

Optimasi Hyperparameter Ensemble Learning untuk Prediksi Perkembangan Penyakit Diabetes dengan Explainable AI

David Suharjanto ^{(1)*}, Muhammad Syafiq Akmal ⁽²⁾, Nur Fikri Khuluq ⁽³⁾, Muh Naufal Muzhaffar ⁽⁴⁾, Maria Ulfah Siregar ⁽⁵⁾

Departemen Informatika, UIN Sunan Kalijaga, Yogyakarta, Indonesia
e-mail : {24206052003,24206052002,24206052001,24206052008}@student.uin-suka.ac.id,
maria.siregar@uin-suka.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 31 Desember 2025, direvisi 21 April 2026, diterima 21 April 2026, dan dipublikasikan 25 Mei 2026.

Abstract

This research focuses on optimizing and assessing ensemble learning models for predicting diabetes progression by combining hyperparameter tuning and explainable artificial intelligence techniques. Experiments were conducted using the scikit-learn diabetes dataset, which contains 442 samples with ten numerical features representing patients' clinical conditions. The data were split into 80% for training and 20% for testing. Two ensemble methods were explored: Random Forest Regressor (bagging) and XGBoost Regressor (boosting). Hyperparameter optimization was carried out using RandomizedSearchCV and BayesianSearchCV under a five-fold cross-validation scheme. Model performance was evaluated using MAE, MSE, RMSE, and R^2 metrics, while interpretability was examined through SHAP summary plots. The results indicate that BayesianSearchCV consistently delivered superior performance gains compared to random search. In particular, the optimized XGBoost model achieved an R^2 score of 0.5018, improving by 19.8% over the baseline model ($R^2 = 0.4188$), and reduced RMSE from 55.49 to 51.37. SHAP analysis showed that serum triglycerides, body mass index, and blood pressure were the most influential features. Overall, the findings suggest that Bayesian-based hyperparameter optimization can effectively improve ensemble regression performance in medical prediction tasks involving limited datasets.

Keywords: Ensemble Learning, Hyperparameter Optimization, Random Forest, XGBoost, Diabetes Prediction

Abstrak

Penelitian ini bertujuan untuk mengoptimalkan dan mengevaluasi model ensemble learning untuk memprediksi perkembangan penyakit diabetes melalui hyperparameter tuning dan teknik explainable AI menggunakan dataset diabetes dari scikit-learn. Dataset terdiri atas 442 sampel yang dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, mencakup sepuluh fitur numerik yang menggambarkan karakteristik klinis pasien. Dua model utama digunakan, yaitu Random Forest Regressor sebagai pendekatan Bagging dan XGBoost Regressor sebagai pendekatan Boosting. Optimasi hyperparameter dilakukan menggunakan RandomizedSearchCV dan BayesianSearchCV dengan five-fold cross-validation. Kinerja model dievaluasi menggunakan metrik MAE, MSE, RMSE, dan R^2 , sementara interpretabilitas dianalisis melalui SHAP (SHapley Additive exPlanations) Summary Plot. Hasil menunjukkan bahwa BayesianSearchCV memberikan peningkatan kinerja yang paling signifikan. Model XGBoost yang telah dioptimalkan mencapai nilai R^2 sebesar 0,5018, meningkat 19,8% dibandingkan dengan baseline ($R^2 = 0,4188$), serta menurunkan RMSE dari 55,49 menjadi 51,37. Analisis SHAP mengungkapkan bahwa fitur 8 (serum trigliserida), 2 (BMI), dan 3 (tekanan darah) memberikan kontribusi terbesar terhadap keluaran prediktif. Temuan ini menegaskan bahwa BayesianSearchCV efektif meningkatkan akurasi model ensemble regression dalam tugas prediksi medis berbasis data terbatas dan berpotensi menjadi strategi optimasi hyperparameter yang andal untuk aplikasi di bidang kesehatan.

Kata Kunci: Ensemble Learning, Optimasi Hyperparameter, Random Forest, XGBoost, Prediksi Diabetes



1. PENDAHULUAN

Diabetes merupakan salah satu penyebab utama kematian dan kecacatan di seluruh dunia, yang dapat menyerang individu tanpa memandang negara, jenis kelamin, maupun usia (Hossain et al., 2024; Ong et al., 2023). Diabetes dipengaruhi oleh berbagai faktor risiko umum, seperti obesitas, pola makan tidak sehat, kurangnya aktivitas fisik, merokok, hipertensi, serta peningkatan kadar glukosa, kolesterol, dan trigliserida (Peters et al., 2019; Wahidin et al., 2024; Zhang et al., 2019). Perkembangan penyakit diabetes yang tidak terkontrol dapat meningkatkan risiko komplikasi serius, sehingga kemampuan untuk memprediksi perkembangan penyakit menjadi sangat penting dalam mendukung perencanaan perawatan dan intervensi klinis yang lebih tepat sasaran (Ong et al., 2023). Dengan kemajuan teknologi kecerdasan buatan, berbagai studi telah memanfaatkan metode machine learning untuk memodelkan dan memprediksi perkembangan penyakit diabetes (Habehh & Gohel, 2021; Mahajan et al., 2023; Ukoba et al., 2025).

Salah satu pendekatan machine learning yang banyak digunakan untuk tugas prediksi medis pada dataset tabular adalah ensemble learning karena kemampuannya menggabungkan banyak model dasar untuk meningkatkan akurasi dan stabilitas prediksi (Khan et al., 2024; Mohammed & Kora, 2023). Dua algoritma ensemble learning yang populer, yaitu Random Forest dan XGBoost, sering menjadi pilihan utama karena performanya yang konsisten pada data klinis dengan hubungan fitur yang kompleks (Gill et al., 2022). Random Forest menawarkan ketahanan terhadap noise melalui mekanisme bagging, sedangkan XGBoost dikenal sebagai metode boosting yang efisien dan mampu memetakan hubungan nonlinier dengan lebih akurat (Imani et al., 2025; Karunakaran et al., 2025). Namun demikian, performa kedua model ini sangat dipengaruhi oleh konfigurasi hyperparameter. Pemilihan hyperparameter yang kurang optimal dapat menyebabkan overfitting, underfitting, atau inefisiensi komputasi, terutama pada dataset berukuran kecil (Wu et al., 2019). Oleh karena itu, optimasi hyperparameter menjadi tahapan krusial untuk memperoleh kinerja model yang optimal (Yan et al., 2025).

Untuk mengatasi permasalahan tersebut, berbagai teknik optimasi hyperparameter telah dikembangkan. Dua pendekatan yang paling banyak digunakan adalah RandomizedSearchCV dan BayesianSearchCV. RandomizedSearchCV menawarkan pencarian ruang parameter yang lebih efisien dibandingkan dengan GridSearchCV melalui pemilihan kombinasi parameter secara acak, sedangkan BayesianSearchCV memanfaatkan pendekatan probabilistik untuk memprediksi kombinasi parameter terbaik secara lebih efektif (Gao et al., 2025). Selain akurasi model, interpretabilitas model juga sangat penting dalam penerapan machine learning di bidang kesehatan karena tenaga medis tidak hanya membutuhkan prediksi yang tepat, tetapi juga penjelasan yang jelas tentang alasan di balik prediksi tersebut (Frasca et al., 2024). Teknik Explainable AI seperti SHAP (SHapley Additive exPlanations) telah banyak digunakan karena dapat memberikan visualisasi kontribusi setiap fitur klinis terhadap prediksi model secara konsisten dan transparan (Santos et al., 2024). Urgensi ini diperkuat oleh temuan terbaru yang menunjukkan bahwa optimasi arsitektur ensemble yang terintegrasi dengan kerangka Explainable AI (XAI) secara signifikan meningkatkan keandalan prediksi klinis (Mienye & Jere, 2024). Selain itu, pendekatan tersebut juga terbukti krusial dalam mencegah bias dalam evaluasi fitur medis yang kompleks (Agrawal et al., 2025; Imans et al., 2024).

Sebagai pembanding, Zhong et al. (2020) menggunakan dataset diabetes scikit-learn dengan pendekatan Linear Regression dan berhasil menurunkan MAE dari 46,17 menjadi 45,18 serta RMSE dari 58,52 menjadi 58,07 setelah dilakukan optimasi. Namun, studi tersebut tidak menyertakan metrik R^2 serta terbatas pada pendekatan linier yang kurang mampu menangkap hubungan fitur yang kompleks, ditambah dengan ketiadaan Explainable AI yang membuat hasil prediksi sulit diinterpretasikan secara klinis. Dataset yang digunakan dalam penelitian ini adalah dataset diabetes scikit-learn yang sama, yang terdiri dari 442 sampel numerik yang merepresentasikan perkembangan penyakit satu tahun setelah baseline. Meskipun menjadi benchmark klasik untuk studi regresi, ukuran dataset yang relatif kecil menghadirkan tantangan bagi model ensemble yang sensitif terhadap konfigurasi hyperparameter. Secara lebih luas, kajian literatur menunjukkan bahwa penelitian yang secara sistematis membandingkan

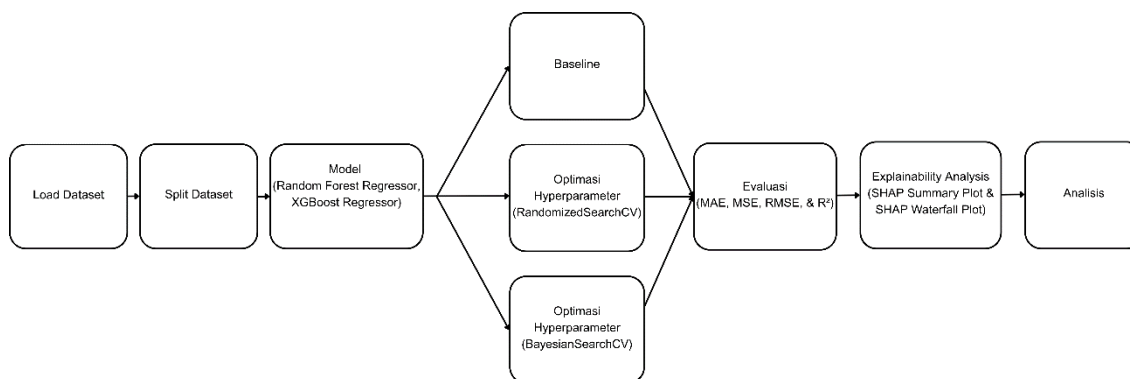


RandomizedSearchCV dan BayesianSearchCV pada dataset berukuran kecil, serta mengintegrasikannya dengan teknik XAI seperti SHAP, masih terbatas. Kondisi ini membuka peluang untuk mengevaluasi bagaimana kombinasi metode optimasi dan XAI dapat meningkatkan akurasi sekaligus interpretabilitas dalam prediksi perkembangan penyakit diabetes.

Melihat kondisi tersebut, penelitian ini dilakukan untuk mengisi celah dalam literatur dengan mengoptimalkan dua model ensemble learning, yaitu Random Forest dan XGBoost, menggunakan teknik optimasi hyperparameter RandomizedSearchCV dan BayesianSearchCV. Penelitian ini juga mengintegrasikan analisis SHAP untuk mengidentifikasi fitur-fitur klinis yang berpengaruh terhadap prediksi perkembangan penyakit diabetes. Secara lebih rinci, kontribusi utama penelitian ini mencakup: (1) analisis komparatif antara RandomizedSearchCV dan BayesianSearchCV dalam mengoptimalkan model Random Forest dan XGBoost pada dataset berukuran kecil; (2) evaluasi performa model menggunakan metrik yang komprehensif, mencakup MAE, RMSE, dan R^2 , guna memberikan penilaian yang lebih menyeluruh dibanding studi sebelumnya; (3) integrasi teknik SHAP untuk menghasilkan interpretasi model secara global maupun lokal, sehingga keluaran numerik dapat ditransformasi menjadi wawasan klinis yang dapat dipertanggungjawabkan; serta (4) justifikasi empiris mengenai relevansi fitur klinis dominan terhadap perkembangan penyakit diabetes. Dengan pendekatan yang terstruktur ini, penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan strategi optimasi model yang lebih efektif, efisien, dan interpretatif, sekaligus menjadi landasan yang lebih kuat bagi pengembangan sistem pendukung keputusan medis yang akurat dan transparan.

2. METODE PENELITIAN

Metode penelitian ini mengikuti alur yang ditunjukkan pada Gambar 1. Proses dimulai dengan memuat dataset diabetes dari pustaka scikit-learn. Setelah data dimuat, dataset dibagi menjadi data pelatihan dan pengujian dengan proporsi 80:20. Meskipun dataset ini secara bawaan telah melalui proses transformasi awal, penelitian ini tetap menerapkan kembali tahap feature scaling menggunakan StandardScaler. Langkah ini dilakukan untuk memastikan setiap variabel memiliki rata-rata nol dan standar deviasi satu guna menjaga stabilitas konvergensi pada model. Tahap berikutnya adalah membangun model regresi menggunakan dua algoritma ensemble, yaitu Random Forest Regressor (bagging) dan XGBoost Regressor (boosting). Kedua model awalnya dilatih sebagai baseline tanpa optimasi hyperparameter sebagai titik acuan kinerja. Setelah itu, performa model ditingkatkan melalui dua metode optimasi hyperparameter, yaitu RandomizedSearchCV dan BayesianSearchCV. Dengan demikian, setiap algoritma menghasilkan tiga varian model, yaitu baseline, hasil RandomizedSearchCV, dan hasil BayesianSearchCV. Seluruh model kemudian dievaluasi menggunakan empat metrik regresi, yaitu MAE, MSE, RMSE, dan R^2 , untuk memberikan gambaran yang komprehensif tentang kinerja prediksi. Pada tahap akhir, hasil pelatihan dianalisis lebih lanjut menggunakan SHAP untuk mengevaluasi kontribusi setiap fitur dan meningkatkan interpretabilitas prediksi.



Gambar 1 Diagram Alir Penelitian



2.1 Dataset

Data yang digunakan dalam penelitian ini adalah dataset diabetes yang disediakan oleh pustaka scikit-learn, dengan proporsi 80% untuk training dan 20% untuk test, serta random state 42. Dataset ini terdiri dari 442 sampel pasien yang dapat diakses melalui https://scikit-learn.org/stable/datasets/toy_dataset.html. Setiap sampel mencakup 10 fitur numerik yang merepresentasikan karakteristik klinis dan demografis pasien (seperti usia, jenis kelamin, BMI, tekanan darah, serta enam pengukuran serum darah). Target variabel adalah ukuran kuantitatif perkembangan penyakit diabetes satu tahun setelah baseline.

2.2 Arsitektur Model

Dua model ensemble regression digunakan dalam penelitian ini, yaitu Random Forest dan XGBoost. Random Forest beroperasi dengan membangun banyak decision tree secara paralel dan melakukan agregasi prediksi, sehingga lebih stabil terhadap overfitting pada dataset berskala kecil. Sebaliknya, XGBoost menerapkan strategi boosting berurutan dengan penambahan pohon secara adaptif, sehingga memiliki performa tinggi namun sangat peka terhadap konfigurasi hyperparameter.

2.3 Lingkungan Pengembangan

Seluruh eksperimen dilakukan menggunakan Python 3.12 dengan pustaka scikit-learn, XGBoost, scikit-optimize (skopt) untuk Bayesian optimization, serta SHAP sebagai metode interpretabilitas. Proses pelatihan dijalankan sepenuhnya di Google Colaboratory menggunakan runtime CPU dengan kapasitas RAM maksimum 12,7 GB, tanpa akselerasi GPU.

2.4 Konfigurasi Pelatihan

Bagian ini menjelaskan konfigurasi pelatihan yang digunakan dalam proses pengembangan dan evaluasi model. Konfigurasi tersebut mencakup pengaturan model baseline sebagai acuan awal, serta konfigurasi optimasi hyperparameter menggunakan metode RandomizedSearchCV dan BayesianSearchCV untuk memperoleh performa model yang lebih optimal. Rincian konfigurasi pada masing-masing pendekatan disajikan pada Tabel 1, Tabel 2, dan Tabel 3.

Tabel 1 Konfigurasi Baseline Model

Model	Parameter	Nilai
Random Forest	n_estimators	300
	max_depth	None
	max_features	1
XGBoost	bootstrap	True
	n_estimators	100
	learning_rate	0.1
	max_depth	6
	subsample	0.8
	colsample_bytree	0.8
	reg_lambda	1
	reg_alpha	0



Tabel 2 Konfigurasi Optimasi Hyperparameter RandomizedSearchCV

Model	Parameter	Tipe	Nilai
Random Forest	n_estimators	integer	100-1000
	max_depth	integer	3-20
	min_samples_split	integer	2-10
	min_samples_leaf	integer	1-10
	max_features	Kategori	['sqrt', 'log2', none, 1]
	bootstrap	Kategori	[True, False]
XGBoost	n_estimators	Integer	100-1000
	learning_rate	uniform	[0.1,0.19]
	max_depth	integer	3-12
	subsample	integer	[0.6, 0.8, 1]
	colsample_bytree	integer	[0.6, 0.8, 1]
	min_child_weight	integer	[1, 3, 5, 7]
	gamma	integer	[0, 0.1, 0.3, 0.5]
	reg_lambda	integer	[0, 0.1, 0.5]
	reg_alpha	integer	[1, 1.5, 2]

Tabel 3 Konfigurasi Optimasi Hyperparameter BayesianSearchCV

Model	Parameter	Tipe	Nilai
Random Forest	n_estimators	integer	100-1000
	max_depth	integer	3-20
	min_samples_split	integer	2-10
	min_samples_leaf	integer	1-10
	max_features	kategori	['sqrt', 'log2', none, 1]
	bootstrap	kategori	[True, False]
XGBoost	n_estimators	Integer	100-1000
	learning_rate	real, prior: log-uniform	[0.01, 0.3]
	max_depth	integer	3-12
	subsample	real	[0.6, 1]
	colsample_bytree	real	[0.6, 1]
	min_child_weight	integer	[1, 8]
	gamma	real	[0, 0.5]
	reg_lambda	real	[0, 0.5]
	reg_alpha	real	[1, 2]

2.5 Metrik Evaluasi

Untuk menilai kinerja seluruh model regresi, penelitian ini menggunakan empat metrik evaluasi utama, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2).

- MAE mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, sehingga memberikan gambaran tentang kesalahan model dalam satuan asli target.
- MSE menghitung rata-rata kuadrat kesalahan prediksi, sehingga memberikan penalti yang lebih besar terhadap kesalahan yang lebih besar.
- RMSE merupakan akar kuadrat dari MSE dan memudahkan interpretasi karena kembali ke skala target.
- R^2 menunjukkan proporsi variansi target yang dapat dijelaskan oleh model, di mana nilai yang mendekati 1 menandakan model memiliki kemampuan prediktif yang lebih baik.

Penggunaan keempat metrik ini memberikan evaluasi yang komprehensif terhadap performa model, mencakup akurasi prediksi, stabilitas, serta kemampuan model dalam menjelaskan variasi data.



3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen Baseline

Tabel 4 menyajikan performa awal (baseline) dari dua model ensemble learning, yaitu Random Forest dan XGBoost. Eksperimen ini dilakukan menggunakan Konfigurasi Baseline Model sebagaimana telah dirinci dalam Tabel 1. Hasil ini berfungsi sebagai titik acuan untuk mengukur efektivitas optimasi hyperparameter yang dilakukan pada tahap berikutnya. Pada baseline Random Forest, model menunjukkan MAE 44,7520, MSE 3016,3541, RMSE 54,9213, dan R² 0,4307. Nilai R² ini sedikit lebih baik dibandingkan dengan baseline XGBoost, yang menunjukkan bahwa Random Forest dapat menjelaskan kurang lebih 43,07% perbedaan dalam skor perkembangan penyakit diabetes, sehingga sedikit lebih unggul daripada XGBoost yang tidak di-tuning. Hal ini sejalan dengan sifat Random Forest yang umumnya stabil dan cukup tangguh meskipun menggunakan pengaturan bawaan, karena proses bagging dan penggabungan banyak pohon keputusan mengurangi varians model.

Tabel 4 Hasil Baseline Model Ensemble Learning

Model	MAE	MSE	RMSE	R ²
Random Forest	44,7520	3016,3541	54,9213	0,4307
XGBOOST	44,2491	3079,4574	55,4929	0,4188

Sementara itu, pada konfigurasi baseline, XGBoost menghasilkan MAE sebesar 44,2491, MSE sebesar 3079,4574, RMSE sebesar 55,4929, dan R² sebesar 0,4188. Angka tersebut menunjukkan bahwa model dapat menjelaskan sekitar 41,88% variasi skor perkembangan penyakit diabetes pada data uji. Nilai RMSE yang masih di atas 55 menunjukkan bahwa rata-rata deviasi antara prediksi dan nilai nyata masih cukup besar, sehingga terdapat peluang perbaikan, baik dalam hal kemampuan model maupun pengaturan hyperparameternya. Meskipun demikian, sebagai dasar tanpa optimasi, kinerja ini menunjukkan bahwa XGBoost sudah dapat menangkap pola hubungan nonlinier antara fitur klinis dan perkembangan diabetes dengan cukup baik. Meskipun demikian, nilai R² dari kedua model yang masih berada di bawah 0,5 menunjukkan bahwa sebagian besar variasi data belum sepenuhnya dapat dijelaskan oleh model. Kondisi ini menegaskan perlunya optimasi hyperparameter agar performa model dapat ditingkatkan secara signifikan.

3.2 Hasil Optimasi Hyperparameter

Tabel 5 menunjukkan hasil optimasi hyperparameter pada kedua model menggunakan metode RandomizedSearchCV dan BayesianSearchCV. Hasil tersebut dibandingkan dengan performa baseline untuk melihat pengaruh proses optimasi terhadap kinerja model. Secara umum, kedua metode optimasi mampu meningkatkan performa model, meskipun tingkat peningkatannya berbeda pada Random Forest dan XGBoost.

Tabel 5 Hasil Optimasi Hyperparameter

Model	Optimasi	MAE	MSE	RMSE	R ²
Random Forest	RandomizedSearchCV	43,3415	2790,9216	52,8292	0,4732
	BayesianSearchCV	43,4272	2788,1631	52,8031	0,4737
XGBOOST	RandomizedSearchCV	42,5728	2791,4117	52,8338	0,4731
	BayesianSearchCV	41,6576	2639,4588	51,3757	0,5018

Pada Random Forest, penerapan RandomizedSearchCV menghasilkan peningkatan kinerja yang konsisten. MAE menurun menjadi 43,3415, MSE turun menjadi 2790,9216, dan RMSE turun menjadi 52,8292. Jika dibandingkan dengan baseline, MAE berkurang sekitar 3,15%, sementara nilai R² meningkat dari 0,4307 menjadi 0,4732 (kenaikan sekitar 0,0425 poin). Peningkatan ini menunjukkan bahwa pemilihan parameter seperti jumlah pohon, kedalaman maksimum, dan



jumlah minimal sampel pada setiap node dapat membantu model menangkap variasi data dengan lebih baik.

Optimasi menggunakan BayesianSearchCV menghasilkan hasil yang hampir sama. MAE sedikit berubah menjadi 43,4272, MSE turun tipis menjadi 2788,1631, dan RMSE menjadi 52,8031, sementara R^2 meningkat menjadi 0,4737. Jika dibandingkan dengan baseline, MAE berkurang sekitar 2,96% dan R^2 naik sekitar 0,043 poin. Perbedaan performa antara RandomizedSearchCV dan BayesianSearchCV hanya sekitar 0,0005 poin pada R^2 , menunjukkan bahwa ruang perbaikan Random Forest relatif sempit. Model baseline Random Forest pada dasarnya sudah cukup stabil, sehingga kedua metode optimasi dengan cepat mencapai titik performa maksimum. Dengan demikian, Random Forest dalam penelitian ini cenderung kokoh tetapi tidak terlalu sensitif terhadap tuning atau optimasi hyperparameter.

Sementara itu, pada XGBoost, efek optimasi terlihat lebih jelas. Penerapan RandomizedSearchCV menurunkan MAE menjadi 42,5728, sementara MSE dan RMSE turun menjadi 2791,4117 dan 52,8338. Penurunan MAE sekitar 3,79% dari baseline menandakan bahwa kesalahan prediksi rata-rata mengecil secara signifikan. Nilai R^2 juga naik dari 0,4188 menjadi 0,4731 (kenaikan sekitar 0,0543 poin), menandakan bahwa XGBoost yang telah dituning mampu menangkap pola hubungan antara fitur dan target dengan lebih efektif.

Optimasi lanjutan menggunakan BayesianSearchCV memberikan peningkatan performa yang paling signifikan di antara semua skenario. MAE turun menjadi 41,6576, MSE menurun drastis menjadi 2639,4588, dan RMSE turun menjadi 51,3757. Nilai R^2 meningkat hingga 0,5018, melewati ambang 0,5, dan menunjukkan bahwa model mampu menjelaskan lebih dari separuh variasi skor perkembangan penyakit diabetes pada data uji. Dibandingkan dengan baseline, kenaikan R^2 sekitar 0,083 poin menunjukkan bahwa pendekatan Bayesian lebih efektif dalam menemukan kombinasi hyperparameter terbaik. Meskipun nilai R^2 berada pada level moderat, angka ini mencerminkan kompleksitas yang tinggi dalam pemodelan data klinis. Keterbatasan varians yang terjelaskan disebabkan oleh penggunaan dataset benchmark yang hanya mencatat indikator fisiologis dan metabolik dasar seperti BMI, glukosa, dan tekanan darah. Selain itu, ukuran dataset yang relatif kecil (442 sampel) membatasi kemampuan model untuk menangkap keragaman pola progresi penyakit yang lebih luas. Sisa varians yang tidak terjelaskan kemungkinan berasal dari faktor risiko eksternal yang tidak tercatat, seperti profil genetik, pola hidup, riwayat diet, serta penyakit komorbid. Kesesuaian hasil ini sejalan dengan temuan bahwa nilai R^2 moderat dapat dianggap bermakna secara klinis karena memengaruhi variabilitas biologis (Gupta et al., 2024). Dengan demikian, XGBoost yang dioptimasi menggunakan BayesianSearchCV dapat dianggap sebagai model dengan performa terbaik dalam penelitian ini.

3.3 Perbandingan Kinerja Semua Model

Tabel 6 menyajikan perbandingan peningkatan performa seluruh model sebelum dan sesudah optimasi hyperparameter berdasarkan metrik R^2 . Secara umum, kedua model ensemble learning, baik Random Forest maupun XGBoost, menunjukkan peningkatan kinerja setelah dioptimalkan. Namun, tingkat sensitivitas terhadap tuning menunjukkan perbedaan yang cukup jelas antara keduanya.

Tabel 6 Perbandingan Peningkatan Performa Model Berdasarkan R^2

Model	Optimasi	R^2	ΔR^2	ΔR^2 (%)
Random Forest	Baseline	0,4307	-	-
	RandomizedSearchCV	0,4732	+0,0425	+9,86
	BayesianSearchCV	0,4737	+0,0430	+9,98
XGBoost	Baseline	0,4188	-	-
	RandomizedSearchCV	0,4731	+0,0543	+12,96
	BayesianSearchCV	0,5018	+0,0830	+19,81



Untuk Random Forest, peningkatan performa tergolong moderat. Dari nilai R^2 baseline sebesar 0,4307, RandomizedSearchCV meningkatkan performa menjadi 0,4732, sementara BayesianSearchCV sedikit lebih tinggi, yaitu 0,4737. Keduanya menghasilkan kenaikan R^2 sekitar 0,0425–0,0430 poin, atau peningkatan relatif 9,86% hingga 9,98%. Kenaikan yang relatif kecil ini mengindikasikan bahwa Random Forest sudah bekerja cukup optimal bahkan pada konfigurasi default. Struktur model berbasis bagging yang stabil membuatnya tidak terlalu sensitif terhadap pencarian kombinasi hyperparameter, sehingga ruang perbaikan cenderung terbatas. Dengan kata lain, Random Forest memberikan performa yang andal tetapi tidak terlalu responsif terhadap tuning lanjutan.

Sebaliknya, XGBoost menunjukkan peningkatan performa yang jauh lebih signifikan setelah dioptimalkan. Dari baseline R^2 sebesar 0,4188, penerapan RandomizedSearchCV meningkatkan nilai R^2 menjadi 0,4731, atau naik sekitar 12,96%. Peningkatan yang lebih besar terjadi ketika menggunakan BayesianSearchCV, yang menghasilkan R^2 0,5018, atau kenaikan 19,81% dari baseline, sehingga menjadikannya satu-satunya model yang melampaui ambang 0,5. Hal ini menunjukkan bahwa XGBoost memiliki ruang perbaikan hyperparameter yang lebih luas dan sangat bergantung pada pemilihan parameter yang tepat, seperti learning rate, depth, subsample, dan regularisasi. Dengan proses optimasi yang lebih sistematis seperti BayesianSearchCV, XGBoost mampu menangkap kompleksitas hubungan nonlinier dalam data secara lebih efektif.

Perbandingan ini menegaskan bahwa meskipun kedua model merupakan metode ensemble, karakteristik internalnya berbeda dalam hal respons terhadap tuning hyperparameter. Random Forest cenderung memiliki performa yang stabil bahkan tanpa optimasi yang berat, sedangkan XGBoost sangat bergantung pada penyesuaian parameter untuk mencapai performa terbaik. Secara keseluruhan, XGBoost dengan teknik optimasi hyperparameter BayesianSearchCV muncul sebagai konfigurasi paling unggul dalam penelitian ini, dengan kemampuan menjelaskan variabilitas data yang lebih baik dibandingkan dengan semua model lainnya.

3.4 Analisis SHAP

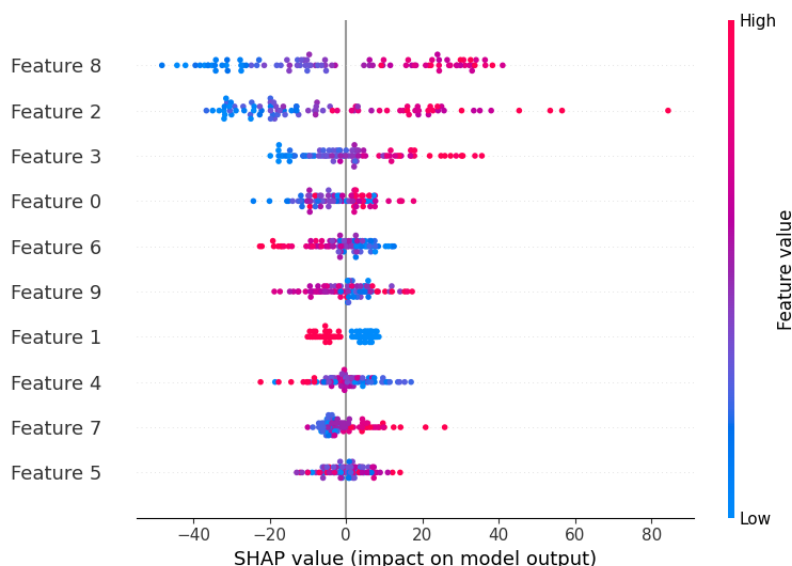
Analisis interpretabilitas dilakukan menggunakan SHAP pada model terbaik, yaitu XGBoost yang dioptimasi dengan BayesianSearchCV. Sebagai acuan perbandingan, analisis juga dilakukan pada model XGBoost baseline untuk mengevaluasi apakah proses optimasi hyperparameter memengaruhi konsistensi interpretasi fitur, bukan hanya performa prediktif. Berdasarkan SHAP summary plot model baseline pada Gambar 2, tiga fitur dengan kontribusi terbesar adalah s5/trigliserida (fitur 8), BMI (fitur 2), dan bp/tekanan darah (fitur 3). Meskipun ketiga fitur teratas ini konsisten, fitur age (fitur 0) berada di posisi keempat dan s3/HDL (fitur 6) berada di posisi kelima, menunjukkan bahwa model baseline masih memberikan bobot yang relatif besar pada faktor demografis dibandingkan biomarker metabolik tertentu.

Pada model yang dioptimasi dengan BayesianSearchCV seperti yang ditampilkan pada Gambar 3, tiga fitur dominan tetap konsisten, yaitu s5/trigliserida, BMI, dan bp/tekanan darah. Namun, terdapat pergeseran yang signifikan pada fitur-fitur berikutnya: s3/HDL (fitur 6) naik dari posisi kelima menjadi keempat, sedangkan age (fitur 0) turun dari posisi keempat menjadi ketujuh. Pergeseran ini mengindikasikan bahwa BayesianSearchCV membantu model menangkap interaksi biomarker metabolik, seperti hubungan antara HDL, BMI, dan tekanan darah, secara lebih akurat, sekaligus mengurangi ketergantungan pada fitur demografis yang kurang spesifik terhadap progresi diabetes. Untuk melihat pergeseran tersebut secara lebih rinci, perbandingan lengkap urutan kontribusi fitur antara model XGBoost baseline dan hasil optimasi BayesianSearchCV disajikan pada Tabel 7.

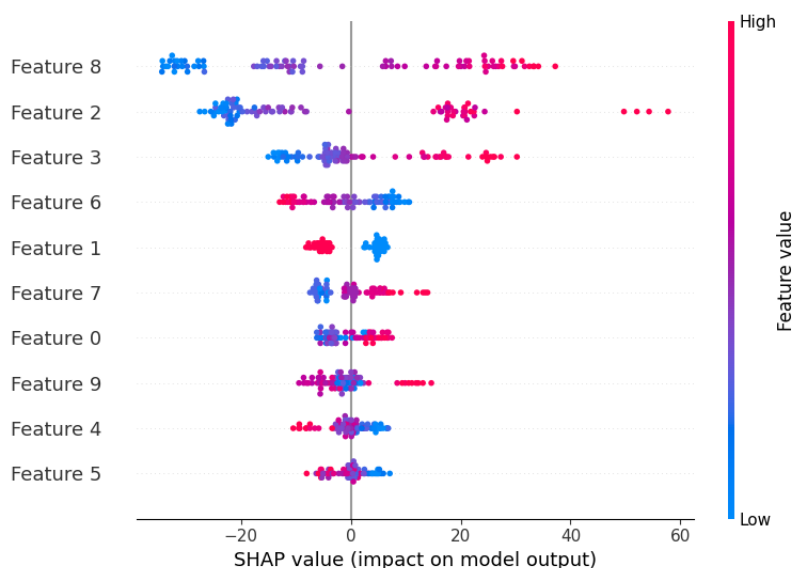
Distribusi nilai SHAP pada ketiga fitur tersebut menunjukkan pola yang konsisten. Titik-titik dengan nilai fitur yang lebih tinggi cenderung menghasilkan nilai SHAP positif dan terletak di sisi kanan grafik, yang berarti model memprediksi progresi diabetes yang lebih besar ketika kadar trigliserida, BMI/indeks massa tubuh, atau tekanan darah meningkat. Sebaliknya, nilai yang lebih rendah pada fitur-fitur tersebut umumnya memiliki SHAP negatif yang menunjukkan kontribusi



terhadap penurunan skor prediksi. Secara visual, summary plot memperlihatkan dua informasi penting sekaligus. Pertama, warna pada tiap titik merepresentasikan nilai fitur, dengan warna biru menunjukkan nilai rendah dan warna merah menunjukkan nilai tinggi. Warna ini membantu mengidentifikasi apakah nilai fitur yang tinggi atau yang rendah yang mendorong prediksi naik. Pada fitur 8, 2, dan 3 terlihat dominasi titik merah pada sisi SHAP positif, yang berarti nilai fitur tinggi secara konsisten meningkatkan prediksi progresi diabetes. Kedua, posisi titik pada sumbu horizontal menggambarkan besarnya dampak fitur terhadap output model, di mana semakin ke kanan, semakin besar kontribusi positif, sedangkan posisi ke kiri mengindikasikan efek penurunan prediksi.



Gambar 2 SHAP Summary Plot untuk Model XGBoost (Baseline)



Gambar 3 SHAP Summary Plot untuk Model XGBoost (BayesianSearchCV)

Hasil ini sejalan dengan pengetahuan klinis, yaitu kadar trigliserida yang meningkat sering berkaitan dengan resistensi insulin, BMI tinggi berhubungan dengan risiko gangguan metabolik, dan tekanan darah tinggi merupakan komorbiditas yang memperburuk kondisi metabolik pasien, yang secara bersamaan dapat mempercepat perkembangan diabetes (Havelda et al., 2025; Wondmkun, 2020). Dengan demikian, SHAP tidak hanya meningkatkan transparansi model,

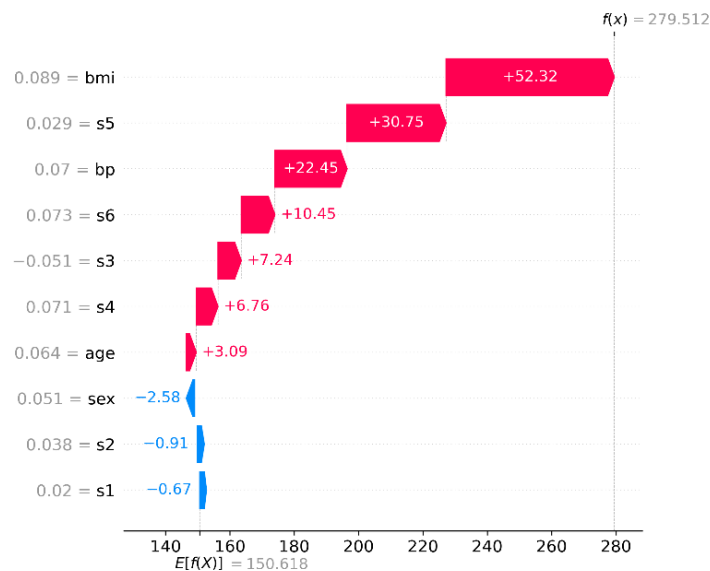


tetapi juga memberikan justifikasi medis yang kuat bahwa model mempelajari pola yang relevan secara klinis. Interpretasi ini memperkuat keandalan model dalam memprediksi perkembangan penyakit dan dapat membantu tenaga medis memahami faktor-faktor yang paling memengaruhi hasil prediksi pada setiap pasien.

Tabel 7 Perbandingan Ranking Fitur SHAP: XGBoost Baseline vs BayesianSearchCV

Rank	Fitur Baseline	Fitur Optimasi
1	s5/Trigliserida	s5/Trigliserida
2	BMI	BMI
3	bp/Tekanan Darah	bp/Tekanan Darah
4	Age/Usia	s3/HDL
5	s3/HDL	Sex/Jenis Kelamin
6	s6/Glukosa	s4/TCH
7	Sex/Jenis Kelamin	Age/Usia
8	s1/Kolesterol Total	s6/Glukosa
9	s4/TCH	s1/Kolesterol Total
10	s2/LDL	s2/LDL

Selain analisis secara global, penelitian ini juga menyertakan analisis interpretabilitas pada tingkat individu untuk memberikan pemahaman yang lebih mendalam bagi para praktisi medis. Analisis lokal dilakukan menggunakan SHAP Waterfall Plot pada sampel pasien dengan indeks ke-17 yang memiliki nilai prediksi tertinggi sebesar 279,512. Berdasarkan Gambar 4, prediksi dimulai dari expected value $E[f(x)] = 150,618$. Pendorong utama yang meningkatkan skor prediksi secara signifikan pada pasien ini selaras dengan tiga fitur utama pada analisis global, yaitu nilai tinggi pada fitur BMI (+52,32), s5/trigliserida (+30,75), dan bp/tekanan darah (+22,45). Sebaliknya, fitur seperti jenis kelamin (sex) dan serum s1-s2 memberikan kontribusi negatif yang kecil terhadap hasil akhir. Analisis ini membuktikan bahwa model XGBoost yang telah dioptimasi mampu memberikan justifikasi medis yang transparan dan logis bagi setiap pasien.



Gambar 4 SHAP Waterfall Plot pada Pasien dengan Prediksi Tertinggi

3.5 Ablation Study

Hasil ablation study menunjukkan pengaruh penghapusan fitur terhadap kinerja model XGBoost yang telah dioptimalkan menggunakan BayesianSearchCV, sebagaimana ditunjukkan pada Tabel 8. Model optimal, yang menggunakan semua fitur dengan konfigurasi hyperparameter



terbaik, menghasilkan performa R^2 sebesar 0,5018, MAE 41,66, MSE 2639,46 dan RMSE 51,38. Model ini dijadikan baseline untuk membandingkan eksperimen lainnya.

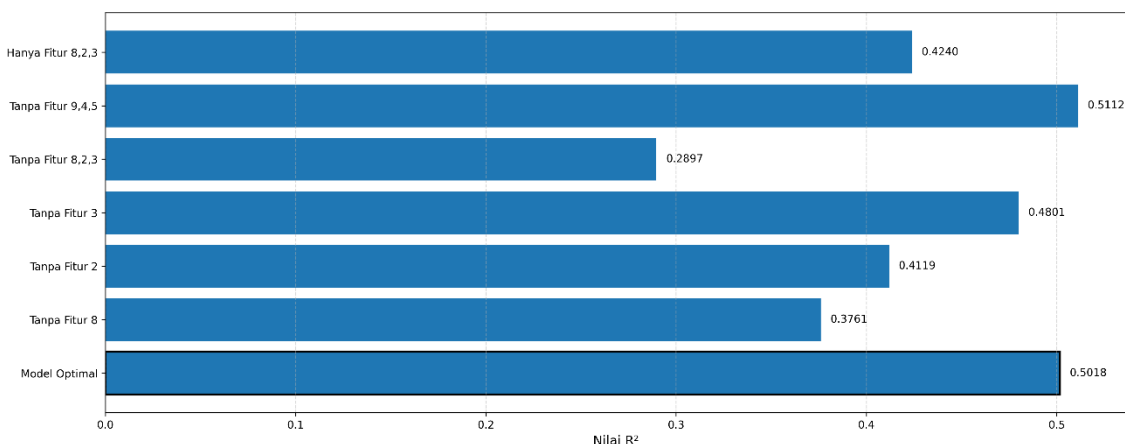
Tabel 8 Hasil Ablation Study pada Model Optimal XGBoost

Eksperimen	MAE	MSE	RMSE	R^2
Model Optimal (XGBoost + Bayesian)	41,6576	2639,4588	51,3757	0,5018
Tanpa Fitur 8	46,2811	3305,4464	57,4930	0,3761
Tanpa Fitur 2	45,9900	3115,7402	55,8188	0,4119
Tanpa Fitur 3	42,8423	2754,2776	52,4812	0,4801
Tanpa Fitur 8, 2, 3	52,1727	3763,4916	61,3473	0,2897
Tanpa Fitur 9, 4, 5	41,3679	2589,5701	50,8878	0,5112
Hanya Fitur 8, 2, 3	45,8291	3051,6620	55,2419	0,4240

Ketika fitur paling penting dihapus satu per satu, terlihat penurunan performa yang bervariasi. Penghapusan fitur 8 menyebabkan R^2 turun drastis menjadi 0,3761, sedangkan penghapusan fitur 2 dan 3 menurunkan R^2 menjadi 0,4119 dan 0,4801. Penurunan signifikan ini menegaskan bahwa ketiga fitur yang teridentifikasi sebagai paling berpengaruh melalui analisis SHAP memberikan kontribusi utama terhadap kemampuan prediktif model. Bahkan ketika ketiga fitur utama dihapus sekaligus mengakibatkan R^2 menurun tajam menjadi 0,2897, hal ini mengindikasikan ketergantungan model yang kuat pada fitur-fitur tersebut.

Sebaliknya, penghapusan fitur-fitur yang kurang berpengaruh, seperti 9, 4, dan 5, justru sedikit meningkatkan kinerja model menjadi R^2 0,5112, menunjukkan bahwa fitur-fitur ini mungkin menambah noise dan penghapusan mereka dapat meningkatkan generalisasi model. Selain itu, eksperimen yang hanya menggunakan tiga fitur paling penting, yaitu 8, 2, dan 3, menunjukkan R^2 sebesar 0,424. Meski hasil ini lebih rendah daripada model optimal, nilai ini tetap menunjukkan kinerja yang cukup baik. Hal ini mengindikasikan bahwa meskipun fitur utama sangat berpengaruh, fitur menengah juga menyumbang informasi tambahan yang dapat meningkatkan akurasi prediksi.

Hasil-hasil tersebut divisualisasikan pada Gambar 5, yang menunjukkan pengaruh penghapusan fitur terhadap nilai R^2 . Pada grafik tersebut, model optimal ditandai sebagai baseline, sehingga perbedaan performa model ablation lainnya dapat dibandingkan secara langsung. Dengan demikian, ablation study ini tidak hanya mengonfirmasi ranking penting fitur dari analisis SHAP, tetapi juga memberikan wawasan praktis mengenai fitur mana yang krusial, fitur mana yang dapat dihapus untuk meningkatkan generalisasi, serta bagaimana kombinasi fitur memengaruhi performa akhir model.



Gambar 5 Hasil Ablation Study pada Model Optimal XGBoost



4. KESIMPULAN

Berdasarkan eksperimen yang telah dilakukan, dapat disimpulkan bahwa baik Random Forest maupun XGBoost efektif dalam memodelkan perkembangan diabetes pada dataset scikit-learn Diabetes, dengan nilai R^2 baseline sekitar 0,42–0,43. Hal ini menunjukkan bahwa kedua model mampu menjelaskan sebagian variasi skor perkembangan penyakit meskipun tanpa optimasi hyperparameter. Setelah diterapkan optimasi menggunakan RandomizedSearchCV dan BayesianSearchCV, performa kedua model meningkat secara konsisten. Pada Random Forest, nilai R^2 meningkat dari 0,4307 menjadi sekitar 0,473, sedangkan pada XGBoost meningkat dari 0,4188 menjadi 0,5018, disertai penurunan MAE, MSE, dan RMSE. Perbandingan kedua teknik optimasi mengindikasikan bahwa BayesianSearchCV lebih efektif dibandingkan RandomizedSearchCV, terutama pada XGBoost, karena mampu menghasilkan peningkatan R^2 yang lebih besar serta mengurangi error secara signifikan. Sementara itu, untuk Random Forest, perbedaan performa antarmetode optimasi relatif kecil, menunjukkan bahwa model ini lebih tahan terhadap variasi hyperparameter.

Selain itu, penerapan Explainable AI menggunakan SHAP berhasil memberikan wawasan yang jelas tentang kontribusi fitur terhadap prediksi. Fitur yang paling berpengaruh adalah kadar serum trigliserida (s5), indeks massa tubuh (BMI), dan tekanan darah (BP). Nilai tinggi pada ketiga fitur ini cenderung menghasilkan nilai SHAP positif yang meningkatkan prediksi progresi diabetes, sedangkan nilai rendah cenderung menghasilkan nilai SHAP negatif. Pola kontribusi ini sejalan dengan pengetahuan klinis, sehingga memperkuat kepercayaan terhadap model serta interpretabilitas prediksi yang dihasilkan.

DAFTAR PUSTAKA

- Agrawal, R., Gupta, T., Gupta, S., Chauhan, S., Patel, P., & Hamdare, S. (2025). Fostering trust and interpretability: integrating explainable AI (XAI) with machine learning for enhanced disease prediction and decision transparency. *Diagnostic Pathology*, 20(1), 105. <https://doi.org/10.1186/s13000-025-01686-3>
- Frasca, M., La Torre, D., Pravettoni, G., & Cutica, I. (2024). Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discover Artificial Intelligence*, 4(1), 15. <https://doi.org/10.1007/s44163-024-00114-7>
- Gao, J., Ren, J., & Wen, Z. (2025). Research on Diabetes Prediction Based on the Randomized Search CV Method. *2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS)*, 1–4. <https://doi.org/10.1109/EEISS65394.2025.11086023>
- Gill, M., Anderson, R., Hu, H., Bennamoun, M., Peterreit, J., Valliyodan, B., Nguyen, H. T., Batley, J., Bayer, P. E., & Edwards, D. (2022). Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction. *BMC Plant Biology*, 22(1), 180. <https://doi.org/10.1186/s12870-022-03559-z>
- Gupta, A., Stead, T. S., & Ganti, L. (2024). Determining a Meaningful R-squared Value in Clinical Medicine. *Academic Medicine & Surgery*. <https://doi.org/10.62186/001c.125154>
- Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291–300. <https://doi.org/10.2174/1389202922666210705124359>
- Havelda, L., Szalai, E. Á., Obeidat, M., Dobszai, D., Veres, D. S., Kóí, T., Sipter, E., Vánicsa, S., Hegyi, P. J., Bucur, M., Molnár, A., Vámosy, K. L., Hegyi, P., & Szentesi, A. (2025). Hypertriglyceridemia is a dose-dependent risk factor for type 2 diabetes mellitus: a systematic review and meta-analysis. *Frontiers in Endocrinology, Volume 16-2025*. <https://doi.org/10.3389/fendo.2025.1710007>
- Hossain, M. J., Al-Mamun, M., & Islam, M. R. (2024). Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused. *Health Science Reports*, 7(3), e2004. <https://doi.org/10.1002/hsr2.2004>
- Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*, 13(3). <https://doi.org/10.3390/technologies13030088>
- Imans, D., Abuhmed, T., Alharbi, M., & El-Sappagh, S. (2024). Explainable Multi-Layer Dynamic Ensemble Framework Optimized for Depression Detection and Severity Assessment.



- Diagnostics*, 14(21). <https://doi.org/10.3390/diagnostics14212385>
- Karunakaran, C., Niranjana, V., & Setlur, A. S. (2025). Random Forest and XGBoost-based ensemble models for colorectal cancer exome variant classification and web application deployment for early prediction. *Computational and Structural Biotechnology Reports*, 2, 100063. <https://doi.org/https://doi.org/10.1016/j.csbr.2025.100063>
- Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 244, 122778. <https://doi.org/https://doi.org/10.1016/j.eswa.2023.122778>
- Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023). Ensemble Learning for Disease Prediction: A Review. *Healthcare*, 11(12). <https://doi.org/10.3390/healthcare11121808>
- Mienye, I. D., & Jere, N. (2024). Optimized Ensemble Learning Approach with Explainable AI for Improved Heart Disease Prediction. *Information*, 15(7). <https://doi.org/10.3390/info15070394>
- Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2), 757-774. <https://doi.org/https://doi.org/10.1016/j.jksuci.2023.01.014>
- Ong, K. L., Stafford, L. K., McLaughlin, S. A., Boyko, E. J., Vollset, S. E., Smith, A. E., Dalton, B. E., Duprey, J., Cruz, J. A., Hagins, H., Lindstedt, P. A., Aali, A., Abate, Y. H., Abate, M. D., Abbasian, M., Abbasi-Kangevari, Z., Abbasi-Kangevari, M., ElHafeez, S. A., Abd-Rabu, R., ... Vos, T. (2023). Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 402(10397), 203 - 234. [https://doi.org/10.1016/S0140-6736\(23\)01301-6](https://doi.org/10.1016/S0140-6736(23)01301-6)
- Peters, R., Ee, N., Peters, J., Beckett, N., Booth, A., Rockwood, K., & Anstey, K. J. (2019). Common risk factors for major noncommunicable disease, a systematic overview of reviews and commentary: the implied potential for targeted risk reduction. *Therapeutic Advances in Chronic Disease*, 10, 2040622319880392. <https://doi.org/10.1177/2040622319880392>
- Santos, M. R., Guedes, A., & Sanchez-Gendriz, I. (2024). SHapley Additive exPlanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis. *Machine Learning and Knowledge Extraction*, 6(1), 316-341. <https://doi.org/10.3390/make6010016>
- Ukoba, O., Joseph, U. O., Charles, O. L., Opirite, P.-K. B., Omamuyovwi, A. E., & Matilda, O.-N. (2025). A systematic review of machine learning methods for diabetes mellitus prediction and classification in Nigeria. *International Journal of Community Medicine and Public Health*, 12(6), 2828-2835. <https://doi.org/10.18203/2394-6040.ijcmph20251734>
- Wahidin, M., Achadi, A., Besral, B., Kosen, S., Nadjib, M., Nurwahyuni, A., Ronoatmodjo, S., Rahajeng, E., Pane, M., & Kusuma, D. (2024). Projection of diabetes morbidity and mortality till 2045 in Indonesia based on risk factors and NCD prevention and control programs. *Scientific Reports*, 14(1), 5424. <https://doi.org/10.1038/s41598-024-54563-2>
- Wondmkun, Y. T. (2020). Obesity, Insulin Resistance, and Type 2 Diabetes: Associations and Therapeutic Implications. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 13, 3611-3616. <https://doi.org/10.2147/DMSO.S275898>
- Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, 17(1), 26-40. <https://doi.org/https://doi.org/10.11989/JEST.1674-862X.80904120>
- Yan, D., Li, X., Wang, Y., & Cai, Z. (2025). Optimized prediction of diabetes complications using ensemble learning with Bayesian optimization: a cost-efficient laboratory-based approach. *Frontiers in Endocrinology, Volume 16-2025*. <https://doi.org/10.3389/fendo.2025.1593068>
- Zhang, H., Ni, J., Yu, C., Wu, Y., Li, J., Liu, J., Tu, J., Ning, X., He, Q., & Wang, J. (2019). Sex-Based Differences in Diabetes Prevalence and Risk Factors: A Population-Based Cross-Sectional Study Among Low-Income Adults in China. *Frontiers in Endocrinology*, 10, 658. <https://doi.org/10.3389/fendo.2019.00658>
- Zhong, H., Zhang, H., & Jia, F. (2020). A computing method of predictive value based on fitting function in linear model. *EAI Endorsed Transactions on Collaborative Computing*, 4(14). <https://doi.org/10.4108/eai.2-10-2020.166542>

