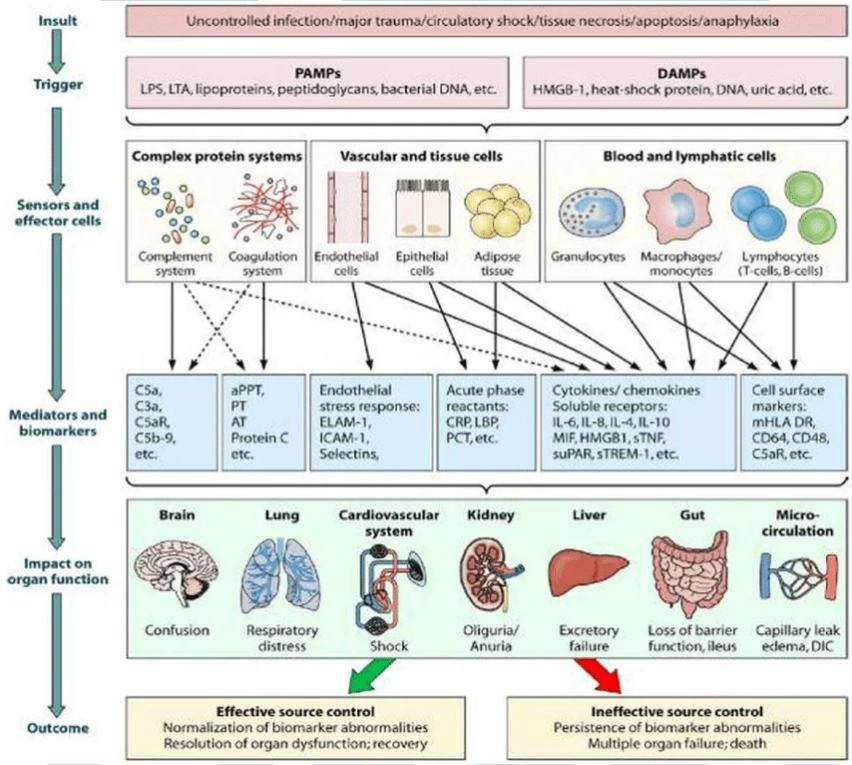
BAB 2
LANDASAN TEORI

2.1 Sepsis



Gambar 2.1. Pathophysiology of sepsis induced organ dysfunction

Sumber: [10]

Sepsis didefinisikan secara medis sebagai bentuk disfungsi organ yang mengancam jiwa yang disebabkan oleh respons inang (*host*) yang tidak teregulasi terhadap infeksi [11, 12, 13]. Secara patofisiologis, alur sepsis secara komprehensif disajikan pada Gambar 2.1. Proses ini diawali dengan adanya *insult* berupa infeksi yang tidak terkendali, trauma mayor, atau nekrosis jaringan yang memicu pelepasan *Pathogen-Associated Molecular Patterns* (PAMPs) dan *Damage-Associated Molecular Patterns* (DAMPs). Molekul-molekul ini bertindak sebagai *trigger* yang diidentifikasi oleh sensor seluler dan sel efektor, termasuk sistem protein kompleks (komplemen dan koagulasi), sel jaringan vaskular, serta sel limfatik darah.

Interaksi tersebut memicu kaskade mediator dan biomarker inflamasi,

seperti sitokin (IL-6, IL-8, TNF), reaktan fase akut (CRP, PCT), serta respons stres endotel. Akumulasi dari mediator ini memberikan dampak sistemik pada berbagai fungsi organ, yang bermanifestasi klinis sebagai kebingungan (*confusion*) pada otak, distress pernapasan pada paru, syok pada sistem kardiovaskular, serta kegagalan ekskresi pada ginjal dan hati. Hasil akhir (*outcome*) dari perjalanan klinis ini sangat ditentukan oleh efektivitas kendali sumber infeksi (*source control*); di mana kegagalan dalam pengendalian tersebut akan menyebabkan kegagalan organ multipel yang berujung pada kematian [14].

Kriteria diagnosis sepsis dalam lingkungan klinis sangat bergantung pada pemantauan parameter tanda vital yang ketat. Indikator utama yang sering diamati meliputi peningkatan detak jantung (*heart rate*), perubahan laju pernapasan (*respiratory rate*), dan penurunan tekanan darah (*blood pressure*) . Selain parameter fisiologis, hasil laboratorium klinis menjadi komponen krusial dalam identifikasi awal. Peningkatan kadar laktat dalam darah serta fluktuasi jumlah sel darah putih (Leukosit/WBC) berfungsi sebagai penanda biokimia penting untuk menilai derajat keparahan infeksi dan risiko mortalitas pasien [15]. Penentuan *onset* sepsis secara klinis umumnya didasarkan pada peningkatan skor *Sequential Organ Failure Assessment* (SOFA) sebesar dua poin atau lebih [11, 16]. Mengingat sepsis merupakan penyebab utama kematian di rumah sakit secara global, identifikasi dini menjadi sangat deterministik bagi kelangsungan hidup pasien [17]. Keterlambatan deteksi di unit perawatan intensif (ICU) dapat berakibat fatal, karena intervensi medis seperti pemberian antibiotik spektrum luas dan resusitasi cairan memiliki efektivitas yang jauh lebih tinggi jika diimplementasikan pada fase awal perjalanan penyakit [18].

2.2 Korelasi Penyakit Kanker dan Risiko Sepsis

Terdapat keterkaitan yang signifikan antara kondisi patologis kanker, intervensi kemoterapi, dan eskalasi risiko terjadinya sepsis. Pasien onkologi dikategorikan sebagai populasi dengan gangguan sistem imun (*immunocompromised*) karena pertahanan biologis mereka sering kali melemah akibat progresi penyakit maupun dampak sitotoksik dari pengobatan agresif [6]. Fenomena kritis yang umum dijumpai pada kelompok ini adalah *neutropenia*, yaitu kondisi penurunan jumlah sel darah putih (*leukosit*) secara ekstrem yang melumpuhkan mekanisme pertahanan tubuh dalam merespons patogen infeksius [19]. Kondisi imunosupresi tersebut menyebabkan pasien kanker memiliki tingkat

fatalitas yang jauh lebih tinggi saat mengalami sepsis dibandingkan dengan populasi pasien umum [6]. Selain itu, tantangan diagnostik pada kelompok ini menjadi sangat kompleks karena manifestasi klinis sepsis pada pasien dengan komorbiditas onkologi cenderung bersifat atipikal atau tidak menunjukkan gejala standar. Hal ini menimbulkan risiko tinggi terjadinya keterlambatan deteksi klinis atau kesalahan diagnosis yang dapat berakibat fatal bagi pasien [12].

2.2.1 Multiple Imputation by Chained Equations (MICE)

Data hilang (*missing data*) merupakan permasalahan yang umum dijumpai pada basis data rekam medis elektronik, termasuk MIMIC-IV, akibat variasi frekuensi pemeriksaan klinis, keterbatasan pencatatan, maupun karakteristik alur perawatan pasien yang tidak seragam [20]. Penanganan data hilang yang tidak tepat, seperti penghapusan observasi secara *listwise* (*complete-case analysis*), berisiko mengurangi ukuran sampel secara signifikan serta memperkenalkan bias apabila pola kehilangan data tidak bersifat acak sepenuhnya [20].

Rubin [21] mengklasifikasikan mekanisme data hilang ke dalam tiga kategori: *Missing Completely at Random* (MCAR), di mana probabilitas suatu nilai hilang tidak berkaitan dengan nilai variabel apa pun; *Missing at Random* (MAR), di mana probabilitas data hilang dapat dijelaskan oleh variabel lain yang teramati dalam *dataset*; dan *Missing Not at Random* (MNAR), di mana probabilitas data hilang berkaitan dengan nilai variabel itu sendiri yang tidak teramati. Multiple imputation, sebagaimana diformulasikan oleh Rubin [21], dirancang untuk menghasilkan estimasi yang valid di bawah asumsi MAR dengan mengganti setiap nilai hilang menggunakan beberapa nilai plausibel yang dihasilkan dari model prediktif, alih-alih satu nilai tunggal seperti pada metode imputasi sederhana (*mean/median imputation*).

Multiple Imputation by Chained Equations (MICE) merupakan salah satu pendekatan multiple imputation yang paling banyak digunakan pada data klinis multivariat, diperkenalkan oleh Van Buuren dan Groothuis-Oudshoorn [21]. MICE bekerja secara iteratif dengan memodelkan setiap variabel yang memiliki data hilang sebagai fungsi dari seluruh variabel lain dalam *dataset*, kemudian mengulangi proses tersebut secara bergantian (*chained*) hingga konvergen. Secara umum, untuk suatu *dataset* dengan p variabel $X_1, X_2, \ldots, X_p$ yang memiliki nilai hilang, algoritma MICE dijalankan melalui tahapan berikut pada setiap iterasi ke-t:

$$X_j^{(t)} \sim f\left(X_j \mid X_1^{(t)}, \ldots, X_{j-1}^{(t)}, X_{j+1}^{(t-1)}, \ldots, X_p^{(t-1)}\right) \tag{2.1}$$

Di mana $X_j^{(t)}$ merupakan nilai imputasi untuk variabel ke-j pada iterasi ke-t, yang diestimasi melalui model regresi $f(\cdot)$ dengan variabel X_j sebagai target dan seluruh variabel lain sebagai prediktor – menggunakan nilai imputasi terbaru dari variabel yang telah diproses pada iterasi saat ini $(X_1^{(t)}, \ldots, X_{j-1}^{(t)})$, dan nilai imputasi dari iterasi sebelumnya untuk variabel yang belum diproses $(X_{j+1}^{(t-1)}, \ldots, X_p^{(t-1)})$. Proses ini diulang untuk seluruh variabel bermasalah secara bergantian sampai nilai imputasi konvergen antar-iterasi. Bentuk model regresi $f(\cdot)$ dapat disesuaikan dengan tipe data variabel target, misalnya regresi linear atau *Bayesian Ridge Regression* untuk variabel numerik kontinu, dan regresi logistik untuk variabel kategorikal biner [21].

Pendekatan MICE telah digunakan secara luas pada studi prediksi klinis berbasis basis data MIMIC, termasuk pada kedua studi acuan penelitian ini [22, 23], sebagai metode standar untuk menangani data laboratorium dan tanda vital yang hilang pada proporsi di bawah ambang batas tertentu (umumnya 20%), sementara variabel dengan proporsi data hilang yang lebih tinggi dieksklusi dari analisis lanjutan.

2.2.2 Seleksi Fitur (*Feature Selection*)

Seleksi fitur merupakan tahapan kritikal dalam pemrosesan data medis guna mereduksi dimensi data, menghilangkan variabel yang bersifat redundan atau tidak relevan, memitigasi risiko *overfitting*, serta meningkatkan efisiensi komputasi dan interpretabilitas model [22]. Dalam studi prediksi klinis berbasis data EHR (*Electronic Health Records*), beragam metode seleksi fitur dapat diterapkan, mulai dari pendekatan statistik inferensial hingga algoritma berbasis *ensemble* dan keutamaan fitur inheren model [23, 22]. Penelitian ini meninjau tiga metode seleksi fitur utama: *Backward Stepwise Selection*, Algoritma *Boruta*, dan *XGBoost Feature Importance*.

A Backward Stepwise Selection

Backward Stepwise Selection merupakan pendekatan eliminasi sekvensial berbasis statistik inferensial yang diawali dengan memasukkan seluruh himpunan variabel prediktor kandidat ke dalam model linier (seperti *Logistic Regression*)

[23, 24, 25, 26]. Pada setiap iterasi, algoritma mengevaluasi kontribusi setiap variabel dan secara bertahap menghapus variabel yang memberikan kontribusi statistik paling tidak signifikan hingga mencapai kriteria penghentian optimal berbasis *Akaike Information Criterion* (AIC) atau *Bayesian Information Criterion* (BIC) [23].

Metode ini mengoptimalkan fungsi AIC untuk menyeimbangkan antara tingkat kelayakan model (*goodness of fit*) dan kompleksitas jumlah parameter:

$$\mathrm{AIC} = 2k - 2\ln(\hat{L}) \tag{2.2}$$

Di mana k merupakan jumlah parameter yang diestimasi dalam model dan $\hat{L}$ menyatakan nilai maksimum dari fungsi *likelihood* model. Variabel yang memiliki nilai p-value terburuk atau yang jika dihapus menghasilkan penurunan nilai AIC terbesar akan dieliminasi secara iteratif. Pendekatan ini sebagaimana diterapkan oleh Hou et al. [23] efektif dalam menghasilkan model prediktif yang ringkas (*parsimonious*).

B Algoritma Boruta (*Boruta Algorithm*)

Algoritma *Boruta* merupakan metode seleksi fitur berbasis *wrapper* yang dibangun di atas algoritma *Random Forest* untuk mengidentifikasi seluruh variabel yang relevan secara signifikan [27, 22, 27, 28]. Berbeda dari metode eliminasi tradisional yang hanya mencari himpunan fitur minimal yang optimal, Boruta mengevaluasi signifikansi setiap fitur riil dengan membandingkannya terhadap fitur acak tiruan (*shadow features*).

Mekanisme kerja algoritma Boruta dilakukan melalui tahapan berikut [27]:

1. Pembuatan Fitur Bayangan (*Shadow Features*): Duplikasi seluruh variabel asli dan acak ulang nilai-nilainya secara independen untuk menghilangkan korelasi dengan variabel dependen y.

2. Pelatihan Model: Latih model *Random Forest* pada dataset gabungan (fitur asli dan fitur bayangan) lalu hitung skor urgensi fitur (*Z-score*) berbasis *Mean Decrease Impurity* (MDI) atau *Mean Decrease Accuracy* (MDA):

$$Z_i = \frac{I_i}{\sigma_I} \tag{2.3}$$

Di mana I_i adalah rerata penurunan akurasi/impuritas dari fitur i, dan σ_I adalah deviasi standarnya.

3. Pengujian Hipotesis: Tentukan nilai Z-score maksimum dari seluruh fitur bayangan ($Z_{\text{max\_shadow}}$). Evaluasi statistik menggunakan uji binomial dua arah diterapkan untuk mengonfirmasi apakah Z-score dari fitur asli secara signifikan lebih tinggi dibanding $Z_{\text{max\_shadow}}$.

Fitur yang secara konsisten berkinerja lebih baik daripada fitur bayangan tertinggi diklasifikasikan sebagai *Confirmed*, sedangkan yang secara signifikan berkinerja lebih buruk dieliminasi (*Rejected*) [22].

C XGBoost Feature Importance

Algoritma *Extreme Gradient Boosting* (XGBoost) memiliki kemampuan inheren untuk mengevaluasi dan mengurutkan kontribusi relative setiap variabel prediktor selama proses pembentukan pohon keputusan [29]. Metrik keutamaan fitur (*feature importance*) dalam XGBoost yang paling umum dan akurat untuk mengukur kontribusi prediktif adalah *Gain* [29, 30, 31, 32].

Gain mengukur peningkatan rata-rata dalam reduksi fungsi kerugian (*loss reduction*) yang dihasilkan oleh pemisahan node (*split*) menggunakan variabel tertentu di seluruh pohon keputusan dalam *ensemble*. Formulasi reduksi kerugian ($\mathscr{L}_{\text{split}}$) saat melakukan pemisahan node dinyatakan sebagai berikut [29]:

$$\text{Gain} = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \tag{2.4}$$

Di mana G_L dan G_R merupakan penjumlahan gradien orde pertama (*first-order gradient*) pada cabang anak kiri dan kanan, sedangkan H_L dan H_R merupakan penjumlahan gradien orde kedua (*hessian*) pada masing-masing cabang. Parameter λ adalah konstanta regularisasi $L2$, dan γ menyatakan penalti kompleksitas daun.

Fitur dengan nilai *Gain* kumulatif yang lebih tinggi menunjukkan peran yang lebih dominan dalam memisahkan observasi klinis secara akurat, sehingga dapat dijadikan acuan dalam menyeleksi subset variabel terbaik sebelum pemodelan akhir.

2.3 Machine Learning

Machine Learning merupakan representasi dari paradigma kecerdasan buatan yang memfasilitasi pengembangan model untuk mengestraksi pola dari data historis guna melakukan prediksi penyakit secara dinamis berdasarkan informasi klinis yang kompleks [33, 25]. Dalam ranah klasifikasi medis, pendekatan *Supervised Learning* diimplementasikan untuk memodelkan korelasi antara variabel independen, seperti data klinis pasien, dengan variabel dependen kategorikal, seperti status mortalitas 28-hari pasien, guna menghasilkan probabilitas prediksi yang akurat [34, 25]. Penelitian ini membandingkan empat algoritma *supervised learning*: *Logistic Regression* sebagai model pembanding klasik (*baseline*), serta tiga algoritma berbasis pohon keputusan (*tree-based*) – *Decision Tree*, *Random Forest*, dan *Extreme Gradient Boosting* (XGBoost). Pemilihan model berbasis pohon dilandasi oleh sifat inheren algoritma tersebut yang tidak sensitif terhadap outlier maupun perbedaan skala antar-fitur, karena keputusan pemisahan (*split*) pada setiap node ditentukan berdasarkan urutan (*ranking*) nilai, bukan magnitudonya secara absolut [35]. Karakteristik ini relevan mengingat nilai ekstrem pada variabel klinis (misalnya kadar laktat atau kreatinin yang sangat tinggi) merepresentasikan keparahan kondisi pasien yang sesungguhnya, sehingga tidak dihilangkan dari data. Skema perbandingan model tradisional versus *tree-based* ini merujuk pada pendekatan yang diterapkan Hou et al. [23], yang membandingkan *Logistic Regression* langsung terhadap XGBoost.

2.3.1 Logistic Regression

Logistic Regression merupakan algoritma klasifikasi linear yang memodelkan probabilitas suatu observasi termasuk ke dalam kelas tertentu melalui fungsi *sigmoid* (*logistic function*), yang memetakan kombinasi linear dari variabel independen ke rentang probabilitas 0 hingga 1 [25]. Model ini banyak digunakan sebagai algoritma pembanding *baseline* dalam studi prediksi klinis karena interpretabilitasnya yang tinggi [22, 23].

Probabilitas kelas positif dimodelkan sebagai berikut:

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}} \tag{2.5}$$

Di mana $\mathbf{w}$ merupakan vektor bobot (koefisien) untuk setiap variabel independen $\mathbf{x}$,

dan b merupakan *intercept*. Parameter model diestimasi melalui minimisasi fungsi kerugian *log-loss* (*binary cross-entropy*) sebagai berikut:

$$J(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(p_i) + (1-y_i)\log(1-p_i)] + \frac{1}{2C}\|\mathbf{w}\|_2^2 \tag{2.6}$$

Suku terakhir merupakan regularisasi *L2* (*ridge penalty*) dengan parameter invers regularisasi C, yang berfungsi mencegah *overfitting* dengan membatasi besaran koefisien model [25].

2.3.2 Decision Tree

Decision Tree merupakan algoritma klasifikasi berbasis struktur pohon yang membagi data secara rekursif berdasarkan nilai ambang variabel independen, mengikuti algoritma *Classification and Regression Trees* (CART) [22, 36, 37, 38]. Setiap pemisahan (*split*) dipilih untuk meminimalkan tingkat ketidakmurnian (*impurity*) pada node anak, diukur menggunakan indeks Gini:

$$\text{Gini}(t) = 1 - \sum_{c=1}^{C} p_c^2 \tag{2.7}$$

Di mana p_c merupakan proporsi observasi kelas c pada node t. Pemisahan optimal pada setiap node dipilih dengan memaksimalkan penurunan Gini antara node induk dan node anak:

$$\Delta\text{Gini} = \text{Gini}(t) - \left(\frac{n_L}{n_t}\text{Gini}(t_L) + \frac{n_R}{n_t}\text{Gini}(t_R)\right) \tag{2.8}$$

Di mana t_L dan t_R merupakan node anak kiri dan kanan hasil pemisahan, dengan n_L, n_R, dan n_t berturut-turut menyatakan jumlah observasi pada masing-masing node.

2.3.3 Random Forest

Random Forest merupakan algoritma *ensemble learning* yang mengonstruksi sejumlah B *decision tree* secara independen menggunakan teknik *bootstrap aggregating* (*bagging*), di mana setiap pohon dilatih pada sampel data hasil *resampling* dengan pengembalian (*bootstrap sample*) dari data latih asli [22, 39, 40, 41]. Selain itu, pada setiap pemisahan node, hanya subset acak dari fitur yang dipertimbangkan, sehingga mengurangi korelasi antar-pohon dan

meningkatkan generalisasi model.

Prediksi akhir untuk klasifikasi ditentukan melalui *majority voting* dari seluruh pohon dalam *ensemble*:

$$\hat{y} = \text{mode}\left(\{h_b(\mathbf{x})\}_{b=1}^{B}\right) \tag{2.9}$$

Di mana $h_b(\mathbf{x})$ merupakan prediksi dari pohon ke-b, dan B merupakan jumlah total pohon dalam *ensemble*. Penelitian ini menggunakan $B = 200$ pohon keputusan.

2.3.4 Extreme Gradient Boosting (XGBoost)

XGBoost merupakan algoritma berbasis *tree-ensemble* yang mengonstruksi sekumpulan pohon keputusan secara berurutan (*sequential*) untuk meningkatkan presisi prediksi, memitigasi risiko *overfitting*, serta memiliki kemampuan inheren dalam menangani data yang hilang (*missing data*) secara efisien [34, 42, 22].

Secara matematis, fondasi dari algoritma XGBoost didasarkan pada optimasi fungsi objektif (*objective function*) pada setiap iterasi ke-t, yang diformulasikan sebagai berikut:

$$\mathscr{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{2.10}$$

Di mana l merupakan fungsi kerugian (*loss function*) yang mengukur perbedaan antara target aktual y_i dan prediksi dari iterasi sebelumnya $\hat{y}_i^{(t-1)}$, sedangkan $f_t(x_i)$ adalah fungsi pohon baru yang ditambahkan pada langkah ke-t. Keunggulan utama XGBoost terletak pada penggunaan ekspansi Taylor orde kedua untuk mendekati fungsi kerugian tersebut secara cepat, serta penambahan suku regularisasi eksplisit $\Omega(f_t)$ yang didefinisikan sebagai:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{2.11}$$

Suku regularisasi ini mengontrol kompleksitas model dengan memberikan penalti terhadap jumlah daun (*leaves*) (T) menggunakan parameter γ, serta melakukan penyusutan (*shrinkage*) terhadap bobot pada setiap daun (w_j) melalui parameter *L2 regularization* (λ). Mekanisme matematika inilah yang menjadikan XGBoost sangat tangguh terhadap *overfitting* saat memodelkan data medis yang

kompleks [29].

Dalam implementasi pada data medis, sering kali muncul tantangan berupa ketidakseimbangan data (*imbalanced data*) [43], sebagaimana ditemukan pada distribusi *outcome* penelitian ini (proporsi kematian lebih kecil dibanding pasien bertahan hidup). Pendekatan *Cost-Sensitive Learning* melalui mekanisme penyesuaian bobot kelas (*class weighting*) menjadi strategi yang efektif untuk meningkatkan sensitivitas model terhadap kelas minoritas tanpa memerlukan augmentasi data sintetis secara berlebihan [44].

Pengukuran ketangguhan model dapat dilakukan melalui analisis nilai probabilitas prediksi yang dihasilkan oleh algoritma XGBoost. Pada proses klasifikasi biner, XGBoost terlebih dahulu menghasilkan skor prediksi mentah (*raw score*) atau *log-odds* yang diperoleh dari akumulasi keluaran seluruh pohon keputusan. Skor tersebut dinyatakan sebagai berikut [29]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{2.12}$$

2.4 Penelitian Terkait

Tabel 2.1 terdapat ringkasan studi terdahulu yang menjadi pendukung bagi penelitian ini. Evaluasi terhadap literatur tersebut difokuskan pada tiga aspek utama: permasalahan, *research gap*, dan hasil. Pencarian permasalahan dapat ditemukan dari latar belakang dan rumusan masalah pada masing-masing studi. Selanjutnya, *research gap* ditemukan melalui analisis pada bagian komponen jurnal, mulai dari abstrak, metode, hingga kesimpulan. Dari analisa tersebut, dapat ditemukan keterbatasan teori maupun kekurangan dari hasil yang ada. Berdasarkan evaluasi terhadap keberhasilan studi-studi tersebut, penelitian ini menyeleksi dan mengadopsi beberapa elemen relevan untuk diimplementasikan. Rincian kajian literatur beserta elemen yang diadaptasi diuraikan sebagai berikut.

Tabel 2.1. Penelitian Terdahulu

Judul	Penulis & Tahun Penerbitan	Dataset	Metode Penelitian	Hasil
On the Intersection of Explainable and Reliable AI for Physical Fatigue Prediction [45]	Narteni et al. (2022)	Simulasi tugas manual (warehousing); 15 partisipan; data IMU	LLM (Logic Learning Machine), DT, NN, SVM, XGBoost, LIME, Anchors	LLM mengungguli DT (Akurasi 0,84 vs 0,76) dan menunjukkan spesifisitas tinggi
An Explainable AI Framework for Parkinson's Disease Prediction Using XGBoost, SHAP, and Large Language Model Integration [46]	Biswas et al. (2025)	Parkinson's Disease Dataset; 195 sampel dengan 23 fitur vokal	XGBoost, SHAP, Meta-LLaMA-3	XGBoost mencapai Akurasi tertinggi 97,43% dan AUC 0,98
Development and prospective implementation of a large language model based system for early sepsis prediction [15]	Shashikumar et al. (2025)	EHR dari dua rumah sakit UCSD Health; 2.500 pasien	COMPOSER (DNN), COMPOSER-LLM (Mixtral 8x7B), RAG	Sensitivitas 72,1%, PPV 52,9%, dan F1-score 61,0%
Enhancing Clinical Decision Support and Patient Communication Using MedGemma and Open Health AI Models [47]	Ugwumba (2025)	Tweet bencana (6.588), teks klinis (32), ringkasan medis (15)	BioGPT, MedGemma	Analisis klinis mencapai success rate 100% dan akurasi identifikasi kondisi kritis 95%

Lanjut pada halaman berikutnya

Tabel 2.1 Penelitian Terdahulu (lanjutan)

Judul	Penulis & Tahun Penerbitan	Dataset	Metode Penelitian	Hasil
Evaluating clinical AI summaries with large language models as judges [48]	Croxford et al. (2025)	Catatan UW Health dan database MIMIC-III (ProbSum)	GPT-o3-mini, GPT-4o, DeepSeek R1, Mixtral 8x22B, Llama 3.1	GPT-o3-mini mencapai reliabilitas antar-penilai (ICC) tertinggi sebesar 0,818
Integrating Classical Classifiers, Explainable AI and Transformer Models for Financial Headline Sentiment [49]	Opoku et al. (2025)	349 headline berita finansial; dataset proprietary	SVC, LogReg, RF, XGBoost, SHAP, LIME, Mistral-7B-Instruct	SVC dengan Word2Vec menghasilkan Akurasi terbaik 0,85 dan F1-score 0,65
Cancer Diagnosis Categorization in Electronic Health Records Using LLMs and BioBERT [50]	Hashtarkhani et al. (2025)	3.456 rekam medis pasien kanker Tennessee Medical Center 2017-2021	GPT-3.5, GPT-4o, Llama 3.2, Gemini 1.5, BioBERT	762 diagnosis unik, BioBERT akurasi 90.8%
Leveraging explainable artificial intelligence for early prediction of bloodstream infections [44]	Bopche et al. (2024)	EHR St. Olavs Hospital, Norwegia; 35.591 pasien	XGBoost, LightGBM, CatBoost, RF, LSTM, GRU, CNN, Transformer, SHAP	Model statis terbaik: RF (AUROC 0,8407); Model sekuensial terbaik: GRU (AUROC 0,7830)

Lanjut pada halaman berikutnya

Tabel 2.1 Penelitian Terdahulu (lanjutan)

Judul	Penulis & Tahun Penerbitan	Dataset	Metode Penelitian	Hasil
Machine learning approach for the prediction of 30-day mortality in sepsis-associated encephalopathy [33]	Peng et al. (2022)	MIMIC-IV; 6.994 pasien	NNET, NB, LR, GBM, RF, ADA, XGB, CatBoost	Model Adapting Boosting dengan akurasi terbesar 0.846 dan AUC 0.834
Machine learning for the prediction of acute kidney injury in patients with sepsis [22]	Yue et al. (2022)	MIMIC-III; 3.176 pasien sepsis	LR, KNN, SVM, DT, RF, XGBoost, ANN	XGBoost memberikan performa terbaik dengan akurasi 0,832 AUC 0,817
Performance of AI and LLMs on Neurosurgical Board Examinations [51]	R.K et al. (2025)	476 pertanyaan ujian bedah saraf; sumber CNS SANS ABNS	22 model LLM (GPT-5, GPT-4.5, Claude 3.7, Gemini 2.5 Pro, Llama-4, dsb)	Gemini 2.5 Pro, Grok, dan GPT-5 melebihi akurasi 80%
Evaluating reasoning large language models in ophthalmic question answering [52]	Wang et al. (2026)	MedQA-Eye; 967 pertanyaan oftalmologi dalam 3 bahasa	DeepSeek-R1, QwQ-32B, LLaMA-3.3-70B-Instruct	DeepSeek-R1 mencapai Akurasi (ACC) tertinggi sebesar 90,59%
Integrating clinical guidelines with large language models for sepsis mortality prediction [14]	Zhao et al. (2025)	MIMIC-IV; 24.237 pasien sepsis ICU	Qwen2.5-72B, LoRA, SFT, SVM, KNN, RF, LSTM, Transformer	Model guideline-enhanced (Ours) mencapai Akurasi 0,819 dan AUC 0,852

Tabel 2.1 menyajikan penelitian-penelitian terdahulu yang menjadi landasan dalam penelitian ini. Berbagai studi menunjukkan bahwa data rekam medis

elektronik (*Electronic Health Records*/EHR), khususnya database MIMIC-III dan MIMIC-IV, telah banyak dimanfaatkan untuk mengembangkan model prediksi luaran klinis pada pasien kritis, seperti mortalitas, sepsis, dan komplikasi organ [23, 22, 35]. Pemanfaatan data tersebut memungkinkan pengembangan model prediksi berbasis *machine learning* dengan cakupan populasi yang besar serta variabel klinis yang beragam.

Dari perspektif metodologi, penelitian terdahulu umumnya menggunakan berbagai algoritma *machine learning*, seperti *Logistic Regression*, *Random Forest*, *Support Vector Machine*, *Artificial Neural Network*, dan XGBoost. Di antara algoritma tersebut, XGBoost secara konsisten menunjukkan performa yang kompetitif dalam memprediksi mortalitas maupun komplikasi pada pasien kritis karena kemampuannya menangani hubungan nonlinier antarvariabel klinis [23, 22]. Selain itu, beberapa penelitian juga menerapkan proses seleksi fitur menggunakan metode seperti Boruta, *Recursive Feature Elimination* (RFE), maupun *stepwise regression* untuk memperoleh kombinasi variabel prediktor yang optimal [23, 22, 35].

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada populasi pasien kritis secara umum atau hanya pada kondisi klinis tertentu, seperti sepsis atau kanker secara terpisah. Penelitian yang secara khusus mengembangkan model prediksi mortalitas pada populasi pasien dengan kanker yang mengalami sepsis masih relatif terbatas, terutama menggunakan data MIMIC-IV dengan pendekatan perbandingan beberapa algoritma *machine learning*.

Berdasarkan tinjauan literatur tersebut, terdapat dua *research gap* utama. Pertama, masih terbatas penelitian yang secara khusus mengembangkan model prediksi mortalitas 28 hari pada pasien kanker dengan sepsis sebagai populasi penelitian utama. Kedua, belum banyak penelitian yang membandingkan performa beberapa algoritma *machine learning* sekaligus mengidentifikasi faktor-faktor klinis yang paling berpengaruh melalui analisis *feature importance* pada populasi tersebut.