

Fitur Penting dan *Explainable AI* dalam Deteksi Penyakit Jantung berbasis *Machine Learning*

Feature Selection and Explainable AI for Heart Disease Detection using Machine Learning

¹Resky Ayu Dewi Talasari*, ²Ayutri Wahyuni, ³Clara Diva, ⁴Muhammad Nur Alamsyah Rajab

^{1,2}Program Teknik Studi Informatika, Fakultas Teknik, Universitas Fajar

³Program Studi Teknik Informatika, Fakultas Informatika dan Komputer, Universitas Kristen Indonesia Paulus

⁴Program Studi Kedokteran, Fakultas Kedokteran dan Ilmu Kesehatan (FKIK), Universitas Islam Negeri Alauddin

^{1,2}Jl. Prof. Dr. H. Abdurrahman Basalamah No. 101 (Ex. Jl. Racing Center), Panakkukang, Kota Makassar, Sulawesi Selatan 90231

³Jl. Perintis Kemerdekaan KM 13, Daya, Kecamatan Mariso, Kota Makassar, Sulawesi Selatan.

⁴Jl. H.M. Yasin Limpo No. 36, Kelurahan Romang Polong

*e-mail: reskytalasari04@gmail.com

(*received*: 28 May 2026, *revised*: 15 June 2026, *accepted*: 16 June 2026)

Abstrak

Deteksi dini penyakit jantung penting untuk mendukung intervensi klinis, mempercepat penanganan, dan meningkatkan kualitas perawatan pasien. Penelitian ini membandingkan tiga algoritma machine learning, yaitu Random Forest, XGBoost, dan Support Vector Machine, yang dikombinasikan dengan dua metode seleksi fitur, yaitu Chi-Square dan Recursive Feature Elimination (RFE), pada dataset Heart Disease UCI. Evaluasi dilakukan terhadap enam skenario pemodelan menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Random Forest memperoleh performa terbaik dengan akurasi 85,2% dan recall 97,0%, sehingga memiliki kemampuan yang baik dalam mendeteksi pasien dengan indikasi penyakit jantung. SHAP digunakan sebagai pendekatan Explainable AI dan diintegrasikan ke dalam sistem berbasis web untuk menjelaskan hasil prediksi secara lebih transparan. Prediksi sistem diposisikan sebagai alat bantu keputusan awal, bukan sebagai pengganti diagnosis dokter.

Kata kunci: kecerdasan buatan yang dapat dijelaskan, seleksi fitur, prediksi penyakit jantung, pembelajaran mesin

Abstract

Early detection of heart disease is essential for supporting timely clinical intervention, improving treatment outcomes, and enhancing the quality of patient care. This study compares the performance of three machine learning algorithms—Random Forest, XGBoost, and Support Vector Machine (SVM)—combined with two feature selection methods, Chi-Square and Recursive Feature Elimination (RFE), using the UCI Heart Disease dataset. Six modeling scenarios were evaluated based on accuracy, precision, recall, and F1-score. The experimental results demonstrate that the Random Forest model achieved the best overall performance, with an accuracy of 85.2% and a recall of 97.0%, indicating a strong capability to identify patients with potential heart disease. To enhance model transparency and interpretability, SHAP (SHapley Additive exPlanations) was employed as an Explainable AI (XAI) technique and integrated into a web-based decision support system to provide intuitive explanations of prediction outcomes. The proposed system is intended to serve as an initial clinical decision-support tool and is not designed to replace diagnosis or clinical judgment by healthcare professionals.

Keywords: explainable AI (XAI), feature selection, heart disease prediction, machine learning

1 Pendahuluan

Penyakit jantung (CVD) merupakan penyebab utama kematian di seluruh dunia, termasuk Penyakit Jantung Iskemik (IHD), yang menempatkannya sebagai masalah kesehatan global yang sangat serius [1]. Oleh karena itu, mendeteksi penyakit jantung pada tahap awal dan mendiagnosisnya secara tepat adalah langkah yang sangat penting untuk mengurangi risiko kematian dan meningkatkan harapan hidup pasien [2].

Dalam upaya meningkatkan akurasi diagnosis, implementasi berbagai metode Machine Learning telah terbukti transformatif dalam menganalisis data medis yang kompleks, seperti rekam medis, hasil pencitraan (EKG, MRI), dan sinyal jantung (PCG)[3][4]. Sistem ini membantu para profesional kesehatan dalam membantu memprediksi risiko dan mendukung proses diagnosis [5].

Meskipun sistem berbasis machine learning menunjukkan potensi besar untuk diagnosis CVD, data medis seringkali sangat besar dan memiliki banyak variabel (faktor diagnostik) [3]. Tidak semua faktor ini sama pentingnya beberapa mungkin tidak relevan atau bahkan mengganggu proses prediksi. Jika terlalu banyak faktor yang tidak penting digunakan, model prediksi bisa menjadi kurang akurat, lebih lambat dan sulit dipahami [1].

Oleh karena itu, langkah utama dalam membangun model prediksi adalah Feature Ranking atau *Feature Selection*. Tujuannya adalah untuk mengidentifikasi dan memilih sekelompok kecil faktor diagnostik yang paling berpengaruh dan signifikan dalam membedakan pasien sehat dari pasien yang berisiko penyakit jantung berdasarkan riwayat kesehatan pasien [6].

Metode *Chi-Square* merupakan salah satu metode *Feature Selection* yang berperan dalam mengukur seberapa kuat hubungan antara setiap faktor diagnostik (misalnya, jenis kelamin, riwayat merokok) dengan hasil penyakit jantung [7]. Metode ini membantu menyaring faktor-faktor yang memiliki keterkaitan statistik yang jelas dengan kondisi target, dengan mengidentifikasi faktor-faktor kategorikal yang paling penting [7] [8].

Penerapan *Chi-Square* dalam deteksi penyakit jantung telah terbukti efektif dalam mengidentifikasi subset fitur optimal yang dapat meningkatkan akurasi model klasifikasi dan juga mempercepat proses training model [9].

Recursive Feature Elimination (RFE) adalah metode yang bekerja dengan menguji faktor-faktor secara langsung pada model klasifikasi [10]. RFE secara bertahap menghilangkan faktor yang paling tidak penting berdasarkan seberapa besar pengaruhnya terhadap kinerja model [11]. Proses ini berulang (rekursif) hingga ditemukan subset faktor optimal yang memberikan akurasi terbaik. RFE sangat efektif untuk menemukan kombinasi faktor terbaik dan telah banyak digunakan dalam studi CVD untuk meningkatkan akurasi prediksi dengan mengurangi dimensi data [12].

SHapley Additive exPlanations (SHAP) adalah salah satu metode *Explainable AI* (XAI) yang tidak hanya memilih faktor, tetapi juga menjelaskan mengapa model membuat keputusan tertentu dan seberapa besar kontribusi setiap faktor terhadap prediksi akhir [13]. SHAP memberikan nilai kontribusi yang adil dan konsisten untuk setiap faktor diagnostik, yang sangat penting untuk membangun kepercayaan klinis [14][15].

Dalam konteks klinis, SHAP sangat penting karena membantu mengatasi masalah black box [13]. Dengan SHAP, kontribusi setiap faktor diagnostik dapat diukur secara adil, memberikan pemahaman yang lebih dalam kepada dokter mengenai faktor risiko mana yang paling memengaruhi hasil prediksi pasien [16].

Penelitian ini bertujuan membandingkan kinerja *Random Forest*, *Support Vector Machine*, dan *XGBoost* dalam deteksi dini penyakit jantung. Penelitian ini juga mengidentifikasi fitur kesehatan yang paling berpengaruh menggunakan SHAP, sehingga hasil prediksi model dapat dijelaskan secara lebih transparan. Model dengan kinerja terbaik kemudian diintegrasikan ke dalam website sederhana agar pengguna dapat memasukkan data kesehatan aktual dan memperoleh hasil deteksi dini secara langsung. Hasil penelitian ini diharapkan dapat menunjukkan model paling efektif serta faktor utama yang berkontribusi terhadap deteksi dini penyakit jantung.

2 Tinjauan Literatur

Penelitian mengenai deteksi penyakit jantung koroner (PJK) berbasis *machine learning* menunjukkan bahwa seleksi fitur berperan penting dalam meningkatkan kualitas prediksi. Pathan *et al.* [9] menggunakan ANOVA-F untuk mengidentifikasi faktor risiko yang paling relevan, seperti

<http://sistemasi.ftik.unisi.ac.id>

usia, hipertensi, riwayat penyakit jantung, dan kadar glukosa rata-rata. Nancy *et al.* [17] menerapkan *Correlation based Feature Selection* (CFS) bersama SVM untuk memilih variabel yang memiliki keterkaitan kuat dengan keluaran prediksi. Sedangkan Mahdi, *et al.* [18] menggunakan *Recursive Feature Elimination* (RFE) untuk memperoleh subset fitur yang lebih optimal. Ketiga penelitian tersebut menegaskan bahwa pemilihan fitur dapat membantu menyederhanakan data dan mendukung performa model. Namun, pembahasannya masih lebih banyak diarahkan pada peningkatan akurasi, belum pada penjelasan mengapa fitur tertentu berpengaruh terhadap hasil prediksi.

Pada sisi algoritma, beberapa penelitian menunjukkan bahwa model machine learning memiliki kemampuan yang berbeda dalam mendeteksi PJK. Teja *et al.* [19] melaporkan bahwa *Random Forest*, *XGBoost*, dan *Bagged Tree* memberikan hasil klasifikasi yang baik dengan tingkat kesalahan yang rendah. Sanjay *et al.* [20] membandingkan *Decision Tree*, *Random Forest*, dan *XGBoost*, dan menunjukkan bahwa model berbasis ensemble cenderung lebih kompetitif dalam menangani pola data penyakit jantung. Sementara itu, Krittanawong *et al.* [21] menekankan potensi SVM dan algoritma boosting dalam prediksi penyakit jantung. Temuan tersebut menunjukkan bahwa pemilihan algoritma sangat menentukan performa klasifikasi. Akan tetapi, kajian yang hanya berfokus pada perbandingan akurasi belum cukup untuk menjelaskan dasar keputusan model, terutama ketika model digunakan pada konteks kesehatan.

Kebutuhan terhadap transparansi hasil deteksi model mulai dijawab melalui penerapan *Explainable Artificial Intelligence* (XAI). Napa *et al.* [22] menggunakan *SHAP* dan *Partial Dependence Plots* untuk menunjukkan kontribusi fitur klinis terhadap prediksi risiko kardiovaskular. *SHAP* memungkinkan identifikasi fitur yang paling mendorong atau menurunkan hasil prediksi, baik secara umum maupun pada kasus individual. Pendekatan ini relevan karena model prediksi di bidang medis perlu memiliki dasar keputusan yang dapat ditelusuri, bukan hanya menunjukkan nilai performa yang tinggi.

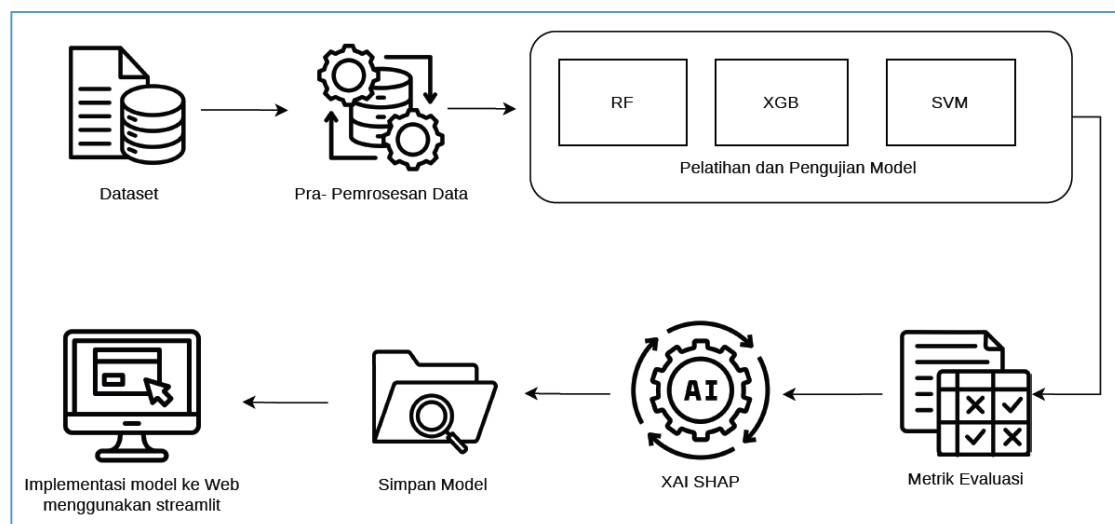
Dari penelitian terdahulu, seleksi fitur telah digunakan untuk meningkatkan efisiensi dan *performa* model, perbandingan algoritma telah dilakukan untuk menentukan model prediksi yang lebih unggul dan XAI telah dimanfaatkan untuk menjelaskan hasil prediksi. Namun, masih terdapat ruang kajian pada hubungan antara ketiga aspek tersebut dalam deteksi PJK. Penelitian sebelumnya belum secara jelas memperlihatkan bagaimana fitur yang telah dipilih berkontribusi terhadap kinerja beberapa algoritma sekaligus bagaimana kontribusi tersebut dapat dijelaskan melalui *SHAP*.

Penelitian ini menguji *Random Forest*, *XGBoost*, dan SVM untuk deteksi PJK. Kinerja ketiga model dibandingkan, sedangkan *SHAP* digunakan untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi. Dengan demikian, penelitian ini tidak hanya menilai performa klasifikasi, tetapi juga mengidentifikasi fitur yang paling berpengaruh dalam pembentukan keputusan model.

3 Metode Penelitian

Dalam eksperimen ini, deteksi penyakit jantung dilakukan melalui beberapa tahap. Proses dimulai dari persiapan dan prapemrosesan dataset, dilanjutkan dengan seleksi fitur untuk menentukan fitur yang paling relevan dalam proses klasifikasi. Selanjutnya, beberapa model *machine learning* dikembangkan dan diuji untuk menilai kemampuan prediksinya. Kinerja model dievaluasi menggunakan sejumlah metrik agar hasil pengujian dapat diukur secara objektif. Pada tahap akhir, model terpilih diimplementasikan ke dalam sistem berbasis web sehingga hasil prediksi dapat digunakan secara lebih praktis. Selain menekankan akurasi, penelitian ini juga menerapkan pendekatan *Explainable AI* untuk menjelaskan keputusan model secara lebih transparan.

Gambar 1 menunjukkan alur eksperimen yang digunakan dalam penelitian ini. Alur tersebut mencakup tahapan utama dalam pengembangan sistem deteksi dini penyakit jantung berbasis *machine learning*, yaitu pengumpulan dataset, prapemrosesan data, seleksi fitur, pengembangan model, evaluasi model, dan implementasi model ke dalam sistem berbasis web. Dataset yang digunakan dalam eksperimen ini diperoleh dari Kaggle, yaitu *Heart Disease UCI dataset*, yang memuat sejumlah fitur klinis pasien serta label target yang menunjukkan ada atau tidaknya *penyakit jantung*.



Gambar 1 Alur penelitian sistem deteksi dini penyakit jantung

Prapemrosesan data dilakukan untuk memastikan bahwa data berada dalam kondisi yang sesuai dan siap digunakan pada proses pemodelan. Pada tahap ini, dataset diperiksa untuk mengidentifikasi *missing value*, memastikan konsistensi tipe data, serta memahami karakteristik setiap fitur yang digunakan. Selain itu, fitur dan label target dipisahkan, kemudian data diproses agar dapat digunakan oleh algoritma *machine learning* yang dipilih. Tahap ini penting karena kualitas data berpengaruh langsung terhadap kemampuan model dalam menghasilkan prediksi yang akurat.

Setelah data siap digunakan, tahap berikutnya adalah seleksi fitur untuk menentukan fitur yang paling relevan dalam prediksi penyakit jantung. Seleksi fitur dilakukan agar model bekerja lebih efisien dengan memanfaatkan fitur yang memberikan kontribusi penting terhadap proses klasifikasi. Selanjutnya, model *machine learning* dikembangkan menggunakan algoritma yang telah ditentukan dan dilatih dengan data latih.

Model yang telah dilatih kemudian dievaluasi menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Metrik tersebut digunakan untuk menilai kemampuan model dalam membedakan pasien yang termasuk kategori Terdapat indikasi ada penyakit jantung dan Tidak ada indikasi penyakit jantung.

Pada tahap akhir, hasil evaluasi model digunakan sebagai dasar untuk menentukan model terbaik. Model dengan performa paling baik kemudian disimpan dan diimplementasikan ke dalam sistem berbasis web menggunakan Streamlit. Melalui sistem ini pengguna dapat memasukkan data klinis pasien untuk memperoleh prediksi awal, kemudian sistem memproses data tersebut dan menghasilkan prediksi awal terkait kemungkinan adanya penyakit jantung. Dengan demikian, metodologi pada Gambar 1 menggambarkan alur eksperimen mulai dari pengolahan data hingga penerapan model dalam bentuk aplikasi prediksi berbasis web. Penjelasan lebih rinci mengenai setiap tahap metodologi disajikan pada Bagian A sampai F.

A. Dataset

Dataset yang digunakan dalam penelitian ini adalah Heart Disease UCI yang diperoleh melalui Kaggle dan merujuk pada basis data UCI Machine Learning Repository. Dataset ini terdiri atas 303 data pasien dengan 14 atribut, yaitu 13 fitur klinis dan 1 label target. Label target menunjukkan kondisi pasien, yaitu 0 untuk pasien tanpa indikasi penyakit jantung dan 1 untuk pasien dengan indikasi penyakit jantung. Berdasarkan distribusi kelas, dataset terdiri atas 165 data pada kelas 1 dan 138 data pada kelas 0. Meskipun basis data asli Heart Disease UCI memiliki 76 atribut, sebagian besar penelitian sebelumnya menggunakan subset 14 atribut untuk analisis prediksi penyakit jantung. Fitur-fitur tersebut mencakup usia, jenis kelamin, tipe nyeri dada, tekanan darah istirahat, kadar kolesterol, gula darah puasa, hasil elektrokardiogram istirahat, detak jantung maksimum, angina akibat olahraga, oldpeak, slope, jumlah pembuluh darah utama, dan thal. Fitur-fitur tersebut digunakan karena relevan dalam menggambarkan kondisi klinis dan faktor risiko yang berkaitan dengan penyakit jantung.

Tabel 1 adalah penjelasan fitur yang terdapat dalam dataset

Fitur dalam Dataset	Tipe data	Deskripsi	Relevansi dengan Prediksi Penyakit Jantung
Age	Int	Usia pasien dalam tahun.	Risiko penyakit jantung meningkat seiring bertambahnya usia [23][24].
Sex	Int	Jenis kelamin pasien (1 = Laki-laki, 0 = Perempuan).	Jenis kelamin memengaruhi risiko dan manifestasi penyakit jantung; laki-laki umumnya memiliki risiko lebih tinggi pada usia muda [25].
Cp	Int	Tipe nyeri dada (Chest Pain Type). Nilai kategorikal (misalnya, 0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic).	Tipe nyeri dada adalah indikator yang sering menjadi fitur paling penting dalam klasifikasi penyakit jantung [26].
Threstbps	Int	Tekanan darah istirahat (Resting Blood Pressure) dalam mmHg.	Tekanan darah tinggi (Hipertensi) adalah faktor risiko utama penyakit jantung koroner dan sering digunakan sebagai prediktor [18][27].
Cho	Int	Kolesterol serum dalam mg/dl.	Kadar kolesterol tinggi (Hiperkolesterolemia) berkontribusi pada pembentukan plak aterosklerosis, yang merupakan penyebab utama penyakit jantung [18][27].
Fbs	Int	Gula darah puasa (Fasting Blood Sugar) > 120 mg/dl (1 = Benar, 0 = Salah).	Digunakan untuk mengidentifikasi pasien dengan diabetes, yang merupakan faktor risiko signifikan untuk penyakit jantung [18].
Restecg	Int	Hasil elektrokardiografi istirahat (Resting Electrocardiographic Results). Nilai kategorikal (misalnya, 0: normal, 1: kelainan gelombang ST-T, 2: hipertrofi ventrikel kiri).	Kelainan pada hasil EKG istirahat dapat mengindikasikan kerusakan atau ketegangan pada otot jantung [27].
Thalach	Int	Detak jantung maksimum yang dicapai (Maximum Heart Rate Achieved).	Detak jantung maksimum yang lebih rendah dari yang diharapkan selama aktivitas fisik sering dikaitkan dengan penyakit jantung [28][29].
Exang	Int	Angina yang diinduksi oleh olahraga (Exercise Induced Angina) (1 = Ya, 0 = Tidak).	Angina yang dipicu oleh aktivitas fisik adalah gejala klasik dari penyempitan arteri koroner yang signifikan [2] [29].
Oldpeak	Float	Depresi ST yang diinduksi oleh olahraga relatif terhadap istirahat.	Mengukur perubahan segmen ST pada EKG selama tes stres, yang merupakan indikator penting iskemia miokard (kurangnya aliran darah ke jantung) [29][27].
Slope	Int	Kemiringan segmen ST puncak latihan (Slope of the Peak Exercise ST Segment). Nilai kategorikal (misalnya, 1: upsloping, 2: flat, 3: downsloping).	Kemiringan segmen ST adalah penemuan diagnostik kunci dalam tes stres; downsloping (3) adalah yang paling mengkhawatirkan [27].
Ca	Int	Jumlah pembuluh darah utama (number of major vessels) (0-3) yang diwarnai oleh fluoroscopy	Jumlah pembuluh darah yang tersumbat atau menyempit adalah ukuran langsung dari tingkat keparahan penyakit arteri koroner [28][29].
Thal	Int	Hasil tes Thallium Stress Test (misalnya, 3: normal, 6: fixed defect, 7: reversible defect).	Menunjukkan distribusi darah ke otot jantung; fixed defect atau reversible defect mengindikasikan masalah aliran darah [28][29].

B. Pemrosesan Data

Dalam proses eksperimen, data terlebih dahulu melalui tahap prapemrosesan agar siap digunakan pada model machine learning. Tahap awal dilakukan dengan memeriksa missing value untuk memastikan tidak terdapat data kosong yang dapat memengaruhi proses pelatihan model.

Selanjutnya, struktur dataset diperiksa berdasarkan tipe data setiap fitur. Dataset ini terdiri atas fitur numerik, seperti usia, tekanan darah, kadar kolesterol, dan detak jantung maksimum, serta fitur kategorikal, seperti jenis kelamin, jenis nyeri dada, dan kadar gula darah puasa. Setelah fitur dan label target dipisahkan, dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% secara stratified agar proporsi kelas 0 dan kelas 1 tetap terjaga pada kedua subset. Setelah pembagian data dilakukan, fitur dinormalisasi menggunakan MinMaxScaler dengan rentang nilai 0 hingga 1. Penerapan scaling setelah pembagian data bertujuan untuk mengurangi potensi data leakage pada proses evaluasi model.

C. Model Development

Penelitian ini mengembangkan enam skenario pemodelan yang berasal dari kombinasi tiga algoritma machine learning, yaitu Random Forest (RF), XGBoost, dan Support Vector Machine (SVM), dengan dua metode seleksi fitur, yaitu Chi-Square dan Recursive Feature Elimination (RFE). Kombinasi tersebut digunakan untuk membandingkan performa setiap algoritma sekaligus melihat pengaruh metode seleksi fitur terhadap hasil prediksi penyakit jantung.

D. Model Evaluation

Evaluasi model menggunakan beberapa metrik penting untuk mengukur performa. Metrik yang dipakai adalah *accuracy*, *precision*, *recall* dan *F1-score* serta membuat *confusion matrix* untuk melihat detail TP (*True Positive*), TN (*True Negative*), FP (*False Positive*) dan FN (*False Negative*). Akurasi mengukur proporsi prediksi yang diklasifikasikan dengan benar terhadap seluruh data, namun dapat menyesatkan pada dataset yang tidak seimbang karena tidak merepresentasikan kinerja tiap kelas secara proporsional [3] yang dapat dilihat pada persamaan (1)

Precision, atau Positive Predictive Value (PPV), mengukur proporsi data yang diprediksi positif dan benar-benar termasuk kelas positif [30] sebagaimana ditunjukkan pada persamaan (2). Recall, juga dikenal sebagai *Sensitivity* atau *True Positive Rate* (TPR), mengukur seberapa banyak dari kasus positif aktual yang berhasil diidentifikasi oleh model. *Recall* sangat penting dalam prediksi penyakit, karena mengukur kemampuan model untuk mendeteksi semua kasus penyakit yang ada [31] sebagaimana ditunjukkan pada persamaan (3) dan *F1-score* merupakan rata-rata harmonik dari precision dan recall yang digunakan untuk menilai keseimbangan kinerja model [31] yang ditunjukkan pada Persamaan (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

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

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

$$f1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

Kemudian *confusion matrix* adalah tabel ringkasan yang memvisualisasikan kinerja algoritma klasifikasi. Matriks ini membandingkan nilai kelas aktual (sebenarnya) dengan nilai kelas yang diprediksi oleh model [32].

- *True Positive* (TP) adalah Hasil positif yang diprediksi dengan benar yaitu pasien yang sakit (Positif) dan diprediksi sakit (Benar).
- *True Negative* (TN) adalah Hasil negatif yang diprediksi dengan benar yaitu Pasien yang sehat (Negatif) dan diprediksi sehat (Benar).
- *False Positive* (FP) adalah Hasil positif yang diprediksi secara salah (*Type I Error*) yaitu pasien yang sehat (Negatif) tetapi diprediksi sakit (Salah).
- *False Negative* (FN) adalah Hasil negatif yang diprediksi secara salah (*Type II Error*) yaitu pasien yang sakit (Positif) tetapi diprediksi sehat (Salah).

E. Explainable AI (XAI) dengan SHAP

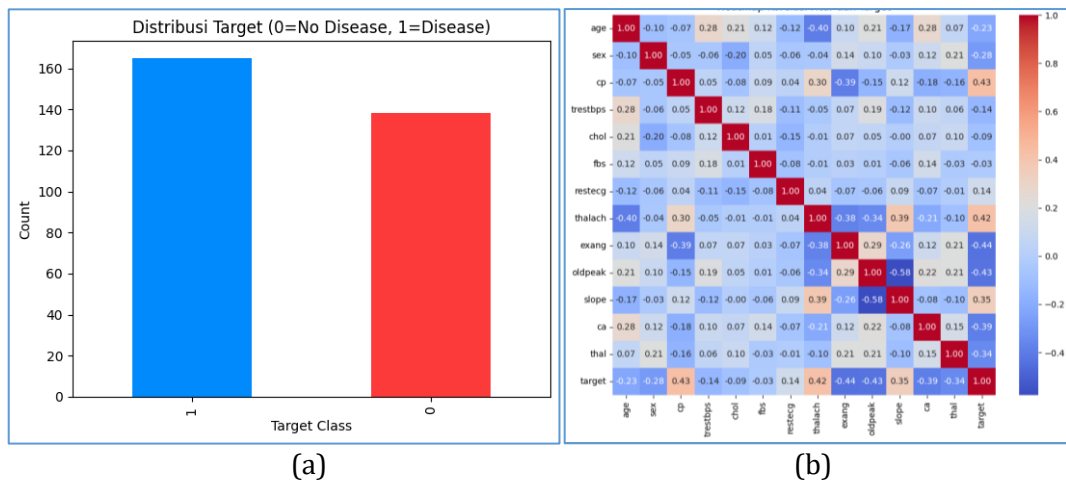
Interpretabilitas model ditingkatkan menggunakan SHAP (*SHapley Additive exPlanations*) yang menjelaskan kontribusi setiap fitur terhadap prediksi. SHAP menghitung seberapa besar pengaruh setiap fitur dengan mempertimbangkan semua kombinasi fitur yang mungkin. *TreeExplainer* digunakan untuk *Random Forest* dan *XGBoost* karena efisien menangani model berbasis tree. *KernelExplainer* digunakan untuk SVM karena bersifat *model agnostic* meskipun komputasinya lebih lambat.

F. Penyimpanan Model untuk Deployment

Setelah *training* dan evaluasi, semua komponen model disimpan menggunakan *joblib* dalam format *pickle* untuk *deployment*. Komponen yang disimpan meliputi *StandardScaler* untuk normalisasi, *feature selectors* (*Chi-Square* dan RFE), serta hasil model ML. Metadata disimpan dalam format *pickle* dan *JSON* yang berisi informasi dataset, daftar fitur, model terbaik, dan metrik evaluasi lengkap setiap model termasuk *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix component*. Aplikasi web dikembangkan menggunakan *Flask* untuk *deployment* model. Sistem memuat model ML, scaler, *feature selectors* dan metadata dari file *pickle*, model menghasilkan kelas prediksi (0/1), probability estimates, dan *confidence level*. Sistem mengintegrasikan SHAP untuk *Explainable AI* dengan *TreeExplainer* untuk *Random Forest* dan *XGBoost*, serta *KernelExplainer* untuk SVM. *Output* SHAP menampilkan 8 fitur paling berpengaruh dengan nilai kontribusi dan arah pengaruhnya.

4 Hasil dan Pembahasan

A. Hasil Analisis Karakteristik Dataset



Gambar 2 (a) karakteristik dataset yang diolah, (b) Heatmap korelasi antar fitur

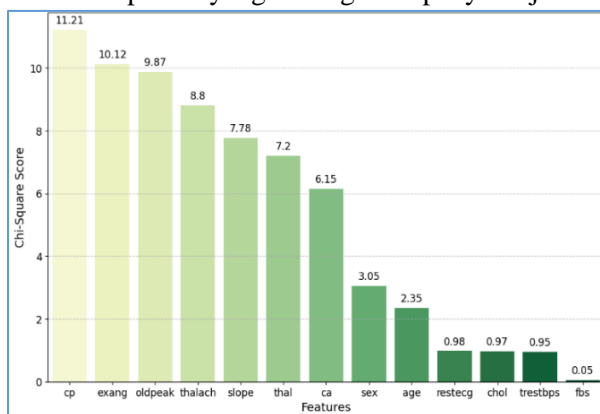
Gambar 2 menampilkan karakteristik awal dataset yang digunakan dalam penelitian ini. Visualisasi sebelah kiri menunjukkan distribusi kelas target, yaitu kelas 0 untuk pasien tanpa indikasi penyakit jantung dan kelas 1 untuk pasien dengan indikasi penyakit jantung. Visualisasi sebelah kanan menunjukkan heatmap korelasi antarfitur dan target untuk melihat hubungan statistik antara variabel klinis dengan label penyakit jantung. Analisis ini digunakan sebagai gambaran awal sebelum proses seleksi fitur dan pemodelan dilakukan.

Heatmap pada Gambar 2 menunjukkan bahwa beberapa fitur memiliki korelasi yang cukup terlihat dengan target penyakit jantung. Fitur jenis nyeri dada (cp) memiliki korelasi positif sebesar 0,43 dengan target, yang menunjukkan bahwa variasi tipe nyeri dada berkaitan dengan kemungkinan adanya indikasi penyakit jantung. Detak jantung maksimum (thalach) juga menunjukkan korelasi positif sebesar 0,42, sedangkan oldpeak memiliki korelasi negatif sebesar -0,43. Selain itu, tekanan darah istirahat (trestbps) menunjukkan korelasi positif sebesar 0,66 dengan kolesterol (chol), dan slope memiliki korelasi positif sebesar 0,39 dengan oldpeak.

Hasil korelasi ini memberikan gambaran awal mengenai hubungan antarfitur klinis, tetapi tidak digunakan sebagai satu-satunya dasar penentuan fitur penting. Penentuan fitur utama dalam penelitian ini tetap mengacu pada hasil Chi-Square dan RFE, sedangkan SHAP digunakan sebagai analisis tambahan pada sistem berbasis web untuk menjelaskan faktor-faktor yang memengaruhi hasil prediksi model, baik ketika sistem menunjukkan adanya indikasi penyakit jantung maupun tidak.

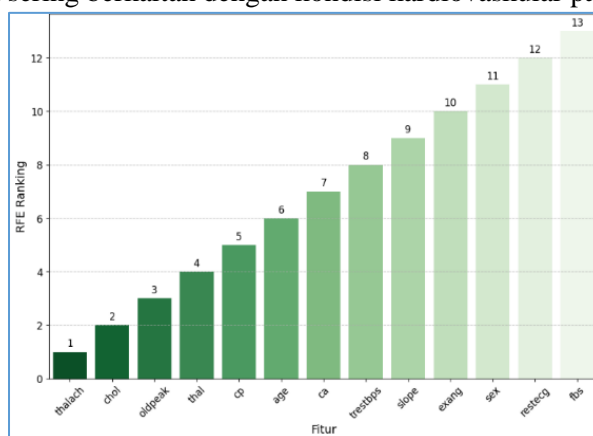
B. Hasil Feature Selection

Hasil pengujian menggunakan metode *Chi-Square* menunjukkan bahwa setiap fitur dalam dataset memiliki tingkat keterkaitan yang berbeda terhadap variabel target. Di antara seluruh variabel yang dianalisis, cp (jenis nyeri dada) muncul sebagai fitur dengan pengaruh paling dominan. Skor *Chi-Square* yang mencapai 11.21 menunjukkan bahwa variasi jenis nyeri dada memainkan peran yang sangat penting dalam membedakan pasien yang terdiagnosis penyakit jantung dengan yang tidak.



Gambar 3 Hasil seleksi fitur menggunakan metode Chi-Square. Fitur cp, exang, dan oldpeak

Gambar 3 Hasil seleksi fitur menggunakan metode Chi-Square. Fitur cp, exang, dan oldpeak memperoleh skor tertinggi dibandingkan fitur lainnya. Temuan ini menunjukkan bahwa fitur cp, exang, oldpeak, dan thalach memiliki keterkaitan yang cukup kuat dengan target klasifikasi, sehingga dapat menjadi indikator penting dalam proses pemodelan penyakit jantung. Hasil ini juga sejalan dengan fakta klinis bahwa tipe nyeri dada, angina akibat aktivitas fisik, perubahan segmen ST, dan detak jantung maksimum sering berkaitan dengan kondisi kardiovaskular pasien [33][34].

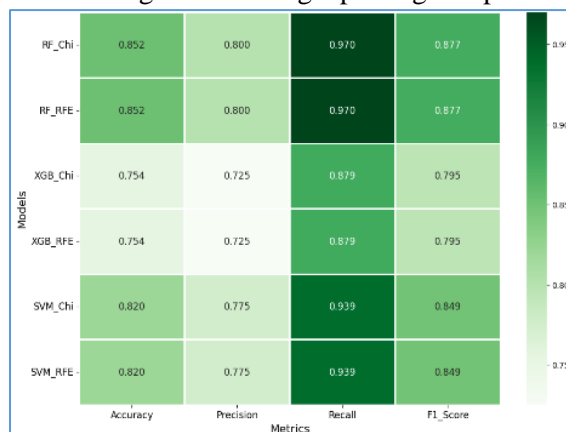


Gambar 4 Grafik hasil seleksi fitur menggunakan metode RFE

Grafik 4 memperlihatkan hasil *Recursive Feature Elimination* (RFE) yang mengurutkan fitur berdasarkan kontribusinya terhadap model prediksi penyakit jantung. Fitur thalach, chol, dan oldpeak memberikan kontribusi paling signifikan terhadap akurasi model dan dipertahankan dalam pemodelan. Sementara itu, fitur fbs, restecg, sex, dan exang memberikan kontribusi lebih kecil terhadap kinerja model, sehingga dapat dipertimbangkan untuk dieliminasi atau dikurangi perannya agar model lebih efisien dan sederhana.

C. Hasil Evaluasi Model

Pada bagian ini, disajikan hasil evaluasi kinerja enam skenario pemodelan yang diperoleh dari kombinasi tiga algoritma klasifikasi, yaitu Random Forest (RF), XGBoost (XGB), dan Support Vector Machine (SVM), dengan dua metode seleksi fitur, yaitu Chi-Square dan RFE. Keberhasilan masing-masing model dievaluasi berdasarkan empat metrik utama: Akurasi, Presisi, Recall, dan *F1-score*, yang secara keseluruhan memberikan gambaran lengkap mengenai performa prediksi model.



Gambar 5 Heatmap perbandingan kinerja enam skenario pemodelan berdasarkan metrik accuracy, precision, recall, dan F1-score

Tabel 2 Ringkasan kinerja enam skenario pemodelan

Skenario	Accuracy	Precision	Recall	F1-score
RF_Chi	0,852	0,800	0,970	0,877
RF_RFE	0,852	0,800	0,970	0,877
XGB_Chi	0,754	0,725	0,879	0,795
XGB_RFE	0,754	0,725	0,879	0,795
SVM_Chi	0,820	0,775	0,939	0,849
SVM_RFE	0,820	0,775	0,939	0,849

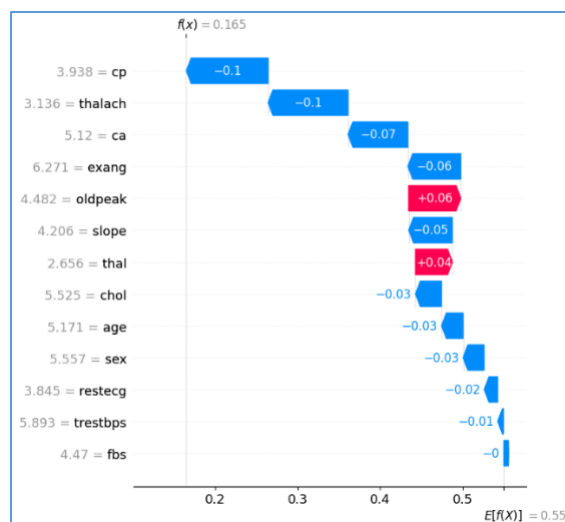
Tabel 2 menunjukkan bahwa RF_Chi dan RF_RFE memperoleh nilai accuracy yang sama, yaitu 0,852, dengan precision 0,800, recall 0,970, dan F1-score 0,877. Kedua skenario Random Forest memberikan performa terbaik dibandingkan XGBoost dan SVM. XGB_Chi dan XGB_RFE memperoleh accuracy 0,754, precision 0,725, recall 0,879, dan F1-score 0,795. Sementara itu, SVM_Chi dan SVM_RFE memperoleh accuracy 0,820, precision 0,775, recall 0,939, dan F1-score 0,849.

Hasil tersebut menunjukkan bahwa Random Forest memiliki performa yang lebih stabil dalam menjaga keseimbangan antara precision dan recall dibandingkan XGBoost dan SVM pada dataset yang digunakan. Nilai recall yang tinggi menunjukkan kemampuan model dalam mengidentifikasi pasien dengan indikasi penyakit jantung secara lebih baik, sehingga potensi false negative dapat ditekan.

D. Hasil Explainable AI (SHAP Analysis)

Pada bagian ini, disajikan interpretasi hasil prediksi yang dihasilkan oleh model RF, SVM, dan XGBoost menggunakan SHAP. Analisis ini bertujuan untuk memberikan pemahaman yang lebih jelas mengenai kontribusi setiap fitur terhadap hasil prediksi model. Pendekatan ini membantu pengguna memahami alasan di balik hasil prediksi yang diberikan sistem. Dengan demikian, sistem tidak hanya menampilkan hasil klasifikasi, tetapi juga memberikan informasi tambahan mengenai fitur-fitur yang paling memengaruhi keputusan model pada setiap data pasien.

Untuk ilustrasi tiap hasil direpresentasikan dengan waterfall plot yaitu cara untuk menjelaskan bagaimana sebuah model membuat keputusan berdasarkan data yang dimilikinya. Di dalam plot ini, kita bisa melihat langkah demi langkah bagaimana setiap informasi yang diberikan, seperti usia, jenis kelamin, atau tekanan darah, mempengaruhi keputusan akhir model.



Gambar 6 Waterfall plot nilai SHAP

Gambar 6 menunjukkan bahwa hasil prediksi model pada Sample 0 mengalami penurunan dari nilai dasar $E[f(x)] = 0,55$ menjadi nilai akhir $f(x) = 0,165$. Penurunan ini menunjukkan bahwa kontribusi fitur pada sampel tersebut secara keseluruhan lebih dominan menurunkan nilai prediksi terhadap kelas Disease. Dengan kata lain, karakteristik pasien pada Sample 0 lebih kuat mengarahkan model pada prediksi Tidak ada indikasi penyakit jantung dibandingkan terdapat indikasi penyakit.

Tabel 3 merupakan fitur yang memberikan kontribusi negatif dan positif, dan fitur yang memiliki kontribusi negatif terbesar terhadap prediksi adalah cp dan thalach, masing-masing dengan nilai SHAP sekitar $-0,1$. Kedua fitur tersebut menjadi faktor utama yang menyebabkan nilai prediksi bergerak turun dari nilai dasar. Selain itu, fitur ca, exang, dan slope juga memberikan kontribusi negatif, sehingga semakin memperkuat kecenderungan model untuk tidak mengklasifikasikan sampel ini ke dalam kelas Disease.

Di sisi lain, fitur oldpeak dan thal memberikan kontribusi positif terhadap kelas Disease, tetapi pengaruhnya relatif lebih kecil dibandingkan akumulasi kontribusi negatif dari fitur lainnya. Oleh karena itu, hasil akhir prediksi menunjukkan bahwa Sample 0 diklasifikasikan sebagai Tidak ada indikasi penyakit jantung. Interpretasi ini menegaskan bahwa keputusan model tidak dipengaruhi oleh satu fitur saja, melainkan oleh gabungan kontribusi seluruh fitur yang membentuk arah prediksi akhir.

Tabel 3 Contoh kontribusi SHAP lokal pada prediksi pasien

Contoh	Fitur	Nilai SHAP	Arah kontribusi
Sample 0	cp	-0,10	Menurunkan prediksi Disease
Sample 0	thalach	-0,10	Menurunkan prediksi Disease
Sample 0	ca	-0,07	Menurunkan prediksi Disease
Sample 0	exang	-0,06	Menurunkan prediksi Disease
Sample 0	oldpeak	+0,06	Mendorong prediksi Disease
Contoh web	oldpeak	+0,482	Mendorong prediksi Disease
Contoh web	ST slope	+0,367	Mendorong prediksi Disease
Contoh web	max heart rate	+0,215	Mendorong prediksi Disease
Contoh web	thalach	-0,236	Menurunkan prediksi Disease
Contoh web	chest pain type	-0,193	Menurunkan prediksi Disease

E. Interpretasi Hasil Feature Selection dan SHAP Analysis

Feature Selection dan SHAP dapat menghasilkan penekanan yang berbeda karena keduanya memiliki tujuan analisis yang tidak sama. Feature Selection digunakan untuk mengidentifikasi fitur-fitur yang paling relevan bagi model secara keseluruhan. Berdasarkan hasil Chi-Square dan RFE, fitur

Di sini, berbagai model machine learning digunakan untuk memprediksi penyakit jantung, dengan XGBoost + RFE yang ditandai sebagai pilihan yang disarankan.

Penjelasan ini membantu pengguna memahami faktor-faktor yang berkontribusi terhadap hasil prediksi, sehingga sistem tidak hanya memberikan keluaran klasifikasi, tetapi juga informasi pendukung yang lebih transparan.

5 Kesimpulan

Penelitian ini membandingkan kinerja tiga algoritma machine learning, yaitu Random Forest (RF), XGBoost (XGB), dan Support Vector Machine (SVM), dalam mendeteksi penyakit jantung menggunakan dataset Heart Disease UCI. Ketiga algoritma tersebut dikombinasikan dengan dua metode seleksi fitur, yaitu Chi-Square dan Recursive Feature Elimination (RFE), sehingga menghasilkan enam skenario pemodelan. Hasil seleksi fitur menunjukkan bahwa beberapa fitur klinis, seperti jenis nyeri dada, detak jantung maksimum, jumlah pembuluh darah utama, oldpeak, dan respons fisiologis terhadap aktivitas fisik, memiliki kontribusi penting dalam proses prediksi. Berdasarkan hasil evaluasi, Random Forest memperoleh performa terbaik dengan accuracy 0,852 dan recall 0,970, sehingga menunjukkan kemampuan yang baik dalam mendeteksi pasien dengan indikasi penyakit jantung. Penerapan SHAP membantu menjelaskan kontribusi setiap fitur terhadap hasil prediksi, sehingga model tidak hanya menghasilkan keluaran klasifikasi, tetapi juga memberikan interpretasi yang lebih transparan. Temuan SHAP pada contoh prediksi menunjukkan bahwa oldpeak, ST slope, dan max heart rate dapat mendorong prediksi ke arah Disease, sedangkan thalach dan chest pain type dapat menurunkan kecenderungan prediksi pada kasus tertentu. Secara umum, sistem yang dikembangkan berpotensi digunakan sebagai alat bantu deteksi dini berbasis web. Namun, hasil prediksi tetap harus diposisikan sebagai informasi pendukung awal dan tidak dapat menggantikan diagnosis dokter.

Referensi

- [1] A. Mondal, B. Mondal, A. Chakraborty, A. Kar, A. Biswas, and A. Banerjee Majumder, "Heart Disease Prediction using Support Vector Machine and Artificial Neural Network," *Artif. Intell. Appl.*, Vol. 2, 2023, DOI: 10.47852/bonviewAIA3202823.
- [2] S. Wan, F. Wan, and X. Dai, "Machine Learning Approaches for Cardiovascular Disease Prediction: A Review," *Arch. Cardiovasc. Dis.*, Vol. 118, Oct. 2025, DOI: 10.1016/j.acvd.2025.04.055.
- [3] L. B. Elvas, A. Almeida, and J. C. Ferreira, "The Role of AI in Cardiovascular Event Monitoring and Early Detection: Scoping Literature Review," *JMIR Med. Informatics*, Vol. 13, 2025, doi: 10.2196/64349.
- [4] A. Kannan, M. J. Saikia, S. Kumar, and S. Datta, "Detection of Valvular Heart Diseases from PCG Signals using Machine and Deep Learning Models: A Review," 2025. DOI: 10.1109/ACCESS.2025.3583263.
- [5] V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain, "An Artificial Intelligence Model for Heart Disease Detection using Machine Learning Algorithms," *Healthc. Anal.*, Vol. 2, Nov. 2022, DOI: 10.1016/j.health.2022.100016.
- [6] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model using Machine Learning and Hyperparameter Optimization Methods," 2022. DOI: 10.1109/ACCESS.2022.3191669.
- [7] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, "Enhanced Heart Disease Prediction based on Machine Learning and χ^2 Statistical Optimal Feature Selection Model," *Designs*, Vol. 6, No. 5, Sep. 2022, DOI: 10.3390/designs6050087.
- [8] A. E. Korial, I. I. Gorial, and A. J. Humaidi, "An Improved Ensemble-based Cardiovascular Disease Detection System with Chi-Square Feature Selection," *Computers*, Vol. 13, No. 6, 2024, DOI: 10.3390/computers13060126.

- [9] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the Impact of Feature Selection on the Accuracy of Heart Disease Prediction," *Healthc. Anal.*, Vol. 2, Nov. 2022, DOI: 10.1016/j.health.2022.100060.
- [10] P. Theerthagiri, "Predictive Analysis of Cardiovascular Disease using Gradient Boosting based Learning and Recursive Feature Elimination Technique," *Intell. Syst. with Appl.*, Vol. 16, Nov. 2022, DOI: 10.1016/j.iswa.2022.200121.
- [11] R. Jahangir, M. N. Islam, M. S. Islam, and M. M. Islam, "ECG-based Heart Arrhythmia Classification using Feature Engineering and a Hybrid Stacked Machine Learning," *BMC Cardiovasc. Disord.*, Vol. 25, Apr. 2025, DOI: 10.1186/s12872-025-04678-9.
- [12] W. Jian, J. P. Li, A. U. Haq, S. Khan, R. M. Alotaibi, S. A. Alajlan, and M. B. Bin Heyat, "Feature Elimination and Stacking Framework for Accurate Heart Disease Detection in IoT Healthcare Systems using Clinical Data," *Front. Med.*, May 2024, DOI: 10.3389/fmed.2024.1362397.
- [13] V. Sapra and L. Sapra, "An Interpretable Approach with Explainable AI for the Detection of Cardiovascular Disease," Nov. 2024. DOI: 10.1109/ICIICS63763.2024.10860265.
- [14] H. Luo, C. Xiang, L. Zeng, S. Li, X. Mei, L. Xiong, Y. Liu, C. Wen, Y. Cui, L. Du, Y. Zhou, K. Wang, L. Li, Z. Liu, Q. Wu, J. Pu, and R. Yue, "SHAP based Predictive Modeling for 1 Year All-Cause Readmission Risk in Elderly Heart Failure Patients: Feature Selection and Model Interpretation," *SCI. Rep.*, Vol. 14, 2024, DOI: 10.1038/s41598-024-67844-7.
- [15] X. Qi, S. Wang, C. Fang, J. Jia, L. Lin, and T. Yuan, "Machine Learning and SHAP Value Interpretation for Predicting Comorbidity of Cardiovascular Disease and Cancer with Dietary Antioxidants," *Redox Biol.*, Vol. 79, Feb. 2025, DOI: 10.1016/j.redox.2024.103470.
- [16] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Ensemble Learning with Explainable AI for Improved Heart Disease Prediction based on Multiple Datasets," *SCI. Rep.*, Vol. 15, Apr. 2025, DOI: 10.1038/s41598-025-97547-6.
- [17] P. Nancy, P. R. Mutkule, K. S. Thakre, A. S. Ladkat, S. B. G. T. Babu, S. L. Bangare, and M. Naved, "Machine Learning and Feature Selection-Enabled Optimized Technique for Heart Disease Classification and Prediction," Vol. 31, No. 4, pp. 419–429, 2024, DOI: 10.24423/comes.2024.602.
- [18] I. S. Al-Mahdi, S. M. Darwish, and M. M. Madbouly, "Heart Disease Prediction Model using Feature Selection and Ensemble Deep Learning with Optimized Weight," *Comput. Model. Eng. SCI.*, 2025, DOI: 10.32604/cmcs.2025.061623.
- [19] M. D. Teja and G. M. Rayalu, "Optimizing Heart Disease Diagnosis with Advanced Machine Learning Models: A Comparison of Predictive Performance," Vol. 6, 2025, DOI: 10.1186/s12872-025-04627-6.
- [20] S. Patidar, D. Kumar, and D. Rukwal, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction, 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 64–69, 2022, DOI: 10.3233/ATDE220723.
- [21] Krittanawong, C, Virk, H.U.H, Bangalore, Z. Wang, T. Kitai, U. Baber, J. L. Halperin, and W. H. W. Tang, "Machine Learning Prediction in Cardiovascular Diseases: A Meta - Analysis," *SCI. Rep.*, pp. 1–11, 2020, DOI: 10.1038/s41598-020-72685-1.
- [22] K. Kumar, R. Govindarajan, S. Sathya, and J. S. Murugan, "Intelligence-based Medicine Comparative Analysis of Explainable Machine Learning Models for Cardiovascular Risk Stratification using Clinical Data and SHapley Additive exPlanations," *Intell. Med.*, Vol. 12, No. June, p. 100286, 2025, DOI: 10.1016/j.ibmed.2025.100286.
- [23] R. Yilmaz, "Early Detection of Coronary Heart Disease based on Machine Learning Methods," Vol. 4, 2022, DOI: 10.37990/medr.1011924.
- [24] B. Dwarakanath, M. Latha, R. Annamalai, J. S. Kallimani, R. Walia, and B. Belete, "A Novel Feature Selection with Hybrid Deep Learning based Heart Disease Detection and Classification in the e-Healthcare Environment," 2022, DOI: 10.1155/2022/1167494.
- [25] H. Yang, "Predicting Coronary Heart Disease using an Improved LightGBM Model: Performance Analysis and Comparison," *IEEE Access*, 2023, DOI: 10.1109/ACCESS.2023.3253885.
- [26] P. Ghadekar, "Predicting Heart Disease Risk in Diabetic Patients using a Pipeline of Ensemble Learning and XAI-Enhanced Approaches." DOI: 10.1109/IC-

- CGU58078.2024.10530665.
- [27] G. P. Oise, S. A. Oyedotun, O. C. Nwabuokeyi, A. E. Babalola, and N. B. Unuigbokhai, "Enhanced Prediction of Coronary Artery Disease using Logistic Regression," *FUDMA J. SCI.*, Mar. 2025, DOI: 10.33003/fjs-2025-0903-3263.
 - [28] H. Alsafi and J. Munilla, "Ensemble Approach for Heart Disease Diagnosis Integrating HHO Algorithm and Machine Learning Techniques," 2025. DOI: 10.1063/5.0258845.
 - [29] K. Mohammad Alfadli and A. Omran Almagrabi, "Feature-Limited Prediction on the UCI Heart Disease Dataset," *Comput. Mater. Contin.*, 2023, DOI: 10.32604/cmc.2023.033603.
 - [30] S. Bajaj, M. Bala, and M. Angurula, "Machine Learning Models for Enhanced Stroke Detection and Prediction," *Egypt. Informatics J.*, Jun. 2025, DOI: 10.1016/j.eij.2025.100705.
 - [31] M. Hameed, M. A. Z. Raja, A. Zameer, H. S. Dar, A. S. Alluhaidan, and R. Aziz, "Acute Myeloid Leukemia Classification using ReLViT and Detection with YOLO Enhanced by Adversarial Networks on Bone Marrow Images," *SCI. Rep.*, Sep. 2025, DOI: 10.1038/s41598-025-17891-5.
 - [32] E. M. Mohammed, A. M. Fakhrudeen, and O. Y. Alani, "Detection of Alzheimer's Disease using Deep Learning Models: A Systematic Literature Review," *Informatics Med. Unlocked*, 2024, DOI: 10.1016/j.imu.2024.101551.
 - [33] 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, Oct. 2024, DOI: 10.1038/s41598-024-74656-2.
 - [34] M. H. Alshayegi and S. Abed, "Heart Disease Prediction by Tabular Modeling with Deep Learning Network and Interpretability," *Mach. Learn. SCI. Technol.*, Vol. 6, Sep. 2025, DOI: 10.1088/2632-2153/adfd39.