

KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN PENDEKATAN HYBRID INFORMATION GAIN dan BACKPROPAGATION NEURAL NETWORK (BPNN)

Heart Disease Classification Using a Hybrid Approach of Information Gain and Backpropagation Neural Network (BPNN)

Azis Wahyudi¹, Haryo Nugroho², Harinto Nur Seha³, Rina Yulida⁴

^{1,2,3,4}Prodi Rekam Medis dan Informasi Kesehatan Politeknik Kesehatan Permata Indonesia
Yogyakarta

Email : azis@permataindonesia.ac.id, haryo@permataindonesia.ac.id
, harinto_ns@permataindonesia.ac.id, rina@permataindonesia.ac.id

Abstrak

Penyakit jantung merupakan penyebab utama kematian di seluruh dunia. Untuk mendeteksi risiko penyakit jantung secara dini, dibutuhkan metode klasifikasi yang akurat dan efisien. Penelitian ini mengusulkan pendekatan hybrid dengan menggabungkan seleksi fitur Information Gain (IG) dan algoritma klasifikasi Backpropagation Neural Network (BPNN). Dataset yang digunakan adalah Heart Disease dari UCI Repository dengan total 303 data pasien. Sebanyak 8 fitur terbaik dipilih menggunakan Information Gain. Model BPNN dilatih menggunakan parameter `hidden_layer_sizes=(16, 8)`, `activation='relu'`, dan `learning_rate_init=0.01`. Hidden layer dengan 16 dan 8 neuron memungkinkan jaringan mempelajari pola kompleks, ReLU mempercepat konvergensi pelatihan, dan learning rate mengatur kecepatan pembaruan bobot. Hasil evaluasi menunjukkan model mencapai akurasi 79.12%, presisi 84.44%, recall 76.00%, F1-Score 80.00%, dan AUC 85.95%. Validasi silang 5-Fold menghasilkan rata-rata akurasi 82.15%. Hasil ini menunjukkan bahwa pendekatan IG + BPNN memberikan performa klasifikasi yang baik dan stabil dalam mendeteksi penyakit jantung.

Kata Kunci: Penyakit Jantung, Information Gain, Backpropagation Neural Network, Klasifikasi, Data Mining

Abstract

Heart disease is the leading cause of death worldwide. To detect the risk of heart disease early, an accurate and efficient classification method is needed. This study proposes a hybrid approach by combining Information Gain (IG) feature selection and the Backpropagation Neural Network (BPNN) classification algorithm. The dataset used is the Heart Disease Dataset from the UCI Repository, consisting of 303 patient records. Eight top features were selected using Information Gain. The BPNN model was trained using parameters `hidden_layer_sizes=(16, 8)`, `activation='relu'`, and `learning_rate_init=0.01`. A hidden layer with 16 and 8 neurons enables the network to learn complex patterns in the data. The ReLU (Rectified Linear Unit) activation function is used to speed up training convergence and avoid the vanishing gradient problem. The `learning_rate_init=0.01` parameter controls the speed of weight updates during the learning process, affecting the model's stability and convergence. The evaluation results show that the model achieves 79.12% accuracy, 84.44% precision, 76.00% recall, 80.00% F1-Score, and 85.95% AUC. The 5-Fold Cross Validation yielded an average accuracy of 82.15%. These results indicate that the IG + BPNN hybrid approach provides good and stable classification performance in detecting heart disease.

Keywords: Heart Disease, Information Gain, Backpropagation Neural Network, Classification, Data Mining

PENDAHULUAN

Penyakit jantung merupakan penyebab kematian tertinggi di dunia. Menurut WHO [1], lebih dari 17 juta orang meninggal setiap tahun karena penyakit kardiovaskular. Deteksi dini penyakit jantung menjadi krusial untuk menekan angka kematian dan beban biaya pengobatan. Kecerdasan buatan, khususnya machine learning dan deep learning, menjadi solusi yang menjanjikan dalam bidang prediksi penyakit [2][3][4]. Algoritma Artificial Neural Network (ANN) telah banyak digunakan dalam diagnosis medis karena kemampuannya mengenali pola kompleks [3]. Backpropagation Neural Network (BPNN) adalah bentuk ANN dengan proses pelatihan berbasis perbaikan bobot menggunakan algoritma propagasi balik kesalahan. Namun, banyaknya fitur dalam dataset dapat menyebabkan overfitting dan memperlambat proses pelatihan. Oleh karena itu, diperlukan metode seleksi fitur seperti Information Gain (IG) [5][6].

Penelitian ini bertujuan membangun model klasifikasi penyakit jantung berbasis hybrid IG + BPNN untuk meningkatkan akurasi dan efisiensi klasifikasi.

METODE

A. Data Mining

Data mining adalah proses sistematis untuk mengekstraksi pola-pola bermakna dan informasi tersembunyi dari kumpulan data besar menggunakan teknik statistik, pembelajaran mesin, dan kecerdasan buatan [4]. Dalam konteks penelitian ini, data mining digunakan untuk mengolah dataset medis guna mendeteksi kemungkinan penyakit jantung secara otomatis dan akurat. Proses data mining meliputi pra-pemrosesan data, seleksi fitur, pelatihan model, dan evaluasi performa klasifikasi

[2][4].

B. Model Hybrid IG dan BPNN

Penelitian ini menerapkan pendekatan hybrid yang terdiri atas dua tahapan utama:

1. Seleksi Fitur dengan Information Gain (IG):

Information Gain adalah teknik seleksi fitur berdasarkan teori informasi yang mengukur seberapa besar informasi yang diperoleh sebuah fitur terhadap target. IG memilih fitur yang memiliki kontribusi tertinggi dalam mengurangi ketidakpastian target kelas. Dalam penelitian ini, dari 13 fitur awal, dipilih 8 fitur terbaik yang digunakan untuk pelatihan model [7]. Adapun fitur yang terpilih adalah: chest pain type, serum cholesterol, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels, dan Thalassemia.

2. Klasifikasi dengan Backpropagation Neural Network (BPNN):

BPNN merupakan jenis jaringan saraf tiruan yang belajar dengan memperbarui bobot melalui propagasi balik. Arsitektur yang digunakan adalah:

- Hidden Layer 1: 16 neuron
- Hidden Layer 2: 8 neuron
- Fungsi aktivasi: ReLU (Rectified Linear Unit)
- Optimizer: Adam
- Learning rate awal: 0.01
- Maksimum iterasi (epoch): 1000
- Fungsi loss: log-loss (cross

entropy)

Skema model ditunjukkan pada rancangan diagram proses berikut (disisipkan di bagian Rancangan Model).

C. Evaluasi Model

Model dievaluasi menggunakan data uji (testing) dan divalidasi menggunakan teknik 5-Fold Cross Validation. Evaluasi dilakukan berdasarkan beberapa metrik klasifikasi sebagai berikut:

1. Akurasi

Akurasi mengukur seberapa banyak prediksi yang benar dari total prediksi yang dilakukan.

$$\text{Akurasi} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

2. Presisi

Presisi mengukur berapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi positif.

$$\text{Presisi} = \text{TP} / (\text{TP} + \text{FP})$$

3. Recall (Sensitivitas)

Recall atau sensitivitas mengukur kemampuan model dalam menemukan data positif secara benar.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

4. F1-Score

F1-Score adalah rata-rata harmonis antara presisi dan recall, yang berguna saat data tidak seimbang.

$$\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$$

5. Spesifisitas

Spesifisitas mengukur kemampuan model dalam mengenali kelas negatif dengan benar.

$$\text{Spesifisitas} = \text{TN} / (\text{TN} + \text{FP})$$

6. AUC (Area Under Curve)

AUC mengukur seberapa baik model dapat membedakan antara kelas positif dan negatif. Dihitung menggunakan:

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}), \text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

7. ROC (Receiver Operating Characteristic)

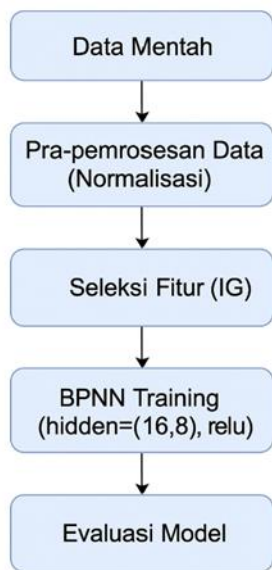
ROC adalah kurva yang menggambarkan True Positive Rate (TPR) terhadap False Positive Rate (FPR) pada berbagai nilai ambang (threshold).

Hasil Validasi

Validasi menggunakan metode 5-Fold Cross Validation menghasilkan rata-rata akurasi sebesar 82.15%, yang menunjukkan performa model yang cukup baik dalam melakukan klasifikasi.

D. Rancangan Model

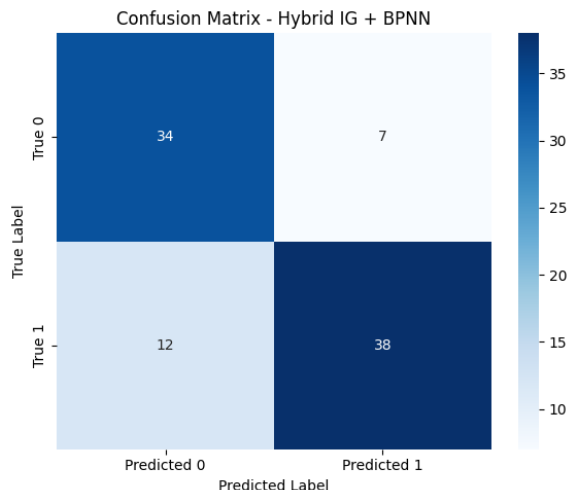
Rancangan atau model pada penelitian ini menggambarkan alur proses klasifikasi penyakit jantung yang dilakukan secara sistematis, mulai dari pengumpulan data hingga evaluasi model. Diagram alir (flowchart) pada gambar 1. berikut menunjukkan tahapan utama dalam metode hybrid Information Gain dan Backpropagation Neural Network (BPNN):



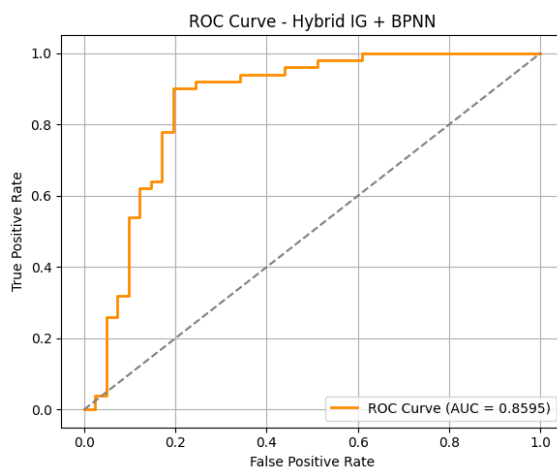
Gambar 1. Flowchart model

HASIL

Model hybrid yang dibangun menggunakan seleksi fitur Information Gain (IG) dan klasifikasi Backpropagation Neural Network (BPNN) berhasil dilatih pada dataset Heart Disease UCI dengan 8 fitur terbaik. Konfigurasi model menggunakan 2 hidden layer (16 dan 8 neuron), fungsi aktivasi ReLU, dan learning rate sebesar 0.01 selama 1000 iterasi. Dapat dilihat pada gambar 2.



Gambar 2. Confusion Matrix-Hybrid + BPNN



Gambar 3. Grafik Kurva ROC

Pada gambar 3. Grafik tersebut menunjukkan kurva ROC (Receiver Operating Characteristic) dari model klasifikasi penyakit jantung yang menggunakan pendekatan hybrid antara Information Gain (IG) sebagai seleksi fitur dan Backpropagation Neural Network (BPNN) sebagai algoritma klasifikasi

Tabel 1. Hasil Evaluasi Model Hybrid IG + BPNN

Akurasi	Presisi	Recall	<i>F1-Score</i>	<i>AUC Score</i>
0.7912	0.8444	0.7600	0.8000	0.8595

Selain itu, dilakukan validasi menggunakan

5-Fold Cross Validation, dengan hasil sebagai pada table 2.

Tabel 2. Hasil 5-Fold Cross Validation

Fold	Akurasi
1	0.8361
2	0.8197
3	0.8852
4	0.8333
5	0.7333

PEMBAHASAN

Hasil menunjukkan bahwa metode hybrid IG + BPNN memberikan kinerja klasifikasi yang baik. Nilai akurasi sebesar 79.12% menunjukkan bahwa mayoritas prediksi model tepat sasaran. Presisi yang tinggi (84.44%) menandakan sedikit kasus false positive, artinya model cukup tepat dalam memprediksi pasien yang memang berisiko. Recall sebesar 76.00% menunjukkan kemampuan model dalam menangkap kasus nyata penyakit jantung dengan cukup baik.

Perbandingan dengan studi sebelumnya yang menggunakan algoritma tunggal menunjukkan bahwa pendekatan hybrid mampu meningkatkan performa. Validasi silang menghasilkan akurasi rata-rata 82.15%, menunjukkan konsistensi dan kemampuan generalisasi model terhadap data baru.

Dengan nilai AUC sebesar 85.95%, model terbukti efektif dalam mengidentifikasi perbedaan antar kelas. Hal ini penting untuk diterapkan dalam sistem pendukung keputusan medis guna membantu diagnosis awal penyakit jantung.

KESIMPULAN

Penelitian ini telah berhasil mengembangkan model klasifikasi penyakit jantung menggunakan pendekatan Hybrid Information Gain (IG) dan Backpropagation Neural Network (BPNN). Dari 13 fitur yang tersedia, sebanyak 8 fitur terbaik dipilih berdasarkan nilai Information Gain untuk meningkatkan efisiensi dan efektivitas klasifikasi.

Model BPNN dengan konfigurasi dua hidden layer (16 dan 8 neuron), fungsi aktivasi ReLU, dan learning rate 0.01 mampu menghasilkan performa yang baik. Hasil validasi dengan 5-Fold Cross Validation juga menunjukkan akurasi rata-rata sebesar 82.15%, yang mengindikasikan model memiliki kestabilan dan generalisasi yang baik terhadap data baru.

Dengan demikian, dapat disimpulkan bahwa metode hybrid IG + BPNN efektif dalam meningkatkan performa klasifikasi penyakit jantung. Model ini berpotensi untuk diterapkan dalam sistem pendukung keputusan medis guna membantu proses diagnosis dini secara cepat dan akurat.

REFERENCE

1. World Health Organization (WHO). Cardiovascular diseases (CVDs) – Key facts. Geneva: WHO; 2021. [Online]. Tersedia di: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
2. Alsharif, M. H., Alsharif, Y. H., & Kim, J. (2020). Artificial Intelligence-Based Heart Disease Prediction System Using Backpropagation Neural Network and Genetic Algorithm. *Journal of Healthcare*

-
- Engineering, 2020, Article ID 1406143. <https://doi.org/10.1155/2020/1406143>
3. Detrano, R. et al. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *The American Journal of Cardiology*, 64(5), 304–310. [https://doi.org/10.1016/0002-9149\(89\)90524-9](https://doi.org/10.1016/0002-9149(89)90524-9)
 4. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann.
 5. Kaur, H., & Wasan, S. K. (2006). Empirical Study on Applications of Data Mining Techniques in Healthcare. *Journal of Computer Science*, 2(2), 194–200. <https://doi.org/10.3844/jcssp.2006.194.200>
 6. Setiawan, R., Nugroho, H. A., & Wicaksono, A. F. (2017). Feature Selection Using Information Gain for Heart Disease Classification. *International Journal of Computer Applications*, 170(8), 10–14. <https://doi.org/10.5120/ijca201791478>
 7. Sony, R. 2020. UCI Heart Disease Data Heart Disease Data Set from UCI data repository. July 10, 2025 from <https://www.kaggle.com/datasets/redwanankarimsony/heart-disease-data>.