Prediksi Penyakit Jantung Menggunakan Model Stacking XGBoost - Random Forest

1st Muhammad Faisal Hibatullah
Program Studi Teknologi Informasi,
Universitas Telkom, Kampus Surabaya,
Surabaya, Jawa Timur, Indonesia
fsalhibtlh@student.telkomuniversity.ac.id

2nd Bernadus Anggo Seno Aji, S.Kom.,M.Kom.
Program Studi Teknologi Informasi,
Universitas Telkom, Kampus Surabaya,
Surabaya, Jawa Timur, Indonesia
bernadusanggosenoaji@telkomuniversity.ac.id

3rd Yohanes Setiawan, S.Si., M.Kom.
Program Studi Teknologi Informasi,
Universitas Telkom, Kampus Surabaya,
Surabaya, Jawa Timur, Indonesia
yohanessetiawan@telkomuniversity.ac.id

Penyakit jantung merupakan isu kesehatan global yang krusial, membutuhkan deteksi dan prediksi akurat untuk penanganan optimal. Tetapi, metode yang ada seringkali tidak mampu menangani kerumitan data medis. Mengatasi masalah utama tersebut, dikembangkan model machine learning yang kuat dan mudah dipahami. Model ini menggunakan dataset Heart Disease UCI yang telah melalui proses persiapan data, termasuk penanganan data yang hilang dengan KNN Imputer.Penelitian ini memperkenalkan model hybrid XGBoost-Random Forest (XGB-RF) yang beroperasi dalam dua mode: sebagai stacking, di mana prediksi XGBoost menjadi input tambahan bagi Random Forest, dan XGB-RF Feature Engineering, tempat XGBoost menghasilkan ciri-ciri baru yang kemudian menjadi data masukan bagi Random Forest. Pemahaman hasil prediksi dilakukan oleh SHapley Additive exPlanations (SHAP), sehingga dapat mengidentifikasi penyebab penyakit jantung pada tiap individu. Dengan model mencapai akurasi 83.69% dan ROC-AUC 86.27%, menunjukkan peningkatan besar dalam kemampuan memprediksi risiko penyakit jantung. Kesimpulannya, model XGB-RF ini efektif dalam memprediksi risiko penyakit jantung dan memberikan penjelasan yang jelas melalui SHAP. Model ini berpotensi besar membantu dokter dalam penilaian risiko pasien yang lebih cepat dan akurat di masa depan, bahkan dapat dikembangkan untuk pemantauan dini melalui integrasi teknologi.

Kata Kunci:. Penyakit Jantung, XGBoost, Random Forest, Stacking, Rekayasa Fitur

I. PENDAHULUAN

Penyakit kardiovaskular (PKV) merupakan penyebab utama kematian di seluruh dunia, sehingga deteksi dini dan identifikasi faktor risiko sangatlah penting[1]. Machine learning (ML) menawarkan solusi potensial untuk mengembangkan model prediksi risiko penyakit jantung yang lebih akurat dan personal[2]. Dataset Penyakit Jantung UCI, khususnya dari Cleveland Clinic, sering digunakan sebagai benchmark dalam penelitian Machine Learning (ML) untuk prediksi penyakit jantung. Dataset ini mencakup berbagai fitur klinis dan demografis esensial seperti usia, jenis kelamin, dan tekanan darah. Akan tetapi, data medis sering kali memiliki nilai yang hilang, yang bisa mengurangi kualitas dan akurasi model. Oleh karena itu, penanganan nilai yang hilang yang tepat, seperti menggunakan metode KNN Imputer, merupakan langkah krusial dalam pra-pemrosesan data untuk memastikan integritas dan kualitas data pelatihan[3].

Untuk membangun model prediksi yang andal, penelitian ini mengusulkan penggunaan kombinasi algoritma ensemble learning yang kuat, yaitu XGBoost dan Random Forest[4]. Kedua algoritma ini dikenal memiliki performa tinggi dalam berbagai tugas klasifikasi, termasuk dalam domain medis. XGBoost dikenal unggul dalam mengoptimalkan bias[5], sedangkan Random Forest efektif dalam mengurangi overfitting dan meningkatkan stabilitas model[6]. Penyatuan dua algoritma ini melalui metode stacking (XGB-RF) diprediksi akan menghasilkan model yang lebih tangguh dan presisi. Meskipun demikian, model ensemble yang kompleks seringkali bersifat seperti "kotak hitam," yang menyulitkan interpretasi. Untuk mengatasi hal ini, penelitian ini mengintegrasikan metode SHapley Additive exPlanations (SHAP) untuk menjelaskan kontribusi setiap fitur terhadap prediksi individu, meningkatkan transparansi dan kepercayaan terhadap hasil model[7].

II. KAJIAN TEORI

Menyajikan dan menjelaskan teori dan konsep yang relevan dengan variabel penelitian, termasuk tinjauan studi sebelumnya, deskripsi dataset, serta metode dan algoritma yang digunakan.

A. Penelitian Terdahulu

Penelitian yang telah dilakukan sebelumnya menjadi landasan penting dalam pengembangan model prediksi penyakit jantung. el Hamdaoui, H., dkk. (2021) menunjukkan efektivitas Random Forest dan kombinasinya dengan AdaBoost, yang berhasil meningkatkan akurasi hingga 96,16%, menegaskan potensi penggunaan algoritma ensemble dalam memprediksi penyakit jantung. Sementara itu, Yoon, T., & Kang, D. (2023) mengembangkan model stacking ensemble multimodal yang menggabungkan beberapa algoritma seperti Random Forest dan XGBoost sebagai meta learner, mencapai akurasi 93,97%, membuktikan superioritas model stacking dibandingkan metode tunggal. Gupta, P., & Seth, D. (2023) juga

melakukan analisis komparatif yang menemukan bahwa Random Forest memberikan akurasi terbaik sebesar 97,13% pada dataset Framingham, serta menekankan pentingnya evaluasi fitur.

Penelitian saat ini secara spesifik mengembangkan arsitektur ensemble kompleks untuk meningkatkan akurasi dan robustas model. Selain itu, penelitian ini mengimplementasikan SHAP untuk menginterpretasi prediksi individu, sebuah aspek yang tidak banyak dieksplorasi dalam penelitian terdahulu, sehingga memberikan penjelasan yang relevan secara klinis. Pendekatan yang lebih komprehensif juga diterapkan dalam pra-pemrosesan data dengan menggunakan KNN Imputer untuk menangani nilai yang hilang.

B. Dataset

Dataset yang digunakan dalam penelitian ini merupakan kompilasi dari lima sumber berbeda (Cleveland, Hungarian, Switzerland, Long Beach VA, dan Stalog), menghasilkan total 918 observasi dengan 12 atribut. Fitur-fitur yang tercakup sangat esensial untuk prediksi penyakit jantung, meliputi usia, jenis kelamin, jenis nyeri dada, tekanan darah dan kolesterol, gula darah puasa, hasil EKG, detak jantung maksimum, angina akibat olahraga, dan oldpeak. Dataset ini unggul karena ukurannya besar dan beragam, sehingga model dapat dilatih dengan data yang lebih bervariasi dan menjadi standar perbandingan dalam banyak penelitian.

C. KNN Imputer

KNN Imputer adalah teknik yang digunakan untuk menangani nilai yang hilang (missing values) dalam sebuah dataset. Metode ini bekerja dengan cara memperkirakan nilai yang hilang tersebut berdasarkan data-data lain yang paling mirip. Prosesnya dimulai dengan mengidentifikasi sebuah titik data yang memiliki nilai hilang. Selanjutnya, algoritma akan mencari k tetangga terdekat, yaitu titik data lain yang paling mirip berdasarkan fitur-fitur yang ada. Setelah ditemukan, nilai yang hilang pada titik data tersebut akan diisi dengan rata-rata atau median dari nilai-nilai tetangga yang sesuai. Pendekatan ini lebih akurat dibandingkan metode sederhana seperti mengisi dengan rata-rata keseluruhan, karena ia memanfaatkan hubungan dan pola yang ada di dalam data.

D. XGBoost (eXtreme Gradient Boosting)

XGBoost adalah algoritma pembelajaran mesin tingkat lanjut yang dikenal karena efisiensi, kecepatan, dan kinerjanya yang superior. Sebagai implementasi yang dioptimalkan dari Gradient Boosting, XGBoost termasuk dalam metode ensemble learning yang membangun model secara berurutan. Setiap pohon keputusan yang baru dilatih berfungsi untuk memperbaiki kesalahan dari gabungan pohon sebelumnya, sehingga sangat efektif dalam meningkatkan akurasi model.

E. Random Forest

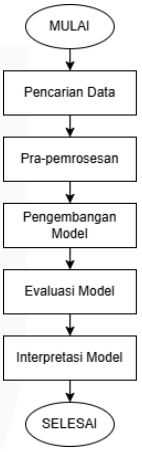
Random Forest adalah algoritma ensemble learning yang efektif dan fleksibel, terutama untuk tugas klasifikasi dan regresi. Selama pelatihan, algoritma ini membuat banyak pohon keputusan secara terpisah. Untuk klasifikasi, hasil prediksi akhirnya ditentukan berdasarkan suara terbanyak dari setiap pohon. Keteracakan tambahan diperkenalkan dengan mencari fitur terbaik dari subset fitur acak saat membagi node, bukan dari semua fitur yang tersedia, yang secara efektif mengurangi overfitting dan meningkatkan akurasi serta stabilitas model.

F. SHAP (SHapley Additive exPlanations)

SHAP adalah metode interpretasi model modern dan terpadu yang didasarkan pada nilai Shapley dari teori permainan kooperatif. SHAP berfungsi untuk menjelaskan hasil prediksi dari model "kotak hitam" dengan mengukur kontribusi unik setiap fitur terhadap prediksi individu. Dalam konteks prediksi penyakit jantung, SHAP dapat menunjukkan seberapa besar pengaruh setiap fitur, seperti usia atau kolesterol, terhadap prediksi risiko pada seorang pasien. Ini memberikan wawasan klinis yang berharga dan meningkatkan kepercayaan pengguna terhadap model, terutama untuk model ensemble yang kompleks.

III. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan rancangan eksperimental untuk membangun dan mengevaluasi model prediksi risiko penyakit jantung. Penelitian ini melibatkan serangkaian tahapan utama, dimulai dari pengumpulan data hingga interpretasi model, yang diilustrasikan dalam Flowchart Sistematika Penyelesaian.



GAMBAR 1 (Flowchart Sistematika Penyelesaian)

A. Pencarian dan Perolehan Data

Dataset yang digunakan adalah Dataset Heart Disease UCI yang akan diperoleh dari repositori UCI Machine Learning. Dataset ini dipilih karena relevansi dan penggunaannya yang luas sebagai tolok ukur dalam penelitian prediksi penyakit jantung.

B. Pra-pemrosesan Data (Preprocessing)

Tahap ini bertujuan untuk menyiapkan data agar siap digunakan untuk melatih model. Langkah-langkahnya adalah sebagai berikut:

- Pembersihan Data: Tahap ini meliputi pemeriksaan dan penanganan inkonsistensi, duplikasi, serta format data yang tidak sesuai untuk menjamin kualitas data yang optimal.
- Penanganan Nilai Hilang: Metode KNN Imputer digunakan untuk mengisi nilai-nilai yang hilang. Pemilihan metode ini didasarkan pada

kemampuannya mengestimasi nilai dengan memanfaatkan data-data terdekat, sehingga struktur lokal data tetap terjaga dan estimasi menjadi lebih akurat.

- Pembagian Data: Dataset yang bersih akan dibagi menjadi data pelatihan (training data) dan data pengujian (testing data) untuk melatih model dan mengevaluasi kinerjanya secara independen.

C. Pengembangan Model Stacking

Penelitian ini akan mengembangkan model stacking XGB-RF untuk meningkatkan akurasi prediksi. Prosesnya adalah sebagai berikut:

- Pelatihan Model XGBoost: Model XGBoost akan dilatih pada data pelatihan.
- Ekstraksi Leaf Feature: Leaf feature dari model XGBoost yang telah dilatih akan diekstrak. Fitur ini merepresentasikan pola dan interaksi non-linear yang kompleks yang ditemukan oleh XGBoost.
- Penggabungan Fitur: Leaf feature dari XGBoost akan digabungkan dengan fitur asli dari dataset, menciptakan dataset yang lebih kaya.
- Pelatihan Random Forest: Model Random Forest akan dilatih menggunakan dataset yang telah diperkaya tersebut.

D. Evaluasi Model

Demi mencapai kinerja model yang optimal, optimasi hyperparameter akan dijalankan pada tiap model menggunakan validasi silang (cross-validation). Kombinasi parameter terbaik yang memberikan skor evaluasi tertinggi akan dipilih sebagai model final. Model akhir ini kemudian dievaluasi pada data pengujian untuk mengukur kemampuan generalisasinya.

E. Interpretasi Model

Setelah model stacking selesai dibangun dan dievaluasi, metode SHapley Additive exPlanations (SHAP) akan diterapkan untuk menginterpretasi prediksi. SHAP akan digunakan untuk mengukur dan memvisualisasikan kontribusi setiap fitur terhadap prediksi risiko penyakit jantung pada setiap individu. Hasilnya akan dianalisis untuk mengidentifikasi faktor-faktor risiko dominan, memberikan wawasan klinis yang berharga, dan mendukung pengambilan keputusan yang lebih terinformasi.

IV. HASIL DAN PEMBAHASAN

Bab ini membahas secara rinci analisis data, proses tuning hyperparameter untuk XGBoost dan Random Forest, serta interpretasi mendalam dari hasil prediksi model menggunakan SHAP.

A. Analisis Data

Sumber Data: Menggunakan Dataset Penyakit Jantung UCI yang merupakan gabungan dari lima database (Cleveland, Hungaria, Swiss, Long Beach V).

Atribut Data: Penelitian ini berfokus pada 14 atribut yang paling umum digunakan untuk prediksi penyakit jantung, seperti usia, jenis kelamin, tipe nyeri dada (cp), tekanan darah istirahat (trestbps), kolesterol serum (chol), gula darah puasa (fbs), hasil EKG (restecg), detak jantung maksimum (thalach), angina akibat olahraga (exang), oldpeak, slope, ca (jumlah pembuluh darah utama), thal (talasemia), dan target (keberadaan penyakit jantung).

Identifikasi dan Penanganan Nilai Hilang: Nilai 0.0 pada trestbps, chol, dan thalach diinterpretasikan sebagai nilai hilang dan ditangani menggunakan KNN Imputer karena secara fisiologis tidak mungkin nol. Fitur ca juga memiliki nilai 0.0 yang perlu dicatat potensi ambiguitasnya.

B. Tuning XGBoost

Hyperparameter yang disetel: learning_rate, max_depth, n_estimators, gamma, dan subsample.

TABEL 1 (EVALUASI HYPERPARAMETER UNTUK XGBOOST)

index	gamma	learning_rate	max_depth	n_estimator	subsample	ROC-AUC	Rank
111	0.1	0.01	3	200	0.7	0.8892596	1
3	0	0.01	3	200	0.7	0.8892222	2
219	0.2	0.01	3	200	0.7	0.8891474	3
220	0.2	0.01	3	200	0.8	0.8885901	4
112	0.1	0.01	3	200	0.8	0.8885895	5
4	0	0.01	3	200	0.8	0.8885521	6
…							
98	0	0.1	5	300	0.9}	0.8624872	322
79	0	0.1	3	300	0.8}	0.8624012	323
97	0	0.1	5	300	0.8}	0.8612169	324

Dampak Hyperparameter terhadap Kinerja Model: Dari Tabel 1, learning_rate 0.01, max_depth 3, n_estimators 200, dan subsample 0.7 menghasilkan kinerja ROC-AUC terbaik. Perubahan gamma tidak terlalu signifikan.

C. Ekstrak Leaf Feature

Proses ini mengambil informasi relevan dari citra daun untuk analisis lebih lanjut. Tabel 4.4 menunjukkan nilai kepentingan fitur:

TABEL 2 (NILAI KEPERCAYAAN FITUR)

Index	Feature	Importance
2	cp	0.199458
8	exang	0.169427
11	ca	0.121395
9	oldpeak	0.077983
1	sex	0.068477
5	fbs	0.060468
12	thal	0.060345
10	slope	0.055246
7	thalach	0.044236
0	age	0.043526
4	chol	0.038360
6	restecg	0.034083
3	trestbps	0.026990

- Fitur paling penting (kepentingan ≥ 0,12) meliputi: cp (nyeri dada), exang (angina yang dipicu olahraga), dan ca (jumlah pembuluh darah utama).
- Fitur dengan kepentingan menengah (0,055 – 0,078) adalah: oldpeak, sex, fbs, thal, dan slope.
- Fitur Kurang Berpengaruh (Di bawah 0.045): thalach, age, chol, restecg, dan trestbps.

D. Tuning Random Forest

Hyperparameter yang disetel: max_depth, max_features, min_samples_split, min_samples_leaf, dan n_estimators.

TABEL 3 (EVALUASI HYPERPARAMETER UNTUK RANDOM FOREST)

index	max_depth	max_features	min_samples_leaf	min_samples_split	n_estimator	ROC-AUC	Rank
105	5	0.8	4	10	100	0.9194381	1
74	5	0.6	4	2	300	0.9193253	2
77	5	0.6	4	5	300	0.9193253	2
84	5	0.8	1	5	100	0.9190241	4
106	5	0.8	4	10	200	0.9189132	5
101	5	0.8	4	2	300	0.9188379	6
..							
351	12	log2	1	2	100	0.9067303	537
432	None	sqrt'	1	2	100	0.9054941	539
459	None	log2	1	2	100	0.9054941	539

Hyperparameter berpengaruh terhadap kinerja model: Berdasarkan Tabel 3, max_depth 5 dan max_features 0.8 atau 0.6 kerap ditemukan pada konfigurasi dengan nilai ROC-AUC tinggi. Ini mengindikasikan bahwa kedalaman pohon yang relatif dangkal serta pembatasan fitur yang efektif berperan penting. Sementara itu, min_samples_leaf, min_samples_split, dan n_estimators memiliki dampak yang kurang signifikan.

E. Evaluasi Akhir

Evaluasi dilakukan pada data pengujian menggunakan metrik Akurasi, Presisi, Recall, F1 Score, dan ROC-AUC, serta waktu eksekusi.

TABEL 4 (HASIL AKHIR)

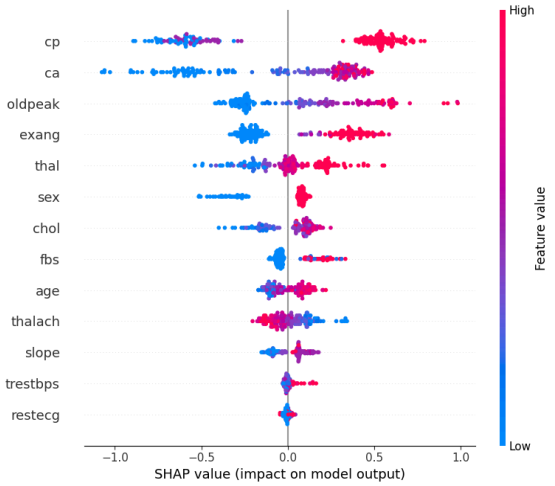
Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	Execution Time
Decision Tree	80,43%	83,33%	81,73%	82,52%	82,29%	0,00118 sec
DT+AdaBoost	82,60%	83,96%	85,57%	84,76%	86,88%	0,02443 sec
LogReg	82,61%	82,14%	**88,46%**	85,18%	87,29%	**0,00117 sec**
SVM	81,52%	83,02%	84,61%	83,81%	86,89%	0,00388 sec
XGBoost	83,15%	82,88%	**88,46%**	85,58%	87,28%	0,00628 sec
RandomForest	83,15%	83,48%	87,50%	85,44%	**88,21%**	0,02849 sec
Stacking (XGB + RF)	82,61%	82,73%	87,50%	85,04%	87,91%	4,70772 sec
Random Forest using XGB Feature *(proposed model)*	**83,69%**	**84,26%**	87,50%	**85,85%**	86,27%	0,02723 sec

Tabel 4 (Hasil Akhir): Model Random Forest + XGB Feature (proposed model) menunjukkan kinerja terbaik dengan Akurasi 83.69%, Presisi 84.26%, Recall 87.50%, F1-score 85.85%, dan ROC-AUC 86.27%, dengan waktu eksekusi 0.02723 detik. Hal ini menunjukkan bahwa memanfaatkan leaf features dari XGBoost sebagai masukan tambahan untuk Random Forest adalah pendekatan yang sangat efisien.

F. Interpretasi Menggunakan SHAP

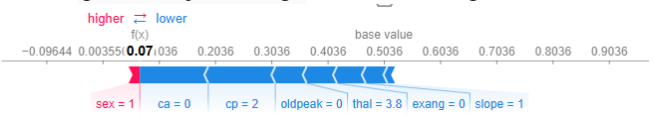
SHAP digunakan untuk menjelaskan mengapa model membuat prediksi tertentu, baik secara global maupun pada tingkat individu.

Summary Plot: Menunjukkan cp (nyeri dada) sebagai fitur paling berpengaruh secara keseluruhan Gambar 2 (Summary Plot).

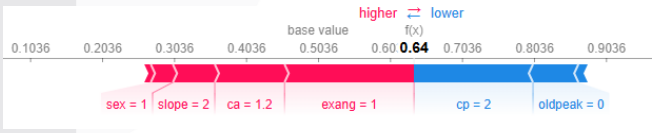


GAMBAR 2 (Summary Plot)

Force plot: Menjelaskan prediksi individu pasien.



GAMBAR 3 (Force Plot Low)



GAMBAR 4 (Force Plot High)

Contoh diberikan untuk pasien dengan risiko rendah (f(x): 0.07) yang didukung oleh fitur seperti ca=0, oldpeak=0, exang=0, dan slope=1, serta pasien dengan risiko tinggi (f(x): 0.64) yang didukung oleh ca=1.2, exang=1, dan slope=2 (Gambar 3 dan Gambar 4).

Analisis SHAP ini meningkatkan kepercayaan pengguna terhadap model dan memberikan pemahaman lebih dalam tentang mekanisme di balik prediksi risiko penyakit jantung, mendukung pengambilan keputusan klinis yang lebih baik.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi penyakit jantung yang akurat dan transparan memanfaatkan metode ensemble stacking XGBoost-Random Forest

(XGB-RF). Setelah melakukan pra-pemrosesan data yang teliti, termasuk penanganan nilai hilang dengan KNN Imputer, model XGB-RF menunjukkan performa unggul. Model ini mencapai metrik kinerja yang tinggi, yaitu akurasi 83.69%, presisi 84.26%, recall 87.50%, F1-score 85.85%, dan ROC-AUC 86.27%. Pencapaian ini didukung oleh kombinasi hyperparameter yang optimal. Selain itu, integrasi SHapley Additive exPlanations (SHAP) memungkinkan interpretasi prediksi model secara mendalam, mengidentifikasi fitur-fitur utama yang memengaruhi risiko penyakit jantung, sehingga meningkatkan kepercayaan dan transparansi model untuk aplikasi klinis.

REFERENSI

[1] Cardiovascular diseases - World Health Organization (WHO), https://www.who.int/health-topics/cardiovascular-diseases

[2] Mensah, G. A., Habtegiorgis Abate, Y., Abbasian, M., Abd-Allah, F., Abdollahi, A., Abdollahi, M., Morad Abdulah, D., Abdullahi, A., Abebe, A. M., Abedi, A., Abedi, A., Olusola Abiodun, O., Ali, H. A., Abu-Gharbieh, E., Abu-Rmeileh, N. M. E., Aburuz, S., Abushouk, A. I., Abu-Zaid, A., Adane, T. D., … Roth, G. A. (2023). Global Burden of Cardiovascular Diseases and Risks, 1990-2022. Journal of the American College of Cardiology, 82(25). https://doi.org/10.1016/j.jacc.2023.11.007

[3] Anderies, A., Tchin, J. A. R. W., Putro, P. H., Darmawan, Y. P., & Gunawan, A. A. S. (2022). Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms. Engineering, MAthematics and Computer Science (EMACS) Journal, 4(3). https://doi.org/10.21512/emacsjournal.v4i3.8683

[4] Saravana Kumar, K., & Ramasubramanian, S. (2023). A clinical decision support system for heart disease prediction with ensemble two-fold classification framework. Journal of Intelligent and Fuzzy Systems, 44(1). https://doi.org/10.3233/JIFS-221165

[5] Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. Journal of Healthcare Engineering, 2022. https://doi.org/10.1155/2022/7351061

[6] Yang, J. C. (2024). The prediction and analysis of heart disease using XGBoost algorithm. Applied and Computational Engineering, 41(1), 61–68. https://doi.org/10.54254/2755-2721/41/20230711

[7] Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information Sciences, 34(7), 4514–4523. https://doi.org/10.1016/j.jksuci.2020.10.013

[8] Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2023). Explainable AI for Retinoblastoma Diagnosis: Interpreting Deep Learning Models with LIME and SHAP. Diagnostics, 13(11). https://doi.org/10.3390/diagnostics13111932

[9] el Hamdaoui, H., Boujraf, S., Chaoui, N. E. H., Alami, B., & Maaroufi, M. (2021). Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms. International Journal of Online and Biomedical Engineering, 17(11). https://doi.org/10.3991/ijoe.v17i11.24781.

[10] Ahamad, G. N., Shafiullah, Fatima, H., Imdadullah, Zakariya, S. M., Abbas, M., Alqahtani, M. S., & Usman, M. (2023). Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease. Processes, 11(3).

[11] Yoon, T., & Kang, D. (2023). Multi-Modal Stacking Ensemble for the Diagnosis of Cardiovascular Diseases. Journal of Personalized Medicine, 13(2). https://doi.org/10.3390/jpm13020373

[12] Gupta, P., & Seth, D. (2023). Comparative analysis and feature importance of machine learning and deep learning for heart disease prediction. Indonesian Journal of Electrical Engineering and Computer Science, 29(1). https://doi.org/10.11591/ijeecs.v29.i1.pp451-45