

## Pengembangan Sistem Pendukung Keputusan Diagnosis Kanker Payudara Menggunakan Regresi Logistik

Rayendra<sup>1</sup>, Ikhsan<sup>2</sup>, Rozi Meri<sup>3</sup>, Ismael<sup>4</sup>, Dian Eka Putra<sup>5</sup>, Rostam Ahmad Efendi<sup>6</sup>, Ryan Ikhbal Salam<sup>7</sup>

<sup>1234</sup>Jurusan Teknologi Informasi, Politeknik Negeri Padang

[1rayendra@pnp.ac.id](mailto:rayendra@pnp.ac.id) , [2ikhsan85@pnp.ac.id](mailto:ikhsan85@pnp.ac.id) , [3rozimeri@pnp.ac.id](mailto:rozimeri@pnp.ac.id) , [4ismael@pnp.ac.id](mailto:ismael@pnp.ac.id) , [5dianekaputra@pnp.ac.id](mailto:dianekaputra@pnp.ac.id) ,  
[6rostamahmadefendi@pnp.ac.id](mailto:rostamahmadefendi@pnp.ac.id) , [7ryan@pnp.ac.id](mailto:ryan@pnp.ac.id)

### Abstract

*Breast cancer is one of the leading causes of women death in Indonesia. Accurate and rapid diagnosis is crucial to improve patient recovery rates. This study aims to develop a decision support system for breast cancer diagnosis using an optimized logistic regression model. The research method uses the Wisconsin Diagnostic Breast Cancer dataset consisting of 569 samples with 30 cell morphological features. The data is divided into 75% training data and 25% testing data. The logistic regression model is implemented with L2 regularization and evaluated by using ROC analysis, threshold optimization with the Youden Index and probability calibration using Expected Calibration Error (ECE). The results show the model achieves 95.8% accuracy, 96.2% sensitivity and 95.6% specificity with an AUC-ROC value of 0.993. Threshold optimization produces an optimal value of 0.560 that balances sensitivity and specificity. Probability calibration shows an ECE of 0.0390 which can be improved to 0.0328 using isotonic regression.*

**Keywords:** breast cancer, decision support systems, logistic regression, probability calibration, threshold optimization

### Abstrak

Kanker payudara merupakan salah satu penyebab kematian tertinggi pada wanita di Indonesia. Diagnosis yang akurat dan cepat sangat penting untuk meningkatkan tingkat kesembuhan pasien. Penelitian ini bertujuan mengembangkan sistem pendukung keputusan untuk diagnosis kanker payudara menggunakan model regresi logistik yang dioptimasi. Metode penelitian menggunakan dataset Wisconsin Diagnostic Breast Cancer yang terdiri dari 569 sampel dengan 30 fitur morfologis sel. Data dibagi menjadi 75% data latih dan 25% data uji. Model regresi logistik diimplementasikan dengan regularisasi L2 dan dievaluasi menggunakan analisis ROC, optimasi ambang batas dengan Indeks Youden, serta kalibrasi probabilitas menggunakan Expected Calibration Error (ECE). Hasil penelitian menunjukkan model mencapai akurasi 95,8%, sensitivitas 96,2%, dan spesifisitas 95,6% dengan nilai AUC-ROC 0,993. Optimasi ambang batas menghasilkan nilai optimal 0,560 yang menyeimbangkan sensitivitas dan spesifisitas. Kalibrasi probabilitas menunjukkan ECE sebesar 0,0390 yang dapat ditingkatkan menjadi 0,0328 menggunakan regresi isotonik.

**Kata kunci:** kanker payudara, sistem pendukung keputusan, regresi logistik, kalibrasi probabilitas, optimasi ambang batas

© 2025 Author

Creative Commons Attribution 4.0 International License



## 1. Pendahuluan

Kanker payudara merupakan masalah kesehatan serius yang dihadapi Indonesia dan dunia. Data Globocan 2020 menunjukkan kanker payudara sebagai jenis kanker dengan kasus baru terbanyak di Indonesia, mencapai 65.858 kasus atau 16,6% dari total kasus kanker[1]. Angka kematian akibat kanker payudara di Indonesia juga tinggi, mencapai 22.430 kasus pada tahun yang sama. Tingginya angka kematian ini salah satunya disebabkan oleh keterlambatan diagnosis, dimana sebagian besar pasien datang dalam stadium lanjut (Kementerian Kesehatan RI, 2022).

Diagnosis kanker payudara saat ini masih sangat bergantung pada interpretasi ahli patologi terhadap hasil pemeriksaan sitologi atau histopatologi. Proses interpretasi ini bersifat subjektif dan dapat menghasilkan variasi diagnosis antar pemeriksa yang berbeda. Penelitian menunjukkan tingkat variabilitas antar-pemeriksa dapat mencapai 20-30% terutama pada kasus-kasus yang sulit atau ambigu[2]. Kondisi ini diperparah dengan terbatasnya jumlah ahli patologi di Indonesia, terutama di daerah-daerah terpencil[3].

Perkembangan teknologi pembelajaran mesin membuka peluang untuk membantu proses diagnosis menjadi lebih objektif dan konsisten. Penelitian yang dilakukan Windarto et al (2023) dapat memandu penelitian di masa mendatang di bidang pembelajaran mesin dan membantu meningkatkan ketepatan diagnosis dan perencanaan pengobatan kanker payudara[4]. Berbagai penelitian telah menunjukkan bahwa algoritma pembelajaran mesin dapat mencapai akurasi yang tinggi dalam klasifikasi kanker payudara. Agarap (2018) melaporkan akurasi 99,04% menggunakan *neural network* pada *dataset* Wisconsin[5]. Penelitian serupa oleh Amrane et al. (2018) menggunakan Naive Bayes dan k-Nearest Neighbor mencapai akurasi masing-masing 96,19% dan 95,57%[6]. Di Indonesia, Tinaliah dan Triana (2024) memprediksi jenis kanker payudara menggunakan *Support Vector Machine* dengan akurasi 98,25%[7]. Pada Tahun 2022 Alshayeji et al (2022) melalui penelitiannya membandingkan metode Artificial Neural Network (ANN) yang menunjukkan kinerja yang menjanjikan dengan akurasi rata-rata 99,85%, untuk deteksi kanker payudara menggunakan the Wisconsin breast cancer dataset, akurasi rata-rata yang dicapai adalah 99,47%[8]

Potensi AI (*Artificial Intelligence*) juga menjadi menarik dalam skrining kanker payudara. Tinjauan ini membutuhkan lebih banyak penelitian dan validasi untuk sepenuhnya memanfaatkan kekuatan AI dalam praktik klinis, memastikan bahwa teknologi ini dapat membantu menyelamatkan nyawa dengan meningkatkan akurasi dan efisiensi diagnosis[9].

Namun demikian, sebagian besar penelitian sebelumnya hanya fokus pada pencapaian akurasi klasifikasi yang tinggi tanpa mempertimbangkan aspek-aspek penting untuk implementasi klinis. Tiga aspek kritis yang sering diabaikan adalah: (1) optimasi ambang batas keputusan yang sesuai dengan konteks klinis, (2) kalibrasi probabilitas untuk memastikan estimasi risiko yang dapat dipercaya, dan (3) interpretabilitas model untuk membangun kepercayaan praktisi medis[10],[11].

Optimasi ambang batas keputusan sangat penting karena dampak kesalahan diagnosis positif palsu (*false positive*) dan negatif palsu (*false negative*) berbeda dalam konteks medis. Kesalahan negatif palsu dimana pasien kanker didiagnosis sehat memiliki konsekuensi yang lebih fatal dibandingkan positif palsu. Oleh karena itu, pemilihan ambang batas harus mempertimbangkan *trade-off* antara sensitivitas dan spesifisitas sesuai dengan prioritas klinis[12].

Kalibrasi probabilitas memastikan bahwa output model dapat diinterpretasikan sebagai estimasi risiko yang bermakna. Model dengan kalibrasi yang baik akan menghasilkan probabilitas prediksi yang sesuai dengan frekuensi kejadian aktual. Hal ini penting untuk komunikasi risiko kepada pasien dan pengambilan keputusan klinis yang melibatkan pertimbangan probabilistik[13].

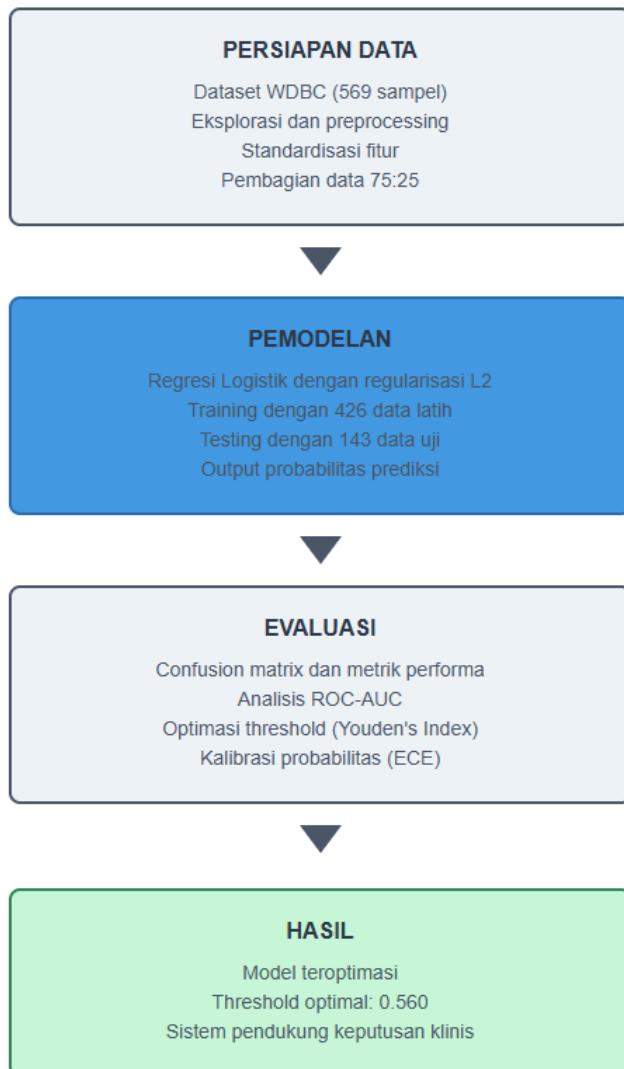
Interpretabilitas model menjadi faktor krusial untuk adopsi teknologi dalam praktik medis. Model yang terlalu kompleks dan tidak dapat dijelaskan akan sulit diterima oleh praktisi medis yang memerlukan pemahaman tentang dasar pengambilan keputusan diagnostik. Regresi logistik menawarkan keseimbangan antara kinerja prediktif dan interpretabilitas yang baik untuk aplikasi medis[14].

Berdasarkan identifikasi kesenjangan tersebut, penelitian ini bertujuan mengembangkan sistem pendukung keputusan diagnosis kanker payudara menggunakan regresi logistik dengan pendekatan evaluasi yang komprehensif. Penelitian tidak hanya fokus pada akurasi klasifikasi, tetapi juga melakukan optimasi ambang batas keputusan, penilaian kalibrasi probabilitas, dan memastikan interpretabilitas model untuk implementasi klinis praktis.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental untuk mengembangkan dan mengevaluasi model regresi logistik sebagai sistem pendukung keputusan diagnosis kanker payudara. Tahapan penelitian mengikuti kerangka

kerja sistematis yang terdiri dari tiga fase utama yaitu persiapan data, pemodelan, serta evaluasi dan optimasi seperti yang digambarkan pada Gambar 1.



**Gambar 1.** Alur Penelitian

*Dataset* yang digunakan dalam penelitian ini adalah Wisconsin Diagnostic Breast Cancer (WDBC) yang diperoleh dari UCI Machine Learning Repository. *Dataset* ini dipilih karena telah menjadi *benchmark* standar dalam penelitian klasifikasi kanker payudara dan memiliki *ground truth* yang valid berdasarkan pemeriksaan histopatologi. *Dataset* terdiri dari 569 sampel hasil aspirasi jarum halus massa payudara dengan distribusi 357 kasus jinak (62,7%) dan 212 kasus ganas (37,3%). Setiap sampel memiliki 30 fitur morfologis kuantitatif yang diekstrak dari citra digital nukleus sel.

**Tabel 1.** Karakteristik *Dataset* dan Fitur Morfologis

Aspek	Keterangan
Total sampel	569 (357 jinak, 212 ganas)
Jumlah fitur	30 fitur morfologis
Kategori fitur	<i>Mean, standard error, dan worst value</i>
Karakteristik dasar	<i>Radius, tekstur, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, fractal dimension</i>
Missing values	Tidak ada

Tahap persiapan data dimulai dengan eksplorasi untuk memahami karakteristik dan distribusi data. Pemeriksaan integritas data menunjukkan tidak ada *missing values* pada seluruh fitur. Label diagnosis yang semula berbentuk kategorikal B untuk benign dan M untuk *malignant* ditransformasi menjadi nilai numerik 0 untuk jinak dan 1 untuk

ganas. Data kemudian dibagi menggunakan *stratified random sampling* dengan proporsi 75% untuk data latih (426 sampel) dan 25% untuk data uji (143 sampel). *Stratified sampling* dipilih untuk mempertahankan proporsi kelas pada kedua *subset* data.

Standardisasi fitur dilakukan menggunakan metode *Z-score transformation* untuk mengatasi perbedaan skala yang signifikan antar-fitur. Sebagai contoh, fitur area memiliki rentang nilai ratusan hingga ribuan, sementara *fractal dimension* hanya dalam rentang desimal. Proses standardisasi menggunakan formula  $z = (x - \mu) / \sigma$ , dimana  $x$  adalah nilai original,  $\mu$  adalah rata-rata, dan  $\sigma$  adalah standar deviasi. Penting untuk dicatat bahwa parameter standardisasi (rata-rata dan standar deviasi) dihitung hanya dari data latih, kemudian diterapkan pada data latih maupun data uji untuk menghindari data *leakage* yang dapat menyebabkan evaluasi yang bias..

Model regresi logistik dipilih sebagai algoritma klasifikasi karena menawarkan keseimbangan optimal antara performa prediktif dan interpretabilitas. Model ini menghitung probabilitas suatu sampel termasuk kelas ganas menggunakan fungsi sigmoid yang mentransformasi kombinasi linear fitur menjadi nilai probabilitas antara 0 dan 1. Implementasi menggunakan *library scikit-learn* dengan konfigurasi solver L-BFGS yang efisien untuk dataset berukuran sedang, regularisasi L2 dengan parameter  $C=1.0$  untuk mencegah *overfitting*, dan maksimal iterasi 100 untuk memastikan konvergensi.

Evaluasi model dilakukan secara komprehensif melalui tiga aspek utama. Pertama, evaluasi performa klasifikasi menggunakan *confusion matrix* untuk menghitung metrik sensitivitas, spesifisitas, presisi, dan akurasi. Sensitivitas mengukur kemampuan model mendeteksi kasus kanker yang sebenarnya, sementara spesifisitas mengukur kemampuan mengidentifikasi kasus jinak dengan benar. Kedua, analisis *Receiver Operating Characteristic* (ROC) dilakukan untuk mengevaluasi kemampuan diskriminasi model secara keseluruhan melalui perhitungan Area Under Curve (AUC). Ketiga, optimasi *threshold* keputusan menggunakan Youden's Index yang didefinisikan sebagai  $J = \text{Sensitivitas} + \text{Spesifisitas} - 1$ , dimana nilai *threshold* yang memaksimalkan  $J$  dipilih sebagai *threshold* optimal.

Aspek kalibrasi probabilitas menjadi fokus khusus dalam penelitian ini karena pentingnya estimasi risiko yang *reliable* untuk pengambilan keputusan klinis. Kalibrasi dievaluasi menggunakan *Expected Calibration Error* (ECE) yang mengukur rata-rata perbedaan absolut antara probabilitas prediksi dengan frekuensi kejadian aktual. Nilai ECE yang rendah menunjukkan bahwa probabilitas *output* model dapat diinterpretasikan langsung sebagai estimasi risiko. Untuk meningkatkan kalibrasi, dua metode dievaluasi yaitu *Platt scaling* yang menggunakan transformasi sigmoid dan *isotonic regression* yang menerapkan transformasi non-parametrik monoton.

**Tabel 2.** Metrik Evaluasi dan Interpretasi Klinis

Metrik	Formula	Interpretasi
Sensitivitas	$TP/(TP+FN)$	Deteksi kasus kanker
Spesifisitas	$TN/(TN+FP)$	Identifikasi kasus jinak
AUC-ROC	<i>Area under ROC curve</i>	Kemampuan diskriminasi keseluruhan
ECE	$\sum  acc(bin) - conf(bin) $	Reliabilitas probabilitas
Youden's Index	$Sens + Spec - 1$	Optimasi <i>threshold</i>

Implementasi penelitian menggunakan bahasa pemrograman Python 3.8 dengan library utama *scikit-learn* untuk pemodelan, NumPy dan Pandas untuk manipulasi data, serta Matplotlib untuk visualisasi. Seluruh eksperimen dilakukan dalam *environment Jupyter Notebook* yang memungkinkan dokumentasi proses secara interaktif. Kode sumber dan data penelitian akan dipublikasikan untuk memastikan reproduksibilitas sesuai prinsip *open science*.

### 3. Hasil dan Pembahasan

Model regresi logistik yang dikembangkan menunjukkan performa klasifikasi yang sangat baik pada data uji independen. Dari 143 sampel data uji, model berhasil mengklasifikasikan 137 sampel dengan benar, menghasilkan akurasi keseluruhan sebesar 95,8%. Analisis *confusion matrix* menunjukkan 86 *true negatives*, 51 *true positives*, 4 *false positives*, dan 2 *false negatives*. Sensitivitas model mencapai 96,2% yang berarti hanya 2 dari 53 kasus kanker yang tidak terdeteksi, sementara spesifisitas sebesar 95,6% menunjukkan kemampuan model mengidentifikasi kasus jinak dengan tepat.

		Prediksi Jinak	Prediksi Ganas	Metrik Performa
Kelas Aktual	Aktual Jinak	<b>86</b> True Negative	<b>4</b> False Positive	Akurasi <b>95.8%</b>
	Aktual Ganas	<b>2</b> False Negative	<b>51</b> True Positive	Sensitivitas <b>96.2%</b>
				Spesifisitas <b>95.6%</b>
				Presisi <b>92.7%</b>
				F1-Score <b>0.944</b>
				False Negative Rate <b>3.77%</b>

Kelas Prediksi

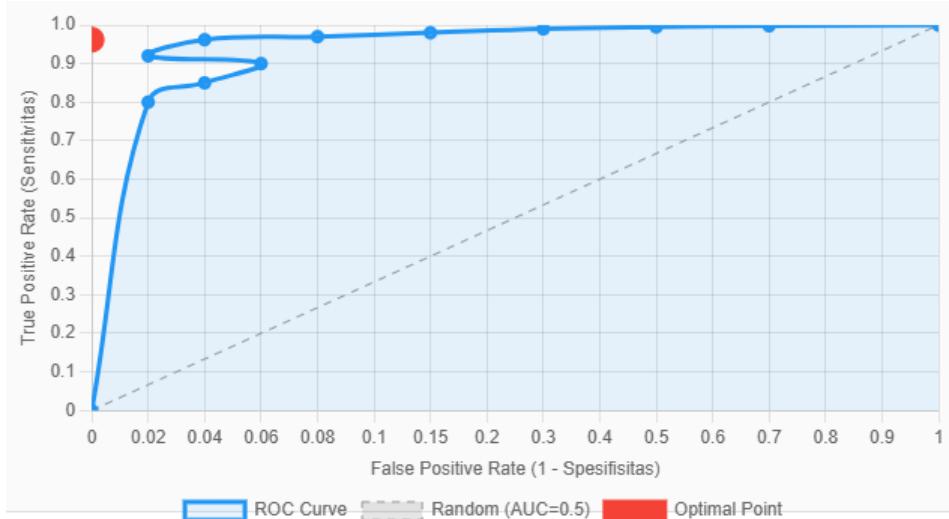
**Gambar 2.** Confusion Matrix dan Metrik Performa

Tingkat *false negative rate* sebesar 3,77% atau hanya 2 dari 53 kasus kanker yang tidak terdeteksi memenuhi standar klinis yang dapat diterima untuk alat bantu skrining kanker. Menurut pedoman American Cancer Society, tingkat *miss rate* di bawah 5% dianggap *acceptable* untuk modalitas skrining kanker payudara[15]. Hasil ini menunjukkan bahwa model memiliki sensitivitas yang memadai untuk aplikasi klinis, meskipun tetap memerlukan validasi lebih lanjut dengan data dari populasi lokal.

**Tabel 3.** Performa Model pada Data Uji

Klasifikasi	Jumlah	Percentase	Interpretasi Klinis
<i>True Negative</i>	86	60,1%	Kasus jinak teridentifikasi dengan benar
<i>True Positive</i>	51	35,7%	Kasus ganas terdeteksi dengan tepat
<i>False Positive</i>	4	2,8%	Kasus jinak salah diagnosis sebagai ganas
<i>False Negative</i>	2	1,4%	Kasus ganas yang terlewat ( <i>miss rate</i> )
Total Benar	137	95,8%	Akurasi keseluruhan model

Analisis *Receiver Operating Characteristic* (ROC) menghasilkan *Area Under Curve* (AUC) sebesar 0,993 yang menunjukkan kemampuan diskriminasi model yang sangat baik. Nilai AUC mendekati 1 mengindikasikan bahwa model dapat membedakan antara kelas jinak dan ganas dengan hampir sempurna. Kurva ROC menunjukkan karakteristik ideal dengan kenaikan cepat menuju sudut kiri atas, menunjukkan bahwa sensitivitas tinggi dapat dicapai dengan *false positive rate* yang minimal.

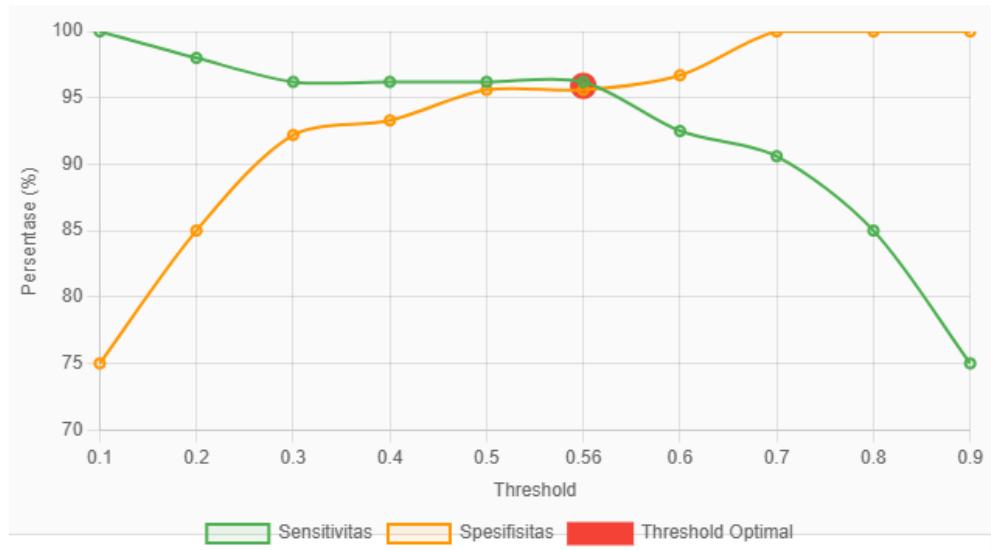


**Gambar 3.** Analisis ROC dan Optimasi Threshold

Optimasi *threshold* menggunakan Youden's Index mengidentifikasi nilai optimal pada 0,560 yang menghasilkan keseimbangan terbaik antara sensitivitas dan spesifisitas. Menariknya, *threshold* optimal ini tidak jauh berbeda dari *threshold default* 0,5 yang umum digunakan, namun perbedaan kecil ini tetap bermakna dalam konteks klinis.

Pada *threshold* optimal, model mempertahankan sensitivitas 96,2% sambil mencapai spesifisitas 95,6%, menghasilkan Youden's Index maksimal sebesar 0,918.

Analisis trade-off pada berbagai nilai *threshold* memberikan wawasan penting untuk implementasi klinis. Threshold yang lebih rendah (0,3-0,4) meningkatkan sensitivitas namun mengorbankan spesifisitas, sesuai untuk program skrining massal dimana prioritas adalah mendeteksi sebanyak mungkin kasus kanker. Sebaliknya, *threshold* yang lebih tinggi (0,7) menghasilkan spesifisitas sempurna 100% namun menurunkan sensitivitas menjadi 90,6%, lebih sesuai untuk diagnosis konfirmatori dimana kepastian diagnosis positif menjadi prioritas.



**Gambar 4.** Trade-off Sensitivitas vs Spesifisitas

Aspek kalibrasi probabilitas menjadi temuan penting dalam penelitian ini. Model original menunjukkan kalibrasi yang baik dengan *Expected Calibration Error* (ECE) sebesar 0,0390, jauh di bawah *threshold* 0,05 yang umumnya dianggap sebagai batas kalibrasi yang *acceptable*. Implementasi teknik kalibrasi menunjukkan bahwa *isotonic regression* memberikan peningkatan terbaik, mengurangi ECE menjadi 0,0328 atau peningkatan sebesar 15,9%. *Platt scaling* memberikan peningkatan yang lebih model dengan ECE 0,0378.

**Tabel 4.** Perbandingan Metode Kalibrasi Probabilitas

Metode Kalibrasi	ECE	Brier Score	Peningkatan ECE	Keterangan
Model Original	0,0390	0,0298	-	Baseline
Platt Scaling	0,0378	0,0296	3,1%	Transformasi sigmoid
Isotonic Regression	0,0328	0,0287	15,9%	Transformasi non-parametrik

Kalibrasi yang baik memiliki implikasi klinis penting karena memungkinkan interpretasi langsung output probabilitas sebagai estimasi risiko. Sebagai contoh, jika model menghasilkan probabilitas 0,7 untuk suatu sampel, maka dapat diinterpretasikan bahwa terdapat risiko 70% sampel tersebut merupakan kasus ganas. Kemampuan ini sangat *valuable* untuk komunikasi risiko kepada pasien dan mendukung *shared decision making* dalam praktik klinis.

**Tabel 5.** Tabel Evaluasi *Threshold*

Threshold	Sensitivitas (%)	Spesifisitas (%)	Youden's Index	False Negative	False Positive	Rekomendasi Klinis
0.300	96.2	92.2	0.884	2	7	Skrining agresif
0.400	96.2	93.3	0.895	2	6	Skrining rutin
0.500	96.2	95.6	0.918	2	4	Default
<b>0.560</b>	<b>96.2</b>	<b>95.6</b>	<b>0.918</b>	<b>2</b>	<b>4</b>	<b>OPTIMAL</b>
0.700	90.6	100.0	0.906	5	0	Diagnosis konfirmatori

Secara umum, sensitivitas model tetap tinggi pada sebagian besar *threshold* yang diuji, yaitu sebesar 96.2%, menunjukkan kemampuan deteksi kanker yang konsisten. Namun, spesifisitas mengalami peningkatan seiring naiknya *threshold*, dari 92.2% pada *threshold* 0.300 hingga mencapai 100% pada *threshold* 0.700. Fenomena ini mencerminkan *trade-off* klasik dalam klasifikasi medis: peningkatan spesifisitas sering kali mengorbankan sensitivitas, dan sebaliknya.

Seperti ditunjukkan pada Tabel 5, *threshold* 0.560 muncul sebagai titik operasi optimal, ditandai dengan nilai Youden's Index tertinggi (0.918), serta jumlah *false negative* dan *false positive* yang minimal (masing-masing 2 dan 4). Menariknya, *threshold* ini tidak hanya diidentifikasi melalui metode Youden, tetapi juga melalui pendekatan minimum *distance to ideal point*, yang keduanya konvergen pada nilai yang sama. Konvergensi ini memperkuat validitas pemilihan *threshold* 0.560 sebagai titik keseimbangan terbaik antara deteksi kanker dan penghindaran alarm palsu.

Dari perspektif klinis, variasi *threshold* memungkinkan adaptasi model terhadap konteks penggunaan yang berbeda. *Threshold* rendah seperti 0.300 cocok untuk skrining agresif, di mana prioritas utama adalah mendekripsi sebanyak mungkin kasus kanker, meskipun dengan konsekuensi meningkatnya *false positive*. Sebaliknya, *threshold* tinggi seperti 0.700 memberikan spesifisitas sempurna, menjadikannya ideal untuk konfirmasi diagnosis sebelum tindakan invasif, meskipun dengan risiko meningkatnya kasus kanker yang terlewati.

Dengan menyediakan analisis *trade-off* yang eksplisit, penelitian ini memberikan kerangka kerja yang dapat digunakan oleh praktisi medis untuk menyesuaikan sistem pendukung keputusan sesuai kebutuhan lokal. Pendekatan ini merepresentasikan pergeseran dari model klasifikasi yang bersifat universal menuju sistem yang adaptif dan kontekstual, sejalan dengan prinsip-prinsip kedokteran presisi dan praktik berbasis bukti.

#### 4. Kesimpulan

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem pendukung keputusan berbasis regresi logistik untuk klasifikasi kanker payudara dengan pendekatan evaluasi yang komprehensif. Model yang dihasilkan mencapai performa klasifikasi yang sangat baik dengan akurasi 95,8%, sensitivitas 96,2%, dan spesifisitas 95,6% pada data uji independen. Kemampuan diskriminasi model yang ditunjukkan dengan nilai AUC-ROC 0,993 mengindikasikan potensi yang sangat baik untuk aplikasi klinis praktis.

Optimasi *threshold* keputusan menggunakan Youden's Index berhasil mengidentifikasi nilai optimal 0,560 yang memberikan keseimbangan terbaik antara sensitivitas dan spesifisitas. Temuan ini menunjukkan bahwa *threshold* default 0,5 yang umum digunakan tidak selalu optimal untuk konteks medis spesifik. Evaluasi kalibrasi probabilitas menunjukkan bahwa model memiliki *Expected Calibration Error* yang rendah (0,0390) dan dapat ditingkatkan lebih lanjut menjadi 0,0328 menggunakan *isotonic regression*. Kalibrasi yang baik ini memastikan bahwa *output* probabilitas model dapat diinterpretasikan langsung sebagai estimasi risiko klinis yang bermakna.

Meskipun hasil yang dicapai sangat menjanjikan, beberapa keterbatasan perlu dipertimbangkan untuk implementasi klinis. Evaluasi menggunakan *single train-test split* tanpa *cross-validation* membatasi kemampuan untuk mengestimasi variabilitas performa model. *Dataset* WDBC yang digunakan, meskipun telah menjadi *benchmark* standar, berasal dari populasi barat yang mungkin memiliki karakteristik berbeda dengan populasi Indonesia. Oleh karena itu, validasi eksternal menggunakan data dari institusi kesehatan lokal mutlak diperlukan sebelum implementasi klinis yang sesungguhnya.

#### Ucapan Terima Kasih

Terima kasih kami sampaikan kepada Direktur Politeknik Negeri Padang melalui Pusat Penelitian dan Pengabdian kepada Masyarakat (P3M) yang telah memfasilitasi pendanaan penelitian melalui Dana Hibah Penelitian DIPA PNP Tahun 2025.

#### Daftar Rujukan

- [1] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.
- [2] J. G. Elmore *et al.*, “Diagnostic concordance among pathologists interpreting breast biopsy specimens,” *JAMA - J. Am. Med. Assoc.*, vol. 313, no. 11, pp. 1122–1132, 2015, doi: 10.1001/jama.2015.1405.
- [3] S. Gondhowiardjo *et al.*, “Five-Year Cancer Epidemiology at the National Referral Hospital: Hospital-Based Cancer Registry Data in Indonesia,” *JCO Glob. Oncol.*, no. 7, pp. 190–203, 2021, doi: 10.1200/go.20.00155.
- [4] A. P. Windarto, A. Wanto, Solikhun, and R. Watrionthos, “A Comprehensive Bibliometric Analysis of Deep Learning Techniques for Breast Cancer Segmentation: Trends and Topic Exploration (2019-2023),” *J. RESTI*, vol. 7, no. 5, pp. 1155–1164, 2023, doi: 10.29207/resti.v7i5.5274.

- [5] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” no. 1, pp. 2–8, 2019, [Online]. Available: <http://arxiv.org/abs/1803.08375>
- [6] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, “Breast cancer classification using machine learning,” *2018 Electr. Electron. Comput. Sci. Biomed. Eng. Meet. EBBT 2018*, pp. 1–4, 2018, doi: 10.1109/EBBT.2018.8391453.
- [7] T. Elizabeth, “Prediksi Jenis Kanker Payudara Menggunakan Metode Support Vector Machine Berbasis Recursive Feature Elimination,” *Julyxxxx*, vol. x, no. x, pp. 1–5, 2024.
- [8] M. H. Alshayeji, H. Ellethy, S. Abed, and R. Gupta, “Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach,” *Biomed. Signal Process. Control*, vol. 71, no. January 2021, p. 103141, 2022, doi: 10.1016/j.bspc.2021.103141.
- [9] S. Al Muhaisen *et al.*, “Artificial Intelligence-Powered Mammography: Navigating the Landscape of Deep Learning for Breast Cancer Detection,” *Cureus*, vol. 16, no. 3, 2024, doi: 10.7759/cureus.56945.
- [10] D. Vickers, L. De Ravee, and J. Graham, “International survey of cochlear implant candidacy,” *Cochlear Implants Int.*, vol. 17, pp. 36–41, 2016, doi: 10.1080/14670100.2016.1155809.
- [11] B. Van Calster *et al.*, “Calibration: The Achilles heel of predictive analytics,” *BMC Med.*, vol. 17, no. 1, pp. 1–7, 2019, doi: 10.1186/s12916-019-1466-7.
- [12] F. Habibzadeh, P. Habibzadeh, and M. Yadollahie, “On determining the most appropriate test cut-off value: The case of tests with continuous results,” *Biochem. Medica*, vol. 26, no. 3, pp. 297–307, 2016, doi: 10.11613/BM.2016.034.
- [13] L. Huang, J. Zhao, B. Zhu, H. Chen, and S. Vanden Broucke, “An Experimental Investigation of Calibration Techniques for Imbalanced Data,” *IEEE Access*, vol. 8, pp. 127343–127352, 2020, doi: 10.1109/ACCESS.2020.3008150.
- [14] E. J. Caruana, M. Roman, J. Hernández-Sánchez, and P. Solli, “Longitudinal studies,” *J. Thorac. Dis.*, vol. 7, no. 11, pp. E537–E540, 2015, doi: 10.3978/j.issn.2072-1439.2015.10.63.
- [15] R. A. Smith *et al.*, “Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening,” *CA. Cancer J. Clin.*, vol. 69, no. 3, pp. 184–210, 2019, doi: 10.3322/caac.21557.