Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

Prediksi Penyakit Jantung Berbasis Random Forest dengan GridSearchCV dan SHAP Menggunakan Dataset Publik UCI Cleveland

Anggi Setiyawan*, Sulistiyasni, Muhammad Akbar Setiawan

Teknik Informatika, STMIK Widyta Utama, Purwokerto
Jl. Sunan Kalijaga, Dusun III, Berkoh, Kec. Purwokerto Selatan, Kabupaten Banyumas, Jawa Tengah, Indonesia
Email: [1,*]anggisetiyawan02733@gmail.com, [2]sulistyasnipwt@swu.ac.id, [3]akbar@swu.ac.id
Email Penulis Korespondensi: anggisetiyawan02733@gmail.com
Submitted: **10/07/2026;** Accepted: **25/07/2026;** Published: **27/07/2026**

Abstrak−Penyakit kardiovaskular merupakan penyebab kematian tertinggi secara global dengan 19,2 juta jiwa pada tahun 2023, sehingga pengembangan sistem prediksi otomatis berbasis data menjadi kebutuhan mendesak. Penelitian ini mengembangkan model prediksi penyakit jantung berbasis Random Forest yang dioptimasi menggunakan Grid Search Cross-Validation (GridSearchCV) dan dilengkapi analisis SHapley Additive exPlanations (SHAP) untuk interpretabilitas klinis pada dataset UCI Cleveland Heart Disease (297 sampel, 13 fitur klinis). Tiga kontribusi metodologis diterapkan: (1) pipeline preprocessing reproducible dengan StandardScaler post-split untuk mencegah data leakage, (2) pencarian hyperparameter deterministik dan exhaustive menggunakan GridSearchCV dengan 216 kombinasi dan Stratified 10-Fold Cross Validation, serta (3) analisis SHAP pada level global berbasis data latih dan lokal berbasis data uji untuk menghasilkan prediksi yang dapat diinterpretasi secara klinis. Konfigurasi hyperparameter optimal yang dihasilkan adalah n_estimators = 100, max_depth = None, min_samples_leaf = 4, min_samples_split = 2, dan max_features = 'sqrt'. Model RF Tuned mencapai AUC-ROC 0,9453 dan CV AUC 0,9010, mengungguli RF Baseline (AUC-ROC 0,9414; CV AUC 0,8804) pada metrik diskriminasi utama dengan stabilitas yang lebih baik. Analisis SHAP mengidentifikasi cp (mean |SHAP| = 0,1031), thal (0,0976), dan ca (0,0805) sebagai tiga fitur klinis paling berpengaruh, selaras dengan indikator diagnostik kardiologi. Integrasi GridSearchCV dan SHAP menghasilkan model yang tidak hanya akurat tetapi juga transparan untuk mendukung pengambilan keputusan medis.

Kata Kunci: Random Forest; GridSearchCV; Hyperparameter Tuning; SHAP; Prediksi Penyakit Jantung

Abstract−Cardiovascular disease remains the leading cause of death globally, with 19.2 million fatalities recorded in 2023, making the development of automated data-driven prediction systems an urgent necessity. This study develops a heart disease prediction model based on Random Forest optimized using Grid Search Cross-Validation (GridSearchCV) and supplemented with SHapley Additive exPlanations (SHAP) analysis for clinical interpretability on the UCI Cleveland Heart Disease dataset (297 samples, 13 clinical features). Three methodological contributions are implemented: (1) a reproducible preprocessing pipeline with post-split StandardScaler to prevent data leakage, (2) deterministic and exhaustive hyperparameter search using GridSearchCV with 216 combinations and Stratified 10-Fold Cross Validation, and (3) SHAP analysis at the global level based on training data and at the local level based on test data to produce clinically interpretable predictions. The optimal hyperparameter configuration obtained is n_estimators = 100, max_depth = None, min_samples_leaf = 4, min_samples_split = 2, and max_features = 'sqrt'. The Tuned RF model achieves an AUC-ROC of 0.9453 and CV AUC of 0.9010, outperforming the RF Baseline (AUC-ROC 0.9414; CV AUC 0.8804) on key discrimination metrics with greater stability. SHAP analysis identifies cp (mean |SHAP| = 0.1031), thal (0.0976), and ca (0.0805) as the three most influential clinical features, consistent with established cardiological diagnostic indicators. The integration of GridSearchCV and SHAP produces a model that is not only accurate but also transparent in supporting medical decision-making.

Keywords: Random Forest; GridSearchCV; Hyperparameter Tuning; SHAP; Heart Disease Prediction

1. PENDAHULUAN

Penyakit kardiovaskular (PKV) merupakan penyebab kematian tertinggi di dunia dan terus menjadi beban kesehatan global yang signifikan. Laporan World Heart Federation mencatat bahwa lebih dari setengah miliar orang hidup dengan penyakit kardiovaskular, dengan angka kematian mencapai 20,5 juta jiwa pada tahun 2021, yakni hampir sepertiga dari seluruh kematian global [1]. Tren peningkatan ini dikonfirmasi oleh data Global Burden of Disease 2023 yang menggunakan metodologi estimasi berbeda, mencatat kematian akibat PKV sebesar 19,2 juta jiwa pada tahun 2023 dengan lintasan peningkatan dari 13,1 juta pada tahun 1990 [2]. Meskipun terdapat perbedaan estimasi antar sumber akibat perbedaan metodologi, kedua laporan secara konsisten menunjukkan tren peningkatan beban PKV secara global. Kondisi ini tidak merata secara global, dengan sekitar 80% kematian akibat PKV terjadi di negara berpendapatan rendah dan menengah [1]. Di kawasan Asia Tenggara, situasinya tergolong kritis angka kematian PKV yang distandarisasi usia di Indonesia bahkan melebihi rata-rata negara dengan indeks pembangunan sosial menengah sebesar 1,5 kali lipat pada tahun 2021 [3]. Tingginya angka kematian ini sebagian besar disebabkan oleh keterlambatan deteksi dan keterbatasan akses terhadap diagnosis klinis yang akurat dan tepat waktu [1], sehingga pengembangan sistem prediksi otomatis berbasis data menjadi kebutuhan yang mendesak.

Perkembangan machine learning (ML) telah membuka peluang baru dalam prediksi dan klasifikasi penyakit jantung secara otomatis. Berbagai algoritma telah dieksplorasi untuk tugas ini, di antaranya Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), dan Extreme Gradient Boosting (XGBoost) [4], [5]. Studi komparatif menunjukkan bahwa performa algoritma

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

sangat bergantung pada ukuran dataset: pada dataset berukuran kecil (303 sampel UCI Cleveland), K-NN dengan hyperparameter tertuning mencapai akurasi testing tertinggi sebesar 87,91%, sementara pada dataset yang lebih besar (1025 sampel gabungan), XGBoost dan Random Forest masing-masing mencapai 99,03% dan 98,05% [4]. Pendekatan hybrid yang menggabungkan Random Forest dengan Logistic Regression juga terbukti meningkatkan akurasi prediksi dibandingkan model tunggal, dengan peningkatan akurasi sebesar 1,32% dari 83,16% menjadi 84,48% pada dataset UCI Cleveland [5]. Hasil-hasil ini menunjukkan potensi besar ML dalam mendukung diagnosis penyakit jantung, namun sekaligus mengungkap perlunya kajian lebih mendalam mengenai strategi optimasi dan interpretabilitas model.

Di antara berbagai algoritma ML yang telah dievaluasi, Random Forest secara konsisten menunjukkan performa kompetitif dan karakteristik yang cocok untuk aplikasi klinis. Sebagai metode ensemble berbasis pohon keputusan, Random Forest mampu menangani data berdimensi tinggi, tahan terhadap overfitting, dan secara inheren menghasilkan informasi feature importance yang dapat digunakan untuk memahami kontribusi setiap variabel klinis terhadap prediksi [6], [7]. Analisis feature importance pada model Random Forest menunjukkan bahwa fitur thal (kondisi thalassemia) secara konsisten teridentifikasi sebagai prediktor klinis paling dominan dalam prediksi penyakit jantung [6]. Karakteristik-karakteristik inilah yang menjadikan Random Forest sebagai pilihan yang relevan untuk sistem pendukung keputusan klinis [7].

Meskipun potensi Random Forest dalam prediksi penyakit jantung telah didemonstrasikan secara luas, tiga kesenjangan penelitian signifikan masih perlu diatasi. Pertama, sebagian besar studi yang menggunakan Random Forest tidak menerapkan hyperparameter tuning yang sistematis, sehingga model beroperasi dengan parameter default yang menghasilkan performa suboptimal. Sebagai contoh, Random Forest tanpa tuning hanya mencapai akurasi 83,16% [5] dan 82,42% [4], padahal pencarian hyperparameter yang terstruktur membuka peluang untuk mengidentifikasi kombinasi parameter yang lebih optimal. Meskipun peningkatan akurasi melalui tuning pada dataset kecil cenderung bersifat marginal [4], nilai utama dari pendekatan deterministik seperti GridSearchCV terletak pada kemampuannya menghasilkan konfigurasi hyperparameter yang dapat diverifikasi dan direplikasi, suatu syarat esensial dalam penelitian klinis yang menuntut konsistensi dan transparansi hasil.

Kedua, studi yang telah menerapkan hyperparameter tuning pada Random Forest umumnya mengandalkan metode metaheuristik seperti FOX, PSO, atau Bat Algorithm [7]. Meskipun metode-metode ini mampu mengeksplorasi ruang pencarian yang luas, sifatnya yang stokastik menyebabkan hasil yang dihasilkan sulit direplikasi secara eksak pada percobaan berbeda. GridSearchCV, sebagai metode pencarian exhaustive yang deterministik, menawarkan jaminan bahwa kombinasi hyperparameter terbaik dalam ruang pencarian yang didefinisikan akan selalu ditemukan dan menghasilkan output yang identik pada setiap eksekusi, karakteristik yang sangat penting dalam validasi model untuk keperluan klinis dan publikasi ilmiah. Ketiga, hampir seluruh model yang ada masih bersifat black-box dan hanya melaporkan metrik performa agregat tanpa menyertakan mekanisme interpretabilitas [4], [5], [6], [7]. Padahal, transparansi dalam proses pengambilan keputusan model merupakan prasyarat mendasar bagi adopsi AI dalam praktik klinis [8], [9], di mana tenaga medis membutuhkan pemahaman tentang faktor-faktor yang mendorong setiap prediksi untuk dapat mempercayai dan menggunakan sistem secara bertanggung jawab [10], [11].

Penelitian ini mengusulkan model klasifikasi penyakit jantung berbasis Random Forest yang dioptimasi menggunakan Grid Search Cross-Validation (GridSearchCV) untuk pencarian hyperparameter yang sistematis, deterministik, dan reproducible, serta dilengkapi dengan analisis SHapley Additive Explanations (SHAP) untuk menghasilkan prediksi yang tidak hanya akurat tetapi juga dapat diinterpretasi secara klinis. Performa model dievaluasi secara komprehensif menggunakan metrik accuracy, precision, recall, F1-score, dan AUC-ROC, dengan validasi K-Fold Cross Validation untuk memastikan generalitas model pada data klinis yang tersedia secara publik. Analisis SHAP diterapkan untuk mengidentifikasi dan mengkuantifikasi kontribusi setiap fitur klinis terhadap hasil prediksi individual maupun global, sehingga model yang dihasilkan dapat mendukung pengambilan keputusan medis secara transparan dan dapat dipertanggungjawabkan.

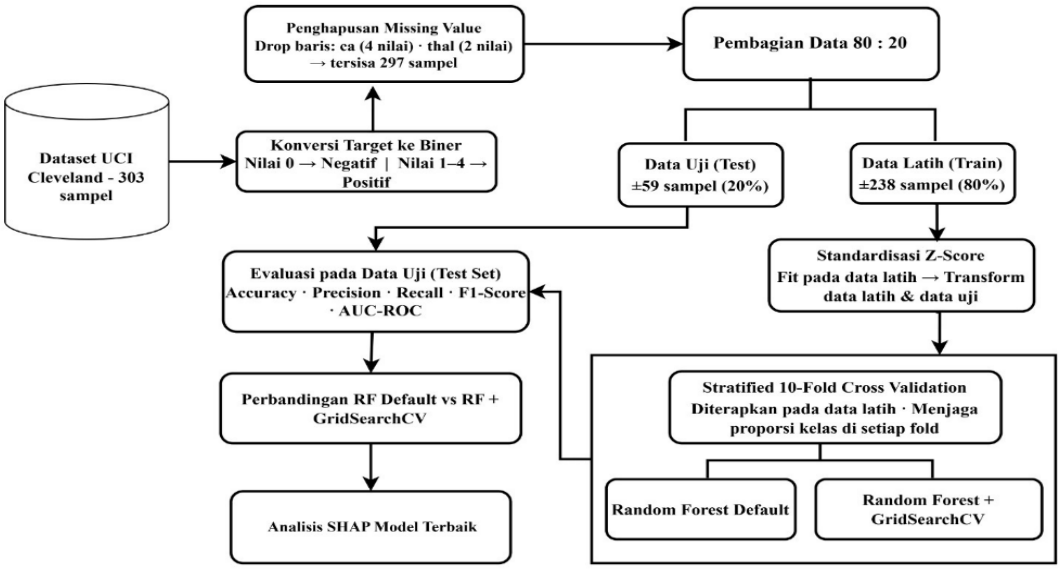
Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit jantung yang akurat dan interpretatif dengan mengintegrasikan Random Forest, GridSearchCV, dan SHAP. Kontribusi utama penelitian ini adalah sebagai berikut: (1) pengembangan model Random Forest yang dioptimasi menggunakan GridSearchCV dengan pencarian hyperparameter yang deterministik dan exhaustive, menghasilkan konfigurasi optimal yang terverifikasi dan reproducible pada dataset UCI Cleveland Heart Disease; (2) perbandingan sistematis antara model Random Forest dengan dan tanpa hyperparameter tuning untuk mengukur dampak optimasi terhadap performa klasifikasi; (3) penerapan SHAP untuk mengidentifikasi dan mengkuantifikasi kontribusi setiap fitur klinis terhadap hasil prediksi, baik pada level global (keseluruhan model) maupun lokal (prediksi individual); dan (4) analisis fitur-fitur klinis paling berpengaruh berdasarkan nilai SHAP sebagai fondasi pengambilan keputusan medis yang transparan dan dapat dipertanggungjawabkan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning untuk membangun model klasifikasi penyakit jantung yang dapat diinterpretasi secara klinis. alur keseluruhan penelitian disajikan pada Gambar 1.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

Secara umum, tahapan penelitian terdiri dari pengumpulan dan pembersihan data (penghapusan missing value dan konversi variabel target menjadi biner), pembagian dataset menjadi data latih dan data uji dengan rasio 80:20, standardisasi fitur menggunakan Z-score, pelatihan model Random Forest baik dengan parameter default (baseline) maupun dengan optimasi GridSearchCV disertai Stratified 10-Fold Cross Validation, evaluasi performa pada data uji, hingga analisis interpretabilitas model menggunakan SHAP.



Gambar 1. Diagram Alur Penelitian

2.1 Dataset

Penelitian ini menggunakan dataset UCI Cleveland Heart Disease yang tersedia secara publik di UCI Machine Learning Repository[12]. Dataset ini terdiri dari 303 sampel dengan 13 fitur klinis dan satu variabel target, di mana variabel target asli memiliki nilai 0 - 4 yang merepresentasikan tingkat keparahan penyakit jantung. Deskripsi lengkap setiap fitur klinis disajikan pada Tabel 1.

Tabel 1. Deskripsi Fitur Dataset Uci Cleveland Heart Disease

Fitur	Deskripsi	Tipe
age	Usia pasien (tahun)	Numerik
sex	Jenis kelamin (1=pria, 0=wanita)	Kategorikal
cp	Tipe nyeri dada (0–3)	Kategorikal
trestbps	Tekanan darah istirahat (mmHg)	Numerik
chol	Kolesterol serum (mg/dL)	Numerik
fbs	Gula darah puasa >120 mg/dL (1=ya, 0=tidak)	Kategorikal
restecg	Hasil ECG istirahat (0–2)	Kategorikal
thalach	Detak jantung maksimum (bpm)	Numerik
exang	Angina akibat olahraga (1=ya, 0=tidak)	Kategorikal
oldpeak	Depresi segmen ST	Numerik
slope	Kemiringan segmen ST (0–2)	Kategorikal
ca	Jumlah pembuluh darah utama (0–3)	Numerik
thal	Kondisi thalassemia (0–3)	Kategorikal

2.2 Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum proses pelatihan model. Seluruh tahap preprocessing diterapkan setelah pembagian data untuk menghindari kebocoran informasi (data leakage) dari data uji ke data latih. Pertama, variabel target dikonversi dari skala ordinal 0–4 menjadi representasi biner, di mana nilai 0 tetap sebagai 0 (negatif) dan nilai 1–4 dikonversi menjadi 1 (positif). Kedua, penanganan missing value dilakukan dengan menghapus baris yang mengandung nilai kosong pada fitur ca (4 baris) dan thal (2 baris), sehingga jumlah sampel berkurang dari 303 menjadi 297 sampel. Pendekatan penghapusan baris dipilih karena jumlah missing value sangat kecil (kurang dari 2% dari total sampel) dan kedua fitur bersifat kategorikal sehingga imputasi berbasis mean tidak tepat secara semantik [13]. Ketiga, standardisasi fitur dilakukan menggunakan metode Z-score untuk memastikan seluruh fitur berada pada skala yang sebanding [11]. Scaler di-fit hanya pada

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

data latih kemudian digunakan untuk men-transform baik data latih maupun data uji, sehingga tidak ada informasi dari data uji yang bocor ke proses pelatihan [13], [14].

2.3 Random Forest

Random Forest merupakan algoritma ensemble berbasis pohon keputusan yang bekerja dengan membangun sejumlah pohon keputusan secara paralel pada subset data latih yang dipilih secara acak, kemudian menggabungkan hasil prediksi seluruh pohon melalui mekanisme majority voting untuk menghasilkan prediksi akhir [15], [16]. Pendekatan ensemble ini menjadikan Random Forest lebih robust dibandingkan pohon keputusan tunggal karena variasi antar pohon yang dibangun dari subset data berbeda secara efektif mengurangi risiko overfitting [15]. Dalam konteks prediksi penyakit jantung, Random Forest memiliki beberapa karakteristik yang menjadikannya relevan untuk aplikasi klinis. Pertama, algoritma ini mampu menangani data berdimensi tinggi dengan fitur yang heterogen, sebagaimana karakteristik data klinis pada umumnya. Kedua, Random Forest secara inheren menghasilkan informasi feature importance yang merepresentasikan kontribusi relatif setiap fitur terhadap prediksi model [7], [17]. Karakteristik ini menjadi fondasi bagi analisis interpretabilitas yang dilakukan pada tahap selanjutnya menggunakan SHAP.

2.4 Hyperparameter Tuning dengan GridSearchCV

GridSearchCV merupakan metode pencarian hyperparameter yang bekerja secara exhaustive dengan mengevaluasi seluruh kombinasi nilai hyperparameter yang telah didefinisikan dalam ruang pencarian [18]. Metode ini telah banyak diadopsi dalam penelitian machine learning karena kemampuannya dalam melakukan fine-tuning model dan meningkatkan akurasi prediksi, termasuk pada kasus prediksi penyakit jantung [19]. Evaluasi setiap kombinasi hyperparameter dilakukan menggunakan Stratified 10-Fold Cross Validation pada data latih. Pada mekanisme ini, dataset dibagi menjadi k lipatan yang berukuran sama, di mana satu lipatan secara bergantian berfungsi sebagai data validasi dan sisanya sebagai data latih, dengan performa rata-rata seluruh lipatan digunakan sebagai dasar pemilihan kombinasi hyperparameter terbaik [17], [18]. Hyperparameter yang dioptimasi beserta ruang pencariannya disajikan pada Tabel 2.

Tabel 2. Ruang Pencarian Hyperparameter GridSearchCV

Hyperparameter	Nilai yang Diuji	Deskripsi
n_estimators	100, 200, 300	Jumlah pohon keputusan dalam ensemble
max_depth	None, 5, 10, 20	Kedalaman maksimum setiap pohon
min_samples_split	2, 5, 10	Minimum sampel yang dibutuhkan untuk melakukan split pada node
min_samples_leaf	1, 2, 4	Minimum sampel yang dibutuhkan pada node daun
max_features	sqrt, log2	Jumlah fitur yang dipertimbangkan saat mencari split terbaik

2.5 Metrik Evaluasi

Performa model dievaluasi menggunakan lima metrik yang umum digunakan dalam klasifikasi biner pada domain medis [20]. Seluruh metrik dihitung berdasarkan komponen confusion matrix True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) yang merepresentasikan hasil prediksi model pada data uji. **Accuracy** mengukur proporsi prediksi yang benar dari keseluruhan sampel [20]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision mengukur proporsi prediksi positif yang benar-benar positif [20]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall mengukur proporsi sampel positif yang berhasil diidentifikasi dengan benar oleh model [20]:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1-Score merupakan rata-rata harmonik antara precision dan recall, memberikan keseimbangan antara keduanya [20]:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) mengukur kemampuan model dalam membedakan kelas positif dan negatif secara keseluruhan pada seluruh ambang klasifikasi, dengan nilai berkisar antara 0 hingga 1 di mana nilai 1 menunjukkan klasifikasi sempurna [20]. Kurva ROC dibentuk dengan memplot True Positive Rate (TPR) terhadap False Positive Rate (FPR) pada berbagai nilai ambang batas, yang masing-masing dirumuskan sebagai:

$$\text{TPR} = \frac{TP}{TP + FN} \quad (5)$$

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

dengan TP (True Positive), FN (False Negative), FP (False Positive), dan TN (True Negative) merepresentasikan elemen dari confusion matrix. Nilai AUC kemudian dihitung sebagai luas area di bawah kurva ROC tersebut:

$$AUC = \int_0^1 T\,PR\left(FPR^{-1}(x)\right) dx \quad (7)$$

Secara umum, nilai AUC 0,5 menunjukkan performa model setara dengan tebakan acak, sedangkan nilai mendekati 1 mengindikasikan kemampuan diskriminasi yang sangat baik antara kelas positif (penyakit jantung) dan negatif (tidak ada penyakit jantung) [20].

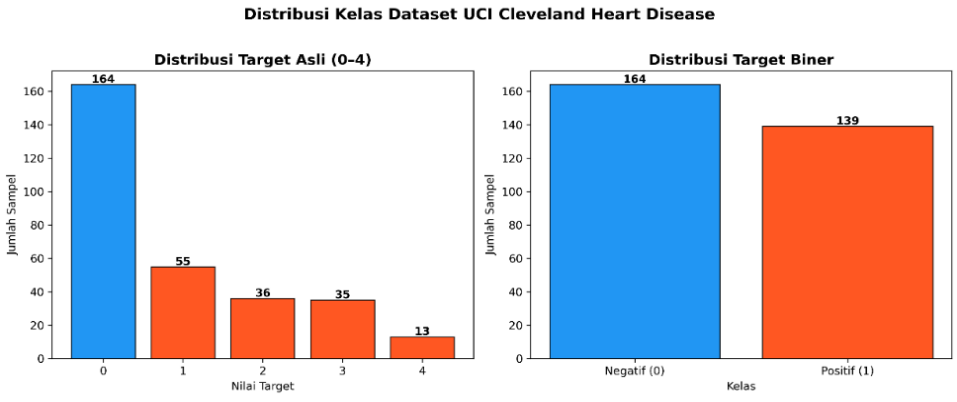
2.6 Explainability dengan SHAP

Model Random Forest yang telah dioptimasi menggunakan GridSearchCV selanjutnya diinterpretasi menggunakan SHapley Additive exPlanations (SHAP), sebuah metode explainable AI (XAI) yang dikembangkan oleh Lundberg dan Lee [10] berbasis teori permainan kooperatif Shapley yang mengukur kontribusi setiap fitur terhadap hasil prediksi secara adil dan konsisten. SHAP dipilih karena mampu menghasilkan interpretasi baik pada level global maupun lokal, serta memiliki landasan matematis yang lebih kuat dibandingkan metode interpretabilitas berbasis perturbasi atau gradien [10], [11], [21]. Untuk memastikan interpretasi global yang representatif, TreeExplainer di-fit pada model RF Tuned dan nilai SHAP dihitung menggunakan data latih (237 sampel). Penggunaan data latih untuk interpretasi global mengikuti praktik yang direkomendasikan dalam literatur XAI, di mana distribusi fitur pada data latih lebih representatif terhadap ruang fitur yang dipelajari model dibandingkan test set yang berukuran terbatas [11]. Perlu dicatat bahwa interpretasi global berbasis data latih mencerminkan pola yang dipelajari model dari data latih dan dapat berbeda dari distribusi populasi yang lebih luas, sehingga generalisasi interpretasi ini ke konteks klinis yang lebih luas memerlukan validasi pada dataset yang lebih besar. Pada level global, summary plot digunakan untuk memvisualisasikan distribusi nilai SHAP seluruh fitur pada seluruh sampel latih, sementara bar plot menampilkan rata-rata nilai absolut SHAP sebagai representasi feature importance global. Pada level lokal, analisis SHAP dilakukan menggunakan data uji (60 sampel) untuk menjelaskan prediksi individual setiap sampel. Waterfall plot digunakan untuk memvisualisasikan kontribusi setiap fitur terhadap prediksi satu sampel tertentu relatif terhadap base value model E[f(x)] = 0,465, menampilkan secara eksplisit besaran dan arah kontribusi masing-masing fitur dalam mendorong prediksi ke arah positif maupun negatif.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksplorasi Data (EDA)

Analisis eksplorasi data dilakukan pada data mentah sebelum proses preprocessing untuk memahami karakteristik dan pola distribusi dataset. Sebagaimana ditunjukkan pada Gambar 2, distribusi target biner menunjukkan dataset terdiri dari 164 sampel negatif (54,13%) dan 139 sampel positif (45,87%) pada data mentah sebelum penghapusan missing value. Setelah penghapusan 6 baris yang mengandung missing value pada tahap preprocessing, jumlah sampel positif berkurang menjadi 133, sehingga total dataset final yang digunakan untuk pemodelan adalah 297 sampel. Distribusi kelas yang relatif seimbang dengan rasio 1,18:1 ini mengindikasikan bahwa dataset tidak memerlukan teknik resampling seperti SMOTE dalam proses pemodelan.

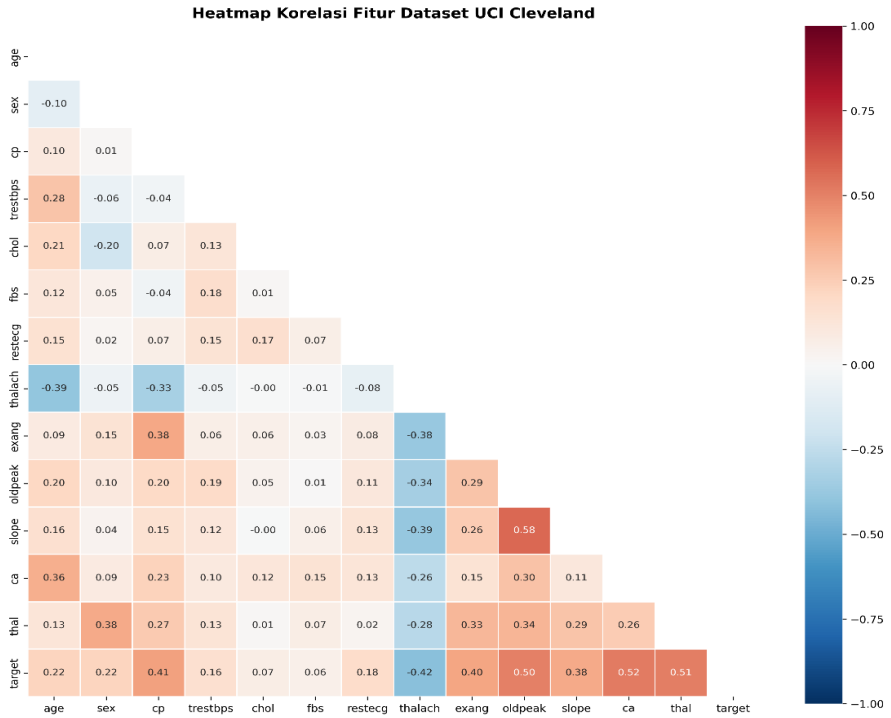


Gambar 2. Distribusi Kelas Target Dataset UCI Cleveland Heart Disease

Analisis korelasi Pearson sebagaimana ditunjukkan pada Gambar 3 mengidentifikasi fitur-fitur dengan korelasi signifikan terhadap variabel target. Thalach menunjukkan korelasi negatif tertinggi (r = -0,42),

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

mengindikasikan bahwa semakin rendah detak jantung maksimum pasien, semakin tinggi probabilitas terdeteksinya penyakit jantung. Fitur oldpeak (r = 0,50), ca (r = 0,52), dan thal (r = 0,51) menunjukkan korelasi positif yang kuat terhadap target, menjadikan ketiganya sebagai kandidat prediktor dominan. Fitur cp (r = 0,41) juga menunjukkan korelasi positif yang cukup signifikan, konsisten dengan perannya sebagai indikator klinis utama dalam diagnosis penyakit jantung. Sebaliknya, fbs (r = 0,07) dan restecg (r = 0,18) menunjukkan korelasi yang sangat lemah, mengindikasikan kontribusi diskriminatif yang terbatas. Temuan korelasi ini akan diverifikasi lebih lanjut melalui analisis SHAP pada sub-bab 3.5



Gambar 3. Heatmap Korelasi Pearson Antar Fitur Dataset UCI Cleveland Heart Disease

3.2 Hasil Preprocessing Data

Tahap preprocessing diawali dengan konversi variabel target dari skala ordinal 0–4 menjadi representasi biner, di mana nilai 0 tetap sebagai kelas negatif (tidak terdeteksi penyakit jantung) dan nilai 1–4 dikonversi menjadi kelas positif (terdeteksi penyakit jantung). Hasil konversi menghasilkan distribusi awal 164 sampel kelas negatif dan 139 sampel kelas positif dari total 303 sampel.

Pemeriksaan nilai hilang (missing value) menunjukkan terdapat 6 nilai hilang yang tersebar pada fitur ca (4 nilai) dan thal (2 nilai). Mengingat proporsi nilai hilang sangat kecil (kurang dari 2% dari total sampel) dan kedua fitur bersifat kategorikal sehingga imputasi berbasis mean tidak tepat secara semantik [15], penanganan dilakukan dengan menghapus baris yang mengandung nilai hilang. Proses ini mengurangi jumlah sampel dari 303 menjadi 297 sampel, dengan distribusi akhir 164 sampel kelas negatif dan 133 sampel kelas positif.

Dataset kemudian dibagi menggunakan metode stratified split dengan rasio 80:20, menghasilkan 237 sampel data latih dan 60 sampel data uji. Stratifikasi dilakukan untuk mempertahankan proporsi distribusi kelas pada kedua subset. Data latih terdiri dari 109 sampel positif (45,99%) dan 128 sampel negatif (54,01%), sementara data uji terdiri dari 28 sampel positif (46,67%) dan 32 sampel negatif (53,33%). Konsistensi proporsi kelas pada kedua subset mengkonfirmasi bahwa stratified split berhasil mempertahankan distribusi kelas secara representatif.

Normalisasi fitur dilakukan menggunakan metode StandardScaler (Z-score normalization) secara post-split untuk menghindari kebocoran data (data leakage). Proses fitting scaler dilakukan secara eksklusif pada data latih menggunakan parameter mean dan standar deviasi yang dipelajari dari data latih saja, kemudian parameter tersebut digunakan untuk mentransformasikan baik data latih maupun data uji. Pendekatan ini memastikan bahwa tidak ada informasi statistik dari data uji yang mempengaruhi proses pelatihan model [13]

3.3 Hasil Hyperparameter Tuning (GridSearchCV)

Proses pencarian hyperparameter dilakukan menggunakan GridSearchCV dengan ruang pencarian 216 kombinasi parameter dan Stratified 10-Fold Cross Validation pada data latih, menghasilkan total 2.160 proses fitting. Hasil pencarian menghasilkan konfigurasi hyperparameter optimal sebagai berikut: n_estimators = 100, max_depth = None, min_samples_split = 2, min_samples_leaf = 4, dan max_features = 'sqrt', dengan nilai CV AUC terbaik sebesar 0,9010.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

Konfigurasi max_depth = None mengindikasikan bahwa pohon keputusan dibiarkan tumbuh tanpa batasan kedalaman, yang merupakan karakteristik optimal untuk dataset berukuran relatif kecil di mana pembatasan kedalaman berpotensi mengorbankan kapasitas representasi model. Nilai min_samples_leaf = 4 berfungsi sebagai regularisasi implisit yang mencegah pembentukan daun pohon dari sampel yang terlalu sedikit, sehingga meningkatkan kemampuan generalisasi model. Perlu dicatat bahwa meskipun ruang pencarian mencakup kombinasi yang berpotensi menghasilkan pohon fully grown seperti max_depth = None dengan min_samples_leaf = 1, mekanisme Stratified 10-Fold Cross Validation dalam GridSearchCV secara implisit mengeliminasi kombinasi tersebut apabila menghasilkan generalisasi yang buruk pada data validasi. Terpilihnya min_samples_leaf = 4 sebagai nilai optimal mengkonfirmasi bahwa proses pencarian secara efektif menghindari konfigurasi yang berisiko overfitting. Penggunaan max_features = 'sqrt' berarti setiap pohon hanya mempertimbangkan akar kuadrat dari jumlah fitur ($\sqrt{13} \approx 3$ fitur) pada setiap split, yang memastikan diversitas antar pohon dalam ensemble tetap terjaga.

Model RF Tuned menghasilkan CV AUC sebesar 0,9010 (±0,0968), meningkat dibandingkan RF Baseline sebesar 0,8804 (±0,1115). Peningkatan CV AUC sebesar 0,0206 mengindikasikan bahwa konfigurasi hyperparameter optimal mampu menghasilkan model yang lebih stabil dan general. Standar deviasi CV AUC yang lebih kecil pada RF Tuned (0,0968 vs 0,1115) mengkonfirmasi bahwa model hasil tuning memiliki konsistensi performa yang lebih baik antar fold dibandingkan model baseline

3.4 Perbandingan Performa Model

Evaluasi performa dilakukan pada test set (60 sampel) yang tidak pernah diakses selama proses pelatihan maupun seleksi model. Hasil perbandingan performa RF Baseline dan RF Tuned disajikan pada Tabel 3.

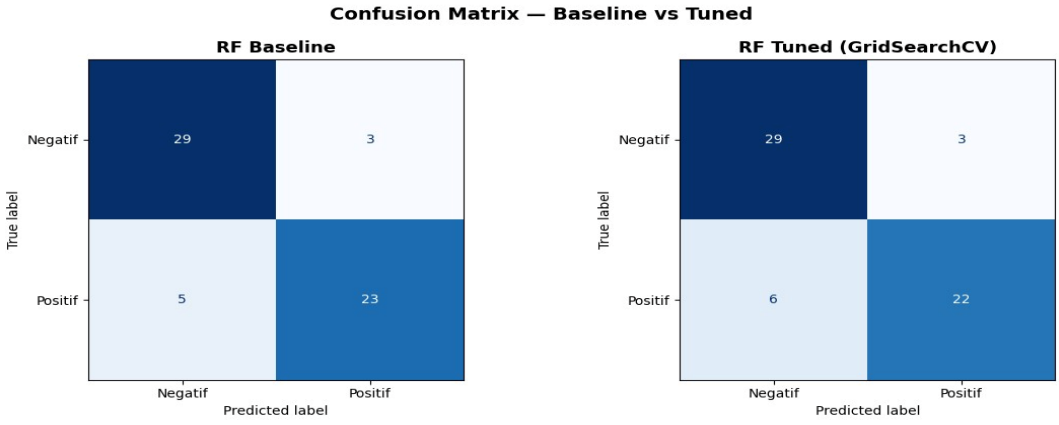
Tabel 3. Perbandingan Performa RF Baseline dan RF Tuned pada Test Set

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	CV AUC
RF Baseline	0,8667	0,8846	0,8214	0,8519	0,9414	0,8804
RF Tuned	0,8500	0,8800	0,7857	0,8302	0,9453	0,9010

Hasil evaluasi menunjukkan pola yang perlu dicermati secara kontekstual. RF Tuned mencapai AUC-ROC tertinggi sebesar 0,9453, melampaui RF Baseline (0,9414) dengan selisih 0,0039, serta CV AUC yang lebih tinggi (0,9010 vs 0,8804) dengan standar deviasi yang lebih kecil (0,0968 vs 0,1115). Kedua metrik ini mencerminkan kemampuan diskriminasi dan stabilitas model secara keseluruhan yang lebih baik pada RF Tuned, dan merupakan indikator yang lebih dapat dipercaya pada dataset berukuran kecil di mana estimasi metrik dari satu kali evaluasi pada test set 60 sampel memiliki variansi yang relatif tinggi. Di sisi lain, RF Baseline mengungguli RF Tuned pada metrik Accuracy (0,8667 vs 0,8500), Recall (0,8214 vs 0,7857), dan F1-Score (0,8519 vs 0,8302). Perbedaan Recall ini perlu diperhatikan secara klinis nilai Recall RF Tuned yang lebih rendah mengindikasikan proporsi False Negative yang lebih tinggi, yakni kasus pasien yang sebenarnya positif penyakit jantung namun tidak terdeteksi oleh model.

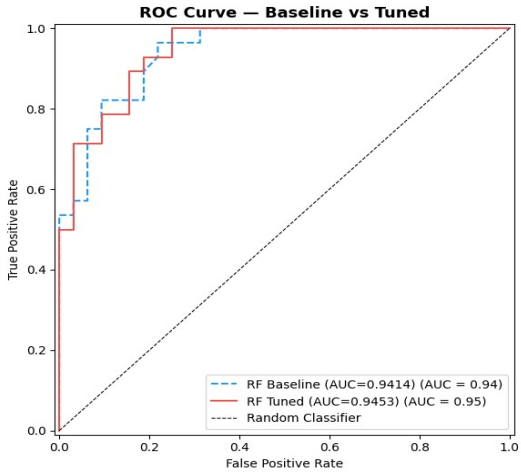
Dalam konteks skrining klinis, False Negative merupakan kesalahan yang lebih kritis dibandingkan False Positive karena berpotensi menunda penanganan medis yang diperlukan. Namun demikian, perbedaan Recall sebesar 0,0357 pada test set 60 sampel setara dengan hanya 2 sampel yang diprediksi berbeda, sehingga interpretasinya harus dilakukan dengan kehati-hatian mengingat variansi estimasi yang tinggi pada ukuran test set yang terbatas ini. Kondisi ini konsisten dengan temuan Ahamad et al. [4] yang menunjukkan bahwa perbedaan performa antar model pada dataset kecil cenderung bersifat marginal dan tidak stabil antar eksekusi.

Sebagaimana ditunjukkan pada Gambar 4, confusion matrix RF Tuned menghasilkan 29 True Negative, 22 True Positive, 3 False Positive, dan 6 False Negative pada test set. Kurva ROC pada Gambar 5 mengkonfirmasi keunggulan RF Tuned secara konsisten di seluruh rentang threshold klasifikasi dengan AUC-ROC 0,9453.



Gambar 4. Confusion Matrix RF Baseline dan RF Tuned pada Test Set

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

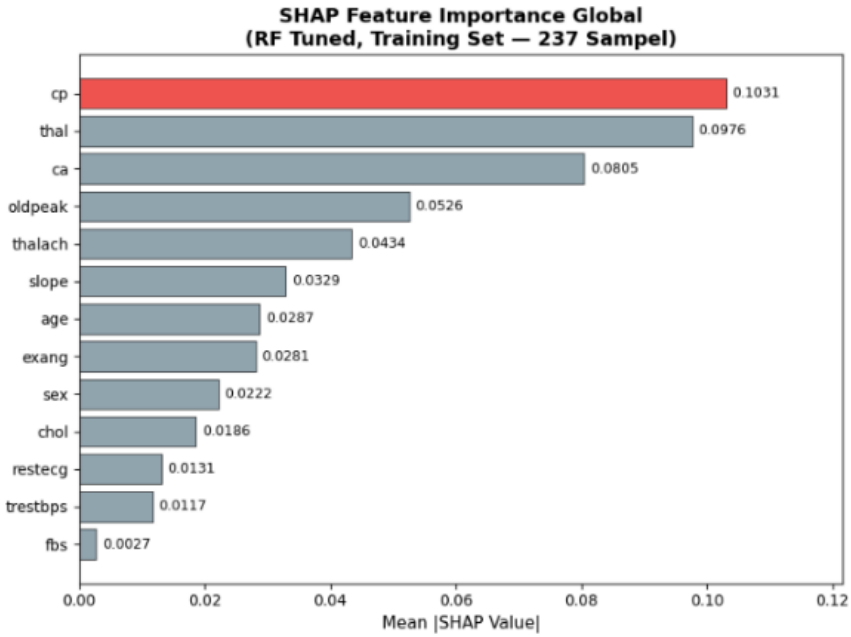


Gambar 5. ROC Curve RF Baseline dan RF Tuned pada Test Set

Berdasarkan pertimbangan tersebut, RF Tuned dipilih sebagai model terbaik dengan justifikasi berikut: pertama, AUC-ROC dan CV AUC merupakan metrik yang lebih robust terhadap variansi ukuran test set kecil dibandingkan Accuracy dan Recall yang sangat sensitif terhadap perubahan satu sampel; kedua, CV AUC yang lebih tinggi (0,9010 vs 0,8804) dengan standar deviasi lebih kecil mencerminkan generalisasi model yang lebih stabil secara keseluruhan; ketiga, nilai utama GridSearchCV bukan semata peningkatan metrik point-estimate, melainkan kemampuan menghasilkan konfigurasi hyperparameter yang deterministik dan reproducible karakteristik yang esensial dalam penelitian klinis. Meskipun demikian, penelitian lanjutan dengan dataset yang lebih besar diperlukan untuk memvalidasi apakah keunggulan AUC RF Tuned diikuti dengan peningkatan Recall yang konsisten, sehingga model dapat diadopsi secara penuh dalam konteks skrining klinis.

3.5 Analisis SHAP

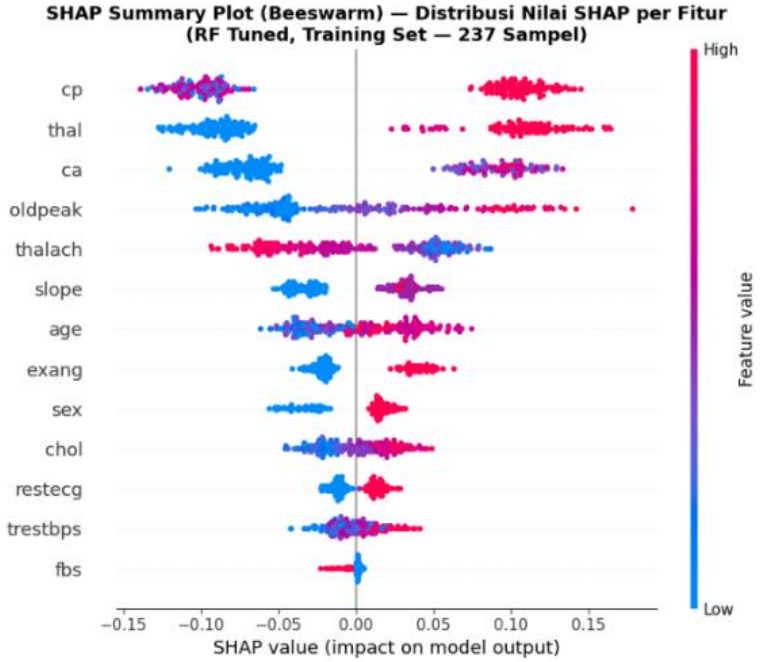
Analisis SHAP diterapkan pada RF Tuned menggunakan TreeExplainer dengan dua skema perhitungan: nilai SHAP global dihitung pada data latih (237 sampel) untuk memastikan representativitas interpretasi, sedangkan nilai SHAP lokal dihitung pada data uji (60 sampel) untuk menjelaskan prediksi individual. Base value model E[f(x)] = 0,465 merepresentasikan rata-rata probabilitas prediksi model pada data latih. Pada level global, Gambar 6 mengidentifikasi cp (tipe nyeri dada) sebagai fitur paling berpengaruh dengan mean |SHAP value| = 0,1031, diikuti thal (kondisi thalassemia) sebesar 0,0976 dan ca (jumlah pembuluh darah utama) sebesar 0,0805. Ketiga fitur ini mendominasi dengan kontribusi yang jauh melampaui fitur lainnya, sementara fbs (gula darah puasa) menunjukkan kontribusi paling minimal (0,0027), konsisten dengan temuan korelasi pada sub-bab 3.1 [22]. menunjukkan pola berlawanan di mana nilai tinggi justru mendorong prediksi ke arah negatif mengkonfirmasi bahwa detak jantung maksimum yang lebih tinggi berasosiasi dengan kondisi jantung yang lebih sehat.



Gambar 6. SHAP Feature Importance Global (RF Tuned, Data latih 237 Sampel)

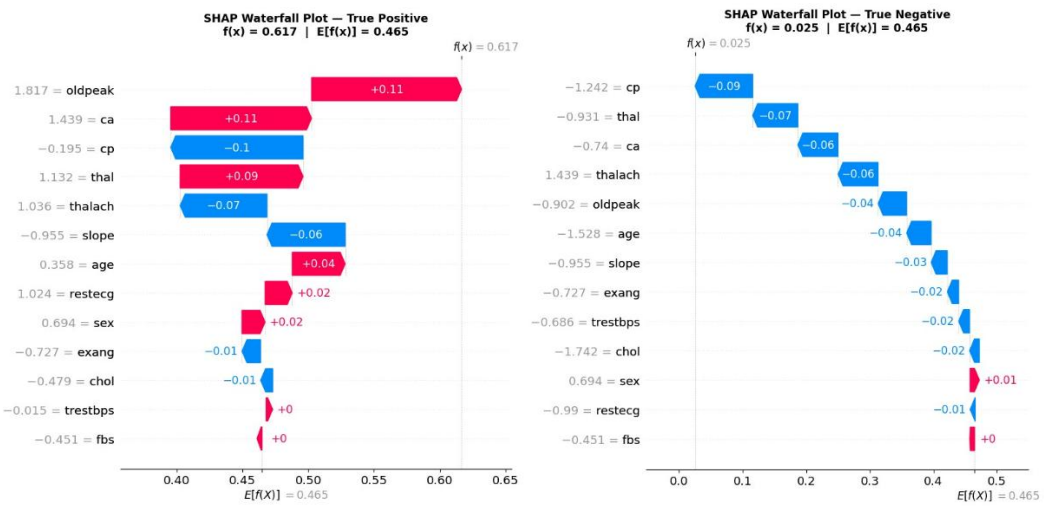
Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

Pada Gambar 7 melengkapi temuan ini dengan mengungkapkan pola arah kontribusi setiap fitur secara simultan: nilai cp, thal, dan ca yang tinggi secara konsisten mendorong prediksi ke arah positif, sedangkan thalach menunjukkan pola berlawanan di mana nilai tinggi justru mendorong prediksi ke arah negatif mengkonfirmasi bahwa detak jantung maksimum yang lebih tinggi berasosiasi dengan kondisi jantung yang lebih sehat.



Gambar 7. SHAP Beeswarm Plot Distribusi Nilai SHAP Seluruh Fitur (RF Tuned, Data latih 237 Sampel)

Pada level lokal, analisis SHAP dilakukan menggunakan data uji (60 sampel) untuk menjelaskan prediksi individual. Gambar 8 menampilkan waterfall plot berdampingan untuk satu sampel True Positive (kiri) dan satu sampel True Negative (kanan) yang representatif dari data uji. Pada sampel True Positive, fitur oldpeak (+0,11) dan ca (+0,11) memberikan kontribusi positif terbesar, mendorong probabilitas prediksi dari base value E[f(x)] = 0,465 menuju f(x) = 0,617 sehingga model mengklasifikasikan sampel tersebut sebagai positif penyakit jantung. Pola ini konsisten secara klinis: depresi segmen ST yang tinggi (oldpeak) merupakan indikator iskemia miokardial yang telah diakui dalam kardiologi, sementara jumlah pembuluh darah utama yang tersumbat (ca) berkorelasi langsung dengan keparahan aterosklerosis koroner. Sebaliknya, pada sampel True Negative, fitur cp (−0,09), thal (−0,07), ca (−0,06), dan thalach (−0,06) secara konsisten mendorong prediksi ke arah negatif, menurunkan probabilitas dari base value 0,465 menuju f(x) = 0,025. Tidak adanya nyeri dada tipikal (cp rendah) dan profil thalassemia normal (thal rendah) mencerminkan profil klinis pasien dengan risiko kardiovaskular rendah. Kemampuan model menghasilkan penjelasan individual yang konsisten secara klinis pada kedua kelas mendemonstrasikan bahwa integrasi SHAP tidak hanya menghasilkan interpretasi global yang agregat, tetapi juga mampu mendukung pengambilan keputusan klinis pada level pasien individual.



Gambar 8. SHAP Waterfall Plot Prediksi Individual: Sampel True Positive (kiri, f(x) = 0,617) dan True Negative (kanan, f(x) = 0,025) pada Data Uji (RF Tuned)

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

3.6 Discussion

Perbandingan performa model dengan studi-studi sebelumnya disajikan pada Tabel 4. Perlu dicatat beberapa perbedaan metodologis yang mempengaruhi komparabilitas hasil antar studi. Al Azhima et al. [5] dan Ahamad et al. [4] menggunakan dataset UCI Cleveland yang identik dengan penelitian ini, dengan akurasi masing-masing 84,48% dan 84,62% tanpa teknik resampling. Masbakhah et al. [7] melaporkan akurasi tertinggi sebesar 97,83% menggunakan RF-FOX, namun hasil ini diperoleh dengan penerapan SMOTE-NC untuk oversampling data training sehingga tidak dapat dibandingkan secara langsung dengan model yang dilatih pada distribusi kelas asli. Alfajr & Defiyanti [6] mencapai akurasi 98% namun menggunakan dataset gabungan berukuran 1026 sampel yang berbeda secara komposisi dari UCI Cleveland 303 sampel yang digunakan penelitian ini.

Tabel 4. Perbandingan Performa Model Prediksi Penyakit Jantung Berbasis Random Forest dengan Studi Terdahulu

Penelitian	Model	Dataset	Sampel	Split	SMOTE	Accuracy	AUC-ROC
Al Azhima et al. [5]	RF + LR Hybrid	UCI Cleveland	303	70:30	Tidak	84,48%	-
Ahamad et al. [4]	RF + GridSearchCV	UCI Cleveland	303	70:30	Tidak	84,62%	-
Masbakhah et al.[7]	RF + FOX	UCI Cleveland	303	80:20	Ya	97,83%	-
Alfajr & Defiyanti [6]	RF + PCA	UCI Gabungan	1026	67:33	Tidak	98,00%	-
Penelitian ini	RF + GridSearchCV + SHAP	UCI Cleveland	297	80:20	Tidak	85,00%	0,9453

Model RF Tuned pada penelitian ini mencapai akurasi 85,00% dengan AUC-ROC 0,9453, satu-satunya metrik AUC yang dilaporkan di antara seluruh studi pembanding. Meskipun akurasi penelitian ini lebih rendah dibandingkan studi yang menggunakan SMOTE atau dataset lebih besar, perbedaan ini tidak mengindikasikan inferioritas model, melainkan mencerminkan pilihan metodologis yang lebih konservatif: tidak menggunakan data sintetis, pipeline preprocessing post-split yang mencegah data leakage, serta integrasi SHAP untuk interpretabilitas klinis yang tidak tersedia pada studi-studi pembanding tersebut.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi penyakit jantung berbasis Random Forest yang dioptimasi menggunakan GridSearchCV (216 kombinasi parameter, Stratified 10-Fold Cross Validation) dengan analisis SHAP untuk interpretabilitas klinis pada dataset UCI Cleveland Heart Disease (297 sampel, 13 fitur klinis). Konfigurasi optimal yang dihasilkan (n_estimators=100, max_depth=None, min_samples_leaf=4, min_samples_split=2, max_features='sqrt') menghasilkan model RF Tuned dengan AUC-ROC 0,9453 dan CV AUC 0,9010, lebih unggul dan lebih stabil dibandingkan RF Baseline. Meskipun nilai Recall RF Baseline sedikit lebih tinggi (0,8214 vs 0,7857), perbedaan ini hanya setara 2 sampel pada test set 60 sampel sehingga tidak dapat diinterpretasikan sebagai keunggulan yang stabil, mengingat keterbatasan ukuran test set dan belum adanya uji signifikansi statistik formal. Nilai utama hyperparameter tuning pada dataset berukuran kecil ini terletak pada peningkatan stabilitas, konsistensi, dan reproducibility model, karakteristik yang esensial bagi penelitian klinis.

Analisis SHAP pada data latih (237 sampel) mengidentifikasi cp (mean |SHAP|=0,1031), thal (0,0976), dan ca (0,0805) sebagai fitur paling berpengaruh, selaras dengan indikator diagnostik kardiologi dan dikonfirmasi melalui waterfall plot yang menunjukkan konsistensi fitur dominan (cp, thal, ca, oldpeak) pada level prediksi individual. Integrasi GridSearchCV dan SHAP menunjukkan bahwa optimasi hyperparameter yang deterministik dan interpretabilitas model dapat dicapai secara bersamaan, sehingga model yang dihasilkan berpotensi mendukung sistem pendukung keputusan klinis yang transparan.

Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan beragam, mengeksplorasi RandomizedSearchCV atau Bayesian Optimization, serta menerapkan kalibrasi probabilitas dan optimasi threshold klasifikasi guna meningkatkan Recall model dalam konteks skrining klinis.

REFERENCES

[1] M. Di Cesare et al., “World-Heart-Report-2023,” Geneva, 2023. [Online]. Available: https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf
[2] Global Burden of Cardiovascular Diseases and Risks 2023 et al., “Global, Regional, and National Burden of Cardiovascular Diseases and Risk Factors in 204 Countries and Territories, 1990-2023,” J. Am. Coll. Cardiol., vol. 86, no. 22, pp. 2167–2243, 2025, doi: 10.1016/j.jacc.2025.08.015.
[3] L. H. Goh et al., “The epidemiology and burden of cardiovascular diseases in countries of the Association of Southeast Asian Nations (ASEAN), 1990–2021: findings from the Global Burden of Disease Study 2021,” Lancet Public Health, vol. 10, no. 6, pp. e467–e479, 2025, doi: 10.1016/S2468-2667(25)00087-8.
[4] G. N. Ahamad et al., “Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease,” Processes, vol. 11, no. 3, 2023, doi: 10.3390/pr11030734.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1267−1277
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10703

[5] S. A. T. Al Azhima, D. Darmawan, N. F. A. Hakim, I. Kustiawan, and M. Al Qibtiya, "Hybrid Machine Learning Model Untuk Memprediksi Penyakit Jantung Dengan Metode Logistic Regression Dan Random Forest," Jurnal Teknologi Terpadu, vol. 8, no. 1, pp. 40–46, 2022
[6] 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, 2024, doi: 10.23960/jitet.v12i3s1.5055.
[7] A. Masbakhah, U. Sa'adah, and M. Muslikh, "Heart Disease Classification Using Random Forest and Fox Algorithm as Hyperparameter Tuning," Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 7, no. 4, pp. 964–976, 2025, doi: 10.35882/jeeemi.v7i4.932.
[8] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," IEEE Access, vol. 6, no. c, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
[9] Md. A. Talukder, A. S. Talaat, M. Kazi, and A. Khraisat, "XAI-HD: an explainable artificial intelligence framework for heart disease detection," Artif. Intell. Rev., vol. 58, no. 12, p. 385, 2025, doi: 10.1007/s10462-025-11385-6.
[10] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017, pp. 4765–4774. doi: 10.48550/arXiv.1705.07874.
[11] 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, pp. 1–18, 2024, doi: 10.1038/s41598-024-74656-2.
[12] A. Janosi, M. Steinbrunn, William Pfisterer, and R. Detrano, "Heart Disease," UCI Machine Learning Repository. [Online]. Available: https://archive.ics.uci.edu/dataset/45/heart+disease
[13] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems, 3rd ed. Sebastopol: O'Reilly Media, 2022.
[14] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
[15] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[16] N. Nasution, M. A. Hasan, and F. Bakri Nasution, "Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset," IT Journal Research and Development, vol. 9, no. 2, pp. 140–150, 2025, doi: 10.25299/itjrd.2025.17941.
[17] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in Forests of randomized trees," in Advances in Neural Information Processing Systems, 2013, pp. 1–9. [Online]. Available: https://www.researchgate.net/publication/264046801_Understanding_variable_importances_in_Forests_of_randomized_trees
[18] R. Muzayanah, D. A. A. Pertiwi, M. Ali, and M. A. Muslim, "Comparison of gridsearchcv and bayesian hyperparameter optimization in random forest algorithm for diabetes prediction," Journal of Soft Computing Exploration, vol. 5, no. 1, pp. 86–91, 2024, doi: 10.52465/joscex.v5i1.308.
[19] S. Rasheed, G. K. Kumar, D. M. Rani, M. V. V. P. Kantipudi, and M. Anila, "Heart Disease Prediction Using GridSearchCV and Random Forest," EAI Endorsed Trans. Pervasive Health Technol., vol. 10, pp. 1–8, 2024, doi: 10.4108/eetpht.10.5523.
[20] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," African Journal of Biomedical Research, vol. 27, no. 4, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
[21] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed. Christoph Molnar, 2022. [Online]. Available: https://christophm.github.io/interpretable-ml-book/
[22] A. Yazdani, K. D. Varathan, Y. K. Chiam, A. W. Malik, W. Azman, and W. Ahmad, "A novel approach for heart disease prediction using strength scores with significant predictors," BMC Med. Inform. Decis. Mak., vol. 21, no. 1, p. 194, 2021, doi: 10.1186/s12911-021-01527-5.