MediMind: An Intelligent Mental Health Monitoring System Using Speech and Text Analytics

Dhairya Korgaonkar[1], Maaz Khan[2], Aryan Shedge[3], Riya Ankush[4], Pranavkumar Badhane[5]

[1]Dhairya Korgaonkar, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic [2]Maaz Khan, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic [3]Aryan Shedge, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic [4]Riya Ankush, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic

[5]Pranavkumar Badhane, Artificial Intelligence and Machine Learning, Fr. Agnel Polytechnic

Abstract - A quiet shift often happens before stress becomes overwhelming. Daily pressures build gradually, reflected in subtle changes in speech, writing, and behavior. Advances in Artificial Intelligence now allow these small signals to be understood through tools designed for Mental Health Monitoring and Preventive Mental Healthcare.

Using Speech Emotion Recognition, systems analyze tone, pitch, pauses, and rhythm to detect emotional variation. At the same time, Natural Language Processing and Sentiment Analysis examine typed words, sentence structure, and vocabulary patterns to identify shifts in mood. Together, these technologies support Mood Prediction and Stress Detection by spotting trends over time rather than reacting only during crises.

Through Machine Learning, models learn individual behavioral patterns, making predictions more personalized and accurate. What appears as normal conversation becomes structured data that reveals emotional trajectories. Integrated into a Mobile Healthcare Application, this process runs quietly in the background, offering gentle nudges, insights, or recommendations when early warning signs appear.

This approach does not replace human care; instead, it strengthens awareness. By identifying emotional changes early, AI-driven systems encourage timely reflection, healthier coping strategies, and proactive support— helping individuals respond before stress escalates into burnout or breakdown.

Key Words: Artificial Intelligence (AI), Mental Health Monitoring, Mood Prediction, Stress Detection, Speech Emotion Recognition (SER), Natural Language Processing (NLP), Sentiment Analysis, Machine Learning, Deep Learning, Behavioral Pattern Analysis, Voice Signal Processing, Emotion Classification, Mobile Healthcare Application, Digital Mental Health, Preventive Mental Healthcare, Real-time Monitoring, Predictive Analytics, Human-Computer Interaction, Psychological Well-being, AI-based Health Assistant.

1. INTRODUCTION

A person's mind matters just as much as their body when it comes to how they feel each day. Lately, more people struggle with worry, constant tension, or deep sadness - pressures from school, long hours at jobs, shifting daily routines, or feeling cut off from others are part of why. Even though these issues happen more now than before, plenty stay quiet instead of reaching out to experts who could assist them; shame, not knowing what's wrong, or clinics being too far away keep them back. When signs show up but get ignored, things can spiral into deeper pain that reshapes thoughts and emotions over time.

Nowhere else has tech moved faster than in how phones track feelings. Phones grab sound and words, layering them into streams machines learn to read. Out of rhythm in talking, shifts in phrasing - clues hide there. Machines notice what people miss when stress builds or moods dip. Tone bends under pressure, word choice tightens or drifts.

Patterns form over days, quiet signals piling up unseen. From typing speed to pauses between replies, details add up. These tools catch slips before they spiral too far. Not magic, just math tuned to human hiccups. Quiet changes in speech often show first when something feels off. Text trails leave traces deeper than most realize. Listening closely means spotting dips early. No flash, no drama - just steady observation

from devices already at hand. Emotions shape expression in ways code now decodes. Every message sent carries hidden markers beneath the surface.

Sometimes quiet moments speak volumes. A new tool listens closely when people talk, picking up subtle shifts in tone. Instead of waiting for symptoms to grow worse, it watches daily patterns across words typed or spoken aloud. Hidden cues emerge through careful study of rhythm, pace, and phrasing. Feelings like sadness or tension leave traces even when unspoken directly. This approach skips uncomfortable tests that feel intrusive. Information flows naturally - no extra effort needed from users. Behind the scenes, smart models

2. LITERATURE REVIEW

New progress in smart computer systems together with tools for phone-based health care has pushed more study into checking mental well-being. Instead of old ways, experts usually depend on face-to-face talks, forms filled by patients themselves, or scheduled visits with specialists. These older techniques work - yet they take too long, can differ between people, plus many cannot reach them because of shame around illness, cost issues, or missing clinics nearby.

Looking into feelings through written words has drawn attention in multiple research efforts. From tweets to survey answers, computer methods break down language by spotting key parts, rating mood tones, or sorting emotions into groups. While some tools show potential when catching signs of tension, worry, or low moods, their success often depends on how much people actually write - and whether they're open about it. What comes out relies heavily on what goes in.

Lately, spotting feelings through voice has drawn interest in mental wellness studies. Tone, pitch, how fast someone talks - these bits hide clues about inner states. From sound clips, machines now pull out patterns using smart math tricks to guess emotions. Though better at guessing moods, plenty of current tools need quiet rooms and fancy gear, making daily use tough. Some work well only where noise stays low and microphones stay top-notch.

Watching moods through both voice and writing works better than using just one. Pulling together different kinds of input lets the technology notice more signs of how someone feels. That leads to steadier guesses about sadness or pressure. Still, many tools today stay stuck in labs. They rarely turn into handy phone apps people can actually use every day.

One thing after another, mobile apps for health now help people manage mental well-being - simple tools that fit into daily life. Through quiet prompts or gentle reminders, they might track moods, lead breathing exercises, or reply with automated comfort when someone feels low. Even so, while lots choose them, most stay basic - no smart learning behind the scenes, just static replies without foresight. Hidden beneath smooth screens, few can spot warning signs before things worsen.

One look at past studies shows something's missing - a smart phone tool that watches mental well-being using both speech and written words. Instead of waiting, MediMind fills the space by quietly tracking emotional shifts through everyday communication. It checks for signs of strain, worry, or shifting moods without asking much effort. Running in the background, it keeps check- ins smooth, unobtrusive, always available. What stands out? The whole process feels natural, built into life, not forced.

3. MATERIALS AND METHODS

Here's how we built MediMind, an AI tool that tracks mental well-being by analyzing speech and written words gathered from a phone app. Voice clips and typed entries feed into models designed to spot shifts in emotion or rising tension over time. Instead of guesswork, patterns emerge from real-world inputs processed with machine learning methods tailored for personal signals. Software components work together behind the scenes - some handle raw input, others refine it before interpretation kicks in. Data comes from everyday users opting in during routine check-ins, offering glimpses into natural emotional rhythms. Algorithms learn gradually, adjusting sensitivity based on repeated exposure to individual habits and expressions. No outside lab setup needed; everything runs within reach, shaped around accessibility and steady observation. Privacy stays central throughout each phase, guiding choices about storage, access, and processing rules. Results aim not at labels but subtle cues

- an uptick here, a dip there - mapped across days or weeks. Design leans on simplicity so attention stays

where it matters: changes hidden beneath surface-level replies.

3. Materials and Methods

3.1 Materials Used

Hardware plus software tools make up what MediMind needs to gather, handle, and study information properly. Sensors capture signals from users while programs run prediction routines behind the scenes. Data moves
through encrypted channels before resting in protected databases. Each piece fits into a larger setup meant to track mental states over time.

3.1.1 Hardware Requirements

Running mostly through a phone, it uses the device's mic to pick up speech for reading emotions. Because data must move safely back and forth, a steady web connection keeps the app linked to its remote base. That base lives in the cloud - handling storage while running smart algorithms without delay. Scaling smoothly comes from processing decisions off-site, where power adjusts on demand.

3.1.2 Software Requirements

Built on tools like Android Studio or React Native, the app offers an easy-to-use experience. Instead of one single method, backend work relies on Flask with Python or Node.js in JavaScript to manage how parts talk and process info. Stored off-site through systems such as Firebase or MongoDB, personal details and results stay accessible yet secure. For machine learning jobs, it pulls from resources like Scikit-learn along with TensorFlow to get tasks done. Libraries including NLTK and SpaCy help handle Natural Language Processing tasks, whereas tools like LibROSA and PyAudio step in when dealing with speech emotion detection through sound analysis. Hosting, user verification, and protected data handling often happen using platforms such as Firebase or AWS in the background.

3.2 Data Collection

MediMind pulls together different kinds of information to get a clearer picture of how someone feels inside.

One kind comes from words - things like notes about mood, messages typed into chats, or thoughts shared each day. Another kind arrives through sound, captured in brief clips where voices show emotion. Before anything gets saved, people confirm they're okay with it. Once approved, every piece stays locked down, kept safe and private behind strong protections. How feelings unfold over time becomes visible only because trust and security come first.

3.3 Data Preprocessing

Starting off, cleaning up data plays a key role in making forecasts more accurate. When it comes to text, special steps clear out clutter before pulling useful details. Voice information follows its own path, using tailored methods to strip away static. Each type gets handled differently, yet both aim at sharper results.

3.3.1 Text Data Preprocessing

Words that add little meaning get stripped out first, cleaning up the raw input. After that comes splitting the text into smaller pieces like single words or full sentences. These bits are adjusted through standardizing forms and reducing verbs to their base versions. Turning language into numbers happens using methods like TF-IDF or embedding vectors instead of simple counts. Numbers shaped this way fit better with algorithms trained to recognize patterns in speech.

3.3.2 Voice Data Preprocessing

Cleaning up sound files begins by removing unwanted background sounds so speech stands out more clearly. What comes next is breaking down each recording into tiny chunks to spot subtle shifts in how someone feels. Mel-Frequency Cepstral Coefficients show up here, along with measurements of pitch and loudness, pulled straight from those clipped bits. After that, everything gets scaled evenly so one person's voice does not overpower another's when compared later.

3.4 Methodology

Starting off, the approach builds a clear path to study mental health. From the app, people share words and speech. Words give up patterns in how they're used; sounds pulled from voices show tone and pace. After that, computers learn from tagged examples - these guide guesses about feelings like calmness, tension,

sadness, joy, or worry. Each step links tightly, feeding into the next without gaps. When the model forecasts align with past emotional patterns, mood estimates emerge alongside alerts for potential stress. On the display people check daily, findings appear clearly - offering reflections on mental state paired with tailored guidance for well-being.

3.5 machine learning techniques used

A computer checks feelings in written words by spotting emotional clues, while separate tools study tone when someone speaks. Instead of just one way to decide, it uses several math-based strategies - like drawing lines between groups or building tree-like decisions - to label mental conditions. Each guess gets tested carefully through numbers that show how often it was right, catching both hits and misses without bias. Results stand up only if they prove steady across different ways of measuring success.

3.6 Security and privacy safeguards

Because mental health details demand care, MediMind builds in strong safeguards right from the start. Access happens only after identity checks that block unapproved entry. Data rests in scrambled format at all times, hidden even if found. Permissions shift based on roles, so one person sees less while another gains more - only what they need. Rules around medical secrecy shape how every feature handles personal facts.

3.7 Ethical Considerations

MediMind helps track mental well-being but won't take the place of expert medical care. Predictions it gives should be seen as guidance, nothing more. When signs of serious emotional strain show up, talking with a trained provider makes sense - this keeps AI use grounded and thoughtful. Because safety matters, leaning on human expertise stays key even when tech offers insights.

4. Experimental Results and Discussion

This part shows how MediMind, a tool for tracking mental well-being, performed in real tests. Its success hinged on predicting emotional states correctly while spotting signs of stress through spoken words and written text. Different algorithms went through training rounds before facing test trials, just to see which one

handled accuracy best. Results focused tightly on how consistently each model classified inputs without slipping up.

4.1 Experimental Setup

Running tests involved a set of tagged examples - some drawn from written expressions showing emotion, others pulled from speech clips tied to conditions like joy, calmness, pressure, worry, sadness. Splitting that collection followed an eighty-to-twenty split between practice material and unseen checkups. Before any learning began, distinct traits got pulled out - one path for words, another for sound waves. Work on each modality moved forward without mixing methods until models started shaping up.

Starting off with words, systems that judge feelings in text learned from word importance scores called TF-IDF. When it came to spoken voice, machines focused on sound traits - things like tone shape, how high or low the voice sits, and loudness patterns. A handful of different decision-making rules got tested against one another, just to see which one fits best when guessing emotional well- being states.

4.2 Performance Evaluation Metrics

Performance of the system got checked through common machine learning measures

Accuracy Precision Recall F1-Score
From the start, these measurements aim to track how often predictions are right while also checking if risks for mental health issues show up accurately. What matters here is balance - catching real cases without missing them, yet staying reliable across different outcomes.
4.3 Mood Prediction Results

Surprisingly accurate, the new system judges mood using both written words and spoken sound. Instead of just one method, it blends two - text clues plus vocal tones - to get better results. Written content helps spot clear emotions like joy or anger without much trouble. On their own, voice patterns catch signs of tension more often than before. When used together, these inputs make predictions steadier and less likely to fail.

4.4 Stress Risk Detected

Every now and then, signs of rising tension showed up when emotions turned sour across several exchanges. Noticing the same uneasy feelings pop up again helped spot trouble before it grew worse. Over time, spotting these repeating moods let the system step in sooner with small suggestions for balance. Moments like these revealed how tracking shifts could quietly guide support.

4.5 Comparative Analysis

Nowhere near as steady, basic linear methods fell short when stacked against grouped approaches like Random Forest. Instead of relying on one path, blended models handled shifts in data more smoothly. Surprisingly sharp at spotting emotions in written words, SVMs stood out clearly. When voices carried the clues, decision trees banded together gave clearer signals. Not every tool fits each task - some shine only under certain conditions.

4.6 Discussion

From lab tests, it turns out machines reading voices and words might help track mental well-being through phones. Combining both sound and written notes boosts how closely guesses match reality. Even if numbers look good, shifts in speaking rhythm or personal wording could sway results one way or another.

4.7 Limits of Test Outcomes

Prediction accuracy depends on data quality and diversity Emotional expression varies across individuals
What happens in tight lab conditions might not match messy everyday reality

Even so, the findings show promise for using this method to spot stress and mental health issues sooner. Though limited, they suggest it could work in real-world settings fairly well.

5. Conclusion

From sounds people make comes a tool called MediMind, built with AI to track feelings through speech and written words. Instead of waiting weeks, help might show up sooner when voices shift or texts change rhythm. Phones carry it now - quietly watching patterns that humans often miss during tough days. When tone dips low or word choice turns sharp, the system notices without asking first. It learns how silence stretches just before someone breaks down. Not every whisper gets flagged, only those hinting at deeper strain beneath normal talk. With time, small signals add up to warnings hidden in everyday messages. Some find comfort knowing machines can spot what friends sometimes ignore. Others question if tech should listen so closely without consent. Still, data flows both ways - users see feedback on their own behavior too. One day fatigue shows louder than sadness; next week anxiety masks itself as irritation. Patterns repeat across thousands yet stay personal enough to matter. Help arrives not because you ask but because your voice gives clues early. Even short sentences hold meaning when spoken twice each morning. No doctor needed right away, though alerts prepare them ahead of crisis. Mental shifts sneak in slowly until one moment everything feels heavy again. This system does not cure - it watches, waits, whispers back when risk climbs.

Testing showed mixing voice and written input makes mood tracking more dependable than using just one type alone. Because it catches shifts in emotion over time, help can start sooner before things get worse. Instead of waiting for crisis moments, people begin noticing feelings earlier with this setup. Watching these signs helps users respond while small changes still matter most. Prevention sits at the core here, quietly guiding attention toward inner balance each day.

Even if it cannot take the place of a doctor's evaluation, this tool acts like a smart helper for mental well-being, making check-ins on emotional health easier to reach. Because findings show artificial intelligence can work inside phone apps to assess psychological states, testing ideas this way makes sense. A step forward in digital care comes through MediMind, which brings together simplicity, growth potential, and thinking-like machines into today's treatment setups.

REFERENCES

1. 1 World Health Organization, "Mental Health: Strengthening Our Response," WHO, Geneva, 2023.

2. World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," WHO, 2017.

3. D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson Education, 2022.

4. R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997.'

5. J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The Development and Psychometric Properties of LIWC," *Behavior Research Methods*, vol. 47, no. 2, pp. 541–559, 2015.

6. S. Bhattacharyya and P. K. Singh, "Speech Emotion Recognition: A Review," *International Journal of Computer Applications*, vol. 96, no. 2, pp. 34–41, 2019.

7. F. Eyben, K. R. Scherer, B. Schuller et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS)," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

8. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *ICLR*, 2013.

9. A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI Workshop*, 1998.

10. Y. Liu, M. Ott, N. Goyal et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692*, 2019.

11. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

12. R. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.

13. A. Miner, L. Milstein, S. Schueller et al., "Smartphone-Based Conversational Agents and Mental Health," *JAMA Internal Medicine*, vol. 176, no. 5, pp. 619–625, 2016.

14. K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering CBT Using Automated Conversational Agents," *JMIR Mental Health*, vol. 4, no. 2, 2017.

15. S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "Wearable Sensors and Systems for Healthcare," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 21, 2012.

16. N. Cummins, S. Amiriparian, G. Hagerer et al., "A Review of Depression Recognition from Speech," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 403–418, 2019.

17. A. M. Hassan, A. R. Ali, and M. A. Hossain, "Emotion Detection Using NLP and Machine Learning," *Journal of Computer Science*, vol. 15, no. 4, pp. 543–550, 2019.

18. S. Schuller, B. Schuller, and F. Eyben, "Computational Paralinguistics: Emotion Recognition," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 97–103, 2010.

19.   M. De Choudhury, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," *ICWSM*, 2013.

20.   J. Torous and I. H. Keshavan, "The Role of Digital Technology in Mental Health Care," *World Psychiatry*, vol. 19, no. 1, pp. 3–4, 2020.

21.   P. Resnik, W. Armstrong, L. Claudino et al., "Beyond LDA: Exploring Supervised Topic Modeling for Depression," *CLPsych Workshop*, 2015.

22.   A. Ben-Zeev, D. Brian, R. Wang et al., "Mobile Behavioral Sensing for Mental Health,"

Psychiatric Services, vol. 66, no. 6, pp. 558–566, 2015.

23.   J. Maalouf, "Logistic Regression in Machine Learning," *International Journal of Data Science*, vol. 2, no. 2, pp. 1–15, 2011.

24.   L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

25.   U. Salma and M. Hasan, "AI-Based Mental Health Monitoring Systems: A Survey," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, 2021.