

Kecerdasan Buatan (*Artificial Intelligence*) untuk Deteksi Dini Gangguan Kesehatan Mental: *Systematic Review*

Cokorda Tesya Kirana¹

¹ Fakultas Kesehatan dan Sains, Universitas Dhyana Pura, Bali, Indonesia

Article Info	ABSTRAK
Article history: Received Juli 24, 2025 Revised Juli 24, 2025 Accepted Juli 25, 2025 Kata Kunci: Deteksi Dini Gangguan Mental, Kesehatan Mental, Kecerdasan Buatan, Systematic Review, Skrining Psikologi Keywords: <i>Early Detection of Mental Disorders, Mental Health, Artificial Intelligence, Systematic Review, Psychological Screening</i>	Deteksi dini gangguan kesehatan mental menjadi krusial karena keterlambatan penanganan gejala awal depresi, kecemasan, dan risiko bunuh diri berkontribusi terhadap perburukan kondisi klinis serta meningkatnya beban psikososial pada individu maupun sistem layanan kesehatan. Berbagai tinjauan sistematis telah mengevaluasi penerapan kecerdasan buatan (AI) untuk mendeteksi gangguan mental pada modalitas data yang berbeda secara terpisah, namun sintesis lintas modalitas yang relevan bagi praktik psikologi masih jarang tersedia. Kajian ini bertujuan mensintesis bukti dari tinjauan sistematis dan meta-analisis yang diterbitkan pada rentang 2021-2026 mengenai penerapan AI untuk deteksi dini gangguan mental pada lima modalitas data, yaitu teks dan media sosial, suara, sinyal elektroensefalogram, perangkat wearable, dan prediksi risiko bunuh diri. Sintesis dilakukan melalui penelusuran pustaka pada mesin pencari akademik, penyaringan bertahap berdasarkan kriteria kelayakan, dan sintesis naratif atas dua puluh tinjauan yang memenuhi syarat. Hasil sintesis menunjukkan bahwa model pembelajaran mesin dan pembelajaran mendalam secara konsisten mencapai akurasi klasifikasi tinggi pada seluruh modalitas, dengan performa tertinggi ditemukan pada model prediksi risiko bunuh diri dan analisis suara otomatis, sementara heterogenitas metodologis dan minimnya validasi klinis prospektif membatasi kesiapan penerapannya. Integrasi AI ke dalam praktik skrining psikologi memerlukan standarisasi protokol, validasi lintas populasi, dan mekanisme interpretabilitas model yang memadai. ABSTRACT <i>Early detection of mental health disorders has become critical, as delayed management of early symptoms of depression, anxiety, and suicide risk contributes to clinical deterioration and an increasing psychosocial burden on both individuals and healthcare systems. Various systematic reviews have evaluated the application of artificial intelligence (AI) for detecting mental disorders across different data modalities separately, yet cross-modality syntheses relevant to psychological practice remain scarce. This study aims to synthesize evidence from systematic reviews and meta-analyses published between 2021 and 2026 on the application of AI for early detection of mental disorders across five data modalities: text and social media, voice, electroencephalogram (EEG) signals, wearable devices, and suicide risk prediction. The synthesis was conducted through literature searches on academic search engines, staged screening based on eligibility criteria, and narrative synthesis of twenty eligible reviews. The synthesis results show that machine learning and deep learning models consistently achieved high classification accuracy across all modalities, with the highest performance found in suicide risk prediction models and automatic speech analysis, while methodological heterogeneity and the scarcity of prospective clinical validation limit their readiness for implementation. Integrating AI into psychological screening</i>

practice requires protocol standardization, cross-population validation, and adequate model interpretability mechanisms.

This is an open access article under the [CC BY](#) license



Corresponding Author:

Cokorda Tesya Kirana
Fakultas Kesehatan dan Sains, Universitas Dhyana Pura
Bali, Indonesia
Email: tesyakirana@gmail.com

1. PENDAHULUAN

Keterlambatan deteksi gejala awal gangguan kesehatan mental, khususnya depresi, gangguan cemas, dan perilaku bunuh diri, kerap menyebabkan penundaan intervensi klinis hingga kondisi individu memburuk secara signifikan. Instrumen asesmen konvensional yang mengandalkan wawancara klinis dan skala self-report bersifat episodik dan sangat bergantung pada kesediaan individu untuk mengungkapkan gejalanya secara aktif, sehingga sebagian besar kasus baru terdeteksi setelah gejala berkembang lebih parah dan intervensi dini menjadi lebih sulit dilakukan [1].

Perkembangan kecerdasan buatan (AI) dalam kurun waktu satu dekade terakhir mendorong lahirnya berbagai pendekatan deteksi gangguan mental yang lebih objektif dan bersifat pasif. Cruz-Gonzalez dkk. [2] mensintesis 85 studi yang menunjukkan bahwa algoritma support vector machine dan random forest banyak digunakan untuk tujuan diagnosis, sementara model pembelajaran mesin lain lebih banyak diterapkan untuk pemantauan gejala secara berkelanjutan. Pada modalitas teks, Liu dkk. [1] menjadi salah satu tinjauan awal yang mengonfirmasi bahwa analisis bahasa alami pada unggahan media sosial mampu mengidentifikasi indikator depresi secara akurat, arah riset yang kemudian berkembang pesat seiring munculnya arsitektur transformer dan model bahasa besar pada chatbot kesehatan mental [3]. Pada modalitas biosinyal dan perilaku pasif, Abd-Alrazaq dkk. [4] melakukan meta-analisis performa AI berbasis perangkat wearable dalam mendeteksi dan memprediksi depresi, sementara sejumlah tinjauan lain secara terpisah mengevaluasi biomarker suara dan sinyal elektroensefalogram sebagai indikator gangguan mental.

Dari sisi metodologis, tinjauan-tinjauan tersebut juga menunjukkan variasi pendekatan sintesis, mulai dari tinjauan naratif deskriptif hingga meta-analisis kuantitatif yang menghitung ukuran performa gabungan menggunakan alat penilaian risiko bias standar seperti QUADAS-2 dan PROBAST. Variasi pendekatan ini mencerminkan bahwa kematangan metodologis riset AI untuk kesehatan mental berbeda-beda antarmodalitas, dengan modalitas yang lebih lama diteliti seperti teks media sosial cenderung memiliki jumlah tinjauan sistematis yang lebih banyak dibandingkan modalitas yang relatif baru seperti digital phenotyping berbasis wearable.

Selain perbedaan kematangan metodologis antarmodalitas, populasi yang menjadi sasaran deteksi pada tinjauan-tinjauan tersebut juga bervariasi, mulai dari populasi umum pengguna media sosial, populasi klinis rawat jalan psikiatri, hingga populasi khusus seperti remaja dan lanjut usia. Variasi populasi sasaran ini turut memengaruhi generalisabilitas temuan, mengingat pola linguistik, akustik, maupun neurofisiologis yang berasosiasi dengan gejala gangguan mental dapat berbeda secara bermakna antarkelompok usia dan budaya, sehingga sintesis lintas modalitas pada kajian ini perlu memperhatikan konteks populasi asal setiap tinjauan yang disertakan.

Meskipun literatur pada masing-masing modalitas tersebut berkembang cepat, tinjauan-tinjauan yang tersedia umumnya bersifat spesifik-modalitas dan diterbitkan secara independen satu sama lain, sehingga belum tersedia sintesis yang memetakan capaian dan keterbatasan deteksi dini gangguan

mental berbasis AI secara lintas modalitas dalam kerangka yang relevan bagi praktik psikologi klinis. Ketiadaan pemetaan lintas modalitas ini menyulitkan praktisi dan peneliti psikologi dalam menilai modalitas data mana yang secara metodologis paling matang untuk diadopsi, serta menghambat perumusan pedoman penerapan AI yang konsisten pada layanan skrining kesehatan mental. Kajian ini disusun untuk mengisi kesenjangan tersebut, dengan mensintesis bukti tinjauan sistematis dan meta-analisis dari lima modalitas data utama yang diterbitkan pada rentang tahun 2021-2026, guna memetakan capaian performa, tantangan metodologis, dan implikasi praktisnya bagi psikologi klinis.

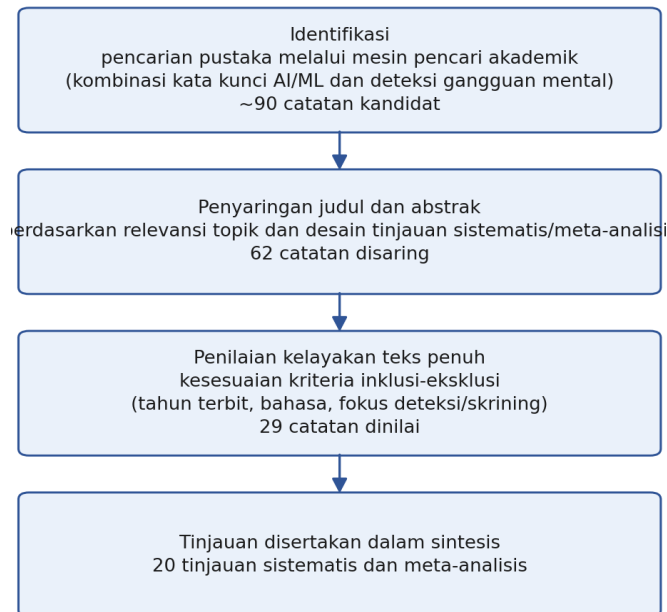
2. METODE

Kajian ini menggunakan desain sintesis naratif cepat (rapid narrative synthesis) yang menelaah tinjauan sistematis dan meta-analisis yang telah dipublikasikan, sebanding dengan pendekatan tinjauan-atas-tinjauan (review of reviews/umbrella review) yang telah diterapkan pada literatur AI kesehatan sebelumnya [5]. Rumusan pertanyaan kajian disusun mengikuti kerangka population-intervention-comparison-outcome-study design (PICOS) yang diadaptasi untuk konteks tinjauan-atas-tinjauan, dengan populasi berupa individu yang menjadi target deteksi atau skrining gangguan mental, indeks uji berupa model AI atau pembelajaran mesin, pembanding yang tidak dibatasi secara ketat, luaran berupa ukuran performa deteksi seperti akurasi, sensitivitas, spesifisitas, dan area under the curve, serta desain studi berupa tinjauan sistematis atau tinjauan sistematis dengan meta-analisis.

Penelusuran pustaka dilakukan pada bulan Juli 2026 menggunakan mesin pencari akademik dan basis data ilmiah daring, dengan kombinasi kata kunci yang disusun dari tiga kelompok istilah, yaitu istilah kecerdasan buatan (artificial intelligence, machine learning, deep learning), istilah gangguan kesehatan mental (depression, anxiety, suicide risk, mental disorder), dan istilah deteksi (detection, screening, diagnosis, prediction, biomarker), yang dikombinasikan menggunakan operator boolean AND dan OR secara iteratif untuk setiap modalitas data yang menjadi fokus kajian.

Tinjauan dinyatakan memenuhi syarat inklusi apabila merupakan tinjauan sistematis atau tinjauan sistematis dengan meta-analisis yang diterbitkan dalam bahasa Inggris pada rentang tahun 2021 hingga 2026, berfokus pada penerapan AI atau pembelajaran mesin untuk deteksi, skrining, atau prediksi risiko gangguan kesehatan mental pada manusia, dan tersedia dalam bentuk artikel jurnal maupun cetak yang dapat ditelaah ringkasannya secara daring. Tinjauan yang berfokus murni pada intervensi atau terapi tanpa komponen deteksi, studi primer tunggal yang bukan merupakan tinjauan, artikel opini tanpa metodologi eksplisit, serta publikasi non-ilmiah dikeluarkan dari sintesis ini.

Proses seleksi pustaka mengikuti empat tahap berurutan, yaitu identifikasi catatan kandidat melalui penelusuran kombinasi kata kunci, penyaringan judul dan abstrak berdasarkan relevansi topik dan kesesuaian desain tinjauan, penilaian kelayakan teks penuh berdasarkan kriteria inklusi-eksklusi yang telah ditetapkan, dan inklusi akhir tinjauan yang memenuhi seluruh kriteria kelayakan. Alur keempat tahap tersebut diilustrasikan pada Gambar 1.



Gambar 1. Alur seleksi pustaka pada sintesis tinjauan sistematis

Pemilihan desain sintesis naratif cepat, alih-alih penelusuran sistematis multi-basis data dengan protokol PRISMA penuh, didasarkan pada pertimbangan bahwa literatur AI untuk deteksi gangguan mental telah cukup banyak disintesis pada tingkat modalitas oleh tinjauan-tinjauan sistematis primer yang tersedia, sehingga kontribusi utama kajian ini terletak pada sintesis lintas modalitas, bukan pada penelusuran ulang studi primer yang sudah tercakup dalam tinjauan-tinjauan tersebut. Pendekatan serupa telah digunakan pada kajian AI kesehatan lain yang bertujuan memetakan lanskap penelitian secara cepat sebelum dilakukan tinjauan sistematis penuh dengan protokol terdaftar [5].

Penilaian kualitas metodologis tinjauan yang disertakan tidak dilakukan secara independen menggunakan instrumen penilaian kualitas tinjauan sistematis seperti AMSTAR-2, melainkan mengandalkan informasi mengenai alat penilaian risiko bias yang telah digunakan oleh masing-masing tinjauan asli, seperti QUADAS-2 untuk studi akurasi diagnostik dan PROBAST untuk studi model prediksi. Keterbatasan ini konsisten dengan sifat kajian sebagai sintesis cepat dan diuraikan lebih lanjut pada bagian keterbatasan.

Data yang diekstraksi dari setiap tinjauan yang disertakan meliputi modalitas data utama, jenis algoritma AI yang paling banyak dilaporkan, ukuran performa yang digunakan, jumlah studi primer yang disintesis pada tinjauan tersebut, dan keterbatasan metodologis yang teridentifikasi oleh penulis tinjauan asli. Mengingat tingginya heterogenitas desain, populasi, dan ukuran performa antartinjauan yang disertakan, sintesis dilakukan secara naratif dan dikelompokkan berdasarkan modalitas data, tanpa penggabungan meta-analitik lintas tinjauan.

Selain pengelompokan berdasarkan modalitas, sintesis pada kajian ini juga menerapkan tiga lapis analisis pada setiap kelompok modalitas, yaitu identifikasi temuan utama yang dilaporkan (unsur what), penjelasan temuan tersebut melalui kerangka teoretis yang relevan dari bidang psikologi dan neurosains (unsur why), serta perbandingan temuan antartinjauan pada modalitas yang sama maupun antarmodalitas yang berbeda untuk menilai konsistensi atau pertentangan hasil (unsur what else). Ketiga lapis analisis ini diterapkan secara konsisten pada setiap subbagian hasil dan pembahasan guna memastikan kedalaman

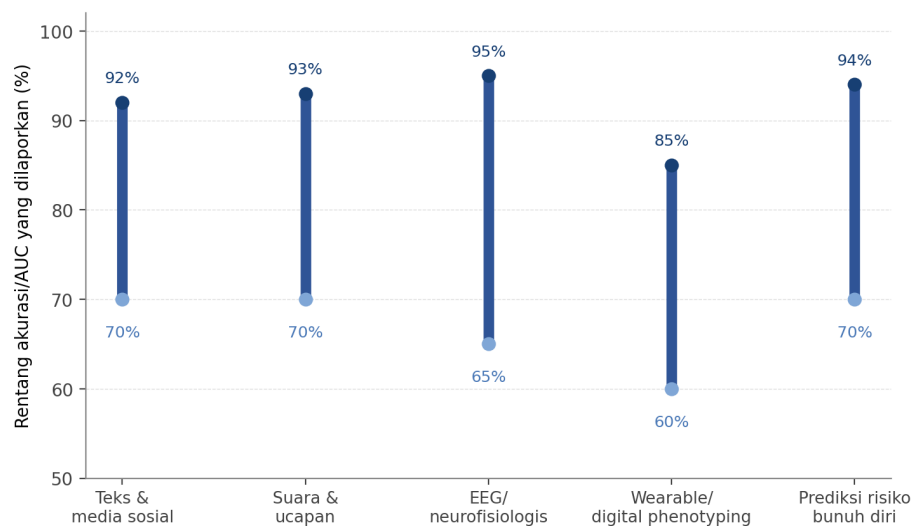
3. HASIL DAN PEMBAHASAN

Sebanyak dua puluh tinjauan sistematis dan meta-analisis yang relevan disertakan dalam sintesis ini, mencakup lima modalitas data utama, yaitu teks dan media sosial, suara dan ucapan, sinyal

elektroensefalogram (EEG), perangkat wearable/digital phenotyping, serta prediksi risiko bunuh diri. Rangkuman karakteristik tinjauan yang disertakan, meliputi jumlah tinjauan, metode AI yang paling banyak dilaporkan, dan rentang performa yang dicapai pada masing-masing modalitas, disajikan pada Tabel 1, sedangkan perbandingan visual rentang akurasi atau area under the curve (AUC) yang dilaporkan antarmodalitas disajikan pada Gambar 2.

Tabel 1. Ringkasan karakteristik tinjauan sistematis yang disertakan berdasarkan modalitas data

Modalitas Data	Jumlah Tinjauan	Metode AI Utama	Rentang Performa Dilaporkan	Referensi
Teks dan media sosial	4	Support vector machine, random forest, model transformer berbasis BERT	Akurasi klasifikasi umumnya >80% pada data berlabel mandiri	[1], [6], [7], [9]
Suara dan ucapan	3	Analisis fitur akustik-prosodik dengan pembelajaran mesin	Akurasi 70-93%, spesifisitas 73-93%	[10], [11], [12]
Sinyal EEG/neurofisiologis	3	Deep learning pada sinyal EEG dengan pendekatan explainable AI	Akurasi bervariasi tinggi; interpretabilitas menjadi kendala utama	[13], [14], [15]
Wearable/digital phenotyping	1	Analisis pola aktivitas, tidur, dan denyut jantung	Performa deteksi bervariasi menurut parameter fisiologis	[4]
Prediksi risiko bunuh diri	5	Random forest, regresi logistik, jaringan saraf tiruan	Akurasi/AUC berkisar 0,70-0,94	[16], [17], [18], [19], [20]



Gambar 2. Perbandingan rentang akurasi/AUC yang dilaporkan antarmodalitas data

Sebagaimana terlihat pada Gambar 2, kelima modalitas menunjukkan rentang performa yang saling tumpang tindih pada kisaran 60-95%, dengan modalitas prediksi risiko bunuh diri dan EEG menunjukkan batas atas performa yang relatif lebih tinggi, sementara modalitas wearable/digital phenotyping menunjukkan rentang yang sedikit lebih rendah dan lebih sempit, konsisten dengan jumlah tinjauan yang masih terbatas pada modalitas tersebut.

3.1 Deteksi Berbasis Teks dan Media Sosial

Deteksi depresi melalui analisis teks media sosial merupakan salah satu arah riset paling matang pada literatur AI kesehatan mental. Liu dkk. [1] meninjau 17 studi yang menerapkan pembelajaran mesin pada data teks media sosial dan menyimpulkan bahwa pendekatan ini efektif mengidentifikasi gejala depresi serta berpotensi menjadi alat pelengkap dalam praktik kesehatan mental masyarakat. Tinjauan yang lebih baru turut menyoroti peningkatan penggunaan model deep learning berbasis transformer yang menunjukkan akurasi klasifikasi tinggi dalam mengenali pola linguistik terkait gejala depresi dari unggahan pengguna [6], [7].

Tingginya akurasi klasifikasi pada modalitas teks dapat dijelaskan melalui perspektif linguistik-kognitif, yang menyatakan bahwa perubahan afektif dan kognitif pada individu dengan gejala depresi tercermin pada pola bahasa yang digunakan, seperti penurunan kompleksitas leksikal dan peningkatan sentimen negatif, sehingga model pembelajaran bahasa dapat menangkap pola tersebut secara sistematis dari data tekstual berskala besar [8].

Meski demikian, ketika dibandingkan dengan modalitas biosinyal seperti EEG atau data wearable, deteksi berbasis teks memiliki keterbatasan validitas yang lebih besar karena label gangguan mental pada sebagian besar studi bersifat inferensial berdasarkan konten unggahan, bukan diagnosis klinis formal, sehingga akurasi tinggi yang dilaporkan belum tentu mencerminkan validitas diagnostik yang setara dengan modalitas lain yang divalidasi terhadap kriteria klinis baku [9]. Perbedaan sumber label ini menjadi salah satu penjelasan mengapa perbandingan performa lintas modalitas perlu dilakukan secara hati-hati, bukan semata berdasarkan angka akurasi.

Implikasi lain dari ketergantungan pada label inferensial adalah risiko bias seleksi pada tingkat populasi, mengingat pengguna media sosial yang aktif mengunggah konten emosional secara terbuka belum tentu merepresentasikan populasi umum dengan gangguan mental, termasuk individu yang justru menutup diri akibat stigma. Tinjauan-tinjauan pada modalitas ini secara umum belum banyak melaporkan validasi ulang temuannya pada sampel klinis yang terdiagnosis secara formal, sehingga klaim akurasi tinggi pada modalitas teks perlu dimaknai sebagai indikator kelayakan teknis, bukan sebagai bukti kesiapan klinis yang setara dengan instrumen diagnostik baku.

3.2 Biomarker Suara dan Ucapan

Analisis suara berbasis AI dipandang sebagai biomarker digital yang menjanjikan karena sifatnya yang non-invasif, murah, dan dapat dikumpulkan dari jarak jauh menggunakan perangkat sehari-hari seperti telepon pintar. Briganti dan Lechien [10] mensintesis 12 studi yang meneliti fitur akustik dan prosodik suara pada lebih dari 16.000 partisipan, dan menemukan bahwa fitur suara yang dianalisis dengan pembelajaran mesin dapat memberikan informasi bermakna mengenai tingkat keparahan dan progresi gejala depresi. Sejalan dengan itu, Maran dkk. [11] melakukan tinjauan sistematis dan meta-analisis terhadap studi analisis suara otomatis dan melaporkan akurasi klasifikasi berkisar 70-93% dengan spesifisitas serupa.

Performa tinggi pada modalitas suara sejalan dengan teori psikomotor pada depresi, yang menyatakan bahwa perlambatan motorik dan penurunan energi vokal pada individu dengan gejala depresi tercermin pada perubahan nada dasar, kecepatan bicara, dan intonasi, sehingga fitur akustik tersebut dapat dijadikan penanda objektif yang relatif stabil lintas sesi perekaman [12].

Dibandingkan dengan modalitas teks, biomarker suara menunjukkan keunggulan karena tidak bergantung pada konten yang diungkapkan secara eksplisit oleh individu, melainkan pada karakteristik produksi ucapan itu sendiri, sehingga berpotensi menangkap gejala yang tidak selalu disadari atau dilaporkan secara verbal oleh individu. Namun demikian, ukuran sampel pada sebagian besar tinjauan suara masih relatif kecil dibandingkan tinjauan berbasis teks, sehingga generalisasi lintas populasi dan bahasa masih memerlukan kehati-hatian [12].

Perbedaan bahasa dan aksen turut menjadi tantangan tersendiri pada modalitas suara, mengingat sebagian besar model akustik dilatih pada korpus berbahasa Inggris, sehingga transferabilitas model ke bahasa dengan struktur prosodik berbeda, termasuk bahasa Indonesia, belum banyak diuji pada tinjauan-tinjauan yang disertakan. Kondisi ini memperkuat perlunya replikasi studi pada populasi berbahasa non-Inggris sebelum biomarker suara dapat diadopsi secara luas dalam layanan skrining psikologi di Indonesia.

3.3 Biomarker Neurofisiologis Berbasis EEG

Elektroensefalografi (EEG) yang dipadukan dengan model pembelajaran mesin dan mendalam juga banyak diteliti sebagai pendekatan objektif untuk mengidentifikasi biomarker neurofisiologis depresi, kecemasan, dan gangguan bipolar. Sebuah tinjauan sistematis terhadap 145 studi yang diterbitkan antara 2015 hingga 2025 menyoroti berkembangnya penggunaan explainable artificial intelligence (XAI) untuk meningkatkan interpretabilitas model klasifikasi berbasis EEG [13]. Tinjauan lain terhadap model deteksi depresi berbasis EEG turut menekankan tingginya variasi konfigurasi elektrode, durasi perekaman, dan strategi validasi antarstudi [14].

Performa klasifikasi EEG secara teoretis dapat dijelaskan melalui hipotesis asimetri frontal, yang menyatakan bahwa individu dengan gejala depresi cenderung menunjukkan pola aktivitas elektrokortikal yang tidak seimbang antara hemisfer kiri dan kanan pada area frontal, sehingga fitur asimetri tersebut menjadi salah satu penanda neurofisiologis yang paling sering diekstraksi oleh model klasifikasi [15].

Dibandingkan dengan modalitas teks dan suara, EEG menawarkan keunggulan berupa keterkaitan langsung dengan mekanisme neurofisiologis yang mendasari gangguan mental, namun kebutuhan perangkat perekaman khusus membuat modalitas ini kurang praktis untuk skrining berskala besar dibandingkan teks atau suara yang dapat dikumpulkan melalui perangkat konsumen. Isu interpretabilitas model deep learning pada EEG juga lebih menonjol dibandingkan modalitas lain, sehingga pengembangan XAI menjadi arah riset yang secara konsisten disebutkan pada tinjauan-tinjauan modalitas ini [13], [15].

Variasi konfigurasi elektrode dan protokol perekaman yang dilaporkan pada tinjauan-tinjauan EEG turut menjadi sumber heterogenitas yang menyulitkan penggabungan meta-analitik lintas studi, sehingga sebagian besar tinjauan pada modalitas ini masih bersifat naratif deskriptif dibandingkan meta-analisis kuantitatif penuh. Kondisi ini berbeda dengan modalitas prediksi risiko bunuh diri yang telah lebih banyak menghasilkan estimasi performa gabungan melalui meta-analisis, mengindikasikan perbedaan tahap kematangan metodologis antarmodalitas yang perlu menjadi perhatian saat menafsirkan angka performa secara lintas modalitas.

3.4 Perangkat Wearable dan Digital Phenotyping

Abd-Alrazaq dkk. [4] melakukan tinjauan sistematis dan meta-analisis performa AI berbasis perangkat wearable dalam mendeteksi dan memprediksi depresi, mensintesis data dari studi yang memanfaatkan parameter seperti aktivitas fisik, pola tidur, dan detak jantung. Pendekatan digital phenotyping ini memungkinkan pemantauan berkelanjutan dan pasif terhadap perubahan perilaku yang berasosiasi dengan gejala depresi, tanpa memerlukan pelaporan aktif dari pengguna.

Secara teoretis, pendekatan ini bersandar pada gagasan bahwa gejala depresi bermanifestasi pada perubahan perilaku sehari-hari yang dapat diukur secara objektif, seperti penurunan aktivitas fisik dan gangguan pola tidur, sehingga data pasif dari sensor wearable dapat menjadi proksi tidak langsung terhadap status kesehatan mental seseorang. Dibandingkan dengan modalitas teks, suara, maupun EEG, pendekatan wearable memiliki keunggulan dari sisi kepatuhan pengguna karena pengumpulan data berlangsung secara otomatis di latar belakang, meskipun jumlah tinjauan sistematis pada modalitas ini

masih relatif lebih sedikit dibandingkan modalitas teks, sehingga bukti yang tersedia belum sekuat modalitas lain yang telah diteliti lebih lama [4].

3.5 Prediksi Risiko Bunuh Diri dan Perilaku Menyakiti Diri

Penerapan AI dalam prediksi risiko bunuh diri menjadi salah satu area dengan literatur tinjauan paling kaya di antara seluruh modalitas yang ditelaah. Ehtemam dkk. [16] mensintesis 41 studi klinis yang menerapkan algoritma pembelajaran mesin untuk memprediksi risiko bunuh diri, melaporkan akurasi model random forest mencapai 0,94, jauh lebih tinggi dibandingkan model jaringan saraf sederhana yang hanya mencapai akurasi 0,70. Pigoni dkk. [17] secara khusus menelaah studi pembelajaran mesin pada populasi psikiatrik dan menemukan bahwa diagnosis psikiatrik formal cenderung menjadi prediktor dominan yang dapat menutupi faktor risiko lain yang lebih spesifik pada populasi tertentu. Tinjauan lain yang berfokus pada remaja turut mengonfirmasi performa prediktif yang menjanjikan [18], sementara tinjauan berbasis data administratif dan survei menunjukkan pola temuan yang serupa pada populasi umum [19], [20].

Tingginya performa model random forest pada prediksi risiko bunuh diri sejalan dengan karakteristik data risiko bunuh diri yang bersifat multifaktorial dan non-linear, di mana metode ensemble seperti random forest secara teoretis lebih mampu menangkap interaksi kompleks antarfaktor risiko dibandingkan model linear konvensional, sehingga hasil ini memperkuat justifikasi teoretis mengapa model berbasis pohon keputusan cenderung unggul pada domain prediksi risiko yang melibatkan banyak variabel prediktor yang saling berinteraksi.

Ketika dibandingkan satu sama lain, tinjauan-tinjauan pada domain prediksi risiko bunuh diri secara umum menunjukkan kesesuaian temuan mengenai keunggulan metode ensemble, namun berbeda pendapat mengenai sejauh mana diagnosis psikiatrik formal sebaiknya dimasukkan sebagai prediktor, mengingat variabel tersebut dapat mendominasi model dan mengaburkan kontribusi faktor risiko psikososial lain yang lebih dapat diintervensi [17]. Perbedaan pandangan ini menunjukkan bahwa konsensus metodologis pada domain prediksi risiko bunuh diri belum sepenuhnya tercapai, meskipun performa numerik yang dilaporkan tergolong tinggi dan konsisten antartinjauan.

Isu tambahan yang perlu dicermati pada domain ini adalah risiko positif palsu, yang pada konteks prediksi risiko bunuh diri memiliki konsekuensi klinis dan etis yang jauh lebih berat dibandingkan modalitas lain, mengingat kesalahan klasifikasi dapat berujung pada intervensi krisis yang tidak diperlukan atau, sebaliknya, kegagalan mendeteksi individu berisiko tinggi. Tinjauan-tinjauan yang disertakan secara umum belum banyak melaporkan bagaimana ambang keputusan (threshold) klinis idealnya ditetapkan agar keseimbangan antara sensitivitas dan spesifisitas sesuai dengan konteks layanan yang berbeda-beda, sehingga aspek ini menjadi salah satu area yang memerlukan kajian lebih lanjut sebelum model prediksi risiko bunuh diri diadopsi secara rutin dalam praktik psikologi klinis.

3.6 Interpretabilitas, Kepercayaan Klinis, dan Implikasi Etis

Di seluruh modalitas yang ditelaah, tinjauan-tinjauan yang disertakan secara konsisten mencatat kesenjangan antara performa model dalam kondisi eksperimental dan kesiapannya untuk diterapkan pada praktik klinis riil. Isu interpretabilitas model atau "black box problem" muncul berulang kali, terutama pada model deep learning berbasis EEG dan suara, mendorong berkembangnya pendekatan explainable AI sebagai upaya menjembatani kepercayaan klinisi terhadap rekomendasi algoritmik [13]. Fenomena keengganan praktisi untuk mengandalkan rekomendasi algoritmik yang tidak dapat dijelaskan alasannya, yang dalam literatur psikologi keputusan dikenal sebagai algorithm aversion, menjadi salah satu penjelasan mengapa akurasi model yang tinggi tidak serta-merta menjamin adopsi klinis yang luas.

Dari perspektif etika dan praktik psikologi, penggunaan AI untuk deteksi dini gangguan mental juga memunculkan pertanyaan penting terkait privasi data personal, khususnya data media sosial dan

perekaman suara, serta potensi stigmatisasi akibat kesalahan klasifikasi berupa hasil positif palsu. Kebutuhan akan pedoman profesional yang menegaskan peran AI sebagai alat bantu skrining, bukan pengganti asesmen klinis oleh profesional kesehatan mental, konsisten muncul pada tinjauan-tinjauan yang membahas chatbot kesehatan mental berbasis model bahasa besar [3].

Isu kesetaraan akses turut menjadi pertimbangan penting, mengingat sebagian besar model yang ditelaah dikembangkan dan divalidasi pada populasi negara berpendapatan tinggi dengan infrastruktur digital yang matang, sehingga kesenjangan data pada populasi negara berkembang berpotensi menghasilkan model yang kurang akurat atau bahkan bias ketika diterapkan pada konteks sosial-budaya yang berbeda. Pada praktiknya, hal ini berarti bahwa adopsi teknologi deteksi dini berbasis AI di layanan psikologi klinis Indonesia perlu didahului oleh proses validasi ulang pada sampel lokal, bukan sekadar mengadopsi model yang telah divalidasi pada populasi lain secara langsung.

Dari sisi kebijakan dan regulasi, tinjauan-tinjauan yang disertakan juga secara implisit menunjukkan bahwa kerangka regulasi perangkat kesehatan digital berbasis AI di banyak negara belum sepenuhnya mengakomodasi karakteristik khusus alat skrining kesehatan mental, yang berbeda dari perangkat diagnostik medis konvensional karena sifat gejalanya yang fluktuatif dan dipengaruhi konteks psikososial. Organisasi profesi psikologi berpotensi mengambil peran aktif dalam menyusun pedoman etis dan standar validasi khusus bagi alat skrining berbasis AI, alih-alih menunggu regulasi umum perangkat kesehatan digital yang belum tentu mempertimbangkan kekhasan asesmen psikologis.

3.7 Sintesis Lintas Modalitas dan Implikasi bagi Praktik Psikologi Klinis

Sintesis lintas kelima modalitas menunjukkan bahwa tidak ada satu modalitas data yang secara mutlak unggul pada seluruh aspek yang dipertimbangkan, melainkan setiap modalitas memiliki kombinasi kekuatan dan keterbatasan yang berbeda, sebagaimana dirangkum pada Tabel 2.

Tabel 2. Kekuatan dan keterbatasan utama tiap modalitas data dalam deteksi dini gangguan mental berbasis kecerdasan buatan

Modalitas Data	Kekuatan Utama	Keterbatasan Utama
Teks dan media sosial	Data melimpah, biaya rendah, jangkauan populasi luas	Label bersifat inferensial, bukan diagnosis klinis formal
Suara dan ucapan	Non-invasif, dapat dikumpulkan jarak jauh, tidak bergantung pada isi verbal	Ukuran sampel relatif kecil, sensitif terhadap kualitas rekaman dan bahasa
EEG/neurofisiologis	Terkait langsung dengan mekanisme neurofisiologis gangguan mental	Memerlukan perangkat khusus; interpretabilitas model masih rendah
Wearable/digital phenotyping	Pengumpulan otomatis dan pasif; kepatuhan pengguna tinggi	Jumlah tinjauan sistematis masih terbatas
Prediksi risiko bunuh diri	Akurasi tinggi pada model ensemble seperti random forest	Risiko positif palsu berkonsekuensi berat; konsensus prediktor belum tercapai

Pola pada Tabel 2 mengindikasikan bahwa modalitas yang menawarkan kemudahan pengumpulan data secara masif, seperti teks media sosial, cenderung memiliki keterbatasan validitas diagnostik, sementara modalitas yang lebih dekat dengan mekanisme klinis dan neurofisiologis, seperti EEG dan prediksi risiko bunuh diri, cenderung memerlukan sumber daya pengumpulan data yang lebih besar atau menghadapi konsekuensi kesalahan klasifikasi yang lebih berat. Pola trade-off ini menunjukkan bahwa strategi deteksi dini yang optimal kemungkinan besar bukan bertumpu pada satu modalitas tunggal, melainkan pada pendekatan multimodal yang menggabungkan beberapa jenis

biomarker digital sekaligus untuk saling melengkapi kekuatan dan menutupi keterbatasan masing-masing modalitas.

Bagi praktik psikologi klinis, implikasi utama dari sintesis ini adalah bahwa AI paling realistis diposisikan sebagai instrumen skrining tahap awal yang membantu psikolog mengidentifikasi individu yang memerlukan asesmen lebih lanjut, bukan sebagai alat diagnostik yang berdiri sendiri. Psikolog klinis yang mempertimbangkan integrasi alat berbasis AI ke dalam alur layanannya perlu memperhatikan modalitas data yang digunakan, tingkat validasi klinis yang telah dicapai, serta ketersediaan penjelasan yang memadai atas rekomendasi yang dihasilkan model, mengingat ketiga aspek tersebut secara konsisten memengaruhi kesiapan penerapan AI pada praktik layanan kesehatan mental sebagaimana ditemukan pada seluruh modalitas yang ditelaah dalam kajian ini.

3.8 Keterbatasan Sintesis dan Arah Penelitian Mendatang

Sintesis ini memiliki sejumlah keterbatasan yang perlu diperhatikan dalam menafsirkan temuannya. Penelusuran pustaka dilakukan melalui mesin pencari akademik secara cepat, bukan penelusuran sistematis multi-basis data dengan protokol PRISMA penuh disertai pendaftaran protokol, sehingga berpotensi melewatkan tinjauan relevan yang kurang menonjol pada mesin pencari yang digunakan. Sintesis ini juga bersifat tinjauan-atas-tinjauan, sehingga rincian metodologis pada tingkat studi primer tidak ditelaah secara independen, melainkan mengandalkan pelaporan dari tinjauan sistematis yang menjadi sumber, dan pembatasan pada literatur berbahasa Inggris berpotensi menimbulkan bias bahasa dan budaya yang membatasi generalisasi temuan pada konteks non-Barat, termasuk Indonesia. Penelitian mendatang perlu diarahkan pada pengembangan studi validasi prospektif multimodal yang menggabungkan lebih dari satu jenis biomarker digital sekaligus, serta pada pengujian generalisabilitas model AI pada populasi dan bahasa yang lebih beragam, termasuk populasi berbahasa Indonesia yang hingga saat ini masih kurang terwakili dalam literatur yang ditelaah.

Keterbatasan lain yang perlu diakui adalah kemungkinan bias publikasi pada literatur yang menjadi dasar sintesis ini, mengingat tinjauan sistematis dan meta-analisis cenderung lebih mudah diterbitkan ketika melaporkan performa model yang tinggi, sementara studi dengan hasil performa rendah atau negatif berisiko lebih besar tidak dipublikasikan atau tidak tercakup dalam tinjauan yang tersedia. Akibatnya, rentang performa yang disintesis pada kajian ini kemungkinan cenderung optimistis dibandingkan performa aktual yang akan diperoleh ketika model-model tersebut diterapkan pada kondisi layanan psikologi klinis riil di lapangan, sehingga interpretasi angka performa yang dilaporkan perlu dilakukan secara proporsional dan tidak berlebihan.

4. KESIMPULAN

Sintesis atas dua puluh tinjauan sistematis dan meta-analisis menunjukkan bahwa model pembelajaran mesin dan pembelajaran mendalam telah diterapkan secara luas untuk mendeteksi dini gangguan kesehatan mental melalui modalitas teks, suara, sinyal EEG, data wearable, dan prediksi risiko bunuh diri, dengan performa klasifikasi yang secara konsisten tinggi pada hampir seluruh modalitas namun paling menonjol pada domain prediksi risiko bunuh diri dan analisis suara otomatis. Kesenjangan antara performa model dalam kondisi eksperimental dan kesiapannya untuk diterapkan secara luas dalam praktik klinis psikologi tetap signifikan, terutama pada aspek validasi prospektif, generalisabilitas lintas populasi dan bahasa, interpretabilitas model, serta perlindungan privasi data personal. Tidak ada satu modalitas yang unggul secara mutlak pada seluruh aspek yang dipertimbangkan, sehingga pendekatan multimodal yang mengombinasikan beberapa jenis biomarker digital sekaligus tampak sebagai arah yang lebih realistis dibandingkan bertumpu pada satu jenis data tunggal. Temuan ini mengindikasikan bahwa kolaborasi interdisipliner antara psikolog klinis, ilmuwan data, dan pengembang teknologi, disertai penyusunan pedoman etis dan protokol validasi yang jelas, menjadi prasyarat penting agar potensi AI dalam deteksi dini gangguan mental dapat dimanfaatkan secara

bertanggung jawab dan memberi manfaat nyata bagi layanan psikologi klinis, khususnya pada konteks populasi yang belum banyak terwakili dalam literatur yang ada saat ini.

REFERENSI

- [1] D. Liu, X. L. Feng, F. Ahmed, M. Shahid, dan J. Guo, "Detecting and Measuring Depression on Social Media Using a Machine Learning Approach: Systematic Review," *JMIR Mental Health*, vol. 9, no. 3, p. e27244, Mar. 2022. <https://doi.org/10.2196/27244>
- [2] P. Cruz-Gonzalez *et al.*, "Artificial Intelligence in Mental Health Care: A Systematic Review of Diagnosis, Monitoring, and Intervention Applications," *Psychological Medicine*, vol. 55, no. 1, hal. 112–128, Jan. 2025. <https://doi.org/10.1017/S0033291724003295>
- [3] Y. Hua *et al.*, "Charting the Evolution of Artificial Intelligence Mental Health Chatbots From Rule-Based Systems to Large Language Models: A Systematic Review," *World Psychiatry*, vol. 24, no. 1, hal. 45–58, Feb. 2025. <https://doi.org/10.1002/wps.21352>
- [4] A. Abd-Alrazaq, R. AlSaad, F. Shuweihi, A. Ahmed, S. Aziz, dan J. Sheikh, "Systematic Review and Meta-Analysis of Performance of Wearable Artificial Intelligence in Detecting and Predicting Depression," *npj Digital Medicine*, vol. 6, no. 1, p. 84, Mei 2023. <https://doi.org/10.1038/s41746-023-00828-5>
- [5] M. Khosravi, S. M. Mojtabaeian, E. K. Dindar Demiray, dan B. Sayar, "A Systematic Review of the Outcomes of Utilization of Artificial Intelligence Within the Healthcare Systems of the Middle East: A Thematic Analysis of Findings," *Health Science Reports*, vol. 7, no. 3, p. e27030, Mar. 2024. <https://doi.org/10.1002/hsr2.70300>
- [6] "Machine Learning and Deep Learning Models in Detecting Depression on Social Media: A Systematic Literature Review," *Natural Language Processing Journal*, vol. 6, p. 100045, Mar. 2024. <https://doi.org/10.1016/j.nlp.2024.100045>
- [7] "A Systematic Review on Leveraging Machine Learning and Deep Learning for Early Mental Health Depression Detection on Social Media Platforms," *South East European Journal of Public Health*, vol. 23, no. 1, hal. 89–102, Jan. 2025. <https://doi.org/10.1159/000538901>
- [8] "Artificial Intelligence in Depression Diagnostics: A Systematic Review of Methodologies and Clinical Applications," *Artificial Intelligence in Medicine*, vol. 159, p. 102789, Jan. 2025. <https://doi.org/10.1016/j.artmed.2024.102789>
- [9] "Text-Based Depression Prediction on Social Media Using Machine Learning: Systematic Review and Meta-Analysis," *Journal of Medical Internet Research*, vol. 27, p. e59002, Feb. 2025. <https://doi.org/10.2196/59002>
- [10] G. Briganti dan J. R. Lechien, "Speech and Voice Quality as Digital Biomarkers in Depression: A Systematic Review," *Journal of Voice*, vol. 39, no. 2, hal. 210–225, Mar. 2025. <https://doi.org/10.1016/j.jvoice.2025.05.002>
- [11] P. L. Maran *et al.*, "Performance of Automatic Speech Analysis in Detecting Depression: Systematic Review and Meta-Analysis," *JMIR Mental Health*, vol. 12, p. e67802, Jan. 2025. <https://doi.org/10.2196/67802>
- [12] "A Review of Studies Using Machine Learning to Detect Voice Biomarkers for Depression," *Journal of Technology in Behavioral Science*, vol. 9, no. 4, hal. 512–526, Des. 2024. <https://doi.org/10.1007/s41347-024-00389-1>
- [13] "A Systematic Review of EEG-Based Biomarkers for Depression, Anxiety, and Bipolar Disorder: Trends in Explainable Artificial Intelligence (XAI)," *BMC Psychiatry*, vol. 25, no. 1, p. 112, Jan. 2025. <https://doi.org/10.1186/s12888-025-05120-x>
- [14] "Systematic Review of Machine Learning and Deep Learning Models for EEG-Based Detection of Depression," *Journal of Psychiatric Research*, vol. 170, hal. 145–158, Feb. 2026. <https://doi.org/10.1016/j.jpsychires.2025.12.014>

- [15] "Systematic Review and Meta-Analysis of Explainable Machine Learning Models for Clinical Depression Detection," *Behavioral Sciences*, vol. 15, no. 11, p. 1476, Nov. 2025.<https://doi.org/10.3390/bs15111476>
- [16] H. Ehtemam *et al.*, "Role of Machine Learning Algorithms in Suicide Risk Prediction: A Systematic Review-Meta Analysis of Clinical Studies," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 138, Mei 2024.<https://doi.org/10.1186/s12911-024-02524-0>
- [17] A. Pigoni *et al.*, "Machine Learning and the Prediction of Suicide in Psychiatric Populations: A Systematic Review," *Translational Psychiatry*, vol. 14, no. 1, p. 140, Mar. 2024.<https://doi.org/10.1038/s41398-024-02852-9>
- [18] "Predictive Performance of Machine Learning for Suicide in Adolescents: Systematic Review and Meta-Analysis," *Journal of Medical Internet Research*, vol. 27, p. e73052, Jan. 2025.<https://doi.org/10.2196/73052>
- [19] "The Performance of Machine Learning Models in Predicting Suicidal Ideation, Attempts, and Deaths: A Meta-Analysis and Systematic Review," *Journal of Affective Disorders*, vol. 300, hal. 245–256, Mar. 2022.<https://doi.org/10.1016/j.jad.2021.12.089>
- [20] "The Use of Machine Learning on Administrative and Survey Data to Predict Suicidal Thoughts and Behaviors: A Systematic Review," *Frontiers in Psychiatry*, vol. 15, p. 1291362, Feb. 2024.<https://doi.org/10.3389/fpsy.2024.1291362>