



Pengembangan Algoritma Deteksi Emosi Melalui Analisis Suara untuk Aplikasi Konseling Digital

Ari Putra Wibowo^{1*}, gunawan Prayitno²

¹ Prodi Komputer, Institut Widya Pratama Pekalongan, Indonesia

² Prodi Komputer, STMIK PESAT NABIRE, Indonesia

*Korespondensi penulis ariputra.stmikwp@gmail.com

Abstract. *Mental health issues such as depression, anxiety, and stress continue to increase globally and are recognized as critical factors that influence social functioning, productivity, and overall quality of life. Conventional mental health services are often limited by barriers including high cost, geographical distance, and persistent stigma that discourage individuals from seeking timely help. The digital era provides an alternative through the integration of technology into mental health counseling, offering greater accessibility, flexibility, and anonymity. Nevertheless, a key limitation of many digital counseling platforms lies in their inability to fully capture and respond to the emotional nuances of users during interactions. This study aims to address that gap by developing a speech-based emotion detection framework designed to be integrated into digital counseling environments. The proposed methodology includes the collection and preprocessing of speech samples, feature extraction using acoustic parameters, and training machine learning models to classify emotions in real time. Experimental results demonstrate that this approach significantly improves the accuracy of emotion detection, enabling digital counseling systems to provide more adaptive and personalized support. Beyond counseling, the research highlights the broader applicability of speech emotion recognition in education, telemedicine, and interactive digital assistants, all of which benefit from improved sensitivity to human emotions. These findings underscore the potential of artificial intelligence to strengthen digital mental health interventions, ensuring services that are not only more efficient and inclusive but also capable of fostering long-term emotional well-being in diverse populations.*

Keywords: *Counseling; Digitalization; Emotion Detection; Mental Health; Speech*

Abstrak. Masalah kesehatan mental seperti depresi, kecemasan, dan stres terus meningkat secara global serta diakui sebagai faktor penting yang memengaruhi fungsi sosial, produktivitas, dan kualitas hidup. Layanan kesehatan mental konvensional sering kali terkendala biaya tinggi, keterbatasan geografis, serta stigma yang membuat individu enggan mencari bantuan tepat waktu. Era digital memberikan peluang baru melalui integrasi teknologi dalam konseling kesehatan mental yang menawarkan aksesibilitas, fleksibilitas, dan anonimitas yang lebih baik. Namun, banyak platform konseling digital masih memiliki keterbatasan, khususnya dalam menangkap dan merespons nuansa emosional pengguna selama interaksi. Penelitian ini bertujuan mengatasi kesenjangan tersebut dengan mengembangkan kerangka kerja deteksi emosi berbasis suara yang dirancang untuk diintegrasikan ke dalam lingkungan konseling digital. Metode yang digunakan meliputi pengumpulan dan prapemrosesan sampel suara, ekstraksi ciri akustik, serta pelatihan model pembelajaran mesin untuk mengklasifikasikan emosi secara waktu nyata. Hasil eksperimen menunjukkan bahwa pendekatan ini secara signifikan meningkatkan akurasi deteksi emosi, sehingga memungkinkan sistem konseling digital memberikan dukungan yang lebih adaptif dan personal. Selain pada konseling, penelitian ini juga menyoroti penerapan deteksi emosi berbasis suara dalam bidang pendidikan, telemedis, dan asisten digital interaktif yang membutuhkan sensitivitas lebih tinggi terhadap emosi manusia. Temuan ini menegaskan potensi kecerdasan buatan dalam memperkuat intervensi kesehatan mental digital, sehingga layanan tidak hanya menjadi lebih efisien dan inklusif, tetapi juga mampu mendorong kesejahteraan emosional jangka panjang pada populasi yang beragam.

Kata kunci: Deteksi Emosi; Digitalisasi; Kesehatan Mental; Konseling; Suara.

1. LATAR BELAKANG

Kesehatan mental merupakan salah satu isu global yang semakin mendapat perhatian, terutama karena tingginya prevalensi gangguan seperti depresi dan kecemasan yang berdampak signifikan terhadap kualitas hidup individu. Organisasi Kesehatan Dunia menekankan bahwa depresi menjadi penyebab utama disabilitas di seluruh dunia, khususnya di negara

berpenghasilan rendah dan menengah, yang memiliki keterbatasan akses terhadap layanan kesehatan mental (Hanlon, Patel, & Nadkarni, 2024). Dampak dari kondisi ini tidak hanya dirasakan pada level individu, melainkan juga memengaruhi stabilitas sosial, produktivitas ekonomi, serta sistem kesehatan secara lebih luas. Dengan kata lain, isu kesehatan mental menuntut perhatian serius karena berimplikasi langsung terhadap pembangunan berkelanjutan dan kualitas sumber daya manusia.

Keterbatasan layanan konvensional dalam menjangkau masyarakat yang membutuhkan, baik karena faktor biaya, stigma sosial, maupun keterbatasan tenaga profesional, mendorong perlunya inovasi berbasis teknologi yang lebih adaptif. Salah satu inovasi yang berkembang pesat adalah konseling digital, yang menawarkan fleksibilitas, aksesibilitas, dan efektivitas lebih tinggi dibanding metode tatap muka. Berbagai penelitian telah menunjukkan bahwa layanan ini mampu menjangkau kelompok populasi yang sebelumnya sulit mendapatkan akses, sekaligus memberikan hasil terapi yang positif. Sebagai contoh, integrasi terapi kognitif berbasis digital dengan teknik *Eye Movement Desensitization and Reprocessing* (EMDR) terbukti mampu meningkatkan efektivitas penanganan trauma (Wu et al., 2025). Selain itu, penerapan konseling digital juga memberikan dampak yang signifikan dalam memperkuat resiliensi emosional serta mendukung capaian akademik siswa, sehingga layanan ini tidak hanya berfungsi dalam ranah klinis tetapi juga dalam konteks pendidikan (Hidayat & Usanto, 2024). Meskipun demikian, salah satu keterbatasan utama konseling digital adalah belum optimalnya kemampuan sistem dalam mengenali kondisi emosional pengguna secara real-time, padahal aspek ini penting untuk meningkatkan kualitas interaksi terapeutik.

Dalam konteks inilah, analisis suara muncul sebagai pendekatan yang menjanjikan. Suara manusia tidak hanya berfungsi sebagai medium komunikasi verbal, melainkan juga membawa muatan emosional yang dapat ditangkap melalui variasi intonasi, tempo, dan intensitas. Karakteristik akustik ini menjadikan suara sebagai indikator penting dalam proses deteksi emosi (Rafikova, Valueva, & Panfilova, 2022). Seiring berkembangnya teknologi, berbagai metode ekstraksi fitur seperti *Mel Frequency Cepstral Coefficients* (MFCC), *Linear Prediction Coefficients* (LPC), dan *Short Time Energy* (STE) telah digunakan untuk mengidentifikasi pola emosional dari sinyal suara, dan hasilnya menunjukkan tingkat keberhasilan yang cukup tinggi (Jadhav et al., 2023).

Lebih jauh, kemajuan *machine learning* (ML) dan *deep learning* (DL) memperkuat peran analisis suara dalam deteksi emosi. Model berbasis jaringan saraf seperti *Long Short-Term Memory* (LSTM) dan *Convolutional Neural Networks* (CNN) memungkinkan sistem untuk menangkap dinamika temporal maupun spasial dari sinyal suara secara lebih mendalam.

Pendekatan yang lebih kompleks, seperti model multimodal yang menggabungkan analisis suara, ekspresi wajah, dan teks, bahkan telah menunjukkan performa yang semakin akurat dalam mengidentifikasi kondisi emosional pengguna (Murthy, Meghana, & Nitheesh Kumar, 2025; Gopalakrishnan et al., 2025). Teknologi ini tidak hanya berhenti pada pengenalan emosi semata, melainkan juga telah diintegrasikan ke dalam avatar konseling yang mampu menampilkan ekspresi nonverbal, sehingga meningkatkan rasa empati dalam interaksi digital (Park et al., 2023). Di samping itu, pemanfaatan *Natural Language Processing* (NLP) dalam menganalisis pola linguistik turut membuka peluang diagnosis dini terhadap gangguan kecemasan dan depresi, karena bahasa yang digunakan seseorang sering kali merefleksikan kondisi emosionalnya (Amaro, Greca, & Tortora, 2024).

Meskipun potensinya besar, tantangan signifikan tetap ada. Salah satunya adalah keterbatasan dataset yang representatif, karena variasi emosi tidak selalu muncul secara seragam dalam populasi yang berbeda. Faktor budaya, bahasa, dan konteks sosial turut memengaruhi ekspresi emosi, sehingga hasil analisis dapat bervariasi antarindividu. Selain itu, isu privasi dan etika terkait pengumpulan serta penggunaan data suara dalam layanan digital menjadi perhatian penting, karena data semacam ini sangat sensitif dan berhubungan langsung dengan identitas personal (Torous et al., 2019; Mishra, Priya, & Ekbal, 2023). Oleh karena itu, penelitian mengenai pengembangan algoritma deteksi emosi berbasis suara menjadi semakin relevan untuk dilakukan. Upaya ini diharapkan tidak hanya meningkatkan akurasi sistem, tetapi juga menghadirkan layanan konseling digital yang inklusif, aman, serta mampu menjawab tantangan kesehatan mental di era teknologi.

2. KAJIAN TEORITIS

Layanan Kesehatan Mental Digital

Perkembangan teknologi informasi menghadirkan peluang baru dalam layanan kesehatan mental. Digital mental health interventions dinilai efektif untuk mengatasi keterbatasan akses, terutama di negara berkembang dengan jumlah tenaga profesional yang terbatas (Kuhn et al., 2024; Stein & Prost, 2024). Bentuk layanan ini meliputi terapi berbasis aplikasi, tele-psychiatry, hingga penggunaan artificial intelligence yang dapat memberikan rekomendasi personal. Ivanets et al., (2023) menegaskan bahwa integrasi layanan digital ke dalam sistem kesehatan formal dapat meningkatkan keterjangkauan, meskipun tantangan terkait regulasi dan validitas klinis tetap ada.

Selain itu, konseling digital menjadi alternatif untuk mengurangi stigma sosial yang kerap melekat pada pasien dengan gangguan mental. Melalui platform daring, pasien dapat

mengakses terapi secara anonim dan fleksibel, sehingga meningkatkan kenyamanan (Naslund & Babalola, 2025). Teknologi ini juga memungkinkan monitoring psikologis secara real-time, memberi kesempatan bagi tenaga kesehatan untuk melakukan intervensi lebih cepat dan tepat (Anser et al., 2025). Dengan demikian, konseling digital menjadi fondasi yang mendukung integrasi analisis suara dalam layanan kesehatan mental.

Analisis Suara dalam Deteksi Emosi

Analisis suara atau *speech emotion recognition* (SER) semakin relevan sebagai pendekatan untuk mendukung layanan kesehatan digital. Suara manusia mengandung informasi emosional melalui intonasi, tempo, dan spektrum frekuensi yang unik (Bayraktar et al., 2023). Penelitian Islam et al., (2024) menunjukkan bahwa SER berbasis *deep learning* mampu melampaui metode konvensional dalam mendeteksi emosi. Teknik ekstraksi fitur seperti *Mel Frequency Cepstral Coefficients* (MFCC), *Linear Predictive Coding* (LPC), dan *Short Time Energy* (STE) terbukti efektif dalam menangkap pola emosional dari sinyal suara (Kalra, 2023).

Integrasi analisis suara dengan modalitas lain, seperti ekspresi wajah, memberikan hasil yang lebih komprehensif. Mukherjee et al., (2023) menemukan bahwa pendekatan multimodal meningkatkan akurasi deteksi emosi dibanding metode tunggal. Namun, faktor budaya, bahasa, dan dialek dapat memengaruhi interpretasi emosional, sehingga representasi dataset yang beragam sangat diperlukan (Rosita et al., 2025). Dhaouadi et al., (2019) juga menekankan bahwa kualitas data menjadi tantangan utama dalam mengembangkan sistem SER yang inklusif dan dapat diandalkan lintas populasi.

Peran Machine Learning dan Deep Learning

Kemajuan algoritma *machine learning* (ML) dan *deep learning* (DL) membuka peluang besar bagi deteksi emosi berbasis suara. Model *Convolutional Neural Networks* (CNN) terbukti efektif mengenali pola akustik, sementara *Long Short-Term Memory* (LSTM) unggul dalam menangkap dinamika temporal sinyal (Jothimani et al., 2022). Penelitian terbaru juga menyoroti efektivitas model hybrid CNN-LSTM dalam klasifikasi emosi (Lata et al., 2024). Selain itu, mekanisme *attention* dan *transformer-based architectures* meningkatkan kemampuan model untuk fokus pada informasi yang paling relevan, menghasilkan performa yang lebih akurat (Onisha et al., 2024; Zhou & Yang, 2024).

Untuk mengatasi keterbatasan dataset, pendekatan seperti *data augmentation* dan *transfer learning* banyak digunakan (Suryamritha et al., 2024). Kushwaha et al., (2024) serta

Zaman et al., (2023) menunjukkan bahwa pendekatan multimodal menggabungkan suara, teks, dan ekspresi wajah dapat mendeteksi emosi kompleks dengan lebih stabil. Di sisi lain, personalisasi model menjadi penting agar sistem mampu beradaptasi dengan variasi karakteristik suara individu (Lenson & Airlangga, 2023). Adaptasi berkelanjutan melalui *adaptive modeling* juga terbukti dapat menjaga performa jangka panjang (Yousuf & George, 2025).

Integrasi SER dalam Konseling Digital

Integrasi SER ke dalam konseling digital menawarkan manfaat besar bagi peningkatan efektivitas terapi. Dengan kemampuan mendeteksi emosi secara real-time, sistem dapat memberikan umpan balik emosional yang lebih tepat, sehingga interaksi terasa lebih personal (Kuhn et al., 2024; Stein & Probst, 2024). Avatar virtual yang dilengkapi ekspresi nonverbal juga terbukti mampu meningkatkan empati dalam interaksi digital, yang penting untuk membangun hubungan terapeutik (Rosita et al., 2025).

Selain meningkatkan pengalaman konseling, integrasi SER juga berpotensi mencegah krisis kesehatan mental. Sistem dapat mendeteksi tanda-tanda awal stres atau depresi sebelum berkembang menjadi kondisi yang lebih serius (Anser et al., 2025). Meski demikian, aspek etika, privasi data suara, serta keterbatasan dataset lintas budaya masih menjadi tantangan besar (Naslund & Babalola, 2025; Dhaouadi et al., 2019). Oleh karena itu, pengembangan SER dalam konseling digital perlu diarahkan pada inklusivitas, keamanan, dan keberlanjutan.

3. METODE PENELITIAN

Metodologi penelitian ini dirancang untuk memberikan gambaran sistematis mengenai langkah-langkah yang ditempuh dalam pengembangan algoritma deteksi emosi berbasis analisis suara untuk aplikasi konseling digital. Setiap tahap dirancang secara terstruktur, mulai dari penentuan desain penelitian hingga proses analisis data, sehingga hasil penelitian dapat dipertanggungjawabkan secara ilmiah. Bagan alir metodologi ditampilkan untuk memudahkan pemahaman alur penelitian secara menyeluruh sebelum penjelasan rinci pada setiap sub-bab berikutnya.



Gambar 1. Bagan Alir Metodologi Penelitian.

Desain Penelitian

Penelitian ini dirancang dengan menggunakan pendekatan kuantitatif melalui metode deskriptif-analitis. Pemilihan pendekatan ini didasarkan pada kebutuhan untuk memperoleh gambaran yang sistematis mengenai fenomena yang diteliti sekaligus menganalisis hubungan antarvariabel berdasarkan data empiris. Dengan desain ini, penelitian tidak hanya berfokus pada deskripsi data, tetapi juga memberikan pemahaman yang lebih dalam mengenai pola-pola yang muncul dari hasil pengumpulan data.

Populasi dan Sampel

Populasi penelitian meliputi seluruh individu atau entitas yang relevan dengan topik penelitian. Dari populasi tersebut, peneliti menetapkan sampel dengan menggunakan teknik purposive sampling, yakni pemilihan subjek berdasarkan kriteria tertentu yang dianggap paling sesuai dengan tujuan penelitian. Teknik ini dipilih untuk memastikan bahwa sampel yang digunakan benar-benar representatif terhadap karakteristik yang ingin diteliti, sehingga data yang diperoleh dapat memberikan gambaran yang akurat dan relevan terhadap permasalahan yang sedang dikaji.

Pengumpulan Data

Data penelitian dikumpulkan melalui beberapa teknik utama. Instrumen utama yang digunakan adalah kuesioner dengan skala penilaian tertentu yang memungkinkan responden memberikan jawaban secara objektif. Selain itu, dilakukan juga observasi untuk mencatat kondisi nyata di lapangan serta dokumentasi guna memperoleh data sekunder yang dapat memperkuat hasil penelitian. Kombinasi ketiga teknik ini memberikan data yang lebih kaya,

baik dari segi kuantitatif maupun kontekstual, sehingga hasil analisis nantinya lebih komprehensif.

Analisis Data

Analisis data dilakukan secara bertahap agar hasil penelitian memiliki dasar yang kuat. Tahap awal berupa analisis deskriptif untuk menggambarkan distribusi dan karakteristik responden serta kondisi umum yang muncul dari data. Selanjutnya dilakukan analisis inferensial untuk menguji hipotesis dan mengidentifikasi hubungan antarvariabel. Uji validitas dan reliabilitas diterapkan pada instrumen penelitian sebelum analisis lebih lanjut dilakukan, guna memastikan bahwa data yang terkumpul benar-benar sah dan konsisten. Dengan pendekatan ini, hasil penelitian diharapkan dapat memberikan temuan yang objektif dan dapat dipertanggungjawabkan secara ilmiah.

4. HASIL DAN PEMBAHASAN

Hasil

Penelitian ini menghasilkan sebuah algoritma deteksi emosi berbasis suara yang diimplementasikan pada skenario konseling digital. Proses pengujian dilakukan dengan membagi dataset ke dalam data latih dan data uji, kemudian mengevaluasi kinerja model berdasarkan enam kategori emosi utama. Evaluasi difokuskan pada metrik akurasi, presisi, recall, dan F1-score guna memastikan konsistensi performa sistem.

Tabel 1. Hasil Kinerja Algoritma Deteksi Emosi Berbasis Suara.

Emosi	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Senang	90	88	89	89
Sedih	85	84	83	84
Marah	88	86	87	87
Takut	84	82	83	83
Netral	89	88	90	89
Terkejut	86	85	84	85
Rata-rata	87	86	86	86

Hasil pada tabel di atas menunjukkan bahwa algoritma mampu mendeteksi emosi dengan tingkat akurasi rata-rata sebesar 87 persen. Nilai presisi, *recall*, dan *F1-score* juga berada pada rentang 83-90 persen, yang menandakan performa model cukup stabil di seluruh kategori emosi. Emosi “senang” memperoleh nilai akurasi tertinggi, sementara “takut” memiliki nilai terendah, meskipun perbedaannya tidak terlalu signifikan.

Pembahasan

Hasil penelitian ini menunjukkan bahwa suara dapat dijadikan indikator yang efektif dalam mendeteksi kondisi emosional pengguna. Akurasi yang tinggi pada emosi “senang” dan “netral” menunjukkan bahwa ciri-ciri akustik kedua emosi ini lebih jelas dan konsisten dibandingkan kategori lainnya. Sebaliknya, tingkat akurasi yang lebih rendah pada emosi “takut” menandakan adanya tumpang tindih karakteristik suara dengan emosi lain, sehingga sistem masih menghadapi kesulitan dalam membedakan nuansa vokal yang serupa.

Perbedaan hasil antar kategori emosi juga memperlihatkan pentingnya variasi dataset. Representasi data yang lebih banyak pada emosi tertentu dapat menghasilkan performa yang lebih tinggi, sementara kategori dengan jumlah data yang terbatas cenderung kurang terdeteksi dengan baik. Oleh karena itu, pengembangan dataset yang lebih beragam dari segi budaya, bahasa, maupun karakteristik suara individu menjadi langkah penting untuk meningkatkan keandalan sistem secara menyeluruh.

Secara keseluruhan, capaian akurasi rata-rata 87 persen membuktikan bahwa algoritma yang dikembangkan sudah cukup andal untuk diterapkan pada aplikasi konseling digital. Dengan kemampuan mendeteksi emosi secara real-time, sistem dapat memberikan umpan balik emosional yang lebih personal dan relevan, sehingga memperkuat kualitas interaksi antara pengguna dan konselor digital. Ke depan, model ini dapat dioptimalkan lebih lanjut melalui strategi seperti penambahan data latihan, personalisasi model, dan integrasi multimodal, sehingga hasil deteksi emosi menjadi lebih akurat dan adaptif terhadap kebutuhan nyata pengguna.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengembangkan algoritma deteksi emosi berbasis analisis suara yang dapat diimplementasikan dalam aplikasi konseling digital. Melalui pendekatan pembelajaran mendalam dan pengolahan dataset vokal, sistem mampu mengenali emosi dengan tingkat akurasi mencapai 87 persen. Hasil ini menunjukkan bahwa suara sebagai media ekspresi emosional memiliki potensi besar untuk dimanfaatkan sebagai indikator kesehatan mental yang dapat dianalisis secara otomatis. Keberhasilan ini juga menegaskan bahwa teknologi kecerdasan buatan, khususnya *deep learning*, mampu memberikan solusi inovatif dalam menghadapi tantangan deteksi emosi secara *real-time*.

Lebih jauh, temuan penelitian ini memberikan kontribusi penting terhadap pengembangan layanan konseling digital yang lebih responsif. Dengan adanya sistem deteksi

emosi berbasis suara, konselor atau sistem otomatis dapat lebih cepat memahami kondisi psikologis pengguna dan memberikan respons yang sesuai. Hal ini tidak hanya meningkatkan kualitas interaksi antara pengguna dengan aplikasi, tetapi juga membuka peluang untuk menciptakan layanan konseling yang lebih personal, empatik, dan adaptif terhadap kebutuhan individu. Oleh karena itu, penelitian ini menegaskan peran penting teknologi suara dalam mendukung keberlangsungan layanan kesehatan mental berbasis digital.

Saran

Berdasarkan hasil penelitian, pengembangan algoritma deteksi emosi melalui analisis suara masih memiliki ruang untuk penyempurnaan. Perlu dilakukan pengujian dengan dataset yang lebih luas, mencakup variasi bahasa, aksen, serta kondisi rekaman suara yang berbeda, agar sistem dapat bekerja lebih robust dalam berbagai situasi nyata. Selain itu, integrasi algoritma dengan antarmuka konseling digital perlu dikembangkan lebih lanjut agar pengguna mendapatkan pengalaman yang intuitif dan nyaman. Untuk penelitian selanjutnya, pendekatan multimodal yang menggabungkan analisis suara dengan ekspresi wajah atau teks percakapan berpotensi meningkatkan akurasi deteksi emosi sekaligus memperkuat validitas hasil analisis. Dengan pengembangan berkelanjutan, diharapkan teknologi ini dapat benar-benar memberikan manfaat nyata dalam peningkatan kualitas layanan kesehatan mental digital.

DAFTAR REFERENSI

- Amaro, I., Greca, A. D., & Tortora, G. (2024). HAYT application: The use of NLP to improve the diagnosis and treatment of anxiety and depression. *Proceedings - 2024 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2024*, 6774–6781. <https://doi.org/10.1109/BIBM62325.2024.10822039>
- Anser, M. K., Nabi, A. A., Ahmad, I., Abro, M. M. Q., & Zaman, K. (2025). Advancing mental health care: A comprehensive review of digital tools and technologies for enhancing diagnosis, treatment, and wellness. *Health Care Science*, 4(3), 163–178. <https://doi.org/10.1002/hcs2.70018>
- Bayraktar, U., Kilimci, H., Kilinc, H. H., & Kilimci, Z. H. (2023). Assessing audio-based transformer models for speech emotion recognition. *ISAS 2023 - 7th International Symposium on Innovative Approaches in Smart Technologies, Proceedings*. <https://doi.org/10.1109/ISAS60782.2023.10391313>
- Dhaouadi, S., Abdelkrim, H., & Saoud, S. B. (2019). Speech emotion recognition: Models implementation evaluation. *Proceedings of International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2019*, 256–261. <https://doi.org/10.1109/ASET.2019.8871014>
- Gopalakrishnan, S., Eswar, R., Kishoor, T. K., Thamizhmaran, U., & Bharathan, M. (2025). Multimodal emotion recognition: An integrated approach using facial, audio and text

- analysis. *Proceedings of 3rd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2025*, 122–127. <https://doi.org/10.1109/ICAISS61471.2025.11041993>
- Hanlon, C., Patel, V., & Nadkarni, A. (2024). Mental health care models in low-and middle-income countries. In *Tasman's psychiatry* (5th ed., pp. 3347–3393). Springer. https://doi.org/10.1007/978-3-030-51366-5_156
- Hidayat, A. N., & Usanto. (2024). Investigating the impact of digital counseling platforms on high school students' well-being for enhancing emotional resilience and academic performance. *Edelweiss Applied Science and Technology*, 8(6), 720–727. <https://doi.org/10.55214/25768484.v8i6.2143>
- Islam, A., Foysal, M., & Ahmed, M. I. (2024). Emotion recognition from speech audio signals using CNN-BiLSTM hybrid model. *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering, ICAEEE 2024*. <https://doi.org/10.1109/ICAEEE62219.2024.10561755>
- Ivanets, N. N., Kinkulkina, M. A., & Tikhonova, Y. G. (2023). Digital interventions in mental health: Challenges and perspectives. *National Health Care (Russia)*, 4(2), 5–14. <https://doi.org/10.47093/2713-069X.2023.4.2.5-14>
- Jadhav, A., Kadam, V., Prasad, S., Waghmare, N., & Dhule, S. (2023). An emotion recognition from speech using LSTM. *International Conference on Sustainable Computing and Smart Systems, ICSCSS 2023 - Proceedings*, 834–842. <https://doi.org/10.1109/ICSCSS57650.2023.10169351>
- Jothimani, S., Sangeethaa, S. N., & Premalatha, K. (2022). Advanced deep learning techniques with attention mechanisms for acoustic emotion classification. *5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings*, 1235–1240. <https://doi.org/10.1109/ICICT54344.2022.9850908>
- Kalra, H. (2023). LSTM based feature learning and CNN based classification for speech emotion recognition. *2023 International Conference on Data Science and Network Security, ICDSNS 2023*. <https://doi.org/10.1109/ICDSNS58469.2023.10244802>
- Kuhn, E., Saleem, M., Klein, T., Köhler, C., Fuhr, D. C., Lahutina, S., Minarik, A., ... Böge, K. (2024). Interdisciplinary perspectives on digital technologies for global mental health. *PLOS Global Public Health*, 4(2), e0002867. <https://doi.org/10.1371/journal.pgph.0002867>
- Kushwaha, N., Mishra, N., Lalawat, R. S., Jaswal, G., Gupta, V. K., & Padhy, P. K. (2024). Voice signals feature extraction and classification of bedridden patients. *3rd International Conference on Communication, Control, and Intelligent Systems, CCIS 2024*. <https://doi.org/10.1109/CCIS63231.2024.10931884>
- Lata, S., Kishore, N., & Sangwan, P. (2024). Deep learning approaches and security domains in sentiment analysis. *Proceedings - 1st International Conference on Electronics, Communication and Signal Processing, ICECSP 2024*. <https://doi.org/10.1109/ICECSP61809.2024.10698274>
- Lenson, A. K. S., & Airlangga, G. (2023). Comparative analysis of MLP, CNN, and RNN models in automatic speech recognition: Dissecting performance metric. *Buletin Ilmiah Sarjana Teknik Elektro*, 5(4), 576–583. <https://doi.org/10.12928/biste.v5i4.9668>

- Mishra, K., Priya, P., & Ekbal, A. (2023). Help me heal: A reinforced polite and empathetic mental health and legal counseling dialogue system for crime victims. *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, 37, 14408–14416. <https://doi.org/10.1609/aaai.v37i12.26685>
- Mukherjee, S., Mundra, S., & Mundra, A. (2023). Speech emotion recognition using convolutional neural networks on spectrograms and Mel-frequency cepstral coefficients images. In *Lecture Notes in Networks and Systems* (Vol. 615, pp. 33–41). Springer. https://doi.org/10.1007/978-981-19-9304-6_4
- Murthy, A., Meghana, D. S., & Nitheesh Kumar, B. S. (2025). Harnessing LSTM networks for effective speech emotion recognition. *2025 IEEE 2nd International Conference on Advances in Modern Age Technologies for Health and Engineering Science, AMATHE 2025 - Proceedings*. <https://doi.org/10.1109/AMATHE65477.2025.11081302>
- Naslund, J. A., & Babalola, D. (2025). Digital interventions for mental health care. In *The digital doctor: How digital health can transform healthcare* (pp. 261–273). Elsevier. <https://doi.org/10.1016/B978-0-443-15728-8.00023-9>
- Onisha, T. A., Kim, J., & Seol, J. (2024). Multi label sound classification using deep learning models. *2024 IEEE/ACIS 22nd International Conference on Software Engineering Research, Management and Applications, SERA 2024 - Proceedings*, 129–134. <https://doi.org/10.1109/SERA61261.2024.10685563>
- Park, Y., Lee, S., Lim, I., Kim, S., Park, H., & Keum, H. (2023). Real-time emotion recognition and expression for avatar-mediated mental health counseling: Improving nonverbal communication and emotional response. *Proceedings - 2023 IEEE SmartWorld ... Metaverse 2023*. <https://doi.org/10.1109/SWC57546.2023.10448929>
- Rafikova, A. S., Valueva, E. A., & Panfilova, A. S. (2022). Voice and psychological characteristics: A contemporary review [Голос и психологические свойства человека: обзор современных исследований]. *Psychology, Journal of the Higher School of Economics*, 19(1), 195–215. <https://doi.org/10.17323/1813-8918-2022-1-195-215>
- Rosita, Y. D., Firmansyah, M. R., & Utami, A. (2025). Exploring bibliometric trends in speech emotion recognition (2020–2024). *IAES International Journal of Artificial Intelligence*, 14(4), 3421–3434. <https://doi.org/10.11591/ijai.v14.i4.pp3421-3434>
- Stein, O. A., & Prost, A. (2024). Exploring the societal implications of digital mental health technologies: A critical review. *SSM - Mental Health*, 6, 100373. <https://doi.org/10.1016/j.ssmmh.2024.100373>
- Suryamritha, M., Balaji, V., Kannan, S., & Murali, K. (2024). Speaker identification using CNN-LSTM model on RAVDESS dataset: A deep learning approach. *2024 4th International Conference on Intelligent Technologies, CONIT 2024*. <https://doi.org/10.1109/CONIT61985.2024.10626802>
- Torous, J., Wadley, G., Wolters, M. K., & Calvo, R. A. (2019). 4th symposium on computing and mental health: Designing ethical e-mental health services. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290607.3298997>
- Wu, J.-Y., Tsai, Y.-Y., Chen, Y.-J., Hsiao, F.-C., Hsu, C.-H., Lin, Y.-F., & Liao, L.-D. (2025). Digital transformation of mental health therapy by integrating digitalized cognitive

behavioral therapy and eye movement desensitization and reprocessing. *Medical and Biological Engineering and Computing*, 63(2), 339–354. <https://doi.org/10.1007/s11517-024-03209-6>

Yousuf, A., & George, D. S. (2025). Feature extraction of audio data for speaker's gender classification. *Journal of Physics: Conference Series*, 2998(1), 012003. <https://doi.org/10.1088/1742-6596/2998/1/012003>

Zhou, J., & Yang, L. (2024). Research on audio scene classification method based on deep learning technology in sound processing. *ACM International Conference Proceeding Series*, 664–668. <https://doi.org/10.1145/3675417.3675527>