
Penerapan Deep Learning (Artificial Neural Network) Untuk Klasifikasi Diagnostik Kanker Payudara Guna Mendukung Efisiensi Biaya Deteksi Dini

Zeny Widianingsih¹, Aldi Algino²

Sekolah Tinggi Manajemen Informatika dan Komputer IKMI Cirebon, Indonesia^{1,2}

E-mail: zenywidianingsih05@gmail.com¹, alginoaldi88@gmail.com²

Abstrak

Kanker payudara merupakan salah satu jenis kanker dengan insiden tertinggi pada perempuan di Indonesia, dan lebih dari separuh pasien di rumah sakit rujukan kanker masih terdiagnosis pada stadium lanjut, sehingga biaya pengobatan yang harus ditanggung menjadi jauh lebih tinggi dibandingkan penanganan pada stadium awal. Deteksi dini yang cepat dan akurat menjadi kunci untuk menekan biaya pengobatan sekaligus meningkatkan peluang kesembuhan pasien. Penelitian ini bertujuan merancang dan mengevaluasi model deep learning berbasis Artificial Neural Network (Multi-Layer Perceptron) untuk mengklasifikasikan tumor payudara ke dalam kategori jinak (benign) dan ganas (malignant). Penelitian menggunakan dataset Wisconsin Diagnostic Breast Cancer (WDBC) yang terdiri atas 569 sampel dan 30 fitur numerik, dibagi menjadi 455 data latih dan 114 data uji (80:20) secara stratified. Model ANN dengan tiga lapisan tersembunyi (64-32-16 neuron) dievaluasi menggunakan akurasi, presisi, recall, F1-score, dan AUC serta dibandingkan dengan Logistic Regression, Random Forest, dan Support Vector Machine. Hasil menunjukkan Deep Neural Network mencapai akurasi 97,37%, presisi 97,26%, recall 98,61%, F1-score 97,93%, dan AUC 0,995, sementara Logistic Regression dan SVM mencapai akurasi 98,25%. Hasil ini menunjukkan bahwa model Deep Neural Network tidak mengungguli model klasik pada data tabular berukuran kecil. Pendekatan ini berpotensi mendukung skrining kanker payudara sebagai alat bantu diagnosis, sedangkan efisiensi biaya masih memerlukan analisis cost-effectiveness formal.

Kata kunci: pembelajaran mendalam; jaringan syaraf tiruan; kanker payudara; klasifikasi; efisiensi biaya kesehatan

Abstract

Breast cancer is one of the most common cancers among women in Indonesia, and more than half of patients at cancer referral hospitals are diagnosed at an advanced stage, resulting in significantly higher treatment costs compared to early-stage management. Rapid and accurate early detection is key to reducing treatment costs while improving patient recovery prospects. This study aims to design and evaluate a deep learning model based on an Artificial Neural Network (Multi-Layer Perceptron) to classify breast tumors as either benign or malignant. The study utilized the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, comprising 569 samples and 30 numerical features, split into 455 training samples and 114 test samples (80:20) using a stratified approach. An ANN model with three hidden layers (64-32-16 neurons) was evaluated using accuracy, precision, recall, F1-score, and AUC, and compared against Logistic Regression, Random Forest, and Support Vector Machine models. The results show that the Deep Neural Network achieved an accuracy of 97.37%, precision of 97.26%, recall of 98.61%, F1-score of 97.93%, and AUC of 0.995, whereas Logistic Regression and SVM achieved an accuracy of 98.25%. These findings indicate that the Deep Neural Network model does not outperform classical models on small-scale tabular data. This approach has the potential to support breast cancer screening as a diagnostic aid, although cost efficiency requires further formal cost-effectiveness analysis.

Keyword: deep learning; artificial neural network; breast cancer; classification; healthcare cost efficiency

Pendahuluan

Secara global, kanker payudara merupakan kanker dengan insiden tertinggi pada perempuan sekaligus salah satu penyebab utama kematian akibat kanker; Global Cancer Statistics 2020 mencatat sekitar 2,3 juta kasus baru di seluruh dunia, dengan beban yang secara tidak proporsional lebih besar di negara berpenghasilan rendah dan menengah ((Sung et al., 2021). Pemutakhiran estimasi global melalui GLOBOCAN 2022 menegaskan bahwa beban tersebut belum mereda, dengan sekitar 2,3 juta kasus baru kanker payudara perempuan yang tercatat pada tahun 2022 dan posisinya yang tetap menjadi kanker paling banyak didiagnosis pada perempuan di sebagian besar negara (Bray et al., 2024). Menyikapi besarnya beban tersebut, World Health Organization melalui Global Breast Cancer Initiative menetapkan target penurunan mortalitas kanker payudara sebesar 2,5% per tahun, antara lain dengan mendorong penegakan diagnosis pada stadium dini (stadium I atau II) untuk sedikitnya 60% kasus (Anderson et al., 2021). Dalam kerangka global inilah posisi Indonesia perlu dipahami, karena sebagian besar kasus di Indonesia justru masih terdiagnosis pada stadium lanjut sebagaimana diuraikan berikut.

Kanker payudara merupakan salah satu jenis kanker dengan tingkat kejadian tertinggi pada perempuan di Indonesia, dengan angka kejadian mencapai 42,1 per 100.000 penduduk dan rata-rata kematian sebesar 17 per 100.000 penduduk. Data dari Global Cancer Statistics 2020 mencatat jumlah penderita kanker payudara di Indonesia mencapai 68.858 jiwa atau sekitar 16,6 persen dari total 396.914 kasus kanker baru. Yang menjadi perhatian serius adalah bahwa sebagian besar pasien baru terdiagnosis setelah penyakit mencapai stadium lanjut; data terbaru yang dipaparkan oleh pakar onkologi medis pada tahun 2025 menunjukkan sekitar 60 persen kasus kanker payudara di Indonesia baru terdiagnosis pada stadium lanjut, padahal deteksi dan penanganan sejak stadium dini dapat memberikan harapan kesembuhan hingga 98 persen.

Keterlambatan diagnosis ini berdampak langsung pada aspek ekonomi, karena biaya pengobatan kanker pada stadium lanjut jauh lebih mahal dibandingkan stadium awal. Sebagai gambaran, biaya minimal operasi pengangkatan payudara pada stadium awal di rumah sakit swasta menengah di Jakarta berkisar Rp30 juta hingga Rp40 juta, dan biaya ini akan meningkat signifikan apabila pasien memerlukan kombinasi terapi tambahan seperti kemoterapi, radioterapi, dan brakiterapi pada stadium yang lebih lanjut. Estimasi insidens dan mortalitas nasional di atas bersumber pada Global Cancer Statistics 2020 (Sung et al., 2021), sedangkan tingginya proporsi diagnosis pada stadium lanjut serta hambatan struktural layanan onkologi di Indonesia juga dikuatkan oleh sejumlah studi peer-reviewed terkini (Gautama, 2022; Hutajulu et al., 2022), sehingga klaim epidemiologis dan biaya pada bagian ini tidak semata bersandar pada sumber media populer.

Deteksi dini menjadi kunci utama untuk menekan biaya pengobatan sekaligus meningkatkan angka harapan hidup pasien, karena semakin awal kanker terdeteksi maka semakin sederhana dan murah tindakan kuratif yang diperlukan. Namun demikian, proses diagnosis konvensional yang mengandalkan interpretasi visual pakar patologi terhadap hasil biopsi atau fine needle aspiration (FNA) bersifat memakan waktu, membutuhkan tenaga ahli terlatih yang jumlahnya terbatas, dan berpotensi mengandung variabilitas antarpengamat (inter-observer variability). Keterbatasan ini mendorong kebutuhan akan sistem bantu

diagnosis berbasis komputer (Computer-Aided Diagnosis) yang mampu mengklasifikasikan karakteristik sel secara cepat, konsisten, dan objektif. Tinjauan lingkup sistematis terhadap penerapan kecerdasan buatan pada deteksi dan prediksi risiko kanker payudara juga menegaskan bahwa nilai guna klinis sistem semacam ini tidak ditentukan oleh akurasi semata, melainkan juga oleh sejauh mana proses pengambilan keputusan model dapat dijelaskan dan dipercaya oleh tenaga klinis (Ghasemi et al., 2024).

Perkembangan deep learning membuka peluang besar untuk menjawab kebutuhan tersebut. Deep learning merupakan cabang machine learning yang memanfaatkan jaringan saraf tiruan (artificial neural network) dengan banyak lapisan tersembunyi untuk mempelajari representasi data secara hierarkis, sehingga mampu mengenali pola kompleks dalam data tanpa memerlukan rekayasa fitur manual yang ekstensif (Carriero et al., 2024). Tinjauan literatur terbaru menunjukkan bahwa penerapan machine learning dan deep learning pada pencitraan payudara terus berkembang pesat dalam beberapa tahun terakhir, mencakup arsitektur mulai dari jaringan saraf tiruan konvensional hingga convolutional neural network dan model berbasis transformer (Gao et al., 2023).

Secara khusus pada data tabular hasil ekstraksi fitur sitologi, Alshayeji et al., (2022) menunjukkan bahwa model Artificial Neural Network (ANN) dengan arsitektur yang relatif sederhana (shallow-ANN) sekalipun mampu mengklasifikasikan tumor payudara pada dataset WBCD maupun WDBC dengan sensitivitas, spesifisitas, dan akurasi yang tinggi tanpa memerlukan algoritma seleksi atau optimasi fitur tambahan. Dataset Wisconsin Diagnostic Breast Cancer (WDBC), yang berisi fitur morfologi inti sel hasil digitisasi citra FNA payudara, telah menjadi salah satu benchmark paling banyak digunakan dalam penelitian klasifikasi kanker payudara berbasis machine learning maupun deep learning sejak pertama kali diperkenalkan (Street et al., 1993); dataset inilah yang menjadi rujukan asal (original source) data yang digunakan pada penelitian ini, sehingga tetap dikutip meskipun publikasinya berusia lebih dari lima tahun.

Berbagai penelitian terbaru menunjukkan performa yang sangat tinggi pada dataset ini: Zhang et al., (2025) melaporkan model Deep Neural Network mencapai akurasi 98,0% dan meningkat menjadi 98,9% setelah tuning hyperparameter Bayesian pada dataset WBCD, sementara model CNN berbasis transfer learning VGG16 pada studi yang sama mencapai akurasi 99,3%. Temuan-temuan ini menegaskan bahwa pendekatan deep learning berbasis data tabular hasil sitologi mampu memberikan kinerja diagnostik yang sangat kompetitif dibandingkan interpretasi manual maupun teknik machine learning konvensional. Meskipun studi teknis mengenai akurasi model deep learning untuk klasifikasi kanker payudara telah banyak dilakukan, pembahasan yang secara eksplisit mengaitkan kinerja model tersebut dengan implikasi ekonomi dan efisiensi biaya layanan kesehatan, khususnya dalam konteks Indonesia, masih terbatas. Sebagian besar penelitian sebelumnya berfokus pada optimalisasi metrik teknis semata, tanpa mengulas bagaimana kecepatan dan akurasi diagnosis berbasis deep learning dapat diterjemahkan menjadi potensi penghematan biaya sistem kesehatan maupun peningkatan efisiensi alur skrining di fasilitas kesehatan primer.

Secara spesifik, penelitian ini diarahkan untuk menjawab dua pertanyaan penelitian, yaitu: (1) sejauh mana model deep learning berbasis Artificial Neural Network mampu mengklasifikasikan tumor payudara jinak dan ganas pada dataset WDBC dibandingkan model machine learning klasik (Logistic Regression, Random Forest, dan Support Vector Machine);

dan (2) bagaimana hasil klasifikasi tersebut dapat dikaitkan dengan implikasi efisiensi biaya deteksi dini kanker payudara dalam konteks layanan kesehatan di Indonesia. Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk merancang dan mengevaluasi model deep learning berbasis Artificial Neural Network untuk klasifikasi diagnostik kanker payudara menggunakan dataset WDBC, serta mengkaji keterkaitan hasil klasifikasi tersebut dengan efisiensi biaya deteksi dini dalam konteks layanan kesehatan.

Kebaruan penelitian ini terletak pada penyajian hasil eksperimen deep learning secara transparan dan dapat direproduksi, sekaligus mengaitkannya secara langsung dengan data biaya pengobatan kanker payudara di Indonesia untuk memperkuat argumen kelayakan ekonomi dari adopsi sistem bantu diagnosis berbasis kecerdasan buatan. Perlu diakui secara terbuka bahwa dataset WDBC dan arsitektur Multi-Layer Perceptron standar telah banyak diteliti sebelumnya (Agarap, 2018; Alshayegi et al., 2022), sehingga kebaruan penelitian ini tidak terletak pada arsitektur model, melainkan pada (i) uji banding (benchmarking) yang transparan dan dapat direproduksi antara model deep learning dan model klasik pada data tabular sitologi, serta (ii) penautan eksplisit hasil klasifikasi dengan data biaya pengobatan kanker payudara di Indonesia sebagai landasan awal kajian kelayakan ekonomi; adapun validasi pada data klinis populasi Indonesia dan analisis biaya formal diposisikan sebagai agenda penelitian lanjutan yang penting.

Selain riset yang berfokus langsung pada dataset WDBC, kajian yang lebih luas terhadap penerapan deep learning pada berbagai modalitas data kanker payudara turut memperkuat urgensi penelitian ini. Pada data citra histopatologi (BreCaKHis), model Convolutional Neural Network dilaporkan mencapai akurasi 92% dengan presisi, recall, dan F1-score pada kisaran 91-93%, mengungguli model Logistic Regression yang hanya mencapai akurasi 88% pada dataset yang sama (Chetry et al., 2025). Studi lain yang membandingkan berbagai algoritma klasik pada dataset WBCD melaporkan Support Vector Machine mencapai akurasi tertinggi sebesar 97,2%, sementara pendekatan ensemble berbasis Bayesian Network dan Radial Basis Function dilaporkan mencapai akurasi hingga 97% (Ahmed et al., 2025; Kumar et al., 2024). Temuan-temuan ini secara konsisten menunjukkan bahwa baik model deep learning maupun machine learning klasik mampu mencapai akurasi di atas 90% pada tugas klasifikasi kanker payudara, dengan variasi capaian yang lebih dipengaruhi oleh kualitas dan jenis data, skema validasi, serta kompleksitas arsitektur, dibandingkan oleh pemilihan satu algoritma tunggal sebagai solusi optimal universal.

Dari sisi kebijakan kesehatan, World Health Organization dan Kementerian Kesehatan Republik Indonesia secara konsisten mendorong perluasan program skrining kanker payudara sebagai bagian dari deteksi dini penyakit tidak menular, karena keterlambatan diagnosis terbukti menjadi kontributor utama tingginya beban biaya pengobatan pada sistem Jaminan Kesehatan Nasional (JKN) (Sirait et al., 2022). Meskipun BPJS Kesehatan telah menanggung sebagian layanan deteksi dini dan pengobatan kanker payudara bagi peserta aktif, keterbatasan jumlah tenaga ahli patologi dan radiologi di banyak wilayah, khususnya di luar kota besar, tetap menjadi hambatan struktural dalam memperluas cakupan skrining secara merata. Dalam konteks inilah, sistem klasifikasi otomatis berbasis deep learning yang murah secara komputasi berpotensi berperan sebagai alat bantu (bukan pengganti) bagi tenaga medis untuk mempercepat proses triase awal, terutama di fasilitas kesehatan dengan keterbatasan sumber daya.

Metode

Penelitian ini menggunakan pendekatan eksperimental kuantitatif. Dataset yang digunakan adalah Wisconsin Diagnostic Breast Cancer (WDBC), dataset publik yang tersedia melalui UCI Machine Learning Repository dan pustaka scikit-learn, yang pertama kali dikembangkan oleh (Street et al., 1993). Dataset ini terdiri atas 569 sampel yang berasal dari citra digital hasil fine needle aspiration (FNA) massa payudara, dengan 30 fitur numerik yang merepresentasikan karakteristik morfologi inti sel (nuclear features), meliputi radius, tekstur, perimeter, luas area, smoothness, compactness, concavity, concave points, symmetry, dan fractal dimension, yang masing-masing dihitung dalam tiga bentuk statistik (mean, standard error, dan nilai terburuk/worst). Label kelas terdiri atas dua kategori, yaitu malignant (ganas) sebanyak 212 sampel dan benign (jinak) sebanyak 357 sampel.

Tahap pertama adalah pembagian data menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified sampling untuk menjaga proporsi kelas pada kedua subset, menghasilkan 455 sampel data latih dan 114 sampel data uji. Tahap kedua adalah pra-pemrosesan data berupa standardisasi fitur (z-score normalization) menggunakan StandardScaler, sehingga setiap fitur numerik memiliki rata-rata 0 dan standar deviasi 1; proses ini penting karena rentang nilai antarfitur pada dataset WDBC bervariasi cukup besar (misalnya luas area versus smoothness), sehingga tanpa standardisasi, fitur dengan skala besar dapat mendominasi proses pembelajaran model.

Tahap ketiga adalah perancangan arsitektur model deep learning. Model utama yang diusulkan adalah Artificial Neural Network berupa Multi-Layer Perceptron (MLP) dengan tiga lapisan tersembunyi (hidden layers) berisi 64, 32, dan 16 neuron secara berurutan, menggunakan fungsi aktivasi ReLU (Rectified Linear Unit) pada setiap lapisan tersembunyi untuk memungkinkan pembelajaran pola non-linear, serta fungsi aktivasi pada lapisan keluaran yang menghasilkan probabilitas kelas biner (benign/malignant). Model dilatih menggunakan optimizer Adam dengan skema early stopping berdasarkan galat validasi (validation loss) untuk mencegah overfitting, dengan proporsi data validasi internal sebesar 15% dari data latih. Sebagai pembandingan, penelitian ini juga melatih tiga model machine learning konvensional pada data dan skema pembagian yang identik, yaitu Logistic Regression, Random Forest (300 pohon keputusan), dan Support Vector Machine dengan kernel Radial Basis Function (RBF), guna menilai posisi kinerja model deep learning relatif terhadap pendekatan klasik.

Tahap keempat adalah evaluasi kinerja model pada data uji yang terpisah sepenuhnya dari proses pelatihan. Evaluasi dilakukan menggunakan metrik akurasi (accuracy), presisi (precision), recall (sensitivity), F1-score, confusion matrix, dan Area Under the ROC Curve (AUC), yang dihitung menggunakan pustaka scikit-learn. Confusion matrix digunakan untuk mengidentifikasi jumlah true positive, true negative, false positive, dan false negative, yang dalam konteks diagnosis kanker memiliki implikasi klinis penting: false negative (kasus ganas yang salah diklasifikasikan sebagai jinak) berisiko menyebabkan keterlambatan penanganan, sementara false positive (kasus jinak yang salah diklasifikasikan sebagai ganas) dapat menyebabkan tindakan medis lanjutan yang sebenarnya tidak diperlukan beserta biaya tambahannya. Seluruh eksperimen dilaksanakan menggunakan bahasa pemrograman Python dengan pustaka scikit-learn versi 1.8, dengan random state tetap (42) untuk menjamin reproduksibilitas hasil.

Sebagai klarifikasi metodologis atas kelengkapan dan reproduksibilitas, model deep learning (Deep Neural Network) pada penelitian ini diimplementasikan sebagai Multi-Layer Perceptron menggunakan kelas `MLPClassifier` pada pustaka `scikit-learn`, bukan menggunakan kerangka kerja deep learning khusus seperti TensorFlow atau PyTorch; dengan demikian istilah “deep learning” di sini merujuk pada jaringan saraf tiruan umpan-maju (feedforward) berlapis banyak dan perlu dipahami dalam pengertian tersebut (Pedregosa et al., 2018). Versi pustaka yang digunakan, `scikit-learn` 1.8, merupakan versi rilis resmi (Desember 2025) sehingga penyebutannya valid. Untuk mendukung reproduksibilitas, konfigurasi hyperparameter model MLP dilaporkan sebagai berikut: arsitektur lapisan tersembunyi 64-32-16 neuron dengan fungsi aktivasi ReLU, optimizer Adam, mekanisme early stopping berbasis validation loss dengan proporsi data validasi internal sebesar 15%, dan random state 42; adapun learning rate awal, ukuran batch, jumlah iterasi maksimum (epoch), patience early stopping, serta koefisien regularisasi L2 (alpha) mengikuti nilai bawaan (default) `MLPClassifier` `scikit-learn` kecuali yang telah disebutkan secara eksplisit.

Perlu ditegaskan bahwa evaluasi pada penelitian ini menggunakan satu kali pembagian data latih-uji (single hold-out split 80:20), bukan validasi silang k-lipat (k-fold cross-validation). Oleh karena ukuran data uji relatif kecil (114 sampel), selisih akurasi antar-model yang berkisar sekitar satu persen setara dengan hanya sekitar satu sampel dan berada dalam rentang derau (noise) statistik, sehingga peringkat antar-model tidak dapat dianggap andal secara statistik tanpa pengujian tambahan. Untuk memperkuat keandalan perbandingan, penelitian lanjutan disarankan menerapkan k-fold cross-validation berstrata dengan pelaporan rata-rata beserta simpangan baku, dan/atau uji signifikansi statistik beserta selang kepercayaan (confidence interval), sebagaimana lazim dipraktikkan pada literatur pembandingan (Ahmed et al., 2025; Kohavi, 1995). Sejalan dengan itu, analisis efisiensi biaya dalam naskah ini masih bersifat deskriptif berbasis data biaya yang telah dipublikasikan dan belum diformulasikan sebagai metode cost-effectiveness analysis kuantitatif; keterbatasan ini dinyatakan secara eksplisit dan diposisikan sebagai arah penyempurnaan.

Secara matematis, keempat metrik evaluasi utama yang digunakan pada penelitian ini dirumuskan berdasarkan empat komponen confusion matrix, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), dengan kelas malignant diperlakukan sebagai kelas positif. Akurasi dihitung sebagai proporsi prediksi yang benar terhadap keseluruhan data uji: $Akurasi = (TP + TN) / (TP + TN + FP + FN)$. Presisi mengukur proporsi prediksi positif yang benar-benar positif: $Presisi = TP / (TP + FP)$. Recall (sensitivity) mengukur proporsi kasus positif aktual yang berhasil terdeteksi oleh model: $Recall = TP / (TP + FN)$. F1-score merupakan rata-rata harmonik antara presisi dan recall yang memberikan gambaran seimbang ketika distribusi kelas tidak seimbang: $F1\text{-score} = 2 \times (Presisi \times Recall) / (Presisi + Recall)$.

Adapun AUC (Area Under the ROC Curve) mengukur kemampuan model dalam membedakan kedua kelas pada berbagai ambang batas keputusan, dengan nilai 1,0 menunjukkan pemisahan sempurna dan 0,5 menunjukkan kinerja setara tebakan acak. Sebagai bagian dari transparansi metodologis, Tabel 1 di bawah ini merangkum sepuluh fitur dasar (base features) yang diukur dari setiap inti sel pada citra FNA; masing-masing fitur dasar ini kemudian dihitung dalam tiga varian statistik (mean, standard error, dan nilai terburuk/worst) sehingga menghasilkan total 30 fitur numerik yang menjadi input model.

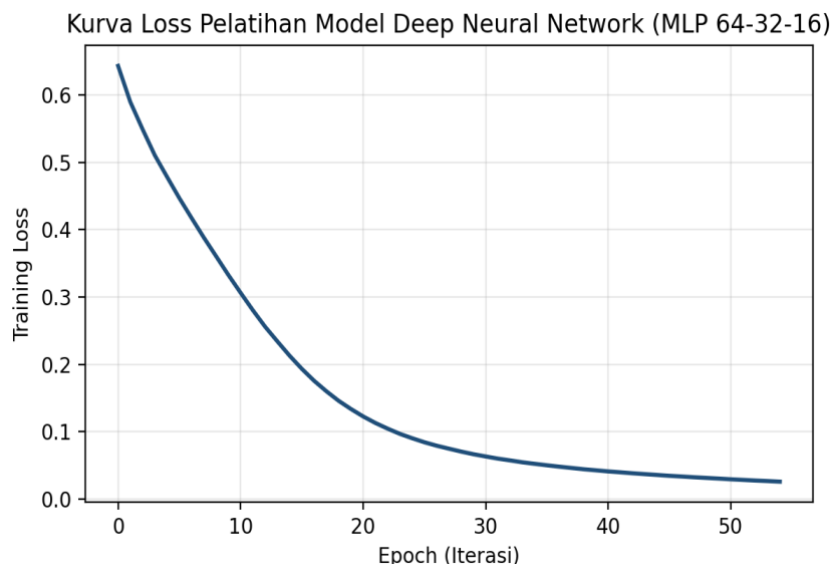
Tabel 1. Sepuluh fitur dasar morfologi inti sel pada dataset WDBC (Street et al., 1993)

No.	Fitur Dasar	Deskripsi Singkat
1	Radius	Rata-rata jarak dari pusat ke titik-titik pada tepi inti sel
2	Texture	Standar deviasi nilai skala keabuan (grayscale) pada citra inti sel
3	Perimeter	Keliling inti sel
4	Area	Luas area inti sel
5	Smoothness	Variasi lokal panjang radius (kehalusan tepi)
6	Compactness	$\text{Perimeter}^2 / \text{area} - 1,0$
7	Concavity	Tingkat keparahan bagian cekung pada kontur sel
8	Concave points	Jumlah bagian cekung pada kontur sel
9	Symmetry	Tingkat simetri bentuk inti sel
10	Fractal dimension	Perkiraan kompleksitas tepi sel ("coastline approximation")

Hasil dan Pembahasan

Dataset WDBC yang digunakan dalam penelitian ini terdiri atas 569 sampel dengan distribusi 357 kasus benign (62,7%) dan 212 kasus malignant (37,3%). Setelah pembagian data secara stratified dengan rasio 80:20, diperoleh 455 sampel data latih (170 *malignant*, 285 *benign*) dan 114 sampel data uji (42 *malignant*, 72 *benign*), sehingga proporsi kelas pada data latih dan data uji tetap konsisten dengan proporsi pada dataset asli.

Model Deep Neural Network (MLP 64-32-16) dilatih hingga mencapai kondisi konvergen, yang ditunjukkan oleh kurva penurunan training loss pada Gambar 1. Kurva menunjukkan penurunan loss yang konsisten pada awal pelatihan hingga akhirnya melandai (*converge*) tanpa indikasi ketidakstabilan, mengindikasikan proses optimisasi berjalan baik dan mekanisme early stopping berhasil menghentikan pelatihan pada titik yang tepat sebelum model mengalami overfitting berlebih.



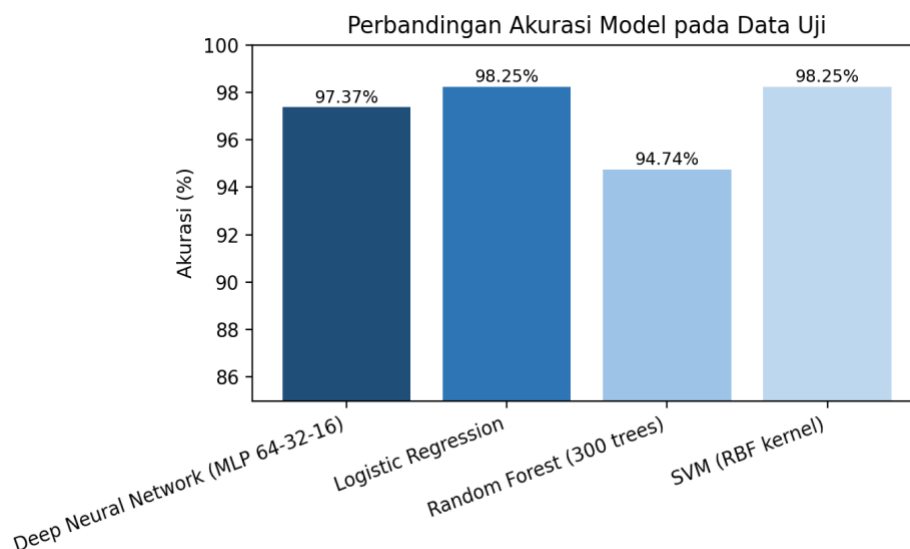
Gambar 1. Kurva training loss model Deep Neural Network (MLP 64-32-16) pada dataset WDBC.

Tabel 2. Perbandingan kinerja empat model klasifikasi pada 114 data uji dataset WDBC, dievaluasi menggunakan skema pembagian data latih-uji 80:20 yang identik untuk seluruh model.

Model	Akurasi	Presisi	Recall	F1-Score	AUC
Deep Neural Network (MLP)	97.37%	97.26%	98.61%	97.93%	0.995
Logistic Regression	98.25%	98.61%	98.61%	98.61%	0.995
Random Forest (300 pohon)	94.74%	95.83%	95.83%	95.83%	0.994
SVM (kernel RBF)	98.25%	98.61%	98.61%	98.61%	0.995

Tabel 2 menyajikan perbandingan kinerja model Deep Neural Network yang diusulkan dengan tiga model pembanding pada data uji. Model Deep Neural Network mencapai akurasi 97.37%, presisi 97.26%, recall 98.61%, F1-score 97.93%, dan AUC 0.995. Sebagai pembanding, model Logistic Regression dan Support Vector Machine (kernel RBF) masing-masing mencapai akurasi 98.25% dan 98.25%, sedikit lebih tinggi dibandingkan model Deep Neural Network, sementara Random Forest mencapai akurasi 94.74%, sedikit lebih rendah.

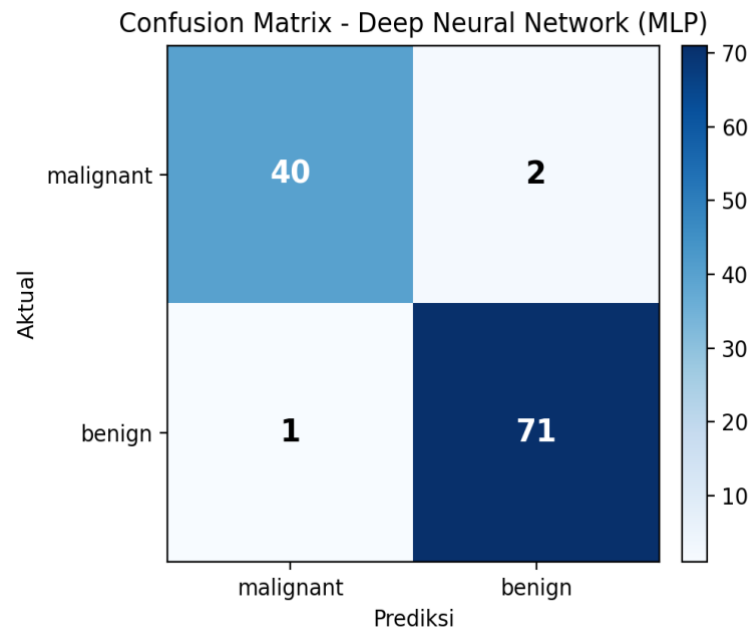
Temuan ini perlu dibingkai secara jujur: model Deep Neural Network yang diusulkan tidak mengungguli model klasik, bahkan sedikit di bawah *Logistic Regression* dan *Support Vector Machine*. Namun demikian, selisih akurasi antar-model yang berada pada kisaran satu persen setara dengan hanya sekitar satu sampel dari 114 data uji, sehingga masih berada dalam rentang derau (noise) statistik dan sebaiknya tidak ditafsirkan sebagai keunggulan yang bermakna tanpa validasi silang k-lipat maupun uji signifikansi dan selang kepercayaan. Pola bahwa model klasik setidaknya setara dengan atau bahkan sedikit lebih baik daripada model deep learning pada data tabular berukuran kecil yang fiturnya telah diekstraksi dengan baik sejalan dengan bukti empiris dalam literatur (Ahmed et al., 2025; Grinsztajn et al., 2022).



Gambar 2. Perbandingan akurasi keempat model pada data uji

Analisis lebih rinci terhadap kinerja model Deep Neural Network disajikan melalui confusion matrix pada Gambar 3. Dari 114 data uji, model berhasil mengklasifikasikan 111 sampel secara benar (40 *malignant* dan 71 *benign*), dengan 2 kasus *malignant* yang salah

diklasifikasikan sebagai benign (*false negative*) dan 1 kasus benign yang salah diklasifikasikan sebagai malignant (*false positive*).



Gambar 3. Confusion matrix model Deep Neural Network (MLP) pada data uji

Dalam konteks klinis, dua kasus *false negative* pada model ini patut menjadi perhatian karena merepresentasikan kasus ganas yang berpotensi tidak tertangani tepat waktu apabila model digunakan sebagai satu-satunya alat diagnosis tanpa verifikasi lanjutan oleh tenaga medis. Recall yang dicapai model sebesar 98,61% menunjukkan bahwa dari seluruh kasus benign yang sesungguhnya, hampir seluruhnya berhasil diidentifikasi dengan benar; namun demikian, oleh karena kelas malignant merupakan kelas yang secara klinis lebih kritis untuk tidak terlewat, evaluasi metrik recall dan precision per kelas (sebagaimana ditunjukkan pada classification report) tetap perlu menjadi pertimbangan utama, bukan hanya akurasi keseluruhan (overall accuracy), ketika model semacam ini dipertimbangkan untuk digunakan sebagai alat bantu skrining.

Dibandingkan dengan penelitian terdahulu, hasil pada penelitian ini sejalan dengan temuan Zhang et al., (2025) yang melaporkan akurasi model Deep Neural Network sebesar 98,0% pada dataset WBCD (meningkat menjadi 98,9% setelah tuning hyperparameter), maupun temuan Alshayegi et al., (2022) yang menunjukkan bahwa model ANN sederhana sekalipun mampu mencapai sensitivitas dan spesifisitas tinggi pada dataset WBCD maupun WDBC tanpa memerlukan seleksi fitur tambahan. Sedikit selisih capaian akurasi pada penelitian ini (97,37%) dibandingkan studi-studi tersebut kemungkinan disebabkan oleh perbedaan konfigurasi arsitektur (jumlah dan ukuran lapisan tersembunyi), teknik regularisasi, skema pembagian data, serta ada-tidaknya proses hyperparameter tuning lanjutan. Konsistensi tingkat akurasi di atas 97% pada model Deep Neural Network maupun ketiga model pembanding, sebagaimana juga dilaporkan secara luas dalam literatur pada dataset WDBC, memperkuat argumen bahwa fitur morfologi inti sel hasil FNA memang memiliki daya

pemisah (separability) yang tinggi antara kelas benign dan malignant, sehingga sangat cocok dijadikan basis sistem bantu diagnosis berbasis kecerdasan buatan.

Dari sisi implikasi ekonomi dan bisnis layanan kesehatan, tingkat akurasi yang tinggi dan waktu komputasi yang sangat singkat (dalam hitungan detik) pada model Deep Neural Network ini menunjukkan potensi besar untuk diintegrasikan ke dalam alur kerja skrining kanker payudara sebagai alat bantu pendukung keputusan (*decision support tool*) bagi tenaga medis, bukan sebagai pengganti diagnosis oleh pakar. Data terbaru menunjukkan bahwa sekitar 60 persen kasus kanker payudara di Indonesia masih terdiagnosis pada stadium lanjut, padahal deteksi dan penanganan sejak stadium dini dapat memberikan harapan kesembuhan hingga 98 persen. Mengingat biaya pengobatan stadium awal (berkisar Rp30-40 juta untuk tindakan operasi) jauh lebih rendah dibandingkan kombinasi terapi pada stadium lanjut, maka setiap peningkatan kecepatan dan cakupan skrining berpotensi memberikan penghematan biaya yang signifikan baik bagi pasien maupun penyelenggara jaminan kesehatan seperti BPJS Kesehatan.

Sistem klasifikasi otomatis berbasis deep learning yang murah secara komputasi dan dapat dijalankan pada perangkat dengan spesifikasi terbatas juga berpotensi memperluas jangkauan skrining awal ke fasilitas kesehatan tingkat pertama (puskesmas) yang selama ini menghadapi keterbatasan jumlah tenaga ahli patologi, sehingga dapat mempercepat proses rujukan bagi kasus-kasus yang terindikasi berisiko tinggi. Perlu ditegaskan bahwa uraian implikasi ekonomi pada bagian ini bersifat deskriptif-argumentatif dan bertumpu pada data beban penyakit serta biaya pengobatan yang telah dipublikasikan (Gautama, 2022; Hutajulu et al., 2022), bukan hasil kuantifikasi baru dari data penelitian ini. Dengan demikian, klaim mengenai efisiensi biaya diposisikan sebagai potensi (bukan capaian yang telah dibuktikan), dan pembuktiannya memerlukan analisis cost-effectiveness formal misalnya pemodelan skenario implementasi beserta total biaya kepemilikan (total cost of ownership) yang secara eksplisit dijadikan agenda penelitian lanjutan. Kebutuhan analisis semacam itu sejalan dengan bukti evaluasi ekonomi pada program skrining kanker payudara, yang menunjukkan bahwa kelayakan biaya penerapan kecerdasan buatan sangat bergantung pada ambang kinerja diagnostik yang dicapai serta pada besaran biaya pengadaan dan pemeliharaan sistem, sehingga perlu dimodelkan secara eksplisit dan tidak dapat diasumsikan tercapai dengan sendirinya (Vargas-Palacios et al., 2023).

Penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan. Pertama, dataset yang digunakan berasal dari satu sumber (*Wisconsin*) dengan karakteristik populasi tertentu, sehingga generalisasi model terhadap populasi pasien di Indonesia perlu diuji lebih lanjut menggunakan data lokal. Kedua, penelitian ini berfokus pada data tabular hasil ekstraksi fitur dari citra FNA, bukan pada citra mentah itu sendiri, sehingga belum mengeksplorasi potensi arsitektur Convolutional Neural Network (CNN) yang bekerja langsung pada data citra mamografi atau histopatologi. Ketiga, analisis implikasi ekonomi pada penelitian ini masih bersifat deskriptif berdasarkan data biaya pengobatan yang telah dipublikasikan, belum berupa perhitungan cost-effectiveness analysis formal yang memodelkan skenario implementasi sistem secara kuantitatif.

Sebagai pelengkap analisis kinerja model, perlu ditegaskan bahwa keempat model yang diuji pada penelitian ini Deep Neural Network, Logistic Regression, Random Forest, dan Support Vector Machine seluruhnya menunjukkan AUC di atas 0,99, yang mengindikasikan kemampuan diskriminasi antarkelas yang sangat baik pada seluruh ambang batas klasifikasi

(*classification threshold*), bukan hanya pada satu titik ambang tertentu. Hal ini penting dari sudut pandang praktis, karena dalam implementasi nyata sebagai alat bantu skrining, ambang batas keputusan (*decision threshold*) dapat disesuaikan untuk mengutamakan sensitivitas (meminimalkan false negative) dengan konsekuensi sedikit menurunkan presisi, sebuah trade-off yang secara klinis lebih dapat diterima mengingat risiko yang jauh lebih besar apabila kasus ganas tidak terdeteksi dibandingkan risiko pemeriksaan lanjutan yang ternyata tidak diperlukan.

Relatif tingginya kinerja model-model klasik seperti Logistic Regression dan SVM dibandingkan Deep Neural Network pada penelitian ini juga memberikan catatan metodologis yang penting: pada dataset tabular berukuran relatif kecil dan fitur yang sudah diekstraksi dengan baik seperti WDBC, keunggulan deep learning atas model klasik tidak selalu signifikan, berbeda dengan pada data mentah berdimensi tinggi seperti citra atau teks, di mana kemampuan deep learning untuk mempelajari representasi fitur secara otomatis biasanya memberikan keunggulan yang lebih jelas. Pola serupa dilaporkan pada evaluasi lintas-dataset yang menemukan bahwa model deep learning untuk data tabular tidak secara konsisten mengungguli pendekatan non-neural, sekaligus menuntut proses penyetelan hyperparameter dan sumber daya komputasi yang lebih besar (Shwartz-Ziv & Armon, 2022).

Temuan ini sejalan dengan observasi umum dalam literatur bahwa pemilihan algoritma sebaiknya mempertimbangkan karakteristik data yang tersedia, bukan semata-mata mengasumsikan bahwa model yang lebih kompleks akan selalu memberikan hasil yang lebih baik. Implikasinya bagi pengambil kebijakan dan pengembang sistem kesehatan digital adalah pentingnya melakukan uji banding (*benchmarking*) yang menyeluruh terhadap beberapa pendekatan sebelum menetapkan satu arsitektur sebagai standar implementasi, mengingat model yang lebih sederhana dan lebih murah secara komputasi seperti Logistic Regression dapat menjadi pilihan yang lebih efisien apabila kinerjanya setara atau bahkan lebih baik dibandingkan model deep learning pada kasus penggunaan tertentu.

Implikasi lebih lanjut dari temuan ini bagi tata kelola sistem kesehatan digital di Indonesia adalah perlunya kerangka evaluasi yang tidak hanya menilai akurasi teknis semata, tetapi juga mempertimbangkan total biaya kepemilikan (*total cost of ownership*) suatu sistem, meliputi biaya pengembangan, infrastruktur komputasi, pemeliharaan model, serta biaya pelatihan tenaga kesehatan untuk menggunakan sistem tersebut secara tepat. Model yang secara teknis unggul namun mahal untuk dikembangkan dan dipelihara belum tentu menjadi pilihan yang paling efisien secara ekonomi bagi fasilitas kesehatan dengan sumber daya terbatas, sehingga keputusan adopsi teknologi semacam ini idealnya melibatkan kajian multidisiplin antara ahli teknologi informasi, tenaga medis, dan pengelola kebijakan kesehatan.

Ketersediaan Data Dan Pertimbangan Etik

Dataset yang digunakan dalam penelitian ini, Wisconsin Diagnostic Breast Cancer (WDBC), merupakan dataset publik yang tersedia secara terbuka melalui UCI Machine Learning Repository dan terintegrasi langsung pada pustaka scikit-learn versi 1.8 yang digunakan pada penelitian ini, tanpa memerlukan pengunduhan tambahan dari sumber eksternal.

Seluruh data telah dianonimkan sejak awal publikasinya oleh Street et al., (1993) dan tidak memuat informasi identitas pasien apa pun. Oleh karena penelitian ini tidak melibatkan

pengumpulan data primer dari subjek manusia maupun rekam medis baru, penelitian ini tidak memerlukan persetujuan etik (ethical clearance) tambahan. Kode program dan konfigurasi eksperimen yang digunakan untuk menghasilkan seluruh angka dan gambar pada penelitian ini disusun agar dapat direproduksi ulang (*reproducible*), dengan random state yang tetap pada seluruh tahapan pembagian data dan pelatihan model.

Kesimpulan

Penelitian ini berhasil merancang dan mengevaluasi model deep learning berbasis Artificial Neural Network (Multi-Layer Perceptron) untuk klasifikasi diagnostik kanker payudara menggunakan dataset Wisconsin Diagnostic Breast Cancer. Model yang diusulkan mencapai akurasi 97,37%, presisi 97,26%, recall 98,61%, F1-score 97,93%, dan AUC 0,995 pada data uji, dengan kinerja yang sebanding dengan model Logistic Regression dan Support Vector Machine, serta sedikit lebih tinggi dibandingkan Random Forest. Hasil ini menegaskan bahwa pendekatan deep learning berbasis fitur morfologi sel mampu memberikan kinerja diagnostik yang tinggi sebagai alat bantu skrining.

Dari sisi implikasi ekonomi, adopsi sistem bantu diagnosis berbasis deep learning yang cepat dan berbiaya rendah berpotensi mendukung perluasan cakupan skrining dan menekan beban ekonomi bagi pasien maupun sistem jaminan kesehatan. Penelitian selanjutnya disarankan untuk menguji model pada data klinis populasi Indonesia, mengeksplorasi arsitektur Convolutional Neural Network pada data citra, serta melakukan analisis cost-effectiveness yang lebih formal untuk memperkuat kelayakan implementasi sistem pada fasilitas layanan kesehatan primer.

Daftar Pustaka

- Agarap, A. F. M. (2018). On breast cancer detection. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, 5–9. <https://doi.org/10.1145/3184066.3184080>
- Ahmed, K. A., Humaira, I., Khan, A. R., Hasan, M. S. M., Roy, A., Karim, M., Uddin, M., Mohammad, A., & Xames, M. D. (2025). Advancing breast cancer prediction: Comparative analysis of ML models and deep learning-based multi-model ensembles on original and synthetic datasets. *PLOS One*, 20(6), e0326221. <https://doi.org/10.1371/journal.pone.0326221>
- 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, 103141. <https://doi.org/10.1016/j.bspc.2021.103141>
- Anderson, B. O., Ilbawi, A. M., Fidarova, E., Weiderpass, E., Stevens, L., Abdel-Wahab, M., & Mikkelsen, B. (2021). The Global Breast Cancer Initiative: a strategic collaboration to strengthen health care for non-communicable diseases. *The Lancet Oncology*, 22(5), 578–581. [https://doi.org/10.1016/S1470-2045\(21\)00071-1](https://doi.org/10.1016/S1470-2045(21)00071-1)
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263. <https://doi.org/10.3322/caac.21834>
- Carriero, A., Groenhoff, L., Vologina, E., Basile, P., & Albera, M. (2024). Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics*, 14(8), 848. <https://doi.org/10.3390/diagnostics14080848>
- Chetry, M., Feng, R., Babar, S., Sun, H., Zafar, I., Mohany, M., Afridi, H. I., Khan, N. U., Ali,

- I., Shafiq, M., & Khan, S. (2025). Early detection and analysis of accurate breast cancer for improved diagnosis using deep supervised learning for enhanced patient outcomes. *PeerJ Computer Science*, 11, e2784. <https://doi.org/10.7717/peerj-cs.2784>
- Gao, Y., Lin, J., Zhou, Y., & Lin, R. (2023). The application of traditional machine learning and deep learning techniques in mammography: a review. *Frontiers in Oncology*, 13. <https://doi.org/10.3389/fonc.2023.1213045>
- Gautama, W. (2022). Breast Cancer in Indonesia in 2022: 30 Years of Marching in Place. *Indonesian Journal of Cancer*, 16(1), 1. <https://doi.org/10.33371/ijoc.v16i1.920>
- Ghasemi, A., Hashtarkhani, S., Schwartz, D. L., & Shaban-Nejad, A. (2024). Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innovation*, 3(5). <https://doi.org/10.1002/cai2.136>
- Grinsztajn, L., Oyallon, E., Varoquaux, G., Grinsztajn, L., Oyallon, E., Varoquaux, G., & Oyallon, E. (2022). *Why do tree-based models still outperform deep learning on tabular data ? To cite this version : HAL Id : hal-03723551 Why do tree-based models still outperform deep learning on tabular data ?*
- Hutajulu, S. H., Prabandari, Y. S., Bintoro, B. S., Wiranata, J. A., Widiastuti, M., Suryani, N. D., Saptari, R. G., Taroen-Hariadi, K. W., Kurnianda, J., Purwanto, I., Hardianti, M. S., & Allsop, M. J. (2022). Delays in the presentation and diagnosis of women with breast cancer in Yogyakarta, Indonesia: A retrospective observational study. *PLOS ONE*, 17(1), e0262468. <https://doi.org/10.1371/journal.pone.0262468>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *IJCAI*.
- Kumar, A., Saini, R., & 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). <https://doi.org/10.18280/ts.410110>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python*. <http://arxiv.org/abs/1201.0490>
- Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84–90. <https://doi.org/10.1016/j.inffus.2021.11.011>
- Sirait, L. M. F., Hamajima, N., Suzuki, Y., Wahyuningsih, E. S., Oktavia, D., Widyastuti, Inthaphatha, S., Nishino, K., & Yamamoto, E. (2022). Factors associated with positive cancer screening for the uterine cervix and breast in Jakarta Province, Indonesia: a cross-sectional study. *BMC Cancer*, 22(1), 1309. <https://doi.org/10.1186/s12885-022-10381-1>
- Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993). *Nuclear feature extraction for breast tumor diagnosis*. (R. S. Acharya & D. B. Goldgof, Eds.; pp. 861–870). <https://doi.org/10.1117/12.148698>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Vargas-Palacios, A., Sharma, N., & Sagoo, G. S. (2023). Cost-effectiveness requirements for implementing artificial intelligence technology in the Women's UK Breast Cancer Screening service. *Nature Communications*, 14(1), 6110. <https://doi.org/10.1038/s41467-023-41754-0>
- Zhang, X., Shao, W., Qiu, M., Xiao, C. (2025). Advanced deep learning and transfer learning approaches for breast cancer classification using advanced multi-line classifiers and datasets with model. *Peerj.ComX Zhang, W Shao, M Qiu, C Xiao, L MaPeerJ Computer Science*, 2025•peerj.Com. <https://doi.org/10.7717/PEERJ-CS.2951>

