Penerapan Explainable Artificial Intelligence(Xai) Dengan Metode Shap Untuk Analisis Faktor Risiko Penyakit Jantung

Tito Purwana Sasmita
Universitas Sebelas April

Received: 7 Januari 2026 Accepted: 21 May 2026 Published: 1 June 2026

*tito.purwana.sasmita@gmail.com

Kata Kunci: *Explainable AI*, Penyakit Jantung, *Random, Forest*, SHAP, *Machine Learning*, *Diagnosis Klinis*

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak: Penyakit kardiovaskular merupakan salah satu penyebab utama kematian di dunia sehingga diperlukan pendekatan deteksi dini yang akurat dan dapat dipahami oleh tenaga medis. Machine Learning dapat digunakan untuk membantu prediksi penyakit jantung, tetapi penerapannya dibidang medis masih menghadapi tantangan interpretabilitas karena sebagian model bekerja seperti kotak hitam (black box). Penelitian ini bertujuan membangun model prediksi penyakit jantung berbasis Random Forest serta meningkatkan transparansi model menggunakan metode SHAP (SHapley Additive exPlanations). Dataset yang digunakan adalah Heart Disease UCI yang telah melalui tahap preprocessing, meliputi penanganan missing value, pembersihan nilai tidak valid, penghapusan data duplikat, analisis outlier, dan transformasi variabel kategorikal. Data dibagi menjadi data latih dan data uji dengan rasio 80:20 serta divalidasi menggunakan 5-fold cross-validation. Hasil pengujian menunjukkan bahwa model Random Forest memperoleh akurasi sebesar 82,61%, presisi 81,82%, recall 88,24%, F1-score 84,91%, dan ROC-AUC sebesar 0,90. Analisis SHAP menunjukkan bahwa fitur exang, cp_atypical angina, thalch, oldpeak, age, dan sex_Male memiliki kontribusi penting terhadap prediksi penyakit jantung. Selain interpretasi global, SHAP juga digunakan untuk memberikan interpretasi lokal pada individu pasien tertentu. Hasil penelitian menunjukkan bahwa integrasi Random Forest dan SHAP tidak hanya menghasilkan performa klasifikasi yang baik, tetapi juga meningkatkan transparansi model sehingga berpotensi mendukung pengembangan sistem pendukung keputusan klinis yang lebih dapat dipahami.

PENDAHULUAN

Penyakit kardiovaskular (Cardiovascular Diseases/CVDs) masih menjadi salah satu tantangan utama dalam bidang kesehatan karena berkontribusi besar terhadap angka kematian global dan memberikan beban signifikan terhadap sistem pelayanan kesehatan [1]. Kondisi ini menunjukkan pentingnya pengembangan metode deteksi dini yang mampu membantu proses identifikasi risiko penyakit jantung secara lebih cepat dan akurat [2]. Dalam beberapa tahun terakhir, pemanfaatan Machine Learning (ML) telah berkembang sebagai pendekatan komputasi yang potensial dalam mendukung prediksi penyakit berbasis data klinis. Berbagai penelitian menunjukkan bahwa algoritma ML mampu memproses data medis yang kompleks dan mengidentifikasi pola yang relevan untuk mendukung proses diagnosis [3], [4]. Namun, salah satu tantangan utama dalam penerapan ML pada domain medis adalah keterbatasan interpretabilitas model. Model yang bersifat black-box sering kali sulit menjelaskan alasan di balik hasil prediksi yang diberikan, sehingga dapat mengurangi tingkat kepercayaan tenaga medis terhadap sistem tersebut [5], [6]. Studi systematic review juga menunjukkan bahwa model Machine Learning banyak digunakan dalam prediksi risiko penyakit kardiovaskular berbasis data klinis dan rekam medis elektronik [7].

Beberapa penelitian terdahulu telah berupaya meningkatkan performa prediksi penyakit jantung menggunakan berbagai algoritma Machine Learning. Sebagian penelitian berfokus pada peningkatan akurasi melalui optimasi algoritma, sedangkan sebagian lainnya mulai mengintegrasikan pendekatan Explainable Artificial Intelligence (XAI). Namun, masih terdapat kebutuhan untuk membangun model prediksi yang tidak hanya memiliki performa klasifikasi yang baik, tetapi juga mampu memberikan penjelasan terhadap keputusan model baik secara global maupun lokal.

Penelitian tidak hanya berfokus pada performa prediksi, tetapi juga pada transparansi model melalui pendekatan Explainable AI menggunakan SHAP. Metode SHAP digunakan untuk memberikan interpretasi global terhadap fitur-fitur yang paling berpengaruh serta interpretasi lokal terhadap prediksi pada individu pasien tertentu. Oleh karena itu, penelitian ini bertujuan membangun model prediksi penyakit jantung menggunakan algoritma Random Forest dan menganalisis interpretabilitas model menggunakan metode SHAP untuk meningkatkan transparansi hasil prediksi.

TINJAUAN LITERATUR

Penerapan teknologi komputasi dalam bidang kardiologi telah berkembang pesat dalam beberapa tahun terakhir. Machine Learning banyak digunakan untuk membantu proses prediksi penyakit jantung karena mampu mengolah data klinis dengan banyak atribut serta menemukan pola yang sulit diidentifikasi secara manual. Studi komparatif yang dilakukan oleh Taher [8] menunjukkan bahwa algoritma supervised learning memiliki potensi dalam meningkatkan performa prediksi penyakit jantung. Berbagai algoritma seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Neural Network, dan Random Forest telah banyak diterapkan dalam penelitian terkait prediksi penyakit jantung. Selain algoritma Machine Learning tradisional, pendekatan Deep Learning juga mulai digunakan dalam prediksi penyakit jantung untuk menangkap pola data yang lebih kompleks [9].

Asih et al. [10] menunjukkan bahwa performa model klasifikasi penyakit jantung sangat dipengaruhi oleh kualitas tahap preprocessing data. Hal ini menunjukkan bahwa penanganan missing value, encoding variabel kategorikal, serta pengolahan data yang tidak valid memiliki peran penting dalam menghasilkan model yang andal. Sementara itu, Almutairi dan Dardouri [11] mengusulkan pendekatan sistem cerdas hibrida untuk meningkatkan performa prediksi penyakit jantung. Temuan-temuan tersebut menunjukkan bahwa Machine Learning memiliki potensi sebagai alat bantu dalam sistem pendukung keputusan medis.

Diantara berbagai algoritma Machine Learning, Random Forest sering digunakan pada data medis berbentuk tabular karena memiliki kemampuan menangani hubungan non-linear antar fitur dan relatif tahan terhadap overfitting. Random Forest bekerja dengan membangun beberapa pohon keputusan dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir. Chicco dan Jurman [12] menunjukkan bahwa algoritma berbasis pohon dapat digunakan untuk menangkap pola klinis yang relevan pada data kesehatan. Meskipun Random Forest mampu memberikan performa klasifikasi yang baik, model ini tetap memiliki keterbatasan dalam hal interpretabilitas apabila hanya dilihat dari hasil prediksi akhir. Dalam konteks medis, interpretabilitas menjadi aspek penting karena tenaga kesehatan membutuhkan penjelasan mengenai faktor-faktor yang memengaruhi hasil prediksi model [13]. Explainable Artificial Intelligence (XAI) hadir sebagai pendekatan untuk meningkatkan transparansi model Machine Learning. Salah satu metode XAI yang banyak digunakan adalah SHAP (SHapley Additive exPlanations), yang menghitung kontribusi masing-masing fitur terhadap hasil prediksi berdasarkan konsep Shapley values [14]. Dalam konteks kardiologi, data elektrokardiogram juga memiliki peran penting dalam proses diagnosis dan pencegahan penyakit jantung, sehingga pendekatan berbasis AI perlu disertai dengan interpretasi yang dapat dipahami oleh tenaga medis [15].

SHAP memiliki keunggulan karena dapat digunakan untuk menjelaskan model pada dua tingkat, yaitu global dan lokal. Interpretasi global digunakan untuk mengetahui fitur-fitur yang paling berpengaruh terhadap prediksi secara keseluruhan, sedangkan interpretasi lokal digunakan untuk menjelaskan keputusan model pada satu individu pasien tertentu. Dalam konteks prediksi penyakit jantung, kemampuan interpretasi ini penting karena dapat membantu memahami hubungan antara fitur klinis dan hasil prediksi model. Untuk memosisikan kontribusi penelitian ini, dilakukan perbandingan terhadap beberapa penelitian terdahulu yang menggunakan pendekatan Machine Learning dan Explainable AI pada kasus penyakit jantung. Ringkasan penelitian terdahulu disajikan pada Tabel 1.

Tabel 1. Perbandingan Penelitian Terdahulu dengan Penelitian Saat Ini

Peneliti (Tahun)	Metode/Algoritma	Fokus Utama Penelitian	Kelemahan/Celah (Research Gap) / Kontribusi Penelitian
Teja & Rayalu (2025) [3]	Optimized Machine Learning	Meningkatkan akurasi prediksi melalui optimasi hyperparameter pada berbagai model.	Hanya berfokus pada performa angka (akurasi), tidak memberikan penjelasan bagaimana model bekerja.

Chicco & Jurman (2020) [12]	Random Forest & Regresi Logistik	Memprediksi tingkat kesintasan pasien gagal jantung menggunakan fitur serum.	Analisis terbatas pada dua fitur utama saja, kurang mencakup variabilitas klinis yang luas.
Guleria et al. (2022) [5]	XAI Framework	Menerapkan kerangka kerja klasifikasi untuk prediksi penyakit kardiovaskular.	Pendekatan interpretasi terkadang kurang stabil secara lokal dibandingkan SHAP yang memiliki landasan Game Theory.
Waqar et al. (2025) [16]	Explainable AI	Melakukan identifikasi fitur dari data kardiak multiparametrik.	Visualisasi yang dihasilkan masih bersifat umum dan belum mendetailkan kontribusi fitur individu secara mendalam.
Penelitian Ini (2026)	Random Forest + SHAP	Membangun model prediksi penyakit jantung berbasis Random Forest serta menerapkan SHAP untuk interpretasi global dan lokal.	Memberikan kontribusi berupa evaluasi model yang lebih komprehensif melalui accuracy, precision, recall, F1-score, ROC-AUC, cross-validation, serta interpretasi fitur menggunakan SHAP.

Metode Penelitian



Gambar 1. Alur metode penelitian yang digunakan

Tahapan penelitian yang dilakukan pada studi ini ditunjukkan pada Gambar 1. Proses penelitian dimulai dari pengumpulan data, preprocessing, pembagian data, pelatihan model Random Forest, evaluasi performa model, hingga analisis interpretabilitas menggunakan SHAP.

Pengumpulan Data

Penelitian menggunakan dataset Heart Disease yang diperoleh dari UCI Machine Learning Repository. Dataset awal terdiri dari 920 data rekam medis pasien. Setelah proses preprocessing, yang meliputi penanganan missing value, penghapusan data duplikat, dan pembersihan nilai tidak valid, jumlah data akhir yang digunakan dalam penelitian ini menjadi 918 data rekam medis pasien dengan 14 atribut klinis yang berkaitan dengan penyakit jantung. Variabel yang digunakan meliputi age, sex, chest pain type (cp), resting blood pressure (trestbps), cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalch), exercise-induced angina (exang), oldpeak, slope, number of major vessels (ca), dan thalassemia (thal).

Variabel target pada penelitian ini berupa klasifikasi biner yang menunjukkan keberadaan penyakit jantung. Nilai target 0 merepresentasikan pasien tanpa penyakit jantung, sedangkan nilai target 1 merepresentasikan pasien dengan penyakit jantung.

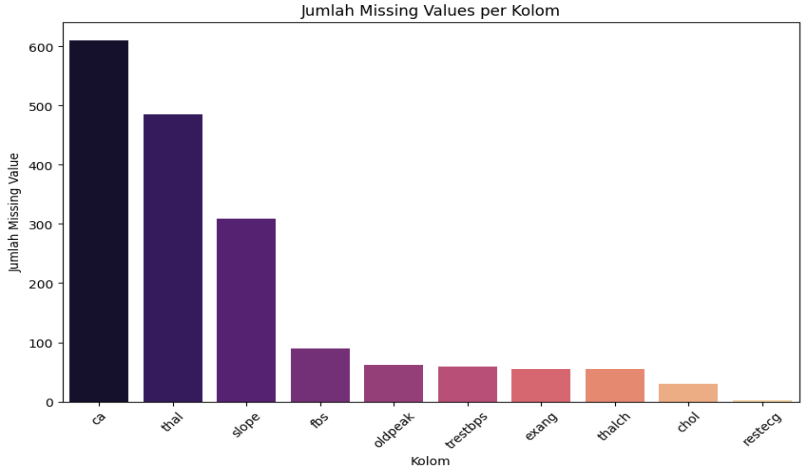
Tabel 2. Spesifikasi Atribut pada Dataset Heart Disease UCI

No	Nama Atribut	Deskripsi Klinis	Tipe Data
1	*age*	Usia pasien dalam satuan tahun	Numerik
2	*sex*	Jenis kelamin (1 = Laki - laki; 0 = Perempuan)	Kategorikal
3	*cp*	*Chest Pain Type* (Tipe nyeri dada): 0: *Typical angina* 1: *Atypical angina* 2: *Non-anginal pain* 3: *Asymptomatic*	Kategorikal
4	*trestbps*	Tekanan darah istirahat (mm Hg saat masuk RS)	Numerik
5	*chol*	Kolesterol serum dalam mg/dl	Numerik
6	*fbs*	Gula darah puasa > 120 mg/dl (1 = Benar; 0 = Salah)	Kategorikal

7	*restecg*	Hasil elektrokardiografi istirahat (skala 0, 1, 2)	Kategorikal
8	*thalch*	Detak jantung maksimal yang dicapai	Numerik
9	*exang*	Angina akibat olahraga (1= Ya; 0 = Tidak)	Kategorikal
10	*oldpeak*	Depresi ST yang diinduksi oleh latihan	Numerik
11	*slope*	Kemiringan puncak latihan segmen ST	Kategorikal
12	*ca*	Jumlah pembuluh darah utama (0-3)	Numerik
13	*thal*	Hasil tes talasemia (3 = Normal; 6 = Cacat tetap; 7 = Cacat reversibel)	Kategorikal
14	*target*	Status Penyakit Jantung (1 = Sakit; 0 = Sehat)	Label

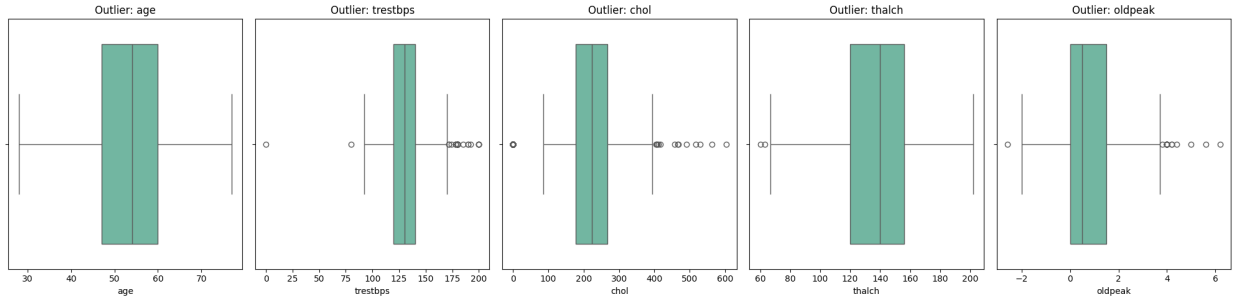
Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum digunakan pada proses pelatihan model Machine Learning. Tahapan preprocessing meliputi penanganan missing value, penghapusan data duplikat, analisis outlier, transformasi variabel kategorikal, dan pembersihan nilai yang tidak valid.



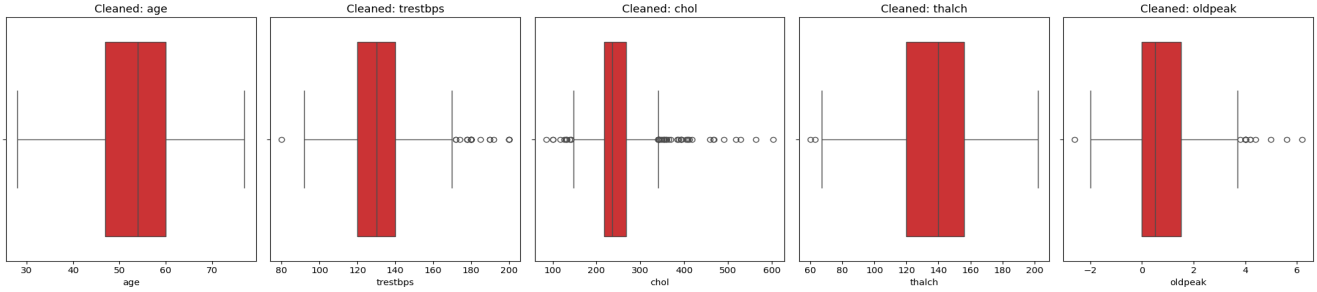
Gambar 2. Distribusi Missing Value Dataset Penyakit Jantung

Gambar 2 menunjukkan distribusi missing value pada dataset penyakit jantung. Beberapa atribut seperti ca, thal, dan slope memiliki jumlah missing value yang lebih tinggi dibandingkan atribut lainnya. Penanganan missing value dilakukan menggunakan metode imputasi berdasarkan tipe data pada masing-masing atribut. Atribut numerik ditangani menggunakan imputasi median karena lebih robust terhadap distribusi data yang tidak normal serta keberadaan outlier. Sementara itu, atribut kategorikal ditangani menggunakan imputasi modus untuk mempertahankan kategori yang paling sering muncul pada dataset. Setelah proses imputasi dilakukan, seluruh missing value pada dataset berhasil ditangani sehingga data siap digunakan pada tahap pelatihan model. Untuk mengidentifikasi keberadaan nilai ekstrem pada atribut numerik, penelitian ini melakukan analisis outlier menggunakan metode boxplot berbasis Interquartile Range (IQR).



Gambar 3. Visualisasi Outlier pada Atribut Numerik Sebelum Preprocessing

Gambar 3 menunjukkan visualisasi boxplot pada atribut numerik sebelum preprocessing dilakukan. Berdasarkan visualisasi tersebut, terlihat adanya beberapa nilai ekstrem (outlier) pada atribut trestbps, chol, thalch, dan oldpeak. Atribut chol menunjukkan jumlah outlier paling dominan dibandingkan atribut lainnya. Dalam konteks medis, nilai ekstrem tidak selalu merepresentasikan noise karena dapat menggambarkan kondisi klinis pasien tertentu. Oleh karena itu, penelitian ini tidak menghapus seluruh outlier secara langsung, melainkan hanya melakukan pembersihan pada nilai yang tidak valid. Setelah proses pembersihan data dan perbaikan nilai tidak valid dilakukan, distribusi data kembali dianalisis untuk melihat perubahan pola distribusi atribut numerik.



Gambar 4. Visualisasi Outlier Setelah Preprocessing Data

Gambar 4 menunjukkan distribusi atribut numerik setelah proses preprocessing dilakukan. Terlihat bahwa nilai yang tidak valid, seperti nilai 0 pada atribut trestbps dan chol, telah berhasil ditangani sehingga distribusi data menjadi lebih stabil dan realistis secara klinis. Meskipun demikian, beberapa nilai ekstrem tetap dipertahankan karena masih dapat merepresentasikan kondisi medis pasien yang valid. Selanjutnya, variabel kategorikal seperti sex, cp, restecg, slope, dan thal ditransformasikan menggunakan metode one-hot encoding untuk mengubah data kategorikal menjadi representasi numerik tanpa memberikan hubungan ordinal antar kategori. Setelah seluruh proses preprocessing selesai dilakukan, dataset akhir yang digunakan dalam penelitian ini terdiri dari 918 data.

Pembagian Data

Proses pembagian data dilakukan sebelum pelatihan model untuk menghindari terjadinya data leakage antara data latih dan data uji. Dataset dibagi menjadi data latih dan data uji menggunakan metode train-test split dengan rasio 80:20. Parameter stratify digunakan untuk menjaga proporsi distribusi kelas pada data latih dan data uji agar tetap seimbang. Selain itu, penelitian ini juga menerapkan 5-fold cross-validation untuk meningkatkan validitas dan generalisasi model. Pendekatan ini memungkinkan model diuji pada beberapa subset data yang berbeda sehingga performa model dapat dievaluasi secara lebih robust. Distribusi kelas pada dataset menunjukkan proporsi yang relatif seimbang sehingga penelitian ini tidak menerapkan teknik penanganan data imbalance tambahan seperti SMOTE atau undersampling.

Pelatihan Model Random Forest

Model klasifikasi pada penelitian ini dibangun menggunakan algoritma Random Forest. Algoritma ini dipilih karena memiliki kemampuan klasifikasi yang baik pada data tabular, mampu menangani hubungan non-linear antar fitur, serta memiliki ketahanan terhadap overfitting. Selain itu, Random Forest juga mampu memberikan performa yang stabil pada data medis dengan jumlah atribut yang beragam. Random Forest dikonfigurasi menggunakan beberapa parameter utama, yaitu n_estimators = 200, max_depth = 10, min_samples_split = 5, min_samples_leaf = 2, serta random_state = 42. Penggunaan parameter tersebut bertujuan untuk meningkatkan stabilitas model sekaligus mengurangi risiko overfitting selama proses pelatihan.

Evaluasi Model

Performa model dievaluasi menggunakan beberapa metrik klasifikasi, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC. Penggunaan beberapa metrik evaluasi bertujuan untuk memberikan gambaran performa model secara lebih komprehensif dalam mendeteksi penyakit jantung. Accuracy digunakan untuk mengukur tingkat ketepatan model dalam melakukan klasifikasi data secara keseluruhan. Nilai accuracy diperoleh dari perbandingan jumlah prediksi benar terhadap seluruh jumlah data yang diuji. Semakin tinggi nilai accuracy, maka semakin baik kemampuan model dalam melakukan klasifikasi data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Keterangan:
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

Precision digunakan untuk mengukur tingkat ketepatan model dalam memprediksi kelas positif. Metrik ini menunjukkan seberapa banyak prediksi positif yang benar-benar termasuk ke dalam kelas positif.

$$Precision = \frac{TP}{TP + FP}$$

Keterangan:
TP = True Positive
FP = False Positive

Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya. Nilai recall yang tinggi menunjukkan bahwa model mampu mengurangi kesalahan prediksi negatif palsu (false negative).

$$Recall = \frac{TP}{TP + FN}$$

Keterangan:
TP = True Positive
FN = False Negative

F1-Score merupakan rata-rata harmonik antara precision dan recall yang digunakan untuk mengevaluasi keseimbangan performa model dalam klasifikasi. Metrik ini sangat berguna ketika distribusi data tidak sepenuhnya seimbang.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:
Precision = tingkat ketepatan prediksi positif
Recall = kemampuan model mendeteksi kelas positif

Selain menggunakan accuracy, precision, recall, dan F1-score, penelitian ini juga menerapkan Receiver Operating Characteristic - Area Under Curve (ROC-AUC) untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif pada berbagai threshold klasifikasi. Nilai ROC-AUC yang mendekati 1 menunjukkan bahwa model memiliki kemampuan diskriminasi yang semakin baik dalam melakukan klasifikasi. Confusion matrix digunakan untuk menganalisis jumlah prediksi benar dan salah yang dihasilkan model. Evaluasi ini terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) sehingga performa model dapat dianalisis secara lebih rinci.

Analisis SHAP

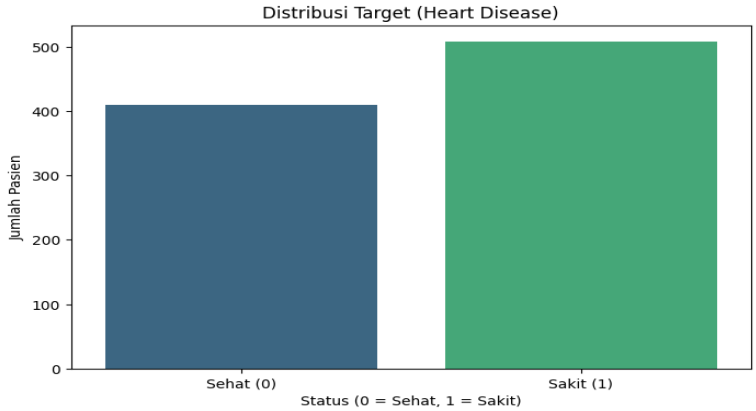
Implementasi menggunakan bahasa pemrograman Python dengan bantuan library Scikit-learn untuk pembangunan model Machine Learning serta library SHAP untuk analisis interpretabilitas model. Untuk meningkatkan transparansi model, penelitian ini menerapkan pendekatan Explainable Artificial Intelligence (XAI) menggunakan SHapley Additive exPlanations (SHAP). Metode SHAP digunakan untuk menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi model Random Forest. Metode ini dipilih karena mampu memberikan interpretasi model yang konsisten berdasarkan kontribusi setiap fitur terhadap hasil prediksi. Analisis interpretabilitas dilakukan pada dua tingkat, yaitu global interpretability dan local interpretability.

Global interpretability digunakan untuk mengidentifikasi fitur-fitur yang paling berpengaruh terhadap prediksi penyakit jantung secara keseluruhan. Sementara itu, local interpretability digunakan untuk menjelaskan alasan model dalam memberikan prediksi pada individu pasien tertentu sehingga proses pengambilan keputusan model dapat dipahami secara lebih transparan.

HASIL PENELITIAN

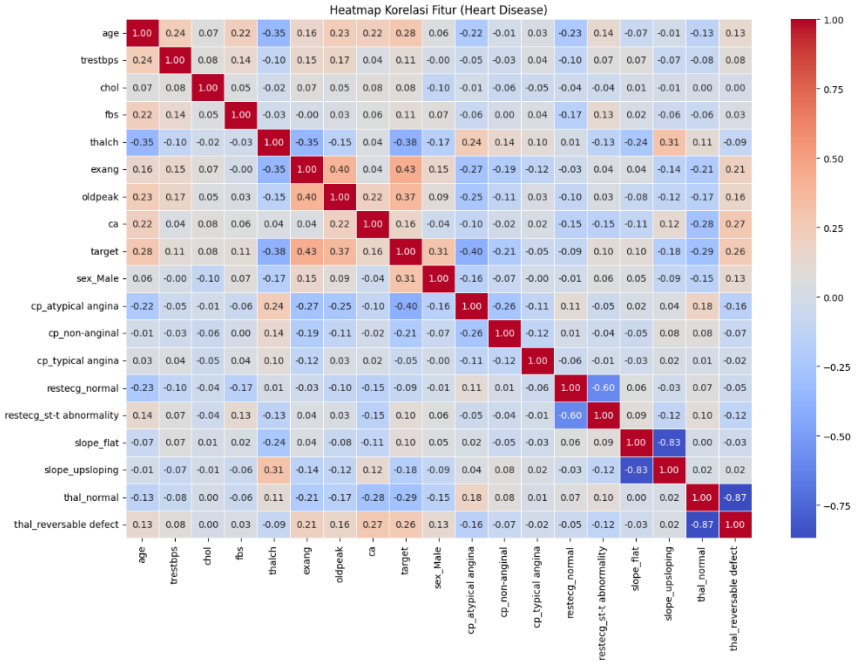
Analisis Distribusi dan Korelasi Data

Sebelum melakukan evaluasi model, penelitian ini terlebih dahulu menganalisis distribusi kelas target untuk mengetahui proporsi data pasien tanpa penyakit jantung dan pasien dengan penyakit jantung.



Gambar 5. Distribusi kelas target pada dataset penyakit jantung.

Gambar 5 menunjukkan distribusi kelas target pada dataset penyakit jantung. Jumlah pasien dengan penyakit jantung lebih banyak dibandingkan pasien tanpa penyakit jantung. Namun, perbedaan jumlah kedua kelas tidak terlalu signifikan sehingga dataset masih dapat dikategorikan relatif seimbang.



Gambar 6. Heatmap korelasi antar fitur pada dataset penyakit jantung.

Gambar 6 menunjukkan heatmap korelasi antar fitur pada dataset penyakit jantung. Berdasarkan visualisasi tersebut, fitur exang memiliki korelasi positif terhadap target, sedangkan fitur thalch memiliki korelasi negatif terhadap target. Hal ini menunjukkan bahwa pasien dengan exercise-induced angina cenderung memiliki risiko penyakit jantung yang lebih tinggi, sementara nilai detak jantung maksimum yang lebih tinggi cenderung berkaitan dengan risiko penyakit jantung yang lebih rendah. Selain itu, fitur oldpeak juga menunjukkan korelasi positif terhadap target, yang mengindikasikan bahwa peningkatan depresi segmen ST memiliki hubungan dengan kemungkinan penyakit jantung. Hasil korelasi ini memberikan gambaran awal mengenai fitur-fitur yang berpotensi berpengaruh terhadap prediksi model.

Evaluasi Model Random Forest

Evaluasi performa model dilakukan pada data uji yang belum pernah dilihat oleh model selama proses pelatihan. Model Random Forest dievaluasi menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC.

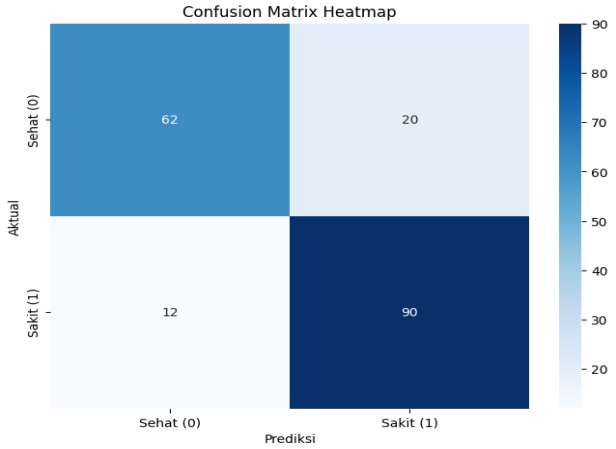
Tabel 3. Hasil Pengujian Model Klasifikasi Random Forest

Metrik Evaluasi	Skor Hasil
Akurasi	82,61%
Presisi	81,82%
Recall	88,24%

F1-Score	84,91%
ROC-AUC	0,90
Rata-rata Akurasi 5-Fold Cross-Validation	81,27%

Keterangan: Akurasi, presisi, recall, F1-score, dan ROC-AUC diperoleh dari pengujian pada data uji, sedangkan rata-rata akurasi 5-fold cross-validation diperoleh dari proses validasi silang.

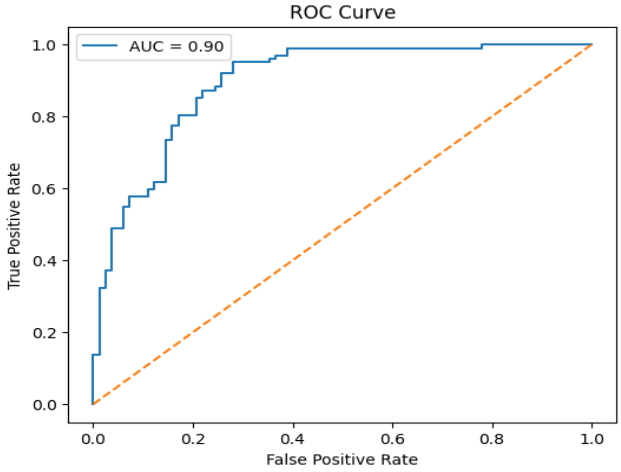
Berdasarkan Tabel 3, model Random Forest memperoleh akurasi sebesar 82,61% pada data uji. Selain itu, hasil 5-fold cross-validation menunjukkan rata-rata akurasi sebesar 81,27%. Hasil ini menunjukkan bahwa performa model relatif stabil pada beberapa subset data yang berbeda.



Gambar 7. Confusion matrix model Random Forest.

Gambar 7 menunjukkan confusion matrix dari model Random Forest pada data uji. Model berhasil mengklasifikasikan 90 pasien dengan penyakit jantung secara benar sebagai kelas sakit (True Positive) dan 62 pasien tanpa penyakit jantung secara benar sebagai kelas sehat (True Negative). Namun, masih terdapat 20 kasus False Positive dan 12 kasus False Negative.

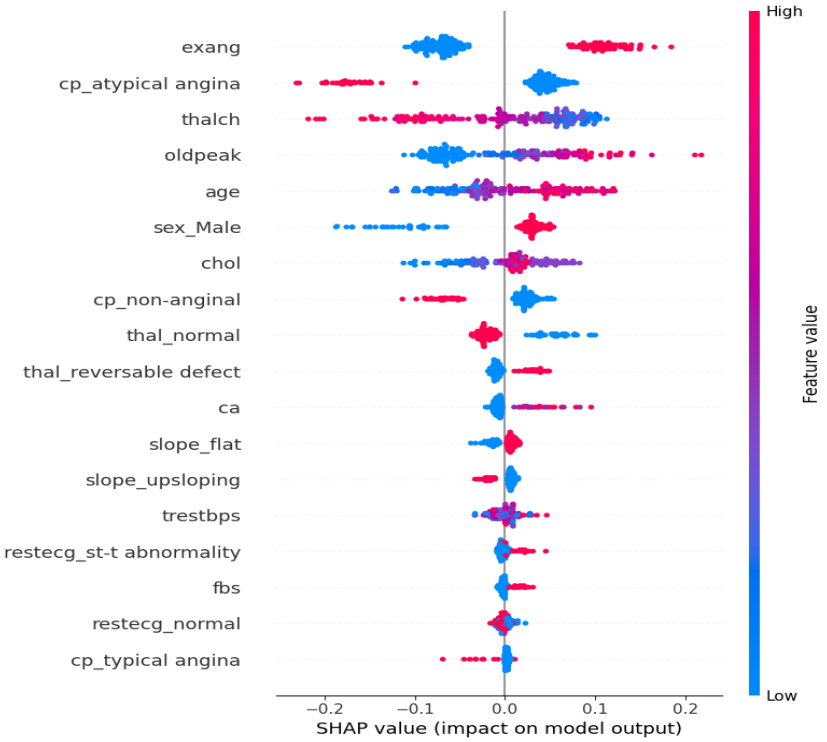
Dalam konteks diagnosis medis, False Negative perlu diperhatikan karena menunjukkan pasien yang sebenarnya memiliki penyakit jantung tetapi diprediksi sebagai sehat oleh model. Meskipun demikian, jumlah False Negative yang relatif lebih rendah dibandingkan True Positive menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mendeteksi pasien dengan penyakit jantung.



Gambar 8. Kurva ROC model Random Forest.

Gambar 8 menunjukkan kurva Receiver Operating Characteristic (ROC) dari model Random Forest. Model memperoleh nilai Area Under Curve (AUC) sebesar 0,90. Nilai tersebut menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat baik dalam membedakan pasien sehat dan pasien dengan penyakit jantung. Kurva ROC yang berada jauh di atas garis diagonal menunjukkan bahwa model memiliki performa klasifikasi yang baik pada berbagai threshold. Dengan demikian, model tidak hanya menunjukkan performa yang baik berdasarkan accuracy, precision, recall, dan F1-score, tetapi juga memiliki kemampuan diskriminatif yang kuat berdasarkan nilai ROC-AUC.

Analisis Interpretabilitas Global SHAP

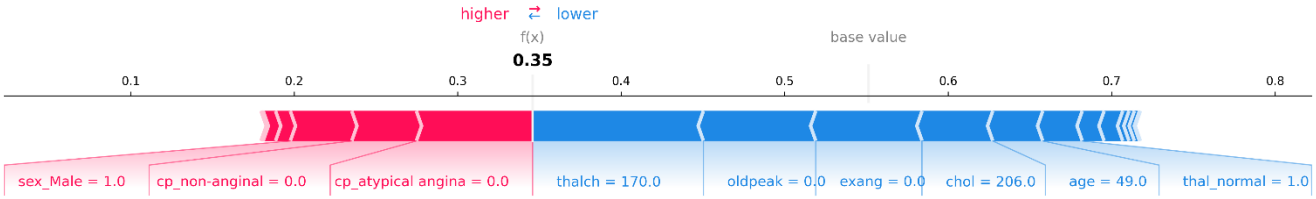


Gambar 9. SHAP summary plot untuk interpretasi global model Random Forest.

Gambar 9 menunjukkan hasil interpretasi global menggunakan SHAP summary plot. Berdasarkan visualisasi tersebut, fitur yang memiliki pengaruh besar terhadap prediksi model antara lain exang, cp_atypical angina, thalch, oldpeak, age, dan sex_Male. Fitur cp_atypical angina dan sex_Male merupakan hasil transformasi one-hot encoding dari atribut kategorikal pada tahap preprocessing. Fitur exang menjadi salah satu fitur paling berpengaruh dalam prediksi penyakit jantung. Nilai exang yang tinggi cenderung memberikan kontribusi positif terhadap prediksi kelas sakit, yang menunjukkan bahwa pasien dengan exercise-induced angina memiliki risiko lebih tinggi. Selain itu, nilai thalch yang rendah cenderung meningkatkan prediksi penyakit jantung, sedangkan nilai thalch yang tinggi cenderung menurunkan probabilitas prediksi sakit.

Fitur oldpeak juga menunjukkan pengaruh penting terhadap prediksi model. Nilai oldpeak yang tinggi cenderung meningkatkan kontribusi terhadap prediksi penyakit jantung karena berkaitan dengan perubahan segmen ST pada pemeriksaan jantung. Temuan ini menunjukkan bahwa model mampu menangkap pola fitur klinis yang relevan dalam prediksi penyakit jantung.

Analisis Interpretabilitas Lokal SHAP



Gambar 10. SHAP force plot untuk interpretasi prediksi individu.

Selain interpretasi global, penelitian ini juga menerapkan interpretasi lokal menggunakan SHAP force plot. Gambar 10 menunjukkan kontribusi masing-masing fitur terhadap prediksi model pada salah satu sampel pasien. Pada visualisasi tersebut, fitur berwarna merah menunjukkan kontribusi yang meningkatkan probabilitas prediksi penyakit jantung, sedangkan fitur berwarna biru menunjukkan kontribusi yang menurunkan probabilitas prediksi penyakit jantung. Pada contoh sampel tersebut, fitur seperti thalch = 170, oldpeak = 0, exang = 0, chol = 206, age = 49, dan thal_normal = 1 memberikan kontribusi dalam

menurunkan probabilitas prediksi penyakit jantung. Hal ini menunjukkan bahwa pada sampel tersebut, faktor-faktor yang menurunkan risiko lebih dominan dibandingkan faktor yang meningkatkan risiko. Hasil menunjukkan bahwa SHAP tidak hanya mampu menjelaskan fitur penting secara global, tetapi juga dapat memberikan penjelasan pada level individu. Pendekatan ini penting dalam konteks medis karena dapat membantu meningkatkan transparansi model dan mendukung proses pengambilan keputusan klinis.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa model Random Forest mampu memberikan performa klasifikasi yang baik dalam prediksi penyakit jantung. Nilai akurasi sebesar 82,61% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan benar. Selain itu, nilai ROC-AUC sebesar 0,90 menunjukkan bahwa model memiliki kemampuan diskriminasi yang baik dalam membedakan pasien sehat dan pasien dengan penyakit jantung. Nilai recall sebesar 88,24% menjadi salah satu hasil penting dalam konteks medis karena menunjukkan kemampuan model dalam mendeteksi pasien yang benar-benar memiliki penyakit jantung. Recall yang tinggi penting karena kesalahan false negative dapat berdampak serius, yaitu pasien yang sebenarnya memiliki penyakit jantung tetapi diprediksi sebagai sehat. Pada penelitian ini, model masih menghasilkan 12 kasus false negative, sehingga aspek ini tetap perlu menjadi perhatian dalam pengembangan model prediksi medis.

Dari sisi interpretabilitas, hasil SHAP menunjukkan bahwa fitur exang, cp_atypical angina, thalch, oldpeak, age, dan sex_Male memiliki kontribusi penting terhadap prediksi model. Fitur exang menjadi salah satu fitur paling berpengaruh, yang menunjukkan bahwa pasien dengan exercise-induced angina cenderung memiliki risiko penyakit jantung yang lebih tinggi. Temuan ini relevan secara klinis karena nyeri dada yang muncul akibat aktivitas fisik dapat menjadi salah satu indikator gangguan kardiovaskular.

Fitur thalch juga menunjukkan kontribusi penting terhadap hasil prediksi. Nilai thalch yang rendah cenderung meningkatkan probabilitas prediksi penyakit jantung, sedangkan nilai thalch yang tinggi cenderung menurunkan probabilitas prediksi sakit. Hal ini menunjukkan bahwa kemampuan jantung dalam mencapai denyut maksimum memiliki hubungan dengan kondisi kardiovaskular pasien. Selain itu, fitur oldpeak yang tinggi juga berkontribusi terhadap peningkatan risiko prediksi penyakit jantung karena berkaitan dengan perubahan segmen ST pada pemeriksaan jantung.

Selain interpretasi global, penggunaan SHAP force plot memberikan penjelasan pada level individu. Interpretasi lokal ini menunjukkan bagaimana setiap fitur dapat meningkatkan atau menurunkan probabilitas prediksi pada satu pasien tertentu. Dengan demikian, model tidak hanya memberikan hasil klasifikasi, tetapi juga menyediakan penjelasan mengenai alasan di balik prediksi yang dihasilkan.

Dibandingkan dengan pendekatan black-box yang hanya menghasilkan prediksi akhir, integrasi Random Forest dan SHAP memberikan nilai tambah berupa transparansi model. Pendekatan ini penting dalam konteks medis karena dapat membantu tenaga kesehatan memahami kontribusi fitur klinis terhadap hasil prediksi. Dengan adanya interpretasi global dan lokal, model prediksi yang dibangun tidak hanya berfokus pada performa, tetapi juga pada aspek interpretabilitas yang diperlukan dalam sistem pendukung keputusan klinis. Meskipun demikian, hasil penelitian ini masih terbatas pada penggunaan dataset sekunder sehingga diperlukan validasi lebih lanjut menggunakan data klinis lokal untuk memastikan kemampuan generalisasi model pada populasi yang berbeda.

KESIMPULAN

Penelitian berhasil menerapkan pendekatan Explainable Artificial Intelligence (XAI) menggunakan metode SHAP untuk meningkatkan transparansi model prediksi penyakit jantung berbasis Random Forest. Berdasarkan hasil pengujian, model Random Forest memperoleh akurasi sebesar 82,61%, presisi sebesar 81,82%, recall sebesar 88,24%, F1-score sebesar 84,91%, serta ROC-AUC sebesar 0,90. Hasil tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam membedakan pasien dengan penyakit jantung dan pasien tanpa penyakit jantung.

Selain performa klasifikasi, penelitian ini juga menunjukkan bahwa penggunaan SHAP mampu memberikan interpretasi terhadap hasil prediksi model. Berdasarkan analisis interpretabilitas global, fitur seperti exang, cp_atypical angina, thalch, oldpeak, age, dan sex_Male memiliki kontribusi penting terhadap prediksi penyakit jantung. Temuan ini menunjukkan bahwa model tidak hanya menghasilkan prediksi, tetapi juga mampu memberikan penjelasan mengenai fitur-fitur yang memengaruhi keputusan model.

Penerapan SHAP pada tingkat lokal juga memberikan nilai tambah karena mampu menjelaskan alasan prediksi pada individu pasien tertentu. Hal ini penting dalam konteks medis karena transparansi model dapat membantu tenaga kesehatan memahami kontribusi setiap fitur klinis terhadap hasil prediksi. Dengan demikian, integrasi Random Forest dan SHAP berpotensi digunakan sebagai dasar pengembangan sistem pendukung keputusan klinis yang lebih transparan dan mudah dipahami.

Meskipun demikian, penelitian masih memiliki keterbatasan. Dataset yang digunakan merupakan dataset sekunder sehingga hasil penelitian masih perlu divalidasi lebih lanjut menggunakan data klinis lokal, khususnya data pasien di Indonesia. Penelitian selanjutnya dapat mengembangkan model dengan validasi eksternal, membandingkan beberapa algoritma Machine Learning lain, serta mengeksplorasi metode interpretabilitas tambahan seperti LIME untuk membandingkan stabilitas dan kualitas penjelasan model.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)." Accessed: May 20, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] R. Kumar *et al.*, "A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions," *Front. Artif. Intell.*, vol. 8, May 2025, doi: 10.3389/frai.2025.1583459.

[3] M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovasc. Disord.*, vol. 25, no. 1, p. 212, Mar. 2025, doi: 10.1186/s12872-025-04627-6.

[4] S. Patidar, D. Kumar, and D. Rukwal, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction," 2022. doi: 10.3233/ATDE220723.

[5] P. Guleria, P. Naga Srinivasu, S. Ahmed, N. Almusallam, and F. K. Alarfaj, "XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques," *Electronics*, vol. 11, no. 24, p. 4086, Dec. 2022, doi: 10.3390/electronics11244086.

[6] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 310, Dec. 2020, doi: 10.1186/s12911-020-01332-6.

[7] T. Liu, A. Krentz, L. Lu, and V. Curcin, "Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis," *Eur. Hear. J. - Digit. Heal.*, vol. 6, no. 1, pp. 7–22, Jan. 2025, doi: 10.1093/ehjdh/ztae080.

[8] H. A. Taher and A. M. Abdulazeez, "Machine Learning Approaches for Heart Disease Detection: A Comprehensive Review," *Int. J. Res. Appl. Technol.*, vol. 3, no. 2, pp. 267–282, 2023, doi: 10.34010/injuratech.v3i2.12052.

[9] T. GEETHA, "Coronary Heart Disease Prediction Using Deep Learning," *Int. J. Sci. Res. Eng. Manag.*, vol. 09, no. 04, pp. 1–9, Apr. 2025, doi: 10.55041/IJSREM46041.

[10] P. S. Asih, Y. Azhar, G. W. Wicaksono, and D. R. Akbi, "Interpretable Machine Learning Model For Heart Disease Prediction," *Procedia Comput. Sci.*, vol. 227, pp. 439–445, 2023, doi: 10.1016/j.procs.2023.10.544.

[11] M. Almutairi and S. Dardouri, "Intelligent Hybrid Modeling for Heart Disease Prediction," *Information*, vol. 16, no. 10, p. 869, Oct. 2025, doi: 10.3390/info16100869.

[12] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 16, Dec. 2020, doi: 10.1186/s12911-020-1023-5.

[13] H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Sci. Rep.*, vol. 14, no. 1, p. 23277, Oct. 2024, doi: 10.1038/s41598-024-74656-2.

[14] I. D. Mienye and N. Jere, "Optimized Ensemble Learning Approach with Explainable AI for Improved Heart Disease Prediction," *Information*, vol. 15, no. 7, p. 394, Jul. 2024, doi: 10.3390/info15070394.

[15] D. Bartusik-Aebisher, K. Rogóż, and D. Aebisher, "Artificial Intelligence and ECG: A New Frontier in Cardiac Diagnostics and Prevention," *Biomedicines*, vol. 13, no. 7, p. 1685, Jul. 2025, doi: 10.3390/biomedicines13071685.

[16] M. Waqar, M. B. Shahnawaz, S. Saleem, H. Dawood, U. Muhammad, and H. Dawood, "Enhancing Heart Attack Prediction: Feature Identification from Multiparametric Cardiac Data Using Explainable AI," *Algorithms*, vol. 18, no. 6, p. 333, Jun. 2025, doi: 10.3390/a18060333.