

SYSTEMATIC LITERATUR REVIEW PADA DIAGNOSA PENYAKIT JANTUNG MENGGUNAKAN MACHINE LEARNING

Akhsani Takwim, Fitri Handayani

Teknik Informatika, Universitas Muhammadiyah Riau,
Akhsanitakwim03@gmail.com

ABSTRAK

Penyakit jantung adalah salah satu penyebab kematian paling umum di seluruh dunia dan merupakan tantangan besar dalam sektor kesehatan. Inovasi dalam teknologi pembelajaran mesin memberikan kesempatan untuk meningkatkan proses prediksi penyakit jantung dengan cara yang lebih cepat, tepat, dan objektif. Namun, banyaknya algoritma, variabel, dan metode evaluasi yang terlibat dalam penelitian membuatnya sulit untuk menemukan metode yang paling efektif. Tujuan dari penelitian ini adalah untuk menemukan variabel yang paling sering digunakan, membandingkan algoritma pembelajaran mesin yang paling efisien, serta mengevaluasi metode pengujian dan penilaian yang diterapkan dalam prediksi kondisi jantung. Metodologi yang diterapkan adalah Tinjauan Literatur Sistematis (SLR) dengan menganalisa 25 artikel riset yang diterbitkan antara tahun 2023 dan 2026. Langkah-langkah penelitian mencakup identifikasi literatur, pemilihan artikel berdasarkan kriteria inklusi dan eksklusi, ekstraksi data, serta analisis terhadap variabel, algoritma, metode pengujian, dan hasil dari penelitian. Temuan dari kajian ini menunjukkan bahwa usia, tekanan darah, dan kadar kolesterol adalah faktor yang paling umum digunakan karena keterkaitannya yang kuat dengan risiko penyakit jantung. Algoritma Random Forest dan Ensemble Learning muncul sebagai metode dominan, menunjukkan performa konsisten dengan tingkat akurasi yang tinggi. Di samping itu, K-Fold Cross Validation adalah metode pengujian paling banyak dipakai, sedangkan akurasi menjadi metrik evaluasi yang paling sering digunakan.

Kata kunci : *Accuracy, Diagnosa Penyakit Jantung, Machine Learning Usia ,Random Forest, Systematic Literature Review (SLR)*

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia dan menjadi masalah kesehatan global yang terus mengalami peningkatan. Penyakit ini sering kali berkembang tanpa gejala yang jelas pada tahap awal sehingga menyebabkan keterlambatan diagnosis dan penanganan. Faktor risiko seperti usia, hipertensi, kadar kolesterol tinggi, diabetes, obesitas, kebiasaan merokok, serta pola hidup yang tidak sehat berkontribusi terhadap meningkatnya risiko penyakit jantung. Oleh karena itu, diperlukan upaya deteksi dini yang efektif untuk membantu tenaga medis dalam mengidentifikasi individu yang berisiko sehingga tindakan pencegahan dapat dilakukan lebih cepat [1][13][16].

Perkembangan teknologi kecerdasan buatan, khususnya machine learning, telah memberikan alternatif solusi dalam bidang kesehatan untuk mendukung proses prediksi dan pengambilan keputusan. Machine learning mampu mempelajari pola dari data medis dalam jumlah besar dan menghasilkan prediksi yang lebih cepat serta konsisten dibandingkan analisis manual. Berbagai algoritma telah diterapkan untuk prediksi penyakit jantung, seperti Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machine (SVM), Random Forest, dan metode Ensemble Learning [3][4][21],[25].

Hasil penelitian menunjukkan bahwa penerapan machine learning dapat meningkatkan kemampuan identifikasi risiko penyakit jantung dengan tingkat akurasi yang cukup tinggi [5][11][15].

Meskipun demikian, penelitian mengenai prediksi penyakit jantung masih menunjukkan hasil yang beragam. Perbedaan jenis dataset, variabel yang digunakan, teknik preprocessing, metode validasi, serta algoritma yang diterapkan menyebabkan variasi performa yang cukup signifikan antar penelitian [6][9][16].

Beberapa penelitian melaporkan bahwa Random Forest menghasilkan performa terbaik [5][22], sementara penelitian lain menunjukkan bahwa SVM, Logistic Regression, maupun metode ensemble tertentu memberikan hasil yang kompetitif [4][21][25].

Selain itu, penelitian-penelitian tersebut berfokus pada pengembangan dan pengujian model secara individual sehingga belum memberikan gambaran yang komprehensif mengenai tren penggunaan variabel, algoritma yang dominan, metode evaluasi yang sering digunakan,

Berdasarkan tersebut, diperlukan penelitian berbasis Systematic Literature Review (SLR) untuk mengintegrasikan, membandingkan, dan menganalisis temuan dari berbagai penelitian yang telah dipublikasikan. Pendekatan SLR dipilih karena mampu menghasilkan sintesis pengetahuan secara sistematis, objektif, dan terstruktur sehingga dapat mengidentifikasi perkembangan penelitian, kelebihan dan kekurangan metode yang digunakan, serta peluang penelitian di masa mendatang.

Dengan adanya kajian yang komprehensif, peneliti dan praktisi dapat memperoleh gambaran yang lebih jelas mengenai metode machine learning yang

paling banyak digunakan dan berpotensi memberikan performa terbaik dalam prediksi penyakit jantung.

Berdasarkan diatas, penelitian ini bertujuan untuk menganalisis penerapan machine learning dalam prediksi penyakit jantung melalui pendekatan Systematic Literature Review. Analisis dilakukan terhadap artikel yang dipublikasikan pada periode 2023–2026 dengan fokus pada identifikasi variabel yang digunakan, algoritma yang diterapkan, metode pengujian, metrik evaluasi, serta hasil performa yang diperoleh. Hasil penelitian ini diharapkan dapat menjadi referensi bagi peneliti dan praktisi dalam menentukan pendekatan yang lebih efektif untuk pengembangan sistem prediksi penyakit jantung di masa mendatang.

2. TINJAUAN PUSTAKA

2.1. Penyakit Jantung

Penyakit jantung merupakan kelompok gangguan yang memengaruhi fungsi dan struktur jantung, termasuk penyakit jantung koroner, gagal jantung, aritmia, dan gangguan katup jantung. Penyakit ini menjadi salah satu penyebab utama kematian di dunia karena sering kali berkembang tanpa gejala yang signifikan pada tahap awal. Faktor risiko penyakit jantung meliputi usia, tekanan darah tinggi, kadar kolesterol yang tinggi, diabetes, obesitas, kebiasaan merokok, serta kurangnya aktivitas fisik. Oleh karena itu, deteksi dini terhadap risiko penyakit jantung menjadi langkah penting dalam upaya pencegahan dan pengurangan angka kematian akibat penyakit kardiovaskular [1][13][16].

2.2. Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer mempelajari pola dari data dan menghasilkan prediksi tanpa diprogram secara eksplisit. Dalam bidang kesehatan, machine learning banyak dimanfaatkan untuk membantu proses diagnosis, klasifikasi penyakit, prediksi risiko pasien, dan pengambilan keputusan medis. Keunggulan machine learning terletak pada kemampuannya dalam mengolah data yang kompleks, menemukan pola tersembunyi, serta menghasilkan prediksi dengan tingkat akurasi yang tinggi [13][16].

Berdasarkan teknik pembelajarannya, machine learning dibagi menjadi beberapa kategori, salah satunya adalah supervised learning yang banyak digunakan dalam prediksi penyakit jantung. Metode ini memanfaatkan data yang telah memiliki label untuk membangun model prediksi sehingga dapat digunakan untuk mengklasifikasikan data baru berdasarkan pola yang telah dipelajari sebelumnya [3][25].

2.3. Algoritma Machine Learning untuk Prediksi Penyakit Jantung

Berbagai algoritma machine learning telah digunakan dalam penelitian prediksi penyakit jantung.

Algoritma yang sering digunakan antara lain Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Random Forest, dan Ensemble Learning [3][4][21][25].

Random Forest merupakan salah satu algoritma yang paling banyak digunakan karena mampu mengatasi overfitting dan menghasilkan performa yang stabil pada data medis yang kompleks [5][11][22].

Selain itu, Support Vector Machine (SVM) memiliki kemampuan yang baik dalam menangani data berdimensi tinggi dan menghasilkan tingkat klasifikasi yang kompetitif [21].

Logistic Regression banyak digunakan karena sederhana dan mudah diinterpretasikan dalam analisis faktor risiko penyakit jantung [4]. Sementara itu, metode Ensemble Learning mengombinasikan beberapa model untuk meningkatkan akurasi dan stabilitas hasil prediksi [25].

2.4. Evaluasi Kinerja Model

Evaluasi kinerja model merupakan tahapan penting dalam pengembangan sistem prediksi penyakit jantung. Tujuannya adalah untuk mengukur kemampuan model dalam melakukan klasifikasi secara tepat. Beberapa metrik evaluasi yang umum digunakan meliputi Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC) [16][25].

Accuracy merupakan metrik yang paling banyak digunakan karena menunjukkan persentase prediksi yang benar terhadap seluruh data yang diuji. Namun, pada kondisi data yang tidak seimbang, penggunaan Precision, Recall, dan F1-Score menjadi penting untuk memberikan gambaran performa model yang lebih komprehensif [17][24].

2.5. Systematic Literature Review (SLR)

Systematic Literature Review (SLR) merupakan metode penelitian yang digunakan untuk mengidentifikasi, mengevaluasi, dan mensintesis hasil penelitian yang relevan secara sistematis dan terstruktur. Metode ini bertujuan untuk mengurangi bias dalam proses pengumpulan dan analisis literatur sehingga menghasilkan kesimpulan yang lebih objektif. Dalam penelitian ini, SLR digunakan untuk mengkaji perkembangan penerapan machine learning dalam prediksi penyakit jantung, termasuk identifikasi variabel yang digunakan, algoritma yang diterapkan, metode pengujian, serta hasil performa yang diperoleh dari berbagai penelitian yang dipublikasikan pada periode 2023–2026.

3. METODE PENELITIAN

3.1. Proses Seleksi dan Analisis Literatur

Penelitian ini menggunakan metode Systematic Literature Review (SLR) untuk mengidentifikasi, mengevaluasi, dan mensintesis hasil penelitian terkait penerapan machine learning dalam prediksi penyakit jantung. Proses penelitian dilakukan melalui beberapa tahapan, yaitu identifikasi literatur, seleksi artikel

berdasarkan kriteria inklusi dan eksklusi, ekstraksi data, serta analisis data.

3.2. Identifikasi Literatur

Pada tahap identifikasi, artikel dikumpulkan dari berbagai sumber publikasi ilmiah yang relevan, seperti Google Scholar, SINTA, yang membahas penerapan machine learning untuk prediksi penyakit jantung. Kata kunci yang digunakan antara lain Accuracy, Diagnosa Penyakit Jantung, Machine Learning Usia, Random Forest, Systematic Literature Review (SLR).

3.3. Kriteria Inklusi dan Eksklusi

Kriteria inklusi digunakan untuk menentukan artikel yang layak dianalisis, sedangkan kriteria eksklusi digunakan untuk mengeliminasi artikel yang tidak sesuai dengan tujuan penelitian.

a. Kriteria Inklusi:

- Artikel membahas prediksi atau klasifikasi penyakit jantung menggunakan metode machine learning.
- Artikel dipublikasikan pada periode 2023–2026.
- Artikel berasal dari jurnal nasional terakreditasi.
- Artikel menyediakan informasi mengenai algoritma yang digunakan dan hasil evaluasi model.
- Artikel dapat diakses secara penuh (full text).

b. Kriteria Eksklusi:

- Artikel yang tidak membahas penyakit jantung.
- Artikel yang tidak menggunakan metode machine learning.
- Artikel yang hanya berupa review, editorial, atau opini.
- Artikel yang tidak menyajikan hasil pengujian atau metrik evaluasi.
- Artikel duplikat dari sumber yang berbeda.

Melalui proses seleksi tersebut diperoleh 25 artikel yang memenuhi kriteria dan digunakan sebagai sumber utama dalam penelitian ini.

3.4. Ekstraksi Data

Tahap ekstraksi data dilakukan untuk memperoleh informasi penting dari setiap artikel yang telah lolos seleksi. Data yang dikumpulkan meliputi:

- a. Nama penulis dan tahun publikasi.
- b. Dataset yang digunakan.
- c. Variabel atau atribut penelitian.
- d. Algoritma machine learning yang diterapkan.
- e. Metode pengujian model.
- f. Metrik evaluasi yang digunakan.
- g. Hasil performa model.

Data tersebut kemudian disusun dalam bentuk tabel ringkasan penelitian untuk memudahkan proses analisis dan perbandingan antar penelitian.

3.5. Analisis Data

Analisis dilakukan secara deskriptif dengan mengelompokkan informasi yang diperoleh dari setiap artikel berdasarkan beberapa aspek berikut:

a. Analisis Variabel

Analisis dilakukan untuk mengidentifikasi variabel yang paling sering digunakan dalam prediksi penyakit jantung. Variabel seperti usia, tekanan darah, kadar kolesterol, denyut jantung maksimum, dan kadar gula darah dianalisis berdasarkan frekuensi kemunculannya pada penelitian yang dikaji.

b. Analisis Algoritma

Analisis algoritma bertujuan untuk mengetahui metode machine learning yang paling banyak digunakan serta performanya dalam prediksi penyakit jantung. Algoritma yang ditemukan kemudian dikelompokkan, seperti Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, Naive Bayes, dan Ensemble Learning.

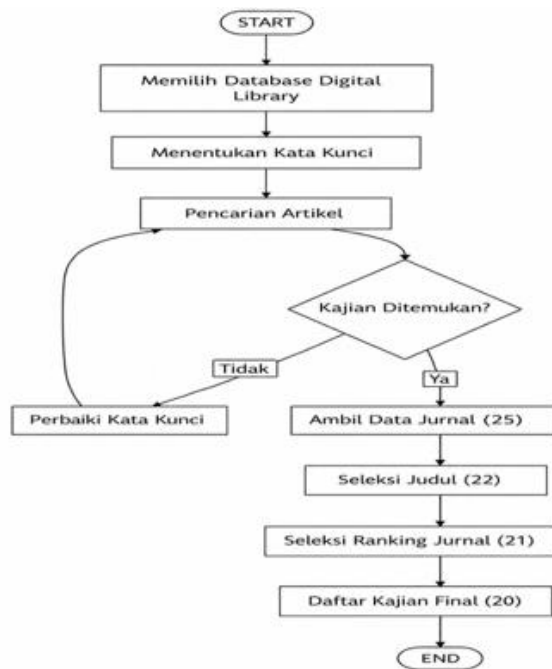
c. Analisis Metode Pengujian

Analisis dilakukan terhadap metode validasi yang digunakan oleh peneliti untuk mengevaluasi model, seperti K-Fold Cross Validation, Holdout Validation, Train-Test Split, dan metode validasi lainnya. Frekuensi penggunaan dan keunggulan masing-masing metode kemudian dibandingkan.

d. Analisis Hasil Penelitian

Analisis hasil penelitian dilakukan dengan membandingkan nilai metrik evaluasi yang digunakan, seperti Accuracy, Precision, Recall, F1-Score, dan AUC. Hasil ini digunakan untuk mengidentifikasi algoritma yang memiliki performa terbaik dan tren perkembangan penelitian prediksi penyakit jantung pada periode 2023–2026. Diagram alir pada gambar berikut menggambarkan rangkaian tahapan yang terstruktur dalam pelaksanaan Systematic Literature Review (SLR) untuk mengkaji penerapan metode machine learning dalam prediksi penyakit jantung. Setiap tahapan dalam alur ini disusun secara sistematis dengan tujuan memastikan bahwa literatur yang digunakan memiliki tingkat relevansi yang tinggi, kualitas akademik yang baik, serta sesuai dengan fokus penelitian.

Selain itu, metode ini juga membantu dalam memahami perkembangan penelitian terkait penggunaan algoritma machine learning dalam mendeteksi penyakit jantung, yang menjadi salah satu penyebab utama kematian sehingga membutuhkan metode prediksi yang akurat. Dengan menggunakan pendekatan yang sistematis seperti PRISMA, proses seleksi literatur dapat dilakukan secara objektif dan terstruktur, sehingga menghasilkan kesimpulan yang lebih valid dan dapat dijadikan dasar dalam pengembangan model prediksi penyakit jantung berbasis machine learning [1].



Gambar 1. Alur Pencarian Dan Seleksi Artikel

Metode penelitian yang digunakan dalam kajian ini adalah Systematic Literature Review (SLR), yaitu pendekatan sistematis untuk mengidentifikasi, mengevaluasi, dan menginterpretasikan seluruh hasil penelitian yang relevan terkait topik prediksi penyakit jantung menggunakan teknik machine learning [2][3].

Tahapan penelitian diawali dengan proses identifikasi sumber literatur melalui pemilihan basis data digital yang kredibel. Dalam penelitian ini, sumber utama yang digunakan adalah Google Scholar dan SINTA, karena keduanya menyediakan akses luas terhadap publikasi ilmiah nasional yang terakreditasi. Pemilihan basis data ini bertujuan untuk menjamin kualitas serta relevansi artikel yang digunakan dalam kajian [4][1].

Selanjutnya, dilakukan perumusan kata kunci pencarian yang disesuaikan dengan fokus penelitian, seperti “prediksi penyakit jantung”, “heart disease prediction”, dan “machine learning”. Kata kunci tersebut dikombinasikan menggunakan operator Boolean (AND, OR) untuk memperoleh hasil pencarian yang lebih spesifik dan relevan. Proses pencarian dilakukan secara iteratif, di mana penyesuaian kata kunci dilakukan apabila hasil yang diperoleh belum memenuhi kebutuhan penelitian [5][6].

Dalam kajian ini, diperoleh sekitar 25 artikel dari hasil pencarian awal. Artikel-artikel tersebut

kemudian melalui tahap seleksi pertama berdasarkan judul dan abstrak, sehingga diperoleh sekitar 22 artikel yang dianggap relevan dengan topik penelitian. Selanjutnya dilakukan seleksi lanjutan dengan mempertimbangkan kualitas jurnal, seperti indeksasi SINTA, tahun publikasi (2023–2026), serta ketersediaan full text. Pada tahap ini, jumlah artikel berkurang menjadi sekitar 21 artikel. Proses ini bertujuan untuk memastikan bahwa literatur yang digunakan memiliki kualitas akademik yang baik dan dapat mendukung analisis secara mendalam [7][8][7][8].

Tahap Sementara itu, kriteria eksklusi mencakup artikel yang tidak relevan, tidak memiliki metode yang jelas, atau tidak menyediakan hasil yang dapat dianalisis. Berdasarkan proses ini, diperoleh sekitar 20 artikel yang digunakan sebagai literatur utama dalam penelitian [9][10].

Seluruh artikel yang telah terpilih kemudian dianalisis secara sistematis untuk mengidentifikasi metode yang digunakan, jenis dataset, algoritma yang diterapkan, serta tingkat akurasi yang dihasilkan. Beberapa algoritma yang banyak digunakan dalam penelitian prediksi penyakit jantung antara lain Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine (SVM), dan Neural Network [11][12][13].

Selain itu, pendekatan SLR dalam penelitian ini juga digunakan untuk memetakan tren penelitian, mengidentifikasi kelebihan dan kekurangan dari masing-masing metode, serta memberikan rekomendasi untuk penelitian selanjutnya [14][15][16].

Secara keseluruhan, metode Systematic Literature Review yang diterapkan dalam penelitian ini memberikan pendekatan yang terstruktur dan transparan dalam proses pengumpulan serta analisis literatur. Hal ini penting untuk memastikan bahwa hasil penelitian memiliki validitas yang tinggi serta dapat dipertanggungjawabkan secara ilmiah [17][18].

4. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil penelitian secara sistematis yang dilengkapi dengan pembahasan analitis berdasarkan tiga pertanyaan penelitian utama (RQ1–RQ3). Pembahasan dilakukan dengan menghubungkan temuan empiris, landasan teori machine learning, serta penelitian terdahulu yang berkaitan dengan prediksi penyakit jantung. Hasil dari proses ekstraksi literatur yang telah lolos tahap penyaringan kemudian disajikan dalam bentuk tabel guna memberikan gambaran yang lebih jelas dan terstruktur.

Tabel 1. Hasil ekstraksi Literatur

No	Judul Penelitian	Tahun	Penulis	Jurnal	Hasil Penelitian
1	Model Machine Learning yang Dioptimalkan untuk Prediksi Penyakit Jantung Menggunakan R Shiny	2026	Y. Dewi Amritha dkk.	Jurnal Teknologi Informasi	Model machine learning yang dioptimalkan dan diimplementasikan melalui R Shiny mampu meningkatkan kemudahan penggunaan sistem prediksi penyakit jantung serta menyajikan visualisasi hasil yang interaktif.

No	Judul Penelitian	Tahun	Penulis	Jurnal	Hasil Penelitian
2	Analisis Klasifikasi Risiko Penyakit Jantung Menggunakan Metode Random Forest	2026	A. Damayanti dkk.	Progresif	Random Forest menunjukkan kemampuan yang baik dalam mengklasifikasikan risiko penyakit jantung dengan tingkat akurasi yang tinggi dan stabil.
3	Prediksi Risiko Penyakit Jantung Sederhana Menggunakan Algoritma Random Forest Classifier Dengan Data Gaya Hidup Siswa	2025	M. Santoso dkk.	Jurnal Manajemen Informatika Jayakarta	Random Forest mampu memprediksi risiko penyakit jantung berdasarkan data gaya hidup siswa dengan tingkat akurasi yang baik.
4	Prediksi Risiko Penyakit Jantung Sederhana Menggunakan Algoritma Random Forest Classifier Dengan Data Gaya Hidup Siswa	2025	M. Santoso dkk.	Jurnal Manajemen Informatika Jayakarta	Random Forest mampu mengidentifikasi risiko penyakit jantung berdasarkan data gaya hidup siswa dengan hasil klasifikasi yang baik.
5	Optimalisasi Preprocessing Data Menggunakan Pendekatan CRISP-DM untuk Meningkatkan Kualitas Klasifikasi Penyakit Jantung	2025	W. Nursahid dkk.	RIGGS Journal	Pendekatan CRISP-DM pada tahap preprocessing berhasil meningkatkan kualitas data dan performa model klasifikasi penyakit jantung.
6	Analisis Prediksi Penyakit Jantung Menggunakan Perbandingan Algoritma Machine Learning	2025	L. Anjeli dan F. Rizki	Jurnal Penelitian Informatika	Hasil perbandingan menunjukkan bahwa algoritma ensemble memiliki performa yang lebih baik dibandingkan beberapa algoritma machine learning lainnya.
7	Prediksi Penyakit Jantung Menggunakan Metode Random Forest	2025	S. P. Penyakit dkk.	JUITIK	Random Forest memberikan hasil klasifikasi yang efektif dalam mendeteksi risiko penyakit jantung.
8	Peningkatan Akurasi Prediksi Penyakit Jantung dengan Teknik SMOTEENN pada Algoritma Random Forest	2025	A. Rahmada dan E. R. Susanto	Jurnal Pendidikan dan Teknologi Indonesia	Teknik SMOTEENN berhasil mengatasi ketidakseimbangan data sehingga meningkatkan akurasi model Random Forest.
9	Penerapan Machine Learning untuk Memprediksi Risiko Pengidap Penyakit Jantung Menggunakan Algoritma Decision Tree	2025	I. Sabily Karima	Jurnal Informatika	Decision Tree menghasilkan model prediksi yang mudah dipahami dengan tingkat akurasi yang cukup baik.
10	Prediksi Penyakit Gagal Jantung Tahun 2025–2028 Rawat Inap di RSUD Gambiran Kota Kediri	2025	Anisa Rahmawati dkk.	Antigen	Penelitian menghasilkan prediksi tren pasien gagal jantung yang dapat digunakan sebagai dasar perencanaan layanan kesehatan rumah sakit.
11	Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung	2024	S. A. Putri dkk.	Seminar Nasional Sains Data	Random Forest menghasilkan akurasi yang tinggi dalam memprediksi penyakit jantung dan mampu mengurangi overfitting.
12	Sistem Artificial Intelligence Deteksi Penyakit THT dan Jantung Menggunakan Forward Chaining dan Image Processing	2024	A. Purno dkk.	Jurnal Sistem Cerdas	Kombinasi Forward Chaining dan Image Processing mampu mendukung deteksi dini penyakit jantung secara otomatis.
13	Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine	2024	R. Hidayat dkk.	BIOS	Algoritma SVM menunjukkan performa yang baik dalam mengklasifikasikan pasien yang berisiko penyakit jantung.
14	Prediksi Penyakit Jantung Menggunakan Metode Random Forest dan Penerapan PCA	2024	N. H. Alfajr dan S. Defiyanti	JITET	PCA berhasil mereduksi dimensi data dan meningkatkan efisiensi model Random Forest.

No	Judul Penelitian	Tahun	Penulis	Jurnal	Hasil Penelitian
15	Optimalisasi Klasifikasi Penderita Penyakit Jantung Koroner dalam Sepuluh Tahun Ke Depan Menggunakan Regresi Logistik, Decision Tree dan Random Forest	2024	I. Hidayatul Imania dkk.	Jurnal Teknologi Informasi	Random Forest menghasilkan akurasi tertinggi dibandingkan Regresi Logistik dan Decision Tree.
16	Optimasi Pemilihan Fitur untuk Prediksi Penyakit Jantung Menggunakan Algoritma Genetika dan Random Forest	2024	T. Gori dan A. Hestiningtyas	Indonesian Journal of Computer Science	Algoritma Genetika mampu meningkatkan akurasi Random Forest melalui seleksi fitur yang optimal.
17	Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier	2023	A. Wijayadhi, M. Makmun, dan S. B. Rahardjo	Bulletin of Information Technology (BIT)	Regresi Linier mampu digunakan untuk memprediksi risiko penyakit jantung berdasarkan faktor-faktor kesehatan pasien.
18	Analisis Kinerja Algoritma Machine Learning untuk Prediksi Penyakit Jantung Menggunakan Metode Data Preprocessing Terintegrasi	2023	A. Rohman dan S. Mujiyono	Jurnal Informatika	Preprocessing data terintegrasi meningkatkan performa berbagai algoritma machine learning dalam prediksi penyakit jantung.
19	Prediksi Penyakit Jantung Menggunakan Machine Learning	2023	A. Hidayat	Jurnal Informatika	Machine learning terbukti dapat membantu proses diagnosis penyakit jantung secara lebih cepat dan akurat.
20	Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung	2023	D. H. Depari dkk.	Jurnal Informatik	Random Forest memperoleh performa terbaik dibandingkan Decision Tree dan Naive Bayes.

4.1. RQ1: Variabel apa saja yang paling sering digunakan dalam penelitian Diagnosa Penyakit Jantung ?

Berdasarkan hasil kajian dari beberapa jurnal yang Berdasarkan 20 artikel yang telah dianalisis, ditemukan bahwa keberhasilan model machine learning dalam mendiagnosis atau memprediksi penyakit jantung sangat dipengaruhi oleh variabel yang digunakan dalam dataset. Variabel tersebut berfungsi sebagai dasar bagi model untuk mengenali pola dan hubungan antar data sehingga dapat menentukan apakah seseorang berisiko mengalami penyakit jantung atau tidak. Semakin relevan dan berkualitas variabel yang digunakan, maka semakin baik pula performa model yang dihasilkan [1][8][13].

Salah satu variabel yang paling sering digunakan adalah usia. Hampir seluruh penelitian menjadikan usia sebagai variabel utama karena risiko penyakit jantung cenderung meningkat seiring bertambahnya usia seseorang. Pertambahan usia dapat menyebabkan penurunan fungsi jantung dan pembuluh darah sehingga meningkatkan kemungkinan terjadinya penyakit jantung [1][13][17].

Selain usia, jenis kelamin juga menjadi variabel yang banyak digunakan. Beberapa penelitian menunjukkan bahwa laki-laki memiliki risiko lebih tinggi mengalami penyakit jantung dibandingkan perempuan pada kelompok usia tertentu. Oleh karena itu, variabel jenis kelamin sering dimasukkan sebagai atribut penting dalam proses klasifikasi [13][14][16].

Tekanan darah merupakan variabel lain yang sangat sering digunakan dalam penelitian diagnosis

penyakit jantung. Tekanan darah tinggi atau hipertensi dapat meningkatkan beban kerja jantung dan mempercepat kerusakan pembuluh darah. Kondisi tersebut menjadikan hipertensi sebagai salah satu faktor risiko utama penyakit jantung sehingga banyak digunakan sebagai variabel prediktor [5][9][18].

Selanjutnya, kadar kolesterol juga termasuk variabel yang dominan digunakan. Tingginya kadar kolesterol dalam darah dapat menyebabkan terbentuknya plak pada pembuluh darah yang menghambat aliran darah menuju jantung. Oleh karena itu, variabel kolesterol sering dijadikan indikator penting dalam proses diagnosis maupun prediksi penyakit jantung [1][14][19].

Variabel jenis nyeri dada (chest pain type) merupakan salah satu atribut yang paling berpengaruh dalam berbagai model machine learning. Karakteristik nyeri dada yang dialami pasien dapat memberikan informasi penting mengenai kondisi jantung sehingga sering digunakan untuk meningkatkan akurasi klasifikasi [4][13][16].

Selain itu, denyut jantung maksimum (maximum heart rate achieved) dan exercise induced angina juga banyak digunakan dalam penelitian. Kedua variabel tersebut berkaitan dengan kemampuan jantung saat melakukan aktivitas fisik. Nilai yang tidak normal pada kedua atribut ini sering dikaitkan dengan adanya gangguan fungsi jantung dan menjadi indikator penting dalam proses diagnosis [13][14][17].

Beberapa penelitian terbaru juga mulai memasukkan faktor gaya hidup seperti kebiasaan merokok, aktivitas fisik, obesitas, diabetes, dan

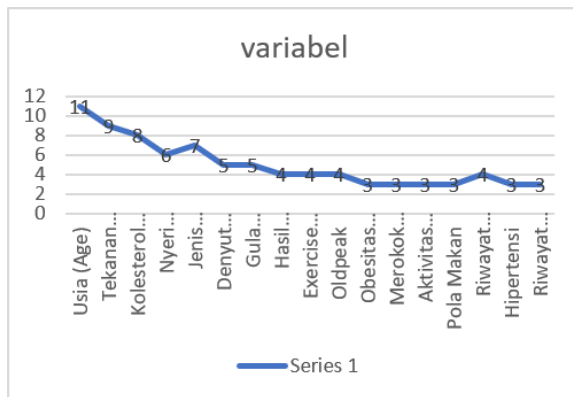
hipertensi sebagai variabel tambahan. Faktor-faktor tersebut terbukti memiliki hubungan yang kuat terhadap peningkatan risiko penyakit jantung sehingga mampu meningkatkan kemampuan model dalam melakukan prediksi secara lebih komprehensif [2][18][20].

Berdasarkan hasil kajian literatur, dapat disimpulkan bahwa variabel yang paling sering digunakan dalam penelitian diagnosis penyakit

jantung adalah usia, jenis kelamin, tekanan darah, kadar kolesterol, jenis nyeri dada, denyut jantung maksimum, dan kadar gula darah. Variabel-variabel tersebut secara konsisten muncul dalam berbagai penelitian dan terbukti memiliki kontribusi yang signifikan terhadap peningkatan akurasi model machine learning dalam mendiagnosis penyakit jantung [1][5][13][14][19].

Tabel 2 : Variabel yang paling sering digunakan dalam penelitian penyakit jantung

Kategori	Variabel / Faktor	Keterangan	Sitasi
Variabel Klinis Utama	Usia (Age)	Risiko penyakit jantung meningkat seiring bertambahnya usia akibat penurunan fungsi jantung dan pembuluh darah.	[1], [4], [5], [6], [8], [13], [14], [16], [17], [18], [19]
	Tekanan Darah (Blood Pressure)	Hipertensi meningkatkan beban kerja jantung dan berkontribusi terhadap penyakit kardiovaskular.	[1], [5], [8], [13], [14], [16], [17], [18], [19]
	Kolesterol (Cholesterol)	Kadar kolesterol yang tinggi dapat menyebabkan penyumbatan pembuluh darah dan meningkatkan risiko penyakit jantung.	[1], [5], [8], [13], [14], [16], [17], [19]
	Nyeri Dada (Chest Pain)	Merupakan gejala utama yang sering digunakan sebagai indikator awal penyakit jantung koroner.	[4], [13], [14], [16], [17], [19]
	Jenis Kelamin (Sex)	Digunakan untuk mengidentifikasi perbedaan tingkat risiko penyakit jantung antara laki-laki dan perempuan.	[4], [5], [13], [14], [16], [17], [19]
Variabel Medis Tambahan	Denyut Jantung Maksimum (Maximum Heart Rate)	Menggambarkan kemampuan jantung saat melakukan aktivitas fisik dan sering digunakan dalam proses diagnosis.	[13], [14], [16], [17], [19]
	Gula Darah (Fasting Blood Sugar)	Kadar gula darah yang tinggi berkaitan dengan diabetes yang merupakan faktor risiko penyakit jantung.	[1], [13], [14], [16], [17]
	Hasil Elektrokardiogram (ECG)	Digunakan untuk mendeteksi gangguan irama dan aktivitas listrik jantung.	[13], [14], [16], [17]
	Exercise Induced Angina	Nyeri dada yang muncul saat aktivitas fisik dan menjadi indikator adanya gangguan jantung.	[13], [14], [16], [17]
	Oldpeak	Menunjukkan depresi segmen ST setelah aktivitas fisik yang berkaitan dengan penyakit jantung.	[14], [16], [17], [19]
Variabel Gaya Hidup	Obesitas (BMI)	Berat badan berlebih berhubungan dengan hipertensi, diabetes, dan peningkatan risiko penyakit jantung.	[2], [18], [20]
	Merokok (Smoking)	Kebiasaan merokok dapat merusak pembuluh darah dan meningkatkan risiko penyakit jantung.	[2], [18], [20]
	Aktivitas Fisik	Kurangnya aktivitas fisik dapat meningkatkan risiko gangguan kardiovaskular.	[2], [18], [20]
	Pola Makan	Konsumsi makanan tinggi lemak, kolesterol, dan garam dapat memicu penyakit jantung.	[2], [18], [20]
Riwayat Penyakit dan Genetik	Riwayat Diabetes	Diabetes merupakan faktor risiko yang dapat mempercepat terjadinya komplikasi penyakit jantung.	[2], [9], [18], [20]
	Hipertensi	Riwayat tekanan darah tinggi merupakan salah satu faktor risiko utama penyakit jantung.	[5], [9], [18]
	Riwayat Keluarga	Faktor genetik dan riwayat keluarga memengaruhi kemungkinan seseorang mengalami penyakit jantung.	[2], [18], [20]



Gambar 2. Grafik Variabel Yang Sering Digunakan

4.2. RQ2: Metode machine learning apa saja yang paling banyak digunakan dan bagaimana karakteristik kinerja dalam Diagnosa Penyakit Jantung?

Berdasarkan hasil kajian dari 20 penelitian yang telah dianalisis, ditemukan bahwa terdapat beberapa metode machine learning yang paling sering digunakan dalam diagnosis dan prediksi penyakit jantung. Pemilihan metode tersebut didasarkan pada kemampuannya dalam mengolah data kesehatan pasien, melakukan klasifikasi risiko penyakit jantung, serta menghasilkan tingkat akurasi yang baik. Metode yang paling banyak ditemukan dalam penelitian adalah Random Forest, Decision Tree, Support Vector Machine (SVM), Regresi Logistik, dan Regresi Linier [1][5][10][13][16][17][18].

Random Forest merupakan metode yang paling dominan digunakan dalam penelitian diagnosis penyakit jantung. Algoritma ini bekerja dengan membangun banyak pohon keputusan (decision tree) dan menggabungkan hasil prediksi dari setiap pohon untuk memperoleh keputusan akhir. Random Forest memiliki keunggulan dalam menangani data yang kompleks, mengurangi risiko overfitting, serta menghasilkan akurasi yang tinggi dan stabil. Oleh karena itu, metode ini menjadi pilihan utama dalam berbagai penelitian diagnosis penyakit jantung [2][5][7][9][14][16][18][19][20].

Selain Random Forest, Decision Tree juga cukup sering digunakan karena memiliki struktur yang sederhana dan mudah dipahami. Algoritma ini mampu menampilkan proses pengambilan keputusan secara visual sehingga memudahkan interpretasi hasil oleh peneliti maupun tenaga medis. Meskipun demikian, Decision Tree memiliki kelemahan yaitu rentan mengalami overfitting apabila digunakan pada dataset yang kompleks atau berukuran besar [10][16][17].

Support Vector Machine (SVM) merupakan metode lain yang banyak digunakan dalam penelitian penyakit jantung. Algoritma ini bekerja dengan

mencari hyperplane terbaik untuk memisahkan data pasien ke dalam kelas yang berbeda. SVM dikenal mampu menghasilkan akurasi yang tinggi, terutama pada dataset yang memiliki pemisahan kelas yang jelas. Namun, metode ini membutuhkan proses tuning parameter yang cukup kompleks dan waktu komputasi yang lebih besar dibandingkan beberapa algoritma lainnya [13].

Selanjutnya, Regresi Logistik (Logistic Regression) masih sering digunakan karena cocok untuk permasalahan klasifikasi biner, yaitu menentukan apakah pasien memiliki penyakit jantung atau tidak. Metode ini memiliki keunggulan dalam interpretasi hasil yang mudah dipahami dan mampu menunjukkan hubungan antara variabel risiko dengan kemungkinan terjadinya penyakit jantung. Akan tetapi, performanya dapat menurun ketika data memiliki hubungan yang kompleks dan nonlinier [17].

Selain itu, beberapa penelitian juga menggunakan Regresi Linier untuk menganalisis hubungan antara faktor-faktor risiko dan kondisi penyakit jantung. Metode ini relatif sederhana dan mudah diterapkan, namun kurang optimal jika digunakan untuk kasus klasifikasi dibandingkan algoritma machine learning lainnya [1].

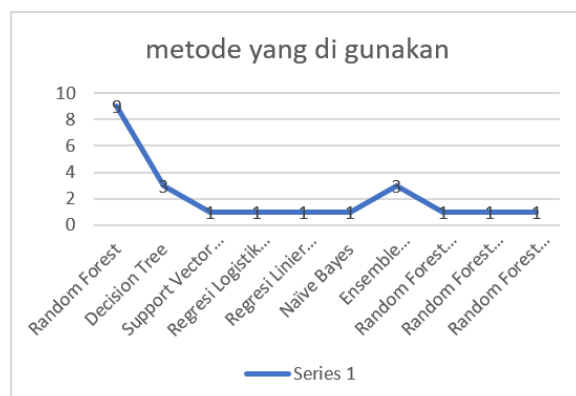
Penelitian terbaru menunjukkan bahwa performa model diagnosis penyakit jantung dapat ditingkatkan melalui penerapan teknik optimasi seperti Principal Component Analysis (PCA), SMOTEENN, Algoritma Genetika, dan CRISP-DM pada tahap preprocessing data. Teknik-teknik tersebut terbukti mampu meningkatkan kualitas data dan akurasi model yang digunakan [3][9][14][19].

Selain model tunggal, beberapa penelitian juga mulai menerapkan pendekatan ensemble learning, terutama melalui Random Forest. Pendekatan ini mampu menghasilkan prediksi yang lebih stabil dan akurat dibandingkan model tunggal karena menggabungkan hasil dari beberapa model secara bersamaan. Oleh karena itu, metode ensemble menjadi salah satu pendekatan yang paling menjanjikan dalam diagnosis penyakit jantung [5][16][18].

Berdasarkan hasil kajian literatur, dapat disimpulkan bahwa Random Forest merupakan metode machine learning yang paling banyak digunakan dan memiliki karakteristik kinerja terbaik dalam diagnosis penyakit jantung. Hal ini didukung oleh kemampuannya dalam menghasilkan akurasi tinggi, menangani data yang kompleks, serta mengurangi risiko overfitting. Sementara itu, Decision Tree unggul dalam interpretabilitas, SVM memiliki kemampuan klasifikasi yang baik pada data berdimensi tinggi, dan Regresi Logistik lebih mudah diinterpretasikan dalam menjelaskan hubungan antar faktor risiko penyakit jantung [2][5][13][16][17][18].

Tabel 3 : Metode machine learning yang paling banyak digunakan dan bagaimana karakteristik kinerja dalam penyakit jantung

No	Metode Machine Learning	Karakteristik Kinerja	Kelebihan	Kekurangan	Sitasi
1	Random Forest	Memiliki akurasi tinggi, stabil, dan mampu menangani data kompleks serta jumlah fitur yang banyak.	Mengurangi overfitting, akurasi tinggi, mampu menangani data besar dan data tidak seimbang.	Membutuhkan waktu komputasi lebih besar dibanding Decision Tree tunggal dan interpretasi model lebih sulit.	[2], [5], [7], [9], [14], [16], [18], [19], [20]
2	Decision Tree	Mampu melakukan klasifikasi dengan struktur pohon keputusan yang mudah dipahami.	Mudah diinterpretasikan, visualisasi sederhana, cepat dalam proses pelatihan.	Rentan mengalami overfitting dan sensitif terhadap perubahan data.	[10], [16], [17]
3	Support Vector Machine (SVM)	Menghasilkan akurasi tinggi terutama pada data dengan pemisahan kelas yang jelas.	Efektif untuk data berdimensi tinggi, memiliki kemampuan generalisasi yang baik.	Membutuhkan tuning parameter yang kompleks dan waktu komputasi lebih lama.	[13]
4	Regresi Logistik (Logistic Regression)	Cocok untuk klasifikasi biner dan menghasilkan probabilitas risiko penyakit jantung.	Mudah dipahami, hasil dapat dijelaskan secara statistik, proses pelatihan cepat.	Kurang optimal untuk hubungan data yang kompleks dan nonlinier.	[17]
5	Regresi Linier (Linear Regression)	Digunakan untuk menganalisis hubungan antara variabel risiko dan penyakit jantung.	Sederhana, mudah diterapkan dan diinterpretasikan.	Kurang sesuai untuk permasalahan klasifikasi dibanding algoritma machine learning lainnya.	[1]
6	Naïve Bayes	Menggunakan pendekatan probabilitas berdasarkan Teorema Bayes untuk klasifikasi data.	Cepat, sederhana, dan efektif pada dataset kecil hingga menengah.	Mengasumsikan setiap variabel independen sehingga kurang sesuai dengan data medis yang saling berkaitan.	[16]
7	Ensemble Learning	Menggabungkan beberapa model machine learning untuk menghasilkan prediksi yang lebih baik.	Akurasi lebih tinggi, stabil, dan mengurangi kesalahan model tunggal.	Membutuhkan sumber daya komputasi lebih besar dan proses pelatihan lebih lama.	[5], [16], [18]
8	Random Forest + PCA	Mengombinasikan Random Forest dengan reduksi dimensi menggunakan PCA.	Mengurangi jumlah fitur, meningkatkan efisiensi dan akurasi model.	Berpotensi menghilangkan informasi penting jika reduksi dimensi tidak optimal.	[14]
9	Random Forest + SMOTEENN	Menggunakan teknik penyeimbangan data sebelum proses klasifikasi.	Meningkatkan akurasi pada dataset tidak seimbang dan mengurangi bias kelas mayoritas.	Membutuhkan proses preprocessing tambahan dan waktu komputasi lebih tinggi.	[9]
10	Random Forest + Algoritma Genetika	Mengoptimalkan pemilihan fitur sebelum proses klasifikasi.	Meningkatkan akurasi dan menghilangkan fitur yang tidak relevan.	Proses optimasi memerlukan waktu komputasi yang lebih lama.	[19]



Gambar 3. Grafik Metode machine learning yang paling banyak digunakan

4.3. RQ3: metode pengujian apa yang di gunakan untuk menilai kinerja model machine learning dalam Diagnosa Penyakit Jantung ?

Berdasarkan hasil kajian terhadap 20 penelitian yang membahas diagnosis dan prediksi penyakit jantung, metode pengujian digunakan untuk mengetahui seberapa baik model machine learning mampu mengklasifikasikan atau memprediksi penyakit jantung pada data yang belum pernah digunakan sebelumnya. Proses pengujian menjadi

tahapan penting karena menentukan tingkat keandalan model sebelum diterapkan pada kasus nyata. Model yang baik tidak hanya memiliki performa tinggi pada data pelatihan, tetapi juga mampu memberikan hasil yang konsisten pada data baru [1][13].

Salah satu metode pengujian yang paling sering digunakan adalah Train-Test Split. Pada metode ini, dataset dibagi menjadi dua bagian, yaitu data pelatihan (training set) dan data pengujian (testing set). Umumnya pembagian data dilakukan dengan rasio 80:20 atau 70:30. Data pelatihan digunakan untuk membangun model machine learning, sedangkan data pengujian digunakan untuk mengevaluasi performa model setelah proses pelatihan selesai. Metode ini banyak digunakan karena sederhana, mudah diterapkan, dan memiliki waktu komputasi yang relatif cepat [1][5][13].

Selain Train-Test Split, beberapa penelitian menggunakan K-Fold Cross Validation untuk memperoleh hasil evaluasi yang lebih stabil. Pada metode ini, dataset dibagi menjadi beberapa bagian (fold), kemudian setiap fold digunakan secara bergantian sebagai data uji sementara fold lainnya digunakan sebagai data latih. Hasil akhir diperoleh dari rata-rata seluruh pengujian. Metode ini mampu mengurangi bias akibat pembagian data dan menghasilkan evaluasi yang lebih representatif dibandingkan Train-Test Split [14][16][17].

Pada penelitian yang menggunakan dataset dengan distribusi kelas yang tidak seimbang, digunakan Stratified K-Fold Cross Validation. Metode ini bekerja dengan menjaga proporsi jumlah data pada setiap kelas agar tetap sama di setiap fold. Teknik ini sangat penting dalam penelitian medis karena jumlah pasien yang menderita penyakit jantung sering kali tidak seimbang dengan jumlah pasien sehat. Dengan demikian, hasil evaluasi yang diperoleh menjadi lebih objektif dan akurat [9][18].

Beberapa penelitian juga menerapkan teknik Repeated Cross Validation, yaitu melakukan proses cross validation berulang kali untuk memperoleh hasil evaluasi yang lebih konsisten. Selain itu, terdapat penelitian yang memanfaatkan teknik Leave-One-Out Cross Validation (LOOCV) pada dataset yang

berukuran kecil. Pada metode ini, satu data digunakan sebagai data uji dan seluruh data lainnya digunakan sebagai data latih secara bergantian hingga seluruh data selesai diuji. Meskipun mampu memaksimalkan penggunaan data, metode ini membutuhkan waktu komputasi yang lebih tinggi dibandingkan metode lainnya [16][17].

Selain metode pembagian data, penilaian kinerja model juga dilakukan menggunakan Confusion Matrix. Confusion Matrix merupakan alat evaluasi yang paling banyak digunakan untuk melihat jumlah prediksi yang benar dan salah berdasarkan kategori True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari Confusion Matrix dapat dihitung berbagai metrik evaluasi yang digunakan untuk mengukur performa model [5][13][16].

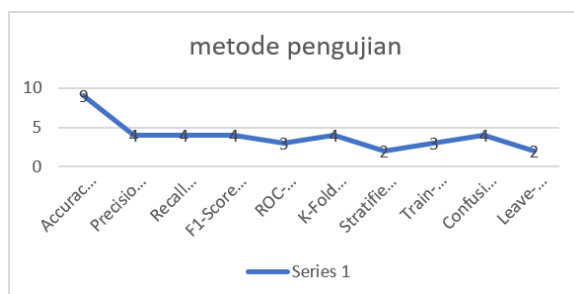
Metrik evaluasi yang paling sering digunakan adalah Accuracy, yaitu persentase prediksi yang benar dibandingkan seluruh data yang diuji. Selain accuracy, penelitian juga menggunakan Precision untuk mengukur ketepatan prediksi positif, Recall (Sensitivity) untuk mengukur kemampuan model dalam mendeteksi pasien yang benar-benar menderita penyakit jantung, serta F1-Score yang merupakan kombinasi antara precision dan recall. Pada beberapa penelitian juga digunakan ROC Curve dan Area Under Curve (AUC) untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif secara keseluruhan [5], [9][13][18].

Berdasarkan hasil kajian literatur, dapat disimpulkan bahwa Train-Test Split, K-Fold Cross Validation, dan Confusion Matrix merupakan metode pengujian yang paling banyak digunakan dalam penelitian diagnosis penyakit jantung. Sementara itu, Accuracy, Precision, Recall, F1-Score, dan AUC menjadi metrik evaluasi utama yang digunakan untuk menilai kinerja model machine learning. Penggunaan kombinasi metode validasi dan metrik evaluasi tersebut bertujuan untuk memastikan bahwa model yang dikembangkan memiliki tingkat akurasi dan keandalan yang tinggi dalam mendiagnosis penyakit jantung [5], [9][13][14] [16][18].

Tabel 4 : metode pengujian yang di gunakan untuk menilai kinerja model machine learning dalam penyakit jantung

No	Metode Pengujian	Cara Kerja	Kelebihan	Kekurangan	Sitasi
1	Accuracy Testing	Mengukur persentase jumlah prediksi yang benar dibandingkan seluruh data pengujian.	Mudah dipahami dan menjadi metrik evaluasi yang paling umum digunakan.	Kurang representatif pada dataset yang tidak seimbang.	[1], [5], [9], [13], [14], [16], [17], [18], [19]
2	Precision Testing	Mengukur ketepatan model dalam memprediksi kelas positif.	Mengurangi kesalahan False Positive (FP).	Tidak menunjukkan kemampuan mendeteksi seluruh pasien yang sakit.	[9], [13], [18], [19]

No	Metode Pengujian	Cara Kerja	Kelebihan	Kekurangan	Sitasi
3	Recall (Sensitivity) Testing	Mengukur kemampuan model dalam mendeteksi pasien yang benar-benar memiliki penyakit jantung.	Sangat penting dalam bidang medis karena meminimalkan False Negative (FN).	Dapat meningkatkan jumlah False Positive.	[9], [13], [18], [19]
4	F1-Score Testing	Menggabungkan nilai Precision dan Recall menjadi satu ukuran evaluasi.	Cocok digunakan pada dataset yang tidak seimbang.	Interpretasinya lebih kompleks dibanding accuracy.	[9], [13], [18], [19]
5	ROC-AUC Evaluation	Mengukur kemampuan model membedakan kelas positif dan negatif pada berbagai threshold.	Memberikan gambaran performa model secara keseluruhan.	Membutuhkan pemahaman statistik yang lebih mendalam.	[13], [18], [19]
6	K-Fold Cross Validation	Dataset dibagi menjadi beberapa fold dan dilakukan pengujian secara bergantian.	Hasil evaluasi lebih stabil dan memanfaatkan seluruh data secara optimal.	Membutuhkan waktu komputasi lebih lama.	[14], [16], [17], [19]
7	Stratified K-Fold Cross Validation	Membagi data dengan mempertahankan proporsi kelas pada setiap fold.	Cocok untuk dataset medis yang tidak seimbang.	Proses implementasi lebih kompleks dibanding K-Fold biasa.	[9], [18]
8	Train-Test Split (Holdout Validation)	Dataset dibagi menjadi data latih dan data uji dengan rasio tertentu (misalnya 80:20 atau 70:30).	Sederhana, cepat, dan mudah diterapkan.	Hasil evaluasi dapat dipengaruhi oleh pembagian data.	[1], [5], [13]
9	Confusion Matrix	Menghitung jumlah TP, TN, FP, dan FN untuk mengevaluasi hasil klasifikasi.	Memberikan informasi detail mengenai kesalahan prediksi model.	Memerlukan metrik tambahan agar interpretasi lebih lengkap.	[5], [13], [16], [18], [19]
10	Leave-One-Out Cross Validation (LOOCV)	Satu data digunakan sebagai data uji dan sisanya sebagai data latih secara bergantian.	Memaksimalkan penggunaan data, terutama pada dataset kecil.	Waktu komputasi sangat tinggi pada dataset besar.	[16], [17]



Gambar 4. Grafik Metode Pengujian

5. KESIMPULAN

Berdasarkan hasil Systematic Literature Review (SLR) yang telah dilakukan terhadap berbagai penelitian mengenai prediksi penyakit jantung menggunakan metode machine learning, dapat disimpulkan bahwa teknologi machine learning memiliki potensi yang sangat besar dalam membantu proses deteksi dini penyakit jantung. Dengan memanfaatkan data pasien, sistem dapat memprediksi kemungkinan seseorang mengalami penyakit jantung secara lebih cepat, akurat, dan efisien dibandingkan proses manual. Melalui proses kajian literatur yang sistematis, penelitian ini berhasil menjawab tiga pertanyaan penelitian utama (RQ1–RQ3) yang berkaitan dengan variabel paling berpengaruh, metode machine learning yang paling sering digunakan, serta

metode pengujian yang digunakan untuk menilai kinerja model.

Variabel yang paling dominan dan paling sering digunakan dalam penelitian prediksi penyakit jantung adalah usia tekanan darah dan kolesterol. Variabel ini menjadi faktor utama karena memiliki hubungan yang sangat kuat terhadap peningkatan risiko penyakit jantung, di mana semakin bertambah usia seseorang maka risiko terjadinya gangguan kardiovaskular juga semakin tinggi.

Oleh karena itu, usia, tekanan dara dan kolesterol menjadi variabel kunci dalam hampir seluruh model prediksi. Metode machine learning yang paling sering dan dominan digunakan adalah Random Forest. Algoritma ini menunjukkan performa yang paling stabil dan memiliki tingkat akurasi yang tinggi dibandingkan metode lainnya.

Kemampuannya dalam menangani data dengan banyak variabel serta mengurangi risiko overfitting menjadikan Random Forest sebagai metode utama dalam prediksi penyakit jantung. Metode pengujian yang paling sering digunakan adalah K-Fold Cross Validation, karena mampu memberikan evaluasi yang lebih stabil dan memanfaatkan data secara optimal. Sementara itu, metrik evaluasi yang paling dominan digunakan untuk mengukur kinerja model adalah accuracy, karena mudah dipahami dan menunjukkan tingkat ketepatan prediksi secara keseluruhan.

DAFTAR PUSTAKA

- [1] L. ' Lu' Anjeli and F. Rizki, "Analisis prediksi penyakit jantung menggunakan perbandingan algoritma machine learning analysis of heart disease prediction using a comparative study of machine learning," vol. 4, no. 2, 2025, [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/heart-failure->
- [2] A. Wijayadhi, M. Makmun, and S. B. Rahardjo, "Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 15–28, 2023, doi: 10.47065/bit.v3i1.
- [3] M. Santoso, M. Zydan K, I. Iwan, and M. Mudrika, "Prediksi Risiko Penyakit Jantung Sederhana Menggunakan Algoritma Random Forest Classifier Dengan Data Gaya Hidup Siswa," *Jurnal Manajemen Informatika Jayakarta*, vol. 5, no. 4, p. 311, Dec. 2025, doi: 10.52362/jmijayakarta.v5i4.2102.
- [4] W. Nursahid, B. I. Nugroho, and S. Syefudin, "Optimalisasi Preprocessing Data Menggunakan Pendekatan CRISP-DM untuk Meningkatkan Kualitas Klasifikasi Penyakit Jantung," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 3621–3626, Aug. 2025, doi: 10.31004/riggs.v4i3.2514.
- [5] S. A. Putri, N. Selayanti, M. Kristanaya, M. P. Azzahra, M. G. Navsih, and K. M. Hindrayani, "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung," *Seminar Nasional Sains Data*, vol. 2024, 2024, [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.
- [6] Y. Dewi Amritha, N. Luh, P. I. Candrawengi, W. Putra Dananjaya, and A. R. Dayanti, "Model Machine Learning yang Dioptimalkan untuk Prediksi Penyakit Jantung Menggunakan R Shiny," 2026.
- [7] S. P. Penyakit *et al.*, "Jurnal Ilmiah Teknik Informatika dan Komunikasi Menggunakan Metode Random Forest," 2025, doi: 10.55606/juitik.v5i1.1330.
- [8] A. Rohman and S. Mujiyono, "Analisis Kinerja Algoritma Machine Learning untuk Prediksi Penyakit Jantung Menggunakan Metode Data Preprocessing Terintegrasi," 2023.
- [9] A. Rahmada and E. R. Susanto, "Peningkatan Akurasi Prediksi Penyakit Jantung dengan Teknik SMOTEENN pada Algoritma Random Forest," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 12, pp. 795–803, Jan. 2025, doi: 10.52436/1.jpti.524.
- [10] I. Sabily Karima, "Penerapan Machine Learning untuk memprediksi Resiko Pengidap Penyakit Jantung menggunakan Algoritma decision tree," 2025.
- [11] Anisa Rahmawati *et al.*, "Prediksi Penyakit Gagal Jantung Tahun 2025-2028 Rawat Inap di RSUD Gambiran Kota Kediri," *Antigen : Jurnal Kesehatan Masyarakat dan Ilmu Gizi*, vol. 3, no. 2, pp. 227–237, May 2025, doi: 10.57213/antigen.v3i2.667.
- [12] A. Purno, W. Wibowo, D. Hamdani, H. Heryono, and R. F. Pratama, "Sistem Artificial Intelligence Deteksi Penyakit THT dan Jantung Menggunakan Forward Chaining dan Image Processing," 2024.
- [13] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, H. T. Saputra, and F. Okmayura, "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 2, pp. 161–168, Sep. 2024, doi: 10.37148/bios.v5i2.152.
- [14] N. H. Alfajr and S. Defiyanti, "Prediksi penyakit jantung menggunakan metode random forest dan penerapan principal component analysis (PCA)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3s1.5055.
- [15] G. Ferry Mandias and I. J. Manoppo, "Penerapan Model Machine Learning untuk Memprediksi Serangan Jantung Dini," *Jurnal Ilmiah MATRIK*, vol. 27, no. 2, 2025.
- [16] J. Homepage *et al.*, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Hyperparameter Tuning for Classification Models in Heart Disease Risk Prediction Penerapan Hyperparameter Tuning pada Model Klasifikasi untuk Prediksi Risiko Penyakit Jantung," vol. 5, pp. 1181–1189, 2025, doi: 10.57152/malcom.v5i4.2138.
- [17] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, H. T. Saputra, and F. Okmayura, "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine," *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 2, pp. 161–168, Sep. 2024, doi: 10.37148/bios.v5i2.152.
- [18] J. Sulaksono, D. Putra Pamungkas, and D. W. Widodo, "Analisis Perbandingan Model Machine Learning untuk Prediksi Penyakit Jantung," 2026.
- [19] A. Wijayadhi, M. Makmun, and S. B. Rahardjo, "Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 15–28, 2023, doi: 10.47065/bit.v3i1.
- [20] A. Hidayat1, "464-Article Text-2073-1-10-20231007," 2023.
- [21] D. Haganta Depari *et al.*, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung,"

- JURNAL INFORMATIK Edisi ke*, vol. 18, p. 2022, 2023.
- [22] I. Hidayatul Imania, N. Dhea, and A. Tamalo, "Optimalisasi Klasifikasi Penderita Penyakit Jantung Koroner dalam Sepuluh Tahun Ke Depan Menggunakan Metode Regresi Logistik, Decision Tree, Dan Random Forest," 2024.
- [23] A. Damayanti, D. Maulana, and M. Z. Abdurrohman, "Analisis Klasifikasi Risiko Penyakit Jantung Menggunakan Metode Random Forest Creative Commons License 4.0 (CC BY-NC)," 2026, doi: 10.35889/progresif.v22i2.3609.
- [24] T. Gori and A. Hestiningtyas, "Optimasi Pemilihan Fitur untuk Prediksi Penyakit Jantung Menggunakan Algoritma Genetika dan Random Forest," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4214.
- [25] Y. Dewi Amritha, N. Luh, P. I. Candrawengi, W. Putra Dananjaya, and A. R. Dayanti, "Model Machine Learning yang Dioptimalkan untuk Prediksi Penyakit Jantung Menggunakan R Shiny," 2026.
- [26] A. Carolina Wibowo, S. Ardi Lestari, S. Informasi, F. Ilmu Komputer, and U. Duta Bangsa Surakarta, "Analisis penggunaan machine learning dalam klasifikasi penentuan penyakit jantung," vol. 9, no. 2, 2024.
- [27] J. K. Nainggolan, F. Sinaga, A. M. Sitorus, A. Khairia, and B. A. Wijaya, "DINAMIK-Volume," vol. 30, no. 2, p. 2025, 2025