iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76085

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Acoustic and Temporal Computational Approaches for Early Identification of Psychological Disorders: A Human-Centric Survey

Vijay Kumar S, Nidhi R.Kashyap, Raveena N., Chandana N.V.
Department of Information Science and Engineering Global Academy of Technology, VTU Bengaluru, India

Abstract: As the world is growing faster, psychological disor- ders are aslo increasing and affecting people of all ages, cultures and financial standards. Since more people face these struggles, earlysigndetectionmatterstoensureanearlyinterventionbefore the problem worsens. Traditional diagnosis usually depends on personal judgment, talking with doctors or patients describing their feelings but that does not work well in cases when one cannot speak up, if a person is too young or is losing memorywithageorfindsithardtocommunicate.Somehidehowtheyfeel out of fear of being judged or even due to the misconception and pressure by others surrounding them. These days, advances in technologies like AI, audio processing, temporal pattern analysis andbehaviormodelingmakeearlydiagnosismoreandmoreeasy by passively observing daily activities of the people. This paper discusses the capability of computers in identifying two types of natural data; voice sounds produced while speaking and words written or shared online across social networks. Voice signs such as in+
-9stability of pitch, irregular pacing, tension in the voice and pausingareamongothermeasurablefeaturesofspeechthat can be indicative of physical or mental states if one doesn't say anything outright. Meanwhile changes in the pattern of writing online, the mood conveyed in these posts, word choices, context clues and manner of expression over days reveal mental health tendencies. A key focus here is timing. Mental health decline usuallydoesn'thitquickly,thesignstendtoshiftslowly,showing up over time instead of in a single moments. Combining voice tech with written behavior clues shows promise in noticing issues sooner without invading their private space or cutting access. This review brings together basic studies of new deep learning ideas, ways to combine different data types and also making models easier to understand while pointing out open issues like smalllabeleddatasets,moralquestions,problemsapplyingresults broadly, plus hurdles in real world medical use. Its objective isto give a clear picture of today's methods, show how sound and analyzing it over time could help spot mental health needs and help build smarter tools that adapt well across different people, include diverse groups, act ethically.

Index Terms: Acousticbiomarkers,Mentalhealthinformatics, Deepneuralarchitectures,Temporalpatternanalysis,Behavioral signal processing, Early risk prediction, Computational psychol- ogy, Speech-based assessment

I. INTRODUCTION

Psychological disorder like depression, anxiety, bipolar problems, memory problems and adhd are now a big concern across the globe, hitting countless lives. These struggles don't just harm individual feelings, they ripple outward, dragging downtheirproductivity,longtermhealthissues,straining familyandincreasesdisabilityrates.Morepeoplenowseehow tough mental health battles can be and that is why spotting signs early matters. This quick helps manage symptoms,boost recovery, prevent chronic progression. Even with the importance of early spotting, traditional diagnosis methods have their drawbacks. Most tests depend on people's self un- derstanding, talking about their feelings and their accessibility to skilled therapists. Some cannot put emotions into words, somedonotnoticetheirstruggleandsomehideitbecause of shame or due to fear of judgment or are uninformed. Key warning signs go unnoticed, especially among older adults, teens, people with different neurodevelopment and people in self isolation Besides these issues, new technology has altered thewaywedetecthumanbehaviorclues;everydaytalkingand typing hide valuable signals about our thoughts, feelings and mentalstates.Currently,deeplearningsystemsdetecteventhe smallest details in voice like pauses, pitch changes, stress and uneven sentences that often represent anxiety, depression or someotherpsychologicalconditions.Inrecenttimesaspeople are using social media more, it has become like a windowinto people's personal life and lets us track and analyse their posting patterns and behavior, tone of their sentences and topics they share and are more interested in. Some researches andstudieslikethee Riskchallengeshaveshownthatlinguistic analysiscanpredicttheriskofdisorderslikedepression,eating disorders, self harm and more well before getting a clinical diagnosis.

Temporal modeling connectes these fields in a very simple way. Mood changes occur with time, so systems that tracksuchtrendsprovidemorevividclues.Modelssensitiveto time detect speech changes by tracking the decline of voice, changes in words used or repetition of emotional cues this makespredictionoffutureproblemspossible,predictionhelps ustoactfasterthandetectingtheproblemonceithasoccurred. This review gives a better perspective by diving deeper into this topic about how technical tools help determine mental health by comparing different methods of detection against one another and showing new areas for study and demon- strates where acoustic and speech analysis meet. This review combines speech-based and text-based detection to show how AI can help identify mental health needs by highlighting tools that are available, non biased, flexible and human centered.Bycombiningtheseapproachesitbecomeseasiertopointout ways technology can support diagnosis of a disorder without replacingthehealthcareprofessionals.Eachpathoffersunique clues, whether from tone or words helping systems respond better over time.

II. PROPOSED METHODOLOGY

The method described here grows from the emotional, cognitive and physiological indicators computers can pick up to help spot mental health issues sooner. This system called “Acoustic AI” is not meant to take over from health care professionals instead, it works quietly alongside them, fitting intodailylifewithoutpushingboundaries.

Whiletypingbased toolsneedpeopletowritethingsout,usingsoundletsuslisten in gently and is ideal for those who say little, talk only now andthenorstruggletoputemotionsintowords.Intheanalysis of the sound, there are four main steps:

- Gatheringandpreparingaudio
- Extractingtheaudiodata
- Developingtemporalmodels
- Classifyingsignals

Breakingdownofthesefourmainprocedureofsoundanalysis into individual tasks we get a raw audio file as a result thatcan be converted into useful information that can be used to predict disorders that will occur in the future.

A. Data Acquisition and Processing

The first step in this method is collecting the audio of the speech that happens naturally and in daily life, it should notbe scripted or staged. These sounds can be collected from normal conversations, calls between caregivers and patient, devices used to help people communicate better, therapy conversations or short bits of conversation captured in daily life. This makes it unique, as it works well even when the peoplegiveonlyshort,scatteredbitsofspeech.Thisflexibility is very important while dealing with toddlers, older adultswho are dealing with memory loss, people with mental health issues and anyone who is holding themselves back to speak duetonervousness.Thesystemusescertainbuiltinsteps to keep the infomation collected private and safe, instead of storing personal details like names, background noise thatmay reveal where they live or what has been said in the conversation,thesethingsareerasedorhiddenbeforethenext step starts. The audio collected is then undergoes processed, it is cleaned, smoothened, divied into smaller parts, gadgets cut downreverb,gadgetquirks,andbackgrounddisturbances.Now tooputeverythinginorderwecutoffquietgaps,boostcertain tones,removehums,sharpentheclarityandshiftsamplerates, thesestepshelpnomatterwhichsourcefileweareusing.This full process turns raw messy audio recording’s into neat input for machines to work with while still protecting who spoke and what they said.

B. Acoustic Feature Extraction

After the audio recordings are processed like we discussed in the earlier section, they are evaluated to see if they show signsofstress,mentalpressureoranypsychologicaldisorders.

These signs are detected from the the audio based on the tim- ing, pitch change pattern, frequency and little vocal changes. Insteadofanalyzingthetypicalmeaningofthewordsused by people, this method analyzes how the person speaks, this showsmajorityofthecclueswhichhelpevenifthepersontalks very less or if their speech is unclear.

- Pitch variation: shows how the volume of the voice changes throughout the recording. If there is too much change in the pitch it indicates stress and anxiety.
- Speechpacing:talksabouttimeofspeech,howfastwords are said and the flow of speech. Slow talking can be linked to feeling sad. Fast speech can be a sign of worry or memory problems.
- Pausebehavior:tellshowoftensomeoneisquiet,thetimeofsuddenstopandhowoftenothersarehittingthetimed.Thiscanoftenbefeltalongwithhavingtothinktoomuch or avoiding feeling.

- Jitter:measuresthepepperofshorttermchangesinsound and can often go up when you have trouble moving and controlling vocal muscles or experiencing high levels of bliss.
- Shimmer: judges the change in sound when talking, itcab often go up when breathing is strained, tired or there is a lack of shine in the voice.

To make stronger robustness, you can use two or more low level descriptors and statistical summaries. Some of thesetoolsareMel-frequencycepstralcoefficients(MFCCs),Teager energyoperators,vibratingnoisesandglottalflowsounds.The above suggests how to match one's acts and mental state.

C. Temporal Modeling and Behavioral Representation

A major step forward in this approach is using temporal voice profiling, this means watching how sound traits change over several clips instead of just one snapshot. Mood shifts don't usually hit overnight; small differences build up slowly, day by day or talk after talk. When we analyze data from different angles by comparing many parts of the audio over time we can see the fading of excitement or interest, contin- uous interruptions or breaks in the flow, there will be a slow andgraduallossofenergyorstrengthintoneandtiredness in voice increases. This method of collecting audio data helps us capture what is usually not noticed by traditional methods, these traditional methods rely on only one audio recording.

The audio is recorded section by section using the sliding window structure. By using recurrent networks, the voice attributes of each person can be synchronized across mul-tiple intervals. With this method, dynamic time adjusts the speech sequences so that rhythm deviations do not affect the sequences. By means of attention mechanisms, transformers identify and correlate connections between different audio parts.Thesemethodswhencombinedtogetherlet'susmonitor changes in speech patterns. Therefore, any unusual changes in how our speaking style will become evident and the detection of the mental disorder can be done much sooner rather than waitinguntilithasaffectedthepersoncompletely.Asaresult of this model rely less on generalised group trends and make more personalised calculations for each individual.

D. Classification and Evaluation

The final stage uses machine learning models that are capable of finding difference between voice patterns and the patternsrepresentingpsychologicalconditionslikedepression, anxietyandmanymore.Deepneuralnetworks,includingcon- volutional network architecture, recurrent models and trans- former based encoders can be used to detect early signs and symptoms. The classification pipeline is used to label risksandalsotocalculatetheseverityprogression,confidencelevels andinterpretsignsthatmedicalprofessionalsorcaregiverscan understand. Real world metrics that changes over time, cross speaker robustness testing, noise tolerance and longitudinal validation are used for calculations and evaluating a person's mental healths condition. The goal here is to create a scalable toolthatisopentonecessaryimprovementsinthefuture andthatcanbeusedwidelybyeveryone,thistoolisbuilt tohelpexpertsandnotreplacethem.Byspottingthesigns of psychological issues early it makes it easier to seek help quicker, to monitor the issue better and to get the right and necessary care that is required.

E. Distinctive Characteristics

There are some important features that makes this method special and unique such as:

- Itstartsbyusingspeechanddoesnotdependonvoice or text.
- Itworkswellforpeoplewhospeaklessandinuneven or small bits
- It focuses on how a person changes over the time and not suddenly or in an instance.
- Ithasprivacyprotectorstokeepallthepersonalinfor- mation safe.
- Itusesmodelsthataredesignedforspecificsignals insteadofrelyingononegeneralclassifierforeverything.
- It handles sensitive mental health and personal informa- tion safely and ethically.

Thesetraitsmakethemodelanewformofworkinthe fast growing field of computer based mental health analytics, providing a route to use when detecting early signs of mental illness in a way that is fair to all, non-invasive and easy to access.

F. Classification and Evaluation

A deep learning based classification pipeline is employedto differentiate between typical and disorder linked vocal patterns. The system aims to detect early signs of depression, anxiety and cognitive decline. The broader goal is to supply clinicians and caregivers with supplemental insight to support timely intervention.

G. Distinctive Characteristics

Theframeworkisdifferentiatedby:

- Exclusiveemphasisonspeechasaprimarydiagnostics.
- Suitabilityforsemi-verbalandnon-verbalgrouping.
- Use of temporal vocal evolution instead of isolated sam- ples.
- Privacysensitivesincenoethicallyquestionableproce- dures are deployed.
- Modalitytailoreddeeplearningintegration

III. NOVELTY AND CONTRIBUTIONS

The work presented within this survey offers a distinct contribution to the growing field of computational mental health analysis by bringing together two separate research areas;acousticbasedpsychologicalusuallytreatedinisolation, detection and linguistic based early risk identification. Many studies have examined each modality on its own, few studies have also examined how these two have similar concepts.This survey shows that time based models are like a bridge between speech signals and digitally written clues, it tells that psychological assessments are more meaningful and helpful when they are tracked over time instead of refering an single moment. This let's the researchers and medical professionals understand how these two areas are connected and how combining them will help in improving early detection of psychological disorders better.

A. Cross-Domain Integration

A new perspective here is of comparing speech analysis tools with clues obtained from digital writings. Most of the past summaries stick to just one area for example ,how word meanings shift over time or odd voice patterns. Putting both these fields next to each helps us see how time based models can track mood swings in text and how the meaning shifts, these tools can also notice voice traits like a steady tone, pauses inbetween or a flat rhythm. Mixing these perspectives opens up new possibilities for helping people who are mental health professionals and technology designers.

B. Temporal Modeling as a Central Construct

This study focuses on how behaviour changes over time since itis important for detecting risks earlier through speech or text, mental health disorder's signs do not appear suddenly, they build up slowly and show up as small shifts and that too step by step. Tools like RNNs or comparing the embeddings alongtimelinesorfollowingvoicetrendshelpintrackingsigns and patterns across a time period, using these tools is not just helpful, it is necessary to get correct prediction. It combines methods from language studies that help track word meanings with similar techniques in voice analysis. Understanding howa person's tone changes over many sessions motivates the expertstouseideasfromonefieldinanotherfield.Lookingat things in this way moves the systems used for early detection ofmentaldisordersawayfromquickjudgementsandtrains it to focus more towards watching how people behave over months or years of time.

C. Broad Survey of Contemporary Research

This survey gives a wide range of research from 2023 to 2025, this was a key period for AI and mental health studiesas there was rapid improvement in methods that use speech, written data and environmental clues to help understand people's mental health better. It has used special tools like transformer structures, graph neural networks and hypergraph autoencoding. Also, the review shows how there were small, simple ways now possible to work on real problems. These small ways are fitting to use in a real world setting and on handheld computers and other medical tools. The review also talks about the merging of big box wide stories and decision making trails with mental health prediction models. For both, their power and how well they explain themselves. By putting these changes in technology together, the review let us see the current depth of the research, the quality of the research, and how useful it is in real life.

D. Challenges and Future Opportunities

Besides showing growth, the survey finds common issues thatnowslowdownworkincomputationalmentalhealthtests. Theseissuesaresmallgroupofdata,troubleofgettinggeneric results for people, high work times as models for long times, moral issues about people's privacy and permission and high explainabilitywhenusingdeepneuralmodelsinrealhospitals.

Knowingtheselimitsindepth,thesurveyoffersawrittenplan for work in the future, one that puts time for understanding models,smallmodels,manytypesofdata,workwithdifferent cultures and people and safe ways of use. These ideas are to leadintobothpeoplewhostudybehaviorandthosewhomake tech to have better and real ways to help, measure and serve patients.

E. Summary of Novel Contributions

Themainpartsofthisreviewareasfollows:

- It brings together the two main ways researchers try to detect mental-health issues through a person's voice and speech.
- It shows that paying attention to how someone's speech changes over time is one of the most important steps for spotting early signs of problems.
- Itcombinesmanymodernresearchtrendslikecombining information from different types of speech and buildsafe and secure models by using powerful deep learning systems.
- It provides a model with practical approach to deal with challenges that will come, these models can understand the data better and tools that more people can use and technologies that can be used securely and responsibly.

Altogether, this review becomes a meaningful guide for any- one trying to create future tools that can detect mental health concerns in a realistic, inclusive and secure way.

IV. LITERATURE REVIEW

Workinsmartwaysofcheckingmentalhealthusingspeech hasgrownupfastinthelastfewyears,fedbysmartlearning, watching humans and their actions, computers that know very well how we act and feel, and models of the brain's signals. Thispartgivesasetwayoflookingatthemainpartsof work, showing how we began working with models that look atsoundandspeechtofindmentalhealthissues.Itthenshows how recent work has combined different ways of learning using several types of data, how behaviors over time can be used, how to create models that work well with less work and how to understand models. By grouping the work by subject rather than in order in time, this review shows how each way ofworkinghelpswiththegreatergoaloffindingmentalhealth issues through the use of smart systems.

A. FoundationalSpeechEmotionRecognition(SER)

Work in this field began to lay the foundation for how people read the state of mind and feelings from speech. These studies mainly deals with recognizing emotional labels suchassadness,anger,fearorneutralbyacousticfeatureextraction and distribution models. In spite of repeated problems, it resultedastohowsuchmodelsconfirmedtodifferentspeakers, cultures, groups, recording situations and cultures. Standard testsrepeatedlyconfirmedSERmodelsmorebyfamiliar core data than on foreign input, there lies a weakness to unequal data sets and variations of groups. By exploring these challengingissues,researchersdiscoverednewpossibilitiesfor usingself-supervisedlearning.ModelssuchasWav2Vec2can learn the patterns of speech directly from large amounts of unlabeled audio because of this, they adapt well to many languagesandaccentsevenwithoutbeingtrainedonemotional datasets.Inaway,themodelpicksuptherhythmandstructure of speech much like understanding communication through patterns rather than explicit emotional signals allowing it to perform strongly across diverse speaking styles. However, ac- curacy is often decreased in underrepresented languages, This highlights that more different emotional speeches are needed. Other research investigated behavioral acoustic perception in individuals who reported depressive symptoms by examining, howspeechisinterpretedratherthanproduced.Althoughthese findings are based on subjective reporting, provided valuable insight into information that would not have been picked upby other neuropsychological measures.

B. Advanced Deep Learning and Multimodal Fusion

As studies grew up, limits of single way speech review madewayformoredeeplearningsystemsandgroupsthat put many ways together. These systems put together sound features with speech content, body signals, extra information or social behaviors for better guesses. Systems based on transformers like the mental perceiver showed good work in judging mental health related states, by reading audio and speechatthesametimebutthesegroupsneedalotof labeleddataandit'sabigchallengewhenpeoplewantto use them. Graph Neural Networks (GNNs) then came along and made it easier to put prosodic features together with data that has to do with relations or contexts. At the same time, hypergraphautoencodermodelswithcontrastivelearningdid better at combing data from many sources but took more todo.TheMEDUSAsystemhadmanylevelsofdeepfusion. It was made to find pure emotions in real-world settings, and there is hope that it could be used to help larger crowds.

In recent times, new research has started, with the use of Large LanguageModels(LLMs),withretrieval-augmentedreasoning to find depressive states, with help from speech timing cues and context. There have been some proposals for multitasking LLMs that can find more than one mental disorder at a time. Althoughsomeworryabouttheirtransparency,ambiguitywith inferences and trustworthiness in a clinical setting together all of these multimodals advances seem to put more focus on representing a combined pictures of mental conditions.

C. EfficientLow-ResourceModels

Temporal analysis has emerged as one of the most influen- tialadvancesinthedetectionofpsychologicaldisorders.Foun- dational studies in linguistic research showed that tracking se- mantic changes in texts over time by using methods like Tem- poral Word Embeddings (TWEC) and Dynamic Contextual Word Embeddings (DCWE) accurately predicted early signs of disorders such as depression, anxiety and self-harm. These findings supported the importance of observing behavioral evolution rather than isolated expressive samples. However, such models required substantial memory resources becauseof their attention based processing structure. Context aware deeplearningmodelsfurtherenhancedpredictivestrength by incorporating situational cues, conversational flow and interactional patterns into classification frameworks. Despite increased complexity, temporal and context aware modelingis very important when one wants to effect change in mental health, from static diagnosis towards dynamic early stage diagnosis.

D. TemporalandContext-AwareModeling

Time analysis has become one of the most important advances in finding out about mental health problems. Basic researchinlinguisticanalysisshowedthatbyfollowingmean- ingful changes in text over time with ways such as Temporal Word Embeddings (TWEC) and Dynamic Contextual Word Embeddings (DCWE), it was possible to carefully forecast early signs of illnesses such as depression, anxiety and self destructive behaviour. This considers behavioral changes not just moments of expression. Transitions in speech analytics madeadvancesthatinvolvedtransformer-basedmodelsbeable to study speech for hours to determine how much depression there was based on longer durations of audio produced. However, these models ended up needing a large amount of memory because their attention mechanisms had to process and keep track of large audio clips. Context aware deep learning models further brought higher accuracy to prediction byincludingproximitycues,speechflowandtheinteractional pattern in classification. These temporal and context-aware modelsarecomplex,theyhaveplayedaimportantroleinshift- ingmental-healthpredictionfromfixed,one-timeassessments to early and responsive diagnosis.

E. Model Explainability

As psychological assessment involves sensitive personal implications, the comprehensibility of computational models hasbecomeacriticaltopicclinicians,caregiversandregulatory bodies require transparency to understand why a model pro- duces a specific inference. Deep learning models particularly use transformer-based and multi-modal architectures, these often lack the ability to explain their internal reasoning, creating challenges in clinical trust, ethical validation and safety requirements.

Recent studies have suggested methods such as post hoc interpretive mapping, which translates internal speech embed- dings into understandable acoustic descriptors. Probabilistic modellingmethodslikeUniversalHiddenMarkovModelsgive clear reasoning but at the cost of the models performance. Attentionbasedmodelsthatcanshowthemostemotionalparts ofarecordinghelps usgetanotherwaytounderstandwhatthe system is focusing on. This shows how explainable AI(XAI) has helped mental health tools be more transparent, reliable and suitable for clinical use.

V. CONCLUSION

This project gives us a new way of understanding how computers can detect the early signs of mental illness from parameters such as speech and language. It looks at how sensors,behaviouralmodels,multimodallearningthattakesin many types of signals at once and machine learning systems help to improve this field. We use many different kinds of signals and time based methods to predict mental health more accurately, instead of only using emotions. This discussion tells us that mood never just jumps and that it takes a slowpath before it becomes easy to see the warning signs. It shows that speech has special benefits as a way to find information that is not painful to get and often occurs by itself. Speech biomarkerslikepitchvariation,pausetimeandtheenergy of the voice can provide clues about emotional pressures or thinkingoverloadswithoutsayingmuch;wecanlookatonline habits and what people are saying to see how they are feeling. When adding a time element to both of these modalities, This makes it much easier to catch signs of abnormality early on and helping identify psychological risks before they grow into moreseriousproblems.

Moreover,itshowsthattheincreasing abilities of machines that can look at mental health are both moreusefulbutalsoraiseopportunities.Ingeneral,deeplearn- ing architectures, multimodal fusion models, and transformer- based systems make it possible to draw conclusions at a much higher level but many of these raise problems that come from being able to give answers about what they mean, knowledge privacy,databias,inclusivityofallculturalgroupsandethical safety. The path ahead needs to follow research that combines innovation with making sure people's concerns are prioritized in the design, so that these tools support mindfulness in people, and do not replace their wisdom, dignity or their wellbeing.Afewgoodideasexistforthefutureenhancements, these include creating simpler lightweight models that can be deployedonmobilesandsupportivedevices,expandingand creating systems that can clearly explain how they reached their conclusions, along with multimodal models that bring speech, text, body signals and even environmental cues to- gether. For helping psychologists, data scientists, clinicians and technology designers work side by side; we move closerto building mental health analysis tools that are easy to use, flexible, transparent and truly empathetic in their function. In the long run, improvements in acoustic and temporal analysis could help us detect problems soon, ease people's distress and shape a mental health ecosystem that feels more responsive, supportive and compassionate.

VI. ACKNOWLEDGMENT

TheauthorsextendtheirdeepappreciationtoProf.Vijayku- mar S, Assistant Professor in the Department of Information Science and Engineering, whose guidance, thoughtful direc- tion and steady encouragement played an important role in shaping the development of this survey and his expertise in this field provided clarity during the project and his feed- backs helped us forge a connection between acoustic analysis, temporal modeling and early stage psychological assessment. The authors acknowledge Global Academy of Technology for providing academic support, research access and institutional resources that helped in the successful completion of this work. The environment fostered curiosity, experimentation and interdisciplinary exploration, contributing to the progres- sion of the ideas presented. The authors also extend their gratitude to their peers, classmates and research colleagues whosediscussionsandsharedproblemsolvingcontributed to deepening the conceptual framing of the paper. Their perspectives helped broaden the scope of analysis and kindled more holistic thinking about mental health technology and societalneed.Finally,thisworkismotivatedbytherecognition that many individuals,especially children, elderly populations and those with limited verbal expression remain underserved bytraditionaldiagnosticmethods.Theauthorsareinspired by the possibility of developing computational tools that can support the people in need of help in ways that are respectful and accessible. It is wished that the research in this survey contributestotheimprovmentofearlydetectionapproachesin detecting psychological disorders while emphasizing on areas such as human dignity, emotional well-being and usage of ethical technology.

REFERENCES

[1] M.Coutoetal.,"TemporalWordEmbeddingsforEarlyDe-tectionofPsychologicalDisordersonSocialMedia,"Journalof Healthcare Informatics Research, 2025. [Online]. Available:https://doi.org/10.1007/s41666-025-00186-9

[2] J. Qin et al., "Mental-Perceiver: Audio-Textual Multi-Modal Learningfor Estimating Mental Disorders," arXiv preprint arXiv:2408.12088,2024. [Online]. Available: https://arxiv.org/abs/2408.12088

[3] D.Kounadis-Bastianetal.,"Wav2Small:DistillingWav2Vec2to72KParametersforLow-ResourceSpeechEmotionRecogni-tion," arXiv preprint arXiv:2408.13920, 2024. [Online]. Available:https://arxiv.org/abs/2408.13920

[4] X. Zhang et al., "SpeechT-RAG: Reliable Depression Detection inLLMs with Retrieval-Augmented Generation Using Speech Timing In-formation,"arXivpreprintarXiv:2502.10950,2025.[Online].Available:https://arxiv.org/abs/2502.10950

[5] A. I. S. Ferreira et al., "Enhancing SER with Graph-Based MultimodalFusion and Prosodic Features," arXiv preprint arXiv:2506.02088, 2025.[Online]. Available: https://arxiv.org/abs/2506.02088

[6] J.Qinetal.,"Context-AwareDeepLearningforMulti-ModalDepressionDetection," in Proceedings of Conference, 2024.

[7] Z. Huang et al., "Efficient Long Speech Sequence Modelling for Time-Domain Depression Level Estimation," in Proceedings of Conference,2025.

[8] Y.Wangetal.,"IdentificationofDepressionStateBasedonMulti-ScaleAcoustic Features," in Proceedings of Conference, 2023.

[9] S. Pal et al., "Interpretable Probabilistic Identification of Depression inSpeech," in Proceedings of Conference, 2025.

[10] M. Atmaja et al., "Explaining Deep Learning Embeddings for SpeechEmotion Recognition by Predicting Interpretable Acoustic Features," inProceedings of Conference, 2024.

[11] S.Latifetal.,"Cross-LingualSpeechEmotionRecognition:Humansvs.Self-Supervised Models," in Proceedings of Conference, 2024 (revised2025).

[12] K. Z. Li et al., "Decoding Emotion: Speech Perception Patterns in Indi-viduals with Self-reported Depression," in Proceedings of Conference,2024.

[13] T. T. Tran et al., "MEDUSA: A Multimodal Deep Fusion Multi-StageTraining Framework for Speech Emotion Recognition in NaturalisticConditions," in Proceedings of Conference, 2025.

[14] Y. Wang et al., "Integration of Text and Graph-based Features forDetecting Mental Health Disorders from Voice," in Proceedings ofConference, 2024.