

ANALISIS MULTI-MODEL KOMPARATIF UNTUK DETEKSI DINI KANKER PAYUDARA MENGGUNAKAN EVALUASI ROC-AUC DAN MCC

Daning Nur Sulistyowati^{1*}; Sri Hadiani²; Ridan Nurfalalah³

Sistem Informasi^{1,3}, Informatika²
Universitas Nusa Mandiri, Jakarta, Indonesia^{1,2,3}
<https://www.nusamandiri.ac.id/1,2,3>

daningnur.dgs@nusamandiri.ac.id^{1*}, sri.shv@nusamandiri.ac.id², ridan.rlh@nusamandiri.ac.id³

(*) Corresponding Author



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract—Breast cancer is one of the leading causes of death in women and relies heavily on early detection to improve the chances of recovery. The main challenges in machine learning-based early detection systems are class imbalance and the limitations of evaluation metrics that often rely solely on accuracy or ROC-AUC. In this context, the Matthews Correlation Coefficient (MCC) offers a more comprehensive assessment because it considers all elements of the confusion matrix. This study analyzes and compares the performance of four classification algorithms: Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM) on the Breast Cancer Wisconsin (Diagnostic) dataset. The dataset was first divided using a stratified train-test split of 80:20 to maintain class proportions. Feature normalization was performed only on the training data to avoid data leakage, then applied to the test data using the same parameters. Furthermore, 5-fold cross-validation was performed on the training data for model evaluation and selection. The results show that SVM provides the best performance with an accuracy of 98.25%, a precision of 1.00, an F1-score of 0.9762, an AUC of 0.9971, and an MCC of 0.9630. Naive Bayes and Random Forest also show excellent performance with AUC values above 0.99 and an MCC of 0.9253, while Decision Tree has a lower performance. Confusion matrix and ROC curve analysis confirm the superiority of SVM in minimizing classification errors. These findings emphasize the importance of a multi-model approach and the use of MCC as a more representative evaluation metric in breast cancer early detection systems.

Keywords: Breast Cancer, Classification, Machine Learning, MCC, ROC-AUC.

Abstrak—Kanker payudara merupakan salah satu penyebab kematian tertinggi pada perempuan dan sangat bergantung pada proses deteksi dini untuk meningkatkan peluang kesembuhan. Tantangan utama dalam sistem deteksi dini berbasis machine learning adalah ketidakseimbangan kelas dan keterbatasan metrik evaluasi yang sering hanya bergantung pada akurasi atau ROC-AUC. Dalam konteks tersebut, Matthews Correlation Coefficient (MCC) menawarkan penilaian yang lebih komprehensif karena mempertimbangkan seluruh elemen confusion matrix. Penelitian ini menganalisis dan membandingkan performa empat algoritma klasifikasi Naive Bayes, Decision Tree, Random Forest, dan Support Vector Machine (SVM) pada dataset Breast Cancer Wisconsin (Diagnostic). Dataset terlebih dahulu dibagi menggunakan stratified train-test split sebesar 80:20 untuk menjaga proporsi kelas. Normalisasi fitur dilakukan hanya pada data latih untuk menghindari data leakage, kemudian diterapkan pada data uji menggunakan parameter yang sama. Selanjutnya dilakukan 5-fold cross-validation pada data latih untuk proses evaluasi dan pemilihan model. Hasil menunjukkan bahwa SVM memberikan performa paling unggul dengan akurasi 98,25%, presisi 1,00, F1-score 0,9762, AUC 0,9971, dan MCC 0,9630. Naive Bayes dan Random Forest juga menunjukkan performa yang sangat baik dengan nilai AUC di atas 0,99 dan MCC 0,9253, sementara Decision Tree memiliki performa lebih rendah. Analisis confusion matrix dan kurva ROC mengonfirmasi keunggulan SVM dalam meminimalkan kesalahan klasifikasi. Temuan ini menegaskan pentingnya pendekatan multi-model serta pemanfaatan MCC sebagai metrik evaluasi yang lebih representatif dalam sistem deteksi dini kanker payudara.

Kata kunci: Kanker Payudara, Klasifikasi, Pembelajaran Mesin, Matthews Correlation Coefficient (MCC), Receiver Operating Characteristic–Area Under the Curve (ROC-AUC).

PENDAHULUAN

Kanker payudara merupakan salah satu penyakit dengan prevalensi tertinggi dan menjadi penyebab kematian utama pada perempuan di seluruh dunia. Organisasi Kesehatan Dunia (WHO) mencatat bahwa kasus kanker payudara meningkat setiap tahunnya seiring dengan perubahan gaya hidup, faktor genetika, serta pola deteksi dini yang belum optimal. (Joy et al., 2023), Kondisi tersebut juga dialami di Indonesia, di mana kanker payudara menduduki urutan pertama dalam jenis kanker yang paling banyak didiagnosis pada perempuan. Upaya deteksi dini menjadi sangat penting karena diagnosis pada stadium awal terbukti dapat menurunkan angka kematian secara signifikan dan meningkatkan peluang keberhasilan terapi.

Seiring perkembangan teknologi berbagai penelitian telah mengeksplorasi beragam algoritma klasifikasi untuk mendukung sistem deteksi dini kanker payudara melalui pemanfaatan metrik evaluasi seperti ROC-AUC dan *Matthews Correlation Coefficient* (MCC). Misalnya beberapa studi melaporkan performa tinggi menggunakan metode SVM, *Random Forest*, *Naive Bayes*, dan *Decision Tree* dengan ROC-AUC > 0,96 serta keseimbangan antara presisi dan sensitivitas yang baik (Singhal et al., 2022).

Studi lain juga menunjukkan bahwa *ensemble learning* dengan optimasi fitur mampu mencapai akurasi hampir 99% dengan ROC-AUC mendekati 1,00 (Sahoo et al., 2024). Selain itu, SVC-RBF terbukti memiliki AUC sekitar 0,96 dan MCC yang sangat kuat dalam membedakan lesi jinak dan ganas (Das, Koley and Saha, 2023). Walaupun demikian, sebagian besar penelitian sebelumnya lebih menitikberatkan pada ROC-AUC tanpa memberikan perhatian yang memadai terhadap MCC. Metrik MCC dinilai lebih informatif karena mempertimbangkan seluruh elemen confusion matrix dan mampu memberikan evaluasi yang lebih seimbang, terutama ketika distribusi kelas tidak sepenuhnya seimbang (Zhang et al., 2025), (Chicco and JURMAN, 2023), (HANIS, ISLAM and MUSA, 2022).

Berdasarkan hal tersebut penelitian ini bertujuan untuk menganalisis dan membandingkan performa beberapa model klasifikasi, yaitu *Naive Bayes*, *Decision Tree*, *Random Forest* dan SVM dalam deteksi dini kanker payudara menggunakan metrik ROC-AUC dan MCC untuk memperoleh gambaran kinerja model yang lebih komprehensif serta representatif. Manfaat penelitian ini diharapkan

mampu memberikan referensi ilmiah dalam pemilihan algoritma klasifikasi yang tepat untuk

mendukung sistem prediksi kanker payudara serta memperkuat kontribusi akademik mengenai pentingnya MCC dalam konteks ketidakseimbangan kelas. Adapun hipotesis yang diajukan adalah bahwa model *Random Forest* dan SVM memiliki performa yang lebih unggul dibandingkan *Naive Bayes* dan *Decision Tree* dalam mendeteksi kanker payudara berdasarkan nilai ROC-AUC dan MCC (SINGHAL et al., 2022), (SAHOO et al., 2024), (DAS, KOLEY and SAHA, 2023).

Metode penelitian mengambil pendekatan analitis kuantitatif eksperimental yang sistematis. *Preprocessing* data dimulai dengan standardisasi fitur dan *encoding* untuk menjamin konsistensi *input* ke semua model. Empat algoritma klasifikasi utama *Naive Bayes*, *Decision Tree*, *Random Forest*, dan SVM dengan kernel RBF dilatih pada data yang telah disiapkan. Evaluasi dilakukan melalui *cross-validation* (minimal 5-fold), dengan ROC-AUC dan MCC sebagai metrik kinerja utama untuk mengukur kemampuan diskriminatif dan korelasi prediksi secara eksplisit. Di samping itu perbedaan akurasi antara data latih dan uji dianalisis untuk mendeteksi kemungkinan *overfitting*. Bentuk metodologis ini mengadaptasi praktik terbaik dari studi-studi serupa dalam mengevaluasi performa klasifikator kanker payudara (Joy Et Al., 2023), (Sahoo Et Al., 2024), (Das, Koley And Saha, 2023), (Pati Et Al., 2024), (Kumar, Saini And Kumar, 2024).

Meskipun berbagai penelitian sebelumnya telah membandingkan performa algoritma *machine learning* untuk deteksi kanker payudara, sebagian besar studi masih berfokus pada metrik akurasi dan ROC-AUC sebagai indikator utama evaluasi model. Pendekatan tersebut berpotensi menghasilkan interpretasi yang terlalu optimistis terutama pada dataset dengan distribusi kelas yang tidak sepenuhnya seimbang. Selain itu masih sedikit penelitian yang secara eksplisit mengevaluasi kemampuan model menggunakan *Matthews Correlation Coefficient* (MCC) sebagai metrik utama membanding antar model klasifikasi. Oleh karena itu nilai tambah penelitian ini terletak pada analisis komparatif multi-model dengan penekanan khusus pada penggunaan MCC untuk memberikan evaluasi klasifikasi yang lebih representatif dan robust dalam konteks data medis.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk melakukan analisis multi-model secara komprehensif dengan membandingkan empat algoritma klasifikasi Naive

Bayes, Decision Tree, Random Forest, dan SVM untuk deteksi dini kanker payudara menggunakan dataset Breast Cancer Wisconsin (Diagnostic). Fokus utama penelitian diarahkan pada evaluasi menggunakan metrik ROC-AUC dan MCC untuk memberikan gambaran yang lebih lengkap mengenai kemampuan diskriminatif dan kualitas prediksi masing-masing model. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam memperkuat landasan ilmiah pemilihan algoritma yang paling efektif serta menegaskan pentingnya MCC sebagai metrik yang

lebih informatif dalam kasus ketidakseimbangan kelas pada data medis.

Literatur Review

Tinjauan literatur dilakukan terhadap penelitian-penelitian yang mengkaji deteksi dini kanker payudara untuk memperkuat dasar teoritis serta menegaskan *urgensi* dari penelitian ini. Fokus utama pada model komparatif serta pemanfaatan metrik evaluasi ROC-AUC dan *Matthews Correlation Coefficient* (MCC) dalam mengidentifikasi perkembangan metodologis serta kesenjangan riset yang menjadi kontribusi dari penelitian ini.

Tabel 1 *Literatur Review*

No	Judul	Deskripsi Singkat	Perbedaan Penelitian
1	<i>Predicting breast cancer types on and beyond molecular subtypes using multimodal deep learning</i> (Zhang et al., 2023).	Model <i>deep learning</i> multimodal (MDL-IIA) menggabungkan mamografi dan ultrasonografi; MCC 0,837, AUC 0,929.	Menekankan MCC dan teknik <i>attention</i> untuk klasifikasi subtype molekuler.
2	<i>Advancements in Breast Cancer Detection: A Review of Global Trends, Risk Factors, Imaging Modalities, Machine Learning, and Deep Learning Approaches</i> (Rahman et al., 2025).	<i>Review</i> global metode <i>machine learning</i> dan <i>deep learning</i> untuk deteksi kanker payudara.	Memberikan gambaran komprehensif berbagai pendekatan dan tantangan deteksi kanker.
3	<i>Breast cancer risk prediction using machine learning: a systematic review</i> (Hussain et al., 2024).	Mengembangkan model prediksi risiko menggunakan data klinis dan citra medis.	Integrasi data klinis dan citra medis untuk meningkatkan akurasi.
4	<i>The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification</i> (Chicco and Jurman, 2023).	Usulan penggunaan MCC sebagai metrik standar dalam evaluasi klasifikasi biner.	Menyoroti pentingnya pemilihan metrik evaluasi, bukan pengembangan model.
5	<i>A Comparison of Techniques for Class Imbalance in Deep Learning for Breast Cancer Classification</i> (Walsh and Tardy, 2023).	Membahas teknik menangani ketidakseimbangan kelas termasuk <i>oversampling/undersampling</i> .	Fokus pada penyeimbangan kelas bukan pengembangan algoritma baru.
6	<i>A systematic review of machine learning algorithms for breast cancer detection</i> (Boddu and Jan, 2025).	<i>Review</i> berbagai algoritma ML (SVM, <i>Naive Bayes</i> , <i>XGBoost</i>) dan performanya.	Memberikan perbandingan langsung antara algoritma berbeda.
7	<i>Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images</i> (Darweesh et al., 2021).	Sistem klasifikasi hierarkis; MCC 0,76, F1-score 0,98.	Fokus pada klasifikasi hierarkis bukan klasifikasi Tunggal.
8	<i>Deep Learning Algorithms for Breast Cancer Detection in a UK Screening Cohort: As Stand-alone Readers and Combined with Human Readers</i> (Hickman et al., 2024).	Evaluasi <i>deep learning</i> pada <i>kohort skrining</i> akurasi lebih tinggi daripada radiolog Tunggal.	Fokus pada setting klinis <i>skrining</i> bukan eksperimental
9	<i>The BCPM method: decoding breast cancer with machine learning</i> (Almarri et al., 2024).	Menggabungkan logika fuzzy dan ML untuk akurasi deteksi tinggi.	Menekankan integrasi logika fuzzy dalam ML.
10	<i>Tissue Classification of Breast Cancer by Hyperspectral Imaging</i> (Jong et al., 2023).	Klasifikasi jaringan menggunakan citra hiperspektral; MCC 0,71.	Fokus pada citra hiperspektral bukan mamografi/ultrasound.

Sumber: (Hasil Penelitian, 2026)

Berdasarkan Tabel 1 menunjukkan bahwa meskipun banyak penelitian yang fokus pada pengembangan model klasifikasi untuk deteksi dini kanker payudara masih terdapat kebutuhan untuk evaluasi yang lebih mendalam menggunakan metrik seperti MCC terutama dalam konteks ketidakseimbangan kelas. Selain itu penelitian pada

dataset berbasis fitur klinis atau data numerik seperti *Breast Cancer Wisconsin* (Diagnostic) menekankan pentingnya pemilihan algoritma klasifikasi yang tepat serta teknik *preprocessing* dan optimasi fitur untuk meningkatkan akurasi dan interpretabilitas model.

BAHAN DAN METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain komparatif untuk mengevaluasi performa beberapa algoritma klasifikasi dalam deteksi dini kanker payudara. Pendekatan ini dipilih karena memungkinkan pengujian efektivitas metode prediktif melalui analisis data numerik secara objektif.

Dataset

Data yang digunakan dalam penelitian ini adalah *Breast Cancer Wisconsin* (Diagnostic) yang diperoleh dari repositori *UCI Machine Learning*. Dataset ini terdiri dari 569 entri dengan 30 fitur numerik yang menggambarkan karakteristik mikroskopis inti sel serta satu label target yang mengindikasikan jenis tumor yaitu "*malignant*" (ganas) atau "*benign*" (jinak).

Pra-pemrosesan Data (Data Preprocessing)

Tahapan pra-pemrosesan data dilakukan untuk memastikan kualitas data dan menghindari bias pada proses pelatihan model. Langkah pertama dilakukan pengecekan terhadap *missing value* dan inkonsistensi data. Dataset kemudian dibagi menggunakan stratified train-test split dengan proporsi 80% data latih dan 20% data uji untuk mempertahankan distribusi kelas pada kedua kelompok data.

Selanjutnya normalisasi fitur dilakukan menggunakan metode *Min-Max Scaling*. Proses normalisasi hanya dilakukan pada data latih untuk menghindari data *leakage*, kemudian parameter hasil normalisasi diterapkan pada data uji menggunakan parameter yang sama. Tahapan ini bertujuan agar seluruh fitur memiliki rentang skala yang sebanding sehingga proses pelatihan model menjadi lebih optimal.

Pelatihan dan Pengujian Model

Penelitian ini menerapkan empat algoritma klasifikasi yaitu *Naive Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM). Seluruh model dilatih menggunakan data latih yang telah melalui proses *preprocessing* dan normalisasi.

Konfigurasi hyperparameter untuk masing-masing algoritma ditentukan berdasarkan parameter umum yang banyak digunakan pada penelitian sebelumnya. Model SVM menggunakan kernel *Radial Basis Function* (RBF) dengan parameter $C=1.0$ dan $\gamma='scale'$. *Random Forest* menggunakan 100 estimator dengan $criterion='gini'$, sedangkan *Decision Tree* menggunakan $criterion='gini'$ dengan parameter *default*. Adapun *Naive Bayes* menggunakan *Gaussian Naive Bayes* dengan parameter *default*. Penelitian

ini belum menerapkan hyperparameter *tuning* secara ekstensif sehingga fokus penelitian diarahkan pada analisis komparatif performa dasar antar model klasifikasi.

Untuk memperoleh evaluasi performa yang lebih stabil dan mengurangi bias estimasi model, penelitian ini menggunakan teknik *5-fold cross-validation* yang diterapkan hanya pada data latih. Pada proses ini data latih dibagi menjadi lima subset, dimana empat subset digunakan untuk pelatihan dan satu subset digunakan untuk validasi secara bergantian hingga seluruh subset digunakan sebagai data validasi.

Setelah proses *cross-validation* selesai model akhir dievaluasi menggunakan data uji terpisah yang tidak terlibat pada proses pelatihan maupun validasi. Pendekatan ini dilakukan untuk mengurangi risiko *overfitting* dan mencegah terjadinya data *leakage* selama proses evaluasi model.

Evaluasi Kinerja Model

Kinerja masing-masing model dievaluasi dengan menggunakan metrik evaluasi berikut:

- 1) Akurasi (*Accuracy*), untuk mengukur rasio prediksi benar terhadap keseluruhan data.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (1)$$

- 2) Presisi (*Precision*) dan *Recall*, untuk menilai keseimbangan antara kesalahan positif dan deteksi kasus nyata.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

- 3) *F1-Score*, yang merupakan rata-rata harmonis dari *presisi* dan *recall*.

$$F1 \text{ Score} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (4)$$

- 4) *Area Under Curve - Receiver Operating Characteristic* (ROC-AUC), yang menggambarkan kemampuan model dalam membedakan antara dua kelas.

Matthews Correlation Coefficient (MCC), yang digunakan untuk mengukur kualitas klasifikasi biner secara menyeluruh, termasuk dalam kondisi distribusi label yang tidak seimbang.

HASIL DAN PEMBAHASAN

Evaluasi Model

Evaluasi performa algoritma klasifikasi *Naive Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM) dilakukan menggunakan metrik akurasi, *presisi*, *recall*, *F1-score*, AUC, dan MCC. Hasil menunjukkan adanya variasi kinerja di antara model-model tersebut yang dapat dilihat pada Tabel 2.

Tabel 2 Hasil Evaluasi Model

Performa	Naive Bayes	Decision Tree	Random Forest	Support Vector Machine (SVM)
Accuracy	0.9649	0.9386	0.9649	0.9825
Precision	0.9756	0.9091	0.9756	1.0000
Recall	0.9302	0.9302	0.9302	0.9535
F1 Score	0.9524	0.9195	0.9524	0.9762
AUC	0.9974	0.9369	0.9972	0.9971
MCC	0.9253	0.8701	0.9253	0.9630

Sumber: (Hasil Penelitian, 2026)

Tabel 2 menunjukan bahwa algoritma *Support Vector Machine* (SVM) memberikan hasil terbaik dibandingkan model lain, dengan akurasi tertinggi sebesar 98,25%, presisi sempurna (1.0000), dan nilai *F1-score* sebesar 0.9762, yang menunjukkan keseimbangan optimal antara presisi dan recall. Model ini juga memperoleh nilai Matthews Correlation Coefficient (MCC) sebesar 0.9630, yang menunjukkan kualitas klasifikasi biner yang sangat baik serta konsistensi prediksi antara kelas *benign* dan *malignant*.

Model *Naive Bayes* dan *Random Forest* menunjukkan performa yang serupa, masing-masing memiliki akurasi 96.49%, *F1-score* 0.9524, dan nilai AUC di atas 0.997, menunjukkan kemampuan yang sangat baik dalam membedakan antara dua kelas (*malignant* dan *benign*). *Decision Tree*, meskipun kinerjanya cukup baik, tercatat memiliki nilai metrik yang sedikit lebih rendah di semua aspek, khususnya nilai AUC (0.9369) yang mengindikasikan area di bawah kurva ROC yang lebih sempit dibandingkan model lainnya.

Analisis Feature Importance

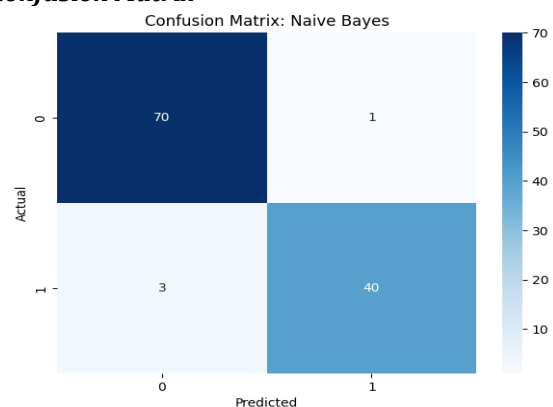
Analisis *feature importance* dilakukan untuk mengidentifikasi fitur-fitur yang paling berkontribusi terhadap proses klasifikasi kanker payudara, khususnya pada model berbasis pohon seperti *Decision Tree* dan *Random Forest*. Berdasarkan karakteristik dataset *Breast Cancer Wisconsin (Diagnostic)* beberapa fitur seperti *radius_mean*, *perimeter_mean*, *area_mean*, *concavity_mean*, dan *texture_mean* memiliki

pengaruh yang cukup dominan dalam membedakan tumor *benign* dan *malignant*.

Fitur-fitur tersebut berkaitan dengan ukuran, bentuk, dan tekstur inti sel yang secara medis memang sering digunakan sebagai indikator dalam diagnosis kanker payudara. Model *Random Forest* menunjukkan kemampuan yang lebih stabil dalam memanfaatkan kombinasi fitur-fitur tersebut dibandingkan *Decision Tree* tunggal sehingga menghasilkan performa klasifikasi yang lebih tinggi.

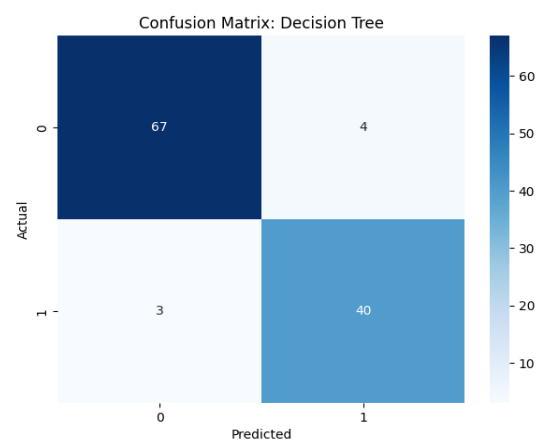
Meskipun demikian penelitian ini belum menerapkan metode interpretabilitas lanjutan seperti SHAP (SHapley Additive exPlanations) atau LIME untuk menjelaskan kontribusi setiap fitur secara lebih detail pada masing-masing prediksi model. Oleh karena itu penelitian selanjutnya disarankan mengintegrasikan pendekatan explainable artificial intelligence (XAI) agar interpretasi model *machine learning* pada data medis menjadi lebih transparan dan mudah dipahami.

Confusion Matrix



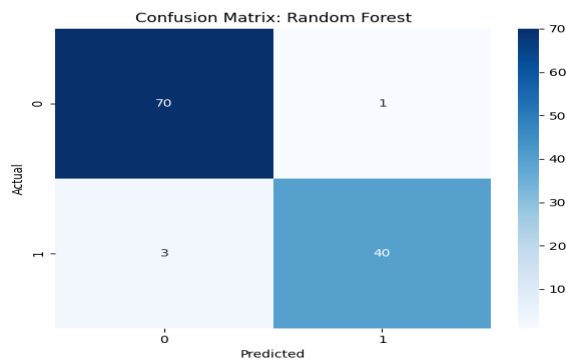
Sumber: (Hasil Penelitian, 2026)

Gambar 1 Confusion Matrix Model *Naive Bayes*



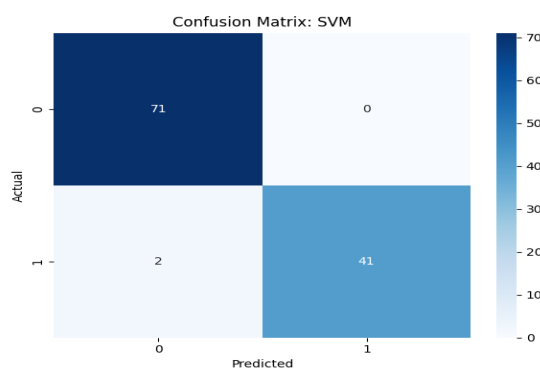
Sumber: (Hasil Penelitian, 2026)

Gambar 2 Confusion Matrix Model *Decision Tree*



Sumber: (Hasil Penelitian, 2026)

Gambar 3 Confusion Matrix Model Random Forest



Sumber: (Hasil Penelitian, 2026)

Gambar 4 Confusion Matrix Model SVM

Berdasarkan hasil *confusion matrix* dari algoritma klasifikasi *Naive Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM) pada gambar 1, 2, 3 dan 4, dilakukan analisis terhadap kemampuan masing-masing model dalam mengklasifikasikan data kanker payudara.

Model *Naive Bayes* berhasil mengklasifikasikan 70 data negatif dengan benar (*true negative*) dan 40 data positif dengan benar (*true positive*), dengan hanya satu kesalahan klasifikasi positif palsu (*false positive*) dan tiga kesalahan klasifikasi negatif palsu (*false negative*). Hasil ini menunjukkan bahwa *Naive Bayes* memiliki performa yang cukup baik dalam menyeimbangkan antara sensitivitas dan spesifisitas.

Sementara itu model *Decision Tree* menunjukkan jumlah *true positive* dan *false negative* yang sama dengan *Naive Bayes*, yaitu masing-masing sebanyak 40 dan 3. Namun, model ini menghasilkan *false positive* yang lebih tinggi, yaitu sebanyak 4, dan *true negative* sebanyak 67. Hal ini menunjukkan bahwa *Decision Tree* cenderung memberikan lebih banyak kesalahan dalam mengklasifikasikan data negatif sebagai positif, yang berisiko menimbulkan *false alarm* dalam konteks medis.

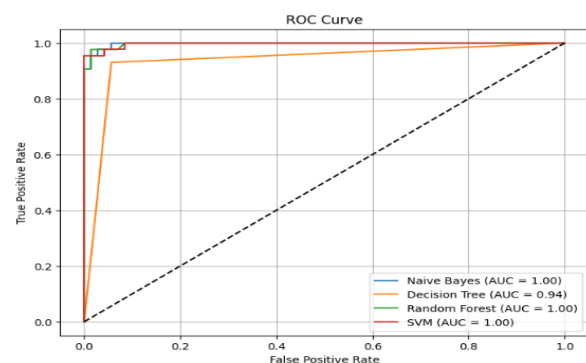
Model *Random Forest* menghasilkan hasil yang identik dengan *Naive Bayes*, yaitu 70 *true negative*, 1 *false positive*, 3 *false negative*, dan 40 *true positive*. Ini menunjukkan bahwa meskipun metode yang digunakan berbeda karena *Random Forest* merupakan model berbasis *ensemble* kinerjanya tetap sebanding dengan *Naive Bayes* dalam hal akurasi klasifikasi.

Adapun model SVM (*Support Vector Machine*) menunjukkan kinerja terbaik diantara keempat model. SVM mampu mengklasifikasikan semua data negatif dengan benar 71 *true negative* tanpa menghasilkan *false positive* sama sekali. Selain itu SVM hanya menghasilkan dua *false negative* dan 41 *true positive* yang merupakan jumlah tertinggi dibanding model lainnya. Hasil ini menunjukkan bahwa SVM memiliki tingkat sensitivitas dan spesifisitas yang sangat tinggi menjadikannya sebagai model yang paling andal dalam deteksi dini kanker payudara berdasarkan *dataset* yang digunakan.

Secara keseluruhan keempat model menunjukkan kinerja yang cukup baik tetapi SVM tampil unggul karena mampu meminimalkan kesalahan klasifikasi secara signifikan. Oleh karena itu SVM dapat dipertimbangkan sebagai pilihan utama dalam sistem pendeteksian dini kanker payudara berbasis klasifikasi.

Interpretasi Kurva ROC

Kurva ROC menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai ambang klasifikasi. Semakin dekat kurva menuju pojok kiri atas grafik semakin baik performa model karena menunjukkan bahwa model memiliki TPR tinggi dan FPR rendah. Selain itu nilai *Area Under the Curve* (AUC) digunakan sebagai indikator kuantitatif untuk menilai kemampuan diskriminatif suatu model: nilai AUC mendekati 1 menunjukkan performa klasifikasi yang sangat baik.



Sumber: (Hasil Penelitian, 2026)

Gambar 5. ROC Curve

Gambar 5 terlihat bahwa model *Naive Bayes*, *Random Forest*, dan *SVM* memiliki kemampuan diskriminatif yang sangat tinggi dengan nilai AUC masing-masing sebesar 0.9974, 0.9972, dan 0.9971. Nilai tersebut menunjukkan bahwa ketiga model mampu membedakan kelas *benign* dan *malignant* dengan sangat baik. Sementara itu *Decision Tree* memperoleh nilai AUC sebesar 0.9369 yang meskipun lebih rendah dibandingkan model lainnya tetap berada dalam kategori performa yang baik. Dengan demikian evaluasi ROC-AUC memperlihatkan bahwa *SVM*, *Random Forest*, dan *Naive Bayes* merupakan model yang paling stabil dan efektif dalam klasifikasi kanker payudara pada penelitian ini

Secara keseluruhan analisis kurva ROC menunjukkan bahwa *Naive Bayes*, *Random Forest*, dan *SVM* merupakan model yang sangat andal untuk deteksi dini kanker payudara dalam penelitian ini. Ketiganya berhasil mencapai nilai AUC yang sangat tinggi (di atas 0,99) menunjukkan kemampuan diskriminatif yang sangat baik dalam membedakan kelas *benign* dan *malignant*. Namun demikian hasil *confusion matrix* masih menunjukkan adanya sejumlah kecil kesalahan klasifikasi, sehingga performa model belum dapat dikategorikan sebagai sempurna.

Penelitian ini belum menerapkan nested cross-validation serta analisis interpretabilitas model seperti feature importance atau SHAP (SHapley Additive exPlanations), sehingga masih terdapat kemungkinan estimasi performa yang cenderung optimistis. Oleh karena itu, penelitian selanjutnya disarankan menggunakan nested cross-validation dan pendekatan explainable artificial intelligence (XAI) untuk meningkatkan reliabilitas evaluasi serta interpretabilitas model klasifikasi.

Selain itu dataset *Breast Cancer Wisconsin (Diagnostic)* yang digunakan dalam penelitian ini memiliki distribusi kelas yang relatif seimbang, sehingga belum merepresentasikan kondisi ketidakseimbangan kelas yang ekstrem sebagaimana sering ditemukan pada data medis di dunia nyata. Oleh karena itu meskipun MCC digunakan sebagai metrik evaluasi utama untuk memberikan penilaian yang lebih komprehensif, penelitian ini belum mengevaluasi pengaruh teknik penanganan ketidakseimbangan kelas seperti *Synthetic Minority Over-sampling Technique* (SMOTE), *Adaptive Synthetic Sampling* (ADASYN), maupun *class weighting*. Penelitian selanjutnya dapat mengkaji efektivitas teknik-teknik tersebut untuk meningkatkan performa model pada dataset dengan tingkat ketidakseimbangan kelas yang lebih tinggi.

KESIMPULAN

Berdasarkan analisis performa model menggunakan kurva ROC, AUC, dan MCC, dapat disimpulkan bahwa algoritma *Naive Bayes*, *Random Forest*, dan *Support Vector Machine* (*SVM*) menunjukkan kinerja klasifikasi yang sangat baik dengan nilai AUC masing-masing sebesar 0,9974, 0,9972, dan 0,9971. Hasil tersebut menunjukkan kemampuan diskriminatif yang sangat tinggi dalam membedakan kasus *benign* dan *malignant* pada dataset *Breast Cancer Wisconsin* (Diagnostic). Di antara seluruh model yang diuji, *SVM* memberikan performa terbaik dengan akurasi 98,25%, F1-score 0,9762, dan MCC 0,9630, serta mampu meminimalkan kesalahan klasifikasi dibandingkan model lainnya. Sementara itu algoritma *Decision Tree* juga menunjukkan performa yang baik dengan nilai AUC sebesar 0,9369 meskipun lebih rendah dibandingkan model lainnya. Hasil penelitian ini menunjukkan bahwa pendekatan multi-model dan penggunaan MCC sebagai metrik evaluasi pelengkap ROC-AUC dapat memberikan gambaran yang lebih komprehensif mengenai kualitas klasifikasi. Dengan demikian pemilihan model yang tepat berpotensi mendukung sistem deteksi dini kanker payudara yang lebih akurat dan andal.

REFERENSI

- Almarri, B. *et al.* (2024) "The BCPM method: decoding breast cancer with machine learning," *BMC Medical Imaging*, 24(1). Available at: <https://doi.org/10.1186/s12880-024-01402-5>.
- Boddu, A.S. and Jan, A. (2025) "A systematic review of machine learning algorithms for breast cancer detection.," *Tissue & cell*, 95, p. 102929. Available at: <https://doi.org/10.1016/j.tice.2025.102929>.
- CHICCO, D. and JURMAN, G. (2023) "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, 16(1), pp. 1–23. Available at: <https://doi.org/10.1186/s13040-023-00322-4>.
- Chicco, D. and Jurman, G. (2023) "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, 16(1). Available at: <https://doi.org/10.1186/s13040-023-00322-4>.

- Darweesh, M.S. *et al.* (2021) "Early breast cancer diagnostics based on hierarchical machine learning classification for mammography images," *Cogent Engineering*, 8(1). Available at: <https://doi.org/10.1080/23311916.2021.1968324>.
- DAS, S., KOLEY, S. and SAHA, T. (2023) "Machine Learning Approaches for Investigating Breast Cancer," *Biosciences Biotechnology Research Asia*, 20(4), pp. 1109–1131. Available at: <https://doi.org/10.13005/bbra/3163>.
- HANIS, T.M., ISLAM, M.A. and MUSA, K.I. (2022) "Diagnostic Accuracy of Machine Learning Models on Mammography in Breast Cancer Classification: A Meta-Analysis," *Iranian Journal of Radiology*, 19(3). Available at: <https://doi.org/10.5812/iranjradiol-121392>.
- Hickman, S.E. *et al.* (2024) "Deep Learning Algorithms for Breast Cancer Detection in a UK Screening Cohort: As Stand-alone Readers and Combined with Human Readers," *Radiology*, 313(2). Available at: <https://doi.org/10.1148/radiol.233147>.
- Hussain, S. *et al.* (2024) "Breast cancer risk prediction using machine learning: a systematic review," *Frontiers in Oncology*. Frontiers Media SA. Available at: <https://doi.org/10.3389/fonc.2024.134362>.
- Jong, L.J.S. *et al.* (2023) "Tissue Classification of Breast Cancer by Hyperspectral Unmixing," *Cancers*, 15(10). Available at: <https://doi.org/10.3390/cancers15102679>.
- JOY, A. *et al.* (2023) "Ensemble Learning for Breast Cancer Lesion Classification: A Pilot Validation Using Correlated Spectroscopic Imaging and Diffusion-Weighted Imaging," *Metabolites*, 13(7). Available at: <https://doi.org/10.3390/metabo13070835>.
- KUMAR, A., SAINI, R. and KUMAR, R. (2024) "A Comparative Analysis of Machine Learning Algorithms for Breast Cancer Detection and Identification of Key Predictive Features," *Traitement du Signal*, 41(1), pp. 127–140. Available at: <https://doi.org/10.18280/ts.410110>.
- PATI, A. *et al.* (2024) "Performance assessment of hybrid machine learning approaches for breast cancer and recurrence prediction," *PLoS ONE*, 19(8 August), pp. 1–29. Available at: <https://doi.org/10.1371/journal.pone.0304768>.
- Rahman, M.A. *et al.* (2025) *Advancements in Breast Cancer Detection: A Review of Global Trends, Risk Factors, Imaging Modalities, Machine Learning, and Deep Learning Approaches*, *BioMedInformatics*. Available at: <https://doi.org/10.3390/biomedinformatics5030046>.
- SAHOO, G. *et al.* (2024) "Predicting Breast Cancer Relapse from Histopathological Images with Ensemble Machine Learning Models," *Current Oncology*, 31(11), pp. 6577–6597. Available at: <https://doi.org/10.3390/currncol31110486>.
- SINGHAL, V. *et al.* (2022) "Breast Cancer Prediction using KNN, SVM, Logistic Regression and Decision Tree," *Ijrasnet Journal For Research in Applied Science and Engineering Technology* [Preprint], (May). Available at: <https://doi.org/https://doi.org/10.22214/ijrasnet.2022.42688>.
- Walsh, R. and Tardy, M. (2023) "A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer," *Diagnostics*, 13(1). Available at: <https://doi.org/10.3390/diagnostics13010067>.
- Zhang, T. *et al.* (2023) "Predicting breast cancer types on and beyond molecular level in a multi-modal fashion," *npj Breast Cancer*, 9(1). Available at: <https://doi.org/10.1038/s41523-023-00517-2>.
- ZHANG, X. *et al.* (2025) "Advanced deep learning and transfer learning approaches for breast cancer classification using advanced multi-line classifiers and datasets with model optimization and interpretability," *PeerJ Computer Science*, 11, pp. 1–35. Available at: <https://doi.org/10.7717/peerj-cs.2951>.