



Analisis Pengaruh Preprocessing Data dan Hyperparameter Tuning pada Backpropagation Neural Network dalam Klasifikasi Stroke

Asrul Gunawan*, Arief Hermawan, Donny Avianto

Program Pascasarjana, Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Sleman, Indonesia

Email: ¹: asrulgunawan5671@gmail.com, ²ariefdb@uty.ac.id, ³donny@uty.ac.id

Email Penulis Korespondensi: asrulgunawan5671@gmail.com

Abstrak—Ketidakseimbangan data dan perbedaan skala antar fitur sering menjadi faktor utama yang menurunkan performa model klasifikasi berbasis jaringan saraf. Penelitian ini bertujuan untuk menganalisis pengaruh *preprocessing* data serta *hyperparameter tuning* terhadap kinerja *Backpropagation Neural Network* (BPNN) dalam klasifikasi penyakit stroke. Penelitian ini menggunakan dataset stroke dari platform *Kaggle* yang terdiri dari 5.110 data pasien dengan 10 fitur klinis. Evaluasi dilakukan dengan menggunakan lima skema dan terdiri dari beberapa teknik penyeimbang data. Teknik-teknik tersebut antara lain tanpa penyeimbangan, SMOTE, dan ADASYN. Selain itu, evaluasi juga melibatkan normalisasi data yang mencakup tanpa normalisasi, *MinMaxScaler*, dan *Z-Score*. Model BPNN yang digunakan memiliki arsitektur 19 neuron input, 29 neuron pada hidden layer, dan 1 neuron output. *Hyperparameter tuning* dilakukan dengan mencari nilai terbaik *learning rate* dan jumlah *epoch*. Hasil evaluasi menunjukkan bahwa model pada skema satu memiliki keterbatasan. Keterbatasan model ini sangat terlihat dalam mengidentifikasi kelas stroke. Penerapan SMOTE dan *MinMaxScaler* pada skema dua membuktikan bahwa hasilnya menjadi lebih baik dan kinerjanya meningkat secara signifikan. Di sisi lain, kombinasi ADASYN dan *Z-Score* pada skema tiga menunjukkan kinerja yang lebih stabil dan mampu mendeteksi kasus stroke dengan lebih akurat. Proses *hyperparameter tuning* pada skema empat dan lima juga terbukti membuat kinerja lebih baik. Hasil terbaik didapatkan pada skema kelima, dengan akurasi sebesar 96,47%, presisi sebesar 97,34%, *recall* sebesar 95,62%, dan *F1-score* sebesar 96,47%. Temuan ini menunjukkan bahwa kombinasi teknik *balancing* adaptif, normalisasi berbasis distribusi, serta *tuning* parameter yang optimal sangat efektif dalam meningkatkan akurasi dan stabilitas BPNN untuk klasifikasi stroke.

Kata Kunci: Backpropagation Neural Network; Stroke; Preprocessing Data; Hyperparameter Tuning; Klasifikasi

Abstract—Data imbalance and scale differences between features are often the main factors that reduce the performance of neural network-based classification models. This study aims to analyze the effect of data preprocessing and hyperparameter tuning on the performance of Backpropagation Neural Network (BPNN) in stroke classification. This study used a stroke dataset from the Kaggle platform consisting of 5,110 patient data with 10 clinical features. The evaluation was conducted using five schemes and consisted of several data balancing techniques. These techniques include no balancing, SMOTE, and ADASYN. In addition, the evaluation also involved data normalization including no normalization, MinMaxScaler, and Z-Score. The BPNN model used has an architecture of 19 input neurons, 29 neurons in the hidden layer, and 1 output neuron. Hyperparameter tuning was performed by finding the best learning rate and number of epochs. The evaluation results showed that the model in scheme one has limitations. This limitation is most visible in identifying stroke classes. The application of SMOTE and MinMaxScaler in scheme two proved that the results were better and its performance increased significantly. On the other hand, the combination of ADASYN and Z-Score in scheme three showed more stable performance and was able to detect stroke cases more accurately. The hyperparameter tuning process in schemes four and five also proved to improve performance. The best results were obtained in scheme five, with an accuracy of 96.47%, a precision of 97.34%, a recall of 95.62%, and an F1-score of 96.47%. These findings indicate that the combination of adaptive balancing techniques, distribution-based normalization, and optimal parameter tuning is very effective in improving the accuracy and stability of BPNN for stroke classification.

Keywords: Backpropagation Neural Network; Stroke; Data Preprocessing; Hyperparameter Tuning; Classification

1. PENDAHULUAN

Stroke merupakan salah satu penyakit tidak menular yang ditandai oleh gangguan fungsi otak yang terjadi secara mendadak akibat terganggunya atau terhentinya aliran darah ke jaringan otak[1]. Kondisi ini dapat menimbulkan dampak serius, mulai dari gangguan fungsional ringan hingga kecacatan permanen dan kematian apabila tidak segera ditangani[2]. Berdasarkan laporan Organisasi Kesehatan Dunia (*World Health Organization/WHO*), stroke masih termasuk dalam lima besar penyebab kematian tertinggi di dunia dan menjadi salah satu penyumbang utama beban disabilitas global. Tingginya angka kejadian serta dampak jangka panjang yang ditimbulkan menjadikan stroke sebagai permasalahan kesehatan yang perlu mendapatkan perhatian khusus, terutama dalam aspek pencegahan dan deteksi dini.

Upaya deteksi dini stroke memiliki peran penting dalam menurunkan risiko kematian dan memperbaiki prognosis pasien. Dalam praktik klinis, keterlambatan diagnosis masih sering terjadi di berbagai fasilitas pelayanan kesehatan. Penanganan yang dilakukan pada fase awal terbukti dapat meningkatkan peluang pemulihan dan mengurangi tingkat kecacatan. Namun demikian, proses identifikasi dini masih menghadapi berbagai kendala. Gejala awal stroke sering kali tidak spesifik dan dapat berbeda pada setiap individu, sehingga sulit dikenali secara cepat, terutama pada tahap pra-klinis. Selain itu, risiko stroke dipengaruhi oleh berbagai faktor seperti usia, hipertensi, diabetes, riwayat penyakit, serta gaya hidup, yang saling berinteraksi secara kompleks. Kondisi ini menuntut adanya pendekatan analisis yang mampu mengolah data medis dalam jumlah besar dan menangkap pola hubungan yang tidak linier. Perkembangan teknologi kecerdasan buatan, khususnya *machine learning*, membuka peluang baru dalam membantu proses analisis dan pengambilan keputusan di bidang medis. Pendekatan ini memungkinkan pemodelan hubungan kompleks antar variabel klinis yang sulit dianalisis menggunakan metode konvensional. Salah satu metode yang banyak diterapkan untuk permasalahan klasifikasi medis adalah Artificial Neural Network (ANN), khususnya *Backpropagation Neural Network* (BPNN). BPNN



memiliki kemampuan untuk mempelajari pola non-linear melalui proses pelatihan berbasis propagasi balik kesalahan, sehingga dinilai sesuai untuk menangani karakteristik data medis yang bersifat multidimensi dan heterogen.

Sejumlah penelitian terdahulu telah menunjukkan efektivitas BPNN dalam berbagai kasus prediksi dan klasifikasi. Penelitian oleh Azhima menunjukkan bahwa metode BPNN dapat digunakan untuk mengklasifikasikan penyakit stroke dengan tingkat akurasi mencapai 96,14% menggunakan kombinasi *Synthetic Minority Over-sampling Technique* (SMOTE) dan normalisasi *MinMaxScaler*[3]. Penelitian selanjutnya yang dilakukan oleh Tang, menggunakan SMOTE untuk mengatasi ketidakseimbangan kelas pada prediksi risiko stroke, sehingga pelatihan model lebih mampu mengenali kelas stroke (minoritas) tanpa menimbulkan kebocoran data karena *oversampling* hanya diterapkan pada data training di setiap fold cross-validation [4]. Selanjutnya, penelitian oleh Solikhun dan Amalya pada tahun 2023 mengkaji prediksi penyakit stroke menggunakan backpropagation metode *Levenberg-Marquardt* dengan 10 atribut (jenis kelamin, umur, hipertensi, dan lainnya.) dari dataset *Kaggle* (5110 record). Studi ini menerapkan arsitektur 10-10-1 pada *epoch* 10000, menghasilkan MSE 0,0550 untuk pelatihan dan pengujian [5]. Selain itu, penelitian oleh Alwaliyanto pada tahun 2025 mengkaji klasifikasi penyakit stroke menggunakan SVM (kernel linear, polynomial, RBF) dengan ADASYN untuk mengatasi *imbalanced* data dan 5-Fold Cross Validation. Studi ini menerapkan dataset stroke yang diseimbangkan, dengan kernel linear sesudah ADASYN memberikan performa terbaik: AUC-ROC 0.8333, *recall* 0.7827, dan *F1-score* 0.2181 untuk kelas stroke[6].

Sebagian besar penelitian terdahulu hanya berfokus pada satu jenis *preprocessing* tanpa melakukan eksplorasi mendalam terhadap interaksi antara teknik *balancing*, normalisasi, dan *hyperparameter tuning* yang dapat memengaruhi stabilitas serta performa model secara keseluruhan. Berdasarkan hal tersebut, penelitian ini berupaya mengisi celah penelitian yang masih terbuka dengan menganalisis secara komprehensif pengaruh *preprocessing data* dan *hyperparameter tuning* terhadap kinerja BPNN dalam klasifikasi penyakit stroke. Penelitian ini tidak hanya berfokus pada peningkatan akurasi, tetapi juga pada bagaimana proses penyeimbangan data dan penyesuaian parameter pelatihan dapat mengubah perilaku model dalam mengenali pola data medis yang kompleks. Dengan pendekatan ini, maka dapat diperoleh pemahaman yang lebih mendalam mengenai peran strategi *data preprocessing* dan *parameter tuning* dalam menghasilkan model kecerdasan buatan yang lebih andal, adaptif, dan memiliki potensi untuk diterapkan dalam sistem pendukung diagnosis medis di masa depan.

Selain itu, penelitian ini memberikan kontribusi penting dengan mengevaluasi secara sistematis kombinasi teknik *balancing* adaptif (SMOTE dan ADASYN) dan normalisasi data bersama *hyperparameter tuning* untuk meningkatkan performa dan stabilitas model. Penelitian ini juga menekankan peningkatan kemampuan deteksi kelas minoritas pada dataset medis yang tidak seimbang, sehingga hasil yang diperoleh tidak hanya meningkatkan akurasi, tetapi juga meningkatkan sensitivitas model dalam mendeteksi kasus stroke. Temuan ini diharapkan dapat menjadi referensi dalam pengembangan model kecerdasan buatan yang lebih robust dan aplikatif pada sistem pendukung keputusan di bidang kesehatan.

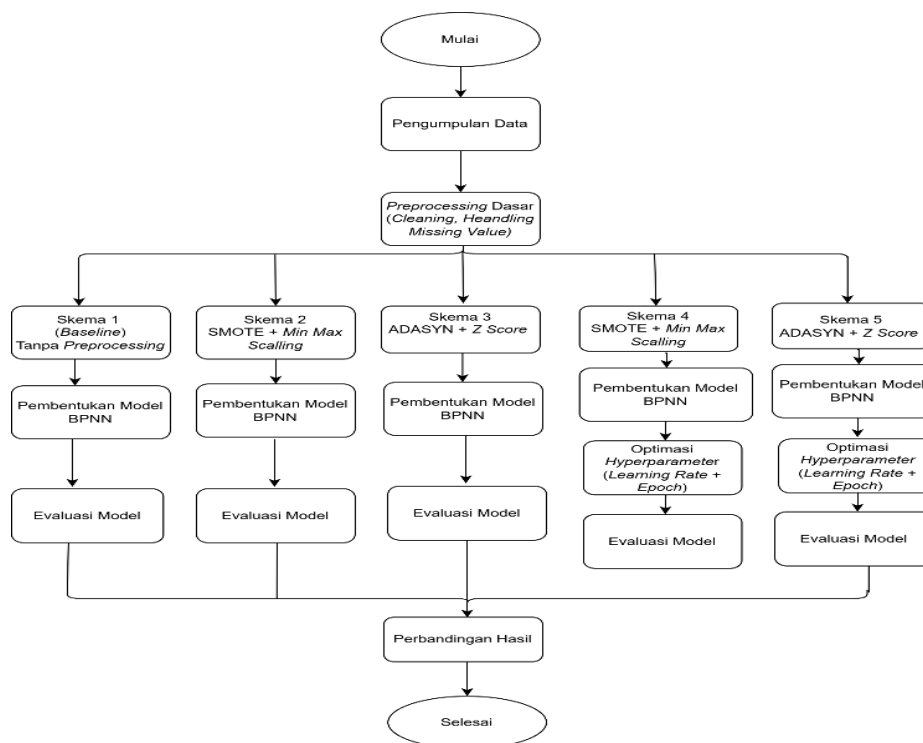
2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Pada Gambar 1 ditunjukkan bahwa penelitian ini dirancang melalui lima tahap utama yang tersusun secara sistematis untuk menghasilkan model klasifikasi stroke yang optimal. Tahap pertama dimulai dengan proses pengumpulan data yang bersumber dari *Kaggle* dataset klasifikasi stroke. Dataset ini menjadi fondasi utama penelitian karena berisi berbagai atribut yang merepresentasikan kondisi kesehatan dan faktor risiko pasien. Setelah seluruh data berhasil dikumpulkan, dilakukan *preprocessing* dasar sebagai tahap persiapan. Proses *preprocessing* meliputi pembersihan data untuk menghilangkan noise atau data yang tidak relevan, penanganan nilai yang hilang (*missing values*), pemeriksaan serta penghapusan data duplikat, dan transformasi awal pada variabel tertentu agar konsisten dan siap diolah. Tahapan ini sangat penting karena kualitas data yang baik akan berpengaruh langsung terhadap akurasi dan keandalan model yang dibangun. Dengan demikian, *preprocessing* dasar bertujuan memastikan dataset berada dalam kondisi optimal sebelum memasuki tahapan eksperimen berikutnya.

Setelah *preprocessing* dasar selesai, dataset kemudian dibagi ke dalam lima skema perlakuan yang berbeda guna menganalisis pengaruh teknik penyeimbangan data, normalisasi fitur, dan optimasi *hyperparameter* terhadap performa model *Backpropagation Neural Network* (BPNN). Skema pertama merupakan skema baseline, yaitu dataset hasil *preprocessing* dasar langsung digunakan untuk pelatihan model tanpa *balancing* data, normalisasi fitur, maupun optimasi *hyperparameter*, sehingga berfungsi sebagai pembandingan utama. Skema kedua menerapkan teknik SMOTE untuk menyeimbangkan distribusi kelas yang tidak seimbang serta *Min-Max Scaling* untuk menormalkan rentang nilai fitur agar berada pada skala yang seragam. Skema ketiga menggunakan metode ADASYN sebagai teknik *oversampling* adaptif untuk mengatasi ketidakseimbangan kelas, dikombinasikan dengan *Z-Score Standardization* untuk menormalkan distribusi fitur sebelum proses pelatihan model. Pada skema keempat, metode SMOTE dan *Min-Max Scaling* digabungkan, kemudian ditambahkan optimasi *hyperparameter* berupa penyesuaian *learning rate* dan jumlah *epoch* guna menemukan konfigurasi pelatihan yang paling optimal. Skema kelima merupakan skema paling komprehensif, yaitu dengan menggabungkan ADASYN, *Z-Score Standardization*, serta optimasi *hyperparameter* untuk memaksimalkan kemampuan pembelajaran model BPNN. Setelah seluruh skema dijalankan, performa model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Tahap akhir penelitian dilakukan dengan membandingkan hasil dari

setiap skema untuk mengidentifikasi sejauh mana penerapan teknik *preprocessing* dan *hyperparameter tuning* mampu meningkatkan kinerja model BPNN dalam klasifikasi stroke.



Gambar 1. Diagram Alur Penelitian

2.2 Dataset

Data yang digunakan sebagai pendukung penelitian ini diambil dari platform *Kaggle*. Data yang digunakan untuk pendukung riset ini adalah data sekunder yang berbentuk 5.110 data yang terdiri dari 2 kelas target yakni stroke dan *non-stroke*. Dataset terdiri dari 10 fitur utama terkait faktor-faktor risiko demam, yaitu: *Gender, Age, Work Type, Residence Type, Hypertension, Heart Disease, Ever Married, Average Glucose Level, Body Mass Index, dan Smoking Status*. Fitur tersebut kemudian dimasukkan sebagai input model BPNN. Fitur stroke dimasukkan sebagai target variabel dengan label binari yakni 1 untuk 'stroke' dan 0 untuk 'non-stroke'. Dataset dibagi menjadi data *training* dan *testing* yang dilakukan dengan 3 frame, yakni 70:30, 80:20, dan 90:10. Untuk menghindari terjadinya data *leakage*, teknik balancing menggunakan SMOTE dan ADASYN hanya diterapkan pada data *training* setelah proses pembagian data, sedangkan data *testing* tetap menggunakan data asli tanpa proses oversampling. Berdasarkan pembagian dataset tersebut, pada *frame* 70:30 menghasilkan 3.577 data *training* dan 1.533 data *testing*. 80:20 menghasilkan 4.088 data *training* dan 1.022 data *testing*. Terakhir adalah 90:10 yang menghasilkan 4.599 data *training* dan 511 data *testing*.

Karena ketidakseimbangan antara *non-stroke* dan stroke di mana ada sekitar 4.881 *non-stroke* dan 249 stroke, teknik *balancing* perlu dilakukan saat *preprocessing*. Pada *frame* sebelum *balancing*, dataset dibagi menjadi *frame-frame training-set ratio*. Pertama adalah 70:30 yang menghasilkan 3.577 data *training* yang terdiri dari sekitar 3.403 *non-stroke* dan 174 stroke, serta 1.533 data *testing* yang terdiri dari sekitar 1.478 *non-stroke* dan 75 stroke. Kedua adalah 80:20 yang menghasilkan 4.088 data *training* yang terdiri dari sekitar 3.889 *non-stroke* dan 199 stroke serta 1.022 data *testing* yang terdiri dari sekitar 972 *non-stroke* dan 50 stroke, sementara yang terakhir 90:10 yang menghasilkan 4.599 data *training* yang terdiri dari sekitar 4.375 *non-stroke* dan 224 stroke serta 511 data *testing* yang terdiri dari sekitar 486 *non-stroke* dan 25 stroke. Pembagian ini dilakukan untuk menguji konsistensi performa model BPNN dalam berbagai komposisi data, sebelum tahap *preprocessing* diterapkan.

2.3 Preprocessing Data

Tahap *preprocessing* bertujuan untuk memastikan bahwa data yang digunakan sesuai dengan kebutuhan model pembelajaran mesin[7]. Data *preprocessing* dilakukan untuk menjamin kualitas data yang akan diolah dalam pemodelan klasifikasi. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

a. Pembersihan Data (*Cleaning*)

Dilakukan pemeriksaan terhadap data kosong (*missing values*) dan duplikasi. Data yang tidak valid atau tidak lengkap dihapus untuk menjaga kualitas dataset.

b. Transformasi Data (*Encoding*)

Variabel kategorikal seperti *gender, work type, residence type, dan smoking status* diubah menjadi representasi numerik menggunakan metode *one-hot encoding*.



c. *Balancing Data*

Untuk menangani ketidakseimbangan distribusi antara kelas mayoritas dan kelas minoritas pada dataset, penelitian ini menerapkan teknik balancing melalui metode oversampling. Resampling (oversampling/undersampling) dapat menyeimbangkan kelas[8]. Dua pendekatan yang digunakan dalam penelitian ini adalah Synthetic Minority Over-sampling Technique (SMOTE) dan Adaptive Synthetic Sampling (ADASYN), yang berfungsi untuk menambah sampel sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Proses ini penting agar model BPNN dapat belajar secara lebih optimal tanpa bias terhadap kelas mayoritas. SMOTE bekerja dengan membuat data sintetis di sepanjang garis lurus antara satu data minoritas dan tetangga terdekatnya[9]. SMOTE merupakan teknik oversampling yang membentuk data sintetis pada kelas minoritas melalui interpolasi linier antar sampel minoritas [4]. Dibandingkan oversampling acak, SMOTE dipandang dapat menekan risiko overfitting karena tidak sekadar menggandakan data yang sama[10]. Metode ini menambah variasi data minoritas tanpa sekadar menduplikasi data, sehingga mengurangi risiko overfitting.

$$X_{\text{new}} = X_i + \lambda (X_{z_i} - X_i) \quad (1)$$

Persamaan (1) menggambarkan proses pembentukan data sintetis pada metode SMOTE. Pada persamaan tersebut, X_i merepresentasikan sampel asli dari kelas minoritas, sedangkan X_{z_i} merupakan salah satu sampel tetangga terdekat dari X_i . Parameter λ adalah bilangan acak antara 0 dan 1 yang menentukan posisi titik baru pada garis interpolasi yang menghubungkan kedua sampel tersebut. Melalui proses ini, dihasilkan sampel sintetis baru X_{new} yang berada di antara X_i dan X_{z_i} . Pendekatan ini memungkinkan pembentukan variasi data minoritas secara lebih alami sehingga membantu menyeimbangkan distribusi kelas tanpa melakukan duplikasi data.

ADASYN memberi lebih banyak data sintetis pada sampel minoritas yang sulit dipelajari, yaitu sampel yang dikelilingi lebih banyak tetangga dari kelas mayoritas[11]. Dengan kata lain, ADASYN adaptif karena fokus pada titik-titik minoritas yang memiliki risiko klasifikasi salah lebih tinggi, sehingga distribusi data menjadi lebih representatif dan proses pembelajaran model lebih efektif[12]. Teknik oversampling ADASYN digunakan untuk mengurangi bias akibat distribusi kelas yang tidak seimbang dengan cara memperbanyak data sintetis pada kelas minoritas, sehingga model menjadi lebih sensitif dalam mengenali kasus stroke[13].

Menghitung tingkat kesulitan tiap sampel minoritas

$$r_i = \frac{\Delta_i}{k} \quad (2)$$

Menentukan jumlah sampel sintesis yang perlu dibuat

$$G_i = r_i \times G \quad (3)$$

Membagikan data sintesis

$$X_{\text{new}} = X_i + \lambda \times (X_{z_i} - X_i) \quad (4)$$

Persamaan (2) digunakan untuk menghitung tingkat kesulitan setiap sampel minoritas berdasarkan proporsi tetangga dari kelas mayoritas di sekitarnya. Semakin besar nilai r_i , semakin sulit sampel tersebut dipelajari oleh model karena berada pada area yang didominasi oleh kelas mayoritas. Selanjutnya, Persamaan (3) menentukan jumlah data sintetis yang perlu dibuat untuk setiap sampel minoritas, di mana nilai G_i menunjukkan jumlah sampel sintetis yang dialokasikan secara adaptif sesuai tingkat kesulitannya. Dengan demikian, sampel yang lebih sulit akan memperoleh lebih banyak data sintetis. Proses pembentukan data sintetis dilakukan menggunakan Persamaan (4), yaitu dengan menginterpolasi sampel minoritas X_i dengan salah satu tetangga terdekatnya untuk menghasilkan sampel baru X_{new} . Pendekatan adaptif ini memungkinkan distribusi data menjadi lebih seimbang sekaligus memperkuat representasi area yang sulit dipelajari oleh model.

d. *Normalisasi Data*

Normalisasi dilakukan untuk menyamakan skala antar variabel numerik agar proses pembelajaran model menjadi lebih stabil. Dua metode yang digunakan adalah MinMaxScaler dan Z-Score Normalization. Metode MinMaxScaler mengubah skala data numerik ke dalam rentang 0 sampai 1. Normalisasi ini sangat cocok untuk algoritma yang sensitif terhadap skala fitur, termasuk BPNN, karena membantu proses pelatihan menjadi lebih stabil dan mempercepat konvergensi.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

Pada Persamaan (5), x merepresentasikan nilai asli suatu fitur sebelum dilakukan normalisasi. Nilai x_{\min} dan x_{\max} masing-masing menunjukkan nilai minimum dan maksimum dari fitur tersebut dalam dataset. Proses normalisasi menghasilkan nilai baru x' , yaitu nilai fitur yang telah diskalakan ke dalam rentang 0 hingga 1. Transformasi ini memastikan bahwa setiap fitur berada pada skala yang sama sehingga tidak ada fitur yang mendominasi proses pembelajaran model.

Z-Score Normalization mengubah data sehingga memiliki mean = 0 dan standar deviasi = 1. Metode ini sangat efektif ketika data memiliki sebaran nilai yang bervariasi atau terdapat outlier, sehingga lebih stabil dibanding MinMaxScaler pada distribusi data yang ekstrem.



$$z = \frac{x-\mu}{\sigma} \quad (6)$$

Pada Persamaan (6), x menyatakan nilai asli suatu fitur sebelum proses standarisasi. Simbol μ menunjukkan nilai rata-rata dari fitur tersebut, sedangkan σ menyatakan simpangan baku yang menggambarkan sebaran data terhadap nilai rata-ratanya. Hasil transformasi menghasilkan nilai baru z , yaitu nilai fitur yang telah distandarasi sehingga memiliki rata-rata mendekati nol dan simpangan baku satu. Proses ini membantu menyamakan skala distribusi data dan meningkatkan stabilitas serta efektivitas proses pelatihan model.

2.4 Skema Eksperimen

Untuk menganalisis pengaruh *preprocessing* dan parameter *tuning*, penelitian ini dibagi ke dalam lima skema utama seperti pada Tabel 1.

Tabel 1. Skema Eksperimen

Skema	<i>Balancing</i>	Normalisasi	<i>Hyperparameter Tuning</i>
1	Tanpa <i>Balancing</i>	Tanpa Normalisasi	Tidak dilakukan <i>hyperparameter tuning</i>
2	SMOTE	<i>MinMax Scaller</i>	Tidak dilakukan <i>hyperparameter tuning</i>
3	ADASYN	<i>Z-Score</i>	Tidak dilakukan <i>hyperparameter tuning</i>
4	SMOTE	<i>MinMax Scaller</i>	($lr = 0.1, 0.01, 0.001, 0.0001$, $epoch = 50, 100, 150$)
5	ADASYN	<i>Z-Score</i>	($lr = 0.1, 0.01, 0.001, 0.0001$, $epoch = 50, 100, 150$)

Berdasarkan Tabel 1 terdapat lima skema eksperimen yang digunakan dalam menganalisis pengaruh dari *preprocessing* dan *hyperparameter tuning* ke BPNN dalam proses kalsifikasi stroke. Skema 1, pada skema ini, tidak dilakukan teknik *balancing*, normalisasi, dan tidak dilakukan *hyperparameter tuning*. Sedangkan skema 2 dilakukan teknik *balancing* dengan menggunakan SMOTE dan normalisasi dengan menggunakan *MinMax Scaller*, namun tidak dilakukan *hyperparameter tuning*. Kemudian, skema 3 dilakukan teknik *balancing* yaitu menggunakan metode ADASYN, dan Normalisaasi *Z-Score* serta tidak dilakukan *hyperparameter tuning*. Pada skema 1 hingga skema 3 tidak dilakukan *hyperparameter tuning*, dimana parameter yang diberikan adalah parameter dasar yaitu *learning rate* 0.01 dan *epoch* 50. Selanjutnya skema 4 dilakukan *balancing* dengan SMOTE dan normalisasi menggunakan *MinMax Scaller*, namun pada skema ini dilakukan *hyperparameter tuning*. Lalu skema 5, pada skema ini dilakukan *balancing* data yaitu menggunakan ADASYN dan normalisasi *Z-Score* serta *hyperparameter tuning*.

2.5 Pemodelan BPNN

Back Propagation Neural Network (BPNN) adalah metode pembelajaran terawasi pada Jaringan Syaraf Tiruan yang tersusun dari input layer, hidden layer, dan output layer, dengan penyesuaian bobot sebagai mekanisme utama pembelajaran [14]. *Backpropagation* merupakan pendekatan pelatihan yang dilakukan secara sistematis untuk mengoptimalkan kinerja jaringan syaraf tiruan yang memiliki lebih dari satu lapisan [15]. Pada lapisan tersembunyi dan lapisan keluaran dilakukan proses komputasi terhadap bobot dan bias, serta perhitungan nilai output masing-masing lapisan berdasarkan fungsi aktivasi yang digunakan. Jaringan syaraf tiruan ini beroperasi melalui dua tahap utama, yaitu fase umpan maju (*feedforward*) dan fase umpan balik (*feedback*). Pada tahap umpan maju, data dari lapisan masukan dipropagasikan menuju lapisan keluaran melalui lapisan tersembunyi. Jika kesalahan yang dihasilkan masih melebihi batas yang ditentukan terhadap target keluaran, maka sinyal keluaran akan dikirim kembali ke lapisan masukan melalui lapisan tersembunyi untuk proses penyesuaian[16].

Tabel 2. *Hyperparameter* Model BPNN

Komponen	Nilai / Pengaturan
Arsitektur Jaringan	19 : 29 : 1
Neuron <i>Input</i>	19
Neuron <i>Hidden</i>	29
Neuron <i>Output</i>	1
Fungsi Aktivasi (<i>Hidden</i>)	ReLU
Fungsi Aktivasi (<i>Output</i>)	<i>Sigmoid</i>
Parameter yang Dioptimasi	<i>Learning Rate, Epoch</i>
Rentang <i>Learning Rate</i>	0.1, 0.01, 0.001, 0.0001
Jumlah <i>Epoch</i>	50, 100, 150

Tabel 2 merupakan model BPNN yang digunakan dalam penelitian, dimana model BPNN digunakan arsitektur dasar [19:29:1], yaitu 19 neuron pada lapisan input, 29 neuron pada hidden layer, dan 1 neuron pada output layer. Kemudian fungsi aktivasi yang digunakan adalah ReLU pada hidden layer untuk membuat proses pembelajaran lebih cepat dan mencegah masalah (*vanishing gradient problem*), sedangkan sigmoid digunakan pada layer output untuk menghasilkan nilai probabilitas biner sesuai kebutuhan klasifikasi stroke. Dalam proses optimasi *hyperparameter* pada skema 4 dan 5 difokuskan pada pencarian kombinasi *learning rate* dan jumlah *epoch* yang paling optimal. Rentang *learning rate* yang diuji adalah 0.1, 0.01, 0.001, dan 0.0001, sedangkan jumlah *epoch* divariasikan pada 50, 100, dan 150.



Model BPNN dibangun menggunakan *TensorFlow* dan *Keras* pada platform *Google Colab*. Algoritma *backpropagation* digunakan untuk memperbarui bobot jaringan berdasarkan kesalahan prediksi (error). BPNN beroperasi dalam dua fase utama, yaitu Fase *Feedforward* dimana data input diproses dari lapisan input menuju output untuk menghasilkan prediksi. Kemudian yang kedua Fase *Backpropagation*, yaitu selisih antara hasil prediksi dan nilai target digunakan untuk memperbarui bobot jaringan secara bertahap hingga error minimum tercapai. Model disimpan pada setiap skema untuk dievaluasi performanya menggunakan metrik yang sama agar hasil perbandingan tetap objektif.

2.6 Evaluasi Model

Evaluasi model dilakukan menggunakan *Confusion Matrix* yang menghasilkan empat metrik utama, yaitu *Accuracy*, *Precision*, *Recall (Sensitivity)*, *F1-score*. Mengingat dataset memiliki ketidakseimbangan kelas yang tinggi (sekitar 19,5:1), metrik *Recall* dan *F1-score* menjadi perhatian utama dalam evaluasi model karena lebih merepresentasikan kemampuan model dalam mendeteksi kelas minoritas (stroke) dibandingkan akurasi. Evaluasi ini digunakan untuk menentukan skema terbaik dengan performa paling stabil dan akurasi tertinggi dalam klasifikasi penyakit stroke. Pada skenario *imbalanced*, nilai *precision/recall/F1* untuk kelas stroke dapat menjadi indikator yang lebih jujur dibanding akurasi agregat [17].

Accuracy (akurasi) adalah ukuran performa model klasifikasi yang menunjukkan seberapa banyak prediksi yang bernilai benar dibandingkan dengan seluruh data yang diuji [18]. Dengan kata lain, akurasi menghitung proporsi prediksi yang bernilai benar (baik positif maupun negatif) dari semua prediksi yang dihasilkan model [19]. *Balanced accuracy* dipakai untuk memastikan performa pada kelas mayoritas dan minoritas ikut berkontribusi secara seimbang dalam penilaian [20].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Persamaan (7) digunakan untuk menghitung tingkat akurasi model, yaitu perbandingan antara jumlah prediksi yang benar terhadap seluruh data yang dievaluasi. Nilai akurasi diperoleh dari penjumlahan *True Positive (TP)* dan *True Negative (TN)*, yang menunjukkan jumlah prediksi yang tepat, kemudian dibagi dengan total seluruh prediksi yang mencakup *TP*, *TN*, *False Positive (FP)*, dan *False Negative (FN)*. Metrik ini memberikan gambaran umum tentang seberapa baik model mengklasifikasikan data secara keseluruhan.

Precision adalah metrik evaluasi dalam klasifikasi yang mengukur seberapa akurat model ketika memprediksi kelas positif. Dengan kata lain, presisi menunjukkan proporsi prediksi positif yang benar-benar positif. Presisi sangat penting dalam kasus medis seperti klasifikasi stroke, karena berkaitan dengan kemampuan model menghindari kesalahan *FP (False Positive)*.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Persamaan (8) digunakan untuk menghitung nilai *precision*, yaitu tingkat ketepatan prediksi positif yang dihasilkan oleh model. *Precision* diperoleh dengan membandingkan jumlah *True Positive (TP)*, yaitu data yang diprediksi positif dan benar, terhadap seluruh data yang diprediksi positif oleh model, yang terdiri dari *TP* dan *False Positive (FP)*. Metrik ini menunjukkan seberapa akurat model dalam mengidentifikasi kelas positif tanpa menghasilkan prediksi positif yang keliru.

Recall (disebut juga *Sensitivity* atau *True Positive Rate*) adalah metrik evaluasi yang mengukur kemampuan model dalam menemukan seluruh data yang benar-benar positif.

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Persamaan (9) digunakan untuk menghitung nilai *recall*, yaitu kemampuan model dalam mengidentifikasi seluruh data yang benar-benar termasuk dalam kelas positif. *Recall* diperoleh dengan membandingkan jumlah *True Positive (TP)*, yaitu data positif yang berhasil diprediksi dengan benar, terhadap total data positif sebenarnya yang terdiri dari *TP* dan *False Negative (FN)*. Metrik ini menunjukkan seberapa baik model dalam mendeteksi kasus positif dan sangat penting pada dataset tidak seimbang, terutama ketika kesalahan dalam melewatkan kasus positif harus diminimalkan.

F1-score merupakan salah satu metrik evaluasi yang menjadi nilai rata-rata antara *Precision* dan *Recall*. Metrik ini digunakan untuk menilai performa model ketika terjadi ketidakseimbangan data (*imbalanced dataset*), seperti pada kasus klasifikasi penyakit stroke.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

Persamaan (10) digunakan untuk menghitung nilai *F1-score*, yaitu metrik yang menggabungkan *precision* dan *recall* dalam satu ukuran keseimbangan. *F1-score* diperoleh dari rata-rata harmonis antara *precision* dan *recall*, sehingga memberikan penilaian yang lebih representatif terhadap performa model, terutama pada dataset yang tidak seimbang. Nilai ini akan tinggi hanya jika *precision* dan *recall* sama-sama tinggi, sehingga *F1-score* sangat berguna untuk mengevaluasi kemampuan model dalam mendeteksi kelas positif secara akurat dan konsisten.



3. HASIL DAN PEMBAHASAN

3.1 Skema 1

Pada skema 1 eksperimen dilakukan tanpa menerapkan teknik *balancing* ataupun teknik normalisasi, Tujuan dari skema ini adalah untuk melihat performa dasar (*baseline*) tanpa intervensi apa pun pada data, serta untuk mengamati sejauh mana ketidakseimbangan data dan perbedaan skala mempengaruhi proses pembelajaran dan hasil klasifikasi.

Tabel 3. Hasil eksperimen tanpa *Balancing* data dan tanpa normalisasi data

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0.95	0.00	0.00	0.00
0.07	0.04	0.94	0.09
0.94	0.00	0.00	0.00
0.95	0.00	0.00	0.00
0.07	0.04	0.94	0.09

Hasil eksperimen pada tahap pertama ditunjukkan pada Tabel 3 terlihat pola yang sangat kontras antara akurasi dan metrik lainnya, yaitu *precision*, *recall*, dan *F1-score*. Model *Backpropagation Neural Network* pada eksperimen tanpa teknik *balancing* dan normalisasi menghasilkan dua jenis performa ekstrem. Pada sebagian besar eksperimen, model mencapai akurasi yang sangat tinggi, yaitu antara 0,94 hingga 0,95. Namun nilai *precision*, *recall*, dan *F1-score* pada kondisi tersebut bernilai 0,00. Hal ini menunjukkan bahwa meskipun model terlihat baik secara akurasi, pada kenyataannya model tidak mampu mengidentifikasi kasus stroke, melainkan hanya mengklasifikasikan seluruh data ke kelas mayoritas, yaitu “tidak stroke”. Nilai akurasi yang tinggi pada skema ini sebanding dengan proporsi kelas mayoritas dalam dataset. Dengan jumlah 4.881 data non-stroke dari total 5.110 data, model dapat mencapai akurasi sekitar 95,5% hanya dengan memprediksi seluruh data sebagai non-stroke. Kondisi ini menunjukkan bahwa model belum mempelajari pola yang bermakna dan hanya mengikuti distribusi kelas mayoritas. Sebaliknya, pada beberapa konfigurasi tertentu akurasi model justru turun drastis, namun nilai *recall* meningkat tinggi hingga 0,94. Hasil ini menandakan bahwa model memprediksi hampir seluruh data sebagai “stroke”, sehingga meskipun *recall* tinggi, prediksinya tidak akurat karena tidak mampu membedakan kedua kelas.

3.2 Skema 2

Pada skema 2, proses *preprocessing* dilakukan, yaitu dengan menerapkan teknik *balancing* menggunakan metode SMOTE. Kemudian dilakukan normalisasi menggunakan *Min-Max Scaler* agar seluruh fitur numerik memiliki rentang nilai yang seragam, sehingga proses pembelajaran jaringan saraf menjadi lebih stabil.

Tabel 4. Hasil eksperimen menggunakan SMOTE dan *MinMax Scaller*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0.820	0.792	0.868	0.828
0.825	0.770	0.926	0.841
0.798	0.843	0.732	0.784
0.792	0.770	0.831	0.800
0.803	0.770	0.864	0.814

Berdasarkan Tabel 4, penerapan SMOTE dan *MinMaxScaler* dapat meningkatkan kinerja model secara konsisten di semua metrik evaluasi. Akurasi tetap stabil di kisaran 0.792–0.825, yang menunjukkan tingkat generalisasi yang bagus. Presisi berada di atas 0.770 dan pada kisaran 0.843 mengindikasikan rendahnya false positif yang dihasilkan sehingga stressor atau kelas target dapat diidentifikasi dengan baik. Kemudian *Recall* sangat tinggi, di atas 0,732 bahkan 0,926 mengindikasikan bahwa model mampu mencakup hampir semua kasus stroke sebenarnya. *F1-score* pada rentang 0.784–0.841 juga menunjukkan bahwa tidak hanya model ini sensitif, tetapi tetap memberikan presisi yang baik.

3.3 Skema 3

Proses *preprocessing* dilakukan pada skema ketiga ini dengan menggunakan teknik *balancing* ADASYN. Kemudian, dilakukan normalisasi dengan menggunakan metode *Z-Score* agar setiap fitur memiliki distribusi dengan rata-rata 0 dan standar deviasi 1.

Tabel 5. Hasil eksperimen menggunakan ADASYN dan *Z Score*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
0.872	0.845	0.914	0.878
0.885	0.857	0.925	0.890
0.883	0.838	0.951	0.891
0.807	0.783	0.853	0.817
0.812	0.776	0.880	0.825



Tabel 5 menampilkan hasil pada skema ketiga. Pada skema tersebut, performa model meningkat sangat signifikan dibandingkan dengan kedua skema sebelumnya, terutama setelah dilakukan penerapan metode ADASYN dan normalisasi dengan *Z-Score*. Akurasi meningkat sekitar 5-7% pada seluruh variasi rasio pelatihan, merujuk pada sampel sintesis adaptif, hal ini menjadi solusi untuk memperbaiki representasi kelas minoritas dan membantu model mempelajari pola dengan lebih efektif. Presisi juga mengalami peningkatan hingga rasio pelatihan paling tinggi, hal ini menunjukkan model semakin akurat dalam memberikan prediksi positif dan sebaliknya, false positive semakin rendah, walaupun kemudian agak berkurang karena model cenderung lebih agresif. Peningkatan paling mencolok adalah *recall* yang mencapai hingga 95%. Hal ini menunjukkan bahwa ADASYN dapat membantu model mendeteksi hampir seluruh kasus stroke dan mereduksi *false negative*. Keseimbangan secara keseluruhan antara presisi dan *recall* tampak dari *F1-score* yang terus meningkat pada seluruh rasio pelatihan, membuktikan model lebih stabil, lebih sensitif, dan akurat.

3.4 Skema 4

Pada skema 4 dilakukan dengan menggunakan metode SMOTE dan *Min-Max Scaler* pada tahap *preprocessing* data, kemudian setelah itu dilanjutkan dengan proses parameter *tuning*. Proses parameter *tuning* dilakukan dengan menggunakan variasi *learning rate* (0.1, 0.01, 0.001, dan 0.0001) serta jumlah *epoch* (50, 100, dan 150). Kemudian, dalam skema ini juga menerapkan teknik *cross-validation* dengan nilai $k = 5$, tujuannya adalah untuk memastikan bahwa hasil evaluasi model lebih stabil, tidak bias terhadap satu subset data tertentu, dan mampu menggambarkan performa model secara lebih menyeluruh. Dengan *5-fold cross-validation*, setiap kombinasi parameter diuji pada lima lipatan data yang berbeda, sehingga nilai akhir yang diperoleh merupakan rata-rata performa model yang lebih akurat.

Tabel 6. Hasil eksperimen menggunakan SMOTE + *Max-Min Scaller* + *hyperparameter tuning*

<i>Learning Rate</i>	<i>Epoch</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0.0100	50	0.963073	0.988335	0.937257	0.962888
0.0100	100	0.961736	0.978937	0.943893	0.961049
0.0010	50	0.961427	0.987632	0.931747	0.959994
0.0010	100	0.961222	0.985731	0.936882	0.960814
0.0001	150	0.960290	0.998800	0.931014	0.959693
0.0001	100	0.959062	0.986190	0.929827	0.956917
0.1000	150	0.958548	0.999578	0.917797	0.956732
0.1000	100	0.958239	0.999776	0.916685	0.956415

Berdasarkan hasil pengujian yang dapat dilihat pada tabel 6, skema 4 menunjukkan bahwa model BPNN mencapai performa paling optimal pada *learning rate* 0.0100 dengan epoch 50–100. Kombinasi ini menghasilkan akurasi tertinggi yaitu sebesar 0.963073, kemudian *precision* 0.988335, *recall* 0.937257, dan *F1-score* 0.962888. *Learning rate* 0.001 dan 0.0001 memberikan performa stabil dengan akurasi di kisaran 0.959–0.961, hal ini menunjukkan bahwa model tetap mampu belajar dengan baik pada nilai *learning rate* yang lebih kecil. Sebaliknya, pada nilai *learning rate* yang lebih besar yaitu 0.1 menghasilkan performa sedikit lebih rendah yaitu dengan akurasi sekitar 0.958 dan *F1-score* sekitar 0.956, hal ini menunjukkan ketidakstabilan pembelajaran yang disebabkan karena nilai *Learning rate* yang terlalu besar.

3.5 Skema 5

Pada skema ini, pendekatan yang digunakan dirancang melalui kombinasi teknik *preprocessing* dan *hyperparameter tuning*. Skema ini dilakukan sebagai tahap eksperimen paling komprehensif untuk menguji pengaruh metode penyeimbangan data yang adaptif serta strategi normalisasi berbasis distribusi terhadap kinerja model. Dalam skema 5 eksperimen dilakukan dengan menggunakan metode ADASYN dan *Z-Score Normalization* pada tahap *preprocessing* data. Selanjutnya, setelah proses *preprocessing* dan normalisasi data, dilakukan parameter *tuning* atau penyesuaian parameter yaitu dengan menggunakan variasi *learning rate* yang sama seperti skema sebelumnya, yaitu nilai *learning rate* 0.1, 0.01, 0.001, dan 0.0001, serta jumlah epoch 50, 100, dan 150 untuk menganalisis pengaruh perubahan parameter terhadap performa model. Sama seperti skema 4, skema ini juga menerapkan *cross-validation* dengan nilai $k = 5$, sehingga setiap kombinasi parameter diuji pada lima lipatan data yang berbeda guna menghasilkan nilai performa yang lebih stabil, tidak bias, dan representatif.

Tabel 7. Hasil eksperimen menggunakan ADASYN + *Z score* + *hyperparameter tuning*

<i>LearningRate</i>	<i>Epoch</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
0.1000	50	0.956288	1.000000	0.913391	0.954574
0.1000	100	0.946709	0.999776	0.894611	0.944254
0.0100	100	0.964745	0.973483	0.956189	0.964749
0.0100	50	0.961382	0.974097	0.948719	0.961234
0.0010	50	0.964337	0.979834	0.948923	0.964084
0.0010	100	0.962910	0.975649	0.950334	0.962764
0.0001	100	0.962095	0.987650	0.936606	0.961436
0.0001	150	0.961586	0.982949	0.940241	0.961083

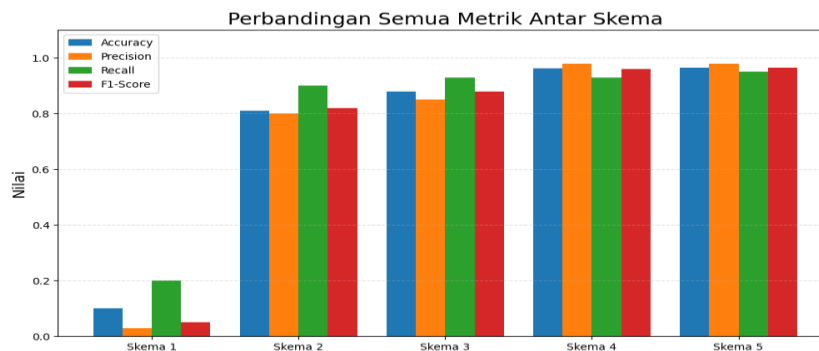
Evaluasi hasil skema 5 berdasarkan pada Tabel 7 menunjukkan bahwa kombinasi antara ADASYN dan *Z-Score* Normalization serta *tuning learning rate* dan *epoch* menghasilkan model BPNN dengan performa yang sangat baik dan stabil. Akurasi tertinggi pada eksperimen ini sebesar 0.9647 terjadi pada *learning rate* 0.01 dan *epoch* 100, dimana akurasi rata-ratanya berada pada rentang 0.946 hingga 0.964. Presisi tertinggi sebesar 1.00 terjadi dalam dua konfigurasi, diantaranya *learning rate* 0.01 dan 100 *epoch*, serta 0.001 dan 50 *epoch*. Presisi dan *recall* memiliki keterkaitan yang sangat tinggi, dimana pada konfigurasi dimana presisi meningkat signifikan, maka *recall* akan mengalami penurunan dan memiliki karakteristik model yang cenderung konservatif. *Recall* tertinggi pada skema ini sebesar 0.956 terjadi pada *learning rate* 0.01. *F1-score* model berada pada rentang 0.97 hingga 0.98 dan bersifat sangat konsisten. Sehingga dapat disimpulkan bahwa model memiliki performa yang sangat baik dengan presisi tinggi dan juga mampu mendeteksi sebagian besar kasus stroke dan kekurangannya. *Learning rate* 0.01 dan 0.001 secara konsisten menghasilkan hasil yang paling stabil di seluruh metrik evaluasi. *Learning rate* besar seperti 0.1 menghasilkan performa yang lebih fluktuatif dan cenderung menurunkan akurasi dan *recall*, terutama pada nilai *epoch* 150.

3.6 Perbandingan Hasil Kelima Skema

Tabel 8. Perbandingan Hasil Kelima Skema

Skema	Accuracy	Precision	Recall	F1-score
1	0.95	0.04	0.94	0.09
2	0.83	0.84	0.93	0.80
3	0.89	0.86	0.95	0.89
4	0.96	1.00	0.94	0.96
5	0.96	1.00	0.95	0.96

Tabel 8 menunjukkan hasil ke 5 skema, dimana pada skema 1, menunjukkan bahwa model hanya memperoleh presisi sebesar 0.04 dan *F1-score* sebesar 0.09, meskipun akurasi mencapai 0.95. Hal ini menunjukkan bahwa tanpa penyeimbangan dan normalisasi, model cenderung mengalami bias terhadap kelas mayoritas sehingga gagal mendeteksi kelas minoritas (stroke) dengan baik. Kinerja mulai meningkat pada skema 2 dengan kombinasi metode SMOTE dan *MinMaxScaler*, dimana hasil akurasi yang diperoleh sebesar 0.83, presisi 0.84, dan *recall* 0.93. Kemudian pada skema ke 3 performa semakin membaik, di mana penerapan metode ADASYN dan normalisasi *Z-Score* memberikan peningkatan akurasi menjadi 0.89 dan *F1-score* menjadi 0.89. Peningkatan paling signifikan terlihat pada skema dengan *hyperparameter tuning*. Dalam skema 4, yang menggunakan metode SMOTE, *MinMaxScaler*, serta *tuning learning rate* dan *epoch*, menunjukkan hasil yang sangat baik dengan akurasi 0.96, presisi 1.00, dan *F1-score* 0.96. Skema 5 juga menunjukkan performa yang sangat baik dengan kombinasi ADASYN, *Z-Score*, dan *tuning* parameter, menghasilkan akurasi 0.96, presisi 1.00, *recall* 0.95, dan *F1-score* 0.96.



Gambar 2. Grafik Perbandingan Semua Metrix Antar Skema

Gambar 2 adalah Grafik perbandingan kelima skema dimana menunjukkan perubahan yang sangat signifikan pada performa setelah dilakukan serangkaian teknik *preprocessing* dan parameter *tuning*. Pada Skema 1, model dilatih tanpa *balancing* dan normalisasi sehingga performanya sangat tidak stabil, akurasi cenderung menipu, dan model hampir selalu gagal mendeteksi kelas stroke karena beratnya ketidakseimbangan data. Kondisi ini menunjukkan bahwa data mentah dalam keadaan tidak seimbang tidak dapat digunakan secara efektif untuk melatih BPNN. Pada Skema 2, dengan diterapkannya SMOTE dan normalisasi menunjukkan perubahan yang sangat signifikan, di mana akurasi, presisi, *recall*, dan *F1-score* meningkat secara stabil, artinya model mulai mampu mengenali kelas stroke pada tingkat yang lebih baik. Selanjutnya, skema 3, yang menggunakan ADASYN dan *Z-Score*, menunjukkan kinerja yang lebih baik daripada SMOTE karena ADASYN mampu mendapatkan sampel sintesis secara adaptif pada area yang memiliki risiko tingkat tinggi, artinya *recall* naik lebih signifikan dan model makin sensitif terhadap kasus stroke.

Performa terbaik dicapai pada skema 4 dan skema 5, dimana proses *parameter tuning* diterapkan bersama *5-fold cross-validation*. Pada Skema 4, kombinasi antara metode SMOTE, *MinMaxScaler*, dan *tuning* dapat memberikan akurasi tinggi dengan performa yang sangat stabil. Skema 5 menunjukkan hasil paling optimal di seluruh metrik, dengan akurasi, presisi, *recall*, dan *F1-score* tertinggi serta paling seimbang. Hal ini menunjukkan bahwa kombinasi



ADASYN dan *Z-Score* lebih efektif dalam membentuk representasi data yang baik sebelum model memasuki proses pelatihan dengan parameter optimal.

3.7 Pembahasan

Hasil dari penelitian ini menunjukkan bahwa kinerja *backpropagation neural network* dalam klasifikasi penyakit stroke sangat dipengaruhi dari strategi *preprocessing* data dan *hyperparameter tuning* pelatihan yang diterapkan. Skema 1 menunjukkan hasil performa yang tidak memuaskan karena model diberikan data secara mentah tanpa *balancing* dan tanpa melalui proses normalisasi data. Terlepas dari tingginya akurasi pada beberapa kali uji yang dilakukan, nilai presisi, *recall*, dan *F1-score* yang dihasilkan sangat rendah, bahkan mendekati nol. Kondisi ini menunjukkan bahwa model masih memiliki kemampuan yang buruk dalam menggeneralisasi data dan cenderung memasukkan semua data yang ada ke dalam kelas mayoritas, sehingga akurasi yang tinggi bersifat menyesatkan. Selain itu, perbedaan skala antar data ternyata juga memberikan dampak pada kinerja model. Karena *feature* yang memiliki nilai tinggi secara mutlak akan tidak berperan dalam pelatihan model, kondisi ini membuat beberapa area yang tidak tercakup oleh pelatihan menyebabkan model tidak dapat mempelajari pola yang kompleks. Untuk alasan-alasan tersebut, hasil yang didapat dari skema 1 tidak bisa dianggap optimal. Model sebenarnya sama sekali tidak mampu membantu dalam mendeteksi kasus penyakit stroke.

Setelah dilakukan *balancing* menggunakan metode SMOTE dan normalisasi data menggunakan *MinMax Scaller*, performa model mulai meningkat. Peningkatan distribusi kelas dan perbaikan skala pada *feature* memungkinkan model untuk belajar lebih banyak mengenai pola stroke. Nilai akurasi pun *start* stabil di atas 0.80, sementara kenaikan presisi, *recall*, dan *F1-score* juga *start* seimbang. Akan tetapi, SMOTE belum optimal dalam representasi variasi pola yang ada di area data minoritas yang kompleks, sehingga pada beberapa kondisi model *start* optimal dalam mentransfer pola kompleks sebagai bagian dari kelas data stroke.

Performa yang lebih baik ditunjukkan pada Skema 3 melalui penerapan ADASYN dan normalisasi *Z-Score*. ADASYN sebagai metode oversampling adaptif mampu menghasilkan sampel sintetis secara lebih selektif pada area data minoritas yang sulit dipelajari, sehingga representasi kelas stroke menjadi lebih kaya. Di sisi lain, normalisasi *Z-Score* membantu menstabilkan proses pelatihan dengan mempertimbangkan distribusi statistik data. Kombinasi kedua teknik ini menghasilkan peningkatan akurasi hingga mendekati 0,90 serta peningkatan signifikan pada *recall*, yang pada beberapa konfigurasi mencapai lebih dari 95%. Peningkatan ini menunjukkan bahwa model semakin efektif dalam mendeteksi kasus stroke dan mampu menekan risiko *false negative*, yang merupakan aspek krusial dalam konteks diagnosis penyakit. Dibandingkan dengan Skema 2, Skema 3 secara konsisten menunjukkan keunggulan pada seluruh metrik evaluasi.

Peningkatan performa yang paling signifikan diperoleh pada Skema 4 dan Skema 5 setelah dilakukan *hyperparameter tuning*, *learning rate* dan jumlah *epoch* menggunakan *5-fold cross-validation*. Adapun pada Skema 4, SMOTE ditambah dengan *MinMaxScaler*, didamping *tuning* parameter, mampu mengungguli nilai akurasi hingga melebihi 0,96, presisi sangat tinggi, *recall* stabil. Hal ini membuktikan bahwa parameter yang tepat dapat meningkatkan generalisasi model dan mengurangi risiko *overfitting*. Secara *overview*, Skema 5 tetap mencapai performa terbaik, di mana ADASYN dan *Z-Score* dikombinasikan dengan proses *tuning* parameter. Pada skema ini, model mencapai akurasi tertinggi sebesar 0,9647 dengan nilai *F1-score* sekitar 0,964, yang menunjukkan keseimbangan yang baik antara presisi dan *recall*.

Berdasarkan hasil penelitian ini, menegaskan bahwa *preprocessing* merupakan faktor kunci dalam menentukan performa BPNN untuk klasifikasi stroke. Model yang dilatih tanpa *balancing* dan normalisasi terbukti tidak mampu mengenali pola kelas minoritas secara memadai. Penerapan metode SMOTE dan *MinMaxScaler* dapat meningkatkan performa model, namun pendekatan yang lebih adaptif seperti ADASYN serta normalisasi berbasis distribusi seperti *Z-Score* memberikan hasil yang lebih konsisten dan optimal. Selain itu, *tuning* parameter berperan penting dalam memaksimalkan kemampuan model. Temuan ini menunjukkan bahwa kombinasi *preprocessing* adaptif, normalisasi yang tepat, dan pengaturan parameter yang optimal sangat diperlukan untuk menghasilkan model klasifikasi medis yang akurat, stabil dan memiliki kemampuan generalisasi yang baik sehingga dapat diandalkan dalam membantu proses pengambilan keputusan klinis secara lebih efektif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa *preprocessing* dan *hyperparameter tuning* memiliki pengaruh yang sangat signifikan terhadap performa BPNN dalam klasifikasi stroke. Tanpa *balancing* dan normalisasi (Skema 1), model gagal mengenali kelas stroke. Penerapan SMOTE + *MinMaxScaler* (Skema 2) dan terutama ADASYN + *Z-Score* (Skema 3) memberikan peningkatan yang kuat pada akurasi, *recall*, dan *F1-score*. Pada skema 4 performa stabil dengan peningkatan akurasi yang lebih signifikan dari skema sebelumnya. Skema 5 adalah skema yang paling terbaik yaitu kombinasi antara ADASYN, *Z-Score*, *tuning learning rate* dan *epoch*, diperoleh dengan akurasi sebesar 96.4%. Keunggulan kombinasi ini dipengaruhi oleh kemampuan ADASYN dalam menghasilkan data sintetis secara adaptif pada area perbatasan kelas yang sulit dipelajari, sehingga representasi kelas minoritas menjadi lebih baik dan kemampuan model dalam mendeteksi kasus stroke meningkat. Berdasarkan hasil ini menegaskan bahwa teknik oversampling adaptif, normalisasi berbasis distribusi, serta *tuning* parameter yang tepat, sangat efektif dalam meningkatkan kualitas model klasifikasi stroke. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Evaluasi model dilakukan hanya pada satu dataset stroke,



sehingga kemampuan generalisasi metode terhadap dataset lain dengan karakteristik berbeda belum dapat dipastikan sepenuhnya. Selain itu, penelitian ini masih berfokus pada satu jenis algoritma jaringan saraf, yaitu BPNN, tanpa melakukan perbandingan langsung dengan algoritma klasifikasi lain. Oleh karena itu, sebagai arah penelitian selanjutnya, performa model dapat ditingkatkan dengan menguji pendekatan *preprocessing* dan *tuning* parameter yang diusulkan pada dataset stroke lain atau dataset medis berbeda, seperti penyakit jantung, diabetes, atau penyakit kronis lainnya, untuk mengevaluasi konsistensi dan generalisasi metode. Penelitian selanjutnya juga dapat membandingkan kinerja BPNN dengan algoritma klasifikasi lain yang telah terbukti efektif pada data medis dan tekstual, seperti *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Random Forest*, atau *Gradient Boosting*, yang pada penelitian terdahulu menunjukkan performa tinggi ketika dikombinasikan dengan teknik *balancing* dan *hyperparameter tuning*.

REFERENCES

- [1] A. F. Eram, A. S. Mahmud, M. M. Khadem, and M. A. Ihsan, "Beyond the numbers: App-enabled stroke prediction system for high-risk individuals in imbalanced datasets," *Neurosci. Informatics*, vol. 5, no. 3, p. 100215, 2025, doi: 10.1016/j.neuri.2025.100215.
- [2] S. Felehgari, P. Sariaslani, S. Shamsizadeh, S. Felehgari, A. Rajabi, and H. Mohammadi, "Multi-classification Deep Learning Approach for Diagnosing Stroke Type and Severity Using Multimodal Magnetic Resonance Images," *J. Med. Signals Sens.*, vol. 15, no. 4, pp. 1–7, 2025, doi: 10.4103/jmss.jmss_37_24.
- [3] M. Azhima, I. Afrianty, E. Budianita, and S. Kurnia Gusti, "KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Metode Backpropagation Neural Network untuk Klasifikasi Penyakit Stroke," *Media Online*, vol. 4, no. 6, pp. 3013–3021, 2024, doi: 10.30865/klik.v4i6.1956.
- [4] X. Tang, M. Tang, W. Liu, and S. Cui, "Explainable machine learning for stroke risk prediction: a comparative study with SHAP-based interpretation," *Front. Neurol.*, vol. 16, no. January, pp. 1–12, 2026, doi: 10.3389/fneur.2025.1716984.
- [5] Solikhun and N. Amalya, "Algoritma Backpropagation Metode Levenberg Marquardt Dalam Memprediksi Penyakit Stroke," *Bull. Comput. Sci. Res.*, vol. 3, no. 2, pp. 191–196, 2023, doi: 10.47065/bulletincsr.v3i2.229.
- [6] Alwaliyanto, Siska Kurnia Gusti, Iis Afrianty, and Fadhilah Syafria, "Penerapan Metode ADASYN Dalam Mengatasi Imbalanced Data Untuk Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine," *Bull. Comput. Sci. Res.*, vol. 5, no. 4, pp. 532–541, 2025, doi: 10.47065/bulletincsr.v5i4.612.
- [7] M. Resa Arif Yudianto, P. Sukmasetya, R. Abul Hasani, and D. Sasongko, "Pengaruh Data Preprocessing terhadap Imbalanced Dataset pada Klasifikasi Citra Sampah menggunakan Algoritma Convolutional Neural Network," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1367–1375, 2022, doi: 10.47065/bits.v4i3.2575.
- [8] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Informatics Med. Unlocked*, vol. 25, p. 100690, 2021, doi: 10.1016/j.imu.2021.100690.
- [9] A. M. Sowjanya and O. Mrudula, "Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms," *Appl. Nanosci.*, vol. 13, no. 3, pp. 1829–1840, 2023, doi: 10.1007/s13204-021-02063-4.
- [10] J. Zhu *et al.*, "Processing imbalanced medical data at the data level with assisted-reproduction data as an example," *BioData Min.*, vol. 17, no. 1, 2024, doi: 10.1186/s13040-024-00384-y.
- [11] V. S. Elangovan, R. Devarajan, O. I. Khalaf, M. S. Sharif, and W. Elmedany, "Analyzing an Imbalanced Stroke Prediction Dataset Using Machine Learning Techniques," *Karbala Int. J. Mod. Sci.*, vol. 10, no. 2, pp. 246–259, 2024, doi: 10.33640/2405-609X.3355.
- [12] E. F. Agyemang *et al.*, "Addressing Class Imbalance Problem in Health Data Classification: Practical Application From an Oversampling Viewpoint," *Appl. Comput. Intell. Soft Comput.*, vol. 2025, no. 1, 2025, doi: 10.1155/acis/1013769.
- [13] S. rizki Zikrillah aulia, O. Okfalisa, E. Haerani, and L. Oktavia, "Application of ADASYN Technique in Classification of Stroke Disease using Backpropagation Neural Network," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 3, pp. 1666–1674, 2025, doi: 10.35314/jdhv9s39.
- [14] R. Azhar, S. K. Gusti, I. Afrianty, and ..., "Perbandingan Teknik Penyeimbang Kelas Pada Multi-Layer Perceptron (MLP) Berbasis Backpropagation Untuk Klasifikasi Diabetes Mellitus," *Bull. Comput. ...*, vol. 5, no. 6, pp. 1304–1314, 2025, doi: 10.47065/bulletincsr.v5i6.804.
- [15] N. Cahyani, R. Irsyada, and R. Mahmuda, "Penerapan Algoritma Neural Network untuk Klasifikasi Diabetes Mellitus: Perbandingan Backpropagation dan Resilient Backpropagation," *Digit. Transform. Technol.*, vol. 4, no. 2, pp. 1067–1074, 2025, doi: 10.47709/digitech.v4i2.5208.
- [16] W. H. Herowati, R. A. Premunendar, and H. Al Azies, "Optimalisasi metode BPNN (Backpropagation Neural Network) menggunakan GA (Genetic Algorithm) dalam menentukan arah offset pada metode ekstraksi fitur GLCM (Gray Level Co-occurrence Matrices)," *Bina Insa. Ict J.*, vol. 10, no. 2, p. 123, 2023, doi: 10.51211/biict.v10i2.2624.
- [17] N. Melnykova, Y. Patereha, S. Skopivskyi, M. Farion, S. Fedushko, and K. Drohomlyretska, "Machine learning for stroke prediction using imbalanced data," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-01855-w.
- [18] A. Firdaus, rif'an, M. Lutfi, and M. Amrulloh, Faishol, "Klasifikasi Jenis Kelengkeng Berdasarkan Morfologi Daun Dengan Ekstraksi Ciri RRGB, GLCM," *J. Keilmuan dan Apl. Tek. Inform.*, vol. 1(2), no. 2, pp. 110–125, 2022, doi: <https://doi.org/10.311102/jatim.v4i2.2341>.
- [19] F. R. Aftha Harianto, Z. Alawi, and I. A. Sa'ida, "Pengaruh Komposisi Split Data Pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 36–44, 2025, doi: 10.47080/simika.v8i1.3663.
- [20] M. A. Saleem *et al.*, "Enhancing stroke risk prediction through class balancing and data augmentation with CBDA-ResNet50," *Sci. Rep.*, vol. 15, no. 1, pp. 1–19, 2025, doi: 10.1038/s41598-025-07350-6.