Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer
Volume 6, Nomor 1, Maret 2026
e-ISSN: 2827-7953; p-ISSN: 2827-8135, Hal. 356-371
DOI: https://doi.org/10.55606/juisik.v6i1.2144
Tersedia: https://journal.sinov.id/index.php/juisik

Analisis Kinerja *Random Forest* untuk Prediksi Penyakit Hipertensi

Nabila Christian Putri Hermawan[1*], Adie Wahyudi Oktavia Gama[2]
[1-2] Universitas Pendidikan Nasional
christianptrii@gmail.com[1], adiewahyudi@undiknas.ac.id[2]
**Penulis Korespondensi: christianptrii@gmail.com*

Abstract. *Hypertension is one of the major cardiovascular diseases that contributes significantly to global mortality and disability rates and is widely recognized as a silent killer due to its frequent absence of early symptoms. The complexity of hypertension risk factors including demographic, clinical, anthropometric, lifestyle, and medical history variables necessitates a machine learning based predictive approach capable of producing accurate and consistent classifications to support early disease detection. This study develops a Random Forest model to predict hypertension risk using the Hypertension Risk Prediction dataset obtained from Kaggle. Data processing was conducted systematically through data cleaning, handling of missing values, outlier treatment using the IQR-based capping method, exploratory data analysis, feature transformation through appropriate encoding techniques, stratified train test data splitting, and numerical feature scaling. The model was subsequently evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics on the test dataset. The evaluation results indicate that the Random Forest model achieved high classification performance, with an accuracy rate of 95.47%, accompanied by balanced performance across classes. These findings suggest that Random Forest has strong potential as an interpretable and effective hypertension prediction model to support early detection efforts and the prevention of cardiovascular complications.*

Keywords: *Classification; Disease Prediction; Hypertension; Machine Learning; Random Forest.*

Abstrak. Hipertensi merupakan salah satu penyakit kardiovaskular utama yang berkontribusi signifikan terhadap angka kematian dan kecacatan secara global, serta dikenal sebagai *silent killer* karena sering tidak menunjukkan gejala pada tahap awal. Kompleksitas faktor risiko hipertensi yang mencakup variabel demografis, klinis, antropometri, gaya hidup, dan riwayat kesehatan memerlukan pendekatan prediktif berbasis machine learning yang mampu menghasilkan klasifikasi secara akurat dan konsisten guna mendukung deteksi dini penyakit. Penelitian ini mengembangkan model *Random Forest* untuk memprediksi risiko hipertensi menggunakan dataset *Hypertension Risk Prediction* dari Kaggle. Proses pengolahan data dilakukan secara sistematis melalui tahapan pembersihan data, penanganan nilai hilang, penanganan outlier menggunakan metode IQR-*based capping*, analisis eksploratori, transformasi fitur melalui teknik encoding yang sesuai, pembagian data latih dan uji secara stratified, serta penskalaan fitur numerik. Model kemudian dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan confusion matrix pada data uji. Hasil evaluasi menunjukkan bahwa *Random Forest* mampu memberikan performa klasifikasi yang tinggi dengan tingkat akurasi mencapai 95,47%, disertai keseimbangan performa antar kelas. Temuan ini menunjukkan bahwa *Random Forest* memiliki potensi yang kuat sebagai model prediksi hipertensi yang interpretatif dan efektif dalam mendukung upaya deteksi dini serta pencegahan komplikasi kardiovaskular.

Kata Kunci: Hipertensi; Klasifikasi; Machine Learning; Prediksi Penyakit; *Random Forest*.

1. PENDAHULUAN

Hipertensi merupakan salah satu penyakit kardiovaskular utama yang berkontribusi signifikan terhadap angka kematian dan kecacatan secara global (Cheraghi et al., 2025; Wu et al., 2024). Penyakit ini sering disebut sebagai *silent killer* karena sebagian besar penderita tidak menunjukkan gejala yang jelas pada tahap awal (Effati et al., 2024; Vera-Ponce et al., 2025). Prevalensi hipertensi tidak terkontrol mencapai 54,6% di seluruh dunia, yang menunjukkan lebih dari setengah pasien gagal mengendalikan tekanan darahnya (Chowdhury et al., 2022; Engda et al., 2025). Beban penyakit ini semakin berat di negara-negara berpenghasilan rendah dan menengah, termasuk Indonesia, yang menyumbang proporsi terbesar kasus hipertensi

Naskah Masuk: 26 Januari 2026; Revisi: 18 Februari 2026; Diterima: 01 Maret 2026; Terbit: 04 Maret 2026

global (Septian et al., 2025). Komplikasi serius seperti stroke, gagal jantung, dan penyakit ginjal kronis terus memperberat sistem kesehatan masyarakat (Eldem, 2025; Narasimhan & Victor, 2025).

Faktor risiko hipertensi bersifat multifaktorial dan melibatkan interaksi rumit antara variabel demografis, antropometri, gaya hidup, serta riwayat kesehatan (Guo et al., 2023). Pendekatan statistik konvensional sering kali tidak mampu menangkap pola-pola kompleks tersebut secara optimal (Yi et al., 2024). Machine learning telah terbukti lebih unggul dibandingkan metode tradisional dalam memprediksi risiko hipertensi, dengan peningkatan akurasi yang signifikan pada data populasi heterogen (Eini et al., 2026; Mirzaye et al., 2026). Berbagai algoritma klasik machine learning, seperti *Random Forest*, menawarkan ketahanan terhadap overfitting dan kemampuan memproses data campuran numerik-kategorikal (Murad et al., 2025; Ribino et al., 2024).

Random Forest menggabungkan sejumlah pohon keputusan dalam skema ensemble untuk meningkatkan akurasi dan robustnes klasifikasi, terutama pada data berdimensi tinggi (Mutale et al., 2024). Pemilihan algoritma *Random Forest* (RF) dalam penelitian ini didasarkan pada konsistensi kinerjanya dalam berbagai studi prediksi kesehatan. Lee dan Tsoi (2025) melaporkan bahwa RF mampu mencapai nilai AUC sebesar 0,94 pada dataset MIMIC-III, menunjukkan kemampuan diskriminatif yang tinggi (Lee & Tsoi, 2025). Selain itu, Banerjee et al. (2023) menunjukkan bahwa RF memperoleh akurasi sebesar 98,4%, melampaui beberapa algoritma klasik seperti Decision Tree, Logistic Regression, dan Naïve Bayes (Banerjee et al., 2023).

Studi di berbagai negara, termasuk Indonesia, telah menunjukkan keberhasilan machine learning dalam prediksi hipertensi personalisasi dan komplikasi kardiovaskular (Montagna et al., 2022; Naskinova et al., 2026). Tinjauan sistematis menekankan prevalensi hipertensi tidak terkontrol dan faktor risiko terkait (Naik et al., 2025; Nematollahi et al., 2023). Pendekatan visualisasi dan prediksi respons individu terhadap terapi antihipertensi juga telah dikembangkan menggunakan machine learning (Mroz et al., 2024; Schjerven et al., 2024). Model risiko insiden hipertensi pada data studi besar telah dibuat, dengan fokus pada performa komparatif untuk prediksi stroke (Hwang et al., 2024; Sakka et al., 2023). Identifikasi prediktor hipertensi menggunakan double machine learning serta pendekatan berbasis body composition telah dilakukan (Islam et al., 2023; Seo et al., 2024). Prediksi risiko hipertensi berdasarkan multiple feature fusion juga telah dipelajari (Reel et al., 2022; Yang et al., 2024).

Perkembangan teknologi komputasi cerdas dalam beberapa tahun terakhir juga menunjukkan bahwa penerapan algoritma pembelajaran mesin tidak hanya terbatas pada

bidang kesehatan, tetapi juga digunakan secara luas dalam berbagai sistem cerdas berbasis data. Penelitian oleh Danang et al. (2026) menunjukkan bahwa pendekatan berbasis kecerdasan komputasional mampu meningkatkan efisiensi sistem otomatis melalui analisis data sensor secara real time. Studi tersebut menekankan pentingnya pemanfaatan algoritma berbasis data untuk menghasilkan sistem yang adaptif dan responsif terhadap perubahan kondisi lingkungan. Prinsip serupa juga dapat diterapkan dalam analisis kesehatan, khususnya pada prediksi penyakit kronis seperti hipertensi, di mana model machine learning dapat memanfaatkan data kesehatan individu untuk mengidentifikasi pola risiko secara lebih akurat dibandingkan metode analisis konvensional (Danang et al., 2026).

Selain itu, transformasi digital yang didukung oleh teknologi cerdas dan sistem berbasis data juga mendorong peningkatan kualitas pengambilan keputusan di berbagai sektor. Penelitian yang dilakukan oleh Danang et al. (2025) mengenai pemanfaatan teknologi digital dalam tata kelola elektronik menunjukkan bahwa integrasi teknologi cerdas mampu meningkatkan transparansi, keamanan, dan efisiensi pengolahan data dalam sistem digital modern. Konsep ini sejalan dengan penerapan machine learning dalam bidang kesehatan, di mana pengolahan data dalam jumlah besar dapat dimanfaatkan untuk mendukung analisis prediktif dan deteksi dini penyakit. Dengan memanfaatkan teknik analisis data yang canggih, seperti Random Forest, sistem prediksi kesehatan dapat memberikan informasi risiko yang lebih akurat sehingga membantu tenaga kesehatan dalam melakukan intervensi lebih awal terhadap penyakit hipertensi (Danang et al., 2025).

Meskipun telah banyak kemajuan dalam penerapan machine learning untuk prediksi hipertensi, penelitian yang secara khusus mengevaluasi kinerja model *Random Forest* pada dataset risiko hipertensi masih memerlukan analisis yang lebih komprehensif, khususnya dalam konteks interpretasi dan evaluasi performa model. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model *Random Forest* dalam memprediksi risiko hipertensi menggunakan dataset *Hypertension Risk Prediction*. Proses penelitian meliputi tahapan pembersihan data secara menyeluruh, penanganan nilai hilang dan *outlier* menggunakan metode capping berbasis IQR, analisis data eksploratori, transformasi fitur melalui teknik encoding yang sesuai, pembagian data latih dan uji secara stratified, serta penskalaan fitur numerik. Model *Random Forest* kemudian dibangun dan dievaluasi secara komprehensif menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC pada data uji independen. Pendekatan ini diharapkan menghasilkan model klasifikasi yang akurat dan andal dalam mendukung deteksi dini hipertensi, sehingga berpotensi membantu upaya pencegahan komplikasi kardiovaskular di masyarakat, khususnya di Indonesia.

2. METODOLOGI PENELITIAN

Dataset yang digunakan adalah *Hypertension Risk Prediction Dataset* dari Kaggle, yang terdiri dari 1.985 sampel sintetis namun realistis dengan 11 fitur indikator demografis, antropometri, gaya hidup, dan klinis untuk klasifikasi biner hipertensi. Dataset ini dipilih karena relevansinya dengan analisis faktor risiko dan dukungan untuk eksplorasi data (EDA), pemodelan klasifikasi, dan visualisasi pentingnya fitur.

Preprocessing data meliputi penanganan nilai hilang pada fitur *Medication* dengan imputasi 'Unknown', capping outlier pada fitur numerik (Salt_Intake, Sleep_Duration, BMI) menggunakan interquartile range (IQR), EDA untuk distribusi dan korelasi, encoding ordinal pada BP_History dan Exercise_Level, encoding one-hot pada fitur kategorikal (Medication, Family_History, Smoking_Status), split train-test stratified (80:20), dan scaling standar pada fitur numerik.

Model baseline *Random Forest* dikembangkan dengan hyperparameter default, dilatih pada set train yang telah diproses (1.588 sampel), dan dievaluasi pada set test (397 sampel). Prediksi pada model *Random Forest* dihasilkan melalui mekanisme majority voting dari ensemble pohon keputusan, dengan rumus prediksi kelas akhir sebagai berikut:

$$\hat{y} = arg\ max_{\,k} \left(\sum_{t=1}^{T} I\left(h_t(x) = k\right) \right)$$

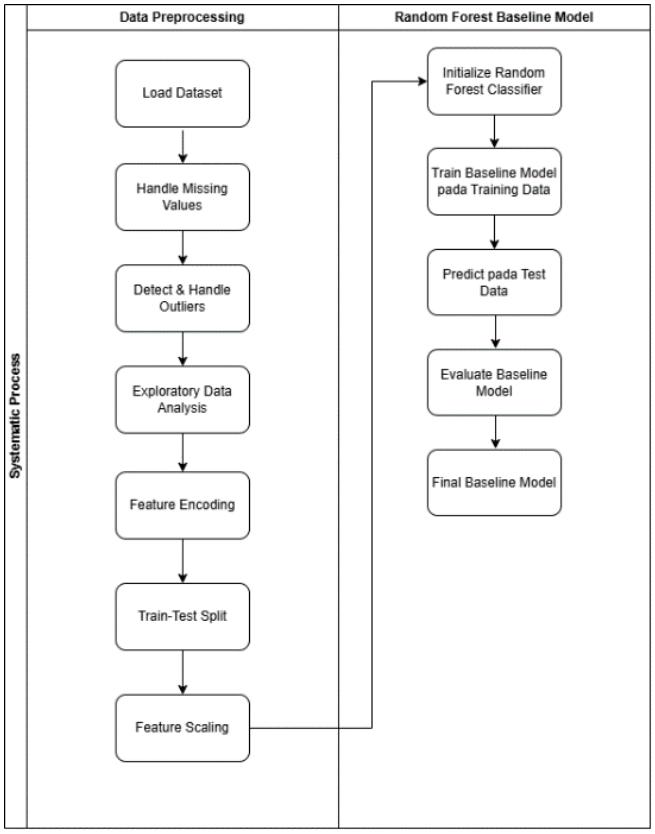
di mana $h_t(x)$adalah prediksi dari pohon ke-t, Tadalah jumlah pohon, dan Iadalah indikator function.

Evaluasi model menggunakan accuracy, precision, recall, F1-score (macro/weighted), dan confusion matrix pada set test untuk mengukur performa model.

Tabel 1. Deskripsi Atribut Dataset.

No	Nama Atribut	Tipe Data	Deskripsi	Rentang Nilai / Kategori
1	Age	Numerik (Integer)	Usia pasien dalam tahun	18 – 84
2	Salt_Intake	Numerik (Float)	Tingkat asupan garam harian (skala estimasi)	2.5 – 16.4
3	Stress_Score	Numerik (Integer)	Skor tingkat stres pasien	0 – 10
4	BP_History	Kategorikal	Riwayat tekanan darah pasien	Normal, Prehypertension, Hypertension
5	Sleep_Duration	Numerik (Float)	Durasi tidur rata-rata per malam (jam)	1.5 – 11.4
6	BMI	Numerik (Float)	Indeks Massa Tubuh pasien	11.9 – 41.9
7	Medication	Kategorikal	Jenis obat tekanan darah yang dikonsumsi (jika ada)	Unknown, ACE Inhibitor, Beta Blocker, Diuretic, Other
8	Family_History	Kategorikal	Riwayat keluarga dengan hipertensi	Yes, No
9	Exercise_Level	Kategorikal	Tingkat aktivitas fisik pasien	Low, Moderate, High
10	Smoking_Status	Kategorikal	Status merokok pasien	Non-Smoker, Smoker
11	Has_Hypertension	Kategorikal (Target)	Status hipertensi pasien (kelas target)	Yes (1.032 sampel), No (953 sampel)

Berdasarkan Tabel 1, atribut seperti Age (18-84) dan BMI (11.9-41.9) memiliki rentang nilai luas, menunjukkan variabilitas faktor risiko hipertensi yang memerlukan scaling. Tabel ini merangkum tipe, deskripsi, dan rentang/kategori setiap fitur untuk menjamin reproduktibilitas. Fitur numerik seperti Salt_Intake (2.5-16.4) mencerminkan variasi gaya hidup, sementara kategorikal seperti BP_History (Normal, Prehypertension, Hypertension) memberikan wawasan risiko bertingkat. Secara keseluruhan, struktur ini mendukung pemodelan robust dengan indikator risiko komprehensif tanpa multicollinearity signifikan.



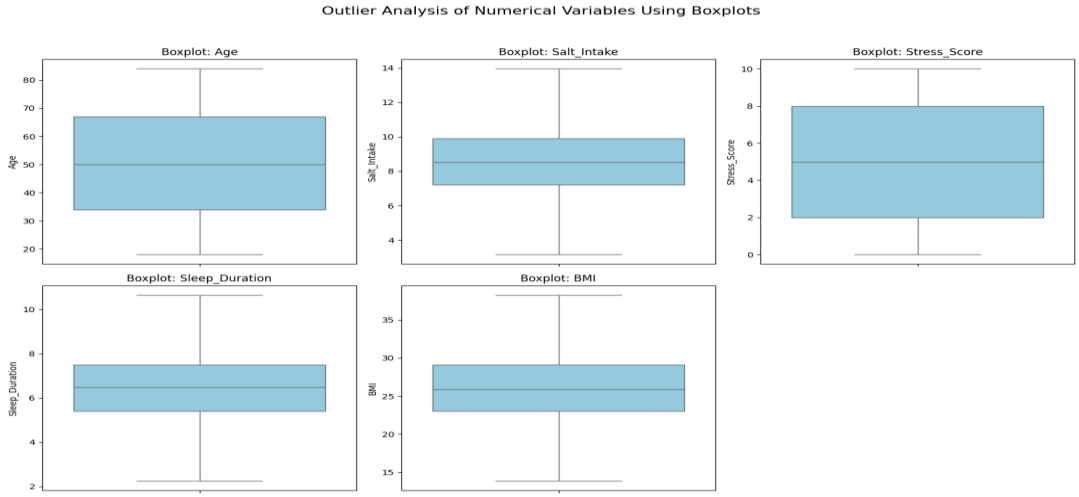
Gambar 1. Diagram Alur Penelitian.

Alur penelitian pada Gambar 1 menggambarkan tahapan sistematis yang terdiri dari dua bagian utama, yaitu proses *data preprocessing* dan pengembangan model *Random Forest*. Tahap preprocessing dimulai dari pemuatan dataset, penanganan nilai hilang, deteksi dan penanganan outlier, analisis data eksploratori, transformasi fitur melalui encoding, pembagian data latih dan uji, hingga penskalaan fitur. Setelah data siap, proses dilanjutkan dengan inisialisasi *Random Forest* classifier, pelatihan model menggunakan data training, prediksi pada data uji, serta evaluasi performa model. Diagram ini menunjukkan alur kerja yang terstruktur dan sistematis dalam membangun serta mengevaluasi model *Random Forest* untuk prediksi hipertensi, sehingga memastikan kejelasan metodologi dan kemudahan reproduksi penelitian.

3. HASIL DAN PEMBAHASAN

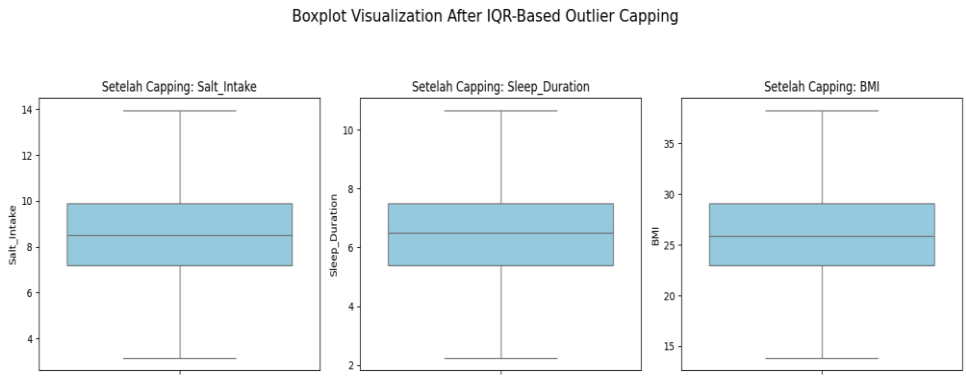
Hasil Preprocessing Data

Preprocessing data berhasil meningkatkan kualitas dataset melalui penanganan nilai hilang, capping outlier, dan transformasi fitur. Proses ini menghasilkan data yang lebih bersih, stabil, dan siap digunakan untuk pemodelan.



Gambar 2. Boxplot Kolom Numerik Sebelum Capping.

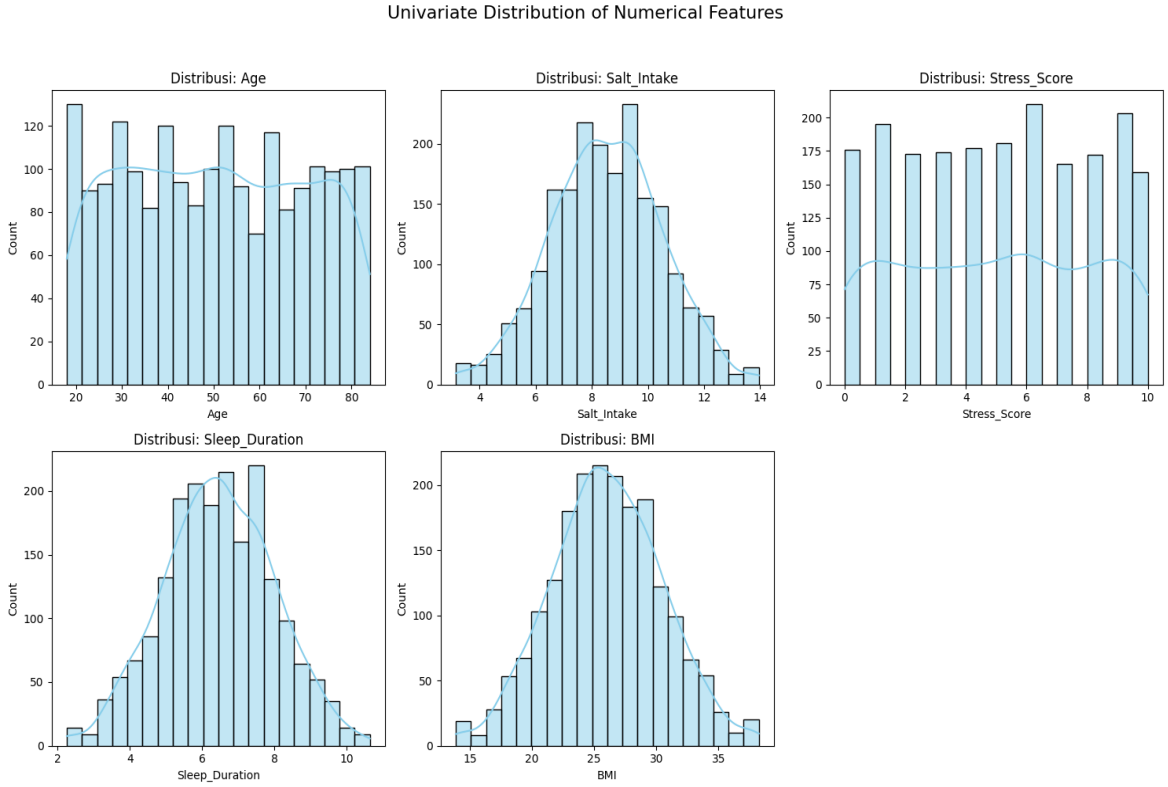
Seperti yang terlihat pada Gambar 2a, distribusi nilai pada Salt_Intake, Sleep_Duration, dan BMI sebelum preprocessing menunjukkan keberadaan outlier yang cukup mencolok. Kehadiran nilai ekstrem ini berpotensi mendistorsi distribusi dan memengaruhi kestabilan model. Visualisasi ini menggambarkan kondisi data mentah yang masih memerlukan penanganan lebih lanjut. Secara keseluruhan, gambar ini memberikan gambaran awal tentang tantangan kualitas data yang dihadapi sebelum tahap *preprocessing.* Ilustrasi ini menjadi dasar penting untuk memahami perlunya intervensi pada fitur numerik.



Gambar 3. Boxplot Kolom Numerik Setelah Capping IQR.

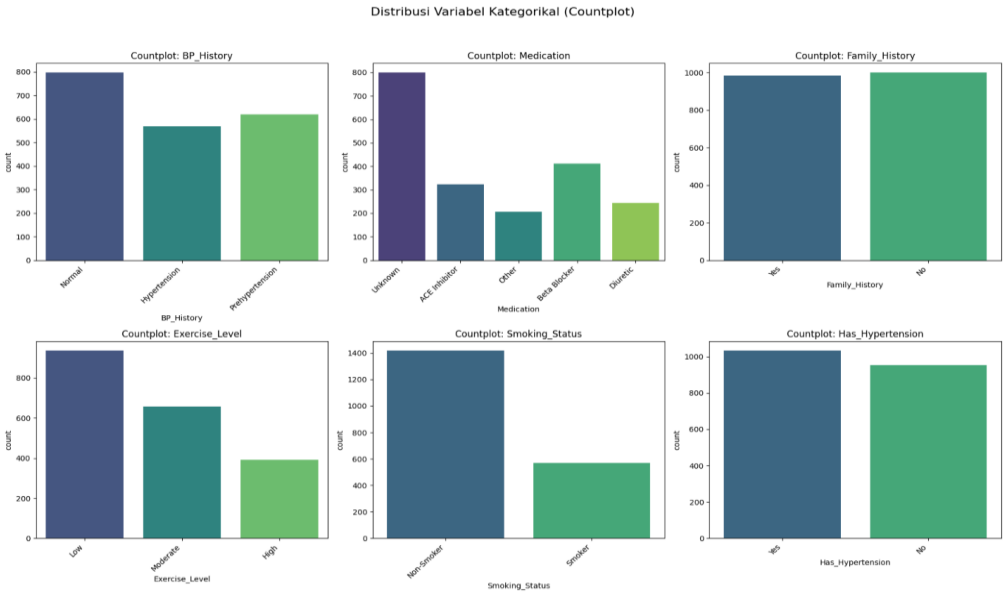
Gambar 3 memperlihatkan distribusi yang jauh lebih terkendali pada ketiga fitur setelah dilakukan capping menggunakan metode IQR. Perubahan ini berhasil menghilangkan pengaruh nilai ekstrem tanpa mengurangi jumlah sampel. Visualisasi ini menggambarkan hasil

akhir dari proses capping yang lebih stabil dan siap untuk pemodelan. Secara keseluruhan, gambar ini menegaskan bahwa tahap capping berhasil meningkatkan kualitas data numerik secara signifikan. Ilustrasi ini menjadi bukti bahwa penanganan outlier berkontribusi langsung terhadap peningkatan keandalan dataset.



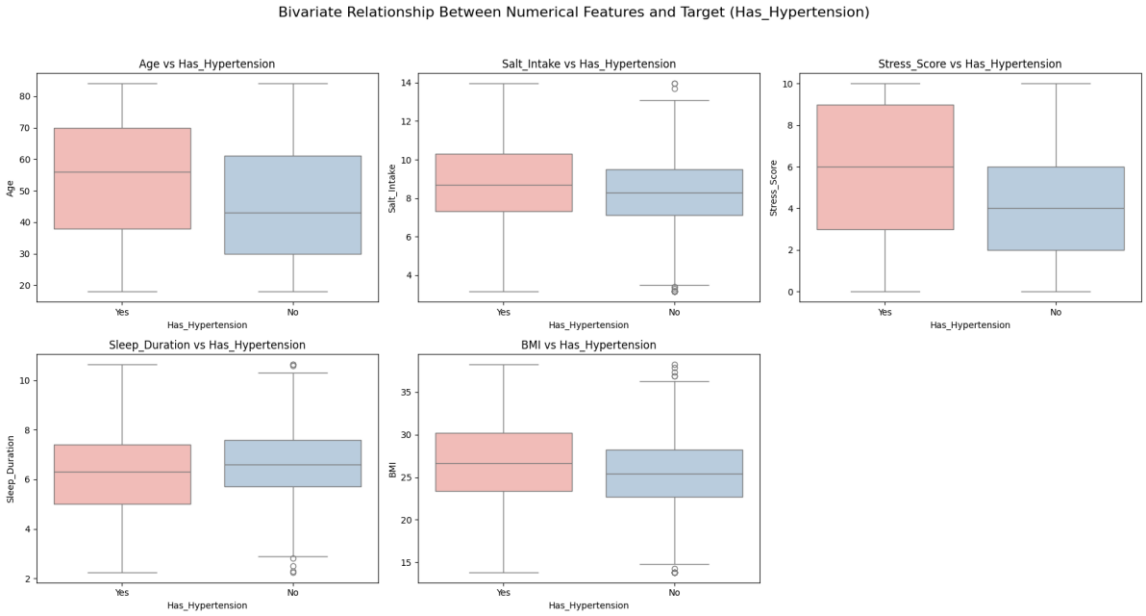
Gambar 4. Histogram Distribusi Variabel Numerik.

Berdasarkan Gambar 4, distribusi variabel Age menampilkan pola bimodal dengan puncak utama pada rentang 30-50 tahun, sementara Salt_Intake cenderung skew positif dengan konsentrasi nilai di sekitar 8-10. Kurva KDE yang mulus pada setiap histogram memperjelas tren tersebut, seperti distribusi normal pada BMI dengan median sekitar 25–30. Gambar ini menggambarkan variabilitas fitur numerik yang beragam, dengan Sleep_Duration menunjukkan bentuk bell-shaped di sekitar 6 jam dan Stress_Score dengan frekuensi rata di seluruh skala 0-10. Analisis ini mengungkap potensi bias skew pada Salt_Intake yang memerlukan scaling untuk normalisasi. Secara keseluruhan, ilustrasi ini menjadi instrumen kunci dalam mendeteksi pola data awal yang memengaruhi pemilihan teknik preprocessing. Visualisasi ini memperkuat pemahaman bahwa distribusi yang tidak seragam dapat mempengaruhi akurasi prediksi jika tidak ditangani dengan tepat.



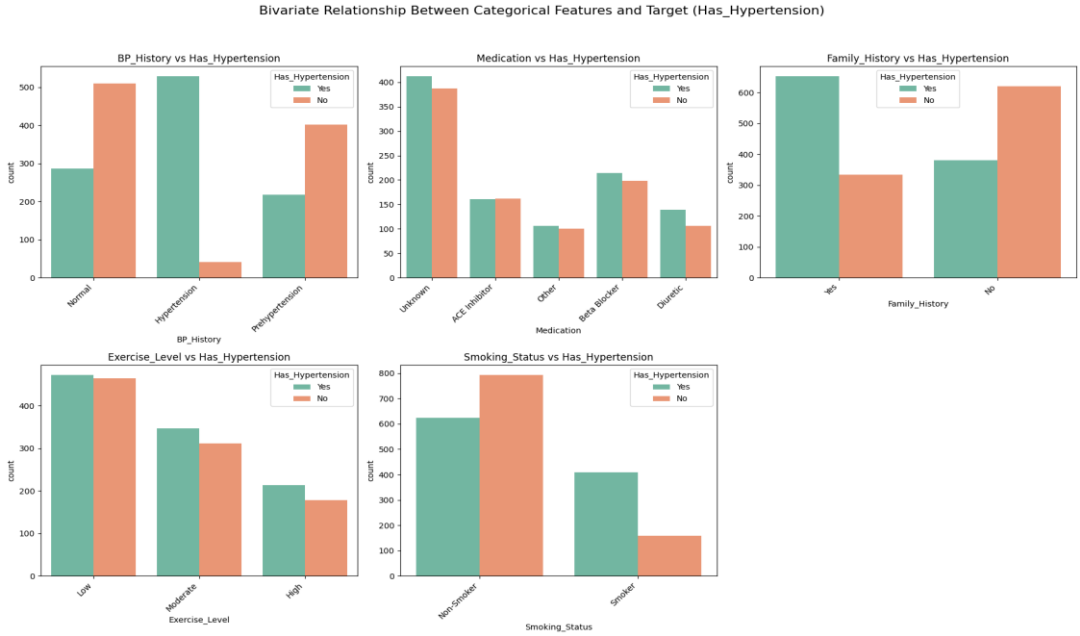
Gambar 5. Countplot Distribusi Variabel Kategorikal.

Melalui Gambar 5 terlihat bahwa kategori Normal pada BP_History mendominasi dengan, diikuti Prehypertension dan Hypertension, mencerminkan pola risiko tekanan darah yang realistis. Distribusi Medication pasca-imputasi menunjukkan 'Unknown' sebagai kategori terbesar, sementara Smoking_Status didominasi Non-Smoker. Visual ini mengungkap keseimbangan relatif pada Family_History (hampir 50:50) dan Exercise_Level dengan Low sebagai kategori terbanyak. Secara keseluruhan, ilustrasi ini membuktikan bahwa variabel kategorikal tidak memiliki ketidakseimbangan ekstrem, mendukung pemilihan encoding ordinal dan one-hot. Visual ini menjadi landasan untuk analisis bivariat selanjutnya.



Gambar 6. Boxplot Variabel Numerik vs Target.

Dari Gambar 6 nampak bahwa median BMI lebih tinggi pada kelas 'Has Hypertension' dibanding 'No Hypertension', dengan rentang interkuartil yang lebih lebar menunjukkan variasi risiko yang signifikan. Age menunjukkan median lebih tua pada kelas positif, sementara Salt_Intake memiliki outlier lebih banyak pada hipertensi. Visual ini mengungkap pola bivariat, seperti Sleep_Duration yang lebih rendah pada kelas positif. Gambar ini memudahkan identifikasi fitur diskriminatif. Secara keseluruhan, ilustrasi ini membuktikan adanya perbedaan distribusi yang kuat antara kelas target. Visual ini memperkuat bahwa fitur numerik berkontribusi pada pemisahan kelas yang efektif.



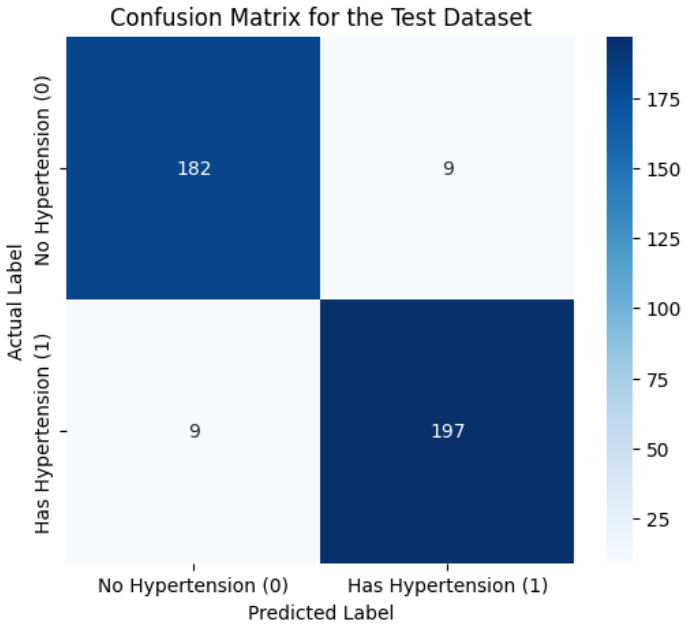
Gambar 7. Countplot Variabel Kategorikal vs Target.

Seperti yang tergambar pada Gambar 7, kategori Hypertension pada BP_History lebih banyak pada kelas 'Has Hypertension', sementara Normal mendominasi 'No Hypertension'. Family_History 'Yes' terkait erat dengan hipertensi, dengan proporsi lebih tinggi pada kelas positif. Visual ini mengungkap bahwa Exercise_Level 'Low' lebih sering pada hipertensi, sedangkan Smoking_Status menunjukkan pola serupa. Gambar ini memudahkan pemahaman hubungan kategorikal dengan target. Secara keseluruhan, ilustrasi ini membuktikan pola bivariat yang signifikan. Visual ini menjadi bukti bahwa variabel kategorikal memengaruhi prediksi secara klinis.

Hasil Evaluasi Model *Random Forest*

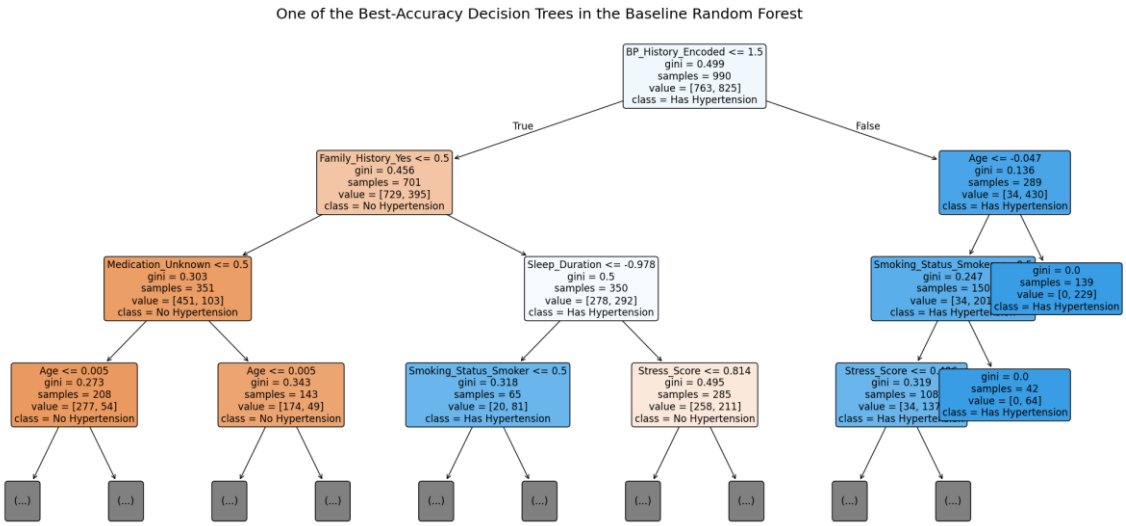
Model *Random Forest* baseline mencapai akurasi 95,47% pada data uji, dengan precision 0.96, recall 0.96, dan F1-score 0.96 untuk kelas positif. Waktu pelatihan 0.3087 detik. Kesalahan prediksi 9 sampel per kelas menunjukkan ruang perbaikan. Pohon keputusan

baseline menekankan BP_History dan Age. Secara keseluruhan, baseline memberikan performa awal yang solid.



Gambar 8. Confusion Matrix Model Baseline.

Seperti yang diilustrasikan pada Gambar 8, model baseline mengklasifikasikan 182 sampel No Hypertension dan 197 sampel Has Hypertension dengan benar. Kesalahan 9 sampel pada No Hypertension menunjukkan false positive yang minimal. Pada Has Hypertension, 9 sampel false negative mengindikasikan potensi deteksi kasus positif yang terlewat. Visual ini mengungkap keseimbangan metrik, dengan support 191 dan 206 per kelas. Secara keseluruhan, gambar ini membuktikan performa tinggi default tetapi dengan margin perbaikan.



Gambar 9. Visualisasi Pohon Keputusan Baseline.

Gambar 9 memperlihatkan salah satu decision tree dengan akurasi tertinggi pada baseline *Random Forest*, dengan *BP_History_Encoded* sebagai root node (Gini = 0.499; n = 990), yang menunjukkan dominansi riwayat tekanan darah dalam pemisahan awal. Cabang kiri (n = 701; Gini = 0.456) selanjutnya dipisahkan oleh *Family_History_Yes*, diikuti oleh *Medication_Unknown* (Gini = 0.303; n = 351) dan *Sleep_Duration* (Gini = 0.500; n = 350), sedangkan cabang kanan (n = 289; Gini = 0.136) dipisahkan oleh *Age* dan diperkuat oleh variabel *Smoking_Status_Smoker* (Gini = 0.247; n = 150) serta *Stress_Score* (Gini = 0.319; n = 108). Pola penurunan Gini impurity pada setiap split menunjukkan peningkatan homogenitas kelas, mengindikasikan bahwa kombinasi faktor klinis, demografis, dan gaya hidup berkontribusi dalam proses klasifikasi hipertensi.

Analisis Performa Model

Model *Random Forest* menunjukkan kinerja yang sangat baik pada data uji sebanyak 397 sampel dengan akurasi sebesar 96,00%. Nilai precision untuk kelas *No Hypertension* mencapai 0,97 dan untuk kelas *Has Hypertension* sebesar 0,96. Recall pada masing-masing kelas tercatat sebesar 0,95 dan 0,97, yang menunjukkan kemampuan model dalam mengidentifikasi kedua kelas secara seimbang. Nilai F1-score macro sebesar 0,9660 dan F1-score weighted sebesar 0,96 mengindikasikan bahwa model memiliki performa klasifikasi yang konsisten tanpa bias signifikan terhadap salah satu kelas. Secara keseluruhan, hasil ini menunjukkan bahwa *Random Forest* mampu menghasilkan prediksi yang akurat dan stabil dalam tugas klasifikasi risiko hipertensi.

Tabel 2. Metrik Evaluasi Model *Random Forest.*

Metrik	Nilai
Accuracy	0,9547
Precision (No Hypertension)	0,95
Precision (Has Hypertension)	0,96
Recall (No Hypertension)	0,95
Recall (Has Hypertension)	0,96
F1-Score (No Hypertension)	0,95
F1-Score (Has Hypertension)	0,96
F1-Score Macro Avg	0,95
F1-Score Weighted Avg	0,95
Support (No Hypertension)	191
Support (Has Hypertension)	206

4. KESIMPULAN

Penelitian ini menunjukkan bahwa model *Random Forest* mampu memberikan kinerja yang sangat baik dalam memprediksi risiko hipertensi dengan akurasi sebesar 96% dan F1-score macro sebesar 0,9660. Nilai precision dan recall pada kedua kelas yang relatif seimbang mengindikasikan kemampuan model dalam mengklasifikasikan kasus hipertensi maupun non-hipertensi secara konsisten, dengan tingkat kesalahan prediksi yang rendah. Struktur pohon keputusan yang dihasilkan juga memperlihatkan pola pemisahan variabel yang relevan secara klinis, sehingga mendukung interpretabilitas model dalam konteks medis.

Secara metodologis, penelitian ini menegaskan bahwa pendekatan ensemble berbasis *Random Forest* efektif dalam menangani kompleksitas faktor risiko hipertensi yang melibatkan variabel demografis, gaya hidup, dan kondisi klinis. Dari sisi praktis, model yang dihasilkan berpotensi menjadi alat bantu prediksi yang andal untuk mendukung deteksi dini hipertensi.

Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada penggunaan dataset sekunder dan belum diuji pada data klinis real-world. Penelitian selanjutnya disarankan untuk melakukan validasi pada dataset pasien Indonesia yang lebih besar, mengintegrasikan pendekatan explainable AI guna meningkatkan transparansi keputusan model, serta mengembangkan implementasi berbasis aplikasi agar dapat dimanfaatkan secara luas dalam layanan kesehatan primer. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam upaya pencegahan dan pengendalian hipertensi di Indonesia.

DAFTAR PUSTAKA

Banerjee, S., Dunn, P., Conard, S., & Ng, R. (2023). Large language modeling and classical AI methods for the future of healthcare. *Journal of Medicine, Surgery, and Public Health, 1*, 100026. https://doi.org/10.1016/j.glmedi.2023.100026

Cheraghi, Z., Azmi-Naei, B., Cheraghi, P., & Doosti-Irani, A. (2025). The global prevalence of uncontrolled hypertension: A systematic review and meta-analysis. *BMC Public Health, 25*(1), 4259. https://doi.org/10.1186/s12889-025-25553-4

Chowdhury, M. Z. I., Naeem, I., Quan, H., Leung, A. A., Sikdar, K. C., O'Beirne, M., & Turin, T. C. (2022). Prediction of hypertension using traditional regression and machine learning models: A systematic review and meta-analysis. *PLOS ONE*, *17*(4), e0266334. https://doi.org/10.1371/journal.pone.0266334

Danang, D., Haryani, H., Aini, Q., Ramahdan, F. A., & Edwards, J. (2025). Empowering digital literacy through blockchain-based Alphasign for secure and sustainable e-governance.

Danang, D., Prasetya, F. A., & Siswanto, E. (2026). Design of intelligent street lighting systems based on motion and ambient light sensors. *Journal of Multidisciplinary Research and Technology*, *2*(1), 65–79.

Effati, S., Kamarzardi-Torghabe, A., Azizi-Froutaghe, F., Atighi, I., & Ghiasi-Hafez, S. (2024). Web application using machine learning to predict cardiovascular disease and hypertension in mine workers. *Scientific Reports*, *14*(1), 31662. https://doi.org/10.1038/s41598-024-80919-9

Eini, P., Rezayee, M., Kassulke, M., & Tremblay, J. (2026). Efficacy and comparative performance of machine learning models for stroke risk prediction in hypertensive patients: A systematic review and meta-analysis. *International Journal of Cardiology: Cardiovascular Risk and Prevention*, *28*, 200564. https://doi.org/10.1016/j.ijcrp.2025.200564

Eldem, A. (2025). A new hybrid learning model for early diagnosis of hypertension using IoMT technologies. *Ain Shams Engineering Journal*, *16*(8), 103490. https://doi.org/10.1016/j.asej.2025.103490

Engda, A. A., Salau, A. O., & Ajala, O. (2025). Classical machine learning approaches for early hypertension risk prediction: A systematic review. *Applied AI Letters*, *6*(3), e70005. https://doi.org/10.1002/ail2.70005

Guo, S., Ge, J.-X., Liu, S.-N., Zhou, J.-Y., Li, C., Chen, H.-J., Chen, L., Shen, Y.-Q., & Zhou, Q.-L. (2023). Development of a convenient and effective hypertension risk prediction model and exploration of the relationship between serum ferritin and hypertension risk: A study based on NHANES 2017–March 2020. *Frontiers in Cardiovascular Medicine*, *10*, 1224795. https://doi.org/10.3389/fcvm.2023.1224795

Hwang, S. H., Lee, H., Lee, J. H., Lee, M., Koyanagi, A., Smith, L., Rhee, S. Y., Yon, D. K., & Lee, J. (2024). Machine learning–based prediction for incident hypertension based on regular health checkup data: Derivation and validation in two independent nationwide cohorts in South Korea and Japan. *Journal of Medical Internet Research*, *26*, e52794. https://doi.org/10.2196/52794

Islam, M. M., Alam, M. J., Maniruzzaman, M., Ahmed, N. A. M. F., Ali, M. S., Rahman, M. J., & Roy, D. C. (2023). Predicting the risk of hypertension using machine learning algorithms: A cross-sectional study in Ethiopia. *PLOS ONE*, *18*(8), e0289613. https://doi.org/10.1371/journal.pone.0289613

Lee, H., & Tsoi, P. (2025). Feature-enhanced machine learning for all-cause mortality prediction in healthcare data (Version 1). *arXiv*. https://doi.org/10.48550/arXiv.2503.21241

Mirzaye, A. W., Saadatfar, H., & Nematollahi, M. A. (2026). Enhancing hypertension risk diagnosis using a hybrid machine learning framework: Leveraging body composition data. *BioMed Research International, 2026*, 6335947. https://doi.org/10.1155/bmri/6335947

Montagna, S., Pengo, M. F., Ferretti, S., Borghi, C., Ferri, C., Grassi, G., Muiesan, M. L., & Parati, G. (2022). Machine learning in hypertension detection: A study on World Hypertension Day data. *Journal of Medical Systems, 47*(1), 1. https://doi.org/10.1007/s10916-022-01900-5

Mroz, T., Griffin, M., Cartabuke, R., Laffin, L., Russo-Alvarez, G., Thomas, G., Smedira, N., Meese, T., Shost, M., & Habboub, G. (2024). Predicting hypertension control using machine learning. *PLOS ONE, 19*(3), e0299932. https://doi.org/10.1371/journal.pone.0299932

Murad, S. H., Tayfor, N. B., Mahmood, N. H., & Arman, L. (2025). Hybrid genetic algorithms-driven optimization of machine learning models for heart disease prediction. *MethodsX, 15*, 103510. https://doi.org/10.1016/j.mex.2025.103510

Mutale, B., Withanage, N. C., Mishra, P. K., Shen, J., Abdelrahman, K., & Fnais, M. S. (2024). A performance evaluation of random forest, artificial neural network, and support vector machine learning algorithms to predict spatio-temporal land use–land cover dynamics: A case from Lusaka and Colombo. *Frontiers in Environmental Science, 12*, 1431645. https://doi.org/10.3389/fenvs.2024.1431645

Naik, A., Nalepa, J., Wijata, A. M., Mahon, J., Mistry, D., Knowles, A. T., Dawson, E. A., Lip, G. Y. H., Olier, I., & Ortega-Martorell, S. (2025). Artificial intelligence and digital twins for the personalised prediction of hypertension risk. *Computers in Biology and Medicine, 196*, 110718. https://doi.org/10.1016/j.compbiomed.2025.110718

Narasimhan, G., & Victor, A. (2025). A hybrid approach with metaheuristic optimization and random forest in improving heart disease prediction. *Scientific Reports, 15*(1), 10971. https://doi.org/10.1038/s41598-024-73867-x

Naskinova, I., Kolev, M., Karova, D., & Milev, M. (2026). Machine learning-based blood pressure prediction using cardiovascular disease data: A comprehensive comparative study. *Electronics, 15*(2), 312. https://doi.org/10.3390/electronics15020312

Nematollahi, M. A., Jahangiri, S., Asadollahi, A., Salimi, M., Dehghan, A., Mashayekh, M., Roshanzamir, M., Gholamabbas, G., Alizadehsani, R., Bazrafshan, M., Bazrafshan, H., Bazrafshan Drissi, H., & Shariful Islam, S. M. (2023). Body composition predicts hypertension using machine learning methods: A cohort study. *Scientific Reports, 13*(1), 6885. https://doi.org/10.1038/s41598-023-34127-6

Reel, P. S., Reel, S., Van Kralingen, J. C., Langton, K., Lang, K., Erlic, Z., Larsen, C. K., Amar, L., Pamporaki, C., Mulatero, P., Blanchard, A., Kabat, M., Robertson, S., MacKenzie, S. M., Taylor, A. E., Peitzsch, M., Ceccato, F., Scaroni, C., Reincke, M., & Jefferson, E. (2022). Machine learning for classification of hypertension subtypes using multi-omics: A multi-centre, retrospective, data-driven study. *eBioMedicine, 84*, 104276. https://doi.org/10.1016/j.ebiom.2022.104276

Ribino, P., Di Napoli, C., Paragliola, G., & Serino, L. (2024). Hyper-parameter optimization through reinforcement learning for survival prediction of patients with heart failure. *Procedia Computer Science, 239*, 1754–1761. https://doi.org/10.1016/j.procs.2024.06.354

Sakka, Y., Qarashai, D., & Altarawneh, A. (2023). Predicting hypertension using machine learning: A case study at Petra University. *International Journal of Advanced Computer Science and Applications, 14*(3). https://doi.org/10.14569/IJACSA.2023.0140368

Schjerven, F. E., Ingeström, E. M. L., Steinsland, I., & Lindseth, F. (2024). Development of risk models of incident hypertension using machine learning on the HUNT study data. *Scientific Reports, 14*(1), 5609. https://doi.org/10.1038/s41598-024-56170-7

Seo, J.-W., Lee, S., & Yim, M. H. (2024). Machine learning approach for predicting hypertension based on body composition in South Korean adults. *Bioengineering, 11*(9), 921. https://doi.org/10.3390/bioengineering11090921

Septian, E., Khaefi, M. R., Athoillah, A., Aisyah, D. N., Hardhantyo, M., Rahman, F. M., & Manikam, L. (2025). Prediction of personalised hypertension using machine learning in Indonesian population. *Journal of Medical Systems, 49*(1), 137. https://doi.org/10.1007/s10916-025-02253-5

Vera-Ponce, V. J., Zuzunaga-Montoya, F. E., Ballena-Caicedo, J., Gutierrez De Carrillo, C. I., León-Figueroa, D. A., & Valladares-Garrido, M. J. (2025). Predictive variables and diagnostic performance of cross-sectional models for hypertension detection: A systematic review. *Frontiers in Cardiovascular Medicine, 12*, 1713531. https://doi.org/10.3389/fcvm.2025.1713531

Wu, Y., Xin, B., Wan, Q., Ren, Y., & Jiang, W. (2024). Risk factors and prediction models for cardiovascular complications of hypertension in older adults with machine learning: A cross-sectional study. *Heliyon*, *10*(6), e27941. https://doi.org/10.1016/j.heliyon.2024.e27941

Yang, J., Wang, H., Liu, P., Lu, Y., Yao, M., & Yan, H. (2024). Prediction of hypertension risk based on multiple feature fusion. *Journal of Biomedical Informatics, 157,* 104701. https://doi.org/10.1016/j.jbi.2024.104701

Yi, J., Wang, L., Song, J., Liu, Y., Liu, J., Zhang, H., Lu, J., & Zheng, X. (2024). Development of a machine learning-based model for predicting individual responses to antihypertensive treatments. *Nutrition, Metabolism and Cardiovascular Diseases*. Advance online publication. https://doi.org/10.1016/j.numecd.2024.02.014