Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 2 (2026) pp: 26029-26036
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Pemanfaatan AI Multimodal Untuk Deteksi Dini Stres Melalui Suara, Teks, dan Fisiologis

Muhammad Tafazzani Addien[1], Apriade Voutama[2], Taufiq Ridwan[3]
[1-3]Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
2310631250102@student.unsika.ac.id, apriade.voutama@staff.unsika.ac.id, taufik.ridwan@cs.unsika.ac.id

Abstrak

Deteksi dini stres menjadi kebutuhan penting karena metode konvensional yang mengandalkan kuesioner dan laporan diri masih rentan terhadap bias subjektif, keterbatasan akurasi, serta belum mampu memantau perubahan kondisi psikologis secara berkelanjutan dan real-time. Perkembangan kecerdasan buatan (Artificial Intelligence/AI) membuka peluang untuk mengembangkan sistem deteksi stres yang lebih objektif melalui pemanfaatan berbagai sumber data secara simultan. Artikel ini bertujuan menyusun rancangan konseptual pemanfaatan AI multimodal dalam mendeteksi stres melalui analisis suara, teks, dan sinyal fisiologis sebagai dasar pengembangan sistem skrining kesehatan mental digital. Metode yang digunakan adalah studi literatur naratif terhadap tujuh belas artikel ilmiah yang membahas kesehatan mental digital, machine learning, speech emotion recognition, analisis teks, sensor fisiologis, serta chatbot berbasis AI. Hasil kajian menunjukkan bahwa setiap modalitas memiliki karakteristik dan keunggulan yang saling melengkapi. Analisis teks mampu mengidentifikasi pola linguistik, pilihan kata, dan sentimen, analisis suara menangkap perubahan prosodi, intonasi, serta fitur akustik yang berkaitan dengan kondisi emosional, sedangkan sinyal fisiologis merepresentasikan respons biologis yang relatif lebih objektif terhadap stres. Integrasi ketiga modalitas melalui pendekatan fusi fitur, fusi keputusan, maupun fusi hybrid berpotensi meningkatkan akurasi, stabilitas, dan keandalan sistem dibandingkan pendekatan unimodal. Meskipun demikian, implementasinya masih memerlukan validasi klinis, perlindungan privasi data, transparansi model AI, serta mekanisme rujukan kepada tenaga profesional. Oleh karena itu, AI multimodal lebih tepat diposisikan sebagai alat skrining dan pendukung deteksi dini kesehatan mental, bukan sebagai pengganti diagnosis psikologis maupun layanan konseling profesional.

Kata Kunci: *AI Multimodal, Deteksi Stres, Kesehatan Mental Digital, Machine Learning, Sinyal Fisiologis*

1. Latar Belakang

Kesehatan mental menjadi persoalan penting karena gangguan psikologis tidak hanya berdampak pada individu, tetapi juga pada keluarga, pendidikan, produktivitas, dan sistem pelayanan kesehatan. Kajian mengenai kondisi kesehatan mental di Indonesia menunjukkan bahwa penanganan kesehatan jiwa masih membutuhkan pendekatan promotif, preventif, kuratif, dan rehabilitatif yang saling terhubung agar masyarakat memperoleh dukungan sebelum gejala berkembang menjadi berat [1]. Dalam konteks remaja dan kelompok usia produktif, stres yang tidak dikenali sejak dini dapat berkembang menjadi depresi, kecemasan, penurunan fungsi sosial, gangguan tidur, hingga perilaku berisiko. Oleh karena itu, sistem deteksi awal yang mudah diakses menjadi kebutuhan penting, terutama bagi individu yang enggan mencari bantuan profesional karena stigma, biaya, atau keterbatasan layanan.

Selama ini, pengukuran stres banyak dilakukan melalui instrumen psikologis, misalnya Perceived Stress Scale (PSS). Instrumen tersebut bermanfaat karena terstruktur, mudah digunakan, dan dapat memberi gambaran awal mengenai tingkat stres seseorang [3]. Namun, metode self-report tetap memiliki kelemahan karena bergantung pada kemampuan individu untuk mengenali, mengingat, dan melaporkan kondisinya secara jujur. Padahal, stres bersifat dinamis dan dapat berubah dalam waktu singkat. Seseorang dapat tampak baik-baik saja pada saat mengisi kuesioner, tetapi menunjukkan tekanan melalui suara, pilihan kata, atau respons fisiologis pada konteks lain.

Perkembangan kecerdasan buatan memberikan peluang untuk mendeteksi stres melalui jejak digital dan sinyal tubuh. Pada modalitas teks, unggahan media sosial dan percakapan digital dapat dianalisis menggunakan Natural Language Processing (NLP), klasifikasi machine learning, serta metode interpretabilitas seperti SHAP untuk mengenali kata atau frasa yang berkaitan dengan stres dan depresi [4]. Tinjauan mengenai deteksi depresi berbasis machine learning juga menunjukkan bahwa data digital dapat membantu skrining awal, meskipun hasilnya tetap perlu ditafsirkan secara hati-hati dan tidak boleh langsung disamakan dengan diagnosis klinis [5].

Pada modalitas suara, emosi dapat dikenali melalui perubahan intonasi, energi, tempo, pitch, dan fitur akustik seperti Mel Frequency Cepstral Coefficients (MFCC). Penelitian speech emotion recognition menunjukkan bahwa model deep learning, seperti CNN, RNN, LSTM, dan BiLSTM, dapat membantu komputer mengenali pola emosi manusia berdasarkan sinyal wicara [6]-[9]. Sementara itu, pada modalitas fisiologis, sinyal EKG, PPG, HRV, electrodermal activity, dan suhu kulit dapat menggambarkan respons tubuh terhadap tekanan psikologis [10]-[12].

Walaupun berbagai penelitian tersebut menjanjikan, sebagian besar sistem masih bergerak pada satu modalitas. Pendekatan unimodal memiliki kelemahan karena stres tidak selalu muncul dengan cara yang sama pada setiap individu. Pengguna dapat menulis kalimat netral tetapi menunjukkan perubahan suara dan detak jantung, atau sebaliknya. Karena itu, sistem yang menggabungkan suara, teks, dan sinyal fisiologis lebih sesuai untuk menggambarkan kondisi stres secara menyeluruh. Integrasi ini juga relevan dengan pengembangan chatbot kesehatan mental berbasis Large Language Model (LLM), deteksi emosi, Retrieval-Augmented Generation (RAG), dan layanan konsultasi digital yang membutuhkan respons empatik, faktual, dan aman [13]-[15].

Kebaruan artikel ini terletak pada penyusunan rancangan konseptual AI multimodal yang menghubungkan tiga sumber utama ekspresi stres, yaitu teks, suara, dan sinyal fisiologis. Rancangan ini juga menempatkan prinsip keamanan, privasi, dan human-in-the-loop sebagai bagian dari arsitektur, sehingga sistem tidak diarahkan untuk menggantikan tenaga profesional. Artikel ini bertujuan menjelaskan kebutuhan deteksi dini stres, menyintesis peran tiap modalitas, serta menawarkan desain sistem yang dapat menjadi dasar pengembangan aplikasi kesehatan mental digital.

2. Metode Penelitian

2.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan metode studi literatur naratif dengan pendekatan perancangan konseptual. Pendekatan ini dipilih karena tujuan artikel bukan menguji model pada dataset baru, melainkan menyusun kerangka ilmiah mengenai bagaimana suara, teks, dan sinyal fisiologis dapat digabungkan dalam sistem deteksi dini stres. Dengan demikian, artikel ini tidak menyajikan klaim akurasi eksperimen baru, tetapi menyajikan sintesis temuan penelitian terdahulu dan rancangan sistem yang dapat diuji dalam penelitian lanjutan.

Studi literatur naratif memungkinkan penulis membandingkan beberapa kelompok penelitian yang berbeda, mulai dari kesehatan mental digital, analisis teks, deteksi emosi berbasis suara, sensor fisiologis, hingga chatbot kesehatan mental. Pendekatan ini sesuai karena isu AI multimodal berada pada irisan ilmu komputer, psikologi, kesehatan digital, dan etika teknologi.

2.2. Sumber Data dan Kriteria Literatur

Sumber data penelitian berupa artikel ilmiah yang dilampirkan dalam bentuk PDF. Literatur yang digunakan berjumlah tujuh belas referensi dan dipilih berdasarkan keterkaitannya dengan lima topik utama, yaitu kesehatan mental dan stres, aplikasi deteksi stres, analisis teks, pengenalan emosi melalui suara, sinyal fisiologis, serta chatbot kesehatan mental. Artikel yang dipakai mencakup penelitian nasional dan internasional agar rancangan sistem tetap relevan dengan konteks Indonesia sekaligus mengikuti perkembangan AI terkini.

Kriteria inklusi literatur adalah artikel yang membahas deteksi stres, kesehatan mental digital, machine learning, deep learning, speech emotion recognition, analisis sentimen, sensor fisiologis, atau chatbot. Artikel yang hanya membahas teori umum tanpa keterkaitan dengan sistem komputasional tidak dijadikan dasar utama. Setiap artikel dibaca untuk memperoleh informasi mengenai tujuan penelitian, jenis data, metode, algoritma, hasil utama, serta keterbatasan.

2.3. Tahapan Analisis Literatur

Tahapan analisis dilakukan melalui empat langkah. Pertama, penulis mengelompokkan literatur berdasarkan modalitas data, yaitu teks, suara, fisiologis, dan chatbot. Kedua, penulis mengekstraksi teknik yang digunakan pada masing-masing modalitas, misalnya TF-IDF, BERT, RoBERTa, SVM, CNN, LSTM, MFCC, HRV, PPG, EKG, SMOTE, dan RAG. Ketiga, penulis membandingkan kekuatan dan keterbatasan masing-masing pendekatan. Keempat, hasil perbandingan digunakan untuk menyusun rancangan arsitektur AI multimodal.

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Sintesis tidak dilakukan dengan menghitung meta-analisis statistik karena artikel yang digunakan memiliki desain, dataset, dan metrik yang beragam. Oleh karena itu, pendekatan yang lebih tepat adalah sintesis konseptual, yaitu menghubungkan hasil penelitian menjadi kerangka rancangan yang logis dan dapat diimplementasikan.

2.4. Rancangan Sistem yang Diusulkan

Rancangan sistem yang diusulkan terdiri atas lima tahap utama. Tahap pertama adalah akuisisi data dari percakapan teks, rekaman suara, dan sensor fisiologis dengan persetujuan pengguna. Tahap kedua adalah prapemrosesan, mencakup pembersihan teks, normalisasi bahasa, noise reduction pada suara, segmentasi audio, filtering artefak fisiologis, serta penyelarasan waktu antar modalitas. Tahap ketiga adalah ekstraksi fitur, yaitu embedding teks berbasis transformer, fitur akustik seperti MFCC, pitch, energi, dan fitur fisiologis seperti HRV. Tahap keempat adalah pemodelan unimodal dan fusi multimodal. Tahap kelima adalah evaluasi menggunakan akurasi, presisi, recall, F1-score, AUC-ROC, confusion matrix, serta uji keamanan respons jika sistem dihubungkan dengan chatbot kesehatan mental.

Karena data yang dikaji berkaitan dengan kondisi psikologis dan sinyal personal, aspek etika ditempatkan sebagai komponen metode. Sistem deteksi stres harus menerapkan persetujuan sadar, anonimisasi data, pembatasan akses, penyimpanan aman, dan pemberitahuan bahwa hasil deteksi bersifat skrining awal. Jika sistem menemukan indikasi risiko tinggi, respons yang diberikan perlu berupa edukasi aman dan rujukan ke tenaga profesional, bukan diagnosis otomatis.

3. Hasil dan Diskusi

3.1. Kebutuhan Sistem Deteksi Dini Stres

Hasil sintesis menunjukkan bahwa kebutuhan deteksi dini stres muncul dari dua persoalan utama, yaitu meningkatnya kebutuhan layanan kesehatan mental dan keterbatasan akses layanan profesional. Studi mengenai kondisi kesehatan mental di Indonesia menekankan pentingnya pendekatan promotif dan preventif agar masalah psikologis tidak baru ditangani ketika sudah berat [1]. Aplikasi digital dapat mendukung kebutuhan ini karena mudah digunakan, dapat diakses kapan saja, serta memberikan rasa privat bagi pengguna yang belum siap bertemu tenaga profesional secara langsung [2].

Namun, sistem digital yang hanya mengandalkan pengisian skala tetap memiliki kelemahan. Instrumen PSS dapat membantu mengukur persepsi stres, tetapi hasilnya tetap dipengaruhi oleh kesadaran diri, kejujuran, dan waktu pengisian [3]. Pada kondisi tertentu, pengguna mungkin meremehkan gejalanya, tidak mampu mengungkapkan tekanan secara verbal, atau justru mengisi jawaban berdasarkan suasana hati sesaat. Dengan demikian, deteksi stres yang lebih adaptif perlu membaca lebih dari satu sumber data.

Analisis unggahan media sosial memperlihatkan bahwa kata seperti stres, lelah, cemas, depresi, kecewa, dan menyerah dapat menjadi indikator penting dalam klasifikasi kondisi mental [4]. Tinjauan machine learning untuk deteksi depresi juga menunjukkan bahwa berbagai algoritma dapat membantu proses identifikasi awal, khususnya ketika data digital diproses secara sistematis [5]. Temuan ini menguatkan bahwa deteksi dini berbasis AI dapat berperan sebagai alat bantu skrining, tetapi tidak dapat berdiri sendiri sebagai alat diagnosis klinis.

3.2. Keterbatasan Pendekatan Unimodal

Pendekatan unimodal adalah sistem yang hanya menggunakan satu jenis data, misalnya hanya teks, hanya suara, atau hanya detak jantung. Pendekatan ini lebih sederhana, tetapi kurang kuat dalam menghadapi variasi ekspresi stres. Pada modalitas teks, seseorang dapat menyembunyikan emosinya melalui pilihan kata yang netral. Pada modalitas suara, kualitas rekaman dapat terganggu oleh kebisingan, mikrofon, atau kondisi lingkungan. Pada modalitas fisiologis, perubahan detak jantung tidak selalu disebabkan oleh stres, tetapi dapat dipengaruhi olahraga, konsumsi kafein, penyakit, atau kurang tidur.

Kelemahan tersebut menunjukkan bahwa stres tidak dapat dipahami hanya sebagai pola kata, pola suara, atau pola tubuh secara terpisah. Stres merupakan pengalaman psikologis yang tampak melalui kombinasi respons kognitif, emosional, perilaku, dan fisiologis. Karena itu, sistem deteksi yang lebih kuat perlu menggabungkan beberapa modalitas agar kesalahan pada satu sumber data dapat dikompensasi oleh sumber data lain.

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 1. Sintesis Peran Modalitas dalam Deteksi Dini Stres

Modalitas	Data/Fitur Utama	Peran dalam Sistem
Teks	Token, sentimen, kata kunci, embedding transformer	Mendeteksi pola linguistik, pilihan kata, perubahan sentimen, dan konteks percakapan yang dapat mengarah pada stres atau depresi [4], [16].
Suara	MFCC, pitch, energi, tempo, spectrogram	Mengenali ekspresi afektif melalui intonasi, tekanan suara, ritme bicara, dan perubahan prosodi [6]-[9].
Fisiologis	HRV, EKG, PPG, EDA, suhu kulit	Mengamati respons tubuh yang relatif objektif dan dapat berlangsung sebelum pengguna menyadari kondisi stres [10]-[12].
Chatbot	Riwayat percakapan, deteksi emosi, basis pengetahuan	Menjadi antarmuka dukungan awal yang dapat memberi respons empatik, edukatif, dan aman [13]-[15].

3.3. Analisis Modalitas Teks

Modalitas teks menjadi salah satu sumber data yang paling mudah diperoleh karena banyak interaksi kesehatan mental digital berlangsung melalui pesan, unggahan media sosial, atau percakapan chatbot. Analisis teks memungkinkan sistem membaca pilihan kata, intensitas emosi, pola negasi, penggunaan kata berisiko, serta perubahan gaya bahasa. Pada penelitian deteksi stres dan depresi unggahan media sosial, tahapan NLP seperti tokenisasi, penghapusan stopword, stemming, TF-IDF, dan klasifikasi SVM digunakan untuk memisahkan kategori stres, depresi, dan netral [4].

Kekuatan modalitas teks terletak pada kemampuannya menangkap isi kognitif. Pengguna yang mengalami stres sering menunjukkan pola bahasa tertentu, misalnya kata-kata yang mengarah pada kelelahan, kecemasan, tekanan akademik, kehilangan harapan, atau ketidakmampuan mengontrol situasi. Model seperti BERT dan RoBERTa juga dapat menangkap konteks kalimat dengan lebih baik dibandingkan metode berbasis kata sederhana, sehingga berguna untuk memahami bahasa Indonesia yang memiliki variasi slang, singkatan, dan bentuk informal [16].

Namun, analisis teks memiliki keterbatasan. Kalimat yang sama dapat memiliki makna berbeda tergantung konteks. Ungkapan seperti “capek banget” dapat berarti kelelahan biasa, stres berat, atau bahkan candaan. Selain itu, pengguna dapat menyensor masalahnya atau menulis dengan gaya ironi. Karena itu, hasil prediksi dari teks perlu diperkuat dengan modalitas lain. Dalam rancangan multimodal, teks berperan sebagai representasi semantik, tetapi tidak menjadi satu-satunya dasar keputusan.

3.4. Analisis Modalitas Suara

Suara memberikan informasi yang tidak selalu muncul dalam teks. Seseorang dapat menulis kalimat pendek dan netral, tetapi nada bicara, tempo, dan energi suaranya menunjukkan tekanan emosional. Penelitian deteksi emosi berdasarkan wicara menunjukkan bahwa fitur seperti Zero Crossing Rate, Chroma, MFCC, RMS, dan MelSpectrogram dapat digunakan untuk mengenali emosi melalui model deep learning [6]. Penelitian lain pada bahasa Indonesia juga menunjukkan bahwa korpus suara, pemilihan fitur, dan penanganan imbalance data menjadi faktor penting dalam peningkatan akurasi pengenalan emosi [7].

Model CNN efektif untuk membaca pola lokal pada spectrogram atau representasi fitur suara, sedangkan LSTM dan BiLSTM membantu menangkap urutan temporal. Pada deteksi emosi berbasis speech bahasa Indonesia, kombinasi fitur MFCC dan frekuensi fundamental dapat memberikan performa yang baik karena kedua fitur tersebut menangkap aspek spektral dan prosodi dari ucapan [8]. Dalam konteks konseling digital, deteksi emosi berbasis suara juga dapat membantu sistem mengenali nuansa emosional pengguna sehingga respons yang diberikan menjadi lebih adaptif [9].

Kelemahannya, data suara sangat sensitif terhadap kualitas rekaman. Kebisingan, jarak mikrofon, jenis perangkat, aksen, logat, serta kondisi fisik pengguna dapat mengubah fitur akustik. Selain itu, ekspresi suara sangat individual. Ada orang yang tetap berbicara datar meskipun sedang stres, sedangkan orang lain tampak emosional meskipun tekanan yang dialami ringan. Karena itu, suara sebaiknya dipakai bersama teks dan sinyal fisiologis agar sistem tidak terlalu bergantung pada satu indikator.

3.5. Analisis Modalitas Sinyal Fisiologis

Sinyal fisiologis menjadi modalitas penting karena stres tidak hanya muncul sebagai pengalaman subjektif, tetapi juga sebagai perubahan tubuh. Saat seseorang mengalami tekanan, sistem saraf otonom dapat memengaruhi detak jantung, variabilitas detak jantung, aktivitas elektrodermal, suhu kulit, dan pola pernapasan. Penelitian klasifikasi tingkat stres mahasiswa menggunakan data EKG menunjukkan bahwa model CNN-LSTM dapat digunakan untuk membaca pola spasial dan temporal dari sinyal fisiologis [10].

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Selain EKG, sinyal PPG dan HRV juga dapat digunakan dalam sistem deteksi stres. Implementasi berbasis sensor Polar Verity Sense menunjukkan bahwa data PPG dapat diproses menggunakan perangkat komputasi seperti Raspberry Pi, sehingga membuka peluang pengembangan sistem edge computing yang lebih ringan dan responsif [11]. Pada konteks wearable healthcare, ensemble learning seperti Random Forest, XGBoost, MLP, dan stacking classifier juga dapat dimanfaatkan untuk meningkatkan stabilitas prediksi dari data sensor [12].

Keunggulan modalitas fisiologis adalah sifatnya yang relatif objektif dan berkelanjutan. Namun, interpretasinya tetap tidak sederhana. Detak jantung yang meningkat dapat berasal dari olahraga, demam, konsumsi kafein, atau aktivitas fisik, bukan hanya stres. Oleh karena itu, sinyal fisiologis memerlukan konteks. Ketika kenaikan detak jantung terjadi bersamaan dengan perubahan suara dan kata-kata yang menunjukkan tekanan, keyakinan sistem terhadap indikasi stres dapat meningkat.

3.6. Strategi Fusi Multimodal

Fusi multimodal merupakan inti dari rancangan sistem karena bertugas menggabungkan informasi dari teks, suara, dan sinyal fisiologis. Terdapat tiga strategi utama yang dapat digunakan, yaitu early fusion, late fusion, dan hybrid fusion. Early fusion menggabungkan fitur dari semua modalitas sebelum proses klasifikasi. Strategi ini memungkinkan model mempelajari hubungan antarmodalitas sejak awal, tetapi membutuhkan sinkronisasi data yang baik. Late fusion membiarkan setiap model unimodal menghasilkan prediksi sendiri, lalu keputusan akhir digabungkan melalui voting, pembobotan, atau meta-classifier. Hybrid fusion menggabungkan kelebihan keduanya.

Dalam deteksi stres, hybrid fusion dinilai paling sesuai karena kondisi data lapangan tidak selalu lengkap. Pengguna mungkin hanya memberikan teks tanpa suara, sensor mungkin tidak aktif, atau rekaman suara memiliki kualitas rendah. Dengan hybrid fusion, sistem tetap dapat menghasilkan prediksi dari modalitas yang tersedia, tetapi akan meningkatkan keyakinan ketika lebih banyak modalitas memberikan sinyal yang konsisten.

Fusi juga perlu memperhatikan bobot adaptif. Tidak semua pengguna mengekspresikan stres dengan cara yang sama. Pada sebagian pengguna, teks sangat informatif karena mereka menuliskan perasaan secara eksplisit. Pada pengguna lain, suara dan fisiologis lebih informatif karena mereka cenderung menekan emosi dalam kata-kata. Oleh karena itu, sistem ideal perlu mempelajari pola individual secara aman dan terbatas, misalnya melalui baseline personal yang dibangun dari data pengguna saat kondisi normal.

Tabel 2. Perbandingan Strategi Fusi Multimodal

Strategi	Kelebihan	Kelemahan	Kesesuaian
Early fusion	Mempelajari hubungan fitur sejak awal.	Membutuhkan data lengkap dan sinkron.	Cocok untuk dataset terkontrol.
Late fusion	Tahan ketika salah satu modalitas lemah.	Hubungan antarmodalitas tidak dipelajari secara mendalam.	Cocok untuk sistem praktis dengan data tidak selalu lengkap.
Hybrid fusion	Menggabungkan kekuatan early dan late fusion.	Lebih kompleks dan membutuhkan tuning.	Paling sesuai untuk rancangan deteksi stres multimodal.

3.7. Rancangan Arsitektur Sistem

Arsitektur sistem yang diusulkan dimulai dari akuisisi data multimodal. Teks diperoleh dari percakapan pengguna dengan sistem, catatan harian digital, atau respons terhadap pertanyaan singkat. Suara diperoleh dari rekaman percakapan atau respons verbal pengguna dengan persetujuan eksplisit. Sinyal fisiologis diperoleh dari wearable seperti smartwatch, sensor PPG, atau perangkat EKG portabel. Seluruh data kemudian diberi timestamp agar dapat diselaraskan dalam jendela waktu yang sama.

Tahap prapemrosesan dilakukan secara spesifik untuk setiap modalitas. Teks dibersihkan dari simbol tidak penting, dinormalisasi, dan diproses menjadi representasi numerik. Suara dibersihkan dari noise, disegmentasi, lalu diekstraksi menjadi fitur akustik. Sinyal fisiologis difilter dari artefak, diresampling, dan diekstraksi menjadi fitur statistik maupun temporal. Setelah itu, masing-masing modalitas diproses oleh model unimodal. Representasi dari model unimodal kemudian masuk ke lapisan fusi untuk menghasilkan skor risiko stres.

Output sistem tidak seharusnya berupa label kaku seperti “stres” atau “tidak stres” saja, tetapi berupa tingkat risiko, misalnya rendah, sedang, dan tinggi. Output juga perlu disertai penjelasan sederhana, misalnya “terdapat peningkatan indikator fisiologis dan pola bahasa yang mengarah pada tekanan.” Penjelasan semacam ini penting

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

agar pengguna tidak merasa divonis oleh sistem, melainkan memahami bahwa hasil tersebut adalah sinyal awal yang perlu ditindaklanjuti secara bijak.

Tabel 3. Rancangan Alur Sistem Deteksi Dini Stres

Tahap	Rancangan Teknis
Akuisisi data	Mengumpulkan teks, suara, dan sinyal fisiologis dengan persetujuan pengguna dan pencatatan waktu.
Prapemrosesan	Membersihkan teks, mengurangi noise audio, memfilter artefak sensor, dan menyelaraskan waktu antar modalitas.
Ekstraksi fitur	Mengubah data menjadi fitur seperti embedding teks, MFCC, pitch, energi, HRV, PPG, EKG, dan fitur temporal.
Pemodelan unimodal	Membangun model terpisah untuk teks, suara, dan fisiologis agar kontribusi tiap modalitas dapat dianalisis.
Fusi multimodal	Menggabungkan fitur atau keputusan model melalui early fusion, late fusion, atau hybrid fusion.
Evaluasi	Mengukur akurasi, presisi, recall, F1-score, AUC-ROC, confusion matrix, uji bias, dan uji keamanan respons.
Tindak lanjut	Memberi umpan balik non-diagnostik, edukasi singkat, rekomendasi self-care umum, dan rujukan profesional jika risiko tinggi.

3.8. Integrasi dengan Chatbot Kesehatan Mental

Integrasi AI multimodal dengan chatbot kesehatan mental memberikan dua fungsi utama. Fungsi pertama adalah skrining awal, yaitu membantu mengenali tanda stres dari percakapan, suara, dan sinyal fisiologis. Fungsi kedua adalah dukungan awal, yaitu memberikan respons edukatif dan empatik. Penelitian mengenai chatbot berbasis LLM dengan deteksi emosi dan RAG menunjukkan bahwa respons sistem dapat dibuat lebih relevan apabila model memiliki konteks emosi pengguna dan sumber pengetahuan yang terkurasi [13].

Pengembangan chatbot berbasis GPT-3.5 melalui WhatsApp menunjukkan bahwa media percakapan populer dapat menjadi jalur dukungan kesehatan mental yang mudah dijangkau [14]. Namun, sistem seperti ini tetap harus dibatasi. Chatbot dapat membantu pengguna mengekspresikan emosi, memperoleh informasi awal, dan melakukan refleksi sederhana. Akan tetapi, chatbot tidak boleh mengambil keputusan klinis, memberi diagnosis, atau menggantikan konselor manusia.

Kajian perbandingan chatbot AI dan layanan konselor menegaskan bahwa chatbot efektif sebagai pendukung awal karena mudah diakses, anonim, dan responsif, tetapi belum mampu menggantikan penilaian profesional dalam kondisi kompleks atau berisiko tinggi [15]. Oleh sebab itu, rancangan sistem harus memiliki safety guardrails. Ketika model mendeteksi risiko tinggi seperti ide menyakiti diri, krisis berat, atau permintaan diagnosis medis, sistem harus memberikan respons aman dan menyarankan rujukan kepada tenaga profesional atau layanan darurat yang sesuai.

3.9. Keamanan, Privasi, dan Penerimaan Pengguna

Sistem deteksi stres berbasis AI memproses data yang sangat sensitif. Teks percakapan dapat berisi masalah pribadi, suara dapat mengandung identitas biometrik, dan sinyal fisiologis dapat menggambarkan kondisi tubuh. Karena itu, prinsip privasi harus diterapkan sejak tahap perancangan. Data perlu diminimalkan, dienkripsi, dianonimkan, dan hanya digunakan sesuai persetujuan pengguna. Pengguna juga harus mengetahui data apa yang dikumpulkan, untuk tujuan apa, berapa lama disimpan, dan bagaimana cara menghapusnya.

Penerimaan pengguna terhadap AI juga perlu diperhatikan. Analisis sentimen terhadap kemajuan kecerdasan buatan di Indonesia menunjukkan adanya kekhawatiran masyarakat terhadap dampak dan batas penggunaan AI [16]. Dalam konteks kesehatan mental, kekhawatiran ini dapat menjadi lebih kuat karena pengguna berhadapan dengan isu pribadi dan rentan. Oleh karena itu, sistem harus transparan, tidak mengklaim sebagai pengganti ahli, serta menyediakan opsi untuk menonaktifkan fitur tertentu.

Keamanan model juga perlu diuji. Jika sistem dihubungkan dengan chatbot, maka evaluasi tidak cukup hanya menggunakan akurasi klasifikasi. Diperlukan safety stress-test untuk memastikan chatbot tidak memberikan saran berbahaya, tidak membuat diagnosis palsu, tidak memperkuat pikiran negatif pengguna, dan tidak mengabaikan tanda krisis. Model perlu dilengkapi mekanisme eskalasi, misalnya pesan rujukan ke profesional ketika risiko tinggi terdeteksi.

3.10. Implikasi Teoritis dan Praktis

Secara teoritis, rancangan AI multimodal memperkuat pemahaman bahwa stres merupakan fenomena multidimensi. Stres tidak hanya dapat dibaca dari satu jenis data, tetapi muncul melalui interaksi antara pikiran,

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

emosi, perilaku, dan respons fisiologis. Machine learning memberikan kerangka untuk mempelajari hubungan kompleks tersebut, terutama ketika data memiliki pola non-linear dan jumlah variabel yang besar [17].

Secara praktis, sistem ini dapat dikembangkan sebagai fitur tambahan dalam aplikasi konseling digital, sistem informasi kampus, wearable kesehatan, atau chatbot pendamping. Pada lingkungan kampus, misalnya, sistem dapat membantu skrining awal mahasiswa yang sedang menghadapi tekanan akademik. Pada layanan digital, sistem dapat membantu chatbot menyesuaikan respons berdasarkan tingkat risiko. Pada wearable, sistem dapat memberi peringatan dini ketika pola fisiologis menunjukkan tekanan yang tidak biasa.

Namun, penerapan praktis harus dilakukan bertahap. Tahap awal dapat berupa uji coba offline menggunakan dataset terkontrol. Tahap berikutnya adalah pengujian terbatas dengan partisipan yang memberikan persetujuan. Setelah itu, sistem perlu divalidasi oleh ahli psikologi, diuji biasnya, dievaluasi keamanan responsnya, dan disesuaikan dengan konteks budaya pengguna. Sistem yang langsung diterapkan tanpa validasi dapat menimbulkan risiko salah deteksi, over-alert, atau rasa aman palsu.

3.11. Keterbatasan Rancangan

Keterbatasan utama artikel ini adalah belum adanya eksperimen langsung terhadap dataset multimodal baru. Karena itu, hasil yang disajikan berupa sintesis literatur dan rancangan konseptual, bukan klaim akurasi akhir. Artikel ini juga menggunakan sumber yang memiliki konteks berbeda, seperti media sosial, suara berbahasa Indonesia, EKG mahasiswa, PPG, wearable pekerja kesehatan, dan chatbot. Perbedaan konteks tersebut menunjukkan bahwa rancangan multimodal perlu diuji kembali pada populasi dan skenario penggunaan yang spesifik.

Keterbatasan lain berkaitan dengan ketersediaan dataset multimodal berbahasa Indonesia. Penelitian suara dan teks Indonesia masih menghadapi tantangan variasi bahasa, logat, slang, serta keterbatasan korpus. Pada sinyal fisiologis, tantangan terletak pada kualitas sensor, noise, dan interpretasi sinyal yang dipengaruhi aktivitas fisik. Oleh karena itu, penelitian lanjutan perlu mengembangkan dataset multimodal yang representatif, beragam secara demografis, dan memiliki label psikologis yang valid.

4. Kesimpulan

AI multimodal memiliki potensi kuat untuk mendukung deteksi dini stres karena dapat menggabungkan tiga sumber informasi utama, yaitu teks, suara, dan sinyal fisiologis. Teks membantu membaca isi pikiran dan sentimen, suara menangkap ekspresi emosional melalui fitur akustik, sedangkan sinyal fisiologis menggambarkan respons tubuh yang relatif objektif. Integrasi ketiga modalitas tersebut dapat meningkatkan ketahanan sistem dibandingkan pendekatan unimodal. Rancangan sistem yang diusulkan terdiri atas akuisisi data, prapemrosesan, ekstraksi fitur, pemodelan unimodal, fusi multimodal, evaluasi, dan tindak lanjut non-diagnostik. Strategi hybrid fusion dinilai paling sesuai karena mampu memanfaatkan hubungan antarmodalitas sekaligus tetap fleksibel ketika salah satu data tidak tersedia. Integrasi dengan chatbot kesehatan mental dapat memperluas akses dukungan awal, tetapi tetap memerlukan batasan yang jelas. Sistem deteksi stres berbasis AI tidak boleh diposisikan sebagai pengganti diagnosis psikologis atau layanan konselor manusia. Sistem lebih tepat digunakan sebagai alat skrining awal dan pendukung keputusan. Validasi klinis, keamanan respons, perlindungan privasi, transparansi model, dan mekanisme rujukan profesional menjadi syarat penting sebelum sistem diterapkan secara luas.

Referensi

[1] A. N. Haryanti, M. B. S. Putra, N. Larasati, V. N. Khairunnisa, and L. D. Dewi A., “Analisis Kondisi Kesehatan Mental di Indonesia dan Strategi Penanganannya,” Student Research Journal, vol. 2, no. 3, pp. 28-40, Jun. 2024, doi: 10.55606/srjyappi.v2i3.1219.

[2] D. Suratmini and T. Afriani, “Pemanfaatan Deteksi Stres Remaja yang Efektif dan Efisien Melalui Aplikasi: Systematic Review,” Jurnal Keperawatan Respati Yogyakarta, vol. 6, no. 1, pp. 542-547, Jan. 2019.

[3] M. Junenda, T. W. Widyaningsih, and A. Rifqi, “Pengembangan Sistem Informasi Kesehatan Mental dengan Mengukur Persepsi Tingkat Stres Menggunakan Instrumen Perceived Stress Scale (PSS),” Jurnal Sistem Komputer dan Kecerdasan Buatan, vol. VIII, no. 1, pp. 34-42, Sep. 2024.

[4] Y. Tolla and Kusrini, “Deteksi Stres dan Depresi Unggahan Media Sosial dengan Machine Learning,” Jurnal FASILKOM, vol. 15, no. 1, pp. 84-92, Apr. 2025.

[5] D. F. M. Dina, T. Haryanti, and M. A. Haq, “Deteksi Gangguan Depresi Menggunakan Machine Learning: Tinjauan Pustaka,” in SNESTIK IV, Surabaya, Indonesia, 2024, doi: 10.31284/p.snestik.2024.5876.

[6] S. Rahmadani, C. S. Rahayu, A. Salim, and K. N. Cahyo, “Deteksi Emosi Berdasarkan Wicara Menggunakan Deep Learning Model,” JINTEKS (Jurnal Informatika Teknologi dan Sains), vol. 4, no. 3, pp. 220-224, Aug. 2022.

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

[7] F. Kasyidi, R. Ilyas, and N. M. Annisa, "Peningkatan Kemampuan Pengenalan Emosi melalui Suara dalam Bahasa Indonesia," MIND Journal, vol. 6, no. 2, pp. 194-204, Dec. 2021, doi: 10.26760/mindjournal.v6i2.194-204.

[8] Y. K. Aini, T. B. Santoso, and T. Dutono, "Pemodelan CNN untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia," Jurnal Komputer Terapan, vol. 7, no. 1, pp. 143-152, May 2021.

[9] A. P. Wibowo and G. Prayitno, "Pengembangan Algoritma Deteksi Emosi Melalui Analisis Suara untuk Aplikasi Konseling Digital," Journal of New Trends in Sciences, vol. 3, no. 2, pp. 36-47, May 2025, doi: 10.59031/jnts.v3i2.792.

[10] A. M. Husein, L. H. Muntaza, G. Sinaga, Khoirulliza, and M. L. Br Sembiring, "Model CNN-LSTM untuk Klasifikasi Tingkat Stres Mahasiswa dalam Menghadapi Ujian Menggunakan Data Elektrokardiogram (EKG)," DECODE: Jurnal Pendidikan Teknologi Informasi, vol. 5, no. 1, pp. 191-201, Mar. 2025, doi: 10.51454/decode.v5i1.1080.

[11] D. A. Firmanda, E. S. Pramukantoro, and B. H. Prasetio, "Implementasi Sistem Deteksi Stres Berbasis Machine Learning berdasarkan Sinyal Photoplethysmogram dari Polar Verity Sense," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 1, no. 1, pp. 1-8, Jan. 2017.

[12] A. Sinhal, A. Sinhal, and A. Sinhal, "Stress Monitoring in Healthcare: An Ensemble Machine Learning Framework Using Wearable Sensor Data," arXiv:2507.07589, 2025.

[13] D. Tawakalna and J. Zeniarja, "Sistem Chatbot Kesehatan Mental Berbasis LLM dengan Deteksi Emosi dan Retrieval Augmented Generation," RABIT: Jurnal Teknologi dan Sistem Informasi Univrab, vol. 11, no. 1, pp. 1398-1412, Jan. 2026, doi: 10.36341/rabit.v11i1.7347.

[14] I. A. Assegaf et al., "Pengembangan Chatbot Konsultasi Kesehatan Mental Berbasis OpenAI Model GPT-3.5 Turbo Menggunakan Media WhatsApp," JINTEKS (Jurnal Informatika Teknologi dan Sains), vol. 6, no. 4, pp. 785-793, Nov. 2024.

[15] N. A. Sutrisno, S. R. Sofiana, and N. Nuurfathina, "Efektivitas Chatbot AI Dibandingkan Layanan Konselor Manusia dalam Dukungan Kesehatan Mental," JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 14, no. 1, pp. 1427-1438, 2026, doi: 10.23960/jitet.v14i1.8587.

[16] N. A. R. Putri and Ardiansyah, "Analisis Sentimen terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," Jurnal Sains dan Informatika, vol. 9, no. 2, pp. 136-145, Nov. 2023, doi: 10.34128/jsi.v9i2.649.

[17] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," IJCIT (Indonesian Journal on Computer and Information Technology), vol. 5, no. 1, pp. 75-82, May 2020.

DOI: https://doi.org/10.31004/riggs.v5i2.10668
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)