

# EVALUASI KINERJA KLASIFIKASI KANKER PAYUDARA DENGAN ALGORITMA GENETIKA DAN XGBOOST DENGAN MANUAL TUNING

Tanwir<sup>1\*</sup>, Anthony Anggrawan<sup>2</sup>, and Khasnur Hidjah<sup>3</sup>

<sup>1</sup> Department of Master of Computer Science, Bumigora University

<sup>2</sup> Department of Master of Computer Science, Bumigora University

<sup>3</sup> Department of Master of Computer Science, Bumigora University

E-Mail:

<sup>1</sup> [tanwir@universitasbumigora.ac.id](mailto:tanwir@universitasbumigora.ac.id)

<sup>2</sup> [anthony.anggrawan@stmikbumigora.ac.id](mailto:anthony.anggrawan@stmikbumigora.ac.id)

<sup>3</sup> [khasnur72.h@universitasbumigora.ac.id](mailto:khasnur72.h@universitasbumigora.ac.id)

## ABSTRACT

Kanker payudara telah menjadi penyebab utama kanker di kalangan wanita di seluruh dunia dengan sekitar 23 juta kasus baru pada tahun 2020. Meskipun kejadian meningkat, angka kematian telah menurun sebesar 43% sejak 1989. Kanker payudara adalah penyakit kompleks yang melibatkan proliferasi sel abnormal dan dapat bermetastasis ke organ lain. Penelitian ini mengeksplorasi potensi integrasi Algoritma Genetika dan XGBoost dalam klasifikasi kanker payudara, fokus pada penerapan teknik manual tuning untuk meningkatkan akurasi dan generalisasi model. Dataset yang digunakan adalah data publik dari Kaggle yang mencakup 569 baris dan 31 kolom. Metode penelitian melibatkan penggunaan Algoritma Genetika untuk pemilihan fitur, diikuti oleh penerapan XGBoost untuk tugas klasifikasi. Hasil penelitian menunjukkan bahwa model XGBoost dengan manual tuning mencapai akurasi tertinggi sebesar 100% pada komposisi data 90:10, dengan performa yang sangat tinggi juga pada komposisi data 70:30 dan 80:20. Temuan ini mengindikasikan bahwa integrasi GA dan XGBoost dengan manual tuning dapat secara signifikan meningkatkan kinerja klasifikasi kanker payudara, menawarkan solusi yang efektif untuk mendukung keputusan medis dalam diagnosis kanker payudara.

## ARTICLE INFO

### Keywords:

Kanker payudara; algoritma genetika; XGBoost; manual tuning; klasifikasi

*Corresponding Author:*

Tanwir, [tanwir@universitasbumigora.ac.id](mailto:tanwir@universitasbumigora.ac.id)

**INTRODUCTION**

Kanker Payudara telah muncul sebagai penyebab utama kejadian kanker di kalangan wanita di seluruh dunia, melampaui kanker paru-paru dengan perkiraan 2,3 juta kasus baru pada tahun 2020, mewakili 11,7% dari semua kasus kanker dan menjadikannya penyebab utama kelima kematian akibat kanker di seluruh dunia dengan 685.000 kematian[1]. Menurut statistik, Kanker payudara tetap menjadi masalah kesehatan masyarakat yang signifikan, dengan statistik mengungkapkan tren bernuansa dalam kejadian, kematian, kelangsungan hidup, dan dampak dari berbagai faktor seperti ras, etnis, dan status sosial ekonomi. Tingkat kejadian kanker payudara telah menunjukkan sedikit peningkatan, terutama dikaitkan dengan stadium lokal dan penyakit reseptor hormon positif, dengan kenaikan tahunan 0,5% dilaporkan dalam beberapa tahun terakhir [1]. Terlepas dari peningkatan kejadian ini, angka kematian telah mengalami penurunan yang stabil sejak 1989, turun 43% hingga 2020, yang berarti 460.000 lebih sedikit kematian akibat kanker payudara [2].

Kanker Payudara adalah penyakit multifaset yang ditandai dengan proliferasi sel abnormal yang tidak terkontrol di dalam dada, sering berasal dari saluran susu dan berpotensi menyebar ke jaringan dan organ lain. Keganasan ini merupakan penyebab utama morbiditas dan mortalitas di kalangan wanita secara global, dengan variasi yang signifikan dalam kejadian dan hasil berdasarkan faktor genetik, lingkungan, dan gaya hidup [3]. Penyakit ini bermanifestasi dalam berbagai bentuk, termasuk tipe noninvasif dan invasif, dengan yang terakhir lebih agresif dan mampu bermetastasis ke organ yang jauh [4]. Sel-sel abnormal pada kanker payudara dapat diklasifikasikan menjadi jenis jinak dan ganas. Sel-sel jinak, meskipun abnormal, tidak menyebar ke bagian lain dari tubuh dan biasanya tidak menyebabkan kerusakan signifikan pada jaringan di sekitarnya. Sebaliknya, sel-sel ganas sangat invasif, mampu berkembang biak dengan cepat, dan dapat menyebabkan kerusakan luas pada jaringan dan organ yang berdekatan [5]. Heterogenitas kanker payudara, dipengaruhi oleh faktor-faktor seperti perubahan hormon, mutasi genetik (misalnya, BRCA1 dan BRCA2), dan paparan lingkungan, mempersulit diagnosis dan pengobatannya [6].

Aspirasi Jarum Halus (Fine Needle Aspiration) adalah alat diagnostik penting dalam manajemen kanker payudara, menawarkan metode yang cepat, hemat biaya, dan invasif minimal untuk mengevaluasi lesi payudara. Ini melibatkan penggunaan jarum tipis untuk mengekstrak bahan seluler dari massa payudara, yang kemudian diperiksa secara sitologis. FNA sangat berharga karena akurasi diagnostiknya yang tinggi, dengan penelitian menunjukkan tingkat sensitivitas hingga 98% dan nilai prediksi positif juga pada 98% [7]. Ini sangat berguna dalam penilaian pra operasi kelenjar getah bening aksila, di mana FNA yang dipandu USG dapat secara akurat menentukan keterlibatan kelenjar getah bening, dengan sensitivitas 83,3% dan spesifisitas 100% [7]. Machine learning, bagian dari kecerdasan buatan, telah menunjukkan harapan yang signifikan dalam mengatasi tantangan terkait kesehatan, terutama dalam deteksi dini dan klasifikasi kanker payudara. Metode klasifikasi payudara telah berkembang secara signifikan, memanfaatkan berbagai teknik pembelajaran mesin (ML) dan kecerdasan buatan (AI) dalam mengidentifikasi dan meningkatkan akurasi kanker payudara untuk melihat hasil analisis sel kanker ganas dan jinak [8]. Sampel gambar digital yang diperoleh melalui aspirasi jarum halus pada massa payudara, bersama dengan fitur inti sel yang dihitung, dapat berfungsi sebagai data pelatihan untuk model pembelajaran mesin yang digunakan untuk tujuan klasifikasi. Model ini memiliki kapasitas untuk membantu ahli patologi dalam membuat keputusan yang lebih efisien dan tepat, sehingga mengurangi potensi kesalahan manusia. Selain itu,

penggabungan teknologi pembelajaran mesin dalam industri layanan kesehatan diharapkan dapat memfasilitasi kemajuan baru, yang pada akhirnya meningkatkan kualitas layanan kesehatan yang diberikan[8].

Penelitian sebelumnya dilakukan oleh Danang Wijayanto 2024 dengan judul “Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara”. Pada penelitian ini dilakukan perbandingan performa antara metode XGBoost dan metode LightGBM sehingga didapatkan persentasi akurasi untuk XGBoost sebesar 97,03% dan LightGBM sebesar 95,59% [9]. Penelitian dilakukan oleh Yahya, dkk pada tahun 2023 dengan judul “Implementasi Data Mining Dalam Klasifikasi Diagnosa Kanker Payudara Menggunakan Algoritma Logistic Regression”. Pada penelitian ini menggunakan metode Logistic Regression sehingga didapatkan persentasi akurasi sebesar 98% [10]. Pada penelitian ini dilakukan oleh Amira J. Zaylaa, dkk pada tahun 2024 dengan judul “Advancing Breast Cancer Diagnosis through Breast Mass Images, Machine Learning, and Regression Models”. Pada penelitian ini dilakukan penggunaan teknologi seperti machine learning dan model regresi dapat meningkatkan akurasi dalam diagnosis kanker payudara melalui analisis gambar massa payudara. Penelitian ini menunjukkan bahwa penggunaan SVM, NB, k-NN, dan DT-C4.5 dalam klasifikasi gambar massa payudara dapat memberikan tingkat akurasi yang tinggi, dengan SVM menunjukkan hasil terbaik dengan akurasi 97.13% [11]. Penelitian dilakukan oleh Alaa Alassaf pada tahun 2024 dengan judul “Genetic Algorithms and Feature Selection for Improving the Classification Performance in Healthcare”. Pada penelitian ini menggunakan delapan model klasifikasi (Random Forest, Logistic, K-Neighbors, Linear SVM, Radial SVM, AdaBoost, Decision Tree, dan Gradient Boosting) untuk mengklasifikasikan kumpulan data yang mengandung empat jenis penyakit: kolon, prostat, payudara, dan sistem saraf pusat. Algoritma Genetika (GA) digunakan untuk lima generasi, dan hasil yang paling akurat dicapai dengan menggabungkan model dengan tingkat akurasi tertinggi dari berbagai klasifikasi. Hasil menunjukkan bahwa diagnosis kanker payudara memiliki akurasi tinggi sebesar 96.15% [12].

Berdasarkan tinjauan literatur komprehensif yang mencakup penelitian yang menggunakan berbagai metode klasifikasi seperti algoritma SVM, XGBoost, LightGBM, dan Regresi Logistik, diketahui bahwa penggunaan algoritma genetika dalam klasifikasi kanker payudara hanya mendapat perhatian yang terbatas. Penelitian ini bertujuan untuk mengeksplorasi potensi integrasi algoritma genetika dan XGBoost dalam klasifikasi kanker payudara, dengan fokus pada penerapan teknik manual tuning untuk meningkatkan akurasi dan generalisasi model. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang berharga bagi pengembangan sistem pendukung keputusan medis yang lebih efektif dan terpercaya dan menjanjikan peningkatan pemahaman kita tentang pemilihan fitur yang optimal dalam pengembangan model klasifikasi yang kuat. Oleh karena itu, penelitian ini berupaya untuk mengatasi kesenjangan pengetahuan yang ada mengenai penerapan pendekatan evolusioner, seperti algoritma genetika dan model regresi, dengan tujuan akhir untuk meningkatkan kinerja klasifikasi.

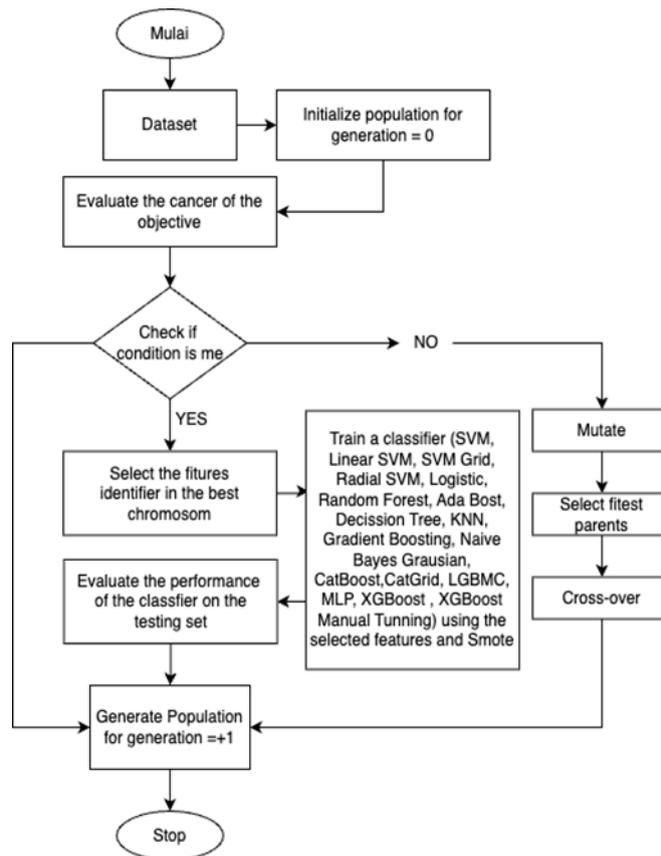
## **METHOD**

Penelitian ini menggunakan Algoritma Genetika untuk tujuan mengidentifikasi fitur-fitur dari dataset kanker. Metodologinya meliputi Seleksi fitur, Synthetic Minority Over-sampling Technique (SMOTE), pembuatan populasi, intersection, dan pemilihan fitur akhir. Dengan menghilangkan fitur yang tidak diperlukan dari kumpulan data, pendekatan ini meningkatkan

performa statistik model pembelajaran mesin. Genetika adalah teknik yang sangat efektif untuk pemilihan fitur karena menghasilkan kumpulan solusi potensial dan kemudian memilih solusi optimal berdasarkan metrik kinerja [13].

**Research Flowchart**

Gambar 1. Menggambarkan fase awal Algoritma Genetika, di mana sekumpulan kromosom acak dihasilkan. Setiap kromosom mewakili solusi potensial untuk masalah pemilihan fitur. Selanjutnya dilakukan evaluasi kanker payudara pada kromosom. Kromosom dengan tingkat kanker payudara tertinggi kemudian dipilih untuk direplikasi. Dengan menggunakan faktor persilangan dan mutasi, kromosom keturunannya terbentuk. Proses berlanjut seiring dengan evaluasi kromosom, dan siklus berulang hingga kriteria penghentian yang telah ditentukan sebelumnya terpenuhi. Kriteria ini dapat mencakup jumlah generasi, kesesuaian kromosom, atau kombinasi keduanya. genetika terbukti menjadi alat yang efektif untuk mengatasi masalah pemilihan fitur di berbagai domain.



**Gambar 1. Flowchat Genetika untuk memilih fitur terbaik.**

**Deskripsi Fungsi untuk Algoritma Genetika**

Split data berfungsi membagi kumpulan data menjadi kumpulan pelatihan dan pengujian yang berbeda, memanfaatkan kumpulan data dan rasio yang ditentukan. Skor Acc digunakan untuk mengevaluasi kinerja berbagai pengklasifikasi pada kumpulan data tertentu. Hal ini memerlukan

satu set pelatihan, satu set pengujian, dan daftar pengklasifikasi sebagai masukannya, dan menghasilkan skor akurasi untuk setiap pengklasifikasi pada set pengujian.

Plotting digunakan untuk memvisualisasikan hasil algoritma genetika. Hal ini memerlukan jumlah generasi dan skor kanker optimal untuk setiap generasi sebagai masukan dan kemudian menghasilkan grafik garis untuk mewakili nilai-nilai tersebut. Initialization-of-population merupakan prosedur mendasar dalam algoritma genetika karena perannya dalam menghasilkan populasi awal kromosom. Fungsi ini mengambil jumlah fitur dan ukuran populasi sebagai parameter masukan dan selanjutnya menghasilkan populasi kromosom melalui proses pengacakan. Cancer-score berfungsi mengevaluasi kanker kromosom dengan mempertimbangkan proses pengumpulan populasi, pelatihan, dan pengujian. Proses ini bertujuan untuk mengidentifikasi skor tertinggi dari penyakit kanker payudara. Selection bertugas untuk memilih fitur yang paling cocok untuk generasi penerus. Termasuk memberikan pelatihan dan pengujian, serta mencari fitur terbaik. Crossover berfungsi melakukan persilangan antara tetua yang berkinerja terbaik untuk menghasilkan keturunan. Orang tua terbaik membutuhkan kromosom untuk reproduksinya. Mutation Fungsi ini memperkenalkan variasi genetik baru ke dalam kromosom keturunannya. Hal ini membutuhkan kromosom keturunan dan laju mutasi untuk menyisipkan dan mereproduksi kromosom keturunan yang bermutasi. Generations Fungsi ini menjalankan semua fungsi yang disebutkan di atas untuk jumlah generasi yang ditentukan. Untuk mengambil kromosom terbaik (kumpulan atribut yang dipilih) dan skor kebugarannya, diperlukan populasi, set pelatihan, set pengujian, generasi, jumlah fitur, dan tingkat mutasi.

### **Tahap Implementasi**

Tahap implementasi membaca data dari file CSV dan membaginya menjadi dua kelompok, yaitu satu untuk pelatihan dan satu lagi untuk pengujian. Model klasifikasi dilatih menggunakan set pelatihan dan performanya dievaluasi menggunakan set pengujian. Kelas-kelas dalam kumpulan data dikodekan menjadi nilai numerik yang sesuai dengan kelas masing-masing. Selanjutnya, fungsi kanker payudara dikembangkan menggunakan algoritma XGBoost dengan manual tuning sebagai fitur yang dipilih dalam tugas klasifikasi. Berbagai algoritma pengklasifikasi, seperti SVM, Linear SVM, SVM Grid, Radial SVM, Logistic Regression, Random Forest, dan teknik lainnya, digunakan untuk memaksimalkan fungsi kanker payudara dan menentukan solusi optimal. Ukuran populasi awal kromosom, yang mewakili kumpulan fitur yang dipilih, ditentukan, dan kemudian kumpulan kromosom acak dibuat untuk memulai algoritma genetika. Skor kanker payudara setiap kromosom dalam populasi dihitung untuk menghasilkan skor dari masing-masing kanker payudara. Kromosom yang paling optimal dipilih untuk menghasilkan generasi baru. Generasi baru diciptakan melalui penerapan faktor persilangan dan mutasi, di mana mutasi melibatkan perubahan fitur tertentu dalam kromosom dan persilangan melibatkan pertukaran fitur antara dua kromosom. Fungsi kanker payudara diterapkan pada kromosom baru untuk menghitung skor kanker payudara mereka. Kromosom terkuat dari generasi sekarang, yang memiliki skor kanker payudara tertinggi, dipilih. Langkah-langkah ini diulang hingga kondisi terminasi terpenuhi, seperti jumlah iterasi yang telah ditentukan sebelumnya. Karakteristik kromosom yang optimal kemudian diidentifikasi sebagai kumpulan karakteristik yang paling sesuai untuk tugas klasifikasi. Model klasifikasi dilatih menggunakan kumpulan fitur yang dipilih, seperti pengklasifikasi SVM. Terakhir, kinerja model dinilai menggunakan set pengujian dengan penghitungan akurasi dan matriks kebingungan untuk mengevaluasi efektivitas model klasifikasi yang dilatih.

**Dataset**

Penelitian kali ini menggunakan data publik Kaggle.com (<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>). Dataset yang akan digunakan memiliki 569 baris dan 31 kolom. Data berasal dari perhitungan gambar digital massa payudara menggunakan teknik aspirasi needle fine.

**Tabel 1. Data Kanker Payudara**

No	Attribute	Type
1	diagnosis	String
2	radius_mean	Float
3	texture_mean	Float
4	perimeter_mean	Float
5	area_mean	Float
6	smoothness_mean	Float
7	compactness_mean	Float
8	concavity_mean	Float
9	concave points_mean	Float
10	symmetry_mean	Float
11	fractal_dimension_mean	Float
12	radius_se	Float
13	texture_se	Float
14	perimeter_se	Float
15	area_se	Float
16	smoothness_se	Float
17	compactness_se	Float
18	concavity_se	Float
19	concave points_se	Float
20	symmetry_se	Float
21	fractal_dimension_se	Float
22	radius_worst	Float
23	texture_worst	Float
24	perimeter_worst	Float
25	area_worst	Float
26	smoothness_worst	Float
27	compactness_worst	Float
28	concavity_worst	Float
29	concave points_worst	Float
30	symmetry_worst	Float
31	fractal_dimension_worst	Float

## RESULT AND DISCUSSION

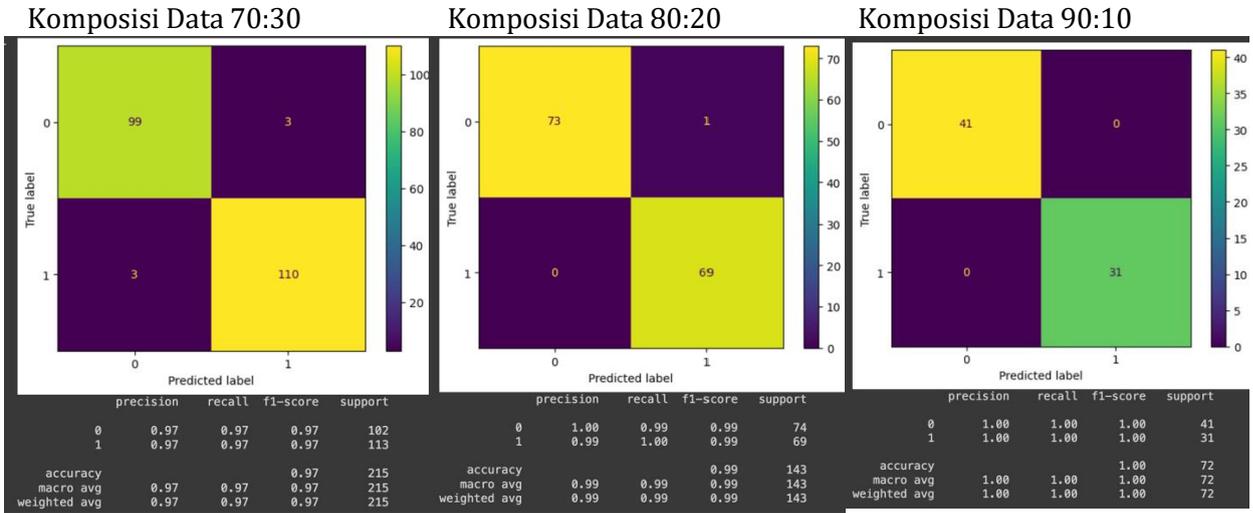
Bagian ini menyajikan dan membahas hasil percobaan yang dilakukan. Hal ini juga mencakup perbandingan antara hasil yang diperoleh dan temuan dari penelitian lain. Metode pemilihan fitur yang diusulkan, yang mengintegrasikan nilai diskrit dan kontinu, diuji menggunakan genetika untuk mengoptimalkan akurasi. Metode ini diterapkan pada beberapa klasifikasi dengan dataset kanker payudara.

Gambar 2. Menunjukkan konfigurasi model XGBoost Manual Tuning menggunakan parameter-parameter khusus yang dirancang untuk mengoptimalkan kinerja klasifikasi. Parameter `base_score` diatur ke 0.41, yang menentukan nilai awal prediksi. `colsample_bytree` sebesar 0.07 menunjukkan proporsi fitur yang akan digunakan pada setiap pohon. `gamma` diatur ke 0.1, memberikan batas minimum untuk split tambahan pada pohon. Tingkat pembelajaran (`learning_rate`) ditetapkan pada 0.3 untuk mengendalikan kecepatan pembaruan bobot. Kedalaman maksimum pohon (`max_depth`) dibatasi pada 3 untuk menghindari overfitting. `min_child_weight` sebesar 0.1 digunakan untuk menghindari overfitting dengan menetapkan berat minimum bagi setiap leaf node. Jumlah estimator (`n_estimators`) yang digunakan adalah 70, yang menentukan jumlah pohon yang akan dibangun. Terakhir, `subsample` sebesar 0.06 digunakan untuk sampling data secara acak sebelum membangun pohon guna mengurangi overfitting. Parameter ini menunjukkan pendekatan tuning manual yang cermat untuk meningkatkan performa model XGBoost dalam tugas klasifikasi.

```
XGBoostManualTunning = XGBClassifier(  
    base_score=0.41,  
    colsample_bytree=0.07,  
    gamma=0.1,  
    learning_rate=0.3,  
    max_depth=3,  
    min_child_weight=0.1,  
    n_estimators=70,  
    subsample=0.06  
)
```

**Gambar 2. Algoritma XGBoost Manual Tuning**

Gambar 3. Menunjukkan hasil akurasi Algoritma XGBoost dengan manual tuning menggunakan tiga komposisi data yang berbeda: 70:30, 80:20, dan 90:10. Setiap komposisi data ditampilkan dalam bentuk confusion matrix yang menunjukkan performa klasifikasi dari model. Pada komposisi data 70:30, model mencapai akurasi, precision, recall, dan f1-score masing-masing sebesar 0.97. Pada komposisi data 80:20, model menunjukkan peningkatan performa dengan nilai akurasi, precision, recall, dan f1-score sebesar 0.99. Terakhir, pada komposisi data 90:10, model menunjukkan performa terbaik dengan semua metrik evaluasi mencapai nilai sempurna 1.00. Hasil ini mengindikasikan bahwa peningkatan proporsi data latih terhadap data uji dapat meningkatkan performa model XGBoost secara signifikan.



Gambar 3. Akurasi Algoritma XGBoost Manual Tuning dengan komposisi berbeda

Tabel 2. Menyajikan hasil akurasi berbagai algoritma klasifikasi ketika dinilai menggunakan tiga komposisi data berbeda: 70:30, 80:20, dan 90:10. Kolom pertama berisi nama algoritma klasifikasi yang digunakan dalam percobaan. Kolom kedua hingga keempat menampilkan nilai akurasi yang dicapai setiap algoritma untuk setiap pembagian data. Hasil yang ditampilkan menawarkan wawasan tentang kemampuan setiap algoritma untuk mengklasifikasikan data secara akurat berdasarkan rasio pembagian data masing-masing. Khususnya, algoritma XGBoost Manual Tuning menunjukkan akurasi tertinggi dengan nilai sempurna (1.000) pada pembagian data 90:10, dan sangat tinggi pada pembagian data 70:30 dan 80:20, Algoritma lain seperti MLPClassifier, Logistic, LinearSVM, dan AdaBoost juga menunjukkan akurasi sempurna pada pembagian data 90:10. Radial SVM dan SVM menunjukkan performa luar biasa pada pemisahan data 70:30 dan 80:20, namun menunjukkan sedikit penurunan ketika dilakukan pemisahan data 90:10. Dibandingkan dengan algoritma lain, Naïve Bayes Gaussian menunjukkan kinerja yang relatif lebih rendah, dengan nilai akurasi terendah sebesar 0,930556 yang diamati dalam konteks pembagian data 90:10.

Tabel 2. Akurasi Penerapan klasifikasi Original

Classifier	Komposisi Data Akurasi		
	70:30	80:20	90:10
<b>XGBoost Manual Tuning</b>	<b>0.993007</b>	<b>0.993006993</b>	<b>1.000.000</b>
<b>XGBoost</b>	0.965035	0,993006993	1.000.000
<b>Radial SVM</b>	0.986014	0,986013986	0.972222
<b>MLP Classifier</b>	0.979021	0,986013986	1.000.000
<b>SVM Grid</b>	0.986014	0,986013986	0.986111

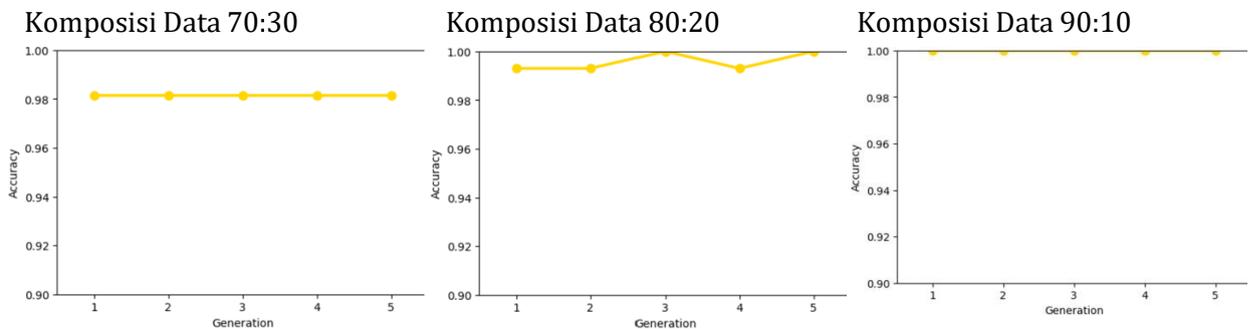
<b>SVM</b>	0.986014	0,9860139 86	0.972222
<b>LGBM Classifier</b>	0.965035	0,9790209 79	0.986111
<b>Linear SVM</b>	0.979021	0,9790209 79	1.000.000
<b>Logistic</b>	0.979021	0,9790209 79	1.000.000
<b>CatBoost Classifier</b>	0.979021	0,9720279 72	0.986111
<b>CatGrid</b>	0.979021	0,9720279 72	0.986111
<b>Gradient Boosting</b>	0.965035	0,9650349 65	0.986111
<b>KNN</b>	0.965035	0,9650349 65	0.972222
<b>AdaBoost</b>	0.972028	0,9650349 65	1.000.000
<b>Random Forest</b>	0.958042	0,9580419 58	0.986111
<b>Decision Tree</b>	0.951049	0,9440559 44	0.972222
<b>Naïve Bayes Grausian</b>	0.944056	0,9370629 37	0.930556

Tabel 3. Menunjukkan akurasi terbaik dari model prediksi cuaca berbasis genetika dengan berbagai komposisi data pelatihan dan pengujian, yakni 70:30, 80:20, dan 90:10. Pada komposisi data 70:30, model mencapai akurasi 0.98 secara konsisten pada lima generasi berturut-turut. Untuk komposisi data 80:20, model menunjukkan peningkatan akurasi, dengan nilai sempurna (1.00) pada generasi ketiga dan akurasi 0.99 pada generasi lainnya. Komposisi data 90:10 menghasilkan akurasi sempurna (1.00) pada semua generasi. Hasil ini mengindikasikan bahwa peningkatan proporsi data pelatihan secara signifikan meningkatkan akurasi model, dengan hasil optimal dicapai pada komposisi data 90:10. Akurasi sempurna pada komposisi 90:10 menunjukkan kemampuan model untuk mempelajari pola dari data pelatihan secara efektif, sehingga mampu melakukan prediksi dengan tingkat akurasi yang sangat tinggi pada data pengujian.

**Tabel 3. Skor terbaik dalam genetika**

Komposisi Data Akurasi	Generation				
	1	2	3	4	5
<b>70:30</b>	0.98	0.98	0.98	0.98	0.98
<b>80:20</b>	0.99	0.99	1.00	0.99	0.99
<b>90:10</b>	1.0	1.0	1.0	1.0	1.0

Gambar 4. Menunjukkan grafik akurasi model machine learning pada berbagai komposisi data pelatihan dan pengujian, yaitu 70:30, 80:20, dan 90:10. Pada grafik pertama (70:30), akurasi model terlihat stabil di angka sekitar 0.98 selama lima generasi. Grafik kedua (80:20) juga menunjukkan stabilitas akurasi di sekitar 0.98 hingga 1.00 dengan sedikit fluktuasi. Pada grafik ketiga (90:10), akurasi model mencapai nilai hampir sempurna, yaitu 1.00, yang menunjukkan konsistensi tinggi. Secara keseluruhan, model menunjukkan performa yang sangat baik pada ketiga komposisi data, dengan peningkatan proporsi data pelatihan yang mengarah pada peningkatan akurasi model. Hasil ini menunjukkan bahwa penggunaan data pelatihan yang lebih banyak dapat meningkatkan kinerja model, yang merupakan temuan penting dalam konteks pengembangan dan evaluasi model machine learning.



**Gambar 4. Skor terbaik dalam genetika**

Tabel 4. Menunjukkan perbandingan akurasi antara metode original dan pendekatan yang diusulkan berdasarkan berbagai komposisi data pelatihan dan pengujian. Dengan rasio data pelatihan terhadap data pengujian 70:30, metode original mencapai akurasi sebesar 0,99, sedangkan pendekatan yang diusulkan mencapai akurasi 0,98. Pada komposisi data 80:20, kedua metode menunjukkan akurasi yang sama yaitu 0,99 untuk metode original dan 1,00 untuk pendekatan yang diusulkan. Sementara itu, pada komposisi data 90:10, kedua metode menunjukkan akurasi sempurna yaitu 1,00. Hasil ini mengindikasikan bahwa pendekatan yang diusulkan mampu memberikan akurasi yang setara atau bahkan lebih baik dibandingkan dengan metode original, terutama pada komposisi data pelatihan yang lebih tinggi.

**Tabel 4. Perbandingan Akurasi metode original dengan metode yang diusulkan**

Komposisi data akurasi	Akurasi menggunakan Metode Original	Akurasi menggunakan pendekatan diusulkan
<b>70:30</b>	0.99	0.98
<b>80:20</b>	0.99	1.00
<b>90:10</b>	1.00	1.00

**CONCLUSION**

Penelitian ini berfokus pada pengembangan metode klasifikasi kanker payudara dengan mengintegrasikan algoritma genetika dan XGBoost dengan manual tuning. Temuan ini menunjukkan peningkatan signifikan dalam akurasi diagnosis kanker payudara. Secara khusus, ketika komposisi data ditetapkan pada rasio 90:10, algoritma XGBoost dengan penyetelan manual mencapai akurasi

tertinggi hingga 100%. Hal ini menunjukkan efektivitasnya dalam deteksi dan klasifikasi kanker payudara. Selain itu, Algoritma Genetika telah terbukti efektif dalam pemilihan fitur. Mereka dapat menghilangkan fitur-fitur yang tidak relevan sambil mempertahankan fitur-fitur penting, sehingga meningkatkan performa model pembelajaran mesin. Peningkatan ini terlihat dari pencapaian akurasi klasifikasi yang lebih baik. Jika dibandingkan dengan metode lain seperti MLP Classifier, Logistic Regression, Linear SVM, dan AdaBoost, algoritma XGBoost dengan manual tuning secara konsisten menunjukkan performa yang unggul. Hal ini menegaskan keunggulan metode yang digunakan dalam penelitian ini. Penerapan pembelajaran mesin, khususnya XGBoost, dalam diagnosis kanker payudara menunjukkan potensi signifikan untuk meningkatkan akurasi dan efisiensi diagnosis, sehingga membantu ahli patologi dalam mengambil keputusan yang lebih tepat dan efisien. Studi ini berkontribusi pada kemajuan sistem pendukung keputusan medis yang lebih efektif dan andal. Selain itu, integrasi Algoritma Genetika dan XGBoost tidak hanya meningkatkan akurasi diagnostik tetapi juga menjanjikan dalam mengurangi kesalahan manusia dalam proses diagnostik. Penelitian ini menunjukkan bahwa penggunaan teknik pembelajaran mesin, khususnya Algoritma Genetika dan XGBoost dengan manual tuning, dapat meningkatkan performa dan akurasi klasifikasi kanker payudara secara signifikan. Hal ini berdampak positif pada kualitas layanan kesehatan.

#### **ACKNOWLEDGEMENT**

Kami ingin menyampaikan rasa terima kasih yang tulus kepada berbagai pihak yang telah memberikan dukungan dan bantuan selama penelitian ini. Pertama-tama, kami berterima kasih kepada Universitas Bumigora Mataram yang telah memberikan fasilitas dan lingkungan akademis yang kondusif untuk penelitian ini. Kami juga ingin mengucapkan terima kasih kepada rekan-rekan di Fakultas Teknik, Program Magister Ilmu Komputer, yang telah berbagi pengetahuan dan wawasan yang berharga. Selain itu, terima kasih kami sampaikan kepada semua pihak yang telah membantu dalam pengumpulan dan pengolahan data, termasuk tim di Kaggle yang menyediakan dataset yang sangat berguna. Terakhir, kami berterima kasih kepada keluarga dan teman-teman kami atas dukungan moral dan semangat yang mereka berikan selama proses penelitian ini.

#### **REFERENCE**

- A. N. Giaquinto et al., "Breast Cancer Statistics, 2022," *CA Cancer J Clin*, vol. 72, no. 6, pp. 524–541, Nov. 2022, doi: 10.3322/caac.21754.
- C. E. DeSantis et al., "Breast cancer statistics, 2019," *CA Cancer J Clin*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.
- Q. Huang, A. Khalil, D. A. Ali, A. Ahmad, R. Luo, and M. Azeem, "Breast cancer chemical structures and their partition resolvability," *Mathematical Biosciences and Engineering*, vol. 20, no. 2, pp. 3838–3853, 2023, doi: 10.3934/mbe.2023180.
- P. D. R. Mehboob, "Breast Cancer- Awareness and Early Detection," *Pakistan BioMedical Journal*, pp. 01–01, Oct. 2022, doi: 10.54393/pbmj.v5i10.814.
- G. A. Palmer, "Breast cancer: diagnosis and treatment.," *Nurse Pract Forum*, vol. 4, no. 2, pp. 100–104, Jun. 1993, doi: 10.47363/jmcn/2022(3)156.

- Y. Feng et al., "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis," *Genes and Diseases*, vol. 5, no. 2. Chongqing University, pp. 77–106, Jun. 01, 2018. doi: 10.1016/j.gendis.2018.05.001.
- R. Shafique et al., "Breast Cancer Prediction Using Fine Needle Aspiration Features and Upsampling with Supervised Machine Learning," *Cancers (Basel)*, vol. 15, no. 3, Feb. 2023, doi: 10.3390/cancers15030681.
- V. Kumari and R. Ghosh, "A magnification-independent method for breast cancer classification using transfer learning," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100207.
- J. Khatib Sulaiman, D. Wijayanto, B. Pulu Hartato, and U. Amikom Yogyakarta, "Analisis Perbandingan Performa Algoritma XGBoost dan LightGBM pada Klasifikasi Kanker Payudara," *Indonesian Journal of Computer Science*.
- Y. Anugerah, D. Khurrota A'yunan, S. Busono, F. Sains, and D. Teknologi, "Implementasi Data Mining Dalam Klasifikasi Diagnosa Kanker Payudara Menggunakan Algoritma Logistic Regression," *Jurnal TEKINKOM*, vol. 6, no. 2, p. 2023, doi: 10.37600/tekinkom.v6i2.948.
- A. J. Zaylaa and S. Kourtian, "Advancing Breast Cancer Diagnosis through Breast Mass Images, Machine Learning, and Regression Models," *Sensors*, vol. 24, no. 7, Apr. 2024, doi: 10.3390/s24072312.
- A. Alassaf et al., "Genetic Algorithms and Feature Selection for Improving the Classification Performance in Healthcare," 2024. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- A. Hichri et al., "Genetic-Algorithm-Based Neural Network for Fault Detection and Diagnosis: Application to Grid-Connected Photovoltaic Systems," *Sustainability (Switzerland)*, vol. 14, no. 17, Sep. 2022, doi: 10.3390/su141710518.