An LLM-based AI engine will then assess risk levels (low, medium, high) and conduct geo-spatial analytics to map and identify suicide hotspots. In the TeleMANAS integration phase, interventions will be tiered based on risk levels: AI-assisted support for low-risk cases, Tier-1 counsellors for medium-risk cases, and mental health professionals for high-risk cases, with inpatient psychiatric care for critical cases. Additionally, geo-spatial insights from the AI pipeline will guide the deployment of crisis outreach teams, particularly in high-incidence areas such as rural regions with limited psychiatric access. These teams, in collaboration with community-based organisations, will provide real-world support beyond digital interventions.

Suicide Severity Rating Scale (C-SSRS) for suicide risk assessment, all of which have been extensively validated in Indian populations [Gururaj *et al.*, 2016]. Additionally, we will leverage EHR data to enhance contextual risk assessment [Barak-Corren *et al.*, 2020].

We propose to employ a federated architecture to ensure data consistency, privacy, and secure aggregation. Social media text will undergo language-specific normalisation, wearable data will be standardised across devices, and clinical assessments will be normalised to consistent scales. A temporal alignment system synchronises diverse data streams, handling varying sampling rates from continuous wearable monitoring to clinical periodic evaluations. Privacy is enforced through differential privacy mechanisms, enabling population-level analysis while preserving individual anonymity [Smith *et al.*, 2023].

LLM-Based Risk Detection System SAHAY leverages LLM advancements for comprehensive suicide risk detection, contextual risk assessment, and culturally adaptive response generation for AI-assisted support. It analyses social media posts and EMA responses to identify linguistic markers of distress (e.g., hopelessness, social withdrawal, self-harm references), enhanced through clinically informed prompt engineering [Chancellor and De Choudhury, 2020; Smith *et al.*, 2023]. The multi-level classification system evaluates both text-based risk indicators and contextual signals from multimodal data, escalating cases to human experts when uncertainty thresholds are exceeded [Khosravi *et al.*, 2024]. A key feature is regional language support, enabling LLMs to process distress signals across multiple Indian languages and addressing a critical accessibility gap in AI-driven mental health systems.

The system incorporates robust safety measures through carefully designed AI guardrails that prevent harmful or inappropriate responses while maintaining the empathetic nature required for mental health support [Dong *et al.*, 2024; Menz *et al.*, 2024]. Rather than developing new language models, SAHAY focuses on creating a framework for deploying existing LLM capabilities in a mental healthcare context, with particular attention to cultural sensitivity and regional language support. SAHAY's LLM framework incorporates specialized components for handling India's linguistic and cultural diversity in mental health contexts. For multilingual processing, we build upon established models like BLOOM [Scao *et al.*, 2022], mT0 [Muennighoff *et al.*, 2023] and Aya [Üstün *et al.*, 2024], and recent India-specific models like OpenHathi, and Airavata [Gala *et al.*, 2024] that demonstrate strong capabilities in processing Hindi contexts. The system implements a domain adaptation pipeline that fine-tunes these base models on culturally relevant mental health conversations, with particular attention to region-specific mental health terminology and expressions. This adaptation is crucial as mental health expressions often vary significantly across cultural contexts – for instance, somatic symptoms are more commonly reported than emotional distress in many Indian communities [Murthy, 2017].

SAHAY's cultural sensitivity framework addresses multiple dimensions: recognising culture-specific stigma markers, understanding family-centric expressions of distress common in Indian contexts, and appropriate interpretation of religious and spiritual references in mental health discussions. The system is designed to understand subtle linguistic markers that might indicate distress, such as indirect references to family conflicts or academic pressure, which are significant stress factors in the Indian context [Gaiha *et al.*, 2020; Singh *et al.*, 2016]. This cultural adaptation extends beyond mere language translation, incorporating an understanding of collectivist social structures, religious beliefs, and traditional support systems that play crucial roles in mental health in India. The framework maintains this cultural sensitivity while adhering to clinical standards, creating a bridge between traditional mental health assessment approaches and culturally relevant support mechanisms.

This approach aligns with our broader goal of creating a scalable, culturally competent mental health support system that effectively serves diverse Indian populations while maintaining the high safety and ethics required in mental healthcare applications [Choudhary *et al.*, 2024].

4.2 Ecological Momentary Assessment (EMA) Integration and Real-Time Risk Monitoring

Building upon SAHAY's LLM-based risk detection system, we plan to implement an EMA framework [Shiffman *et al.*, 2008] that enables continuous psychological state monitoring through interactive micro-assessments. The system delivers clinically validated instruments (PHQ-9, GAD-7, C-SSRS) through an engaging interface with gamification elements [Mikus *et al.*, 2018], maintaining therapeutic validity while encouraging sustained participation. When the LLM identifies potential risk patterns, it triggers targeted EMA interactions to gather specific information about the user's current state [Torous *et al.*, 2018]. The system integrates EMA responses with passive monitoring data from smartphone sensors and wearables, tracking behavioural indicators like sleep patterns and activity levels. This multi-modal integration enables the LLM to perform more comprehensive risk assessments by comparing self-reported states against objective measures [De Choudhury *et al.*, 2013]. The continuous feedback loop between EMA responses and LLM analysis creates a personalised monitoring system that adapts to individual risk patterns while maintaining engagement [Khosravi *et al.*, 2024].

4.3 AI-Expert Collaboration Framework

Following LLM-based risk detection and real-time risk monitoring, SAHAY aims to implement a structured handoff protocol transitioning identified cases to TeleMANAS' human-led intervention network. Upon risk detection, the system generates a comprehensive assessment package combining the LLM's findings with supporting evidence from multimodal data sources, which is securely transmitted to TeleMANAS Tier-1 counsellors with the AI-generated risk assessment, behavioural patterns, and the temporal sequence of concerning signals that triggered the alert. The system implements priority routing for high-risk cases, ensuring immediate counsellor connection. Tier-1 counsellors provide initial psychological first aid and determine escalation needs to Tier-2 specialists. The system prepares automated referral documentation for specialised intervention by clinical psychologists and psychiatrists for escalated cases. The framework maintains structured clinical documentation of all interactions and interventions, serving as comprehensive care records and feedback data for refining AI detection parameters. The system manages follow-up scheduling based on risk levels and clinical recommendations while monitoring engagement patterns through automated reminders. Throughout this process, AI components function solely as support tools, with all therapeutic decisions residing with mental health professionals. This structured approach ensures a seamless transition from AI-based detection to human-led intervention while maintaining clinical autonomy and accountability through detailed audit logs.

4.4 Longitudinal Care Management System

Following the initial TeleMANAS intervention, SAHAY's longitudinal care system will implement structured protocols for sustained mental health monitoring and support. The system coordinates long-term care through an adaptive scheduling algorithm that manages follow-up appointments based on risk assessments, intervention outcomes, and clinical recommendations. The system integrates with TeleMANAS' callback infrastructure, preparing contextual summaries of interventions and behavioural patterns for each follow-up session. For cases requiring specialised care, it facilitates seamless transitions to District Mental Health Programs (DMHPs) and psychiatric facilities through automated referral documentation and care coordination. The system continuously monitors recovery trajectories throughout these transitions through integrated data modalities, including daily activities and physiological markers. The framework maintains comprehensive longitudinal records with strict access controls, enabling authorised providers to access relevant care histories while ensuring privacy. Progress reports are automatically generated for clinical review, allowing care teams to assess intervention effectiveness and adjust treatment plans. The system continues active monitoring until clinical teams determine intensive follow-up is no longer required, transitioning individuals to maintenance protocols while maintaining the capability for reactivation if needed.

5 Deployment Plans with TeleMANAS

Existing Capabilities of TeleMANAS We plan to deploy SAHAY using TeleMANAS, for which NIMHANS is the apex centre. NIMHANS is the premier mental health institution in India and they have demonstrated expertise in scaling digital mental health interventions through technology-driven training and outreach [Malathesh *et al.*, 2021b]. This includes initiatives that successfully trained accredited social health activists (ASHAs) and primary care doctors, reducing the treatment gap for severe mental disorders (SMDs) and substance use disorders (SUDs) [Gakkhar *et al.*, 2023; Malathesh *et al.*, 2021a]. Large-scale programs like TORENT have equipped thousands of rural health workers, leveraging technology to

identify, refer to, and manage psychiatric cases in underserved areas [Thakur *et al.*, 2025]. This experience establishes a strong foundation for integrating SAHAY with TeleMANAS, ensuring seamless AI-human collaboration for nationwide mental health support.

TeleMANAS is a centralised mental health support system operating in a two-tier system: Tier-1 consists of trained counsellors providing psychological first aid, psychoeducation, and crisis de-escalation, while Tier-2 comprises mental health professionals handling complex cases. It currently supports mental health care in 20+ regional languages, TeleMANAS processes 7,000+ follow-up calls monthly nationwide.

SAHAY's Deployment Strategy We propose to deploy SAHAY using TeleMANAS's infrastructure, focusing on integrating multi-faceted data sources and AI-driven intervention. Our strategy encompasses three integrated phases as follows:

1. **Phase 1: Risk Detection and Immediate Support.** The initial phase establishes multiple data collection and analysis pathways. We will develop partnerships with social media platforms to monitor opt-in public discourse for distress indicators through sentiment analysis. Simultaneously, we will collaborate with health technology firms to integrate wearable biosensors that track physiological correlates of mental health deterioration, including sleep disturbances, heart rate variability (HRV) patterns, and physical activity deviations. The system incorporates clinically validated psychometric assessment tools: PHQ-9, GAD-7 and C-SSRS, all embedded within a mobile health (mHealth) application. LLMs perform continuous NLP to identify high-risk expressions and generate real-time alerts for human review. The system deploys personalised interventions through EMA – a real-time data collection methodology that provides contextual behavioural insights and enables dynamic adaptation of coping strategies.
2. **Phase 2: Human-led Intervention and Crisis Response.** The second phase activates SAHAY's comprehensive crisis response mechanisms. The system employs AI-driven geospatial analytics to map high-risk suicide hotspots, particularly identifying underserved rural areas with limited access to psychiatric facilities. In response to identified needs, crisis outreach teams are mobilised through partnerships with community-based organisations, extending support beyond digital interventions to provide in-person assistance. Through the TeleMANAS platform, Tier-1 counsellors provide immediate psychological first aid using evidence-based crisis intervention protocols. Tier-2 mental health professionals develop specialised interventions for sustained therapeutic engagement in complex cases. This creates a multi-layered support system where AI-assisted risk assessment guides human-led interventions while ensuring appropriate resource allocation based on severity levels.
3. **Phase 3: Long-Term Care Transition and Monitoring.** The final phase establishes comprehensive pathways for sustained mental healthcare. The system implements automated referral protocols that seamlessly transition high-risk users from digital interventions to human-led tele-consultations through TeleMANAS' tiered network. Based on standardised clinical assessment protocols, critical cases are directed to DMHPs or inpatient psychiatric facilities. Post-transition care includes AI-augmented follow-up systems that leverage TeleMANAS' extensive callback infrastructure, monitoring treatment adherence and early warning signs of relapse. The system employs longitudinal predictive analytics frameworks to track mental health trajectories over time, enabling continuous optimisation of prevention strategies and intervention protocols.

TeleMANAS Integration and Future Scalability Building upon the established deployment phases, SAHAY's future expansion focuses on deepening integration with TeleMANAS while maintaining system independence. The enhanced integration framework will be implemented through an advanced API-driven workflow system automatically connecting AI-detected high-risk cases with appropriate TeleMANAS counsellors. This integration leverages TeleMANAS' extensive linguistic capabilities, supporting mental health interventions across 22 Indian languages to increase the accessibility of suicide prevention services for diverse linguistic demographics. This is scrutinised by an anonymised data-sharing protocol that conducts continual training to refine the AI models further while maintaining strict adherence to privacy regulations and ethical guidelines.

While SAHAY will be deeply integrated with TeleMANAS, it is designed with an independent operational architecture that enables it to function autonomously across various healthcare ecosystems. This flexibility ensures that the AI-driven risk detection and initial intervention components can operate effectively even in settings without immediate TeleMANAS access. Within this framework, TeleMANAS serves as an optimal closure partner, facilitating the transition from AI-based detection and intervention to sustained human-led care. SAHAY's modular and adaptable design enables it to function as a plug-and-play with DPI, which is a scalable, technology-driven backbone designed to address large-scale societal challenges through interoperable and inclusive systems. DPI enables seamless integration across digital and non-digital environments, ensuring accessibility for diverse user groups. Common examples include India's Aadhaar for identity verification and UPI for digital payments, allowing multiple access modes, including online platforms and offline alternatives. By adopting this approach, SAHAY aims to seamlessly integrate with both public and private mental healthcare ecosystems, including platforms like Wysa, Amaha, and Sneha, ensuring broad accessibility and interoperability across the mental health landscape.

6 Project Evaluation

We plan to evaluate the performance of SAHAY based on multiple factors as mentioned below:

- **Risk Detection Accuracy**: The model should effectively identify at-risk individuals with high precision, recall, and F1-score while minimizing false positives. Evaluation metrics such as AUC-ROC and PR-AUC will be utilized to measure the trade-off between sensitivity and specificity.

- **Cross-Lingual Generalization**: The model should perform consistently across multiple Indian languages. We will assess its effectiveness across various cross-lingual benchmarking datasets.
- **Intervention Effectiveness**: The system should provide timely and effective AI-driven interventions, measured through user engagement metrics (AI interaction rates, EMA survey adherence), crisis resolution rates (successful transitions to human-led interventions), and reduction in distress levels (PHQ-9, GAD-7 pre/post intervention scores).
- **Ethical Compliance & Bias**: The model should remain fair and impartial, ensuring that suicide risk predictions are unbiased across different demographics. We will conduct fairness audits, analyze false alarm rates, and ensure compliance with privacy regulations (DPDP Act, GDPR, HIPAA-equivalent standards) [Singh Sethi *et al.*, 2025].
- **Privacy and Safety**: Given the sensitive nature of suicide prevention, SAHAY should maintain strict data security protocols through end-to-end encryption, anonymized data pipelines, and AI guardrails to prevent misinformation or harmful responses.

7 Challenges

There are several challenges involved in deploying an AI-based suicide prevention system. Moreover, integrating multimodal data poses a complex challenge: social media text, wearable sensor data, and EMA responses are not only coming from different sources but often show inconsistencies that require robust aggregation and validation (measured by data completeness and inter-source agreement scores). Linguistic and cultural adaptability is another barrier, with most existing AI models being English-centric and failing to capture regional dialects and culturally specific distress signals (evaluated by risk classification accuracy across languages). Risk detection, reporting false positives and false negatives, can either overburden crisis response teams or miss those in crisis, requiring high-precision AI models (measured by F1-score and recall). It is also crucial to ensure seamless AI-human collaboration, where delays in handoffs to TeleMANAS could impact intervention effectiveness (tracked by response time from AI flagging to counsellor engagement). Additionally, privacy and ethical concerns demand strong encryption, anonymisation, and consent frameworks (assessed through compliance with GDPR and the DPDP Act). Finally, long-term care is still challenging as it is important to continue with the engagement post-crisis to prevent relapse, measured by follow-up retention rates and long-term intervention success. Overcoming these challenges will enable SAHAY to scale as a privacy-preserving, clinically effective solution for suicide prevention in India.

8 Risks, Limitations, Ethical Considerations

SAHAY's primary challenges include false positives and false negatives, which require a meticulously designed high-precision AI model with structured AI-human handoff to mitigate such scenarios. Furthermore, as suicide assessments are highly sensitive, data privacy and security remain major concerns. While SAHAY has implemented end-to-end encryption, anonymisation, and strict compliance with GDPR, DPDP Act, and HIPAA-equivalent standards, data misuse and regulatory adherence concerns remain [Singh Sethi *et al.*, 2025]. Although SAHAY integrates multimodal data and TeleMANAS interventions, its effectiveness depends on real-world resource availability, the geolocation in which crisis events unfold, and the generalizability of AI across different population demographics. Ethical frameworks focus on user autonomy, informed consent, and bias mitigation. Additionally, there are challenges of AI decision-making being non-transparent, users being unable to override an AI decision, and third-party auditors being unable to observe AI's inner workings to validate fairness. Addressing these concerns is essential for SAHAY/'s scalable, ethical, and effective deployment.

9 Expected Results and Long-Term Plans

SAHAY aims to improve early identification of suicide risk, AI-assisted crisis response, and privacy-preserving mental health support. The expected outcomes include precise patient risk classification, seamless AI-human intervention, and better accessibility through multilingual adaptation. Structured human-led support with TeleMANAS integration and longitudinal follow-up to prevent relapse. Eventually, SAHAY will develop into a country-wide system of AI-powered suicide intervention that folds in regional language support, personalised interventions, and geo-spatial risk mapping. Future directions of research include global adaptation targeting low-resource settings, near-zero treatment gap [Kumar *et al.*, 2022], cost-effective intervention with high return on investment [Maheshwari *et al.*, 2024] and policy advocacy for AI-assisted mental health care. SAHAY envisions redefining mental health intervention in India and the global context by bridging gaps in detection, scalability, and ethical AI deployment.

10 Conclusion

In this proposal, we proposed a comprehensive plan to deploy SAHAY, the first-of-its-kind scalable AI-predictive suicide prevention model that fully integrates prediction, prevention and access to treatment, thus bridging the gap between risk identification and real-world intervention. Unlike existing AI models focusing on either prediction or intervention, SAHAY operates as a closed-loop system by integrating multimodal risk detection and assessment, AI-driven early intervention and structured human-led support. Additionally, the seamless integration with TeleMANAS will help create structured referrals to ensure timely and accessible mental health care. In the long term, we intend SAHAY to be accessible to the broad diversity of populations and scalable through a modular DPI as a deployment platform that can integrate into existing public health initiatives, influencing AI-driven mental health policy. Finally, it will ensure early detection, real-time intervention, and continuous support and potentially redefine AI-assisted suicide prevention in India and beyond.

Acknowledgments

We acknowledge the generous support from Tower Research Capital Markets for funding our work on applying machine learning for social good.

References

[Armstrong and Vijayakumar, 2018] Gregory Armstrong and Lakshmi Vijayakumar. Suicide in india: a complex public health tragedy in need of a plan. *The Lancet Public Health*, 3(10):e459–e460, Oct 2018.

[Arya, 2024] Vikas Arya. Suicide prevention in india. *Mental Health and Prevention*, 33:200316, 2024.

[Barak-Corren *et al.*, 2020] Yuval Barak-Corren, Victor M. Castro, Matthew K. Nock, Kenneth D. Mandl, Emily M. Madsen, Ashley Seiger, William G. Adams, R. Joseph Applegate, Elmer V. Bernstam, Jeffrey G. Klann, Ellen P. McCarthy, Shawn N. Murphy, Marc Natter, Brian Ostasiewski, Nandan Patibandla, Gary E. Rosenthal, George S. Silva, Kun Wei, Griffin M. Weber, Sarah R. Weiler, Ben Y. Reis, and Jordan W. Smoller. Validation of an electronic health record–based suicide risk prediction modeling approach across multiple health care systems. *JAMA Network Open*, 3(3):e201262–e201262, 03 2020.

[Chakraborty *et al.*, 2025] Tanmoy Chakraborty, Koushik Sinha Deb, Himanshu Kulkarni, Sarah Masud, Suresh Bada Math, Gayatri Oke, Rajesh Sagar, and Mona Sharma. The promise of generative ai for suicide prevention in india. *Nature Machine Intelligence*, Feb 2025.

[Chancellor and De Choudhury, 2020] Stevie Chancellor and Munmun De Choudhury. Methods in predictive techniques for mental health status on social media: a critical review. *npj Digital Medicine*, 3(1):43, Mar 2020.

[Chatterjee *et al.*, 2022] Moumita Chatterjee, Piyush Kumar, Poulomi Samanta, and Dhrubasish Sarkar. Suicide ideation detection from online social media: A multi-modal feature based technique. *International Journal of Information Management Data Insights*, 2(2):100103, 2022.

[Chaudhary *et al.*, 2024] Sofia Chaudhary, Jennifer A. Hoffmann, Christian D. Pulcini, Mark Zamani, Matt Hall, Kristyn N. Jeffries, Rachel Myers, Joel Fein, Bonnie T. Zima, Peter F. Ehrlich, Elizabeth R. Alpern, Stephen Hargarten, Karen M. Sheehan, Eric W. Fleegler, Monika K. Goyal, and Children's Hospital Association Research in Gun Related Events (CHARGE) Collaborative. Youth suicide and preceding mental health diagnosis. *JAMA Network Open*, 7(7):e2423996–e2423996, 07 2024.

[Choudhary *et al.*, 2024] Soumya Choudhary, Siddharth Dutt, John Torous, and Urvakhsh Meherwan Mehta. Bridging the mental health gap: The scope, challenges, and future of digital psychiatry in india. *Asian Journal of Psychiatry*, 101:104215, 2024.

[Dandona *et al.*, 2018] Rakhi Dandona, G. Anil Kumar, R. S. Dhaliwal, Mohsen Naghavi, Theo Vos, D. K. Shukla, Lakshmi Vijayakumar, G. Gururaj, J. S. Thakur, Atul Ambekar, Rajesh Sagar, Megha Arora, Deeksha Bhardwaj, Joy K. Chakma, Eliza Dutta, Melissa Furtado, Scott Glenn, Caitlin Hawley, Sarah C. Johnson, Tripti Khanna, Michael Kutz, W. Cliff Mountjoy-Venning, Pallavi Muraleedharan, Thara Rangaswamy, Chris M. Varghese, Mathew Varghese, K. Srinath Reddy, Christopher J. L. Murray, Soumya Swaminathan, and Lalit Dandona. Gender differentials and state variations in suicide deaths in india: the global burden of disease study 1990–2016. *The Lancet Public Health*, 3(10):e478–e489, Oct 2018.

[De Choudhury *et al.*, 2013] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, page 47–56, New York, NY, USA, 2013. Association for Computing Machinery.

[Dong *et al.*, 2024] Yi Dong, Ronghui Mu, Gaojie Jin, Yi Qi, Jinwei Hu, Xingyu Zhao, Jie Meng, Wenjie Ruan, and Xiaowei Huang. Building guardrails for large language models. *arXiv preprint arXiv:2402.01822*, 2024.

[Gaiha *et al.*, 2020] Shivani Mathur Gaiha, Tatiana Taylor Salisbury, Mirja Koschorke, Usha Raman, and Mark Petticrew. Stigma associated with mental health problems among young people in india: a systematic review of magnitude, manifestations and recommendations. *BMC Psychiatry*, 20(1):538, Nov 2020.

[Gakkhar *et al.*, 2023] Shivam Gakkhar, P. Lakshmi Nirisha, Gajanan Sabhahit, Patley Rahul, Nithesh Kulal, Nisha R. Harshitha, N. Manjunatha, Jagadisha Thirthalli, Naveen C. Kumar, Adarsha Alur Manjappa, Rajani Parthasarthy, Prabhat Kumar Chand, Sanjeev Arora, and Suresh Badamath. Reducing the treatment gap for psychiatric disorders – the role of accredited social health activists in south india. *World Social Psychiatry*, 5(1), 2023.

[Gala *et al.*, 2024] Jay Gala, Thanmay Jayakumar, Jaavid Aktar Husain, Mohammed Safi Ur Rahman Khan, Diptesh Kanojia, Ratish Puduppully, Mitesh M Khapra, Raj Dabre, Rudra Murthy, Anoop Kunchukuttan, et al. Airavata: Introducing hindi instruction-tuned llm. *arXiv preprint arXiv:2401.15006*, 2024.

[Garg *et al.*, 2019] Kabir Garg, C Naveen Kumar, and Prabha S Chandra. Number of psychiatrists in india: Baby steps forward, but a long way to go. *Indian J Psychiatry*, 61(1):104–105, January 2019.

[Gururaj *et al.*, 2016] G Gururaj, Mathew Varghese, VNRKP Benegal, G NIMHANS Rao, K Pathak, LK Singh, and R Misra. National mental health survey of india, 2015-16: Summary. *Bengaluru: National Institute of Mental Health and Neurosciences*, pages 1–48, 2016.

[Khosravi *et al.*, 2024] Mohsen Khosravi, Zahra Zare, Seyyed Morteza Mojtabaeian, and Reyhane Izadi. Artificial intelligence and decision-making in healthcare: A thematic analysis of a systematic review of reviews. *Health Services Research and Managerial Epidemiology*, 11:23333928241234863, 2024.

[Kumar *et al.*, 2022] Channaveerachari Naveen Kumar, K. Rakesh Chander, Shweta Gowda, Kudumallige K. Suresha, D. M. Harisha, Patley Rahul, Sharad Philip, Prateek Varshney, Vinay Basavaraju, Narayana Manjunatha, Suresh Bada Math, and Jagadisha Thirthalli. Near zero treatment gap for schizophrenia: Can it be reality in rural india? *Indian Journal of Psychiatry*, 64(Suppl 3), 2022.

[Maheshwari *et al.*, 2024] Prerna Maheshwari, Sukriti Mukherjee, Hari Hara Suchandra, Rahul Patley, Ferose Azeez Ibrahim, Sumi Jain, AH Harsha, Narayana Manjunatha, Channaveerachari Naveen Kumar, and Suresh Bada Math. Transforming mental healthcare: A crude return-on-investment analysis of the chhattisgarh community mental healthcare tele-mentoring program (champ), india. 2024.

[Malathesh *et al.*, 2021a] Barikar C. Malathesh, Bhavya K. Bairy, Channaveerachari Naveen Kumar, P. Lakshmi Nirisha, Gopi V. Gajera, Praveen Pandey, Narayana Manjunatha, Aurobind Ganesh, Kanika Mehrotra, Binukumar Bhaskarapillai, Durai Murukan Gunasekaran, Sanjeev Arora, Narendra Kumar Sinha, and

Suresh Bada Math. Impact evaluation of technology driven mental health capacity building in bihar, india. *Psychiatric Quarterly*, 92(4):1855–1866, Dec 2021.

[Malathesh *et al.*, 2021b] Barikar C. Malathesh, Ferose Aziz Ibrahim, P. Lakshmi Nirisha, Channaveerachari Naveen Kumar, Prabhat Kumar Chand, Narayana Manjunatha, Suresh Bada Math, Jagadisha Thirthalli, Adarsha Alur Manjappa, Rajani Parthasarathy, Shanivaram Reddy, and Sanjeev Arora. Embracing technology for capacity building in mental health: New path, newer challenges. *Psychiatric Quarterly*, 92(3):843–850, Sep 2021.

[Menz *et al.*, 2024] Bradley D. Menz, Natansh D. Modi, Michael J. Sorich, and Ashley M. Hopkins. Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance: Weapons of Mass Disinformation. *JAMA Internal Medicine*, 184(1):92–96, 01 2024.

[Mikus *et al.*, 2018] Adam Mikus, Mark Hoogendoorn, Artur Rocha, Joao Gama, Jeroen Ruwaard, and Heleen Riper. Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data. *Internet Interventions*, 12:105–110, 2018.

[Muennighoff *et al.*, 2023] Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. Crosslingual generalization through multitask finetuning. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *ACL*, pages 15991–16111, July 2023.

[Murthy, 2017] R. Srinivasa Murthy. National mental health survey of india 2015-2016. *Indian journal of psychiatry*, 59(1):21–26, Jan 2017. PMC5419008.

[Omboni *et al.*, 2022] Stefano Omboni, Raj S. Padwal, Tourkiah Alessa, Béla Benczúr, Beverly B. Green, Ilona Hubbard, Kazuomi Kario, Nadia A. Khan, Alexandra Konradi, Alexander G. Logan, Yuan Lu, Maurice Mars, Richard J. McManus, Sarah Melville, Claas L. Neumann, Gianfranco Parati, Nicolas F. Renna, Philippe Ryvlin, Hugo Saner, Aletta E. Schutte, and Jiguang Wang. The worldwide impact of telemedicine during covid-19: current evidence and recommendations for the future. *Connected Health And Telemedicine*, 1(1), 2022.

[Patel *et al.*, 2018] Vikram Patel, Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y. Collins, Janice L. Cooper, Julian Eaton, Helen Herrman, Mohammad M. Herzallah, Yueqin Huang, Mark J. D. Jordans, Arthur Kleinman, Maria Elena Medina-Mora, Ellen Morgan, Unaiza Niaz, Olayinka Omigbodun, Martin Prince, Atif Rahman, Benedetto Saraceno, Bidyut K. Sarkar, Mary De Silva, Ilina Singh, Dan J. Stein, Charlene Sunkel, and JÜrgen UnÜtzer. The lancet commission on global mental health and sustainable development. *The Lancet*, 392(10157):1553–1598, Oct 2018.

[Pigoni *et al.*, 2024] Alessandro Pigoni, Giuseppe Delvecchio, Nunzio Turtulici, Domenico Madonna, Pietro Pietrini, Luca Cecchetti, and Paolo Brambilla. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Translational Psychiatry*, 14(1):140, Mar 2024.

[Scao *et al.*, 2022] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, et al. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.

[Shiffman *et al.*, 2008] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. Ecological momentary assessment. *Annual review of clinical psychology*, 4:1–32, 2008.

[Singh *et al.*, 2016] Aakansha Singh, Surendra K. Mattoo, and Sandeep Grover. Stigma associated with mental illness: Conceptual issues and focus on stigma perceived by the patients with schizophrenia and their caregivers. *Indian Journal of Social Psychiatry*, 32(2), 2016.

[Singh Sethi *et al.*, 2025] M. I. Singh Sethi, Rakesh C. Kumar, Narayana Manjunatha, Channaveerachari Naveen Kumar, and Suresh Bada Math. Mental health apps in india: regulatory landscape and future directions. *BJPsych International*, 22(1):2–5, 2025.

[Smith *et al.*, 2023] Katharine A Smith, Charlotte Blease, Maria Faurholt-Jepsen, Joseph Firth, Tom Van Daele, Carmen Moreno, Per Carlbring, Ulrich W Ebner-Priemer, Nikolaos Koutsouleris, Heleen Riper, Stephane Mouchabac, John Torous, and Andrea Cipriani. Digital mental health: challenges and next steps. *BMJ Ment Health*, 26(1), 2023.

[Thakur *et al.*, 2025] Monika Thakur, Radhakrishnan Govindan, Prerna Maheshwari, Satish Suhas, Hari Hara Suchandra, Sumi Jain, Narayana Manjunatha, Channaveerachari N. Kumar, and Suresh B. Math. Bridging the mental health divide: The torent initiative in chhattisgarh, india. *Indian Journal of Community Medicine*, 50(1), 2025.

[Torous *et al.*, 2018] John Torous, Mark E. Larsen, Colin Depp, Theodore D. Cosco, Ian Barnett, Matthew K. Nock, and Joe Firth. 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, Jun 2018.

[Üstün *et al.*, 2024] Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, et al. Aya model: An instruction finetuned open-access multilingual language model. *arXiv preprint arXiv:2402.07827*, 2024.

[Yadav *et al.*, 2023] Suryakant Yadav, Aathavan K K, Solveig Argeseanu Cunningham, Pravat Bhandari, Udaya Shankar Mishra, Aditi Aditi, and Ravita Yadav. Changing pattern of suicide deaths in india. *The Lancet Regional Health - Southeast Asia*, 16, Sep 2023.

[Yosep *et al.*, 2024] Iyus Yosep, Rohman Hikmat, Ai Mardhiyah, and Taty Hernawaty. A scoping review of digital-based intervention for reducing risk of suicide among adults. *Journal of multidisciplinary healthcare*, 17:3545–3556, Jul 2024. PMC11283240.

[Zang *et al.*, 2024] Chengxi Zang, Yu Hou, Daoming Lyu, Jun Jin, Shane Sacco, Kun Chen, Robert Aseltine, and Fei Wang. Accuracy and transportability of machine learning models for adolescent suicide prediction with longitudinal clinical records. *Translational Psychiatry*, 14(1):316, Jul 2024.