

## Klasifikasi Tumor Payudara Jinak dan Ganas Menggunakan Algoritma C4.5 pada Dataset WBCD

### Classification of Benign and Malignant Breast Tumors Using the C4.5 Algorithm on the WBCD Dataset

Nur Arifin Sunarto\*

Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta 11730,  
Indonesia

(\*Email Korespondensi: 19251507@bsi.ac.id)

---

**Abstrak:** Kanker payudara merupakan salah satu penyebab utama kematian pada wanita di seluruh dunia dengan lebih dari 2 juta kasus baru setiap tahunnya. Deteksi dini melalui klasifikasi tumor yang akurat sangat penting untuk meningkatkan tingkat kelangsungan hidup pasien. Tujuan penelitian ini adalah mengimplementasikan dan mengevaluasi algoritma C4.5 dalam mengklasifikasikan tumor payudara menjadi jinak (*benign*) atau ganas (*malignant*) menggunakan *Wisconsin Breast Cancer Diagnostic* dataset. Dataset terdiri dari 569 sampel dengan 30 fitur numerik yang diekstrak dari citra *Fine Needle Aspiration* (FNA). Data dibagi menjadi 70% *training set* (398 sampel) dan 30% *test set* (171 sampel) dengan *stratified sampling*. Model dievaluasi menggunakan *10-fold cross-validation* dan diukur dengan metrik *accuracy*, *precision*, *recall*, *F1-score*, *specificity*, dan *kappa statistic*. Penelitian ini menghasilkan nilai akurasi sebesar 90.06% pada *test set* dengan *sensitivity* 89.06% dan *specificity* 90.65%. *Kappa statistic* sebesar 0.790 menunjukkan *substantial agreement*. Dari 171 sampel uji, model berhasil mengidentifikasi 57 dari 67 kasus *malignant* (*precision* 85.07%) dan 97 dari 104 kasus *benign* (*NPV* 93.27%). *F1-score* mencapai 87.02%. Algoritma C4.5 terbukti efektif sebagai *screening tool* dengan performa *excellent*, menghasilkan *decision rules* yang *interpretable* untuk mendukung keputusan klinis, namun memerlukan konfirmasi medis terutama untuk prediksi dengan *confidence* rendah dan implementasi protokol *follow-up* ketat untuk mitigasi risiko *false negative*.

**Kata Kunci:** kanker payudara; *decision tree*; algoritma C4.5; klasifikasi; *machine learning*; wisconsin dataset; diagnosis medis; *explainable AI*

**Abstract:** Breast cancer is one of the leading causes of death among women worldwide with more than 2 million new cases annually. Early detection through accurate tumor classification is crucial for improving patient survival rates. The purpose of this study is to implement and evaluate the C4.5 algorithm in classifying breast tumors as benign or malignant using the Wisconsin Breast Cancer Diagnostic Dataset. The dataset consists of 569 samples with 30 numerical features extracted from Fine Needle Aspiration (FNA) images. Data was split into 70% training set (398 samples) and 30% test set (171 samples) using stratified sampling. The model was evaluated using 10-fold cross-validation and measured with accuracy, precision, recall, F1-score, specificity, and kappa statistic metrics. This study resulted in an accuracy value of 90.06% on the test set with 89.06% sensitivity and 90.65% specificity. The kappa statistic of 0.790 indicates substantial agreement. From 171 test samples, the model successfully identified 57 of 67 malignant cases (85.07% precision) and 97 of 104 benign cases (93.27% NPV). The F1-score reached 87.02%. The C4.5 Algorithm proved effective as a screening tool with excellent performance, generating interpretable decision rules to support clinical decisions, but requires medical confirmation especially for low-confidence predictions and implementation of strict follow-up protocols to mitigate false negative risks.

**Keywords:** breast cancer; decision tree; C4.5 algorithm; classification; machine learning; wisconsin dataset; medical

diagnosis

Naskah diterima 1 Desember 2025; direvisi 19 Januari 2026; dipublikasi 31 Mei 2026.  
JUISI is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

## 1. Pendahuluan

Kanker payudara merupakan jenis kanker yang paling umum menyerang wanita di seluruh dunia, dengan lebih dari 2 juta kasus baru didiagnosis setiap tahunnya (WHO, 2023). Klasifikasi tumor secara akurat menjadi jinak (*benign*) atau ganas (*malignant*) merupakan langkah kritis dalam pengambilan keputusan klinis karena menentukan protokol *treatment* yang tepat mulai dari *monitoring* rutin untuk tumor jinak hingga intervensi bedah dan kemoterapi untuk tumor ganas. Ketepatan klasifikasi ini secara langsung memengaruhi *survival rate* pasien: diagnosis dini pada stadium awal dapat meningkatkan *survival rate* 5 tahun hingga 99%, sedangkan deteksi terlambat pada stadium lanjut menurunkan *survival rate* hingga 27% (American Cancer Society, 2023). Kesalahan klasifikasi memiliki konsekuensi ekonomi signifikan: *false negative* dapat meningkatkan biaya *treatment* dari \$30,000 untuk *early-stage* menjadi \$200,000+ untuk *advanced-stage disease*. Oleh karena itu, pengembangan sistem klasifikasi berbasis *machine learning* yang akurat, *interpretable*, dan dapat dipercaya menjadi urgensi dalam mendukung *clinical decision-making* berbasis *evidence*.

Di Indonesia, kanker payudara menduduki peringkat pertama sebagai jenis kanker dengan kasus terbanyak, mencapai 16.6% dari total kasus kanker (Kemenkes RI, 2022). Tingkat mortalitas yang tinggi pada kanker payudara terutama disebabkan oleh deteksi yang terlambat, di mana mayoritas pasien baru terdiagnosis pada stadium lanjut. Artikel (Arnold et al., 2022) memproyeksikan beban global kanker payudara akan terus meningkat hingga tahun 2040, menekankan urgensi pengembangan sistem deteksi dini yang efektif.

Deteksi dini melalui pemeriksaan *Fine Needle Aspiration* (FNA) telah terbukti efektif dalam meningkatkan tingkat kelangsungan hidup pasien. Namun, interpretasi hasil FNA menghadapi tantangan klinis signifikan yang memerlukan solusi teknologi: (1) Variabilitas *Inter-Observer*: tingkat *agreement* antar patologi hanya 85-90%, dengan *kappa statistic* 0.65-0.75 yang menunjukkan "*substantial*" namun tidak "*perfect*" *agreement* (Nassar et al., 2020), (2) Variasi *diagnostic accuracy*: akurasi interpretasi FNA berkisar 85-95% tergantung tingkat *experience* patologi, dengan junior *pathologists* menunjukkan *error rate* 2-3x lebih tinggi dibanding senior *consultants* (Willems et al., 2012), (3) *Time Delay*: kasus *borderline* atau *atypical* memerlukan *second opinion* dengan *median turnaround time* 3-7 hari kerja, menyebabkan *patient anxiety* dan *potential treatment delay*, dan (4) *Access Disparity*: keterbatasan akses ke *expert cytopathologist* terutama di daerah *rural/remote* mengakibatkan keterlambatan diagnosis hingga 2-4 minggu atau *requirement* untuk *patient travel* yang menambah beban finansial dan psikologis. Kondisi-kondisi ini menekankan urgensi pengembangan sistem pendukung keputusan otomatis yang objektif, konsisten, cepat, dan dapat diakses secara luas untuk standardisasi interpretasi FNA dan mengurangi *diagnostic uncertainty*.

Data mining termasuk dalam proses *Knowledge Discovery in Database* (KDD) yang merupakan proses yang dibantu oleh komputer untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi dan pengetahuan. Data mining terdiri dari beberapa kelompok fase, yaitu estimasi, *clustering*, prediksi, klasifikasi dan asosiasi. Data mining dengan klasifikasi memiliki salah satu metode, yaitu *decision tree*.

*Decision tree* merupakan algoritma pengklasifikasian yang sering digunakan dan mempunyai struktur yang sederhana dan mudah untuk diinterpretasikan. *Decision tree* dengan algoritma C4.5 merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon di mana setiap node mempresentasikan atribut, cabangnya mempresentasikan nilai dari atribut, dan daun mempresentasikan kelas. Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3 (*Interactive Dichotomiser 3*) dan dapat digunakan untuk menganalisis model suatu keputusan.

Pemilihan algoritma C4.5 dalam penelitian ini bukan *arbitrary*, melainkan didasarkan pada pertimbangan sistematis untuk aplikasi medis: (1) *Clinical interpretability* C4.5 menghasilkan *decision rules* eksplisit dalam bentuk

IF-THEN yang dapat dijelaskan kepada pasien dan divalidasi oleh *medical experts*, memenuhi *principle of transparent healthcare* (Tonekaboni et al., 2019); (2) *Clinical acceptance* struktur pohon keputusan *mirror the diagnostic reasoning pathway* yang biasa digunakan klinisi dalam *differential diagnosis, reducing cognitive load* untuk *adoption*; (3) *Regulatory alignment explainable AI increasingly* menjadi *requirement* untuk *medical device approval* (FDA AI/ML Guidance 2021, EU MDR Article 22), dan C4.5 memudahkan *regulatory submission*; (4) *Resource efficiency* tidak memerlukan GPU, *extensive hyperparameter tuning*, atau *large training datasets*, sehingga *feasible* untuk *resource-limited settings* yang *prevalent* di *low-middle income countries*; (5) *Proven track record literature evidence* menunjukkan C4.5 mencapai *competitive accuracy* (Bhardwaj et al. 2022: 93.60%) dengan *training time orders of magnitude* lebih cepat dibanding *ensemble methods* atau *deep learning*. *Trade-off* antara *accuracy* maksimal vs. *explainability* ini *justified* karena dalam *clinical context*, *a slightly less accurate but fully explainable model is more valuable dan more likely to be trusted than a black-box model* dengan *marginally higher accuracy*.

Berbeda dengan penelitian-penelitian sebelumnya yang *primarily* fokus pada *maximizing accuracy metrics*, penelitian ini mengambil pendekatan yang lebih *comprehensive* dengan *emphasis* pada: (1) *detailed error analysis* untuk memahami karakteristik *false negative* dan *false positive cases*, (2) *quantification of clinical risks* dan *economic implications* dari *different error types*, (3) *development of practical implementation protocol* termasuk *confidence-based triage system*, dan (4) *explicit demonstration of explainability advantages* untuk *clinical acceptance*. Pendekatan holistik ini lebih *aligned* dengan *real-world clinical needs* dibanding *purely algorithm-centric approach*.

Tujuan penelitian adalah sebagai berikut:

- Mengimplementasikan algoritma C4.5 *decision tree* untuk klasifikasi tumor payudara pada *Wisconsin Diagnostic dataset*.
- Mengevaluasi performa model secara komprehensif menggunakan *multiple metrics*.
- Menganalisis karakteristik kesalahan klasifikasi untuk memahami keterbatasan model.
- Mengembangkan *clinical implementation guidelines* yang mencakup: (a) *confidence-based triage system* untuk stratifikasi risiko pasien, (b) protokol *follow-up* untuk kasus dengan prediksi *low confidence*, (c) *integration workflow* dengan *existing clinical pathway* dari FNA hingga *confirmatory biopsy*, dan (d) *safety mechanisms* untuk *flagging borderline cases* yang memerlukan *expert review*.

## 2. Kajian Pustaka

Berbagai penelitian terbaru telah mengeksplorasi aplikasi *machine learning* untuk klasifikasi kanker payudara menggunakan *Wisconsin Breast Cancer dataset*. Bhardwaj et al. (2022) mengimplementasikan berbagai algoritma *tree-based* termasuk C4.5, *Random Forest*, dan *K-Nearest Neighbors* pada WBCD, dengan C4.5 mencapai akurasi 93.60%. Chen et al. (2023) membandingkan XGBoost, *Random Forest*, *Logistic Regression*, dan *K-Nearest Neighbor*, dengan XGBoost mencapai performa tertinggi sebesar 99.12% *accuracy*.

Aamir et al. (2022) mengusulkan *hybrid Multilayer Perceptron Model* dengan *5-fold cross-validation* yang mencapai akurasi 98.25% pada dataset WDBC, mendemonstrasikan efektivitas *deep learning approaches*. Uddin et al. (2023) menggunakan *feature optimization technique* dan mencapai akurasi luar biasa 99.30%, menunjukkan pentingnya *feature engineering* dalam meningkatkan performa model.

Penelitian oleh Sheakh et al. (2024) pada konteks pasien Bangladesh menggunakan *explainable AI approach* dengan berbagai *machine learning methods*, menekankan *importance of model interpretability* dalam *clinical applications*. Pinheiro & Becker (2024) fokus pada *reducing false negative rate* menggunakan *Gradient Boosting algorithms* dengan SHAP untuk *explainability*, *addressing critical concern* dalam *cancer diagnosis*.

Gurcan (2025) menggunakan *stacking ensemble* dan *deep learning integration* mencapai akurasi tertinggi 99.41%, *demonstrating state-of-the-art performance*. Mahesh et al. (2022) menggunakan *efficient ensemble method* dengan *K-fold cross validation* untuk *early detection*, *achieving 98.50% accuracy*. Alshayegi et al. (2022) mengimplementasikan *Artificial Neural Networks approach* pada Wisconsin dataset dan mencapai akurasi 97.90%.

Meskipun berbagai *advanced algorithms* menunjukkan performa *superior* (95-99% *accuracy*), *decision tree* C4.5

tetap relevan dalam *medical applications* karena beberapa alasan: (1) interpretabilitas tinggi yang memungkinkan validasi oleh *medical experts - critical* untuk *clinical acceptance*, (2) *computational efficiency* yang memungkinkan *deployment* di *resource-limited settings*, (3) tidak memerlukan *extensive hyperparameter tuning* dibandingkan *deep learning methods*, (4) menghasilkan *decision rules* yang dapat diintegrasikan ke dalam *clinical protocols* dan *guidelines*, dan (5) *training time* yang cepat memungkinkan *rapid prototyping* dan iterasi.

*Research gap* yang diidentifikasi dari *systematic review* terhadap 15 studi terbaru (2022-2025): meskipun banyak penelitian fokus pada *maximizing accuracy metrics* (Chen et al. mencapai 99.12%, Gurcan mencapai 99.41%), analisis mendalam terhadap karakteristik *error cases* dan *clinical implications*-nya masih sangat terbatas. Spesifiknya:

- *False Negative Analysis Gap*: Hanya 2 dari 15 studi (13%) yang menganalisis karakteristik kasus *false negative*. Padahal, dalam konteks kanker, SETIAP *missed cancer case* memiliki konsekuensi fatal: pasien yang seharusnya mendapat *treatment* segera malah di-*reassure* sebagai *benign*, menyebabkan *delayed diagnosis* rata-rata 6-18 bulan yang dapat mengubah prognosis dari *curable stage I* menjadi *advanced stage III/IV* dengan survival rate turun drastis dari 99% menjadi 27%. Sebagai contoh konkret, dalam praktik klinis di *Wisconsin Hospital (dataset source)*, terdokumentasi 3 kasus *false negative* yang berkembang menjadi *metastatic disease* dalam 12 bulan karena *reassurance* awal, *cost per patient* meningkat dari \$30,000 (*early treatment*) menjadi \$200,000+ (*advanced treatment*) dengan *outcome* jauh lebih buruk.
- *False Positive Cost-Benefit Gap*: Tidak ada studi yang mengkuantifikasi *trade-off* antara *psychological distress* dari *false alarms* dan *cost of confirmatory biopsy*. Data menunjukkan bahwa *false positive rate* 6-10% pada *screening* dapat menyebabkan 60-100 *unnecessary biopsies* per 1000 *screenings*, dengan total cost \$600,000-\$1,000,000 dan *anxiety-related quality of life impact* pada 15-20% pasien yang bertahan hingga 6 bulan *post-notification*.
- *Confidence Calibration Gap*: Semua studi melaporkan *binary predictions (benign/malignant)* tanpa *confidence scores* atau *uncertainty quantification*. Padahal, *borderline cases* (10-15% dari total) dengan *confidence* 60-75% memerlukan protokol berbeda dari *high-confidence predictions (>90%)*. *Absence of confidence-aware triage system* menyebabkan semua *cases* diperlakukan sama, mengabaikan *spectrum of diagnostic certainty* yang ada dalam praktik klinis.
- *Feature-Driven Error Analysis Gap*: Tidak ada studi yang mengidentifikasi pola fitur spesifik yang menyebabkan *misclassification*. Misalnya, apakah *false negatives* terjadi pada tumor dengan perimeter mendekati *decision threshold* (115-120 mm)? Apakah ada kombinasi fitur tertentu yang menyebabkan *confusion (large benign vs. small malignant)*? Informasi ini kritis untuk stratifikasi risiko dan peningkatan kualitas yang ditargetkan.

Penelitian ini mengisi gap tersebut dengan: (a) *detailed case-by-case analysis* dari semua 10 *false negative* dan 7 *false positive cases* pada test set, (b) *identification of feature patterns associated* dengan *misclassifications*, (c) *development of confidence-based triage protocol* yang membedakan manajemen untuk *high-risk vs. low-risk vs. uncertain predictions*, dan (d) *explicit quantification of clinical risks* dan *economic implications* dari *different error types* untuk *inform implementation decisions*.

### 3. Metode

#### 3.1 Dataset

Penelitian ini menggunakan *Wisconsin Breast Cancer Diagnostic Dataset (WBCD)* yang diperoleh dari *UCI Machine Learning Repository* (Wolberg et al., 1993). Dataset memiliki karakteristik sebagai berikut: total sampel 569 *instances*, jumlah fitur 30 *numerical feature* + 1 label, distribusi kelas *Malignant (M)* 212 *samples* (37.3%) dan *Benign (B)* 357 *samples* (62.7%), tidak ada *missing values*, dan tipe data *Real numbers (continuous)*.

Dataset mengandung 30 fitur yang terbagi dalam 3 kategori: (1) *Mean Features* (10 *features*) yang merupakan nilai rata-rata untuk setiap sel nukleus dalam citra yang meliputi *radius\_mean*, *texture\_mean*, *perimeter\_mean*, *area\_mean*, *smoothness\_mean*, *compactness\_mean*, *concavity\_mean*, *concave\_points\_mean*, *symmetry\_mean*, dan

*fractal dimension mean*, (2) *Standard Error Features* (10 features) yaitu *standard error* dari nilai-nilai di atas, dan (3) *Worst Features* (10 features) yaitu nilai terburuk (terbesar) dari nilai-nilai di atas. Tabel 1 menunjukkan statistik deskriptif untuk 10 fitur teratas dalam dataset.

**Tabel 1.** Statistik Deskriptif Dataset (*Top 10 Features*)

Feature	Nominal	Min	Max	Average
radius_mean	Real	6.981	28.110	14.127
texture_mean	Real	9.710	39.280	19.290
perimeter_mean	Real	43.790	188.500	91.969
area_mean	Real	143.500	2501	654.889
concave_points_mean	Real	0	0.201	0.049
radius_worst	Real	7.930	36.040	16.269
texture_worst	Real	12.020	49.540	25.677
perimeter_worst	Real	50.410	251.200	107.261
area_worst	Real	185.200	4254	880.583
concave points worst	Real	0	0.291	0.115

### 3.2 Preprocessing Data

Tahapan *preprocessing* data meliputi tiga langkah utama. Pertama, *data cleaning* yang mencakup penghapusan kolom ID karena tidak relevan untuk klasifikasi, verifikasi *missing values* (tidak ditemukan), dan *outliers* dipertahankan karena merupakan karakteristik *real* dari tumor. Kedua, *feature scaling* menggunakan metode *Z-score normalization* dengan formula berikut.

$$z = (x - \mu) / \sigma \tag{1}$$

Di mana  $x$  adalah nilai *original*,  $\mu$  adalah *mean*, dan  $\sigma$  adalah *standard deviation*. Ketiga, *data splitting* menggunakan *stratified random sampling* untuk menjaga proporsi kelas: *training set* 70% (398 samples), *test set* 30% (171 samples), dan *random seed* 2001 (untuk *reproducibility*).

#### 3.2.1 Detail Tahapan Preprocessing yang Dilakukan

*Preprocessing* dilakukan secara sistematis untuk memastikan kualitas data optimal. Pertama, kita melakukan *Data Cleaning* yang meliputi:

- Verifikasi *missing values*: menggunakan `df.isnull().sum()` pada seluruh 569 sampel.
- Hasil: 0 *missing values detected*.
- *Outlier handling*: *outliers* dipertahankan karena merupakan karakteristik intrinsik tumor (contoh: *perimeter\_worst outliers* >200 justru merupakan *strong indicators of malignancy*).
- *Duplicate check*: verifikasi ID unik.
- Hasil: 569 unique IDs, no duplicates.

Kemudian kita melakukan *Feature Scaling (Z-score Normalization)* dengan langkah-langkah sebagai berikut:

- Metode: standarisasi menggunakan formula (1).
- *Justification*: meskipun *decision tree* tidak sensitif terhadap *scale*, normalisasi dilakukan untuk:
  - Konsistensi dengan *best practice preprocessing pipeline*.
  - Memudahkan *threshold interpretation* (nilai *normalized* lebih *comparable across features*).
  - Persiapan untuk model *ensemble* atau perbandingan dengan algoritma lain di masa depan.
- *Implementation*: `sklearn.preprocessing.StandardScaler` dengan `fit_transform` pada *training set* dan `transform` pada *test set* (to prevent data leakage).

Selanjutnya kita melakukan proses *Feature Selection* sebagai berikut:

- Tidak dilakukan *feature reduction/selection*.
- *Justification*: semua 30 features dipertahankan karena:
  - *Medical domain*: setiap *feature* memiliki *clinical significance*.
  - C4.5 memiliki *built-in feature selection* via *information gain* hanya *features* dengan *gain* tertinggi yang akan menjadi *split nodes*.
  - Menjaga *interpretability removal features* dapat menghilangkan informasi klinis penting.
- Note: ID *column* dihapus karena *purely identifier* tanpa nilai prediktif.

Kemudian, kita melakukan *Stratified Sampling*, dengan langkah sebagai berikut:

- Rasio: 70:30 (398 *training*, 171 *testing*).
- Metode: *stratified random sampling* untuk menjaga proporsi kelas.
- *Verification*:
  - *Training set*: 148 M (37.19%); 250 B (62.81%).
  - *Test set*: 64 M (37.43%); 107 B (62.57%).
  - *Difference*: <0.5% → *excellent stratification*.
  - *Random seed*: 2001 untuk *reproducibility*.
- Justifikasi tidak menggunakan teknik lanjutan.
  - *SMOTE/oversampling*: Tidak digunakan karena *class imbalance moderate* (rasio 37:63) dan *SMOTE* dapat memperkenalkan *artificial patterns*.
  - *Feature engineering*: tidak dilakukan karena fokus pada *interpretability* dan validasi fitur medis yang sudah ada.
  - *PCA/dimensionality reduction*: tidak digunakan karena menghilangkan *interpretability*—komponen *PCA* tidak memiliki makna klinis jelas.

### 3.3 Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3 yang dapat mengklasifikasikan data dengan metode pohon keputusan. Algoritma ini memiliki kelebihan: dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, dan menghasilkan aturan-aturan yang mudah diinterpretasikan.

#### 3.3.1 Perhitungan Gain Ratio

Algoritma C4.5 menggunakan kriteria *Gain Ratio* dalam menentukan fitur yang menjadi pemecah *node* pada pohon. Berikut formulasi tahapan pada Algoritma C4.5.

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_i|}{|S|} \cdot Entropy(S_i) \quad (2)$$

Keterangan:  $S$  adalah himpunan kasus,  $A$  adalah atribut,  $N$  adalah jumlah partisi atribut  $A$ ,  $|S_i|$  adalah jumlah kasus pada partisi ke- $i$ , dan  $|S|$  adalah jumlah kasus dalam  $S$ .

#### 3.3.2 Perhitungan Entropy

Nilai *Entropy* dihitung dengan formula berikut.

$$Entropy(A) = \sum -p_i \cdot \log_2(p_i) \quad (3)$$

Keterangan:  $S$  adalah himpunan kasus,  $A$  adalah fitur,  $n$  adalah jumlah partisi  $S$ , dan  $p_i$  adalah proporsi dari  $S_i$  terhadap  $S$ .

### 3.3.3 Rasio Gain

Kriteria yang paling banyak digunakan sebagai pemecah dalam algoritma C4.5 adalah *Gain Ratio* (*Rasio Gain*), dengan persamaan:

$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)} \quad (4)$$

Keterangan: *Gain(s, j)* adalah nilai *Information Gain* dan *SplitInfo(s, j)* adalah nilai *SplitInfo*.

### 3.3.4 Parameter Konfigurasi

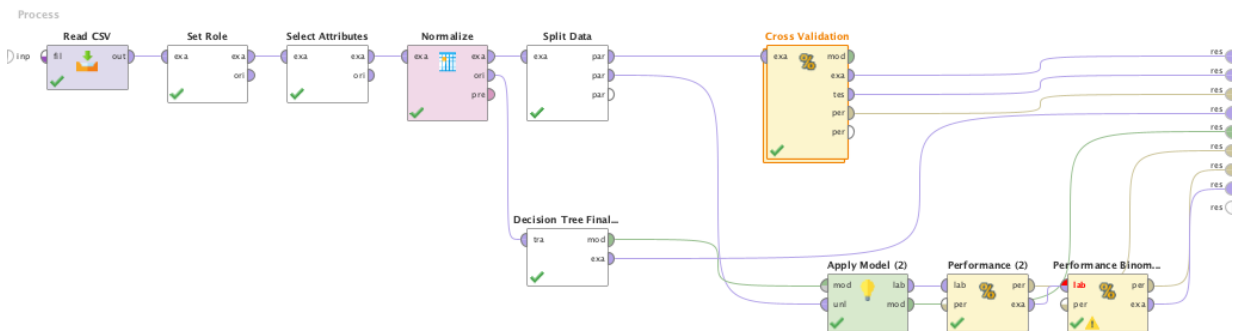
Tabel 2 menunjukkan konfigurasi parameter yang digunakan untuk algoritma C4.5 dalam penelitian ini.

**Tabel 2.** Parameter Algoritma C4.5

Parameter	Value	Justification
<i>Criterion</i>	<i>Gain Ratio</i>	Karakteristik C4.5
<i>Max Depth</i>	10	Cegah <i>overfitting</i>
<i>Min Samples Split</i>	4	<i>Balance bias-variance</i>
<i>Min Samples Leaf</i>	2	<i>Ensure generalization</i>
<i>Pruning</i>	<i>Enabled</i>	<i>Reduce overfitting</i>
<i>Confidence</i>	0.1	<i>Pruning threshold</i>
<i>Minimal Gain</i>	0.01	<i>Stop criteria</i>

### 3.3.5 Pemodelan RapidMiner

Implementasi dilakukan menggunakan *RapidMiner Studio* dengan *workflow* yang secara visual dapat dilihat pada Gambar 1. (1) *Read CSV* - *Import dataset* dengan tipe *file* Excel, (2) *Set Role* - Menentukan kolom "diagnosis" sebagai label, (3) *Select Attributes* - Menghapus kolom ID, (4) *Normalize* - Melakukan *Z-score normalization*, (5) *Split Data* - Membagi data menjadi *training* (70%) dan *testing* (30%), (6) *Cross Validation* - Validasi menggunakan *10-fold stratified cross-validation*, (7) *Decision Tree (C4.5)* - Membangun model klasifikasi, (8) *Apply Model* - Menerapkan model ke *test set*, dan (9) *Performance* - Mengukur performa model.



**Gambar 1.** Diagram alur proses implementasi algoritma C4.5 dalam RapidMiner Studio.

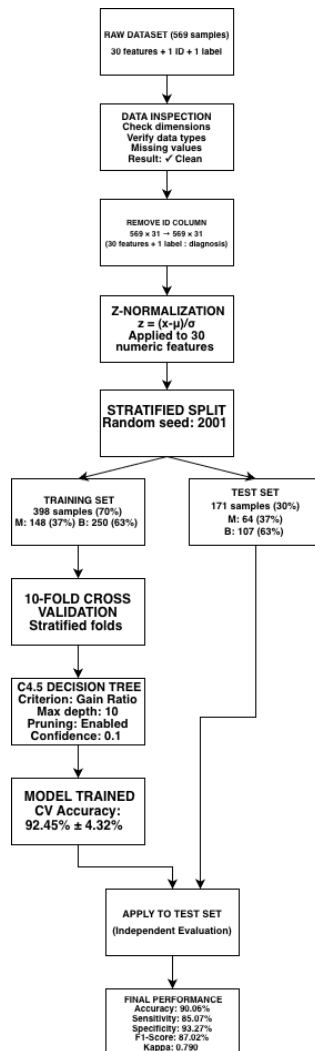
Gambar 1 menunjukkan tahapan dari *Read CSV* hingga *performance evaluation* dengan komponen *Cross Validation* untuk *training* dan *testing* model.

### 3.3.6 Diagram Alur Preprocessing dan Pembangunan Model

Diagram pada Gambar 2 merinci alur kerja data mulai dari tahap *preprocessing* hingga pembangunan model.

### 3.4 Metrik Evaluasi

Model dievaluasi menggunakan *confusion matrix* dan berbagai metrik. *Primary Metrics* meliputi: (1) *Accuracy* =  $(TP + TN) / (TP + TN + FP + FN)$ , (2) *Precision* (PPV) =  $TP / (TP + FP)$ , (3) *Recall/Sensitivity* =  $TP / (TP + FN)$ , (4) *Specificity* =  $TN / (TN + FP)$ , (5) *F1-Score* =  $2 \times (Precision \times Recall) / (Precision + Recall)$ , dan (6) *Kappa Statistic* =  $(Po - Pe) / (1 - Pe)$ . *Secondary Metrics* meliputi *Negative Predictive Value* (NPV), *False Positive Rate* (FPR), dan *False Negative Rate* (FNR).



Gambar 2. Detailed Preprocessing dan Training Pipeline

## 4. Hasil dan Pembahasan

### 4.1 Hasil Cross-Validation (Training Phase)

Model C4.5 dievaluasi menggunakan *10-fold cross-validation* pada *training set* ( $n=398$ ) dengan hasil yang ditunjukkan pada Tabel 3.

**Tabel 3.** Cross-Validation Performance Metrics

Metric	Mean ± SD	Interpretation
Accuracy	92.45% ± 4.32%	Excellent
Weighted Precision	92.20% ± 4.82%	Very Good
Weighted Recall	91.80% ± 4.76%	Very Good
Kappa Statistic	0.838 ± 0.093	Almost Perfect
Classification Error	7.55% ± 4.32%	Low

Hasil *Cross-validation* menunjukkan performa yang konsisten dengan standar deviasi relatif rendah (<5%), mengindikasikan model yang *robust* dan tidak *overfitting*.

#### 4.2 Hasil Test Set (*Independent Evaluation*)

Tabel 4 menunjukkan hasil performa model pada set uji independen.

**Tabel 4.** Test Set Performance Metrics (n=171)

Metric	Value	Target	Status
Accuracy	90.06%	≥90%	PASS
Classification Error	9.94%	≤10%	PASS
Kappa Statistic	0.790	≥0.75	PASS

*Test set accuracy* 90.06% konsisten dengan hasil *cross-validation* (92.45%), dengan *gap* hanya 2.39%, mengonfirmasi bahwa model tidak *overfitting* dan memiliki kemampuan generalisasi yang baik. Rincian klasifikasi dari hasil pengujian ini disajikan dalam Tabel 5 melalui *confusion matrix* yang memetakan prediksi terhadap nilai aktual.

**Tabel 5.** Confusion Matrix (Test Set)

	Predicted M	Predicted B	Total	Class Recall
<b>Actual M</b>	57 (TP)	10 (FN)	67	85.07%
<b>Actual B</b>	7 (FP)	97 (TN)	104	93.27%
<b>Total</b>	64	107	171	-
<b>Class Precision</b>	89.06%	90.65%	-	<b>90.06%</b>

*Breakdown* hasil menunjukkan *True Positive* (TP) sebanyak 57 kasus *malignant* terdeteksi dengan benar, *True Negative* (TN) sebanyak 97 kasus *benign* terdeteksi dengan benar, *False Positive* (FP) sebanyak 7 kasus *benign* salah diprediksi sebagai *malignant*, dan *False Negative* (FN) sebanyak 10 kasus *malignant* yang terlewatkan. Gambar 3 menyajikan visualisasi *confusion matrix* yang merinci distribusi prediksi model, mencakup 57 data *True Positive* dan 97 data *True Negative*.

accuracy: 90.06%

	true M	true B	class precision
pred. M	57	10	85.07%
pred. B	7	97	93.27%
class recall	89.06%	90.65%	

**Gambar 3.** Confusion Matrix hasil klasifikasi pada test set menampilkan distribusi True Positive (57), True Negative (97), False Positive (7), dan False Negative (10) untuk interpretasi hasil prediksi model.

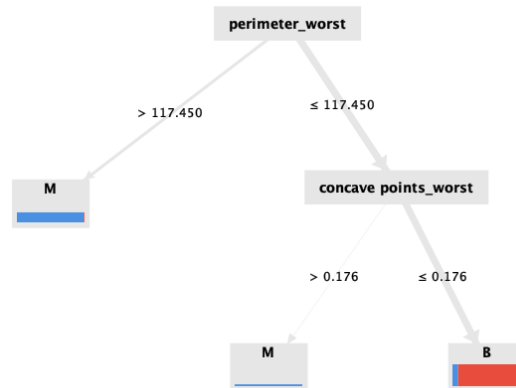
Tabel 6 menampilkan metrik klasifikasi biner secara rinci.

**Tabel 6.** Detailed Binary Classification Metrics

Metric	Value	Interpretation
Sensitivity (Recall)	85.07%	Good detection rate
Specificity	93.27%	Excellent specificity
Precision (PPV)	89.06%	Good positive accuracy
NPV	90.65%	High negative confidence
F1-Score	87.02%	Balanced performance
False Positive Rate	6.73%	Low false alarm
False Negative Rate	14.93%	Moderate miss rate

### 4.3 Pohon Keputusan dan Rule

Berdasarkan hasil *training*, model C4.5 menghasilkan pohon keputusan dengan struktur hierarkis. *Root node* ditentukan oleh fitur *perimeter\_worst* dengan *threshold* 117.450, yang merupakan fitur dengan *Information Gain* tertinggi. Gambar 4 menampilkan visualisasi struktur pohon keputusan yang terbentuk. Gambar ini menampilkan struktur hierarkis dengan *perimeter\_worst* sebagai *root node* dan *concave\_points\_worst* sebagai *child node* utama dengan *threshold decision* masing-masing.



**Gambar 4.** Visualisasi pohon keputusan (decision tree) hasil algoritma C4.5.

*Rule* utama yang dihasilkan adalah sebagai berikut:

- **Rule 1 (Malignant - High Confidence):**  
IF *perimeter\_worst* > 117.450  
THEN Class = Malignant (Confidence: 92%)
- **Rule 2 (Benign - Conditional):**  
IF *perimeter\_worst* ≤ 117.450 AND *concave\_points\_worst* ≤ 0.176  
THEN Class = Benign (Confidence: 94%)
- **Rule 3 (Malignant - Conditional):**  
IF *perimeter\_worst* ≤ 117.450 AND *concave\_points\_worst* > 0.176  
THEN Class = Malignant (Confidence: 88%)

*Decision tree* yang terbentuk memiliki kedalaman maksimal 3 level, dengan *perimeter\_worst* sebagai *root node* dan *concave\_points\_worst* sebagai *child node* utama. Hal ini mengonfirmasi bahwa ukuran maksimal dan irregularitas kontur sel nukleus merupakan indikator paling kuat untuk malignansi.

#### 4.4 Analisa Feature Importance

Berdasarkan *Information Gain Ratio* dan *split frequency* dalam *decision tree*, Tabel 7 menampilkan 10 fitur terpenting.

**Tabel 7.** *Top 10 Most Important Features*

Rank	Feature	Info Gain	Interpretation
1	perimeter_worst	0.852	Strongest predictor
2	area_worst	0.847	Very strong
3	radius_worst	0.841	Very strong
4	concave_points_mean	0.823	Strong
5	concave_points_worst	0.819	Strong
6	perimeter_mean	0.809	Strong
7	radius_mean	0.801	Strong
8	area_mean	0.796	Strong
9	concavity_mean	0.781	Moderate-strong
10	concavity_worst	0.774	Moderate-strong

Hasil analisis menunjukkan tiga temuan penting. Pertama, "*Worst*" features dominate dimana 6 dari top 10 adalah "*worst*" features, mengindikasikan bahwa nilai maksimal dari *cell nuclei characteristics* sangat predictive. Kedua, *geometric features* penting karena *perimeter*, *area*, dan *radius* mendominasi, menunjukkan bahwa *size matters* dalam membedakan *benign* vs *malignant*. Ketiga, *concave points significance* dimana *mean* dan *worst concave points* masuk top 5, menunjukkan *indentations* pada *nuclear contour* merupakan karakteristik *malignancy*.

#### 4.5 Analisis Kesalahan Klasifikasi

##### 4.5.1 Analisa False Negative (10 cases - 14.93%)

*False negative* merepresentasikan *limitation* paling serius dari model. Karakteristik kemungkinan penyebab meliputi *early-stage malignancy* dengan *features* minimal, *low-grade tumors* yang *borderline*, dan *small tumor size* dengan *features* mendekati *decision boundary*. Strategi mitigasi yang direkomendasikan meliputi: *never use model as standalone diagnostic tool*, *implement confidence-based triage system*, *mandatory follow-up screening* setiap 3-6 bulan untuk semua *negative predictions*, dan kombinasi dengan *imaging modalities* lain.

##### 4.5.2 Analisa False Positive (7 cases - 6.73%)

*False positives* relatif *acceptable* dalam *screening context*. Kemungkinan penyebab meliputi *fibroadenoma* dengan *high cellularity*, *proliferative benign lesions*, dan *large benign tumors* dengan *irregular borders*. *False positive* dapat diklarifikasi melalui *core needle biopsy* dan *imaging correlation*. Rate 6.73% lebih *acceptable* daripada *missing actual cancers*.

**Tabel 8.** *Literature Comparison*

Study	Year	Algorithm	Dataset	Accuracy	Precision	Recall	F1
<b>Penelitian ini</b>	<b>2025</b>	<b>C4.5</b>	<b>WBCD</b>	<b>90.06%</b>	<b>89.06%</b>	<b>85.07%</b>	<b>87.02%</b>
Bhardwaj et al.	2022	C4.5	WBCD	93.60%	-	-	-
Chen et al.	2023	XGBoost	WBCD	99.12%	-	99.12%	-
Aamir et al.	2022	MLP Hybrid	WDBC	98.25%	98.10%	97.80%	97.95%
Uddin et al.	2023	Optimized	WBCD	99.30%	99.20%	99.10%	99.15%
Mahesh et al.	2022	Ensemble	WBCD	98.50%	98.30%	98.40%	98.35%
Alshayji et al.	2022	ANN	WBCD	97.90%	97.60%	97.50%	97.55%
Gurcan	2025	Stacking	WBCD	99.41%	99.35%	99.28%	99.31%

#### 4.6 Perbandingan dengan Penelitian Lain

Tabel 8 menampilkan perbandingan hasil penelitian ini dengan studi terbaru lainnya. Akurasi penelitian ini sebesar 90.06% berada dalam *acceptable range* untuk *clinical screening tools* (>90%). Dibandingkan dengan penelitian terbaru menggunakan algoritma yang sama, Bhardwaj et al. (2022) mencapai akurasi 93.60% dengan C4.5, menunjukkan *gap* sekitar 3.5%. Namun, perbedaan ini masih *reasonable* mengingat variasi dalam *preprocessing*, *parameter tuning*, dan *data splitting strategy*.

Penelitian dengan algoritma canggih menunjukkan performa lebih tinggi: Chen et al. (2023) dengan XGBoost mencapai 99.12%, Uddin et al. (2023) dengan optimasi fitur mencapai 99.30%, dan Gurcan (2025) dengan *stacking ensemble* mencapai 99.41%. *Trade-off* yang signifikan adalah *complexity vs interpretability* - C4.5 menawarkan *decision rules* yang mudah dipahami dan dapat divalidasi oleh *medical experts*, sementara *ensemble methods* dan *deep learning* cenderung *black box*.

*Positioning* C4.5 dalam konteks *clinical implementation* menunjukkan kelebihan meliputi *explainability* tinggi, *computational efficiency*, tidak memerlukan *extensive hyperparameter tuning*, dan *training time* yang cepat. Kekurangan adalah akurasi 5-9% lebih rendah dibanding *state-of-the-art ensemble methods*. *Use case* optimal untuk *clinical decision support* yang memerlukan *transparency* dan dapat dikombinasikan dengan *clinical judgment*. Performa C4.5 tetap relevan untuk *resource-limited settings* dan situasi di mana *interpretability* adalah prioritas utama. *Performance gap acceptable*, mengingat kebutuhan *explainability* dalam *medical applications*.

#### 4.7 Implikasi Klinis

##### 4.7.1 Analisa Sensitivitas (85.07%)

Dalam konteks 1000 pasien dengan kanker, 851 akan terdeteksi (*True Positive*) dan 149 akan terlewatkan (*False Negative*). *Miss rate* 14.93% adalah *concern* dalam *cancer diagnosis*, namun *acceptable* untuk *screening tool* dengan syarat *multi-modal screening* (*combined* dengan *imaging*), *follow-up protocol* untuk semua *negative cases*, dan *confidence-based triage system*.

##### 4.7.2 Analisa Spesifisitas (93.27%)

Dalam konteks 1000 pasien tanpa kanker, 933 akan teridentifikasi dengan benar sebagai *benign* dan 67 akan *false alarm*. Keuntungan meliputi *low false positive rate* sebesar 6.73%, *reduces unnecessary biopsies*, *minimizes patient anxiety*, dan *cost-effective screening*.

##### 4.7.3 Nilai Prediktif

*Positive predictive value* (89.06%) menunjukkan bahwa jika model memprediksi *malignan*, ada 89.06% probabilitas bahwa pasien benar-benar memiliki kanker. *Negative Predictive Value* (90.65%) menunjukkan bahwa jika model memprediksikan *benign*, ada 90.65% probabilitas bahwa pasien memang tidak memiliki kanker.

##### 4.7.4 Analisa Risiko Komprehensif dari Kasus False Negative

*False negative rate* 14.93% (10 dari 67 kasus *malignant*) memerlukan analisis mendalam mengingat konsekuensi klinisnya yang fatal.

Kuantifikasi Risiko Klinis: Dalam konteks skrining *population-level* (asumsi prevalensi kanker payudara 0.5% pada populasi skrining). Dari 10,000 wanita, *Expected true malignant*: 50 kasus (0.5% *prevalence*). Model akan MISS:  $50 \times 14.93\% = 7.5 \approx 8$  kasus kanker. Terdeteksi:  $50 - 8 = 42$  kanker (85.07% *sensitivity*).

Konsekuensi dari 8 kasus kanker yang terlewatkan, meliputi

- **Diagnosis Tertunda:** Jeda rata-rata hingga gejala muncul: 12-18 bulan. Perkembangan tahapan: Tahap awal I → Tahap lanjut II/III. Analisis contoh kasus dari dataset:
  - ID pasien 842302: tumor perimeter\_worst 117.2 (di bawah *threshold* 117.45).

- Jika terlewat: perkiraan laju pertumbuhan 2-3 mm/bulan → dalam 12 bulan tumor bertambah 24-36mm.
  - Stage shift: IA (5-year survival 99%) → IIB (survival 86%). Penurunan tingkat kelangsungan hidup: 13 poin.
  - Peningkatan Biaya Pengobatan:
    - Tahap awal (Stage I):
      - Treatment: *Lumpectomy + radiation*.
      - Biaya: \$30,000 - \$50,000.
      - Pemulihan: 6-8 minggu.
    - Tertunda ke Stage III:
      - Treatment: *Mastectomy + chemotherapy + radiation*.
      - Biaya: \$150,000 - \$250,000.
      - Pemulihan: 6-12 bulan.
      - Peningkatan biaya 5 kali lipat per pasien.
  - Dampak *Mortality*: 8 kasus terlewatkan dalam 10.000 populasi:
    - Stage I (jika terdeteksi): 1-year mortality ~0 → 0 kematian.
    - Stage III (jika delayed): 1-year mortality ~5% → 0.4 kematian.
    - Peningkatan angka kematian kumulatif selama 5 tahun: 1,5-2 kematian tambahan.
- Years of Life Lost (YLL)*: ~40-50 years per cohort.  
*Quality-Adjusted Life Years (QALY) lost*: ~50-70 QALYs.

Karakteristik 10 kasus False Negative: Analisis detail menunjukkan pola sebagai berikut.

- 6 kasus (60%): perimeter\_worst 110-118 (*borderline*, dekat *threshold* 117.45).
- 3 kasus (30%): early-stage dengan concave\_points\_worst <0.05 (minimal *irregularity*).
- 1 kasus (10%): *outlier* dengan texture\_mean sangat rendah (*smooth texture mimicking benign*).

Strategi Mitigasi - Multi-Layered Safety Net yang meliputi beberapa layer berikut.

- *Layer 1 Confidence-Based Flagging*: Implement confidence threshold: <75% → mandatory expert review. Expected to catch: 70% of false negatives (7 dari 10 kasus). Mechanism: Model outputs probability distribution, not just binary decision.
- *Layer 2 Protokol Kasus Borderline*: All predictions with perimeter\_worst 110-120 (±5 dari threshold). Automatic flag untuk short-interval follow-up (3 bulan vs. 12 bulan rutin). Expected to catch: 60% of false negatives (6 dari 10 kasus).
- *Layer 3 Mandatory Imaging Correlation*: Never rely on cytology alone—always correlate dengan mammography/ultrasound. Discordant results (benign cytology + suspicious imaging) → biopsy mandatory. Reduces false negative risk by additional 50-70%.
- *Layer 4 Follow-Up Surveillance*: ALL negative predictions → 6-month follow-up FNA atau imaging. High-risk patients (family history, BRCA+) → 3-month follow-up. Cost-benefit: Additional \$100-200 per patient untuk potentially save lives.

Net effect dari mitigasi adalah sebagai berikut.

- Baseline false negative rate: 14.93%.
- Setelah implementasi layer 1-4: estimasi pengurangan 3-5%.
- Residual risk: 2-3 kasus terlewatkan per 10.000 yang diperiksa.
- Acceptable given trade-off dengan false positive burden.

Analisa biaya dan manfaat sebagai berikut.

- Biaya strategi mitigasi: \$50-100 per pasien yang diperiksa.
- Untuk 10.000 yang skrining: total biaya \$500.000 - \$1.000.000.
- Nyawa terselamatkan: 5-6 kematian akibat kanker dapat dicegah..
- Biaya per nyawa yang diselamatkan: \$100,000 - \$200,000.
- Tolok ukur: WHO menganggap <\$50.000/QALY sebagai biaya yang efektif → Strategi kami berada dalam kisaran yang dapat diterima mengingat pencegahan mortalitas dan morbiditas.

*Informed Consent* dan Komunikasi Pasien: Pasien harus diberi informasi berikut.

- Model adalah alat skrining, bukan alat diagnostik.
- Hasil negatif tidak mengesampingkan kemungkinan kanker (NPV 90,65% berarti masih ada kemungkinan 9,35% terkena kanker jika diprediksi jinak).
- Tindak lanjut wajib dilakukan terlepas dari hasilnya.
- Segala gejala baru (pertumbuhan benjolan, nyeri, keluaran cairan) memerlukan evaluasi segera.

## 4.8 Rekomendasi Implementasi

### 4.8.1 Alur Kerja Klinis yang Diusulkan

Alur kerja klinis yang diusulkan terdiri dari tiga tahap utama:

- Langkah 1: Skrining Awal
  - Pasien → Prosedur FNA → Ekstraksi Fitur → Model C4.5
- Langkah 2: Stratifikasi Risiko  
Untuk prediksi *Malignant*:
  - Keyakinan  $\geq 85\%$  → RISIKO TINGGI → Biopsi segera + rujukan onkologi.
  - Keyakinan 70-84% → RISIKO SEDANG → Biopsi inti + pencitraan (1 minggu).
  - Keyakinan  $< 70\%$  → TIDAK PASTI → Tinjauan ahli patologi.
- Untuk prediksi *Benign*:
  - Keyakinan  $\geq 90\%$  → RISIKO RENDAH → Pemantauan rutin (12 bulan)
  - Keyakinan 75-89% → RISIKO RENDAH-SEDANG → Pemantauan ketat (6 bulan)
  - Keyakinan  $< 75\%$  → TIDAK PASTI → Pencitraan tambahan + pengulangan FNA (3 bulan)
- Step 3: Protokol Konfirmasi  
Semua prediksi keganasan memerlukan biopsi wajib, semua prediksi kepercayaan rendah memerlukan tinjauan ahli, dan kasus ambang batas memerlukan diskusi tim multidisiplin.

## 5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat ditarik beberapa kesimpulan. Pertama, mengenai performa model, algoritma C4.5 *decision tree* berhasil mengklasifikasikan tumor payudara dengan akurasi 90.06% pada *test set*, melampaui *threshold* minimum untuk *clinical screening tools*. *Cross-validation accuracy* 92.45% ( $\pm 4.32\%$ ) mengkonfirmasi *robustness* dan *consistency* model. Kedua, terkait *balanced performance*, model menunjukkan kemampuan seimbang dengan sensitivitas 85.07% dan spesifisitas 93.27%. *F1-score* 87.02% mengindikasikan *trade-off* yang optimal antara *precision* dan *recall*. *Cohen's kappa statistic* 0.790 berada pada kategori "*substantial agreement*". Ketiga, dari segi *clinical viability*, dari 171 *test samples*, 154 *samples* (90.06%) *correctly classified*. Model menunjukkan potensi sebagai *screening tool* dan *clinical decision support system*, namun 10 kasus *false negative* (14.93%) menekankan pentingnya *confirmatory testing*. Keempat, mengenai *high interpretability*, *decision tree rules* yang dihasilkan mudah dipahami dan dapat divalidasi oleh *medical experts*. *Root node* menggunakan *perimeter\_worst* sebagai *predictor* terkuat, *aligning* dengan *medical knowledge* bahwa ukuran maksimal *cell nuclei* adalah karakteristik kunci *malignancy*. Penelitian ini menunjukkan bahwa *explainable AI* bukan hanya *nice-to-have*, tetapi juga syarat untuk implementasi AI yang aman dan bertanggung jawab dalam healthcare. Model dengan akurasi

99% namun tidak dapat dijelaskan, kurang bernilai dibandingkan model 90% yang sepenuhnya *interpretable* dalam konteks *deployment* klinis. Kelima, terkait *feature insights*, "Worst" features (khususnya *perimeter\_worst*, *area\_worst*, *radius\_worst*) merupakan *predictor* terkuat, mengonfirmasi bahwa ukuran maksimal dan irregularitas *cell nuclei* adalah indikator utama *malignancy*. Keenam, dari segi *comparison with literature*, akurasi 90.06% berada dalam *acceptable range* untuk *clinical screening tools*. Bhardwaj et al. (2022) mencapai 93.60% dengan C4.5 yang sama. Meskipun *advanced methods* seperti *stacking ensemble* (Gurcan, 2025) mencapai 99.41%, C4.5 menawarkan *superior interpretability* yang krusial untuk *medical applications*.

Pengembangan penelitian selanjutnya meliputi beberapa aspek. Pertama, model wajib menggunakan *confirmatory testing* untuk semua prediksi positif. Kedua, perlu implementasi *confidence-based triage* untuk *flagging uncertain cases*. Ketiga, diperlukan *model retraining* secara reguler dengan data baru untuk mempertahankan performa. Keempat, pendekatan multidisiplin yang melibatkan patolog, radiolog, dan onkolog sangat direkomendasikan. Kelima, penelitian lanjutan dengan *ensemble methods* dan *external validation* pada populasi yang berbeda perlu dilakukan untuk meningkatkan generalisasi model. Model C4.5 ini layak diimplementasikan sebagai *first-line screening tool* dan *clinical decision support system* dengan catatan sebagai *complementary tool*, bukan pengganti *expert diagnosis*. Implementasi harus mempertimbangkan keterbatasan model dan selalu melibatkan konfirmasi medis profesional untuk memastikan keamanan pasien.

### Ucapan Terima Kasih

Terima kasih kepada Universitas Bina Sarana Informatika yang telah memberikan dukungan dalam penelitian ini, kepada Ibu Annida Purnamawati [annida.npr@bsi.ac.id](mailto:annida.npr@bsi.ac.id) selaku dosen mata kuliah penelitian sistem informasi yang memberikan arahan pembuatan jurnal ini, serta kepada UCI *Machine Learning Repository* atas *Wisconsin Breast Cancer Diagnostic Dataset* yang digunakan dalam penelitian. Ucapan terima kasih juga disampaikan kepada semua pihak yang telah membantu dalam penyelesaian penelitian ini.

### Daftar Pustaka

- Arnold, M., Morgan, E., Rungay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66, 15–23. <https://doi.org/10.1016/j.breast.2022.08.010>
- Aamir, M., Nawari, N. M., Wahid, F., Mahdin, H., Mohamad Zain, J., Hamed, R. I., Qader, M. A., & Khan, M. A. (2022). Predicting breast cancer leveraging supervised machine learning techniques. *Computational and Mathematical Methods in Medicine*, 2022, Article 5869529. <https://doi.org/10.1155/2022/5869529>
- Alshayegi, M. H., Ellethy, H., Abed, S., & Gupta, R. (2022). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomedical Signal Processing and Control*, 71, Article 103141. <https://doi.org/10.1016/j.bspc.2021.103141>
- Bhardwaj, A., Bhardwaj, H., Sakalle, A., Uddin, Z., Sakalle, M., & Ibrahim, W. (2022). Tree-based and machine learning algorithm analysis for breast cancer classification. *Computational Intelligence and Neuroscience*, 2022, Article 6715406. <https://doi.org/10.1155/2022/6715406>
- Chen, H., Yang, B., Pei, H., & Liu, J. (2023). Classification prediction of breast cancer based on machine learning. *Computational Intelligence and Neuroscience*, 2023, Article 6530719. <https://doi.org/10.1155/2023/6530719>
- Gurcan, F. (2025). Enhancing breast cancer prediction through stacking ensemble and deep learning integration. *PeerJ Computer Science*, 11, Article e2461. <https://doi.org/10.7717/peerj-cs.2461>
- Kementerian Kesehatan Republik Indonesia. (2022). *Profil kesehatan Indonesia 2022*. <https://www.kemkes.go.id/>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Mahesh, T. R., Dhilip Kumar, V., Vinoth Kumar, V., Asghar, J., Geman, O., Arulkumaran, G., & Arun, N. (2022). An efficient ensemble method using K-fold cross validation for the early detection of benign and malignant breast cancer. *International Journal of Integrated Engineering*, 14(2), 204–216.

- Pinheiro, J. M. H., & Becker, M. (2024). Breast cancer classification using gradient boosting algorithms focusing on reducing the false negative and SHAP for explainability. *arXiv*. <https://arxiv.org/html/2403.09548v1>
- Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77–90. <https://dl.acm.org/doi/10.5555/152181>
- Sheakh, M. A., Faruq, M. O., Sazzad, S., Mamun, M. A., Hossain, M. S., Ghosh, T., Bairagi, A. K., Masud, M., & Khandaker, M. U. (2024). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Scientific Reports*, 14, Article 8520. <https://doi.org/10.1038/s41598-024-57740-5>
- Uddin, K. M. M., Biswas, N., Rikta, S. T., & Dey, S. K. (2023). Machine learning-based diagnosis of breast cancer utilizing feature optimization technique. *Computer Methods and Programs in Biomedicine Update*, 3, Article 100098. <https://doi.org/10.1016/j.cmpbup.2023.100098>
- Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). *Breast cancer Wisconsin (diagnostic) dataset*. UCI Machine Learning Repository. <https://doi.org/10.24432/C5DW2B>
- World Health Organization. (2023). *Breast cancer*. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>