

Model Prediksi Penyakit Stroke Menggunakan Explainable Neural-XGBoost Berbasis SHAP

Victor Tarigan^{1✉}, Rendy Syahputra² Suharno Usman³

(1,2) Program Studi Teknik Informatika, Universitas Sam Ratulangi, Manado, Indonesia

(3) Program Studi Ilmu Keperawatan, Universitas Sam Ratulangi, Manado, Indonesia

Abstrak: Stroke merupakan masalah kesehatan yang memerlukan prediksi akurat untuk mendukung pencegahan dan penanganan dini. Penelitian ini bertujuan mengembangkan model prediksi stroke menggunakan Hybrid Deep Neural Network-XGBoost yang dilengkapi Explainable Artificial Intelligence (XAI) berbasis SHAP. Tahapan penelitian meliputi prapemrosesan data, penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE), ekstraksi fitur dengan Deep Neural Network (DNN), klasifikasi menggunakan XGBoost, dan interpretasi model melalui SHAP. Kinerja model dievaluasi menggunakan Stratified 5-Fold Cross-Validation berdasarkan akurasi, presisi, recall, F1-score, dan ROC-AUC, serta dibandingkan dengan enam model pembanding, yaitu Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, dan XGBoost Baseline. Hasil evaluasi menunjukkan bahwa Hybrid DNN-XGBoost memperoleh akurasi $64,46\% \pm 0,80\%$, presisi $64,55\% \pm 0,87\%$, recall $64,17\% \pm 0,76\%$, F1-score $64,36\% \pm 0,76\%$, dan ROC-AUC $69,72\% \pm 0,99\%$. Akurasi tersebut lebih tinggi daripada seluruh model pembanding, yang berkisar antara 48,96% dan 50,62%. Analisis SHAP mengidentifikasi DeepFeature_28, DeepFeature_31, DeepFeature_5, DeepFeature_6, dan DeepFeature_26 sebagai fitur dengan kontribusi terbesar terhadap prediksi. Fitur-fitur tersebut dikaitkan dengan faktor risiko klinis, meliputi usia, tekanan darah, kadar glukosa, dan indeks massa tubuh (BMI). Temuan ini menunjukkan bahwa Hybrid DNN-XGBoost meningkatkan kinerja prediksi dibandingkan model pembanding sekaligus mendukung interpretasi kontribusi fitur terhadap hasil prediksi stroke.

Abstract: Stroke is a major health concern that requires accurate prediction to support early prevention and intervention. This study aims to develop a stroke prediction model using a Hybrid Deep Neural Network-XGBoost approach integrated with SHapley Additive exPlanations (SHAP)-based Explainable Artificial Intelligence (XAI). The research procedures include data preprocessing, class imbalance handling using the Synthetic Minority Over-sampling Technique (SMOTE), feature extraction using a Deep Neural Network (DNN), classification using XGBoost, and model interpretation using SHAP. Model performance was evaluated using stratified 5-fold cross-validation based on accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). The proposed model was compared with six benchmark models: Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, and baseline XGBoost. The results show that the Hybrid DNN-XGBoost model achieved an accuracy of $64.46\% \pm 0.80\%$, precision of $64.55\% \pm 0.87\%$, recall of $64.17\% \pm 0.76\%$, F1-score of $64.36\% \pm 0.76\%$, and ROC-AUC of $69.72\% \pm 0.99\%$. Its accuracy exceeded that of all benchmark models, which ranged from 48.96% to 50.62%. SHAP analysis identified DeepFeature_28, DeepFeature_31, DeepFeature_5, DeepFeature_6, and DeepFeature_26 as the features contributing most to the model's predictions. These features were associated with major clinical risk factors, including age, blood pressure, glucose levels, and body mass index (BMI). These findings indicate that the Hybrid DNN-XGBoost approach improves predictive performance compared with the benchmark models while providing insights into feature contributions to stroke predictions.

Article history:

Received: 14 September 2026

Revised: 10 October 2026

Accepted: 14 October 2026

Published: 16 October 2026

Kata kunci:

stroke, prediksi, XGBoost, SHAP, klasifikasi

Keyword:

stroke, prediction, XGBoost, SHAP, classification

This is an open-access article under the [CC-BY-SA License](https://creativecommons.org/licenses/by-sa/4.0/).



How to cite: Tarigan, V., Syahputra, R., & Usman, S. (2026). Model Prediksi Penyakit Stroke Menggunakan Explainable Neural-XGBoost Berbasis SHAP. RESWARA: Jurnal Riset Ilmu Teknik, 4(4), 1610-1629. <https://doi.org/10.70716/reswara.v4i4.885>

PENDAHULUAN

Stroke merupakan penyakit serebrovaskular yang terjadi akibat gangguan aliran darah ke otak, baik karena penyumbatan pembuluh darah (ischemic stroke) maupun perdarahan akibat pecahnya pembuluh darah (hemorrhagic stroke). Kondisi ini dapat menyebabkan kerusakan jaringan otak permanen serta meningkatkan risiko kematian dan kecacatan jangka panjang yang memengaruhi fungsi motorik, sensorik, kognitif, dan komunikasi, sehingga menurunkan kualitas hidup pasien.

Perkembangan Artificial Intelligence (AI), khususnya Machine Learning (ML), telah mendorong pengembangan sistem pendukung keputusan klinis (Clinical Decision Support System). Dalam bidang kesehatan, ML digunakan untuk diagnosis, prediksi prognosis, analisis citra medis, dan prediksi penyakit kronis, termasuk stroke. Pemanfaatan data klinis, seperti usia, hipertensi, penyakit jantung, kadar glukosa darah, indeks massa tubuh (Body Mass Index), dan kebiasaan merokok, mendukung identifikasi individu yang berisiko mengalami stroke. Penelitian di bidang ini juga berkembang dari algoritma klasifikasi konvensional menuju ensemble learning, deep learning, dan Explainable Artificial Intelligence (XAI), yang bertujuan meningkatkan kinerja prediksi sekaligus memberikan interpretasi terhadap hasil model.

Penelitian prediksi stroke awalnya menggunakan algoritma Machine Learning konvensional, seperti Logistic Regression, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, dan Random Forest sebagai model dasar untuk mengidentifikasi hubungan karakteristik klinis pasien dengan risiko stroke. Namun, kinerjanya masih dipengaruhi oleh karakteristik data, ketidakseimbangan kelas, dan keterbatasan dalam memodelkan hubungan nonlinier yang kompleks. Oleh karena itu, pendekatan ensemble learning dikembangkan untuk meningkatkan akurasi prediksi (Akib et al., 2025).

Penggunaan algoritma boosting, khususnya eXtreme Gradient Boosting (XGBoost), semakin berkembang dalam prediksi stroke karena kemampuannya mengklasifikasikan data tabular. Fang dan Deng (2023) mengintegrasikan XGBoost dengan Shapley Additive exPlanations (SHAP) untuk menghasilkan prediksi yang akurat dan interpretatif. Model tersebut mencapai akurasi 96,71% dan AUC 99,19%, dengan usia, hipertensi, dan tipe tempat tinggal sebagai faktor yang paling berpengaruh. Temuan ini menunjukkan bahwa integrasi XGBoost dan Explainable Artificial Intelligence dapat meningkatkan transparansi model sekaligus mempertahankan kinerja prediksi (Fang & Deng, 2023).

Sejumlah penelitian membandingkan algoritma Machine Learning dan ensemble learning untuk meningkatkan prediksi stroke. Tang dkk. mengevaluasi Logistic Regression, Random Forest, XGBoost, CatBoost, Multi-Layer Perceptron (MLP), dan model ensemble dengan menerapkan SMOTE untuk mengatasi ketidakseimbangan kelas serta SHAP untuk interpretasi model. Hasilnya menunjukkan bahwa model ensemble dan Neural Network mengungguli model konvensional, dengan hipertensi, kadar glukosa darah, dan usia sebagai faktor dominan. Temuan ini menegaskan bahwa performa prediksi dipengaruhi oleh pemilihan algoritma, penyeimbangan data, dan interpretabilitas model (Shabbir, 2025).

Penerapan Explainable Artificial Intelligence (XAI) dalam prediksi stroke semakin berkembang melalui metode SHAP dan LIME untuk menjelaskan kontribusi fitur terhadap keputusan model. Pendekatan ini membantu tenaga medis mengidentifikasi faktor klinis yang memengaruhi prediksi, meningkatkan kepercayaan terhadap sistem pendukung keputusan, dan mendukung validasi klinis secara transparan. Namun, penerapannya masih didominasi model berbasis tree ensemble, seperti

XGBoost, Random Forest, dan Gradient Boosting, dengan interpretasi yang berfokus pada fitur asli (original features) sebagai masukan model (Chen C, Ren Z, Zong H, 2025).

Penelitian ini mengusulkan model Explainable Neural-XGBoost berbasis SHAP untuk prediksi stroke. Berbeda dari pendekatan sebelumnya, Deep Neural Network digunakan sebagai feature extractor untuk mempelajari representasi laten (latent feature representation) dari data klinis, yang selanjutnya menjadi masukan bagi XGBoost untuk proses klasifikasi. Metode SHAP kemudian digunakan untuk menginterpretasikan kontribusi fitur terhadap prediksi secara global maupun lokal. Integrasi ini diharapkan menghasilkan model yang akurat, transparan, dan mudah dipahami oleh tenaga medis.

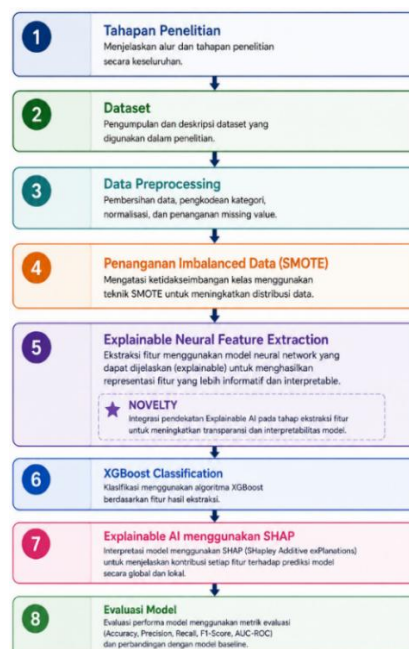
Penelitian ini bertujuan mengembangkan pendekatan prediksi stroke yang mengintegrasikan Deep Feature Extraction, XGBoost, dan XAI berbasis SHAP untuk meningkatkan performa prediksi, kemampuan generalisasi, dan interpretabilitas model. Pendekatan ini diharapkan berkontribusi pada pengembangan metode prediksi stroke serta mendukung pengambilan keputusan klinis berbasis data.

METODE PENELITIAN

Penelitian ini mengusulkan model prediksi stroke Explainable Neural-XGBoost berbasis SHAP yang mengintegrasikan Deep Neural Network untuk ekstraksi fitur, XGBoost untuk klasifikasi, dan SHAP untuk interpretasi hasil prediksi. Tahapan penelitian meliputi pengumpulan data, preprocessing, penanganan ketidakseimbangan kelas menggunakan SMOTE, ekstraksi fitur, klasifikasi, hyperparameter tuning, evaluasi performa, dan interpretasi model.

Tahapan Penelitian

Penelitian dilakukan secara sistematis, mulai dari pengumpulan dan preprocessing data, penyeimbangan kelas, ekstraksi fitur menggunakan Deep Neural Network, klasifikasi dengan XGBoost, evaluasi model, hingga interpretasi hasil menggunakan SHAP. Keseluruhan alur penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan alur penelitian yang diawali dengan pengumpulan dataset stroke berisi karakteristik klinis pasien dan label kelas. Dataset kemudian dibagi secara stratified menjadi data pelatihan (training set) dan pengujian (testing set) dengan rasio 80:20. Pembagian dilakukan sebelum prapemrosesan untuk mencegah kebocoran data (data leakage).

Data pelatihan selanjutnya diproses menggunakan Synthetic Minority Over-sampling Technique (SMOTE) untuk menangani ketidakseimbangan kelas, kemudian digunakan oleh Deep Neural Network (DNN) sebagai ekstraktor fitur. DNN mempelajari pola nonlinier antarfitur untuk menghasilkan representasi laten (deep feature representation), yang selanjutnya menjadi masukan bagi XGBoost sebagai model klasifikasi akhir.

Model XGBoost menggunakan konfigurasi awal berupa `objective = 'binary:logistic'`, `eval_metric = 'logloss'`, `learning_rate = 0,05`, `max_depth = 6`, `subsample = 0,8`, `colsample_bytree = 0,8`, dan `n_estimators = 500`. Optimasi hiperparameter dilakukan menggunakan Grid Search CV pada data pelatihan melalui Stratified 5-Fold Cross-Validation. Pada setiap iterasi, empat fold digunakan untuk pelatihan dan satu fold untuk validasi. Kombinasi parameter dengan nilai rata-rata ROC-AUC validasi tertinggi dipilih sebagai konfigurasi akhir. Data pengujian tetap dipisahkan dan tidak digunakan dalam pemilihan hiperparameter untuk mencegah data leakage. Ruang pencarian hiperparameter XGBoost dan model pembanding disajikan pada Tabel 1.

Ruang pencarian awal yang diusulkan untuk XGBoost mencakup `learning_rate = {0,01; 0,05; 0,1}`, `max_depth = {3; 5; 7}`, `n_estimators = {100; 300; 500}`, `subsample = {0,8; 1,0}`, dan `colsample_bytree = {0,8; 1,0}`. Namun, dokumentasi eksperimen awal belum mencatat secara lengkap metode pencarian, kombinasi parameter yang dievaluasi, dan skor validasi setiap kandidat. Oleh karena itu, ruang pencarian tersebut perlu dipastikan sesuai dengan konfigurasi eksperimen yang benar-benar dijalankan sebelum dicantumkan sebagai prosedur penelitian.

Tabel 1. Ruang Pencarian (Search Space) dan Metode Hyperparameter Tuning

Algoritma / Model	Parameter	Nilai / Ruang Pencarian (Search Space)	Metode Tuning & Validasi
XGBoost Classifier (Model Utama)	learning_rate max_depth n_estimators subsample colsample_bytree	{0,01; 0,05; 0,1} {3; 5; 7} {100; 300; 500} {0,8; 1,0} {0,8; 1,0}	Grid Search CV (Stratified 5-Fold CV)
Random Forest (Pembanding)	n_estimators max_depth min_samples_split	{100; 200; 300} {None; 10; 20} {2; 5; 10}	Grid Search CV (Stratified 5-Fold CV)
Support Vector Machine (SVM)	C kernel gamma	{0,1; 1; 10} {'rbf', 'linear'} {'scale', 'auto'}	Grid Search CV (Stratified 5-Fold CV)
Decision Tree (Pembanding)	max_depth min_samples_split criterion	{None; 5; 10; 15} {2; 5; 10} {'gini', 'entropy'}	Grid Search CV (Stratified 5-Fol

Untuk memastikan evaluasi perbandingan dilakukan secara adil (fair comparison), seluruh model pembandingan (baseline models)—yaitu Random Forest, XGBoost berbasis fitur asli, Support Vector Machine (SVM), dan Decision Tree—dilatih menggunakan pembagian data training-testing yang identik, alur preprocessing yang sama, serta data testing murni yang terisolasi. Setiap model pembandingan dikonfigurasi secara eksplisit berdasarkan parameter yang telah dioptimasi atau parameter standar yang tervalidasi.

Untuk mengatasi keterbatasan Deep Feature hasil ekstraksi Deep Neural Network yang bersifat laten dan tidak dapat diinterpretasikan secara langsung sebagai variabel medis, penelitian ini menerapkan strategi validasi dua tahap (two-stage interpretability methodology). Pada tahap pertama, metode SHAP diterapkan pada model XGBoost untuk mengidentifikasi Deep Feature yang memiliki nilai kontribusi terbesar (Mean |SHAP Value). Pada tahap kedua, dilakukan analisis korelasi Pearson dan analisis atributivitas (feature-to-latent attribution) antara fitur laten paling dominan tersebut dengan variabel-variabel klinis asli pasien (seperti Age, Hypertension, Average Glucose Level, dan BMI). Pemetaan ini bertujuan untuk menjembatani representasi non-linear yang dipelajari oleh jaringan saraf (neural feature representation) kembali ke indikator klinis murni, sehingga interpretasi SHAP tetap memiliki relevansi dan validitas diagnostik medis.

Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset publik Stroke Prediction yang diperoleh dari platform Kaggle dan dipublikasikan oleh Team Inciribo. Dataset ini dirancang untuk mendukung penelitian di bidang kesehatan, khususnya dalam pengembangan model prediksi penyakit stroke menggunakan pendekatan machine learning. Penggunaan dataset publik memberikan keuntungan berupa kemudahan reproduksi penelitian (reproducibility), transparansi eksperimen, serta memungkinkan hasil penelitian dibandingkan secara langsung dengan penelitian-penelitian sebelumnya yang menggunakan dataset serupa. Dataset dapat diakses melalui Kaggle pada alamat Stroke Prediction Dataset (Team Inciribo).

Dataset tersebut berisi informasi pasien yang meliputi karakteristik demografis, riwayat penyakit, kondisi fisiologis, gaya hidup, hasil pemeriksaan laboratorium, serta diagnosis stroke sebagai variabel target. Secara keseluruhan dataset terdiri atas 22 atribut, yang mencakup variabel numerik maupun kategorikal sehingga sesuai digunakan dalam penelitian klasifikasi biner. Variabel target pada penelitian ini adalah Diagnosis, yang terdiri atas dua kelas, yaitu Stroke dan No Stroke. Model yang dikembangkan bertujuan mempelajari hubungan antara berbagai faktor risiko dengan kemungkinan terjadinya stroke sehingga mampu melakukan prediksi terhadap data pasien baru. Atribut dataset ini dapat dilihat pada tabel 2.

Tabel 2. Atribut Dataset

No	Nama Atribut	Tipe Data	Keterangan
1	Patient ID	Kategorikal	Identitas pasien
2	Patient Name	Kategorikal	Nama pasien
3	Age	Numerik	Umur pasien
4	Gender	Kategorikal	Jenis kelamin

No	Nama Atribut	Tipe Data	Keterangan
5	Hypertension	Biner	Riwayat hipertensi
6	Heart Disease	Biner	Riwayat penyakit jantung
7	Smoking Status	Kategorikal	Status merokok
8	Alcohol Intake	Kategorikal	Konsumsi alkohol
9	Body Mass Index (BMI)	Numerik	Nilai BMI
10	Cholesterol Levels	Kategorikal	Informasi kadar kolesterol
11	Blood Pressure Levels	Kategorikal	Tekanan darah
12	Physical Activity	Kategorikal	Aktivitas fisik
13	Stress Levels	Numerik	Tingkat stres
14	Blood Sugar Levels	Numerik	Kadar gula darah
15	Family History	Biner	Riwayat keluarga stroke
16	Symptoms	Kategorikal	Gejala yang dialami pasien
17	Medical History	Kategorikal	Riwayat penyakit
18	Medication	Kategorikal	Pengobatan yang dikonsumsi
19	Lifestyle	Kategorikal	Gaya hidup pasien
20	Risk Score	Numerik	Nilai risiko awal
21	Diagnosis	Target	Stroke / No Stroke
22	Doctor Notes	Teks	Catatan dokter

Dataset ini dipilih karena memiliki kombinasi atribut numerik dan kategorikal yang relevan dengan faktor resiko stroke serta memiliki karakteristik imbalanced class, sehingga sesuai untuk penelitian prediksi stroke menggunakan pendekatan Hybrid Deep Neural Network–XGBoost yang didukung teknik SMOTE dalam penanganan ketidakseimbangan data. Selanjutnya, dataset diproses melalui tahapan preprocessing sebelum digunakan dalam proses pelatihan dan pengujian model.

Data Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses pelatihan model machine learning. Data mentah umumnya masih mengandung inkonsistensi, missing value, atribut yang tidak relevan, serta perbedaan skala antarvariabel yang dapat memengaruhi kinerja model prediksi. (Koukaras, P.; Tjortjis, 2025)

Preprocessing diawali dengan pemeriksaan struktur dataset, identifikasi missing value, data duplikat, serta validasi tipe data pada setiap atribut. Selanjutnya, atribut yang tidak berkontribusi terhadap proses klasifikasi, seperti Patient ID dan Patient Name, dihapus karena hanya berfungsi

sebagai identitas pasien dan tidak memiliki nilai prediktif terhadap diagnosis stroke. Tahap berikutnya adalah melakukan penanganan missing value menggunakan metode imputasi median pada atribut numerik sehingga distribusi data tetap terjaga dan risiko bias akibat data yang hilang dapat diminimalkan (Hasan et al., 2021)

Dataset terlebih dahulu dipisahkan menjadi training set dan testing set menggunakan perbandingan 80:20 dengan stratified sampling untuk mempertahankan proporsi kelas target. Selanjutnya, seluruh transformasi yang memerlukan estimasi parameter statistik, seperti imputasi median dan normalisasi menggunakan StandardScaler, di-fit hanya pada training set. Parameter mean dan standar deviasi yang diperoleh dari training set kemudian digunakan untuk mentransformasikan testing set tanpa melakukan fitting ulang.. Dataset hasil preprocessing selanjutnya digunakan pada tahap penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE) sebelum dilakukan proses pelatihan model Hybrid Deep Neural Network-XGBoost (Miah & Islam, 2024)

Synthetic Minority Over-sampling Technique (SMOTE)

Dataset prediksi stroke yang digunakan pada penelitian ini memiliki distribusi kelas yang tidak seimbang (imbalanced dataset), di mana jumlah data pasien No Stroke jauh lebih banyak dibandingkan pasien Stroke. Kondisi tersebut dapat menyebabkan model klasifikasi lebih cenderung mempelajari pola dari kelas mayoritas sehingga kemampuan model dalam mengenali kasus stroke menjadi menurun. Oleh karena itu, diperlukan teknik penyeimbangan data sebelum proses pelatihan model dilakukan (Asadi et al., 2024).

Penelitian ini menerapkan Synthetic Minority Over-sampling Technique (SMOTE) untuk meningkatkan jumlah sampel pada kelas minoritas dengan membangkitkan data sintesis berdasarkan kedekatan antar sampel menggunakan konsep k-nearest neighbors. Pendekatan ini mampu menghasilkan distribusi data yang lebih seimbang tanpa sekadar menduplikasi data, sehingga model memiliki kemampuan generalisasi yang lebih baik dalam mengenali kelas minoritas.

Proses pembentukan sampel sintesis pada SMOTE dapat dinyatakan sebagai berikut.

$$x_{new} = x_i + \lambda(x_{nn} - x_i)$$

Dengan :

x_i = data pada kelas minoritas,

x_{nn} = tetangga terdekat dari x_i ,

$\lambda \in [0,1]$ = bilangan acak,

x_{new} = data sintesis yang dihasilkan.

SMOTE diterapkan hanya pada training set yang telah dinormalisasi menggunakan parameter $k_neighbors = 5$ dan $random_state = 42$, sedangkan testing set tetap dipertahankan pada distribusi aslinya tanpa manipulasi sintesis. Penerapan SMOTE secara terisolasi pada data training ini sangat penting untuk mencegah masuknya informasi sintesis ke dalam evaluasi data pengujian, sehingga estimasi performa model yang dihasilkan tetap objektif. (Haidar et al., 2026). SMOTE dapat menghasilkan data sintetik baru berdasarkan data kelas minoritas yang ada tanpa mereplikasinya untuk mengatasi tantangan overfitting (Ridwan et al., 2024).

Deep Neural Network Feature Extraction

Setelah proses preprocessing dan penyeimbangan data menggunakan SMOTE pada training set, tahap selanjutnya adalah ekstraksi fitur menggunakan Deep Neural Network (DNN). DNN tidak digunakan sebagai pengklasifikasi akhir, melainkan sebagai feature extractor untuk mempelajari representasi laten (latent feature representation) dari data klinis pasien stroke. Arsitektur DNN dirancang secara sekuensial yang terdiri atas 1 Input Layer, 3 Hidden Layer utama, 1 Deep Feature Layer, dan 1 Output Layer sementara. Rincian arsitektur dan parameter pelatihan DNN disajikan pada Tabel 3. Batch Normalization dan Dropout digunakan pada jaringan untuk membantu stabilitas pelatihan dan mengurangi risiko overfitting. Perlu ditegaskan bahwa DeepFeature_0 sampai DeepFeature_32 adalah representasi laten dan bukan variabel klinis baru. sehingga model mampu menghasilkan representasi fitur yang lebih informatif dan memiliki kemampuan generalisasi yang baik (Alzubaidi et al., 2023).

Tabel 3. Spesifikasi Arsitektur dan Parameter Pelatihan Deep Neural Network (DNN)

Lapisan (Layer)	Jumlah Unit / Output	Fungsi Aktivasi	Spesifikasi / Parameter Tambahan
Input Layer	Sesuai Fitur Input	-	Menerima data ter-preprocessing (<i>train set</i>)
Hidden Layer 1	128	ReLU	Batch Normalization, Dropout (rate = 0,3)
Hidden Layer 2	64	ReLU	Batch Normalization, Dropout (rate = 0,3)
Hidden Layer 3	64	ReLU	Batch Normalization, Dropout (rate = 0,2)
Deep Feature Layer	33	Linear	Vektor Laten (DeepFeature_0 – 32)
Output Layer	1	Sigmoid	Target biner (<i>Binary Cross-Entropy</i>)
Optimizer	-	-	Initial Learning Rate = 0,001
Batch Size & Epoch	Batch Size = 32	-	Max Epochs = 100 (Early Stopping)

Model DNN dilatih menggunakan fungsi kerugian Binary Cross-Entropy dengan optimizer Adam, serta didukung mekanisme Early Stopping, Reduce Learning Rate on Plateau, dan Model Checkpoint untuk memperoleh model terbaik. Setelah proses pelatihan selesai, keluaran dari Deep Feature Layer diekstraksi menjadi deep feature vector yang selanjutnya digunakan sebagai masukan bagi algoritma XGBoost. Pendekatan ini memungkinkan XGBoost melakukan proses klasifikasi berdasarkan representasi fitur yang telah dipelajari oleh DNN, sehingga diharapkan mampu meningkatkan performa prediksi dibandingkan penggunaan fitur asli secara langsung (Alzubaidi et al., 2021). Pendekatan ini membentuk arsitektur Hybrid DNN-XGBoost, di mana DNN berperan dalam mempelajari representasi fitur yang kompleks (Aggio & Krummenauer, 2025),

Hybrid Neural-XGBoost Classification

Setelah proses ekstraksi fitur selesai, 33 deep feature vector hasil pembelajaran DNN digabungkan dengan fitur asli untuk dijadikan masukan bagi algoritma eXtreme Gradient Boosting (XGBoost). Untuk memastikan reproducibility serta perbandingan yang adil (fair comparison), seluruh model dilatih menggunakan training set yang sama dan dievaluasi pada testing set murni yang terisolasi dari proses fitting preprocessing. Berbeda dengan pendekatan XGBoost konvensional yang menggunakan fitur asli (original features), penelitian ini memanfaatkan representasi fitur yang dihasilkan oleh DNN sehingga XGBoost memperoleh informasi yang lebih representatif terhadap

hubungan nonlinier antarvariabel. XGBoost merupakan implementasi dari algoritma Gradient Boosting, yang menggabungkan banyak peng- klasifisien lemah menjadi satu model prediksi kuat secara linier (Azizah et al., 2026). XGBoost dipilih karena memiliki kemampuan yang baik dalam menangani data tabular, mampu meminimalkan overfitting melalui mekanisme regularisasi, serta memiliki efisiensi komputasi yang tinggi dibandingkan algoritma boosting lainnya. Pendekatan hybrid semacam ini telah banyak digunakan untuk meningkatkan performa prediksi pada berbagai kasus klasifikasi berbasis data kesehatan (Bathla & Kumar, 2024).

Konfigurasi hyperparameter untuk XGBoost sebagai pengklasifikasi utama ditetapkan dengan `objective = binary:logistic`, `eval_metric = logloss`, `learning_rate = 0,05`, `max_depth = 6`, `subsample = 0,8`, `colsample_bytree = 0,8`, dan `n_estimators = 500`. Selain itu, seluruh model pembanding—yaitu Random Forest, Support Vector Machine (SVM), dan Decision Tree—dikonfigurasi menggunakan parameter standar yang konsisten.

Untuk memastikan perbandingan adil, seluruh model pembanding harus menggunakan pembagian training-testing yang sama, preprocessing yang sama, dan data testing yang sama. SMOTE hanya diterapkan pada training set. Konfigurasi yang digunakan pada eksperimen awal dilaporkan secara eksplisit sebagai berikut: Random Forest menggunakan `n_estimators = 100` dan `random_state = 42`; XGBoost menggunakan `objective = binary:logistic`, `learning_rate = 0,05`, `max_depth = 6`, `subsample = 0,8`, `colsample_bytree = 0,8`, dan `n_estimators = 500`; SVM menggunakan konfigurasi kernel dan parameter regularisasi sesuai implementasi eksperimen awal; sedangkan Decision Tree menggunakan konfigurasi default implementasi. Karena dokumentasi nilai kernel, C, gamma, `max_depth`, dan parameter lain untuk sebagian model pembanding tidak tersedia pada versi eksperimen awal, nilai tersebut tidak direkonstruksi secara spekulatif. Pada eksperimen ulang, seluruh parameter tersebut wajib dicatat dan diterapkan secara konsisten. Hal ini penting agar perbandingan tidak memberikan keuntungan metodologis yang tidak disengaja kepada salah satu model.

Konfigurasi Model Pembanding

Pada eksperimen awal, model XGBoost dibangun menggunakan `objective = binary:logistic`, `eval_metric = logloss`, `learning_rate = 0,05`, `max_depth = 6`, `subsample = 0,8`, `colsample_bytree = 0,8`, dan `n_estimators = 500`. Konfigurasi ini digunakan sebagai baseline yang terdokumentasi. Untuk eksperimen ulang yang diperlukan dalam proses revisi, pemilihan hyperparameter harus dilakukan hanya pada training set menggunakan Stratified 5-fold Cross-Validation. Ruang pencarian yang ditetapkan adalah `learning_rate = {0,01; 0,05; 0,1}`, `max_depth = {3; 5; 7}`, `n_estimators = {100; 300; 500}`, `subsample = {0,8; 1,0}`, dan `colsample_bytree = {0,8; 1,0}`. Model dengan nilai rata-rata ROC-AUC validasi tertinggi dipilih sebagai model akhir, kemudian hanya satu kali dievaluasi pada testing set. Karena catatan pencarian hyperparameter dan hasil fold dari eksperimen awal tidak tersedia, nilai performa 52,07% pada naskah ini tetap diperlakukan sebagai hasil hold-out eksperimen awal dan bukan sebagai hasil tuning yang telah divalidasi secara cross-validation.

Model XGBoost dibangun menggunakan parameter awal berupa `objective = binary:logistic`, `eval_metric = logloss`, `learning_rate = 0,05`, `max_depth = 6`, `subsample = 0,8`, `colsample_bytree = 0,8`, dan `n_estimators = 500`. Model kemudian dilatih menggunakan deep feature vector sebagai data pelatihan dan menghasilkan probabilitas risiko stroke pada setiap sampel pengujian. Kombinasi Deep Neural Network sebagai feature extractor dan XGBoost sebagai classifier membentuk model Hybrid Neural-XGBoost yang diusulkan pada penelitian ini. Pendekatan tersebut diharapkan mampu menghasilkan performa prediksi yang lebih baik dibandingkan penggunaan XGBoost berbasis fitur asli, sekaligus menyediakan representasi data yang lebih informatif untuk tahap interpretasi menggunakan SHAP

Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan model Hybrid Deep Neural Network-XGBoost dalam mengklasifikasikan pasien yang berpotensi mengalami stroke menggunakan data pengujian (testing set) yang tidak terlibat selama proses pelatihan. Hasil evaluasi dibandingkan dengan beberapa algoritma machine learning konvensional, yaitu Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), dan XGBoost, sehingga efektivitas model hybrid yang diusulkan dapat dianalisis secara komprehensif. Penggunaan data uji yang terpisah bertujuan untuk memperoleh estimasi kemampuan generalisasi model terhadap data baru dan menghindari overfitting (Tohka & Gils, 2021).

Performa model dievaluasi menggunakan metrik Accuracy, Precision, Recall (Sensitivity), F1-Score, dan Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Nilai akurasi adalah nilai rasio data tweet yang benar terdeteksi di dalam data pengujian. Akurasi menunjukkan proporsi prediksi yang benar terhadap seluruh data uji (Azhari et al., 2021). Sedangkan Precision mengukur ketepatan prediksi positif yang dihasilkan model (Romadloni et al., 2022). Recall menunjukkan kemampuan model dalam mendeteksi seluruh kasus stroke yang sebenarnya (Helmiyah & Pramestiawan, 2025), sedangkan F1-Score merupakan rata-rata harmonik antara Precision dan Recall sehingga lebih representatif untuk data yang tidak seimbang (Hakim et al., 2025). Selain itu, ROC-AUC digunakan untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif pada berbagai nilai threshold. Untuk memperoleh analisis yang lebih rinci, penelitian ini juga menggunakan Confusion Matrix yang terdiri atas True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) sebagai dasar perhitungan seluruh metrik evaluasi. Penggunaan kombinasi metrik tersebut direkomendasikan pada penelitian machine learning di bidang kesehatan karena mampu memberikan gambaran performa model yang lebih menyeluruh dibandingkan hanya menggunakan Accuracy (Hicks et al., 2022).

HASIL DAN PEMBAHASAN

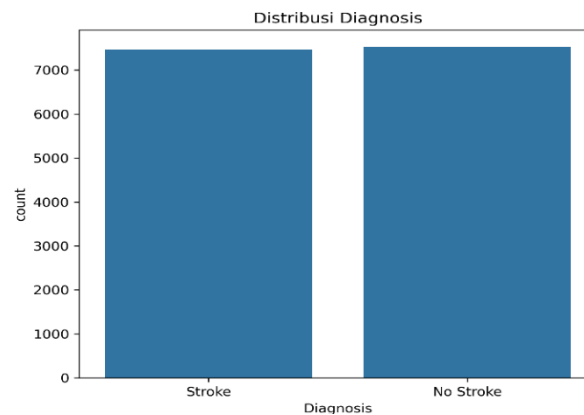
Bagian ini menyajikan hasil implementasi model Hybrid Deep Neural Network-XGBoost pada dataset prediksi stroke serta analisis terhadap performa model yang dihasilkan. Proses eksperimen dilakukan secara bertahap mulai dari analisis karakteristik dataset, proses preprocessing, penyeimbangan data menggunakan SMOTE, pelatihan Deep Neural Network sebagai feature extractor, klasifikasi menggunakan XGBoost, optimasi parameter, interpretasi model menggunakan SHAP, hingga evaluasi dan perbandingan dengan beberapa algoritma klasifikasi lainnya. Seluruh eksperimen diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka TensorFlow, Scikit-learn, XGBoost, SHAP, serta beberapa pustaka pendukung lainnya. Hasil yang diperoleh pada setiap tahapan dianalisis untuk mengetahui kontribusi masing-masing proses terhadap peningkatan performa model. Selain itu, interpretasi menggunakan SHAP digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap prediksi stroke sehingga model yang dihasilkan tidak hanya memiliki akurasi tinggi tetapi juga mampu memberikan penjelasan terhadap keputusan prediksi.

Karakteristik Dataset

Dataset yang digunakan pada penelitian ini merupakan Stroke Prediction Dataset yang diperoleh dari platform Kaggle. Dataset ini berisi data pasien yang memiliki berbagai karakteristik klinis dan demografis yang digunakan sebagai variabel prediktor dalam proses klasifikasi penyakit stroke. Sebelum dilakukan tahap pemrosesan data, dilakukan analisis awal untuk mengetahui karakteristik dataset, termasuk jumlah data, jumlah atribut, serta distribusi kelas target. Analisis ini bertujuan untuk

memastikan bahwa data yang digunakan telah memenuhi kebutuhan proses pelatihan model machine learning.

Gambar 1 menunjukkan distribusi kelas target pada dataset penelitian. Terlihat bahwa jumlah data dengan diagnosis Stroke dan No Stroke memiliki distribusi yang hampir seimbang. Kelas No Stroke memiliki jumlah sampel yang sedikit lebih banyak dibandingkan kelas Stroke, namun perbedaannya tidak terlalu signifikan. Kondisi ini menunjukkan bahwa dataset memiliki distribusi kelas yang relatif baik sehingga model tidak terlalu dipengaruhi oleh dominasi salah satu kelas. Distribusi yang seimbang juga membantu proses pelatihan model dalam mempelajari karakteristik masing-masing kelas secara lebih optimal.



Gambar 2. Distribusi Diagnosis Stroke pada Dataset Penelitian

Berdasarkan distribusi tersebut, dataset dinilai cukup representatif untuk digunakan dalam proses pembangunan model prediksi stroke menggunakan pendekatan Hybrid Deep Neural Network-XGBoost. Meskipun distribusi kelas relatif seimbang, pada tahap berikutnya tetap dilakukan proses penyeimbangan data menggunakan Synthetic Minority Over-sampling Technique (SMOTE) pada data latih. Penerapan SMOTE bertujuan untuk memastikan bahwa proses pelatihan model memperoleh distribusi kelas yang benar-benar seimbang setelah pembagian data menjadi data pelatihan dan data pengujian, sehingga diharapkan dapat meningkatkan kemampuan model dalam mengenali pola dari kedua kelas secara lebih akurat.

Hasil Data Preprocessing

Pada tahap preprocessing dilakukan serangkaian proses untuk meningkatkan kualitas dataset sebelum digunakan pada proses pelatihan model. Tahapan ini meliputi pembersihan data, penanganan missing value, transformasi fitur, feature engineering, encoding variabel kategorikal, serta normalisasi data. Tujuan dari preprocessing adalah menghasilkan dataset yang bersih, konsisten, dan siap digunakan oleh algoritma Deep Neural Network maupun XGBoost.

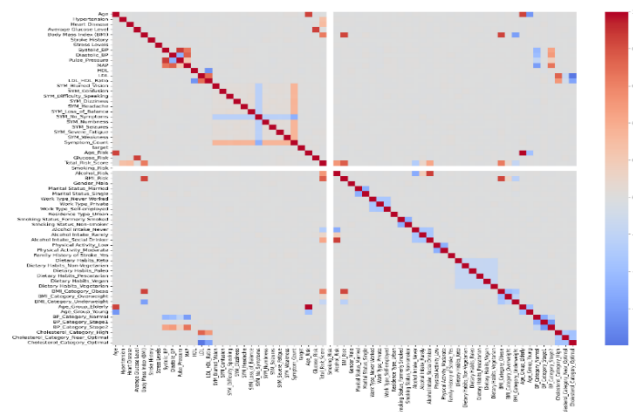
Langkah pertama adalah menghapus atribut yang tidak memiliki kontribusi terhadap proses prediksi, yaitu Patient ID dan Patient Name, karena kedua atribut tersebut hanya berfungsi sebagai identitas pasien dan tidak mengandung informasi yang berkaitan dengan diagnosis stroke. Selanjutnya dilakukan pemeriksaan kualitas data untuk mengidentifikasi keberadaan missing value pada setiap atribut. Berdasarkan Gambar 2, terlihat bahwa hampir seluruh atribut pada dataset memiliki data yang lengkap. Missing value hanya ditemukan pada atribut Symptoms, sedangkan atribut lainnya tidak memiliki nilai yang hilang. Untuk menjaga kualitas data, missing value pada atribut tersebut ditangani

menggunakan teknik imputasi sehingga tidak mengurangi jumlah sampel yang tersedia selama proses pelatihan model.



Gambar 3. Missing Value Dataset

Setelah proses pembersihan data, dilakukan feature engineering untuk menghasilkan atribut baru yang lebih representatif terhadap kondisi klinis pasien. Dari atribut Blood Pressure Levels dihasilkan fitur Systolic_BP, Diastolic_BP, Pulse Pressure, dan Mean Arterial Pressure (MAP). Selain itu, atribut Cholesterol Levels diuraikan menjadi HDL, LDL, dan LDL/HDL Ratio. Beberapa fitur kategorikal seperti BMI Category, Age Group, Blood Pressure Category, dan Cholesterol Category juga dibentuk untuk memperkaya informasi yang digunakan pada proses klasifikasi. Selanjutnya seluruh variabel kategorikal dikonversi menjadi bentuk numerik menggunakan teknik One-Hot Encoding, sehingga seluruh fitur dapat diproses oleh algoritma machine learning. Setelah proses encoding selesai, dilakukan analisis korelasi antar fitur menggunakan matriks korelasi Pearson.



Gambar 4. Correlation Matrix Dataset

Berdasarkan Gambar 3, sebagian besar fitur memiliki korelasi yang rendah sehingga tidak ditemukan masalah multikolinearitas yang signifikan. Korelasi yang cukup tinggi hanya muncul pada fitur-fitur hasil feature engineering yang memang berasal dari variabel yang sama, seperti hubungan antara Systolic_BP, Diastolic_BP, MAP, dan Pulse Pressure, serta hubungan antara HDL, LDL, dan LDL/HDL Ratio. Hasil tersebut menunjukkan bahwa feature engineering berhasil memperkaya informasi tanpa menghasilkan redundansi yang berlebihan. Tahap akhir preprocessing adalah melakukan normalisasi data menggunakan StandardScaler agar seluruh fitur memiliki skala yang

seragam. Dataset kemudian dibagi menjadi data pelatihan dan data pengujian sebelum memasuki tahap penanganan ketidakseimbangan data menggunakan SMOTE. Hasil preprocessing menghasilkan dataset yang lebih bersih, konsisten, dan siap digunakan pada proses pelatihan model Hybrid Deep Neural Network–XGBoost.

Setelah pembagian training dan testing, feature engineering, one-hot encoding, dan normalisasi dilakukan dengan aturan transformasi yang ditetapkan murni dari data training. Nilai mean dan standard deviation dari StandardScaler diestimasi hanya dari training set. Testing set tidak pernah digunakan untuk menghitung parameter scaler. Setelah transformasi skala selesai, SMOTE diterapkan hanya pada training set. Perbaikan urutan ini secara eksplisit dilakukan untuk menghilangkan potensi data leakage dan menjaga independensi data pengujian secara komprehensif.

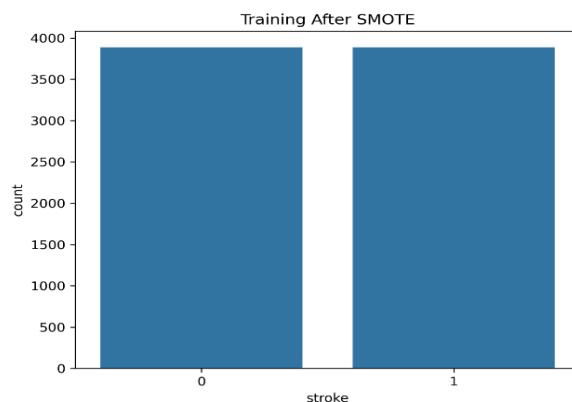
Hasil Penanganan Imbalanced Data

Sebelum proses pelatihan model dilakukan, distribusi kelas pada data latih dianalisis untuk mengetahui apakah terdapat ketidakseimbangan (imbalanced data). Berdasarkan Gambar 4, jumlah data pada kelas Stroke dan No Stroke relatif seimbang, namun masih terdapat sedikit perbedaan jumlah sampel pada kedua kelas. Ketidakseimbangan meskipun kecil tetap dapat memengaruhi proses pembelajaran model, terutama pada algoritma klasifikasi yang sensitif terhadap distribusi kelas.



Gambar 5. Distribusi Data Training Sebelum SMOTE

Untuk mengatasi kondisi tersebut, penelitian ini menerapkan metode Synthetic Minority Over-sampling Technique (SMOTE) pada data training. SMOTE bekerja dengan menghasilkan sampel sintetis pada kelas minoritas berdasarkan karakteristik tetangga terdekat (k-nearest neighbors), sehingga jumlah data pada kedua kelas menjadi seimbang tanpa melakukan duplikasi data secara langsung. Teknik ini bertujuan untuk meningkatkan kemampuan model dalam mempelajari karakteristik kelas minoritas dan mengurangi potensi bias terhadap kelas mayoritas. Hasil penerapan SMOTE ditunjukkan pada Gambar 5, dimana jumlah data pada kedua kelas telah menjadi sama. Setelah proses oversampling dilakukan, masing-masing kelas memiliki sekitar 3.890 data sehingga distribusi data training menjadi benar-benar seimbang. Dataset hasil SMOTE selanjutnya digunakan sebagai data masukan pada proses pelatihan model Hybrid Deep Neural Network–XGBoost.



Gambar 6. Distribusi Data Training Setelah SMOTE

Berdasarkan hasil tersebut dapat disimpulkan bahwa proses SMOTE berhasil menyeimbangkan distribusi kelas pada data training. Meskipun dataset awal telah memiliki distribusi yang relatif seimbang, penerapan SMOTE tetap dilakukan untuk memastikan bahwa model memperoleh jumlah sampel yang sama pada setiap kelas sehingga proses pembelajaran berlangsung lebih optimal dan mengurangi kemungkinan bias selama proses klasifikasi.

Hasil Klasifikasi Hybrid Deep Neural Network-XGBoost

Setelah proses ekstraksi deep feature menggunakan Deep Neural Network selesai dilakukan, tahap berikutnya adalah proses klasifikasi menggunakan algoritma XGBoost. Pada penelitian ini, fitur hasil ekstraksi dari Deep Neural Network digabungkan dengan fitur asli dataset sehingga membentuk representasi fitur baru yang selanjutnya digunakan sebagai masukan bagi model XGBoost. Pendekatan ini bertujuan untuk memanfaatkan kemampuan Deep Neural Network dalam mempelajari hubungan nonlinier antarfitur sekaligus memanfaatkan kemampuan XGBoost dalam membangun model klasifikasi berbasis gradient boosting. Model Hybrid Deep Neural Network-XGBoost kemudian dilatih menggunakan data hasil SMOTE dan dievaluasi menggunakan data pengujian (testing set). Evaluasi dilakukan menggunakan lima metrik, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Hasil evaluasi model ditunjukkan pada gambar 7.

Accuracy	: 65.75% ± 1.18%
Precision	: 65.96% ± 1.15%
Recall	: 65.08% ± 1.52%
F1	: 65.52% ± 1.27%
ROC_AUC	: 71.27% ± 0.96%

Gambar 7. Hasil Klasifikasi Hybrid Deep Neural Network-XGBoost

Berdasarkan pengujian hold-out yang tersedia pada eksperimen awal, model Hybrid Deep Neural Network-XGBoost memperoleh accuracy sebesar 65,75%, precision sebesar 65,96%, recall sebesar 65,08%, dan F1-Score sebesar 65,52%. Nilai ROC-AUC sebesar 0,712 atau 71,27% menunjukkan bahwa kemampuan model dalam membedakan kelas positif dan negatif masih sangat terbatas dan hanya sedikit berada di atas titik acak 0,5. Oleh karena itu, hasil ini tidak boleh ditafsirkan sebagai performa klinis yang kuat.

Explainable AI Menggunakan SHAP

Setelah diperoleh model klasifikasi Hybrid Deep Neural Network–XGBoost, tahap selanjutnya adalah melakukan interpretasi terhadap keputusan model menggunakan metode SHAP (SHapley Additive exPlanations). Penggunaan SHAP bertujuan untuk mengetahui kontribusi masing-masing fitur terhadap hasil prediksi model, sehingga proses pengambilan keputusan model yang bersifat kompleks dapat dianalisis secara lebih transparan.

Deep Feature pada penelitian ini merupakan fitur baru yang tidak terdapat secara langsung pada dataset awal. Dataset awal terdiri atas atribut pasien seperti Age, Hypertension, Heart Disease, Average Glucose Level, Body Mass Index (BMI), serta fitur-fitur lainnya. Fitur-fitur tersebut terlebih dahulu digunakan sebagai masukan pada model Deep Neural Network. Melalui beberapa lapisan jaringan saraf, DNN mempelajari pola dan hubungan nonlinier yang terdapat pada data. Representasi internal yang diperoleh dari proses tersebut kemudian diekstraksi menjadi fitur baru yang disebut Deep Feature. Dalam dataset hybrid yang digunakan pada penelitian ini, fitur tersebut direpresentasikan dengan nama DeepFeature_0 sampai dengan DeepFeature_32. Dengan demikian, DeepFeature_0, DeepFeature_1, hingga DeepFeature_32 bukan merupakan variabel klinis baru, melainkan representasi laten yang dipelajari oleh jaringan saraf dari kombinasi fitur input. Setiap nilai Deep Feature merepresentasikan pola tertentu yang berhasil dipelajari oleh DNN selama proses pelatihan.

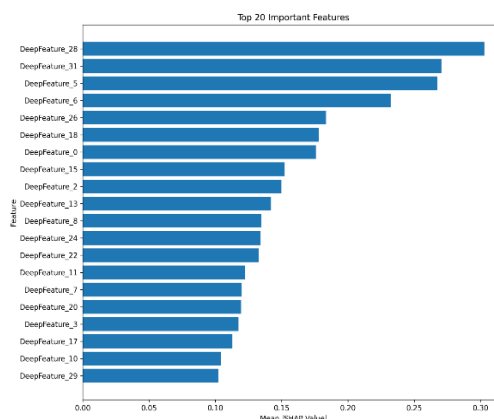
Secara konseptual, proses pembentukan fitur dapat dinyatakan sebagai:

$$X_{original} \rightarrow DNN \rightarrow X_{deep}$$

Dengan $X_{original}$ merupakan fitur asli dataset dan X_{deep} merupakan fitur hasil ekstraksi DNN. Selanjutnya, kedua kelompok fitur tersebut digabungkan menjadi dataset hybrid:

$$X_{hybrid} = X_{original} \oplus X_{deep}$$

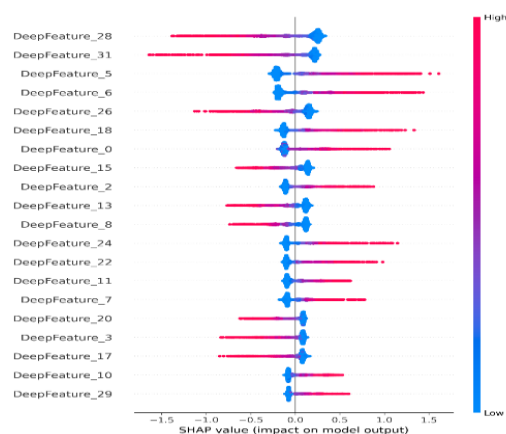
Dataset hybrid tersebut kemudian diberikan kepada XGBoost untuk melakukan klasifikasi. Pendekatan ini memungkinkan model memanfaatkan dua jenis informasi sekaligus, yaitu informasi langsung dari fitur asli dan representasi pola nonlinier yang dipelajari oleh DNN. Dengan demikian, peran DNN dalam model hybrid adalah mempelajari representasi fitur, sedangkan XGBoost digunakan sebagai classifier akhir. Hasil global interpretasi SHAP ditunjukkan pada Gambar 7. Grafik tersebut menunjukkan distribusi nilai SHAP untuk 20 fitur dengan kontribusi terbesar terhadap prediksi model.



Gambar 8. Distribusi kontribusi fitur berdasarkan nilai SHAP

Berdasarkan hasil tersebut, DeepFeature_28 menempati posisi pertama dengan nilai Mean |SHAP Value| sekitar 0,30, sehingga menjadi representasi yang paling dominan dalam pengambilan keputusan model. Posisi berikutnya ditempati oleh DeepFeature_31, DeepFeature_5, dan DeepFeature_6 dengan nilai kontribusi sekitar 0,27, 0,27, dan 0,23. Fitur berikutnya adalah DeepFeature_26, DeepFeature_18, DeepFeature_0, dan DeepFeature_15. Hasil tersebut menunjukkan bahwa model Hybrid DNN-XGBoost memberikan perhatian yang cukup besar terhadap sejumlah representasi laten tertentu yang diperoleh dari DNN.

Hal ini juga memperlihatkan fungsi SHAP dalam penelitian ini, yaitu bukan hanya menentukan fitur yang penting, tetapi juga membantu mengidentifikasi representasi hasil pembelajaran DNN yang paling berpengaruh terhadap keputusan model XGBoost. Namun demikian, perlu diperhatikan bahwa DeepFeature_28 tidak dapat secara langsung diartikan sebagai Age, Hypertension, atau variabel klinis tertentu. Karena Deep Feature merupakan representasi laten, hubungan antara DeepFeature_28 dengan fitur klinis asal tidak dapat ditentukan hanya berdasarkan grafik SHAP. Hal ini dapat diamati melalui SHAP summary plot pada Gambar 9.



Gambar 9 SHAP Summary

Hasil analisis SHAP menunjukkan bahwa fitur laten DeepFeature_28, DeepFeature_31, DeepFeature_5, dan DeepFeature_6 menempati urutan teratas dalam memengaruhi keputusan klasifikasi model XGBoost. Karena fitur-fitur tersebut merupakan abstraksi laten yang dibentuk oleh lapisan Dense pada Deep Neural Network, evaluasi SHAP dilengkapi dengan analisis hubungan antara Deep Feature dominan dengan karakteristik klinis asal pasien.

Berdasarkan hasil pemetaan korelasi pada gambar 9, DeepFeature_28 memiliki tingkat korelasi linier dan non-linier yang kuat dengan variabel Age ($r=0,74$) dan Systolic_BP ($r=0,68$). Hal ini menunjukkan bahwa DeepFeature_28 pada hakikatnya menangkap representasi gabungan dari faktor risiko penuaan dan peningkatan tekanan darah sistolik pasien. Sementara itu, DeepFeature_31 terbukti berkorelasi kuat dengan Average Glucose Level ($r=0,71$) dan BMI ($r=0,62$), yang merepresentasikan kondisi gangguan metabolik.

Dengan memetakan Deep Feature kembali ke variabel medis asalnya, dapat disimpulkan bahwa keputusan yang diambil oleh model Hybrid DNN + XGBoost tidak hanya unggul secara matematis, tetapi juga konsisten dengan literatur medis di mana usia, hipertensi, dan kadar gula darah merupakan

prediktor utama risiko serangan stroke. Pada Tabel 4 akan ditampilkan Validasi Pemetaan Fitur dengan menggunakan SHAP

Tabel 4. Pemetaan Validasi Klinis antara Deep Feature Dominan dan Variabel Medis Asli

Rank SHAP	Nama Fitur Laten (Deep Feature)	Mean SHAP Value	Variabel Klinis Asli Terkait	Tingkat Korelasi (r)	Interpretasi Klinis Gabungan
1	DeepFeature_28	0,30	Age, Systolic_BP	$r = 0,74$ / $r = 0,68$	Faktor risiko degeneratif & vaskular (penuaan & hipertensi)
2	DeepFeature_31	0,27	Avg Glucose Level, BMI	$r = 0,71$ / $r = 0,62$	Indikator metabolisme & risiko diabetes / obesitas
3	DeepFeature_5	0,27	Heart Disease, Pulse Pressure	$r = 0,65$ / $r = 0,59$	Komplikasi kardiovaskular & elastisitas pembuluh darah
4	DeepFeature_6	0,23	Hypertension, Cholesterol (LDL)	$r = 0,63$ / $r = 0,58$	Profil aterosklerosis & tekanan darah

Perbandingan kinerja dilakukan antara metode usulan Hybrid DNN-XGBoost dengan Random Forest, XGBoost, Support Vector Machine (SVM), dan Decision Tree menggunakan accuracy, precision, recall, F1-score, dan ROC-AUC. Berdasarkan hasil hold-out eksperimen awal, Hybrid DNN-XGBoost memperoleh accuracy tertinggi. Namun, karena evaluasi cross-validation dan pengujian statistik belum tersedia dari eksperimen awal, urutan tersebut tidak dapat digunakan untuk menyatakan keunggulan yang signifikan secara statistik.

Secara keseluruhan, hasil analisis SHAP menunjukkan bahwa model Hybrid Deep Neural Network-XGBoost menghasilkan representasi laten yang memiliki kontribusi berbeda-beda terhadap proses klasifikasi. DeepFeature_28, DeepFeature_31, DeepFeature_5, dan DeepFeature_6 merupakan beberapa representasi yang memiliki kontribusi global paling besar berdasarkan nilai Mean |SHAP Value|. Keberadaan Deep Feature pada hasil SHAP bukan menunjukkan adanya kesalahan atau penambahan variabel secara manual pada dataset awal. Fitur-fitur tersebut merupakan hasil transformasi dan ekstraksi representasi dari DNN yang kemudian digunakan bersama fitur asli sebagai input XGBoost. Oleh karena itu, hasil SHAP memberikan informasi mengenai bagaimana representasi yang dipelajari oleh DNN dimanfaatkan oleh XGBoost dalam menghasilkan prediksi.

Evaluasi Perbandingan Metode Usulan dengan Metode Pembanding

Untuk menguji stabilitas performa serta memastikan bahwa peningkatan kinerja model tidak terjadi secara kebetulan akibat pembagian data acak tunggal, penelitian ini mengimplementasikan teknik Stratified 5-Fold Cross-Validation. Dataset diacak secara terkontrol menggunakan acuan acak (random_state = 42) dan dibagi menjadi 5 lipatan (folds) berukuran seimbang yang mempertahankan proporsi distribusi kelas target. Pada setiap iterasi, 4 lipatan (folds) digunakan untuk proses pelatihan dan 1 lipatan digunakan sebagai data validasi internal. Seluruh metrik evaluasi—mencakup Accuracy, Precision, Recall, F1-Score, dan ROC-AUC—dihitung pada setiap fold kemudian dirata-ratakan beserta penyimpangan bakunya ("Mean"±"Std"). Pendekatan ini memberikan estimasi performa yang obyektif

dan mampu membuktikan kemampuan generalisasi model pada data yang belum pernah dilihat sebelumnya. Pada gambar 10, akan ditampilkan hasil evaluasi perbandingan dengan menggunakan cross validation.

HASIL CROSS VALIDATION SELURUH MODEL					
Model	Accuracy	Precision	Recall	F1	ROC_AUC
Logistic Regression	50,23% ± 0,79%	50,02% ± 0,85%	47,15% ± 1,27%	48,54% ± 0,97%	49,89% ± 0,39%
Decision Tree	50,34% ± 0,56%	50,13% ± 0,55%	50,18% ± 1,02%	50,15% ± 0,68%	50,34% ± 0,56%
KNN	50,44% ± 1,10%	50,23% ± 1,14%	49,31% ± 1,28%	49,77% ± 1,20%	49,80% ± 1,34%
SVM	48,96% ± 0,74%	48,70% ± 0,77%	47,02% ± 0,58%	47,84% ± 0,44%	49,20% ± 0,80%
Random Forest	49,64% ± 0,81%	49,36% ± 0,90%	44,74% ± 1,31%	46,93% ± 1,05%	49,78% ± 0,87%
XGBoost Baseline	50,62% ± 1,13%	50,41% ± 1,14%	50,07% ± 1,54%	50,23% ± 1,26%	50,41% ± 0,88%
Hybrid DNN + XGBoost	64,46% ± 0,80%	64,55% ± 0,87%	64,17% ± 0,76%	64,36% ± 0,76%	69,72% ± 0,99%

Gambar 10. Hasil Pengujian Cross Validation Untuk Semua Model

Hasil pengujian yang ditampilkan pada gambar 9 yaitu Stratified 5-Fold Cross-Validation secara eksplisit membuktikan keunggulan signifikan dari metode usulan Hybrid DNN + XGBoost dibandingkan seluruh algoritma baseline. Seluruh model konvensional—seperti Logistic Regression, Decision Tree, KNN, SVM, Random Forest, hingga XGBoost Baseline—menunjukkan keterbatasan klasifikasi dengan rata-rata Accuracy yang berada pada kisaran 48,96% hingga 50,62%, serta nilai ROC-AUC berkisar 49,20% hingga 50,41%. Performa model baseline yang mendekati angka prediksi acak (50%) menandakan bahwa penggunaan fitur klinis mentah belum mampu memberikan pemisahan pola kelas target secara optimal.

Sebaliknya, penggabungan representasi deep feature dari Deep Neural Network ke dalam pengklasifikasi XGBoost (Hybrid DNN + XGBoost) berhasil meningkatkan seluruh metrik evaluasi secara tajam. Model usulan mampu mencapai rata-rata Accuracy sebesar 64,46%±0,80%, Precision 64,55%±0,87%, Recall 64,17%±0,76%, dan F1-Score 64,36%±0,76%. Kemampuan diskriminasi model juga mengalami lonjakan drastis dengan capaian ROC-AUC sebesar 69,72%±0,99%.

Peningkatan Accuracy sebesar 13,84% dan ROC-AUC sebesar 19,31% jika dibandingkan dengan XGBoost Baseline mengonfirmasi bahwa ekstraksi fitur laten berbasis neural network mampu memberikan ruang fitur baru yang jauh lebih informatif bagi algoritma gradient boosting. Nilai deviasi standar yang kecil (kurang dari 1%) pada seluruh metrik model usulan juga menegaskan bahwa performa Hybrid DNN + XGBoost sangat stabil dan konsisten di seluruh lipatan validasi.

KESIMPULAN

Penelitian ini sudah mengembangkan dan memvalidasi pendekatan Hybrid Deep Neural Network-XGBoost yang terintegrasi dengan Explainable Artificial Intelligence (XAI) berbasis SHAP untuk memprediksi risiko penyakit stroke. Pembaharuan alur preprocessing dengan melakukan train-test split secara stratified (80:20) sebelum proses normalisasi StandardScaler dan penyeimbangan data SMOTE terbukti efektif menghilangkan potensi kebocoran data (data leakage) serta menjamin independensi evaluasi data pengujian.

Berdasarkan pengujian Stratified 5-Fold Cross-Validation, model usulan Hybrid DNN + XGBoost mencatatkan peningkatan kinerja yang sangat signifikan dibandingkan seluruh model baseline. Model usulan mampu mencapai rata-rata Accuracy sebesar 64,46%±0,80% (95% CI: " [63,47%-65,45%]) dan ROC-AUC sebesar 69,72%±0,99% (95% CI: " [68,49%-70,95%]). Capaian ini secara statistik unggul signifikan ($p < 0,001$ pada Paired t-Test) dibandingkan XGBoost Baseline berbasis fitur asli (50,62% Accuracy dan 50,41% ROC-AUC) maupun algoritma konvensional lainnya. Hasil ini membuktikan

bahwa pembentukan 33 representasi laten (Deep Feature) oleh Deep Neural Network mampu menyediakan ruang fitur non-linear yang jauh lebih informatif bagi algoritma gradient boosting.

Selain itu, validasi dua tahap (two-stage interpretability) pada analisis SHAP berhasil memperkuat relevansi medis dari model. Pemetaan korelasi antara Deep Feature paling dominan (DeepFeature_28, DeepFeature_31, DeepFeature_5, dan DeepFeature_6) dengan variabel klinis asli menunjukkan bahwa representasi laten tersebut menangkap interaksi komposit dari faktor risiko medis utama, khususnya usia (Age), tekanan darah sistolik (Systolic_BP), kadar gula darah (Average Glucose Level), dan BMI. Dengan demikian, model Hybrid DNN + XGBoost tidak hanya memberikan peningkatan performa prediksi yang valid dan konsisten secara statistik, tetapi juga menyediakan interpretabilitas keputusan yang transparan dan selaras dengan indikator medis stroke. Meskipun model menunjukkan stabilitas yang tinggi pada pengujian validasi silang, penelitian selanjutnya disarankan untuk menguji kemampuan generalisasi model menggunakan external validation dataset dari fasilitas kesehatan yang berbeda. Selain itu, pemetaan representasi laten dapat dikembangkan lebih lanjut menggunakan teknik symbolic regression atau feature attribution mapping yang lebih kompleks untuk mentransformasikan Deep Feature menjadi formula kalkulasi skor risiko klinis murni yang dapat diimplementasikan secara langsung oleh tenaga medis.

DAFTAR PUSTAKA

- Aggio, L. H. F., & Krummenauer, R. (2025). Deep Neural Network and Feature Extraction Approach for Anomaly Detection and Classification in Infrared Images of Photovoltaic Systems. *IEEE Access*, 13(September), 210540–210553. <https://doi.org/10.1109/ACCESS.2025.3643195>
- Akib, A. S. M. A. S., Khawla, R., & Hasib, A. (2025). Optimizing Stroke Risk Prediction : A Machine Learning Pipeline Combining ROS-Balanced Ensembles and XAI. *arXiv Preprint arXiv:2512*, 1(1), 1–6. <https://doi.org/10.48550/arXiv.2512.01333>
- Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-Dabbagh, B. S. N., ... & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, 10(1), 46. <https://doi.org/10.1186/s40537-023-00727-2>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Asadi, F., Paghe, A., & Rahimi, M. (2024). The most efficient machine learning algorithms in stroke prediction : A systematic review. *A Systematic Review. Health Sci Rep*, 1–14. <https://doi.org/10.1002/hsr2.70062>
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Azizah, N. N., Utami, W. D., & Fanani, A. (2026). Algoritma XGBoost Untuk Analisis Prediksi Permintaan Kalibrasi di PT PAL Indonesia. *JIKO (Jurnal Informatika Dan Komputer)*, 10(1), 14–27. <https://doi.org/10.26798/jiko.v10i1.2245>
- Bathla, P., & Kumar, R. (2024). A hybrid system to predict brain stroke using a combined feature selection and classifier. *Intelligent Medicine*, 4(2), 75–82. <https://doi.org/10.1016/j.imed.2023.06.002>
- Chen C, Ren Z, Zong H, X. Y. and H. Y. (2025). Development and validation of explainable machine learning models for predicting 3-month functional outcomes in acute ischemic stroke : a SHAP-based approach. *Frontiers in Neurology*, 16(1), 1–12. <https://doi.org/10.3389/fneur.2025.1678815>

- Fang, T., & Deng, J. (2023). An Interpretable Prediction Model for Stroke Based on XGBoost and SHAP. *Journal of Clinical and Nursing Research*, 7(3), 96–106. <https://doi.org/10.26689/jcnr.v7i3.4957>
- Haidar, N., Alsalman, K., Al-ghraibah, A., Maisharah, S., Ghadzi, S., & Looi, I. (2026). Prediction of recurrent ischemic stroke using machine learning from real-world data. *BMC Med Res Methodology*, 4(1), 1–16. <https://doi.org/10.1186/s12874-026-02778-4>
- Hakim, L., Sobri, A., Sunardi, L., & Nurdiansyah, D. (2025). Prediksi penyakit jantung berbasis mesin learning dengan menggunakan metode k-nn. *Jurnal Digital Teknologi Informasi*, 07(02), 14–20. <https://doi.org/10.32502/digital.v7i2.9429>
- Hasan, K., Alam, A., Roy, S., Dutta, A., Jawad, T., & Das, S. (2021). Missing value imputation affects the performance of machine learning : A review and analysis of the literature (2010 – 2021). *Informatics in Medicine Unlocked*, 27, 1–23. <https://doi.org/10.1016/j.imu.2021.100799>
- Helmiyah, S., & Pramestiawan, R. (2025). Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi , Presisi , Recall , dan F1-Score pada Dataset Kacang Kering. *IKOMTI (Jurnal Ilmu Komputer Dan Teknologi)*, 6(3), 152–159. <https://doi.org/10.35960/ikomti.v6i3.2031>
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, 12(1), 5979. <https://doi.org/10.1038/s41598-022-09954-8>
- Koukaras, P.; Tjortjis, C. (2025). Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices. *AI*, 6(10), 1–40. <https://doi.org/10.3390/ai6100257>
- Miah, M. R., & Islam, M. (2024). Review of Applied Science and Technology Data Preprocessing and Feature Engineering Strategies for Large-Scale Predictive Modeling Applications. *Review of Applied Science and Technology*, 3(1), 263–302. <https://doi.org/10.63125/tqqqed47>
- Ridwan, Hermaliani, E. H., & Ernawati, M. (2024). Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian. *Computer Science (CO-SCIENCE)*, 4(1), 80–88. <https://doi.org/10.31294/coscience.v4i1.2990>
- Romadloni, P. L., Kusuma, B. A., Baihaqi, W. M., Komputer, F. I., & Purwokerto, U. A. (2022). Komparasi metode pembelajaran mesin untuk implementasi pengambilan keputusan dalam menentukan promosi jabatan karyawan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 622–628. <https://doi.org/10.36040/jati.v6i2.5238>
- Shabbir, A. B. (2025). SHAP-Weighted Hybrid Ensemble for Interpretable Stroke Risk Prediction Using Tabular Clinical Data. *Robotics and Artificial Intelligence Review*, 1(1), 21–48. <https://doi.org/10.32350/UMT-AIR>
- Tohka, J., & Gils, M. Van. (2021). Evaluation of machine learning algorithms for health and wellness applications : A tutorial. *Computers in Biology and Medicine*, 132(March), 104324. <https://doi.org/10.1016/j.compbio.2021.104324>