

PREDIKSI PROGNOSIS KANKER PAYUDARA MENGUNAKAN HYBRID ANN-KNN

Revania Juniarta Siahaan✉, Margaretha Yohanna, Harlen Gilbert Simanullang

Fakultas Ilmu Komputer, Universitas Methodist Indonesia, Medan, Indonesia

Email: revasiahaan66@gmail.com

DOI: <https://doi.org/10.46880/methoda.Vol16No2.pp221-229>

ABSTRACT

Breast cancer is a disease with a high incidence and mortality rate worldwide. Determining patient prognosis is often challenging due to the complex clinical factors involved. This study aims to develop a breast cancer prognosis prediction model using a hybrid Artificial Neural Network (ANN) and K-Nearest Neighbors (KNN) method. The dataset used is the Breast Cancer Wisconsin Prognostic (WPBC) obtained from the UCI Machine Learning Repository, comprising 198 patient data with 35 attributes. The research steps include data cleaning, label transformation, normalization using Min-MAX Scaler, feature extraction using ANN and classification using KNN. The optimal K-value was determined by testing K=1, K=3, K=5, and K=7 using Stratified K-Fold Cross-Validation. Test results indicate that the highest accuracy of 84,62% was achieved in the fifth fold with K=7.

Keywords: Breast Cancer, Artificial Neural Network, K-Nearest Neighbors.

ABSTRAK

Kanker payudara merupakan salah satu penyakit yang memiliki tingkat kejadian dan angka kematian yang cukup tinggi di dunia. Penentuan prognosis pasien sering kali menjadi tantangan karena melibatkan banyak faktor klinis yang kompleks. Penelitian ini bertujuan untuk membangun model prediksi prognosis kanker payudara menggunakan metode hybrid Artificial Neural Network (ANN) dan K-Nearest Neighbors (KNN). Dataset yang digunakan adalah Breast Cancer Wisconsin Prognostic (WPBC) yang diperoleh dari UCI Machine Learning Repository sebanyak 198 data pasien dengan 35 atribut. Tahapan penelitian meliputi pembersihan data, transformasi label, normalisasi menggunakan Min-Max Scaler, ekstraksi fitur menggunakan ANN, dan klasifikasi menggunakan KNN. Penentuan nilai K dilakukan dengan menguji K=1, K=3, K=5, K=7 menggunakan Stratified K-Fold Cross-Validation. Hasil pengujian menunjukkan bahwa akurasi tertinggi sebesar 84,62% diperoleh pada fold ke-5 dengan K=7.

Kata Kunci: Kanker Payudara, Artificial Neural Network, K-Nearest Neighbors.

PENDAHULUAN

Kanker payudara merupakan salah satu jenis kanker yang paling sering dialami oleh perempuan di seluruh dunia. Penyakit ini tidak hanya berdampak pada kondisi fisik, tetapi juga memberikan pengaruh besar terhadap kondisi psikologis, sosial, dan kualitas hidup penderitanya. Banyak pasien yang harus menjalani proses pengobatan yang panjang, mulai dari diagnosis hingga pemantauan setelah pengobatan selesai, sehingga menjadikan kanker

payudara sebagai salah satu masalah kesehatan yang serius.

Menurut data dari World Health Organization (WHO), pada tahun 2022 tercatat sekitar 2,3 juta kasus baru kanker payudara secara global. Angka ini menunjukkan bahwa kanker payudara masih menjadi jenis kanker dengan tingkat kejadian yang sangat tinggi di dunia.

Di Indonesia, kanker payudara juga menjadi salah satu jenis kanker dengan jumlah kasus yang tinggi. Berdasarkan data dari Kementerian

Kesehatan Republik Indonesia, pada tahun 2022 terdapat sekitar 66.271 kasus baru kanker payudara dengan angka kematian mencapai 22.298 jiwa. Hal ini menunjukkan bahwa kanker payudara masih menjadi ancaman serius dalam sistem kesehatan nasional.

Selain tingginya angka kejadian, risiko kekambuhan setelah pengobatan juga menjadi permasalahan penting dalam kanker payudara. Kekambuhan dapat terjadi dalam waktu yang berbeda-beda, baik dalam jangka pendek maupun setelah beberapa tahun pasien dinyatakan sembuh. Kondisi ini membuat kanker payudara menjadi penyakit yang perlu mendapatkan perhatian khusus (Cahyani et al., 2025)(Arianda et al., 2023).

Penentuan prognosis bukanlah hal yang mudah karena melibatkan banyak faktor yang kompleks. Faktor-faktor tersebut meliputi ukuran tumor, bentuk sel kanker, serta karakteristik lainnya. Kompleksitas ini dapat menyebabkan perbedaan dalam hasil penilaian antar tenaga medis (Samarov & Barotova, 2025).

Perkembangan teknologi kecerdasan buatan atau Artificial Intelligence (AI) telah memberikan solusi dalam membantu menganalisis data yang kompleks. AI mampu mengolah data dalam jumlah besar dan menemukan pola yang sulit dikenali secara manual, sehingga dapat meningkatkan akurasi dalam proses prediksi (Arianda et al., 2023).

Salah satu pendekatan dalam AI yang banyak digunakan adalah machine learning. Machine learning memungkinkan sistem untuk belajar dari data yang tersedia dan menghasilkan model yang dapat digunakan untuk melakukan prediksi. Dalam bidang kesehatan, metode ini telah banyak digunakan untuk membantu diagnosis dan prognosis berbagai penyakit, termasuk kanker payudara (Kurniati et al., 2024).

Dalam penelitian ini, digunakan pendekatan model hybrid yang menggabungkan Artificial Neural Network (ANN) dan K-Nearest Neighbors (KNN). ANN digunakan karena kemampuannya dalam mempelajari pola data yang kompleks dan menghasilkan representasi fitur yang lebih baik (Kurniati et al., 2024).

Sementara itu, KNN merupakan metode klasifikasi yang sederhana namun cukup efektif.

KNN bekerja dengan menentukan kelas suatu data berdasarkan kedekatannya dengan data lain menggunakan perhitungan jarak tertentu (Ujianto et al., 2025).

Penggunaan model hybrid dipilih karena masing-masing metode memiliki keterbatasan. ANN dapat mengalami penurunan performa pada data tertentu, sedangkan KNN dapat mengalami penurunan kinerja jika digunakan pada data dengan jumlah fitur yang besar. Dengan menggabungkan kedua metode ini, diharapkan dapat meningkatkan akurasi model dalam memprediksi prognosis kanker payudara (Samarov & Barotova, 2025).

Beberapa penelitian sebelumnya telah menerapkan metode machine learning untuk prediksi dan klasifikasi kanker payudara. Jika dibandingkan dengan penggunaan metode tunggal, pendekatan hybrid cenderung memberikan hasil yang lebih stabil. Hal ini sejalan dengan penelitian yang menyatakan bahwa kombinasi dua metode dapat mengurangi kelemahan masing-masing metode. ANN mampu menangani kompleksitas data, sedangkan KNN mampu melakukan klasifikasi dengan pendekatan berbasis kedekatan data, sehingga keduanya saling melengkapi (Samarov & Barotova, 2025).

Penelitian oleh Cahyani dkk, menyatakan bahwa performa model klasifikasi sangat dipengaruhi oleh kualitas fitur dan distribusi data. Hal ini mendukung penggunaan ANN sebagai ekstraksi fitur dalam penelitian ini, karena mampu menghasilkan representasi data yang lebih baik sebelum proses klasifikasi dilakukan (Cahyani et al., 2025).

Selanjutnya, penelitian oleh Fahrurrozi dkk, menunjukkan bahwa ANN mampu mencapai akurasi hingga 97% dalam klasifikasi kanker payudara. Hal ini menunjukkan bahwa ANN memiliki kemampuan yang sangat baik dalam mempelajari pola kompleks pada data medis, sehingga sangat tepat digunakan sebagai tahap ekstraksi fitur (Fahrurrozi et al., 2025).

Di sisi lain, penelitian oleh Rihastuti dkk, menunjukkan bahwa KNN mampu mencapai akurasi hingga 96,49% dengan pemilihan nilai K yang tepat. Hal ini memperkuat alasan penggunaan KNN sebagai metode klasifikasi dalam penelitian ini (Rihastuti et al., 2025).

Selain itu, penelitian Cristianes dkk, menyatakan bahwa KNN memiliki performa yang stabil dan kompetitif dibandingkan algoritma lain, terutama ketika fitur yang digunakan memiliki kualitas yang baik. Hal ini menunjukkan bahwa kombinasi ANN-KNN dalam penelitian ini sudah tepat, karena ANN meningkatkan kualitas fitur dan KNN memanfaatkan fitur tersebut untuk klasifikasi (Cristianes et al., 2023).

Penelitian oleh Jamaluddin dkk, juga menunjukkan bahwa ANN efektif dalam mempelajari pola data kanker payudara (Jamaluddin et al., 2024), sedangkan penelitian Kurniati dkk, menyatakan bahwa hidden layer pada ANN mampu menghasilkan fitur baru yang lebih informatif. Hal ini semakin memperkuat bahwa penggunaan ANN sebagai ekstraksi fitur memberikan kontribusi besar terhadap performa model (Kurniati et al., 2024).

Menurut Zuriati dkk, KNN tidak memerlukan proses pelatihan yang kompleks sehingga cocok digunakan sebagai metode klasifikasi akhir. Namun, performanya sangat bergantung pada kualitas fitur yang digunakan. Oleh karena itu, kombinasi dengan ANN menjadi solusi yang efektif (Qomariyah, 2023).

Penelitian oleh Yordanus dan Wenny menggunakan pendekatan machine learning dan deep learning untuk diagnosis serta prognosis kanker payudara dan menunjukkan bahwa ANN mampu memberikan performa yang baik terutama setelah dilakukan penanganan terhadap ketidakseimbangan data (Yordanus & Wenny, 2024).

Meskipun demikian, sebagian besar penelitian sebelumnya masih menggunakan ANN atau KNN sebagai metode tunggal. Penggunaan ANN sebagai ekstraksi fitur yang kemudian dikombinasikan dengan KNN sebagai metode klasifikasi masih relatif jarang diterapkan pada prediksi prognosis kanker payudara. Oleh karena itu, penelitian ini mengusulkan model hybrid ANN-KNN, dimana ANN digunakan untuk membentuk representasi fitur yang lebih informatif melalui hidden layer, sedangkan KNN digunakan untuk melakukan klasifikasi prognosis berdasarkan fitur yang telah diekstraksi serta mengevaluasi kinerja model yang dihasilkan.

TINJAUAN PUSTAKA

Kanker Payudara

Kanker payudara merupakan salah satu penyakit yang masih menjadi penyebab utama kematian pada perempuan. Penyakit ini ditandai dengan pertumbuhan sel yang tidak normal pada jaringan payudara sehingga dapat berkembang menjadi tumor ganas dan menyebar ke organ lain apabila tidak ditangani dengan baik. Tingginya jumlah kasus serta risiko kekambuhan setelah pengobatan menyebabkan proses penentuan prognosis menjadi bagian penting dalam penanganan pasien. Oleh karena itu, diperlukan suatu metode yang mampu membantu tenaga medis dalam memprediksi kemungkinan kekambuhan secara lebih cepat dan akurat (Ali et al., 2025).

Seiring berkembangnya teknologi komputasi, penerapan machine learning dalam bidang kesehatan semakin banyak digunakan untuk membantu proses analisis data medis. Pemanfaatan machine learning memungkinkan sistem mengenali pola-pola pada data pasien yang sulit diamati secara manual sehingga dapat mendukung proses pengambilan keputusan klinis. Pada kasus kanker payudara, metode machine learning telah banyak diterapkan baik untuk diagnosis maupun prediksi prognosis pasien (Nafiiyah et al., 2026).

Artificial Neural Network (ANN)

ANN merupakan salah satu algoritma yang dirancang menyerupai cara kerja jaringan saraf manusia dalam memproses informasi. ANN memiliki kemampuan untuk mempelajari hubungan yang kompleks antarvariabel melalui proses pelatihan sehingga mampu menghasilkan model prediksi yang baik pada data yang bersifat nonlinier.

Selain digunakan sebagai metode klasifikasi, ANN juga dimanfaatkan sebagai metode ekstraksi fitur. Pada proses ini, hidden layer menghasilkan representasi data baru yang lebih ringkas tetapi tetap mempertahankan informasi penting dari data asli. Representasi tersebut dapat digunakan sebagai masukan bagi algoritma klasifikasi lain sehingga mampu meningkatkan performa model secara keseluruhan (Ali et al., 2025).

Penelitian (Nafiiyah et al., 2026) menunjukkan bahwa jaringan saraf tiruan mampu menghasilkan representasi fitur yang lebih baik pada proses klasifikasi kanker payudara. Kemampuan tersebut menjadikan ANN banyak digunakan dalam penelitian berbasis kecerdasan buatan untuk membantu analisis data medis yang memiliki karakteristik kompleks.

K-Nearest Neighbors (KNN)

KNN merupakan algoritma klasifikasi yang menentukan kelas suatu data berdasarkan kedekatan dengan sejumlah data latih terdekat. Kedekatan tersebut umumnya dihitung menggunakan Euclidean Distance, kemudian kelas mayoritas dari tetangga terdekat digunakan sebagai hasil prediksi.

KNN memiliki kelebihan berupa proses implementasi yang sederhana dan tidak memerlukan proses pelatihan model yang rumit. Akan tetapi, performa KNN sangat dipengaruhi oleh kualitas fitur yang digunakan serta pemilihan nilai K yang tepat. Semakin baik kualitas fitur yang digunakan, semakin tinggi pula tingkat akurasi klasifikasi yang dihasilkan (Santoso et al., 2025).

Penelitian (Arifin et al., 2023) menunjukkan bahwa optimasi parameter pada KNN mampu meningkatkan performa klasifikasi kanker payudara. Hasil tersebut menunjukkan bahwa KNN masih menjadi salah satu metode yang efektif dalam menyelesaikan permasalahan klasifikasi pada data medis apabila didukung oleh proses prapemrosesan dan pemilihan fitur yang baik.

Hybrid ANN-KNN

Model hybrid adalah pendekatan yang menggabungkan dua atau lebih algoritma dengan tujuan memanfaatkan kelebihan masing-masing metode. Pada penelitian ini, ANN digunakan sebagai ekstraksi fitur, sedangkan KNN digunakan sebagai metode klasifikasi. Pendekatan tersebut dipilih karena ANN mampu menghasilkan representasi fitur yang lebih informatif, sementara KNN memiliki kemampuan klasifikasi yang baik ketika menggunakan fitur yang berkualitas.

Beberapa penelitian menunjukkan bahwa kualitas fitur memiliki pengaruh yang besar terhadap keberhasilan proses klasifikasi. Oleh karena itu, penggunaan ANN sebelum proses klasifikasi KNN diharapkan mampu meningkatkan akurasi prediksi dibandingkan apabila KNN digunakan secara langsung pada data awal. Pendekatan serupa juga banyak diterapkan pada penelitian klasifikasi kanker berbasis machine learning karena mampu meningkatkan kemampuan model dalam mengenali pola data yang kompleks ((Putri & Nur, 2023), (Arifin et al., 2023)).

METODE PENELITIAN

Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan data sekunder, yaitu data yang diperoleh dari sumber yang telah tersedia dan tidak dikumpulkan secara langsung oleh peneliti. Penggunaan data sekunder dipilih karena memungkinkan penelitian dilakukan dengan memanfaatkan dataset yang telah terdokumentasi dan dapat diakses secara terbuka untuk kepentingan analisis serta pengembangan model.

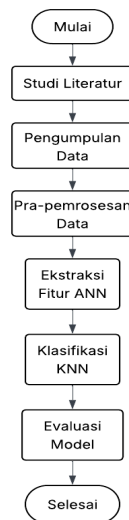
Dataset yang digunakan dalam penelitian ini adalah Breast Cancer Wisconsin Prognostic (BCWP) yang diperoleh dari UCI Machine Learning Repository. Dataset tersebut berisi sejumlah karakteristik atau fitur yang berkaitan dengan kondisi pasien kanker payudara dan digunakan sebagai dasar dalam proses pengolahan data, analisis, serta pengujian metode yang diterapkan dalam penelitian..

Dataset tersebut berisi informasi karakteristik sel tumor payudara dalam bentuk atribut numerik, seperti ukuran, bentuk, tekstur, serta karakteristik lainnya. Data ini digunakan untuk memprediksi prognosis kanker payudara, yaitu kemungkinan terjadinya kekambuhan (recurrence) atau tidak (non-recurrence).

Flowchart Penelitian

Flowchart penelitian menggambarkan alur proses yang dilakukan dalam penelitian prediksi prognosis kanker payudara menggunakan metode hybrid ANN dan KNN. Alur penelitian dimulai dari pengumpulan data hingga evaluasi model

yang dihasilkan. Flowchart penelitian dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Penelitian

Penjelasan tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Studi Literatur

Pada tahap ini dilakukan pengumpulan referensi berupa jurnal, buku, dan penelitian terdahulu yang berkaitan dengan kanker payudara, machine learning, serta metode ANN dan KNN. Tahap ini bertujuan untuk memahami konsep dasar serta menentukan metode yang digunakan dalam penelitian.

2. Pengumpulan Data

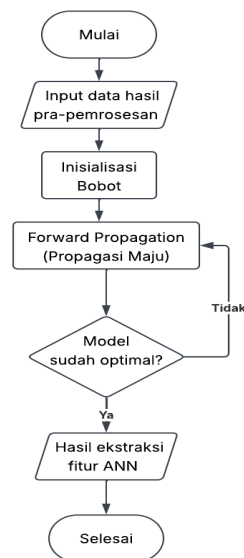
Pada tahap ini dilakukan pengambilan dataset Breast Cancer Wisconsin Prognostic dari UCI Machine Learning Repository. Dataset ini terdiri dari 198 data dengan 35 atribut.

3. Pra-pemrosesan Data

Tahap ini dilakukan untuk menyiapkan data sebelum digunakan dalam proses permodelan. Proses yang dilakukan meliputi pengecekan data, penanganan nilai yang tidak valid dengan mengganti menggunakan median, serta penghapusan atribut yang tidak diperlukan seperti ID. Selanjutnya, label kelas diubah dari bentuk kategorikal menjadi numerik, yaitu recurrence menjadi 1 dan non-recurrence menjadi 0. Data kemudian dinormalisasi menggunakan Min-Max Scaler agar memiliki skala yang sama.

4. Ekstraksi Fitur Menggunakan ANN

Pada tahap ini dilakukan pelatihan ANN yang digunakan sebagai metode ekstraksi fitur. ANN yang digunakan terdiri dari satu hidden layer dengan 16 neuron. Proses dimulai dengan inisialisasi bobot dan bias, kemudian dilakukan forward propagation untuk menghasilkan output pada hidden layer. Fungsi aktivasi yang digunakan adalah ReLU. Hasil dari proses ini berupa fitur baru yang lebih ringkas dan digunakan sebagai input pada tahap selanjutnya. Tahapan ini dapat dilihat pada Gambar 2.



Gambar 2. Flowchart ANN

5. Klasifikasi Menggunakan KNN

Fitur hasil ekstraksi dari ANN kemudian digunakan sebagai input pada metode KNN. Pada tahap ini, dilakukan pengujian terhadap nilai K untuk memperoleh hasil yang optimal. Nilai K yang diuji adalah K=1, K=3, K=5, K=7. Proses klasifikasi dilakukan dengan menghitung jarak antara data uji dan data latih menggunakan Euclidean Distance. Berdasarkan jarak tersebut, beberapa nilai K dengan jarak terdekat dipilih sebagai tetangga. Kelas data uji ditentukan berdasarkan kelas mayoritas dari tetangga terdekat. Nilai K terbaik selanjutnya ditentukan melalui proses Stratified K-Fold Cross-Validation dan digunakan sebagai konfigurasi model akhir.

6. Stratified 5-Fold Cross-Validation

Untuk memperoleh hasil evaluasi yang lebih stabil, penelitian ini menambahkan pengujian

menggunakan Stratified 5-Fold Cross-Validation. Metode ini membagi dataset menjadi lima bagian dengan tetap mempertahankan proporsi kelas recurrence dan non-recurrence pada setiap fold. Penggunaan Stratified 5-Fold Cross-Validation dilakukan untuk melihat konsistensi kinerja model pada beberapa pembagian data dan mengurangi ketergantungan hasil terhadap satu pembagian data latih dan data uji.

HASIL DAN PEMBAHASAN

Tahapan ini menjelaskan mengenai sumber data yang digunakan, tahapan pra-pemrosesan data, ekstraksi fitur menggunakan ANN, klasifikasi menggunakan KNN, dan evaluasi model.

Sumber Data

Data yang digunakan dalam penelitian ini adalah dataset WPBC yang terdiri dari 198 data pasien dengan total 35 atribut, termasuk atribut target. Atribut target dalam dataset ini adalah Outcome, yang menunjukkan apakah pasien mengalami kekambuhan kanker (recurrence) atau tidak (non-recurrence). Data kanker payudara dapat dilihat pada Tabel 1.

Tabel 1. Data Kanker Payudara

No	ID	...	Lymph_node_status
1	119513	...	5
2	8423	...	2
6	843786	...	0
7	844359	...	?
...	...	...	...
198	947489	...	0

Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahap awal yang dilakukan untuk mempersiapkan data sebelum digunakan dalam proses permodelan. Tahapan pra-pemrosesan pada penelitian ini meliputi beberapa langkah sebagai berikut:

1. Cleaning

Pada tahap ini dilakukan pengecekan terhadap data untuk memastikan tidak terdapat nilai yang tidak valid seperti data kosong atau

tanda seperti “?”. Jika ditemukan nilai tidak valid, maka nilai tersebut diganti menggunakan nilai median dari atribut yang bersangkutan. Selain itu, dilakukan penghapusan atribut yang tidak digunakan dalam proses permodelan, seperti atribut ID karena tidak memiliki pengaruh terhadap hasil prediksi.

2. Pengubahan Label

Pada tahap ini dilakukan perubahan data target dari bentuk kategorikal menjadi numerik. Kelas recurrence (R) diubah menjadi nilai 1, sedangkan non-recurrence (N) diubah menjadi nilai 0.

3. Normalisasi Data

Tahap normalisasi dilakukan untuk menyamakan skala nilai antar fitur. Hal ini penting karena setiap fitur memiliki rentang nilai yang berbeda, sehingga tanpa adanya normalisasi dapat mempengaruhi kinerja model. Normalisasi fitur menggunakan Min-Max Scaler, dimana nilai setiap fitur diubah ke rentang 0 hingga 1.

Hasil dari tahap cleaning, pengubahan label, dan normalisasi data, dapat dilihat pada Tabel 2.

Tabel 2. Data Prapemrosesan

No	X ₁	X ₂	...	X ₃₂	X ₃₃
1	0,2419	0,6467	...	0,4792	0,1852
2	0,4839	0,6455	...	0,2708	0,0741
3	0,9274	0,7753	...	0,2188	0,0000
4	0,9839	0,3932	...	0,1667	0,0000
5	0,2097	0,7339	...	0,3229	0,0000
...	...	...	...	...	...
198	0,0403	0,0188	...	0,3229	0,0000

3. Ekstraksi Fitur Menggunakan ANN

Setelah tahap pra-pemrosesan, data selanjutnya diproses menggunakan ANN untuk melakukan ekstraksi fitur. Pada tahap ini, data yang telah dinormalisasi dimasukkan ke dalam jaringan satu hidden layer.

Jumlah neuron pada hidden layer perlu diperhatikan karena dapat memengaruhi kemampuan ANN dalam membentuk representasi

fitur. Tidak ada jumlah neuron yang dapat dianggap paling tepat untuk setiap dataset. (Gilang et al., 2025) menyatakan bahwa jumlah neuron pada hidden layer tidak memiliki aturan yang tetap karena menyesuaikan dengan kebutuhan dan parameter penelitian. Pada penelitian ini digunakan 16 neuron dengan jumlah fitur yang digunakan, yaitu 34 fitur dan jumlah data sebanyak 198 pasien. Jumlah neuron yang terlalu sedikit dapat membatasi kemampuan jaringan dalam mempelajari pola data, sedangkan jumlah neuron yang lebih banyak akan membuat arsitektur jaringan menjadi lebih kompleks.

Hasil dari proses ini berupa fitur baru yang merupakan representasi dari data awal. Fitur yang dihasilkan memiliki jumlah yang lebih sedikit dibandingkan dengan fitur awal, namun tetap mampu mempertahankan informasi penting dari data.

Berdasarkan hasil yang diperoleh, nilai keluaran dari hidden layer menunjukkan bahwa setiap data memiliki pola yang berbeda setelah melalui proses ekstraksi fitur. Hal ini menunjukkan bahwa ANN mampu menangkap karakteristik penting dari data yang sebelumnya tidak terlihat secara langsung.

4. Pengujian Beberapa Nilai K

Pengujian beberapa nilai K dilakukan untuk mengetahui pengaruh jumlah tetangga terhadap kinerja klasifikasi KNN sekaligus menentukan nilai K yang optimal. Nilai K yang diuji dalam penelitian ini adalah K=1, K=3, K=5, K=7. Pengujian setiap nilai K dilakukan pada lima fold. Setiap nilai K memperoleh lima hasil pengujian yang berasal dari pembagian data yang berbeda. Hasil pengujian kemudian digunakan untuk menghitung rata-rata dan standar deviasi dari masing-masing metrik evaluasi.

5. Stratified 5-Fold Cross-Validation

Untuk mengetahui kestabilan kinerja model dan menentukan nilai K yang optimal, penelitian ini melakukan pengujian tambahan menggunakan Stratified K-Fold Cross-Validation. Pengujian dilakukan dengan membagi 198 data menjadi lima fold dengan tetap mempertahankan proporsi kelas recurrence dan non-recurrence. Pada setiap iterasi, empat fold digunakan sebagai data latih dan satu

fold digunakan sebagai data uji. Proses tersebut dilakukan sebanyak lima kali sehingga setiap fold memperoleh kesempatan menjadi data uji. Pada setiap fold dilakukan pengujian terhadap empat nilai K, yaitu K=1, K=3, K=5, dan K=7. Dengan demikian, terdapat 20 kombinasi pengujian yang terdiri dari lima fold dan empat nilai K. Hasil pengujiannya dapat dilihat pada Tabel 5.

Tabel 3. Hasil 5-Fold Cross Validation

Fold	K	Akurasi (%)	Presisi (%)	Recall (%)	F1-score (%)
1	1	77,50	50	66,67	57,14
1	3	80	55,56	55,56	55,56
1	5	75	40	22,22	28,57
1	7	77,50	50	33,33	40
2	1	72,50	44,44	40	42,11
2	3	62,50	22,22	20	21,05
2	5	65	25	20	22,22
2	7	67,50	28,57	20	23,53
3	1	60	25	30	27,27
3	3	70	37,50	30	33,33
3	5	72,50	42,86	30	35,29
3	7	77,50	60	30	40
4	1	61,54	20	22,22	21,05
4	3	71,79	33,33	22,22	26,67
4	5	69,23	28,57	22,22	25
4	7	71,79	33,33	22,22	26,67
5	1	58,97	23,08	33,33	27,27
5	3	64,10	30,77	44,44	36,36
5	5	76,92	50	44,44	47,06
5	7	84,62	80	44,44	57,14

Berdasarkan hasil pengujian pada masing-masing fold, akurasi tertinggi diperoleh pada fold 5 dengan K=7 sebesar 84,62%. Hasil ini menunjukkan bahwa penggunaan K=7 pada fold 5 memberikan performa klasifikasi paling baik dibandingkan kombinasi fold dan nilai K lainnya.

6. Perbandingan ANN, KNN, dan hybrid ANN-KNN

Untuk melihat pengaruh penggunaan model hybrid, dilakukan pengujian dengan membandingkan kinerja ANN, KNN dan hybrid ANN-KNN. Pengujian pada ketiga metode menggunakan dataset dan tahapan evaluasi yang sama. Nilai K yang digunakan pada KNN ditentukan berdasarkan nilai K terbaik, yaitu K=7 dan dengan fold 5. Perbandingan akurasi antara

ANN, KNN, dan hybrid ANN-KNN dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan ANN, KNN, dan Hybrid ANN-KNN

Metode	Akurasi
ANN	71,79%
KNN	74,36%
Hybrid ANN-KNN	84,62%

Berdasarkan hasil perbandingan ANN, KNN, dan hybrid ANN-KNN menunjukkan bahwa model hybrid ANN-KNN memberikan performa terbaik dibandingkan ANN dan KNN. Hal tersebut menunjukkan bahwa ekstraksi fitur menggunakan ANN sebelum proses klasifikasi KNN dapat membantu meningkatkan kemampuan model dalam melakukan klasifikasi.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat diperoleh kesimpulan sebagai berikut:

1. Penelitian ini berhasil menerapkan model hybrid ANN-KNN untuk memprediksi prognosis kanker payudara menggunakan dataset Breast Cancer Wisconsin Prognostic yang terdiri dari 198 data dan 35 atribut.
2. Pada pengujian menggunakan Stratified 5-Fold Cross-Validation dengan variasi K=1, K=3, K=5, dan K=7 menunjukkan bahwa akurasi tertinggi model hybrid ANN-KNN adalah 84,62% yang diperoleh pada fold ke-5 dengan K=7.
3. Pada pengujian perbandingan kinerja metode, model hybrid ANN-KNN memperoleh akurasi lebih tinggi dibandingkan ANN dan KNN.

DAFTAR PUSTAKA

Ali, R. M., Nurdin, N., Khaidar, A., Azzanna, M., & Rusadi, A. (2025). Comparative analysis of Random Forest algorithms, artificial neural networks, and logistic regression in breast cancer prediction with machine learning approach. *Journal of Artificial Intelligence and Software Engineering*, 5(3), 1285-1295.
<https://doi.org/10.30811/jaise.v5i3.7028>
Ardiansyah, G. R., Musayyanah, M., Aqvirandy, W., Farady, M. D., Cahya, M. N., &

Hadiono, T. (2023). Deteksi kanker payudara menggunakan artificial neural network pada deep learning. *INFOTECH: Jurnal Informatika & Teknologi*, 4(2), 259-269.
Arifin, T., Agung, I. W. P., Wibowo, I. R., & Junianto, E. (2026). Optimasi K-NN dengan Bayes Search CV untuk Klasifikasi Kanker Payudara. *Jurnal Informatika Polinema*, 12(3), 553-560.
Cahyani, N., Irsyada, R., & Kartini, A. Y. (2025). Implementasi Machine Learning Model sebagai Sistem Prediksi Penyakit Breast Cancer. *Digital Transformation Technology*, 4(2), 1112-1120.
<https://doi.org/10.47709/digitech.v4i2.5209>
Fahrurrozi et al., 2025. (2025). *Fasttext Dan Support Vector Machine*. 14(1), 90-101.
Fajar, A. K., Mutaqin, M. Z., Mutoffar, M. M., & Setiyadi, D. (2024). Klasifikasi Kanker Payudara Menggunakan Algoritma Neural Network dan Random Forest. *Jurnal Manajemen Informatika dan Sistem Informasi*, 7(1), 74-80.
Gilang, A., Irianto, H., & Sudarmilah, E. (2025). Pengaruh Variasi Jumlah Neuron dalam Hidden Layer Algoritma Pelatihan Levenberg-Marquardt Jaringan Backpropagation: A Systematic Literature Review. *Jurnal Al-Azhar Indonesia Seri Sains dan Teknologi*, 10(2), 180-189.
Kurniati, A. B., Sidik, W. A., & Jajang. (2024). Model Artificial Neural Networks (ANN) untuk Prediksi COVID-19 di Indonesia. *JST (Jurnal Sains Dan Teknologi)*, 12(3), 833-844.
<https://doi.org/10.23887/jstundiksha.v12i3.53437>
Nafiiyah, N., Elsony, M. L., Harjoko, A., & Hidayanto, A. N. (2026). Pendekatan Deep Learning untuk Deteksi Kanker Payudara Menggunakan Concatenated ResNet50 dan ResNet152. *Jurnal Algoritma*, 23(1), 420-428.
<https://doi.org/10.33364/algoritma/v.23-1.3259>
Pakpahan, S., & Arianto, D. B. (2025). Perbandingan performa algoritma machine learning pada prediksi penyakit kanker payudara. *Journal of Information System and Computer*, 5(1), 33-38.
Putri, A. K., & Nur, D. I. (2023). Penggunaan bahasa python untuk analisis dan visualisasi data penduduk di Desa Sumberjo, Nganjuk. Karya: *Jurnal Pengabdian kepada Masyarakat*, 3(3), 206-217.

- Qomariyah, N. (2023). Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN). *ROUTERS: Jurnal Sistem dan Teknologi Informasi*, 1-8.
- Rihastuti, S., Rosyidi, A., & Handoko, H. (2025). Implementasi K-Nearest Neighbours Dengan Google Collab Untuk Klasifikasi Penyakit Kanker Payudara. *Jurnal Teknologi Informasi*, 10(2), 79–85.
<https://doi.org/10.52643/jti.v10i2.5223>
- Samarov, A. P. X., & Barotova, Z. (2025). A Robust Hybrid Model Based on ANN and KNN for Multi Class Network Attack Detection and Classification. *International Journal of Inventive Engineering and Sciences*, 12(9), 1–6.
<https://doi.org/10.35940/ijies.h1115.12090925>
- Ujianto, N. T., Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis. *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, 4(1), 33-43.
- Yordanus, M. I., & Wenny, F. S. (2024). Prediksi Diagnosis dan Prognosis Breast Cancer menggunakan Machine Learning. *Jurnal Strategi*, 6(2), 284–290.
- Zubairi, A., Hermanto, H., Santoso, H., Samad, A., & Homaidi, A. (2025). Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Status Kanker. *Justify: Jurnal Sistem Informasi Ibrahimy*, 3(2), 117-121.