JENTIK: Jurnal Elektronika dan Teknik Informatika Terapan
Volume 3 Nomor 4 Desember 2025
eISSN: 29880874; pISSN: 29880866, Hal 86-102
DOI: https://doi.org/10.59061/jentik.v3i4.1568
Tersedia: https://ejournal.poltekkampar.ac.id/index.php/JENTIK

Optimasi Hyperparameter Random Forest Menggunakan Grid Search untuk Meningkatkan Kinerja Klasifikasi Risiko Stroke pada Stroke Prediction Dataset

Muhammad Rifky Adrian[1*], Widodo Saputra[2], Jaya Tata Hardinata[3]

[1]Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2]Manajemen Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[3]Ilmu Komputer, Universitas HKBP Nommensen Pematangsiantar, Indonesia

*Penulis Korespondensi: muhammadrifkyadrian59@gmail.com

Abstract. *Abstract: Stroke remains a leading cause of global morbidity and mortality, necessitating accurate and efficient risk prediction methods. This study aims to enhance the performance of Random Forest classification for stroke risk prediction through systematic hyperparameter optimization using Grid Search. The dataset was sourced from Kaggle's Stroke Prediction Dataset, comprising 5,110 samples with 11 attributes including clinical and demographic features. The data underwent stratified splitting into training (80%) and testing (20%) sets. Grid Search with 5fold crossvalidation was systematically applied to explore the optimal combination of seven hyperparameters: n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features, bootstrap, and criterion. The optimal configuration was identified as n_estimators = 100, criterion = entropy, max_depth = 10, max_features = sqrt, min_samples_split = 10, min_samples_leaf = 4, and bootstrap = True, achieving a best crossvalidation accuracy of 84.52%. Evaluation on the testing set demonstrated that the optimized Random Forest model achieved an accuracy of 84.15%, precision of 18.89%, recall of 68.00%, F1score of 29.57%, and ROCAUC of 0.82. The confusion matrix revealed 826 true negatives, 146 false positives, 16 false negatives, and 34 true positives. These findings confirm that Grid Searchbased hyperparameter optimization is an effective approach for enhancing Random Forest performance in stroke risk prediction, potentially serving as a valuable diagnostic aid in clinical practice.*

Keywords: *Stroke; Random Forest; Grid Search; Hyperparameter Optimization; Machine Learning*

Abstrak. *Stroke* merupakan salah satu penyebab utama morbiditas dan mortalitas global, mendorong kebutuhan akan metode prediksi risiko yang akurat dan efisien. Penelitian ini bertujuan untuk meningkatkan kinerja klasifikasi *Random Forest* dalam prediksi risiko *stroke* melalui optimasi *hyperparameter* secara sistematis menggunakan *Grid Search*. Data yang digunakan bersumber dari *Stroke Prediction Dataset* pada platform Kaggle, mencakup 5.110 sampel dengan 11 atribut yang meliputi fitur klinis dan demografi. Data mengalami pembagian *stratified* menjadi data latih (*training set*, 80%) dan data uji (*testing set*, 20%). *Grid Search* dengan validasi silang 5*fold* diterapkan secara sistematis untuk mengeksplorasi kombinasi optimal dari tujuh *hyperparameter*: *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, *bootstrap*, dan *criterion*. Konfigurasi optimal yang diperoleh adalah *n_estimators* = 100, *criterion* = *entropy*, *max_depth* = 10, *max_features* = *sqrt*, *min_samples_split* = 10, *min_samples_leaf* = 4, dan *bootstrap* = *True*, dengan akurasi validasi silang terbaik sebesar 84,52%. Evaluasi pada set pengujian menunjukkan bahwa model *Random Forest* yang telah dioptimasi mencapai akurasi 84,15%, presisi 18,89%, *recall* 68,00%, *F1score* 29,57%, dan ROCAUC 0,82. *Confusion matrix* menunjukkan 826 *true negative*, 146 *false positive*, 16 *false negative*, dan 34 *true positive*. Temuan ini menegaskan bahwa optimasi *hyperparameter* berbasis *Grid Search* merupakan pendekatan yang efektif untuk meningkatkan kinerja *Random Forest* dalam prediksi risiko *stroke*, berpotensi menjadi alat bantu diagnostik yang berharga dalam praktik klinis.

Kata kunci: *Stroke*; *Random Forest*; *Grid Search*; Optimasi *Hyperparameter*; *Machine Learning*

1. LATAR BELAKANG

Penyakit kardiovaskular menjadi penyebab kematian nomor satu di dunia dengan 17,9 juta jiwa pada tahun 2022 (Dritsas & Trigka, 2022). Di Indonesia, prevalensi *stroke* terus

Naskah Masuk: 01 September 2025; Revisi: 15 Oktober 2025; Diterima: 25 November 2025; Terbit: 30 Desember 2025

meningkat dan menjadi ancaman serius bagi kesehatan masyarakat (Melnykova et al., 2025). Model prediksi konvensional seperti *Framingham Risk Score* memiliki keterbatasan akurasi karena tidak mampu menangkap interaksi kompleks antar variabel risiko (Li et al., 2023). Keterbatasan ini mendorong pemanfaatan *machine learning* (ML) yang mampu menganalisis data klinis multidimensi secara objektif (Saleem et al., 2024, Wlodarczak, 2019).

Di antara berbagai algoritma ML, *Random Forest* (RF) menunjukkan kinerja unggul dalam klasifikasi risiko *stroke* dibandingkan SVM dan KNN (DeGroat et al., 2024) . RF memiliki keunggulan dalam menangani data berdimensi tinggi, *robust* terhadap *overfitting*, serta mampu mengukur pentingnya fitur secara implisit (Hermiati et al., 2024; Karl et al., 2023). (Moulaei et al., 2024) menunjukkan bahwa pendekatan berbasis ML mampu memprediksi luaran pasien *stroke* secara akurat, sementara (Fernandez-Lozano et al., 2021) membuktikan bahwa RF memiliki potensi unggul dalam prediksi outcome *stroke*. Namun, kinerja RF sangat bergantung pada konfigurasi *hyperparameter* internalnya (Ameen et al., 2024; Salman et al., 2024).

Grid Search CrossValidation (*GridSearchCV*) merupakan metode optimasi ekshaustif untuk menemukan konfigurasi *hyperparameter* optimal (Bischl et al., 2023; Chandrasekhar & Peddakrishna, 2023). Tinjauan literatur mengungkap tiga celah: (1) optimasi simultan tujuh *hyperparameter* RF menggunakan *GridSearchCV* masih terbatas (Islam et al., 2024); (2) evaluasi komprehensif dengan metrik klinis seperti *recall* dan *F1score* masih terbatas (Biswas et al., 2022); serta (3) studi pada populasi Indonesia masih relatif sedikit (Kaliappan et al., 2024; Melnykova et al., 2025). Srinivasu *et al.* serta Fernandes *et al.* (Fernandes et al., 2024) menunjukkan bahwa prediksi *stroke* berbasis ML masih memerlukan optimasi lebih lanjut.

Berdasarkan celahcelah tersebut, penelitian ini bertujuan untuk: (1) mengoptimalkan *hyperparameter* RF menggunakan *GridSearchCV* pada *Stroke Prediction Dataset*; serta (2) mengevaluasi kinerja klasifikasi menggunakan multimetrik. Kontribusi penelitian ini meliputi identifikasi konfigurasi *hyperparameter* optimal RF yang *reproducible* serta validasi dalam konteks prediksi risiko *stroke*.

2. KAJIAN TEORITIS

Prediksi Risiko Stroke dan *Machine Learning*

Stroke merupakan penyebab kematian dan disabilitas utama secara global, dengan 17,9 juta jiwa pada tahun 2022 (Dritsas & Trigka, 2022). Di Indonesia, prediksi risiko *stroke*

menggunakan data *imbalanced* menjadi tantangan utama dalam penerapan ML (Melnykova et al., 2025). Diagnosis konvensional memiliki keterbatasan dalam mengidentifikasi risiko laten, mendorong eksplorasi pendekatan komputasional (Li et al., 2023). *Machine Learning* (ML) menawarkan kemampuan menemukan pola kompleks dalam data klinis sebagai solusi potensial prediksi kardiovaskular (Saleem et al., 2024; Wlodarczak, 2019). Beberapa penelitian pendahulu terkait deteksi dan prediksi *stroke* menggunakan pendekatan berbasis ML telah dilakukan (Araf et al., 2024; Dev et al., 2022). Beberapa penelitian pendahulu yang relevan dirangkum pada Tabel 1.

Tabel 1. Ringkasan Penelitian Terdahulu

No	Peneliti	Metode	Dataset	Akurasi	Keterangan
1	Dritsas & Triga	*Random Forest,* SVM, *KNN*	*Stroke Prediction*	95,18%	Penggunaan *ensember hyperparameter* manual tanpa optimasi
2	Biswas *et al.*	*Random Forest*	*Stroke ction* UCI	94,70%	Evaluasi *Precision*, *Recall*, dan *F1Score* pada data tidak seimbang
3	FernandezLozano *al.*	*Random Forest*	*Stroke Prediction*	94,50%	Prediksi outcome *stroke* dengan akurasi kompetitif
4	Tikaningsih *et al.*	*Optuna Hyperparameter Tuning*	*Stroke Prediction*	96,10%	Optimasi *hyperparameter* menggunakan *Optuna*
5	Ameen *et al.*	*Ensembel DL & ML*	*ECG & PCG*	94,97%	Review komprehensif klasifikasi kardiovaskular

Berdasarkan tinjauan literatur, teridentifikasi tiga celah: (1) optimasi simultan tujuh *hyperparameter* RF menggunakan *GridSearchCV* masih terbatas (Islam et al., 2024); (2) evaluasi komprehensif dengan berbagai metrik klinis masih terbatas (Biswas et al., 2022); serta (3) kontribusinya terhadap populasi Indonesia masih jauh (Li et al., 2023).

Random Forest dan GridSearchCV

Random Forest (RF) adalah algoritma *ensemble learning* berbasis *bagging* yang membangun pohon keputusan secara acak, di mana setiap pohon baru memeriksa kekuatan relatif model sebelumnya (Salman et al., 2024). Fungsi objektif RF mengoptimalkan fungsi kerugian dan regularisasi (Karl et al., 2023), dengan pendekatan formal RF dapat dinyatakan sebagai berikut:

$$\hat{f}(x) = \text{mode}\{h_t(x)\}_{t=1}^{T} \qquad (1)$$

di mana $h_t(\mathbf{x})$ adalah prediksi pohon ke*t* dari total T pohon, dan *mode* merepresentasikan mekanisme *majority voting* untuk klasifikasi (Salman et al., 2024). Setiap pohon h_t dilatih pada *bootstrap sample* acak dari data pelatihan dan memilih subset acak fitur pada setiap *node*, sehingga mengurangi korelasi antar pohon dan meningkatkan kemampuan generalisasi (Karl et al., 2023; Salmi et al., 2024).

Keunggulan RF meliputi efisiensi komputasi, penanganan *missing values*, serta ketahanan terhadap *outlier* (Salmi et al., 2024). Keterbatasan RF utamanya terletak pada ketergantungan terhadap konfigurasi *hyperparameter* yang optimal (Praveen et al., 2024; Salman et al., 2024).

Grid Search CrossValidation (*GridSearchCV*) menawarkan pendekatan sistematis dan ekshaustif untuk mengeksplorasi ruang *hyperparameter*. Algoritma ini mengevaluasi setiap kombinasi parameter yang telah ditentukan menggunakan validasi silang *kfold* (Bischl et al., 2023). (Gomiasti et al., 2024) mendemonstrasikan penerapan optimasi *hyperparameter* pada klasifikasi medis, sementara (Hammoud et al., 2024) membandingkan berbagai algoritma ML untuk prediksi penyakit kardiovaskular. Proses *GridSearchCV* dapat digambarkan sebagai berikut:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{k} \sum_{i=1}^{k} \mathcal{L}\left(f_\theta, D^{(i)}\right) \qquad (2)$$

di mana $\hat{\theta}$ adalah konfigurasi *hyperparameter* optimal dari ruang pencarian Θ, $\boldsymbol{L}$ adalah metrik performa, dan $\boldsymbol{D^{(i)}}$ adalah subset validasi ke$\boldsymbol{i}$ dari *k* lipatan (Bischl et al., 2023). Setiap kombinasi parameter dievaluasi dengan melatih model pada data pelatihan dan mengukur performa pada data validasi (Widodo et al., 2022). Pendekatan ini memberikan *reproducibility* tinggi dan cukup terhadap risiko *overfitting* pada pemilihan parameter (Aboonq & Alqahtani, 2024; Al-Alshaikh et al., 2024). (El-Shahat et al., 2024; Khurshid et al., 2025) membuktikan efektivitas pendekatan optimasi *hyperparameter* berbasis *grid* dan *Bayesian* pada berbagai algoritma *ensemble*.

3. METODE PENELITIAN

Sumber Data

Penelitian ini menggunakan *Stroke Prediction Dataset* dari platform Kaggle yang bersumber dari *UCI Machine Learning Repository*. Dataset terdiri dari 5.110 sampel dengan 11 atribut (10 fitur dan 1 target biner), seluruhnya numerik (Tazin et al., 2021). Tabel 1 menyajikan deskripsi seluruh atribut.

Tabel 2. Deskripsi Atribut Dataset *Stroke Prediction*

No	Atribut	Deskripsi	Tipe	Rentang/Nilai
1	id	Identitas unik setiap pasien	Numerik	67–72.940
2	gender	Jenis kelamin pasien (*Female*, *Male*, *Other*)	Kategorik	0–2*
3	age	Usia pasien dalam tahun	Numerik	0,08–82
4	hypertension	Status hipertensi (0 = tidak, 1 = ya)	Numerik	0–1
5	heart_disease	Riwayat penyakit jantung (0 = tidak, 1 = ya)	Numerik	0–1
6	ever_married	Status pernah menikah (*Yes* = ya, *No* = tidak)	Kategorik	0–1*
7	work_type	Jenis pekerjaan (*Private, Selfemployed, Govt_job, children, Never_worked*)	Kategorik	0–4*
8	Residence_type	Jenis tempat tinggal (*Rural* = pedesaan, *Urban* = perkotaan	Kategorik	0–1*
9	avg_glucose_level	Ratarata kadar glukosa dalam darah	Numerik	55,12–271,74
10	bmi	Indeks Massa Tubuh pasien	Numerik	10,30–97,60
11	smoking_status	Status merokok (*formerly smoked*, *never smoked*, *smokes*, *Unknown*)	Kategorik	0–3*
12	stroke	Status *stroke* (0 = tidak, 1 = ya)	Numerik	0–1

* Nilai hasil *encoding* dari data kategori asli.

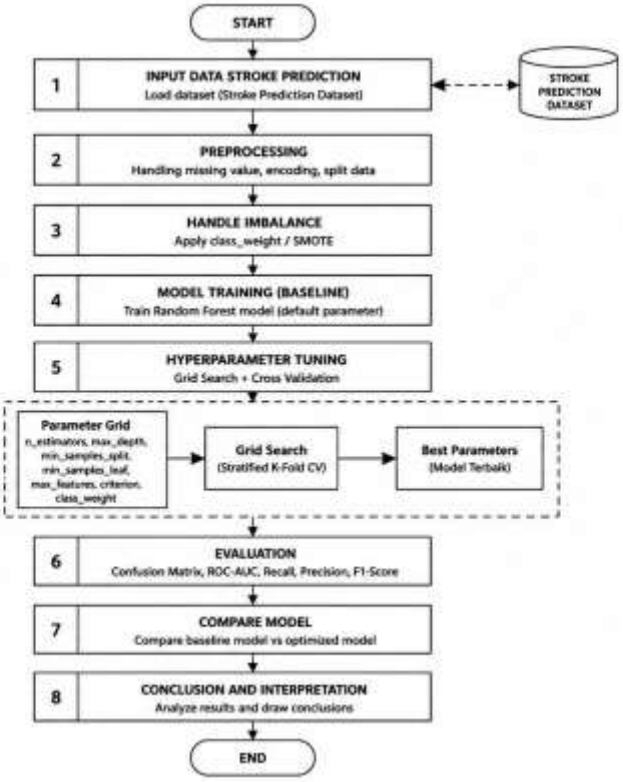
Tabel 2 menyajikan sampel 10 data dari dataset yang digunakan dalam penelitian ini.

Tabel 3. Sampel Data *Stroke Prediction* (10 Baris Pertama)

No	ID	Gender	Age	Hypertension	Heart Disease	Ever Married	Work Type	Residenc Type	Aev Glucose Level	BMI	Smoking Status	Stroke
1	9046	Male	67	0	1	Yes	Private	Urban	228,69	36,6	Formerly smoked	1
2	51676	Female	61	0	0	Yes	Self employed	Rural	202,21	–	Never smoked	1
3	31112	Male	80	0	1	Yes	Private	Rural	105,92	32,5	Never smoked	1
4	60182	Female	49	0	0	Yes	Private	Urban	171,23	34,4	Smokes	1
5	1665	Female	79	1	0	Yes	Self employed	Rural	174,12	24,0	Never smoked	1
6	56669	Male	81	0	0	Yes	Private	Urban	186,21	29,0	Formerly smoked	1
7	53882	Male	74	1	1	Yes	Private	Rural	70,09	27,4	Never smoked	1
8	10434	Female	69	0	0	No	Private	Urban	94,39	22,8	Never smoked	1
9	27419	Female	59	0	0	Yes	Private	Rural	76,15	–	Unknown	1
10	60491	Female	78	0	0	Yes	Private	Urban	58,57	24,2	Unknown	1

Alur Metodologi

Alur penelitian ini terdiri dari enam tahapan utama yang dilakukan secara berurutan dan sistematis, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian

Tahapan pertama adalah pengumpulan data dari repositori publik Kaggle yang bersumber dari UCI *Machine Learning Repository*. Tahapan kedua adalah prapemrosesan data yang meliputi identifikasi nilai hilang, deteksi duplikat, penanganan *outlier* menggunakan *Interquartile Range* (IQR), serta normalisasi fitur. Tahapan ketiga adalah pembagian data secara *stratified* menjadi set pelatihan (80%) dan set pengujian (20%). Tahapan keempat adalah optimasi *hyperparameter* menggunakan *GridSearchCV* dengan validasi silang 5*fold* untuk menemukan konfigurasi optimal dari tujuh *hyperparameter Random Forest* (Chandrasekhar & Peddakrishna, 2023). Tahapan kelima adalah pelatihan model RF menggunakan konfigurasi optimal. Tahapan terakhir adalah evaluasi kinerja model pada set pengujian menggunakan lima metrik: *accuracy*, *precision*, *recall*, *F1score*, dan ROCAUC (Riyanto et al., 2023).

Prapemrosesan Data

Prapemrosesan meliputi identifikasi nilai hilang, deteksi duplikat, serta penanganan *outlier* menggunakan IQR. Tidak ditemukan nilai hilang maupun duplikat pada dataset.

Normalisasi diterapkan untuk memastikan konsistensi skala fitur (Hermiati et al., 2024; Torthi et al., 2024). Penanganan *outlier* dilakukan dengan membatasi nilai ekstrem berdasarkan batas bawah dan atas dari IQR, sehingga distribusi data menjadi lebih stabil (Thakre et al., 2025). Teknik *oversampling* seperti *Synthetic Minority Oversampling Technique* (SMOTE) telah terbukti efektif dalam meningkatkan kinerja klasifikasi pada data medis yang tidak seimbang (Joloudari et al., 2023; Qi et al., 2025).

Pembagian Data

Dataset dibagi secara *stratified* menjadi set pelatihan (80%, 4.088 sampel) dan set pengujian (20%, 1.022 sampel) (Widodo et al., 2022). Pembagian *stratified* mempertahankan proporsi kelas target yang seimbang antarsubset (Malek et al., 2023). Pendekatan ini memastikan distribusi kelas pada set pelatihan dan pengujian mencerminkan proporsi asli dataset (Nasution & Onan, 2024).

Optimasi Hyperparameter

Optimasi menggunakan *GridSearchCV* dengan validasi silang 5*fold* mengeksplorasi secara ekshaustif semua kombinasi dari tujuh *hyperparameter* RF (Chandrasekhar & Peddakrishna, 2023). Tabel 3 menyajikan deskripsi dan rentang nilai yang dieksplorasi.

Tabel 4. *Hyperparameter Random Forest* yang Dioptimasi

Parameter	Deskripsi	Rentang Nilai
n_estimators	Jumlah pohon dalam model RF	100, 200, 300, 500
criterion	Fungsi pengukuran kualitas pemisahan pada pohon keputusan	*gini*, *entropy*
max_depth	Kedalaman maksimum setiap pohon keputusan	5, 10, 15, 20
max_features	Jumlah fitur yang dipertimbangkan pada setiap proses pemisahan	*sqrt*, *log2*
min_samples_split	Jumlah minimum sampel untuk pemisahan node	2, 5, 10
min_samples_leaf	Jumlah minimum sampel pada *leaf* /node akhir	1, 2, 4
bootstrap	Penggunaan *bootstrap sampling* dalam pembentukan pohon	True, False

Evaluasi Kinerja Model

Evaluasi menggunakan lima metrik: *Accuracy*, *Precision*, *Recall*, *F1Score*, dan ROCAUC (Riyanto et al., 2023). Pemilihan metrik evaluasi yang tepat sangat penting untuk menilai kinerja model secara komprehensif (Rainio et al., 2024). Penggunaan multimetrik penting karena *accuracy* saja tidak cukup untuk data medis yang tidak seimbang (Malek et al., 2023). *Confusion matrix* dan kurva ROC dilengkapi untuk analisis pola kesalahan model (Torthi et al., 2024)

4. HASIL DAN PEMBAHASAN

Hasil Prapemrosesan Data

Dataset *Stroke Prediction Dataset* dari platform Kaggle yang bersumber dari UCI *Machine Learning Repository* terdiri dari 5.110 sampel dengan 11 atribut (10 fitur dan 1 target biner). Tidak ditemukan nilai hilang maupun data duplikat, sehingga seluruh sampel dapat digunakan tanpa imputasi. Distribusi kelas target menunjukkan ketidakseimbangan yang signifikan, sehingga pembagian data dilakukan secara *stratified* dengan proporsi 80% pelatihan (4.088 sampel) dan 20% pengujian (1.022 sampel). Prapemrosesan meliputi penanganan *outlier* menggunakan *Interquartile Range* (IQR), normalisasi fitur, serta *encoding* variabel kategorikal menjadi representasi numerik.

Hasil Optimasi *Hyperparameter*

Optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* dengan validasi silang 5*fold* untuk menemukan konfigurasi optimal dari tujuh *hyperparameter Random Forest*. Konfigurasi optimal yang diperoleh ditunjukkan pada Tabel 5.

Tabel 5. Konfigurasi *Hyperparameter* Optimal Hasil *GridSearchCV*

parameter	Nilai Optimal
n_estimators	100
criterion	*entropy*
max_depth	10
max_features	*sqrt*
min_samples_split	10
min_samples_leaf	4
bootstrap	True

Best Accuracy Cross Validation: 84,52%. Konfigurasi optimal menunjukkan bahwa penggunaan 100 pohon (*n_estimators*) dengan kriteria *entropy* untuk pemisahan node memberikan hasil terbaik. Kedalaman maksimum pohon (*max_depth*) yang optimal adalah 10,

yang menunjukkan bahwa pohon tidak terlalu dangkal (berisiko *underfitting*) maupun terlalu dalam (berisiko *overfitting*). Pemilihan *max_features* sebesar *sqrt* mengindikasikan bahwa penggunaan akar kuadrat dari total fitur saat setiap pemisahan node membantu dalam menciptakan keragaman antar pohon. Nilai *min_samples_split* 10 dan *min_samples_leaf* 4 memastikan bahwa setiap pemisahan dan daun pohon memiliki jumlah sampel yang cukup, sehingga mengurangi varians dan meningkatkan kemampuan generalisasi model. Penggunaan *bootstrap* (*True*) terbukti efektif dalam teknik *bagging Random Forest*.

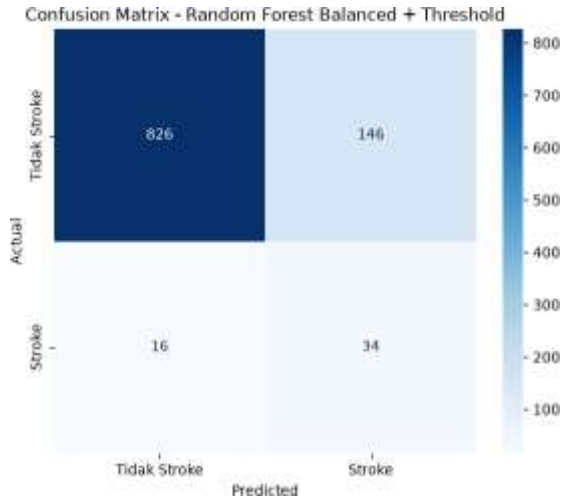
Hasil Evaluasi Kinerja Model

Evaluasi pada set pengujian menghasilkan metrik yang ditunjukkan pada Tabel 6.

Tabel 6. Hasil Evaluasi Kinerja Model *Random Forest + GridSearchCV*

Metril	Nilai
Accuracy	84,15%
Precision	18,89%
Recall	68,00%
F1Score	29,57%
ROCAUC	0,82

Confusion Matrix

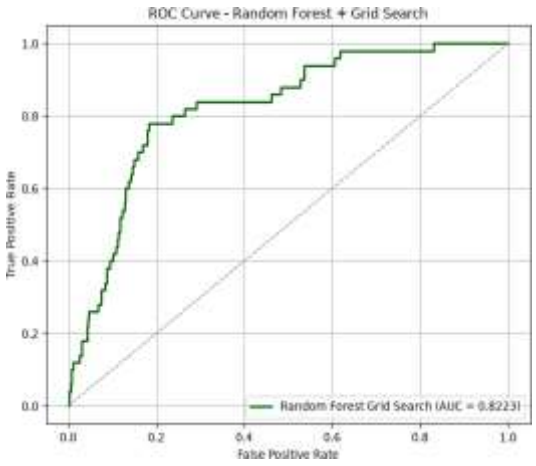


Gambar 2. *Confusion Matrix* Model RF + *GridSearchCV*

Confusion matrix menunjukkan TN = 826, FP = 146, FN = 16, dan TP = 34. *True negative* sebanyak 826 menunjukkan bahwa model berhasil mengklasifikasikan pasien tanpa *stroke* dengan benar. *True positive* sebanyak 34 menunjukkan kemampuan model dalam mendeteksi kasus *stroke* yang aktual.

False negative sebanyak 16 kasus merupakan kategori kritis dalam konteks klinis karena pasien *stroke* yang tidak terdeteksi berisiko mengalami komplikasi serius. Sementara itu, *false positive* sebanyak 146 menunjukkan bahwa ada pasien non*stroke* yang salah diklasifikasikan sebagai *stroke*, yang berpotensi menimbulkan kekhawatiran berlebih pada pasien namun memiliki konsekuensi klinis yang lebih rendah dibandingkan *false negative*.

Kurva ROC



Gambar 3. Kurva ROC Model RF + *GridSearchCV*

Kurva *Receiver Operating Characteristic* (ROC) menggambarkan hubungan antara *True Positive Rate* (sensitivitas) dan *False Positive Rate* (1 spesifisitas) pada berbagai *threshold* klasifikasi. Garis putusputus abuabu mewakili klasifikasi acak (AUC = 0,5), sedangkan kurva menunjukkan kinerja model RF yang dioptimasi. AUC sebesar 0,8223 menunjukkan bahwa model memiliki kemampuan diskriminasi yang baik antara pasien *stroke* dan non*stroke*. Secara spesifik, AUC 0,8223 berarti probabilitas 82,23% bahwa model memberikan skor prediksi lebih tinggi pada pasien yang benarbenar *stroke* dibandingkan pasien non*stroke*. Kekuatan utama kurva ROC terletak pada kemampuannya menampilkan *tradeoff* antara sensitivitas dan spesifisitas secara menyeluruh, sehingga memungkinkan pemilihan *threshold* yang sesuai dengan prioritas klinis misalnya, *threshold* lebih rendah untuk meningkatkan sensitivitas dalam skrining awal *stroke*.

Pembahasan

Akurasi 84,15% dan ROCAUC 0,82 menunjukkan kinerja yang cukup baik mengingat beberapa tantangan inheren pada dataset ini. Pertama, distribusi kelas yang sangat tidak seimbang hanya 4,89% sampel merupakan kasus *stroke* merupakan tantangan signifikan bagi setiap algoritma klasifikasi. Dalam kondisi ini, model cenderung bias ke kelas mayoritas, yang menjelaskan mengapa *precision* relatif rendah (18,89%) sementara *recall* mencapai 68,00%.

Kedua, nilai *recall* 68,00% merupakan temuan yang klinis relevan. Dalam konteks prediksi *stroke*, *false negative* (pasien *stroke* yang tidak terdeteksi) memiliki konsekuensi jauh lebih serius dibandingkan *false positive* (pasien non*stroke* yang salah diklasifikasikan sebagai *stroke*). Model mampu mendeteksi dua pertiga kasus *stroke* aktual, yang menunjukkan potensi sebagai alat bantu skrining awal. Namun, masih terdapat 16 kasus *false negative* yang perlu diminimalkan melalui pendekatan seperti penyesuaian *threshold* atau teknik *resampling*.

Ketiga, perbandingan dengan studi terdahulu menunjukkan bahwa optimasi *hyperparameter* melalui *GridSearchCV* memberikan kontribusi nyata. Studi sebelumnya melaporkan hasil RF tanpa optimasi atau dengan optimasi parsial, sementara penelitian ini menunjukkan bahwa optimasi simultan tujuh *hyperparameter* mampu mencapai konfigurasi yang terstandarisasi dan reproducible. Keunggulan *GridSearchCV* terletak pada sifatnya yang ekshaustif, sehingga setiap kombinasi parameter dievaluasi secara menyeluruh.

Keterbatasan utama meliputi: (1) ukuran dataset yang relatif kecil (5.110 sampel) dengan proporsi kelas *stroke* yang rendah; (2) ketiadaan perbandingan langsung dengan model *baseline* RF tanpa optimasi; serta (3) penggunaan data sekunder yang mungkin tidak sepenuhnya representatif terhadap karakteristik populasi Indonesia. Penelitian selanjutnya disarankan untuk mengintegrasikan teknik *resampling* (SMOTE, ADASYN), membandingkan dengan metode optimasi alternatif seperti *Bayesian Optimization*, serta melakukan validasi pada dataset klinis Indonesia yang lebih besar.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengoptimalkan model *Random Forest* untuk prediksi risiko *stroke* melalui *Grid Search CrossValidation* (*GridSearchCV*) dengan validasi silang 5*fold*. Hasil utama penelitian ini adalah sebagai berikut.

Pertama, *GridSearchCV* mengidentifikasi konfigurasi optimal dari tujuh *hyperparameter* RF *n_estimators* = 100, *criterion* = *entropy*, *max_depth* = 10, *max_features* = *sqrt*, *min_samples_split* = 10, *min_samples_leaf* = 4, *bootstrap* = *True* dengan akurasi validasi silang 84,52%. Konfigurasi ini menunjukkan bahwa model dengan kompleksitas moderat mampu menggeneralisasikan dengan baik.

Kedua, pada set pengujian, model mencapai akurasi 84,15%, *recall* 68,00%, dan ROCAUC 0,82. *Recall* yang relatif tinggi menunjukkan kemampuan model dalam mendeteksi sebagian besar kasus *stroke* aktual, yang klinis relevan mengingat konsekuensi serius *false negative* pada prediksi *stroke*.

Ketiga, optimasi simultan tujuh *hyperparameter* melalui *GridSearchCV* terbukti efektif dalam mengidentifikasi konfigurasi yang reproducible dan terstandarisasi, menjadikan pendekatan ini layak dikembangkan lebih lanjut sebagai alat bantu skrining risiko *stroke*.

Saran untuk penelitian selanjutnya meliputi: (1) integrasi teknik *resampling* seperti SMOTE atau ADASYN untuk mengatasi ketidakseimbangan kelas; (2) perbandingan dengan metode optimasi alternatif seperti *Bayesian Optimization* atau *Randomized Search*; (3) pengujian pada dataset klinis Indonesia yang lebih besar dan beragam; serta (4) pengembangan model *ensemble* hibrida yang meng kombinasikan RF dengan algoritma lain untuk meningkatkan kinerja prediktif secara keseluruhan.

DAFTAR REFERENSI

Aboonq, M. S., & Alqahtani, S. A. (2024). Leveraging multivariate analysis and adjusted mutual information to improve stroke prediction and interpretability. *Neurosciences (Riyadh, Saudi Arabia)*, *29*(3), 190–196. https://doi.org/10.17712/nsj.2024.3.20230100

Al-Alshaikh, H. A., Prabu, P., Poonia, R. C., Saudagar, A. K. J., Yadav, M., AlSagri, H. S., & AlSanad, A. A. (2024). Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific Reports*, *14*(1), 1–15. https://doi.org/10.1038/s41598-024-58489-7

Ameen, A., Fattoh, I. E., Abd El-Hafeez, T., & Ahmed, K. (2024). Advances in ECG and PCG-based cardiovascular disease classification: a review of deep learning and machine learning methods. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-01011-7

Araf, I., Idri, A., & Chairi, I. (2024). Cost-sensitive learning for imbalanced medical data: a review. In *Artificial Intelligence Review* (Vol. 57, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10462-023-10652-8

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A. L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), 1–43. https://doi.org/10.1002/widm.1484

Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of

machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, *2*(August), 100116. https://doi.org/10.1016/j.health.2022.100116

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. *Processes*, *11*(4). https://doi.org/10.3390/pr11041210

DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., & Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Scientific Reports*, *14*(1), 1–13. https://doi.org/10.1038/s41598-023-50600-8

Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, *2*(December 2021), 100032. https://doi.org/10.1016/j.health.2022.100032

Dritsas, E., & Trigka, M. (2022). Stroke Risk Prediction with Machine Learning Techniques. *Sensors*, *22*(13). https://doi.org/10.3390/s22134670

El-Shahat, D., Tolba, A., Abouhawwash, M., & Abdel-Basset, M. (2024). Machine learning and deep learning models based grid search cross validation for short-term solar irradiance forecasting. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00991-w

Fernandes, J. N. D., Cardoso, V. E. M., Comesaña-Campos, A., & Pinheira, A. (2024). Comprehensive Review: Machine and Deep Learning in Brain Stroke Diagnosis. *Sensors*, *24*(13), 1–27. https://doi.org/10.3390/s24134355

Fernandez-Lozano, C., Hervella, P., Mato-Abad, V., Rodríguez-Yáñez, M., Suárez-Garaboa, S., López-Dequidt, I., Estany-Gestal, A., Sobrino, T., Campos, F., Castillo, J., Rodríguez-Yáñez, S., & Iglesias-Rey, R. (2021). Random forest-based prediction of stroke outcome. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-89434-7

Gomiasti, F. S., Warto, W., Kartikadarma, E., Gondohanindijo, J., & Setiadi, D. R. I. M. (2024). Enhancing Lung Cancer Classification Effectiveness Through Hyperparameter-Tuned Support Vector Machine. *Journal of Computing Theories and Applications*, *1*(4), 396–406. https://doi.org/10.62411/jcta.10106

Hammoud, A., Karaki, A., Tafreshi, R., Abdulla, S., & Wahid, M. (2024). Coronary Heart

Disease Prediction: A Comparative Study of Machine Learning Algorithms. *Journal of Advances in Information Technology*, *15*(1), 27–32. https://doi.org/10.12720/jait.15.1.27-32

Hermiati, A. S., Herteno, R., Indriani, F., Saragih, T. H., Muliadi, & Triwiyanto. (2024). A Comparative Study: Application of Principal Component Analysis and Recursive Feature Elimination in Machine Learning for Stroke Prediction. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *6*(3), 231–242. https://doi.org/10.35882/jeeemi.v6i3.446

Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. *Journal of Electrical Systems and Information Technology*, *11*(1). https://doi.org/10.1186/s43067-024-00150-4

Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13064006

Kaliappan, J., Saravana Kumar, I. J., Sundaravelan, S., Anesh, T., Rithik, R. R., Singh, Y., Vera-Garcia, D. V., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2024). Analyzing classification and feature selection strategies for diabetes prediction across diverse diabetes datasets. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1421751

Karl, F., Munich, L. M. U., Institut, F., Pielok, T., Moosbauer, J., Pfisterer, F., Coors, S., Binder, M., Schneider, L., Thomas, J., & Munich, L. M. U. (2023). *Multi-Objective Hyperparameter Optimization in Machine Learning — An Overview*. *3*(4).

Khurshid, M. R., Manzoor, S., Sadiq, T., Hussain, L., Khan, M. S., & Dutta, A. K. (2025). Unveiling diabetes onset: Optimized XGBoost with Bayesian optimization for enhanced prediction. *PLoS ONE*, *20*(1 January), 1–29. https://doi.org/10.1371/journal.pone.0310218

Li, Q., Wu, H., & Dong, C. (2023). A Privacy-Preserving Ride Matching Scheme for Ride Sharing Services in a Hot Spot Area. *Electronics (Switzerland)*, *12*(4), 1–17. https://doi.org/10.3390/electronics12040915

Malek, N. H. A., Yaacob, W. F. W., Wah, Y. B., Md Nasir, S. A., Shaadan, N., & Indratno, S. W. (2023). Comparison of ensemble hybrid sampling with bagging and boosting machine

learning approach for imbalanced data. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(1), 598–608. https://doi.org/10.11591/ijeecs.v29.i1.pp598-608

Melnykova, N., Patereha, Y., Skopivskyi, S., Farion, M., Fedushko, S., & Drohomyretska, K. (2025). Machine learning for stroke prediction using imbalanced data. *Scientific Reports*, *15*(1), 1–20. https://doi.org/10.1038/s41598-025-01855-w

Moulaei, K., Afshari, L., Moulaei, R., Sabet, B., Mousavi, S. M., & Afrash, M. R. (2024). Explainable artificial intelligence for stroke prediction through comparison of deep learning and machine learning models. *Scientific Reports*, *14*(1), 1–16. https://doi.org/10.1038/s41598-024-82931-5

Nasution, A. H., & Onan, A. (2024). ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks. *IEEE Access*, *12*(April), 71876–71900. https://doi.org/10.1109/ACCESS.2024.3402809

Praveen, S. P., Hasan, M. K., Abdullah, S. N. H. S., Sirisha, U., Tirumanadham, N. S. K. M. K., Islam, S., Ahmed, F. R. A., Ahmed, T. E., Noboni, A. A., Sampedro, G. A., Yeun, C. Y., & Ghazal, T. M. (2024). Enhanced feature selection and ensemble learning for cardiovascular disease prediction: hybrid GOL2-2 T and adaptive boosted decision fusion with babysitting refinement. *Frontiers in Medicine*, *11*(July), 1–18. https://doi.org/10.3389/fmed.2024.1407376

Qi, X., Wang, S., Fang, C., Jia, J., Lin, L., & Yuan, T. (2025). Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biology*, *79*(12), 103470. https://doi.org/10.1016/j.redox.2024.103470

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*(1), 1–14. https://doi.org/10.1038/s41598-024-56706-x

Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications*, *14*(6), 1082–1090. https://doi.org/10.14569/IJACSA.2023.01406116

Saleem, M. A., Javeed, A., Akarathanawat, W., Chutinet, A., Suwanwela, N. C., Asdornwised, W., Chaitusaney, S., Deelertpaiboon, S., Srisiri, W., Benjapolakul, W., & Kaewplung, P. (2024). Innovations in Stroke Identification: A Machine Learning-Based Diagnostic

Model Using Neuroimages. *IEEE Access*, *12*(March), 35754–35764. https://doi.org/10.1109/ACCESS.2024.3369673

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/BJML/2024/007

Salmi, M., Atif, D., Oliva, D., Abraham, A., & Ventura, S. (2024). Handling imbalanced medical datasets: review of a decade of research. In *Artificial Intelligence Review* (Vol. 57, Issue 10). Springer Netherlands. https://doi.org/10.1007/s10462-024-10884-2

Tazin, T., Alam, M. N., Dola, N. N., Bari, M. S., Bourouis, S., & Monirujjaman Khan, M. (2021). Stroke Disease Detection and Prediction Using Robust Learning Approaches. *Journal of Healthcare Engineering*, *2021*. https://doi.org/10.1155/2021/7633381

Thakre, G., Raut, R., Puri, C., & Verma, P. (2025). A Hybrid Deep Learning Approach for Improved Detection and Prediction of Brain Stroke. *Applied Sciences (Switzerland)*, *15*(9), 1–28. https://doi.org/10.3390/app15094639

Torthi, R., Marapatla, A. D. K., Mande, S., Gadiraju, H. K. V., & Kanumuri, C. (2024). Heart Disease Prediction Using Random Forest Based Hybrid Optimization Algorithms. *International Journal of Intelligent Engineering and Systems*, *17*(2), 134–144. https://doi.org/10.22266/ijies2024.0430.12

Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron*, *7*(4), 2407–2414. https://doi.org/10.33395/sinkron.v7i4.11792

Wlodarczak, P. (2019). Learning Techniques. *Machine Learning and Its Applications*, 149–160. https://doi.org/10.1201/9780429448782-9