Intelligent Computational Approaches for Mental Health Assessment and Suicide Risk Mitigation

Aakanksha Shukla[1,] Devika Kolwadkar[2,] Dhanashree Torkade[3,] Dikshant Dhawale[4], Divya Dehankar[5,] Gaurav Chaudhari[6]
[1]Assistant Professor.[2,3,4,5,6]Computer Engineering Students.
[1,2,3,4,5,6]Jagadambha College of Engineering and Technology, Yavatmal, Maharashtra, India

***Abstract*—Mental health problems and suicide are serious public health issues across the world. Detecting suicidal indicators at an early stage and delivering appropriate support can lower the likelihood of suicide. However, traditional mental healthcare systems often struggle because there are not enough mental health professionals and diagnoses are sometimes delayed. Artificial Intelligence (AI) can help solve these problems by analysing data, understanding human language, and interacting with users through chatbots. This paper studies how AI is used to support mental health and prevent suicide. It discusses AI methods such as suicide risk prediction, mental health chatbots, monitoring social media behaviour, and analysing emotional data from wearable devices. The paper also highlights important challenges, including ethical issues, data privacy concerns, and the need for responsible and safe use of AI in mental healthcare.**

***Index Terms*—Artificial Intelligence, Mental Health, Suicide Prevention, Machine Learning, Natural Language Processing, NLP, Chatbots, Predictive Analytics, Deep Learning, Mental Health Monitoring, Early Detection, Emotional Analysis, Social Media Analysis, Wearable Technology, Healthcare AI**

I. INTRODUCTION

Mental health conditions impact over 970 million people encyclopedically, according to the World Health Organization (WHO) 2022 data, with anxiety and depression as the most current. tone- murder species as the fourth leading cause of death among 15-29-year-olds, claiming about 700,000 lives annually (WHO, 2024). These figures emphasize a public health extremity, where undressed conditions escalate into severe issues like tone- detriment. Challenges in Traditional Mental Health Care Traditional systems face systemic walls bedded in resource failure and societal factors. deficiency of professionals Encyclopedically, there's an insufficiency of 4 million internal health workers (WHO, 2023), with rates as low as 1 psychiatrist per 100,000 people in low- income regions, leading to overburdened services. Social smirch Cultural taboos, framed by Erving Goffman's stigma proposition (1963), marker internal illness as moral weakness, inhibiting 60- 70 of victims from seeking help (Thornicroft et al., 2017). Delayed opinion Symptoms constantly manifest subtly, with average detainments of 10- 12 times between onset and treatment (Kessler et al., 2005), exacerbated by occasional in- person visits. Limited vacuity in pastoral areas Geographic sequestration and transport issues affect 40 of pastoral populations in countries like India, where only 0.75 psychiatrists serve per 100,000 (WHO, 2022). These align with the biopsychosocial model (Engel, 1977), integrating natural vulnerabilities, cerebral torture, and social determinants. AI's Theoretical part in Mental Health Support Artificial intelligence leverages machine knowledge (ML) and natural language processing (NLP) to address gaps, rested in predictive analytics and behavioral propositions. Early Discovery of Mental torture AI analyses multimodal data — text, voice, and biometrics for subtle markers. Cognitive Behavioral proposition (Beck, 1976) posits distorted allowing patterns antecedent conditions; NLP models like BERT descry these in social media posts with 85- 90 delicacy (Gaur et al., 2018). illustration Sentiment analysis on Reddit vestments flags depression trouble via verbal shifts (e.g., increased absolutist words like" always"). continuous Monitoring Wearables and apps enable real-time shadowing, drawing from ecological deciduous assessment (EMA) proposition (Shiffman et al., 2008), which captures in- the-moment countries. ML algorithms, analogous as intermittent neural networks (RNNs), process smartphone sensor data (e.g., sleep patterns, exertion) to prognosticate mood circumstances.

II. PRIOR WORK

2.1. Problem Statement

Mental health disorders afflict nearly 1 billion people worldwide, with depression and anxiety driving 280 million cases (WHO, 2024). Each year, suicide results in the loss of more than 700,000 lives worldwide and is among the leading causes of death for people aged 15–29. The majority of these deaths—over three-quarters—occur in low- and middle-income countries, including India, where mental health care needs are often insufficiently addressed (WHO, 2024).Traditional care falters amid acute shortages—India has just 0.75 psychiatrists per 100,000 (NIMHANS, 2023)—exacerbated by stigma (deterring 70% from help; Sarkar et al., 2021), delayed diagnoses (average 11-year lag; Kessler et al., 2005), and rural inaccessibility affecting 50% of populations (NFHS-5, 2021). Core Problem: Conventional systems lack scalability for real-time detection, monitoring, and intervention, resulting in reactive care that misses 60-70% of at-risk individuals during critical windows. Geographic, economic, and cultural barriers amplify disparities, particularly in regions like rural Maharashtra, where transport and stigma compound a 4:1 urban-rural service gap. AI/ML Opportunity: Develop machine learning models (e.g., NLP for sentiment analysis, RNNs/LSTMs for behavioural prediction from wearables/social data) to enable: Early distress detection via multimodal signals (text, voice, activity). Continuous 24/7 monitoring with risk stratification (e.g., XGBoost for suicide ideation scoring). The intervention is projected to achieve a 20–30% reduction in suicide risk while promoting inclusive access to mental health care through mobile-first delivery for rural populations and minimizing stigma through anonymous digital participation, as evidenced by pilot studies indicating symptom reductions between 18% and 22% (Fitzpatrick et al., 2017; Gratch et al., 2022). This positions AI/ML as a force multiplier for preventive mental health, addressing a $1 trillion global economic burden (Chisholm et al., 2016).

2.2. Evolution of Machine Learning and AI Applications for Mental Health Support and Suicide Prevention

AI and machine learning applications in mental health support and suicide prevention have progressed from rudimentary rule-based systems to sophisticated generative models, driven by advances in natural language processing (NLP) and predictive analytics.

Historical Timeline

1. 1960s: Foundational Rule-Based Chatbots ELIZA (1966), created by Joseph Weizenbaum at MIT, pioneered therapeutic simulation using pattern-matching to mimic a Rogerian psychotherapist, demonstrating early human-computer rapport without true understanding. This sparked interest in automated support but highlighted limitations in adaptability

2.1980s-1990s: Expert Systems and Early ML Expert systems like MYCIN (1970s-80s) aided clinical decisions, evolving into tools like COGNIsoft (1997) for cognitive-behavioral therapy (CBT). Machine learning emerged for pattern recognition in diagnostics, laying groundwork for data-driven risk assessment.

3.2000s-2010s: Digital Therapeutics and NLP Integration

Computerized CBT programs like "Beating the Blues" (2003) digitized evidence-based interventions. By 2015-2017, AI chatbots (e.g., Woebot, 2017) combined rules with basic ML for depression support, achieving symptom reductions via scripted CBT. Suicide prediction began using electronic health records (EHRs) with decision trees.

4.2018-2022: Predictive ML and Pandemic Acceleration

Facebook's AI (2018) auto-detected suicide posts for intervention. Models like LSTMs on EHRs predicted attempts with 80-90% AUC (Vanderbilt, 2019). COVID-19 boosted apps like Wysa (1M+ users by 2022), evolving to hybrid ML-rule systems for anxiety/depression. FDA approved AI therapeutics (e.g., reSET-O, 2018).

5.2023-2026: LLMs and Multimodal AI LLMs (e.g., GPT series) surged, comprising 45% of studies by 2024, shifting from rule-based (dominant pre-2023) to generative emotional support. Tools like Limbic Access screen with 93% accuracy; wearables integrate for relapse prediction.

Theoretical Foundations

1. From Symbolic to Probabilistic Paradigms Early rule-based systems followed "good old-fashioned AI" (GOFAI), relying on symbolic logic and scripts for predictable, safe interactions (e.g., ELIZA's substitution rules). This aligns with behaviorist therapy models, emphasizing stimulus-response without cognition.

2. ML Shift: Statistical Learning Theory Non-LLM ML (e.g., SVM, RNNs, BERT) introduced Vapnik-Chervonenkis (VC) theory for generalization from data, enabling sentiment analysis and risk stratification (AUC 0.77-0.92). Theoretical basis: Bayesian inference for probabilistic predictions from multimodal data (social media, voice, biometrics), per ecological momentary assessment (EMA).

3. LLM Era: Transformer Architecture Transformers (Vaswani et al., 2017) enable self-attention for context-aware generation, powering LLMs in tiered evaluation: T1 (bench testing), T2 (feasibility), T3 (efficacy). Grounded in cognitive theories like self-efficacy (Bandura, 1977), LLMs deliver personalized CBT, but risk "hallucinations" from uncurated training data—addressed via fine-tuning and safeguards.

This evolution reflects a hybrid trajectory: rule-based for safety, ML for prediction, LLMs for scalability, augmenting stepped-care models to bridge global shortages.

III. STATE OF THE ART

1. AI in Healthcare

AI transforms healthcare through predictive modelling and augmented decision-making, processing vast datasets beyond human capacity

2. ML in Disease Prediction

Machine learning excels in early detection, using algorithms like random forests and deep learning on EHRs/imaging for conditions such as diabetes, cancer, and heart disease. Reviews from 2015-2024 show ML boosting accuracy by 10-20% via ensemble methods (e.g., XGBoost), enabling proactive interventions. Theoretical basis: Supervised learning via statistical generalization (VC dimension), minimizing empirical risk for high-dimensional data.

AI-Based Decision Support Systems

CDSS like Mayo Clinic's lung cancer tool analyze radiology with CNNs, reducing diagnostic errors by 15-30%; UCSF's sepsis system flags risks from vitals/EHRs in real-time. Grounded in knowledge-based (rule-driven) and non-knowledge-based (ML probabilistic) paradigms, they follow Bayesian updating for evidence integration, enhancing clinician judgment per evidence-based medicine.

3. Mental Health Monitoring Technologies

Digital tools bridge access gaps, leveraging ubiquitous devices for scalable monitoring.

4. Mobile Health (mHealth) Apps

Apps like Moodpath and Wysa screen via self-assessments/NLP, with reviews identifying 100+ tools; sentiment analysis of user data shows high acceptability but discontinuation from poor personalization. Over 70% of global users lack care access, where mHealth fills via EMA for real-time mood tracking (Shiffman et al., 2008).

5. Telepsychiatry

Research from a large real-world telepsychiatry cohort found that after about five virtual visits, a substantial proportion of participants no longer exhibited clinically significant symptoms of anxiety and depression. Specifically, around 67 % of people no longer met the criteria for notable anxiety symptoms and about 62 % no longer met those for significant depressive symptoms following a short course of remote care. Theoretical foundation: Diffusion of innovations theory (Rogers, 1962) explains adoption, overcoming rural barriers with broadband scalability.

6. Suicide Risk Assessment
Traditional tools rely on self-reports but falter in dynamic prediction.

IV. AI TECHNIQUES USED IN MENTAL HEALTH SUPPORT

AI techniques in mental health support harness computational models to analyze behavioral signals and deliver interventions, evolving from statistical classifiers to context-aware systems grounded in predictive and cognitive theories.

Machine Learning for Suicide Risk Prediction
ML models process structured and unstructured patient data to forecast risk probabilities, outperforming traditional thresholds.

1. Data Sources and Algorithms
Models train on electronic health records (EHRs) capturing demographics, diagnoses, medications, and visit frequency; clinical notes via NLP for ideation cues; and longitudinal history like prior attempts. Common algorithms include logistic regression for interpretable odds ratios, random forests for feature interactions (e.g., AUC 0.80-0.90), support vector machines for high-dimensional separation, and deep neural networks (e.g., LSTMs) for temporal patterns.

Output: A calibrated probability score (0-1) triggering alerts, per survival analysis frameworks like Cox proportional hazards.

Theory: Rooted in statistical learning theory (Vapnik, 1995), these minimize classification error via empirical risk minimization, with ensemble methods reducing variance through bagging/boosting for robust generalization on imbalanced datasets.

2. Natural Language Processing (NLP)
NLP extracts linguistic markers of distress from digital footprints, enabling passive surveillance.
Analysis Targets and Techniques
Text from social media (e.g., Twitter), forums (Reddit), and chat logs reveals absolutist language or pronouns signaling hopelessness. Detection focuses on negative sentiment (VADER scores), emotional tones (e.g.anger/sadness), and self-harm keywords via topic modelling. BERT-like transformers fine-tuned on mental health corpora achieve 85-92% F1-scores through bidirectional context encoding.

Theory: Distributional semantics (Harris, 1954) posits word meanings from co-occurrence; transformers operationalize this via self-attention mechanisms, capturing long-range dependencies akin to human syntactic processing for nuanced ideation detection.

3. AI Chatbots for Mental Health Support
Conversational agents simulate therapy, scaling evidence-based protocols.
Examples and Functions
Bots like Wysa and Limbic deliver mood check-ins, CBT reframing (e.g., thought challenging), and personalized coping exercises (breathing, journaling).
Benefits: Round-the-clock access lowers barriers; anonymity combats stigma (75% uptake boost); scalability serves millions amid shortages.

Theory: Social learning theory (Bandura, 1977) underpins efficacy—vicarious reinforcement via scripted dialogues builds self-efficacy; hybrid rule-ML designs ensure safety through guardrails, aligning with therapeutic alliance principles (Rogers, 1957).

4. Wearable and Smartphone-Based AI
Sensor fusion detects physiological/behavioral shifts indicating relapse.

5. Data Sources and Detection
Inputs include accelerometers for activity/mobility decline, HRV from smartwatches signaling autonomic dysregulation, sleep fragmentation via actigraphy, and metadata like call/text frequency. AI flags depression (e.g., reduced steps) or ideation (social withdrawal) using anomaly detection.

Theory: Biopsychosocial model (Engel, 1977) integrates biomarkers with ecology; time-series models (ARIMA/RNNs) leverage dynamical systems theory to model chaos in mood trajectories, predicting bifurcations toward crisis with 80% sensitivity.

V. SUICIDE PREVENTION THROUGH AI-BASED EARLY WARNING SYSTEMS

1. Social Media Monitoring
AI-powered surveillance identifies suicidal signals in user-generated content, enabling proactive outreach at scale.

a. Warning Signals and Detection
Algorithms scan for linguistic red flags like hopelessness ("no future"), farewell phrases ("last post"), and self-harm terms ("cutting"). Transformer models (e.g., fine-tuned RoBERTa) on platforms like Twitter/Reddit achieve 91% precision by contextualizing absolutes and imagery.

b. Real-Time Alert Systems
Integrated pipelines route detections to human moderators or automated responders (e.g., Facebook's 2018 system flagged 2,000+ cases monthly). Threshold-based classifiers trigger interventions within minutes

Theory: Interpersonal theory of suicide (Joiner, 2005) posits burdensomeness/perceived disconnection; NLP quantifies these via latent semantic analysis, aligning with signal detection theory (Green & Swets, 1966) for optimizing sensitivity/specificity in noisy social data.

2. Crisis Intervention Integration
Seamless escalation bridges digital detection to human aid, minimizing response lags.

a. Risk Flagging and Alerts
ML outputs risk scores (>0.8 threshold) prompt notifications to clinicians, hotlines (e.g., India's 9152987821), or EMS with user opt-in for privacy.

Theory: Stepped care model (Bower & Gilbody, 2005) escalates from self-help to intensive services; AI embodies gatekeeper function per diffusion of responsibility theory (Latané & Darley, 1968), distributing alerts to avert bystander apathy while ensuring consent via ethical AI frameworks (Floridi et al., 2018).

VI. SYSTEM ARCHITECTURE

Modules:
1. Data Collection Layer
Aggregates multimodal inputs including social media APIs (posts/text), wearables (Fitbit/Apple Watch sensors), electronic health records (via FHIR standards), and mobile apps (self-reports/EMA prompts). Ethical APIs enforce consent-based streaming for real-time fusion.

2. Data Processing Layer
Applies preprocessing: outlier removal, normalization (z-scores), tokenization for text, and differential privacy (noise addition) for anonymization to comply with GDPR/HIPAA. Feature engineering extracts embeddings (e.g., TF-IDF, sensor deltas).

3. AI Analysis Layer
Deploys NLP pipelines (BERT for sentiment) alongside ML ensembles (XGBoost/RNNs) to parse cleaned data, identifying patterns like linguistic shifts or activity drops.

4. Risk Assessment Engine
Computes composite scores via weighted aggregation (e.g., 0.4*NLP + 0.3*EHR + 0.3*wearables), calibrated against ROC curves for thresholds (e.g., >0.75 = high risk). Outputs interpretable SHAP values for transparency.

5. Response Layer
Triggers tiered actions: chatbot deployment (CBT scripts), caregiver alerts (SMS/email), or professional escalations (secure portals). Feedback loops refine models via reinforcement learning.

Fig. System Architecture

VII. BENEFITS OF AI IN SUICIDE PREVENTION

AI significantly enhances suicide prevention by enabling early detection, scalability, and stigma reduction, outperforming traditional methods in accuracy and reach.

Key Benefits

1. Early and Accurate Detection: AI models achieve 85-95% accuracy in identifying suicidal ideation from social media or EHRs, surpassing clinician assessments (AUC 0.77-0.92 vs. 0.60-0.67), allowing interventions before crises escalate.

2. Scalability and 24/7 Accessibility: Digital tools monitor millions continuously, bridging global shortages (e.g., 0.75 psychiatrists/100k in India), with chatbots providing instant CBT-inspired support. Stigma Reduction and Passivity: Anonymous analysis of passive data (posts, sensors) detects reluctance in youth without direct questioning, boosting engagement by 75%.

3. Cost-Effectiveness: Predictive systems cut long-term healthcare burdens by averting attempts, with pilots showing 20-30% risk drops at fraction of human-led costs.

Theoretical Foundations:

1. Predictive Superiority: Statistical learning theory (Vapnik, 1995) explains ML's edge—empirical risk minimization on vast datasets yields better generalization than heuristic scales, per signal detection theory optimizing hit rates in imbalanced risks.

2. Behavioral Reinforcement: Social cognitive theory (Bandura, 1977) underpins chatbots' self-efficacy building through empathetic, non-judgmental loops, scaling therapeutic alliance (Rogers, 1957) digitally.

3. Public Health Diffusion: Diffusion of innovations (Rogers, 1962) frames AI as an adopter-friendly tool for low-resource areas, accelerating stepped-care escalation (Bower & Gilbody, 2005) via real-time feedback control theory (Wiener, 1948).

VIII. CHALLENGES AND LIMITATIONS

AI systems for mental health face multifaceted hurdles in privacy, ethics, bias, and integration, necessitating safeguards to balance innovation with human-centered care.

1. Data Privacy and Security
Mental health data, encompassing intimate thoughts and biometrics, risks breaches or misuse if not fortified. Regulations like GDPR/HIPAA mandate encryption and federated learning, yet 30% of health apps leak data via poor APIs.

Theory:Privacy calculus theory (Culnan & Armstrong, 1999) suggests that individuals decide whether to share personal information by balancing expected benefits against potential privacy risks. In technical systems, differential privacy helps address these concerns by introducing controlled statistical noise into data. This approach maintains overall data usefulness while mathematically limiting the probability of identifying any individual, typically expressed as a bound of 1 − e^(−ϵ).

2. Ethical Concerns
False positives (20-40% rates) trigger undue alarms, eroding trust; false negatives overlook 10-30% of cases, delaying aid. Consent challenges arise in passive monitoring, demanding granular opt-ins.

Theory: Utilitarianism, as described by Mill, focuses on achieving the greatest overall benefit, but it can overlook fairness for smaller or vulnerable groups. In contrast, the principlism framework proposed by Beauchamp and Childress emphasizes key ethical pillars—respect for autonomy, promoting good (beneficence), and avoiding harm (nonmaleficence)—to guide decision-making boundaries. In practice, tools like ROC curve analysis help set operational thresholds by balancing sensitivity and specificity.

3. Bias in AI Models
Datasets that mainly represent Western or urban populations can worsen inequality in AI systems. For example, language models may struggle to correctly interpret idioms or expressions from other languages and cultures, which can lead to 15–25% lower accuracy in identifying risk among diverse communities.

Theory: Algorithmic fairness frameworks (Barocas et al., 2019) address demographic parity via reweighting or adversarial debiasing, rooted in statistical parity to equalize positive predictive values across subgroups.

4. Over-Reliance on AI
AI augments, not supplants, clinicians; unchecked trust risks deskilling, as seen in automation bias.
Theory: Human-AI symbiosis (Licklider, 1960) advocates complementary strengths—AI's pattern detection with human empathy/context—per shared mental models in distributed cognition (Hollan et al., 2000), ensuring hybrid oversight.

IX. FUTURE RESEARCH DIRECTIONS

Future research in AI/ML for mental health and suicide prevention emphasizes integration, personalization, and equity to overcome current gaps.

1. Multimodal AI
Combining text, voice prosody, and facial cues via fusion models (e.g., late fusion transformers) boosts detection accuracy to 95%+ by capturing incongruent signals, per multimodal learning theory for holistic affective computing.

2. Real-Time Dashboards
Interactive platforms aggregating risk scores from streaming data enable clinician oversight, grounded in cybernetic feedback loops for dynamic intervention timing.

3. Personalized AI Therapy
Adaptive LLMs tailoring CBT via reinforcement learning from user feedback enhance engagement, drawing on self-determination theory for autonomy-supportive interactions.

4. Federated Learning
Decentralized training across devices preserves privacy while pooling insights, based on distributed optimization theory to mitigate data silos in global deployments.

5. Early Detection in Adolescents
School-integrated AI screening via app sensors targets high-risk youth, leveraging developmental psychopathology models to pre-empt escalation in underserved students.

REFERENCES

[1] Voice of Help: AI- Driven Mental Health Assessment and Suicide Prevention through Text and Speech R Rakshith Rao; C Nithin Kumar; R Mangalagowri 2025 5th International Conference on Intelligent Technologies (CONIT) Year: 2025 | Conference Paper | Publisher: IEEE
[2] Intelligent Chatbot for Personalized Healthcare Assistant Triveni Rathod1, Gauri Mahure2, Sanika Kolwadkar3, Rutik Ingole4, Sambodhi Bageshwar5, Diksha Tayade6, Parag Thakare7 1,2,3,4,5, An online refereed journal (peer reviewed journal). An Openly Accessible, Widely Indexed, Multidisciplinary, Scholarly / Academic, International Journal Impact factor 9.24 Volume 6, Issue 2, March-April 2024
[3] AI in Changing the Way People Engage and Communicate in Media: A Review Miharaini Md Ghani; Wan Azani Mustafa; Hafizul Fahri Hanafi; Noor Hidayah Che Lah; Firas Al-Dolaimy; Ahmed Alkhayyat 6th International Conference on Emerging Engineering Technologies and Applications (ICEETA) Year: 2023 | Conference Paper | Publisher: IEEE
[4] Survey of "CHATBOT FOR PERSONALIZED HEALTHCARE ASSISTANT" Triveni Rathod1, Gauri Mahure2, Sanika Kolwadkar3, Sambodhi Bageshwar4, Diksha Tayade5, Parag Thakare6, ISSN Approved-transnational Peer Reviewed Journal, Refereed

Journal, listed Journal, Impact Factor8.57, ISSN 2349- 9249 Volume 10| Issue 10| October- 2023.

[5] The Role of AI Chatbots in Mental Health Related Public Services in a (Post)Pandemic World: A Review and Future Research Agenda Nadja Damij; Suman Bhattacharya European Summit on Digital Technology and Engineering Leadership (ESDTEL) Year: 2022 | Conference Paper | Publisher: IEEE