

A Multilingual AI Framework for Real-Time Psychological Stress and Trauma Assessment in Emergency Helpline Systems

Sahil Kshirsagar*, Dharna Totlani**, Vikas Gupta*, Gayatri Oman*, Shripad S. Bhide**

**(Student, Department of Master of Computer Applications, PES Modern College of Engineering, Shivajinagar Pune-05, India*

***(Guide, Department of Master of Computer Applications, PES Modern College of Engineering, Shivajinagar Pune-05, India*

Email: sahil_kshirsagar_mca@moderncoe.edu.in | dharna_totlani_mca@moderncoe.edu.in |

vikas_gupta_mca@moderncoe.edu.in | gayatri_oman_mca@moderncoe.edu.in | shripad.bhide@moderncoe.edu.in

Abstract

Victims and complainants who come to the National Helpline Against Atrocities (NHAA 14566) and its integrated grievance portal are routinely reporting extreme levels of trauma, PTSD, emotional distress and co-existing suicidal thoughts as a result of systemic harassment, caste-based atrocities and other socio-legal emergencies. The intake triage process is under considerable strain due to the sheer volume of calls, the cognitive fatigue of the intake officers, the prevalence of multilingual and code-switched regional narratives (particularly in Marathi and Hindi), and the inherently subjective nature of manual psychological screening. We present an end-to-end, multimodal, real-time framework for the assessment of stress and trauma which is specifically designed for the NHAA telephony and digital complaint portal channels. The architecture combines an Acoustic Stream and a Textual Natural Language Processing (NLP) Stream through a cross-attentive fusion classifier. The acoustic stream utilises a range of features from the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) to pitch jitter, shimmer, formant dispersion and Mel-Frequency Cepstral Coefficients (MFCCs), which are processed through a Bidirectional Gated Recurrent Unit (BiGRU) to extract the affective biomarkers of voice stress. The textual stream takes the spoken transcripts from the telephony channel and the written complaints from the portal as input and uses a cross-lingual transformer (XLM-RoBERTa / Sentence-BERT) coupled with a zero-shot DeBERTa-v3 Natural Language Inference (NLI) cross-encoder to map the disclosures directly to the DSM-5 criteria for PTSD and the PHQ-9 criteria for depression. The experimental results show that the multimodal fusion approach achieves an overall accuracy of 93.84% (Macro F1: 0.9261, ROC-AUC: 0.9712) in the classification of trauma and distress, which is significantly better than the unimodal text (86.12%) and unimodal voice (81.75%) baselines. The framework also incorporates the Integrated Ethical Approach for Computational Psychiatry (IEACP) and participatory injustice principles with bias-audited risk stratification, explainable TreeSHAP symptom attribution for the caseworkers and zero latency emergency routing in the case of an imminent suicidal crisis.

Keywords — National Helpline Against Atrocities (NHAA 14566), Multimodal Stress Assessment, Post-Traumatic Stress Disorder (PTSD), Acoustic Correlates, Marathi Speech Processing, Cross-Lingual Transformers, DSM-5 PHQ-9, Computational Psychiatry, Participatory Justice, Suicidal Ideation Prevention.

I. INTRODUCTION

The National Helpline Against Atrocities (NHAA 14566) run by the Department of Social Justice and Empowerment, Government of India is an emergency redressal mechanism for SCs, STs and other marginalised people and victims of caste atrocities and other civil atrocities. It is available 24x7 across the country via the toll free number 14566 and through its integrated grievance web portal. The helpline is used as a primary mechanism for filing of FIRs, seeking protective intervention and for requesting immediate psychosocial assistance. The people who come to the helpline are invariably in a state of acute psychological distress, suffering from severe trauma, severe anxiety and PTSD as a result of recent physical violence, coercion and/or generations of social subjugation. In some of the most extreme cases of desperation, complainants have suicidal ideation [15], [16]. Running an emergency

helpline of this scale creates enormous problems on the frontline. The frontline staff (call handlers and intake officers) are faced with hundreds of emotionally laden audio calls and written complaints from across India's linguistic landscape. A huge volume of calls are in regional languages (Marathi, Hindi etc) and often in a code-mixed form. When the call queue is at its worst, manual triage is very susceptible to intake officer cognitive fatigue, variable subjective scoring of severity and a delay in the identification of a life-threatening trauma spike. The complainant may well present their trauma in a non-linear fashion. Their words may be very guarded, very cautious for fear of local reprisal, but their voice will be screaming with panic and affective decompensation [6], [10] in terms of the acoustic parameters (pitch tremor, vocal fry, micro-pauses, erratic F0 etc). Recent advances in speech emotion recognition (SER) [8], [9], [12], [13] and clinical NLP [1], [19] have the

potential to enable automated crisis stratification. Deep neural networks and Large Language Models (LLMs) can be used to quickly process clinical interview transcripts [16], [19] and detect emotional valence from speech acoustic features [10], [14]. The problem with current off-the-shelf mental health AI is that they are overwhelmingly English based, they typically work on either the acoustic or the textual modality in isolation [17], [18], they don't make use of any standard clinical framework for risk assessment (e.g. the DSM-5) [1], [9] and they lack any real ethical guardrails to prevent participatory injustice in their interactions with historically marginalised populations [22], [23]. To address the gap between the operational and technological aspects of the NHAA 14566 system we have developed an explainable, multimodal, real time stress and trauma assessment module. The module is designed to take in both the voice stream and the written complaint at the same time using a single multi-task architecture. It uses acoustic feature extraction (aGeMAPS, MFCCs and BiGRU networks) in conjunction with cross-lingual transformer inference (XLM-RoBERTa and sliding window DeBERTa-v3 NLI) to produce an objective Trauma and Psychological Distress Index (TPDI), to flag any immediate suicidal risk and to provide the human officer with an auditable clinical justification for their decision.

II. Literature review & taxonomy The automated screening of trauma, distress and affective states is at the nexus of acoustic speech signal processing, cross-lingual NLP and computational psychiatry. This section reviews the state of the art across four key research strands.

A. Acoustic correlates and Speech Emotion Recognition (SER) The human voice is a rich source of physiological and autonomic markers of acute stress. There have been a number of seminal surveys in the area of speech emotion recognition [9],[11],[12] which have systematically catalogued emotion databases, preprocessing techniques and acoustic representations. As Schuller et al. and a number of subsequent systematic reviews have established [8],[13], the vocal tract is a direct conduit for the effects of sympathetic nervous system activation and will automatically respond to this with changes in pitch fundamental frequency (F0), jitter (cycle-to-cycle variation in frequency), shimmer (amplitude perturbation) and spectral flux. A 2025 review of negative emotions and stress via acoustic correlates [10] found that speakers in a depressed or traumatised state have reduced pitch dynamics, increased glottal leakage and longer response latencies. Voice based stress detection methodologies from 2021 onwards [14] have shown that by combining a range of standardised acoustic feature sets (e.g. the extended Geneva Minimalistic Acoustic Parameter Set eGeMAPS) with deep temporal models you can capture both the micro-prosodic tremors and the macro-utterance level affective decline. Finally, studies which have modelled physiological and appraisal indicators using audio features have shown that the acoustic stress correlates are robust across a range of conversational contexts [6].

B. Regional and Marathi speech stress modelling A major limitation of the current SER and voice stress literature is the overwhelming concentration on

high-resource languages (i.e. English and Standard German). In a significant 2026 study a hybrid deep learning framework for Marathi speech based stress detection was proposed using Gated Recurrent Units (GRUs) and hand-crafted acoustic features [21]. The study found that the combination of traditional mel-frequency features with temporal GRU layers effectively captures the variations in phoneme duration and the dialectal stress characteristics of Marathi speech, resulting in good discriminative sensitivity for the detection of stress in the context of real world, regional conversational speech. It would be vital to adapt such a language specific approach for use in emergency helplines in Western India where the caller is likely to be speaking in Marathi, Hindi and/or a mix of both.

C. Cross-Lingual & Clinical Trauma/PTSD Detection Post Traumatic Stress Disorder (PTSD) is a serious psychiatric disorder that arises following a traumatic event and is typified by intrusive memories, hyperarousal and emotional numbing. Systematic reviews and meta-analyses [1],[2] have demonstrated the burgeoning use of machine learning in the automated diagnosis of PTSD, with the finding that while traditional classifiers have achieved a reasonable degree of accuracy they have been found to perform poorly in terms of cross-site generalisation. In the context of clinical dialogue, analysis of speech in the course of a virtual agent conversation has been shown to be effective in distinguishing between patients with PTSD and healthy controls [5], with vocal flattening and reduced rate of speech being the key indicators. Progress has also been made in the area of speech based recognition and assessment of PTSD severity [4] with machine learning models being able to assign a level of severity on a continuous clinical scale. Systematic reviews of text, audio and video data [3] have all concluded that multimodal modelling is vastly superior to any single modality in the prediction of anxiety and PTSD. More recently a head to head comparison of NLP methods and Large Language Models (LLMs) on clinical interview transcripts [19] has shown that transformer based representations are superior to classical lexical baselines in the detection of subtle markers of cognitive avoidance. To address the linguistic limitations of such models, a comparison of BERT and XLM-RoBERTa with machine translation adaptation as a form of cross-lingual knowledge transfer has been carried out [20] and it has been found that cross-lingual transfer is a very effective way of screening for PTSD in non-English crisis narratives.

D. Multimodal Depression and Suicidal Ideation Screening Trauma and ongoing distress frequently lead to the co-existence of Major Depressive Disorder and an acute suicidal crisis. A number of systematic reviews looking at the use of NLP to detect and prevent suicide ideation [15] have shown that transformer-based semantic models are able to detect suicidal intent well in advance of any manual triage. A number of cross-sectional clinical studies [16] have demonstrated that machine learning applied to intake interviews can successfully identify latent markers of suicide in the form of spoken despair. At the same time, a range of multimodal depression detection systems using text, audio and video [17] as well as more sophisticated multimodal voice and text fusion approaches [18], have shown

that the combination of early acoustic blunting with the expression of worthlessness in the lexicon provides a far more robust diagnostic signal than text mining alone. E. Ethical AI Governance and Participatory Justice AI deployment in the sensitive areas of mental health and trauma carries with it enormous moral and socio-technical weight. The IEACP (Integrated Ethical Approach for Computational Psychiatry) framework [23] states that any clinical AI system must be procedurally transparent, empirically validated and clinically interpretable rather than a black box. There is also a significant body of critical work on the implications of AI for the mental health of vulnerable populations [22] which warns against 'participatory injustice' in which marginalised communities are subjected to algorithmic surveillance or to algorithmic triage models that are biased and which are not the product of inclusive participatory design, thus reinforcing existing inequalities. In the context of the NHAA 14566 infrastructure the triage algorithm must be equitable, unbiased across caste and language groups and fully interpretable to the human caseworkers.

III. PROBLEM FORMULATION AND SYSTEM ARCHITECTURE

The proposed real time assessment module is intended to be a non-intrusive integration into the existing NHAA 14566 telephony exchange and the web based grievance logging portal. Fig 1 shows the high level dual stream processing architecture.

[NHAA Intake Channels] --> (1) Voice Telephony (14566) & (2) Integrated Web Grievance Portal

|--> Stream A: Acoustic Feature Extraction (eGeMAPS + MFCCs -> BiGRU)

|--> Stream B: Multilingual NLP Pipeline (Indic ASR -> XLM-RoBERTa + DeBERTa-v3 DSM-5/PHQ-9 NLI)

|--> Cross-Attentive Multimodal Fusion -> Multimodal Distress & Trauma Classifier (XGBoost / Dense)

|--> Clinical Decision Support & Escalation Engine (TPDI Index, TreeSHAP Explainability, Emergency Triage)

Fig. 1. Architectural blueprint of the proposed multimodal trauma and stress assessment system for NHAA 14566.

A. Dual-Channel Intake Stream

The system has two data ingestion paths: (i) the Telephony Voice Stream (8kHz / 16kHz G.711) which is the audio from the live NHAA emergency helpline calls; and (ii) the Web Complaint Text Stream which is the written narrative of the complaint, the FIR and the structured grievance tickets that are all submitted via the web portal in Devanagari, Marathi or in Hindi and Romanised code-switched text.

B. Indic Speech-to-Text & Regional Dialect Normalisation

For spoken audio calls, the raw speech is fed in parallel to both the acoustic feature extractor and a high performance fine-tuned Indic ASR engine (Marathi and Hindi dialects supported). The transcripts that result go through four stages of automated text sanitisation: regex cleaning to remove transmission artefacts, phonetic repair of distress terms (e.g. fixing mistranscribed homophones which may be related to

suicide or assault), and code-switched colloquial mapping to preserve the emotional tone.

IV. Methodology & Multimodal Fusion Pipeline

To achieve clinical validity the framework models both the involuntary physiological voice tremors and the cognitive narrative expression of trauma.

A. Acoustic Feature Engineering & BiGRU Temporal Modelling

As per the established acoustic stress literature [10], [14], [21] the speech pipeline uses OpenSMILE to extract an 88 dimensional extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), with 40 dimensional MFCCs and delta-delta MFCCs (computed over 25ms Hamming windows with a 10ms hop size). The key parameters are:

1) Fundamental Frequency (F0) and Micro-Prosody: Shows vocal cord tension and small changes in pitch. Micro-tremors are measured using jitter, which tracks pitch changes from one cycle to the next, and shimmer, which measures changes in loudness.

2) Spectral & Energy Dynamics: spectral flux, spectral slope and harmonic to noise ratio (HNR) which reflect glottal leakage and vocal exhaustion as would be expected in the case of clinical trauma [10].

3) Bidirectional GRU Temporal Encoder: The frame level feature vectors are fed into a 2 layer Bidirectional Gated Recurrent Unit (BiGRU) [21] with hidden dimension $h=128$ to produce a 256 dimensional summary acoustic vector a_t which captures the emotional trajectory over the duration of the call.

B. Textual cross-lingual & DSM-5 NLI feature extraction

The textual stream handles both ASR transcripts and web grievance submissions through a dual-representation approach:

1) Dense Semantic Embeddings: The input text is tokenised and encoded using XLM-RoBERTa and 384 dimensional multilingual Sentence-BERT [20] to produce dense semantic embeddings which capture the holistic semantic representation of the text across the Indic languages without any translation penalty.

2) Sliding Window Zero-Shot Clinical Entailment: To deal with the token truncation issue and to anchor the predictions in the realm of diagnostic psychiatry [1],[9],[19] long complaint narratives are split up using a 400 word sliding window with a 50 word overlap. Each window is then assessed against a set of standard clinical hypothesis templates derived from the DSM-5 PTSD criteria (intrusion, avoidance, negative alterations in cognition and mood, hyperarousal) and the PHQ-9 depression/suicide items [9] via a DeBERTa-v3 NLI cross-encoder. The window level entailment probabilities are max-pooled to produce a structured 13 dimensional clinical symptom vector c_t .

C. Cross-Attentive Multimodal Fusion

The acoustic temporal embedding a_t (256-dim), the dense textual embedding e_t (384-dim) and the clinical symptom vector c_t (13-dim) are combined via a gated cross-attention mechanism. The cross-modal attention weights dynamically gate the contribution of voice prosody against verbal semantics: in the case of a caller who is speaking in very subdued, guarded

terms but with very pronounced acoustic tremor (e.g. high levels of jitter and very erratic pitch when under threat), the gating network will boost the acoustic stream weight. The resulting 653 dimensional fused representation is classified by an Optuna tuned gradient boosted decision ensemble (Multi-Class XGBoost) and a calibrated Softmax layer.

V. EXPERIMENTAL EVALUATION AND RESULTS

The model was evaluated on a benchmark comprising a synthesis of clinical and helpline interaction corpora. This includes the Marathi Speech Stress Corpus [21], clinical PTSD interview recordings [5], [19] and multilingual crisis helpline text data [16], [20]. The classification taxonomy is four tiered: (1) Acute Trauma / Active PTSD, (2) Severe Crisis & Suicidal Ideation, (3) Moderate Psychological Distress / Depression and (4) Non-Crisis / Procedural Grievance. TABLE I. COMPARATIVE PERFORMANCE OF MODALITIES ON VALIDATION BENCHMARK

Model / Modality Architecture	Accuracy (%)	Macro F1	Sensitivity	ROC-AUC
Acoustic Only (eGeMAPS + BiGRU) [21]	81.75	0.8042	0.8120	0.8874
Textual Only (XLM-RoBERTa + NLI) [20]	86.12	0.8530	0.8655	0.9238
Late Fusion (Average Probability)	89.40	0.8872	0.8910	0.9450
Proposed Multimodal Gated Attention	93.84	0.9261	0.9412	0.9712

As per Table I the proposed multimodal cross-attentive framework achieves an overall accuracy of 93.84% and an ROC-AUC of 0.9712 (a 7.72pp improvement over text alone and a 12.09 point improvement over acoustic alone). Most importantly in the high stakes Suicidal Ideation and Severe Trauma categories the sensitivity is 94.12% demonstrating that vocal biomarkers are able to detect the panic states that the caller is attempting to mask in their language.

A. Feature Sensitivity and TreeSHAP Interpretability

TreeSHAP analysis across all feature dimensions shows that in the acoustic stream F0 std, jitter local and low frequency spectral flux were the most predictive features during emotional outbursts. In the textual stream DeBERTa-v3 clinical entailment for DSM-5 Hyperarousal and PHQ-9 item 9 (thoughts of self-harm) were among the top 5 global predictors, again confirming that the clinical inference features are very much in line with the assessments of the clinical expert [1], [9], [19].

VI. ETHICAL DECISION-MAKING & PARTICIPATORY JUSTICE

Deploying AI in the context of the National Helpline against Atrocities (14566) requires robust ethical governance. The victims of atrocities are a historically marginalised and

vulnerable group and are particularly prone to being technologically marginalised [22].

A. Operationalizing the IEACP Ethical Framework

The module is built directly on the Integrated Ethical Approach for Computational Psychiatry (IEACP) [23] with three hard and fast guardrails: (1) Human in the Loop Supremacy: the AI model is strictly an advisory decision support tool and will never reject a complaint, will never delay a call going into the queue and will never automate the administrative dismissal of a complaint. (2) Algorithmic Transparency: the caseworker is given a clear visual dashboard showing the Trauma and Psychological Distress Index (TPDI) and plain language clinical justifications (e.g. 'Acoustic Panic: High Pitch Jitter detected; Narrative: DSM-5 Intrusive Trauma thoughts detected'). (3) Independent Auditability: all inferences are logged in an immutable way for post hoc psychiatric review.

B. Mitigating Participatory Injustice

To prevent participatory injustice [22] we evaluated the system across a range of dialects and socio-cultural speech patterns to ensure that people from non-urban or marginalised communities aren't disadvantaged by miscalibrated ASR or by biased sentiment thresholds. In the event of an acute suicidal crisis being identified the module will immediately route to the priority helplines (Tele-MANAS 14416 and KIRAN 1800-599-0019) and also alert the supervising officer.

VII. CONCLUSION AND FUTURE SCOPE

This is a new, explainable multimodal real-time framework for the assessment of stress and trauma for the National Helpline Against Atrocities (NHAA 14566) and its integrated grievance portal. It combines acoustic biomarkers of vocal distress (eGeMAPS, BiGRU) with cross-lingual transformer NLP (XLM-RoBERTa) and sliding window zero-shot DSM-5/PHQ-9 entailment to achieve 93.84% accuracy in the classification of trauma and acute psychological distress. It is built on the IEACP ethical architecture and the principles of participatory justice so that the most vulnerable victims of atrocities can be rapidly and appropriately supported with an empathic and objective crisis response. Future work will be to enable on-device edge deployment to reduce the telephony latency and to incorporate longitudinal tracking for those callers who are making multiple calls over time.

ACKNOWLEDGMENT

The authors would like to extend their thanks to the Department of Master of Computer Applications at P.E.S. Modern College of Engineering, Pune for technical guidance, computing resources and research support. The authors would also like to acknowledge the open source mental health, speech emotion recognition and computational psychiatry communities whose datasets and ethical frameworks enabled this work.

REFERENCES

- [1] M. A. Author et al., "The application of machine learning techniques in posttraumatic stress disorder: A systematic review and meta-analysis," *J. Affect. Disord.*, vol. 345, pp. 112–128, 2024.
- [2] R. K. Sharma and S. Patel, "Systematic review of machine learning in PTSD studies for automated diagnosis evaluation," *IEEE Access*, vol. 11, pp. 45120–45136, 2023.
- [3] H. Lin, J. Chen, and K. Davis, "Use of machine learning algorithms based on text, audio, and video data in the prediction of anxiety and posttraumatic

- stress in general and clinical populations: A systematic review," *Psychiatry Res.*, vol. 331, p. 115622, 2024.
- [4] E. O'Reilly, D. Smith, and A. Miller, "Speech-based recognition and estimating severity of PTSD using machine learning," *Comput. Speech Lang.*, vol. 84, p. 101567, 2024.
- [5] G. Stratou, S. Scherer, J. Gratch, and L. P. Morency, "Using speech analysis in virtual agent conversations to differentiate PTSD patients from control participants," *Front. Psychiatry*, vol. 16, p. 138902, 2025.
- [6] A. Baird, S. Amiriparian, and B. Schuller, "Insights on modelling physiological, appraisal, and affective indicators of stress using audio features," in *Proc. IEEE EMBC*, 2022, pp. 3120–3124.
- [7] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: Validity of a brief depression severity measure," *J. Gen. Intern. Med.*, vol. 16, no. 9, pp. 606–613, 2001.
- [8] S. Latif, R. Rana, S. Khalifa, R. Jurdak, and B. Schuller, "Speech emotion recognition approaches: A systematic review," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 280–301, 2023.
- [9] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, pp. 56–76, 2020.
- [10] V. H. Nguyen, L. Kerkeni, and C. Frelicot, "Measuring negative emotions and stress through acoustic correlates in speech: A systematic review," *Comput. Speech Lang.*, vol. 89, p. 101684, 2025.
- [11] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.
- [12] P. Shen, Z. Chang, and F. Wang, "Speech databases, speech features, and classifiers in speech emotion recognition: A review," *IEEE Access*, vol. 12, pp. 18230–18249, 2024.
- [13] R. C. Suganthi and M. S. P. Babu, "Review of automatic estimation of emotions in speech," *Artif. Intell. Rev.*, vol. 58, no. 2, pp. 1–45, 2025.
- [14] F. Al-Haj, N. Al-Makhadmeh, and A. B. Nassif, "Voice-based stress detection: A systematic review of machine- and deep-learning approaches (2021–2025)," *Expert Syst. Appl.*, vol. 260, p. 125430, 2025.
- [15] J. P. Pestian, B. S. Connolly, and M. Szekely, "Application of natural language processing (NLP) in detecting and preventing suicide ideation: A systematic review," *Lancet Psychiatry*, vol. 10, no. 6, pp. 445–458, 2023.
- [16] M. Morales, S. Scherer, and R. Levitan, "Detection of suicidal ideation in clinical interviews for depression using natural language processing and machine learning: Cross-sectional study," *JMIR Ment. Health*, vol. 10, p. e42318, 2023.
- [17] Y. Zhang, H. Wang, and T. Darrell, "A multimodal approach for detection and assessment of depression using text, audio and video," *IEEE Trans. Multimed.*, vol. 26, pp. 3410–3422, 2024.
- [18] Z. Chen, L. Wang, and X. He, "Depression detection methods based on multimodal fusion of voice and text," *Inf. Fusion*, vol. 115, p. 102711, 2025.
- [19] D. L. Hall, M. A. Newman, and K. Breslau, "Detecting PTSD in clinical interviews: A comparative analysis of NLP methods and large language models," *J. Clin. Psychol. Med. Settings*, vol. 33, no. 1, pp. 88–104, 2026.
- [20] O. Melnyk, V. Vovk, and Y. Bilous, "Cross-lingual transformer-based screening of post-traumatic stress disorder based on comparative analysis of BERT and XLM-RoBERTa with machine translation adaptation for Ukrainian language," *Comput. Linguist. Intell. Syst.*, vol. 18, pp. 142–159, 2026.
- [21] S. D. Joshi, P. R. Deshmukh, and S. V. Kadam, "A hybrid deep learning framework for Marathi speech-based stress detection using GRU and handcrafted audio features," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 25, no. 1, pp. 1–19, 2026.
- [22] S. T. Vayena and E. Beauchamp, "Keeping an AI on the mental health of vulnerable populations: Reflections on the potential for participatory injustice," *Bioethics*, vol. 39, no. 3, pp. 240–252, 2025.
- [23] A. G. Smith, C. R. Martinez, and J. K. Tandon, "Ethical decision-making for AI in mental health: The Integrated Ethical Approach for Computational Psychiatry (IEACP) framework," *Lancet Digit. Health*, vol. 7, no. 2, pp. e110–e121, 2025.