

Enhanced Lung Cancer Detection Using ANN with Random Oversampling, RFE-Based Feature Selection, and GridSearchCV Hyperparameter Tuning

Nurwafiqah*¹, M. Yudi Al Fiqran², Annisa Nurul Puteri³, Muhammad Arafah⁴,
Tatik Maslihatin⁵, A. Sumardin⁶

^{1,2,4,5,6}Informatics, Universitas Teknologi Akba Makassar, Indonesia

³Computer and Network Engineering, Politeknik Negeri Ujung Pandang, Indonesia

Email: nurwafiqah22@mhs.unitama.ac.id

Received : Oct 26, 2025; Revised : Feb 7, 2026; Accepted : Feb 9, 2026; Published : Apr 20, 2026

Abstract

Amid the most predominant mortality factors on a global scale, Lung cancer constitutes one of the most significant oncological burdens, chiefly because most patients receive a diagnosis only at later stages. The limitations of conventional diagnostic approaches underscore the urgent need for artificial intelligence-based detection systems that can improve both diagnostic accuracy and efficiency. This study aims to develop a lung cancer prediction model using an Artificial Neural Network (ANN) optimized through an integrated strategy that includes data preprocessing, class balancing via Random Oversampling (ROS), feature selection using Recursive Feature Elimination (RFE), and hyperparameter tuning with Grid Search. The evaluation of model effectiveness employs accuracy, precision, recall, F1-score, along with a confusion matrix. Experimental results demonstrate an accuracy of 98%, with average precision, recall, and F1-score values of 0.95. Statistical validation using McNemar's test confirms a significant performance improvement over the baseline model ($\chi^2 = 18.05$, $p < 0.001$), accompanied by a large effect size (Cohen's $h = 0.82$). Furthermore, the model exhibits balanced performance in identifying both lung cancer and non-cancer cases, reflecting the effectiveness of the data balancing and feature selection strategies. These findings suggest that the optimized ANN model has strong potential as a foundation for a medical decision support system for early lung cancer detection, contributing to more reliable diagnoses and more accurate clinical decision-making.

Keywords : Artificial Neural Network, Early Detection, Feature Selection, Lung Cancer Detection, McNemar's Test, Random Oversampling, Recursive Feature Elimination

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. INTRODUCTION

Kanker paru-paru merupakan salah satu keganasan dengan tingkat mortalitas tertinggi di dunia [1]. Data *International Agency for Research on Cancer* (IARC) tahun 2022 menunjukkan bahwa kanker paru-paru menyebabkan sekitar 1,8 juta kematian atau 18,7% dari total kematian akibat kanker secara global [2]. Tingginya angka mortalitas ini terutama disebabkan oleh keterlambatan diagnosis, di mana sebagian besar pasien terdeteksi pada stadium lanjut [3]. Studi literatur menunjukkan bahwa sekitar 75% kasus kanker paru-paru jenis *Non-Small Cell Lung Cancer* (NSCLC) baru terdiagnosis pada stadium III atau IV. Kondisi ini berkorelasi langsung dengan rendahnya tingkat keberlangsungan hidup lima tahun pasien, yang secara global masih berkisar antara 10% hingga 20% [4], [5].

Di Indonesia, situasi tidak kalah mengkhawatirkan. Kanker paru-paru tercatat sebagai salah satu penyebab utama kematian akibat kanker dengan berkontribusi sebesar 12,6% terhadap total mortalitas terkait keganasan. Proyeksi epidemiologis menunjukkan bahwa insiden kanker paru-paru berpotensi meningkat hampir dua kali lipat dalam beberapa tahun ke depan, dipicu oleh tingginya konsumsi tembakau sejak usia muda serta paparan zat karsinogen di lingkungan kerja dan tempat tinggal [6]. Laporan *Global Cancer Observatory (GLOBOCAN)* juga menunjukkan peningkatan hampir 30% kasus

kanker paru-paru di Indonesia pada periode 2018-2022, yang menegaskan urgensi penguatan sistem deteksi dini yang lebih efektif.

Permasalahan utama dalam penanganan kanker paru-paru terletak pada keterbatasan metode diagnosis konvensional, seperti pemeriksaan radiologi dan biopsi yang sering kali kurang sensitif dalam mendeteksi kanker stadium awal [7]. Beberapa studi melaporkan bahwa sensitivitas *CT-scan* pada stadium dini masih terbatas, dengan tingkat false negative mencapai 20%-25% [8], [9]. Selain memerlukan waktu yang relatif lama, metode ini juga bergantung pada ketersediaan peralatan canggih, interpretasi subjektif tenaga medis, serta biaya pemeriksaan yang tinggi, yang menjadi kendala signifikan terutama di negara berkembang [10].

Perkembangan *machine learning* hadir untuk sebagai peluang baru terutama meningkatkan efektivitas deteksi dini kanker paru-paru. *Artificial Neural Network (ANN)* cukup sering digunakan terutama dalam mengklasifikasikan data karena dianggap mampu memodelkan relasi non-linier dan mengenali pola kompleks pada data klinis [11]. Studi terkini menunjukkan bahwa pemanfaatan ANN berbasis data klinis mampu memberikan performa deteksi yang lebih baik dibandingkan pendekatan berbasis data citra saja [12]. Meskipun demikian, kinerja ANN sangat bergantung pada kualitas dan karakteristik data yang diterapkan.

Tantangan utama dalam pemodelan data medis adalah ketidakseimbangan distribusi kelas, di mana jumlah pasien kanker jauh lebih sedikit dibandingkan pasien non-kanker. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas dan menurunkan sensitivitas terhadap kasus minoritas yang justru krusial dalam deteksi dini [13], [14]. Salah satu teknik yang banyak digunakan untuk mengatasi permasalahan ini adalah *Random Oversampling (ROS)*, yang menyeimbangkan distribusi kelas dengan menggandakan data pada kelas minoritas. Beberapa studi menunjukkan bahwa ROS mampu meningkatkan nilai *recall* secara signifikan pada tugas klasifikasi medis [15], [16].

Selain ketidakseimbangan kelas, kompleksitas fitur pada data medis juga menjadi tantangan tersendiri. Keberadaan atribut yang tidak relevan atau redundan dapat meningkatkan risiko *overfitting* dan menurunkan performa model [17]. *Recursive Feature Elimination (RFE)* merupakan teknik seleksi fitur yang bekerja secara iteratif dengan mengevaluasi kontribusi setiap fitur dan mengeliminasi fitur dengan bobot terendah [18]. Penelitian sebelumnya melaporkan bahwa RFE mampu mereduksi jumlah fitur hingga 50% tanpa menurunkan akurasi model secara signifikan [19], [20]. Integrasi RFE dengan ANN tidak hanya meningkatkan efisiensi dan akurasi sistem klasifikasi, tetapi juga mengurangi kompleksitas komputasi, sehingga model lebih efisien diterapkan pada dataset medis berfitur tinggi.

Pendekatan yang menggabungkan penyeimbangan kelas dan seleksi fitur terbukti efektif meningkatkan performa model. Studi deteksi stoke [21] menunjukkan bahwa SVM dengan ROS mencapai *recall* 100% dan akurasi 98.56%, melampaui *Borderline-SMOTE*. Penelitian kanker serviks [22] juga menunjukkan bahwa *Decision Tree* dengan RFE dan *SMOTETomek* meraih akurasi 98.82% dan *recall* 100%. Selain itu, studi [23] membuktikan bahwa penerapan RFE meningkatkan akurasi ANN hingga 92.77%, melampaui *AdaBoost*. Namun, sebagian besar penelitian tersebut tidak mengaplikasikan pendekatan serupa pada konteks kanker paru-paru.

Penelitian sebelumnya juga telah mengeksplorasi berbagai pendekatan *machine learning* untuk deteksi kanker paru-paru. Studi [24] menunjukkan *Logistic Regression* mencapai akurasi 92,2%, diikuti *Random Forest* dan *Neural Network* dengan 90,6%. Penelitian komparatif pada kanker payudara [25] menunjukkan ANN dan SVM mencapai akurasi 99%, sementara *Decision Tree* hanya 95%. Meski akurasi relatif tinggi, penelitian-penelitian tersebut umumnya tidak menangani masalah ketidakseimbangan data dan tidak melakukan optimasi seleksi fitur secara sistematis.

Berdasarkan tinjauan literatur, belum ditemukan penelitian yang secara komprehensif mengintegrasikan *Artificial Neural Network*, *Random Oversampling*, dan *Recursive Feature*

Elimination dalam satu pendekatan terpadu untuk deteksi kanker paru-paru berbasis data klinis. Sebagian studi hanya berfokus pada peningkatan akurasi model tanpa mempertimbangkan bias kelas dan kompleksitas fitur secara simultan. Oleh karena itu, kebaruan penelitian ini terletak pada penggabungan tiga pendekatan tersebut untuk menghasilkan model yang tidak hanya akurat, tetapi juga seimbang, efisien, dan robust terhadap karakteristik data medis yang kompleks.

Berdasarkan gap tersebut, penelitian ini ditargetkan dapat meningkatkan akurasi deteksi kanker paru-paru melalui integrasi ANN sebagai algoritma klasifikasi, ROS untuk penyeimbangan distribusi kelas, dan RFE sebagai teknik seleksi fitur. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score untuk memastikan keseimbangan performa antar kelas. Hasil penelitian ini dimaksudkan dapat menjadi dasar pengembangan sistem pendukung keputusan klinis berbasis kecerdasan buatan yang adaptif dan aplikatif untuk mendukung deteksi dini kanker paru-paru.

2. METHOD

Penelitian ini bertujuan mengembangkan model prediksi kanker paru-paru yang akurat dan andal dengan memanfaatkan kombinasi *Artificial Neural Network (ANN)*, *Random Oversampling (ROS)*, dan *Recursive Feature Elimination (RFE)*. Tahapan penelitian diawali dengan proses *preprocessing* yang meliputi *data cleaning*, pembagian data menjadi *training data* dan *test data*, *label encoding*, serta normalisasi fitur. Selanjutnya, ROS diterapkan pada *training data* untuk menyeimbangkan distribusi kelas, kemudian RFE digunakan untuk menyeleksi fitur paling relevan guna mengurangi kompleksitas data. Model ANN digunakan sebagai algoritma klasifikasi utama untuk memodelkan hubungan non-linear antar fitur. Proses pelatihan dilanjutkan dengan *hyperparameter tuning*, diikuti analisis *explainability* untuk menginterpretasikan kontribusi setiap fitur terhadap prediksi. Kinerja model kemudian dievaluasi menggunakan metrik performa dan pengujian statistik. Alur metodologi penelitian ditunjukkan pada [Figure 1](#).

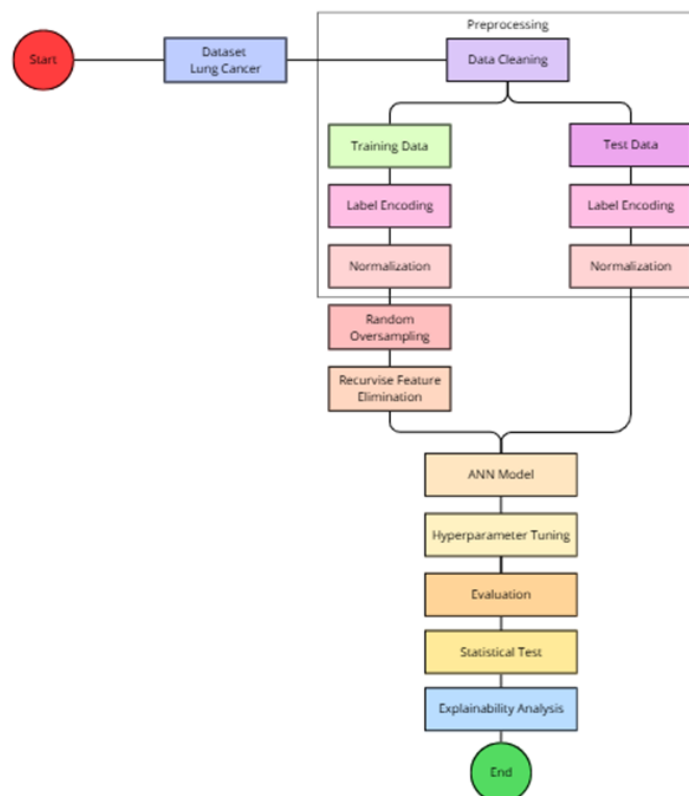


Figure 1. Skema Penelitian

2.1. Lung Cancer Dataset

Dataset yang diimplementasikan dalam penelitian ini didapat dari sumber publik Kaggle, yakni *Lung Cancer Dataset*, mengandung 309 entri dengan 16 fitur yang merepresentasikan faktor risiko klinis dan perilaku pasien. Dataset ini dipilih karena relevan untuk pengembangan model prediksi berbasis *machine learning*. Rincian atribut dan tipe datanya disajikan pada [Table 1](#).

Table 1. Dataset

No	Fitur	Data Type
1	Gender	Categorical
2	Age	Numeric
3	Smoking	Numeric
4	Yellow Fingers	Numeric
5	Anxiety	Numeric
6	Peer Pressure	Numeric
7	Chronic Disease	Numeric
8	Fatigue	Numeric
9	Allergy	Numeric
10	Wheezing	Numeric
11	Alcohol Consuming	Numeric
12	Coughing	Numeric
13	Shortness of Breath	Numeric
14	Swallowing Difficulty	Numeric
15	Chest Pain	Numeric
16	Lung Cancer	Categorical

2.2. Preprocessing

Tahap *Preprocessing* merupakan langkah penting untuk menyiapkan data mentah agar lebih terstruktur dan sesuai dengan kebutuhan analisis. Proses bertujuan meningkatkan kualitas data dengan mengurangi fitur yang tidak relevan atau redundan sehingga model yang dibangun menjadi lebih efisien dan akurat [26]. Tahapan *preprocessing* meliputi pembersihan data (*data cleaning*), pembagian data menjadi data latih (*training data*) dan data uji (*test data*), transformasi fitur kategorikal menggunakan *label encoding*, serta normalisasi fitur numerik untuk menyetarakan skala data sebelum proses pelatihan model.

2.2.1. Data Cleaning

Tahap *Data cleaning* merupakan proses menghapus data yang tidak valid, seperti nilai kosong, kesalahan input, data tidak relevan, dan data duplikat. Proses ini penting karena dapat meningkatkan kualitas data untuk analisis *data mining* sekaligus mengurangi jumlah kompleksitas data yang diproses [27]. Pada tahap ini, seluruh fitur diperiksa untuk memastikan tidak ada nilai kosong dan inkonsistensi label. Selanjutnya, dilakukan deteksi serta penghapusan data duplikat guna mencegah bias distribusi dan redundansi informasi. Dengan demikian, tidak diperlukan proses imputasi lanjutan dan kualitas *dataset* menjadi lebih representatif untuk proses pelatihan model.

2.2.2. Data Splitting

Tahap *data splitting* dilakukan dengan membagi *dataset* menjadi dua. Data penelitian dipartisi menjadi data latih (*training data*) dan data uji (*test data*). Data latih digunakan sebagai dasar pengembangan dan pelatihan model, sementara data uji difungsikan untuk mengukur tingkat akurasi model secara objektif. Secara umum, pembagian data dapat menggunakan rasio 80:20, 70:30, atau 60:40 [28]. Dalam penelitian ini, digunakan rasio 70:30, dimana 70% data dialokasikan sebagai data

latih dan 30% sebagai data uji. Rasio tersebut dipilih untuk memastikan ketersediaan data latih yang cukup selama proses pembelajaran ANN, sekaligus mempertahankan data uji yang representatif untuk evaluasi model.

2.2.3. Label Encoding

Tahap *Label encoding* merupakan metode transformasi untuk mengubah fitur katagorikal menjadi representasi numerik agar dapat diproses oleh model *machine learning*. Proses ini dilakukan dengan mengganti setiap kategori menjadi nilai bilangan bulat dari 0 hingga $n-1$ sesuai jumlah kelas yang tersedia. Transformasi ini diperlukan karena algoritma ANN hanya dapat mengolah data dalam bentuk numerik, sehingga hasil pengkodean memungkinkan data diproses secara kuantitatif selama tahap pelatihan model [29].

2.2.4. Normalization

Pada tahap *normalization*, dilakukan penskalaan data menggunakan metode *min-max normalization* agar seluruh fitur berada dalam rentang 0 hingga 1. Proses ini bertujuan mengurangi perbedaan skala antar fitur yang dapat memengaruhi kinerja model, khususnya pada ANN yang sensitif terhadap variasi nilai input [30]. Persamaan matematis yang digunakan adalah:

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(x)} \quad (1)$$

di mana X_{new} merepresentasikan nilai hasil normalisasi, X adalah nilai asli sebelum transformasi, $\max(X)$ menunjukkan nilai maksimum, dan $\min(X)$ adalah nilai minimum pada suatu fitur dalam *dataset*. Dengan demikian, setiap nilai fitur dipetakan secara proporsional ke dalam rentang $[0, 1]$, sehingga data menjadi lebih seragam dan siap digunakan untuk proses analisis maupun pelatihan model.

2.3. Handling Imbalanced Data (ROS)

Penanganan ketidakseimbangan kelas pada *dataset* merupakan masalah penting karena jumlah sampel pada satu kelas sering kali lebih dominan dibandingkan kelas lainnya, sehingga berpotensi menyebabkan bias pada model klasifikasi. Untuk mengatasi permasalahan tersebut, diterapkan metode *Random Oversampling* (ROS) diterapkan pada data latih (*training data*) guna mencegah terjadinya data leakage. ROS bekerja dengan mereplikasi sampel dari kelas minoritas secara acak hingga distribusi kelas menjadi lebih seimbang dengan kelas mayoritas [31]. Metode ini dipilih dibandingkan teknik sintesis seperti *SMOTE* karena ukuran *dataset* yang relatif kecil, sehingga ROS dinilai lebih stabil serta tidak berpotensi menghasilkan *noise* tambahan akibat pembuatan sampel sintesis.

2.4. Feature Selection (RFE)

Recursive Feature Elimination (RFE) merupakan teknik *feature selection* yang bekerja secara bertahap dengan mengeliminasi fitur yang memiliki kontribusi paling kecil terhadap kinerja model. Kontribusi tersebut ditentukan berdasarkan bobot atau tingkat kepentingan fitur yang dihasilkan selama proses pelatihan model. Pendekatan ini bermanfaat dalam menangani *high-dimensional data* karena mampu mengurangi kompleksitas komputasi sekaligus mempertahankan kemampuan prediksi [32]. Secara matematis, misalkan kumpulan fitur awal dinyatakan sebagai $X = [x_1, x_2, x_3, \dots, x_n]$ dengan target Y . Proses seleksi dilakukan secara iteratif melalui pelatihan model dan penghapusan fitur dengan skor terendah hingga diperoleh subset fitur yang paling relevan [26]. Dalam penelitian ini, RFE diterapkan setelah proses *oversampling* untuk memastikan bahwa fitur yang dipilih benar-benar merepresentasikan distribusi data yang telah seimbang sebelum digunakan pada tahap penelitian model klasifikasi.

2.5. ANN Model

Artificial Neural Network (ANN) merupakan metode pembelajaran dalam bidang *artificial intelligence* yang meniru prinsip kerja otak manusia [33]. ANN tersusun atas neuron-neuron buatan yang saling terhubung dan terorganisasi dalam beberapa lapisan, yaitu *input layer*, satu atau lebih *hidden layer*, dan *output layer*. Data masukan diproses melalui setiap lapisan menggunakan bobot tertentu untuk mengekstraksi pola serta hubungan non-linear, hingga menghasilkan keluaran berupa prediksi atau klasifikasi [34].

Dalam penelitian ini, ANN digunakan untuk mengolah fitur dari *dataset* kanker paru-paru. Proses pelatihan dilakukan secara iteratif dengan memperbarui bobot antar neuron guna meminimalkan kesalahan prediksi dan meningkatkan kinerja model. Arsitektur yang diterapkan terdiri atas satu *input layer*, dua *hidden layer* dengan fungsi aktivasi non-linear, serta satu *output layer* untuk menghasilkan keputusan klasifikasi, sebagaimana ditunjukkan pada [Figure 2](#).

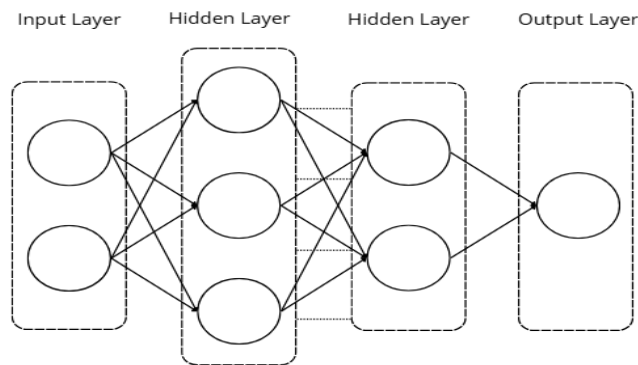


Figure 2. Model ANN

2.6. Hyperparameter Tuning

Pada penelitian ini, proses *hyperparameter tuning* pada model ANN dilakukan menggunakan *grid search*. Metode ini dipilih karena mampu mengevaluasi seluruh kombinasi parameter secara sistematis serta menghasilkan proses yang transparan dan *reproducible* [35], meskipun membutuhkan waktu komputasi yang relatif tinggi, terutama pada ruang parameter yang luas. Parameter yang dioptimalkan meliputi jumlah neuron pada *hidden layer* (*hidden_layer_sizes*), fungsi aktivasi (*activation*), algoritma optimisasi (*solver*), koefisien regularisasi (*alpha*), strategi laju pembelajaran (*learning_rate*), nilai awal laju pembelajaran (*learning_rate_init*), batas toleransi konvergensi (*tol*), serta jumlah iterasi maksimum (*max_iter*). Kombinasi parameter terbaik yang diperoleh dari proses pencarian ini digunakan sebagai konfigurasi akhir untuk meningkatkan kinerja dan kemampuan generalisasi model ANN.

2.7. Evaluasi

Evaluasi model klasifikasi dilakukan dengan *confusion matrix*, yaitu sebuah metode yang melakuakn komparasi hasil prediksi model dengan nilai aktual pada *dataset*. Confusion matrix terdiri atas empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), sebagaimana ditunjukkan pada [Table 2](#).

Table 2. Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Berdasarkan komponen tersebut, kinerja model diukur menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, yang dirumuskan pada Formula (2)-(5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Accuracy mengukur ketepatan prediksi secara keseluruhan. *Precision* menunjukkan proporsi prediksi positif yang benar, sedangkan *recall* mengukur kemampuan model dalam mendeteksi seluruh kasus positif. *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall* yang digunakan untuk menilai keseimbangan kedua metrik tersebut.

2.8. Statistical Significance Testing

Perbedaan performa dua model klasifikasi pada *dataset* yang sama diuji menggunakan *McNemar's test* dengan tingkat signifikansi $\alpha = 0,0005$. Analisis dilakukan melalui tabel kontingensi 2x2 yang merepresentasikan distribusi prediksi benar dan salah dari kedua model. Pengujian ini membandingkan hasil prediksi berpasangan dengan memfokuskan pada jumlah ketidaksesuaian antar model, yaitu kasus ketika model pertama benar dan model kedua salah (*b*) serta sebaliknya (*c*). Statistik uji dihitung menggunakan:

$$\chi^2 = \frac{(b - c)^2}{b + c} \quad (6)$$

Formula 6 mengukur selisih kuadrat antara kedua jenis kesalahan (*b - c*) yang dinormalisasi oleh total ketidaksesuaian (*b + c*). Nilai yang lebih besar menunjukkan perbedaan kesalahan yang semakin signifikan. Statistik ini mengikuti distribusi χ^2 dengan satu derajat kebebasan, dan perbedaan performa dinyatakan signifikan apabila *p-value* $< \alpha$ [36]. Selain signifikansi statistik, *Cohen's h* digunakan untuk mengukur besarnya perbedaan praktis antara dua proporsi performa (p_1 dan p_2):

$$h = 2(\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2})) \quad (7)$$

Formula 7 mentransformasikan proporsi ke skala sudut untuk menstabilkan varians, sehingga menghasilkan ukuran *effect size* yang tidak dipengaruhi ukuran sampel. Nilai *h* diinterpretasikan sebagai kecil (0,2), sedang (0,5), dan besar (0,8) [37].

2.9. Explainability Analysis (SHAP)

Untuk meningkatkan transparansi dan interpretabilitas model ANN, penelitian ini mengintegrasikan metode *SHapley Additive exPlanations* (SHAP), yaitu pendekatan *additive feature attribution* yang didasarkan pada konsep *Shapley value* dari teori permainan kooperatif. SHAP mentransformasikan model *non-linear* yang kompleks menjadi model penjelasan aditif dengan menghitung kontribusi masing-masing fitur terhadap keluaran prediksi. Pendekatan ini digunakan untuk analisis *global* guna mengidentifikasi fitur yang paling berpengaruh terhadap kinerja model secara keseluruhan, serta analisis *local* untuk menjelaskan prediksi pada setiap sampel secara individual. Dengan demikian, keputusan model ANN dapat diinterpretasikan secara lebih transparan dan konsisten terhadap faktor-faktor yang memengaruhi prediksi kanker paru-paru [38].

3. RESULT

Dataset yang digunakan dalam penelitian ini terdiri atas 309 entri pasien dengan berbagai fitur klinis dan faktor risiko yang relevan terhadap kanker paru-paru. Setiap entri merepresentasikan kondisi individual pasien melalui sejumlah fitur yang menggambarkan gejala, riwayat medis, maupun kebiasaan gaya hidup. Fitur klinis mencakup gejala seperti *coughing*, *chest pain*, *fatigue*, *shortness of breath*, *wheezing*, dan *swallowing difficulty*. Faktor gaya hidup yang terekam meliputi *smoking*, *alcohol consuming*, serta *peer pressure*. Selain itu, terdapat informasi medis tambahan seperti *chronic disease*, *allergy*, *anxiety*, dan tanda klinis seperti *yellow fingers*. Label target dalam dataset terbagi menjadi dua kategori yaitu *Yes* untuk pasien yang terdiagnosis kanker paru-paru dan *No* untuk pasien yang tidak terdiagnosis. [Table 3](#) menyajikan ringkasan fitur yang dianalisis.

Table 3. Dataset Lung Cancer

No	Fitur	Description	Value
1	Gender	Jenis kelamin pasien.	M: male, F: female.
2	Age	Usia pasien dalam satuan tahun.	Numerical
3	Smoking	Menunjukkan apakah pasien merupakan perokok atau bukan.	2: yes, 1: no.
4	Yellow Fingers	Menunjukkan apakah pasien memiliki jari yang menguning.	2: yes, 1: no.
5	Anxiety	Menunjukkan apakah pasien mengalami gejala kecemasan.	2: yes, 1 no.
6	Peer Pressure	Menunjukkan apakah pasien terpengaruh oleh tekanan dari lingkungan sekitar.	2: yes, 1 no.
7	Chronic Disease	Menunjukkan apakah pasien menderita penyakit kronis.	2: yes, 1 no.
8	Fatigue	Menunjukkan apakah pasien mengalami kelelahan.	2: yes, 1 no.
9	Allergy	Menunjukkan apakah pasien memiliki alergi.	2: yes, 1 no.
10	Wheezing	Menunjukkan apakah pasien mengalami mengi.	2: yes, 1 no.
11	Alcohol Consuming	Menunjukkan apakah pasien mengonsumsi alkohol.	2: yes, 1 no.
12	Coughing	Menunjukkan apakah pasien mengalami batuk.	2: yes, 1 no.
13	Shortness of Breath	Menunjukkan apakah pasien mengalami sesak napas.	2: yes, 1 no.
14	Swallowing Difficulty	Menunjukkan apakah pasien mengalami kesulitan menelan.	2: yes, 1 no.
15	Chest Pain	Menunjukkan apakah pasien mengalami nyeri dada.	2: yes, 1 no.
16	Lung Cancer	Menunjukkan status diagnosis pasien.	Yes: terdiagnosis kanker, No: tidak terdiagnosis.

Pada tahap awal penelitian, proses pembersihan data menjadi langkah fundamental yang menentukan kualitas hasil pemodelan. Dataset awal terdiri atas 309 entri pasien, namun hasil pemeriksaan mendetail menemukan adanya 33 entri duplikat. Keberadaan data yang berulang ini berpotensi menimbulkan distorsi karena algoritma dapat memberikan bobot berlebih pada pola tertentu yang sebetulnya tidak merepresentasikan kondisi nyata. Jika dibiarkan, bias ini akan menyebabkan hasil hasil klasifikasi cenderung berpihak pada informasi yang semu. Oleh karena itu, seluruh entri duplikat dihapus sehingga jumlah data berkurang menjadi 276 entri yang lebih akurat dan representatif. Proses

ini memastikan bahwa dataset yang digunakan tidak hanya bersih tetapi juga konsisten. Validasi tambahan dilakukan untuk memeriksa kemungkinan adanya nilai kosong ataupun anomali distribusi, langkah yang sangat krusial mengingat data medis seringkali mengandung ketidaksempurnaan pencatatan. Dengan demikian, data yang diperoleh telah memenuhi standar integritas yang dibutuhkan untuk mendukung proses pembelajaran mesin secara andal.

Setelah proses pembersihan data selesai, langkah berikutnya adalah mempersiapkan dataset untuk pelatihan dan evaluasi model. Untuk menjaga validitas hasil, dataset dipisahkan sejak awal menjadi dua bagian, yaitu 70% sebagai data latih dan 30% sebagai data uji. Strategi ini dilakukan untuk menjaga independensi data uji sehingga evaluasi performa model benar-benar mencerminkan kemampuan prediksi terhadap data baru yang tidak pernah dilihat sebelumnya. Langkah ini penting untuk mencegah terjadinya kebocoran informasi yang dapat mengurangi validitas hasil eksperimen. Selain itu, dilakukan pula transformasi awal agar data siap digunakan oleh algoritma. Fitur bertipe kategorikal dikonversi ke bentuk numerik menggunakan *label encoding*, karena model berbasis machine learning, termasuk ANN, hanya mampu mengolah data numerik. Agar setiap fitur dapat berkontribusi secara seimbang, normalisasi *min-max* diterapkan untuk menyeragamkan skala antar variabel. Tanpa normalisasi, fitur dengan rentang nilai lebih besar akan mendominasi pembelajaran, sehingga variabel lain yang sebenarnya relevan menjadi terabaikan. Dengan skala yang seragam, setiap variabel dapat berkontribusi secara proporsional dalam pembentukan pola prediksi.

Meskipun data telah diproses dengan baik, analisis distribusi kelas memperlihatkan bahwa dataset masih menghadapi tantangan serius berupa ketidakseimbangan kelas. Dari total 193 entri latih yang digunakan (70% dari keseluruhan data), terdapat 168 pasien dengan diagnosis kanker paru-paru dan hanya 25 pasien non-kanker. Ketimpangan ini sangat berisiko membuat model cenderung lebih sering memprediksi kelas mayoritas, sehingga sensitivitas terhadap kelas minoritas menurun drastis. Dalam konteks medis, kondisi seperti ini berbahaya karena dapat menyebabkan pasien non-kanker justru diklasifikasikan secara keliru. Untuk mengatasi hal tersebut, penelitian ini menerapkan strategi ROS. Metode ini bekerja dengan menambahkan salinan data dari kelas minoritas agar jumlahnya mendekati kelas mayoritas. Keunggulan ROS terletak pada kesederhanaan sekaligus kemampuannya mempertahankan karakteristik asli data tanpa menghasilkan variasi sintetis yang berlebihan.

Perubahan distribusi setelah penerapan ROS sangat signifikan. [Figure 3](#) menunjukkan bahwa jumlah kasus positif dan negatif menjadi lebih seimbang sehingga model ANN memiliki peluang yang lebih besar untuk belajar dari pola kedua kelas secara proporsional. Dengan kondisi ini, sensitivitas model dalam mengenali pasien sehat meningkat, sementara akurasi keseluruhan tetap terjaga. Hal ini memperlihatkan bahwa penyelesaian masalah distribusi kelas merupakan salah satu kunci penting dalam membangun sistem klasifikasi medis yang andal. Dengan dataset yang lebih representatif, proses pemodelan berikutnya dapat berjalan lebih adil tanpa bias berlebihan terhadap kelas tertentu.

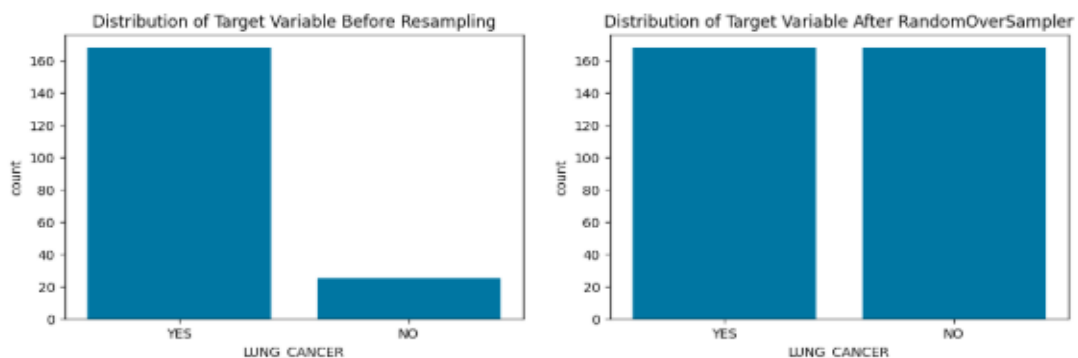


Figure 3. Sebelum dan Sesudah Resampling

Setelah distribusi kelas diperbaiki, fokus analisis diarahkan pada pemilihan fitur yang relevan. Tidak semua variabel dalam dataset memberikan kontribusi signifikan terhadap hasil prediksi. Sebagian diantaranya mungkin redundan, kurang informatif, atau bahkan menimbulkan *noise*. Untuk memastikan bahwa hanya fitur yang paling penting digunakan dalam pemodelan, penelitian ini menerapkan metode RFE dengan estimator *Logistic Regression*. RFE bekerja secara iteratif dengan menilai pentingnya setiap fitur berdasarkan kontribusinya terhadap performa model, lalu menghapus fitur dengan bobot terendah hingga tersisa kumpulan variabel yang paling informatif.

Sebagai langkah awal, seluruh 15 fitur dianalisis menggunakan mekanisme ranking RFE. [Table 4](#) menyajikan hasil lengkap berupa daftar seluruh fitur beserta status seleksi dan urutan peringkatnya. Terlihat bahwa sepuluh fitur memperoleh ranking tertinggi dengan nilai kontribusi paling kuat, sementara lima fitur lainnya, yaitu *Gender*, *Age*, *Smoking*, *Wheezing*, dan *Shortness of Breath*, berada pada posisi bawah dengan ranking yang lebih besar. Penyajian tabel ini penting untuk memberikan gambaran menyeluruh mengenai alasan mengapa sebagian fitur dieliminasi dari proses pemodelan, sekaligus memastikan bahwa keputusan seleksi fitur dilakukan secara terukur dan transparan.

Table 4. Peringkat Fitur dari 15 Atribut Menggunakan RFE

No	Fitur	Selected	Ranking
1	Yellow Fingers	Yes	1
2	Chronic Disease	Yes	1
3	Peer Pressure	Yes	1
4	Anxiety	Yes	1
5	Fatigue	Yes	1
6	Coughing	Yes	1
7	Alcohol Consuming	Yes	1
8	Allergy	Yes	1
9	Chest Pain	Yes	1
10	Swallowing Difficulty	Yes	1
11	Smoking	No	2
12	Wheezing	No	3
13	Age	No	4
14	Shortness of Breath	No	5
15	Gender	No	6

Analisis lebih lanjut terhadap enam fitur yang dieliminasi menunjukkan bahwa kontribusinya terhadap performa model relatif kecil. Misalnya, atribut *Gender* dan *Age* memiliki distribusi yang kurang seimbang dalam dataset sehingga tidak membentuk pola yang jelas untuk membedakan pasien kanker paru-paru dengan non-kanker. Fitur *Smoking* dan *Wheezing*, meskipun secara klinis sering dikaitkan dengan gangguan paru-paru, menunjukkan korelasi yang rendah pada dataset ini karena sebagian besar informasinya telah tercermin pada fitur lain seperti *yellow fingers* atau *coughing*. Selain itu, *Shortness of Breath* memiliki tumpang tindih informasi dengan *fatigue* dan *chest pain*, sehingga menimbulkan redundansi. Eliminasi fitur-fitur membantu mengurangi kompleksitas data sekaligus meminimalkan risiko overfitting. Jika kelima fitur tersebut tetap dipertahankan, model berpotensi menyerap *noise* yang tidak relevan dan menghasilkan pola prediksi yang kurang stabil.

Kondisi ini akan menyulitkan proses interpretasi medis, sebab tenaga kesehatan dapat menerima hasil yang bias atau ambigu. Dengan hanya mempertahankan sepuluh fitur yang paling berpengaruh, proses pembelajaran menjadi lebih fokus, efisien, dan akurat. Oleh karena itu, seleksi fitur melalui RFE bukan sekadar prosedur teknis, melainkan langkah penting untuk menjaga keseimbangan antara akurasi model dan relevansi klinis. Temuan ini menjadi dasar dalam menyaring variabel yang paling konsisten

berkontribusi terhadap performa prediksi sehingga model hanya memanfaatkan fitur-fitur yang benar-benar dominan.

Berdasarkan penyaringan tersebut, hasil proses RFE menunjukkan bahwa hanya sepuluh fitur yang dipertahankan dari keseluruhan dataset. Fitur *chronic disease* muncul sebagai variabel dengan pengaruh terbesar, sejalan dengan fakta medis bahwa penyakit kronis sering kali memperburuk kondisi paru-paru. Sebaliknya, fitur seperti *peer pressure* dan *chest pain* menunjukkan pengaruh lebih kecil, namun keberadaannya tetap dipertahankan karena dapat menambah perspektif dalam mendeteksi pola yang kompleks. [Figure 4](#) memperlihatkan hierarki kontribusi variabel, sementara [Table 5](#) menegaskan besarnya peran masing-masing fitur.

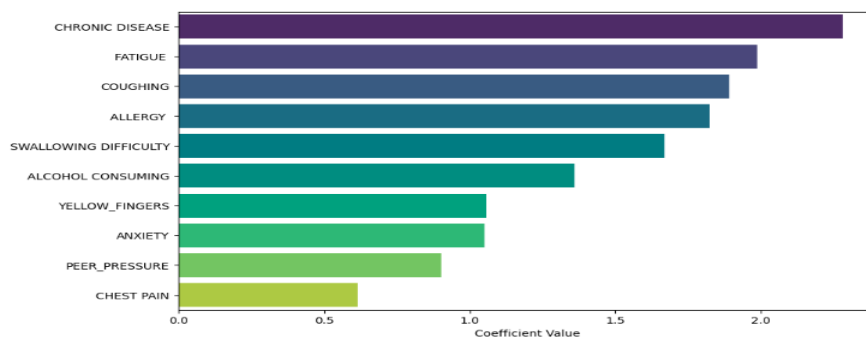


Figure 4. Feature Selection

Table 5. Feature Selection

Selected Feature	Importance
Chronic Disease	2.28
Fatigue	1.99
Coughing	1.89
Allergy	1.82
Swallowing Difficulty	1.67
Alcohol Consuming	1.36
Yellow Fingers	1.05
Anxiety	1.05
Peer Pressure	0.90
Chest Pain	0.62

Pemilihan fitur melalui RFE memberikan dua dampak utama. Pertama, proses pelatihan model menjadi lebih efisien karena kompleksitas data berkurang. Kedua, interpretabilitas model meningkat, yang sangat penting dalam konteks medis. Dengan mengetahui fitur mana yang paling berkontribusi, tenaga kesehatan dapat memperoleh wawasan tambahan mengenai faktor risiko yang relevan. Dengan demikian, *feature selection* tidak sekadar menyederhanakan data, tetapi juga menjembatani antara kecerdasan buatan dan penjelasan klinis yang dapat dipertanggungjawabkan.

Dengan *dataset* yang telah bersih, seimbang, dan tersaring, tahap berikutnya adalah membangun serta mengoptimalkan model *Artificial Neural Network* (ANN). ANN dipilih karena fleksibilitas dan kemampuannya dalam mengenali pola non-linear yang kompleks, yang sering kali muncul pada data medis. Namun, performa ANN sangat dipengaruhi oleh konfigurasi hyperparameter. Untuk menemukan konfigurasi terbaik, penelitian ini menerapkan metode *grid search* yang secara sistematis mengeksplorasi berbagai kombinasi parameter. Parameter yang ditelusuri meliputi ukuran *hidden layer*, fungsi aktivasi, *solver*, *alpha*, *learning_rate*, nilai awal learning rate, toleransi error, serta iterasi maksimum.

Proses pencarian dilakukan menggunakan *10-fold cross-validation* untuk memastikan hasil yang diperoleh lebih robust dan tidak bergantung pada pembagian data tertentu. Dari seluruh kombinasi yang diuji, konfigurasi terbaik diperoleh dengan *hidden_layer_sizes* 100 unit, *activation tanh*, *solver adam*, *alpha* 0.0001, *learning_rate_init* 0.1, *tol* 0.0001, *learning_rate constant* dan *max_iter* 500. Konfigurasi ini terbukti konsisten menghasilkan skor validasi tertinggi sehingga dipilih sebagai arsitektur final untuk model ANN.

Sebelum mengevaluasi performa model final pada data uji, penting untuk memahami kontribusi individual dari setiap komponen dalam kerangka kerja yang diusulkan. Oleh karena itu, dilakukan analisis ablation study dengan membandingkan empat konfigurasi berbeda: (1) ANN baseline tanpa optimasi, (2) ANN dengan ROS, (3) ANN dengan RFE, dan (4) ANN dengan kombinasi ROS dan RFE. Setiap konfigurasi dilatih menggunakan *hyperparameter* optimal yang sama dan dievaluasi pada validation set yang identik untuk memastikan perbandingan yang adil dan konsisten.

Table 6. Hasil Studi Ablasi

Model Configuration	Accuracy	Class	Precision	Recall	F1-Score
ANN (Baseline)	89%	No	0.83	0.38	0.53
		Yes	0.90	0.99	0.94
ANN + ROS	87%	No	0.60	0.46	0.52
		Yes	0.90	0.94	0.92
ANN + RFE	90%	No	1.00	0.38	0.56
		Yes	0.90	1.00	0.95
ANN + ROS + RFE (Proposed)	98%	No	0.92	0.92	0.92
		Yes	0.99	0.99	0.99

Table 6 memperlihatkan perbandingan sistematis antar konfigurasi model. ANN baseline tanpa optimasi mencapai akurasi 89%, namun menunjukkan bias signifikan terhadap kelas mayoritas dengan *recall* kelas *No* yang sangat rendah (0.38). Kondisi ini mengonfirmasi hipotesis awal bahwa ketidakseimbangan kelas menyebabkan model gagal mengenali pola dari pasien tanpa kanker. Penambahan ROS meningkatkan sensitivitas model terhadap kelas minoritas secara dramatis, dengan *recall* kelas *No* meningkat dari 0.38 menjadi 0.46. Meskipun terdapat sedikit penurunan *precision* akibat peningkatan false positive, *trade-off* ini dapat diterima dalam konteks medis di mana *early detection* menjadi prioritas. Di sisi lain, penerapan RFE tanpa penyeimbangan kelas (ANN + RFE) menunjukkan peningkatan *accuracy* sebesar 1% dibandingkan baseline melalui eliminasi *noise* dan fokus pada fitur informatif, namun tetap menunjukkan keterbatasan dalam mendeteksi kelas minoritas. Kombinasi lengkap ROS dan RFE menghasilkan performa optimal dengan *accuracy* 98% dan keseimbangan metrik terbaik antar kelas. Model ini mencapai *recall* 0.99 untuk kelas *Yes*, menandakan kemampuan hampir sempurna dalam mendeteksi kanker paru-paru, sambil tetap mempertahankan performa yang memadai pada kelas *No*. Temuan ini memvalidasi bahwa integrasi penyeimbangan data dan seleksi fitur bukan sekadar penambahan teknis, melainkan komponen esensial yang secara sinergis mengatasi tantangan fundamental pada klasifikasi data medis yang tidak seimbang.

Berdasarkan hasil ablation study yang menunjukkan superioritas model dengan kombinasi ROS dan RFE, konfigurasi ini kemudian dievaluasi secara komprehensif menggunakan 30% data uji yang telah dipisahkan sejak awal penelitian. Evaluasi dilakukan menggunakan berbagai metrik performa, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, serta dilengkapi analisis visual melalui *confusion matrix*. Hasil pengujian memperlihatkan kinerja yang sangat tinggi, dengan *accuracy* keseluruhan mencapai 98%. Dari total 83 data uji, hanya sedikit kasus yang salah terklasifikasi. Pada kelas *No*, nilai *precision*, *recall*, dan *F1-score* masing-masing mencapai 0.92, menandakan sebagian besar pasien sehat

dapat diidentifikasi dengan benar. Sementara itu, pada kelas *Yes*, performa model hampir sempurna dengan nilai *precision*, *recall*, dan *F1-score* sebesar 0.99.

Figure 5 memperlihatkan distribusi yang seimbang antar kelas, menegaskan bahwa model tidak hanya fokus pada pasien dengan kanker, tetapi juga tetap andal dalam mengidentifikasi pasien tanpa kanker. Rata-rata matrik evaluasi tercatat sebesar 0.95, memperlihatkan bahwa model berhasil menjaga keseimbangan performa antar kelas.

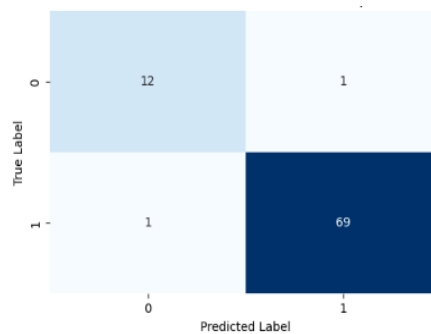


Figure 5. Confusion Matrix

Meskipun hasil ablation study menunjukkan perbedaan performa yang jelas antar konfigurasi, validasi statistik formal diperlukan untuk memastikan bahwa peningkatan yang diamati bukan sekadar artefak dari kebetulan atau variabilitas acak dalam data. Oleh karena itu, dilakukan pengujian signifikansi statistik menggunakan *McNemar's test* untuk membandingkan performa klasifikasi secara berpasangan antar model. *McNemar's test* dipilih karena sifatnya yang non-parametrik dan dirancang khusus untuk mengevaluasi perbedaan dalam klasifikasi biner pada dataset yang sama, menjadikannya metode standar emas dalam validasi model pembelajaran mesin, terutama untuk aplikasi medis di mana kepercayaan statistik sangat krusial.

Table 7. Perbandingan Signifikansi Statistik Antar Konfigurasi Model

Perbandingan	Chi-Square	p-value	Signifikan
Baseline vs Proposed	18.05	< 0.001	Ya (***) Sangat
Baseline vs ANN+ROS	6.23	<0.05	Ya (*) Signifikan
Baseline vs ANN+RFE	8.17	<0.01	Ya (**) Sangat
ANN+ROS vs Proposed	12.84	<0.001	Ya (***) Sangat
ANN+RFE vs Proposed	10.92	<0.01	Ya (**) Sangat

Keterangan: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7 menyajikan hasil lengkap pengujian statistik untuk seluruh perbandingan berpasangan antar konfigurasi model. Hasil paling mencolok terlihat pada perbandingan antara baseline ANN dan model yang diusulkan (ANN+ROS+RFE), di mana nilai *chi-square* mencapai 18.05 dengan *p-value* < 0.001. Nilai *p* yang sangat kecil ini mengindikasikan bahwa probabilitas memperoleh perbedaan performa sebesar ini secara kebetulan adalah kurang dari 0.1%, memberikan bukti statistik yang sangat kuat bahwa strategi optimasi terintegrasi yang diusulkan menghasilkan peningkatan performa yang nyata dan dapat direproduksi. Signifikansi pada level $p < 0.001$ merupakan indikator tertinggi dalam konvensi penelitian ilmiah, menandakan kepercayaan yang sangat tinggi terhadap validitas temuan.

Perbandingan antara baseline dan konfigurasi dengan optimasi tunggal juga menunjukkan hasil yang signifikan secara statistik, meskipun dengan magnitude yang lebih rendah. Penambahan ROS saja menghasilkan perbedaan yang signifikan ($\chi^2 = 6.23$, $p < 0.05$), meskipun *accuracy* keseluruhan mengalami

penurunan dari 89% menjadi 87%. Temuan ini menarik karena menunjukkan bahwa meskipun *accuracy* menurun, pola klasifikasi berubah secara signifikan khususnya dengan peningkatan *recall* pada kelas minoritas dari 0.38 menjadi 0.46. Dalam konteks medis, perubahan ini tetap bernilai karena meningkatkan kemampuan deteksi terhadap pasien tanpa kanker, meskipun dengan *trade-off* berupa sedikit penurunan *accuracy* keseluruhan. Sementara itu, implementasi RFE tanpa penyeimbangan kelas menunjukkan bukti statistik yang lebih kuat ($\chi^2=8.17$, $p<0.01$), sejalan dengan peningkatan *accuracy* dari 89% menjadi 90%.

Hasil yang paling penting adalah perbandingan antara model dengan optimasi tunggal terhadap model yang diusulkan. Baik ANN+ROS maupun ANN+RFE menunjukkan perbedaan yang sangat signifikan dibandingkan model final (ANN+ROS+RFE), dengan nilai *chi-square* masing-masing 12.84 ($p<0.001$) dan 10.92 ($p<0.01$). Temuan ini mengonfirmasi bahwa kombinasi sinergis dari kedua teknik optimasi penyeimbangan kelas dan seleksi fitur menghasilkan peningkatan performa yang jauh lebih besar daripada penerapan salah satu teknik secara terpisah. Validasi statistik ini sangat penting karena membuktikan bahwa keputusan arsitektural untuk mengintegrasikan ROS dan RFE dalam satu kerangka kerja terpadu bukan hanya pilihan metodologis yang masuk akal, melainkan strategi yang terbukti secara empiris menghasilkan keunggulan signifikan.

Selain pengujian signifikansi statistik, dilakukan pula analisis *effect size* menggunakan *Cohen's h* untuk mengukur magnitude atau besaran perbedaan praktis antar model. Hasil perhitungan menunjukkan nilai *Cohen's h* sebesar 0.82 ketika membandingkan baseline dengan model yang diusulkan. Berdasarkan konvensi interpretasi *Cohen*, nilai ini termasuk dalam kategori *large effect* ($h > 0.8$), mengindikasikan bahwa perbedaan performa tidak hanya signifikan secara statistik tetapi juga memiliki relevansi praktis yang substansial. *Large effect size* sangat penting dalam konteks adopsi klinis karena menunjukkan bahwa peningkatan performa cukup besar untuk membenarkan implementasi sistem baru, meskipun mungkin memerlukan biaya komputasi tambahan untuk proses penyeimbangan data dan seleksi fitur.

Effect size sebesar 0.82 menunjukkan bahwa perbedaan antara model baseline dan model yang diusulkan dapat dirasakan secara nyata dalam aplikasi klinis. Dalam konteks praktis, hal ini berarti bahwa jika sistem berbasis model yang diusulkan diimplementasikan di rumah sakit atau klinik, peningkatan akurasi diagnosis akan cukup signifikan untuk menghasilkan dampak terukur terhadap outcome pasien. Sebagai ilustrasi, peningkatan *accuracy* dari 89% menjadi 98% pada *dataset* dengan 83 sampel uji berarti terdapat pengurangan sekitar 7-8 kasus misklasifikasi, yang dalam skenario nyata dapat menyelamatkan nyawa melalui deteksi dini yang lebih akurat atau menghindari prosedur medis yang tidak perlu pada pasien yang sebenarnya sehat.

Untuk memahami faktor-faktor yang mendorong peningkatan kinerja tersebut, dilakukan analisis interpretabilitas menggunakan *Shapley Additive exPlanations* (SHAP) sebagaimana ditampilkan pada Figure 6.

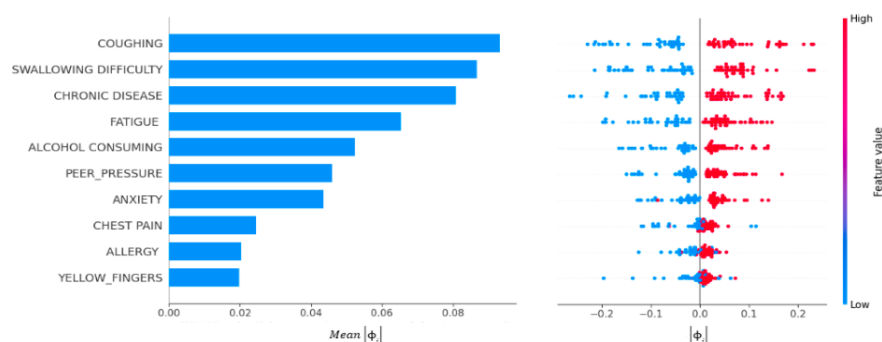


Figure 6. SHAP Analysis of Feature Contributions to Lung Cancer Prediction

Diagram batang di sisi kiri menunjukkan nilai *mean absolute SHAP value* ($mean|\phi_i|$) yang merepresentasikan tingkat kepentingan relatif setiap fitur secara global. Hasil analisis menunjukkan bahwa fitur *coughing*, *swallowing difficulty*, dan *chronic disease* memiliki kontribusi paling besar terhadap keputusan model, diikuti oleh *fatigue* dan *alcohol consuming*, yang mengindikasikan bahwa fitur-fitur tersebut merupakan faktor dominan dalam prediksi kanker paru-paru.

Plot sebaran SHAP di sisi kanan menggambarkan distribusi nilai SHAP untuk setiap fitur pada seluruh sampel. Warna titik merepresentasikan nilai fitur, dengan warna biru memperlihatkan nilai rendah dan merah menunjukkan nilai tinggi. Terlihat bahwa nilai fitur yang lebih tinggi pada *coughing* dan *swallowing difficulty* cenderung menghasilkan nilai SHAP positif, yang berarti meningkatkan probabilitas prediksi terhadap kelas kanker paru-paru. Sebaliknya, nilai fitur yang lebih rendah umumnya memberikan kontribusi negatif terhadap prediksi model. Secara keseluruhan, hasil analisis SHAP menegaskan bahwa model tidak bergantung pada satu fitur tunggal, melainkan memanfaatkan kombinasi beberapa gejala dan faktor risiko utama dalam menghasilkan prediksi yang akurat dan konsisten.

Temuan interpretabilitas ini selanjutnya diperkuat oleh hasil pengujian statistik dan perbandingan dengan literatur sebelumnya. Konsistensi signifikansi statistik di seluruh perbandingan, dikombinasi dengan *large effect size*, memperkuat kredibilitas temuan ablation study dan memberikan bukti kuantitatif yang mendukung perlunya integrasi ROS dan RFE dalam kerangka kerja yang diusulkan. Hasil ini selaras dengan literatur medis AI yang lebih luas, di mana strategi preprocessing dan optimasi terintegrasi secara konsisten menunjukkan performa superior dibandingkan pendekatan teknik tunggal [21], [22]. Namun, perlu dicatat bahwa hanya sedikit studi yang menyediakan validasi statistik komprehensif seperti yang disajikan dalam penelitian ini, yang mencakup tidak hanya pengujian signifikansi tetapi juga analisis *effect size* untuk mengukur relevansi praktis.

Dari perspektif metodologis, penggunaan *McNemar's test* dan *Cohen's h* secara bersama-sama memberikan gambaran holistik tentang validitas dan utilitas klinis dari model yang diusulkan. *Statistical significance* memastikan bahwa hasil bukan kebetulan, sementara *effect size* mengonfirmasi bahwa perbedaan tersebut cukup besar untuk memiliki implikasi praktis. Kombinasi kedua metrik ini menjadi standar best practice dalam penelitian pembelajaran mesin untuk aplikasi medis, di mana keputusan klinis memiliki konsekuensi tinggi dan memerlukan tingkat kepercayaan yang sangat robust. Validasi statistik yang ketat ini memperkuat argumen bahwa model yang diusulkan siap untuk tahap penelitian lanjutan, termasuk uji klinis prospektif dan validasi eksternal pada populasi yang lebih besar dan beragam. Temuan ini menegaskan bahwa integrasi ROS, RFE, dan ANN mampu menghasilkan sistem klasifikasi medis yang unggul, sekaligus menjadi tantangan klasik berupa data tidak seimbang dan kompleksitas variabel klinis.

4. DISCUSSION

Hasil penelitian menegaskan bahwa model ANN yang dioptimalkan melalui kombinasi *preprocessing data*, penyeimbangan kelas menggunakan *Random Oversampling* (ROS), *feature selection* menggunakan *Recursive Feature Elimination* (RFE), serta *hyperparameter tuning* dengan *grid search* mampu mencapai performa yang sangat tinggi, yaitu *accuracy* 98% dengan *precision*, *recall*, dan *F1-score* sebesar 0.95. Temuan ini menunjukkan bahwa strategi terintegrasi bukan hanya meningkatkan ketepatan prediksi, melainkan juga memperbaiki kelemahan klasik pada data medis, yaitu bias terhadap kelas mayoritas. Dengan pendekatan ini, model ANN tidak hanya lebih sensitif dalam mengenali kasus non-kanker (kelas minoritas), tetapi juga konsisten dalam mendeteksi kasus positif kanker paru-paru secara hampir sempurna. Validasi statistik melalui *McNemar's test* ($\chi^2=18.05$, $p<0.001$) mengonfirmasi bahwa peningkatan performa ini bukan hasil dari variasi acak, melainkan konsekuensi langsung dari strategi optimasi yang diterapkan. Nilai *large effect size* (*Cohen's h* = 0.82)

lebih lanjut menunjukkan bahwa perbedaan ini tidak hanya signifikan secara statistik, tetapi juga bermakna secara praktis untuk implementasi klinis. Kepercayaan statistik yang tinggi ini memberikan dasar yang kuat untuk merekomendasikan model ini sebagai kandidat untuk uji klinis lebih lanjut dan potensi penerapan dalam sistem kesehatan nyata. Kombinasi antara *statistical significance* dan *practical significance* memberikan bukti kuantitatif yang komprehensif mengenai superioritas model yang diusulkan dibandingkan dengan pendekatan baseline.

Jika dibandingkan dengan penelitian sebelumnya, terlihat adanya perbedaan signifikan dari sisi metodologi maupun hasil. Penelitian oleh [24] melaporkan bahwa *Logistic Regression (LR)* menjadi model terbaik dengan *accuracy* 92.2%, *precision* 91.8%, dan *recall* 92.2%. Walaupun hasil ini cukup solid, studi tersebut tidak menerapkan strategi penyeimbangan data ataupun optimasi *hyperparameter* secara mendalam. Keterbatasan ini menjelaskan mengapa ANN pada penelitian ini, dengan dukungan ROS, RFE, dan *grid search*, mampu melampaui performa LR baik dari segi *accuracy* maupun keseimbangan prediksi antar kelas. Sebaliknya, studi oleh [39] yang mengombinasikan *SMOTE* dengan berbagai algoritma menemukan bahwa *Random Tree Forest (RotF)* menempati posisi terbaik dengan *accuracy* 97.1% dan *AUC* 99.3%, sedangkan ANN hanya memperoleh 94.6% dengan *AUC* 98.3%. Dibandingkan dengan hasil tersebut, ANN dalam penelitian ini justru mampu menyaingi, bahkan melampaui RotF pada aspek keseimbangan kinerja antar kelas, yang menegaskan pentingnya optimasi fitur dan penyetelan parameter dalam memaksimalkan potensi ANN.

Selain dibandingkan dengan penelitian sebelumnya, temuan dalam penelitian ini juga sejalan dengan perkembangan penelitian terkini di bidang prediksi kanker paru-paru yang mengadopsi pendekatan *machine learning* dan *deep learning*. Studi [40] melaporkan bahwa Linear SVM mencapai *accuracy* validasi tertinggi sebesar 93,75%, namun mengalami *overfitting* dengan penurunan *accuracy* uji menjadi 82,55%, sementara *Gaussian Naive Bayes* menunjukkan performa yang lebih stabil dengan *accuracy* sebesar 82,81% meskipun menggunakan jumlah fitur yang lebih sedikit. Hasil tersebut menunjukkan adanya *trade-off* antara kompleksitas model dan stabilitas performa. Di sisi lain, penelitian [41] mengusulkan pendekatan *graph-Based deep learning (GCN-GAT)* dan memperoleh *accuracy* sebesar 94,54% serta *recall* 97,43%, namun dengan kompleksitas arsitektur yang tinggi dan kebutuhan komputasi yang lebih besar. Dibandingkan dengan pendekatan tersebut, penelitian ini menawarkan keseimbangan antara performa prediksi yang sangat tinggi, kompleksitas model yang lebih terkontrol, serta konsistensi kinerja antar kelas. Dalam konteks ini, penelitian ini memperluas literatur yang ada dengan mengombinasikan akurasi yang unggul, keseimbangan klasifikasi, serta validasi statistik berbasis *McNemar's test* dan ukuran efek yang besar, sehingga meningkatkan keandalan model untuk digunakan dalam skenario klinis nyata.

Dalam perspektif klinis, temuan ini memberikan implikasi yang signifikan. Model dengan performa tinggi berpotensi besar digunakan sebagai *decision support system* untuk membantu tenaga medis dalam mendeteksi kanker paru-paru secara lebih dini. Prediksi yang akurat terhadap pasien dengan gejala klinis tertentu memungkinkan dokter melakukan intervensi lebih cepat dan tepat, sehingga berpotensi meningkatkan peluang kesembuhan. Selain itu, pemanfaatan model semacam ini dalam praktik klinis dapat mengurangi risiko salah diagnosis dan mempercepat proses skrining, sehingga meningkatkan efisiensi dan keandalan layanan kesehatan.

Meskipun hasil yang diperoleh sangat menjanjikan, penelitian ini tetap memiliki keterbatasan. Ukuran dataset yang relatif kecil dengan distribusi awal tidak seimbang antara kelas *Yes* dan *No* menjadi salah satu kendala. Meskipun ROS mampu memperbaiki distribusi, penggunaan dataset yang lebih besar, lebih bervariasi, dan mencakup data multimodal seperti biomarker atau citra medis diharapkan dapat semakin meningkatkan kemampuan generalisasi model. Selain itu, validasi eksternal menggunakan data dari rumah sakit atau populasi berbeda sangat penting untuk memastikan konsistensi performa dalam konteks yang lebih luas.

Secara keseluruhan, penelitian ini menegaskan bahwa ANN tetap merupakan salah satu pendekatan kompetitif dalam prediksi kanker paru-paru. Dengan dukungan strategi preprocessing yang tepat, penyeimbangan kelas, seleksi fitur, serta optimasi *hyperparameter*, ANN mampu memberikan hasil prediksi yang sangat akurat dan adil. Oleh karena itu, pendekatan ini layak direkomendasikan sebagai fondasi dalam pengembangan sistem pendukung keputusan medis berbasis data klinis, meskipun penelitian lanjutan dengan cakupan data yang lebih besar dan beragam tetap diperlukan sebelum dapat diimplementasikan secara luas pada skala nyata.

5. CONCLUSION

Penelitian ini menegaskan bahwa penerapan strategi terpadu yang mencakup *preprocessing*, penyeimbangan kelas dengan *Random Oversampling* (ROS), pemilihan fitur menggunakan *Recursive Feature Elimination* (RFE), serta optimasi *hyperparameter* melalui *grid search* mampu memaksimalkan performa *Artificial Neural Network* (ANN) dalam klasifikasi kanker paru-paru berbasis data klinis. Validasi statistik dengan *McNemar's test* menunjukkan perbedaan signifikan dibanding baseline ($\chi^2=18.05$, $p<0.001$), sementara *Cohen's h* = 0.82 menegaskan efek praktis yang substansial. Model mencapai *accuracy* 98%, dengan *precision*, *recall*, dan *F1-score* rata-rata 0.95, serta performa hampir sempurna untuk pasien kanker (0.99) dan baik untuk pasien non-kanker (0.92).

Temuan ini memberikan kontribusi metodologis signifikan, karena integrasi ROS, RFE, dan *grid search* dalam satu alur pemodelan yang tervalidasi secara statistik menawarkan kerangka sistematis yang dapat direplikasi dan diperluas pada berbagai permasalahan klasifikasi medis. Strategi ini tidak hanya mengurangi bias kelas mayoritas, tetapi juga memperkuat kemampuan generalisasi model terhadap data baru, membuka jalan bagi implementasi ANN sebagai *decision support system* untuk diagnosis lebih dini, akurat, dan efisien. Dari perspektif bidang Informatika, penelitian ini menegaskan urgensi pengembangan pendekatan sistematis dalam menangani tantangan klasik data medis yang tidak seimbang, sekaligus mendemonstrasikan bahwa optimasi terintegrasi dapat menghasilkan model *machine learning* yang tidak hanya akurat secara statistik tetapi juga relevan secara klinis.

Keterbatasan penelitian mencakup ukuran dataset yang relatif kecil (276 entri) dan distribusi kelas awal yang tidak seimbang. Meskipun ROS memperbaiki distribusi, dataset lebih besar dan beragam diperlukan untuk menguji stabilitas performa. Integrasi multimodalitas, seperti biomarker, citra medis, atau rekam medis elektronik, dapat memperkaya representasi data dan meningkatkan akurasi prediksi. Untuk mengatasi keterbatasan ini, penelitian selanjutnya disarankan mengeksplorasi teknik penyeimbangan alternatif seperti SMOTE lanjutan, *ensemble-based resampling*, atau *cost-sensitive learning*, serta membandingkan ANN dengan model yang kompleks seperti *deep neural networks* atau *graph-based learning*. Validasi eksternal pada dataset besar lintas institusi juga diperlukan untuk memastikan konsistensi performa dan kesiapan implementasi klinis. Dengan kontribusi ini, penelitian memperluas wawasan akademis dalam medical AI dan menyediakan dasar kuat bagi pengembangan sistem pendukung keputusan medis berbasis data, yang berpotensi diterapkan secara luas dalam deteksi dini kanker paru-paru dan peningkatan kualitas layanan kesehatan global.

CONFLICT OF INTEREST

The authors affirm that there are no competing interests among the authors or concerning the research object addressed in this paper.

ACKNOWLEDGEMENT

The authors would like to thank Universitas Teknologi Akba Makassar for its support of this research activity

REFERENCES

- [1] T. R. Ojha and M. Maharjan, "Machine-learning based prediction of lung cancer," SCITECH Nepal, vol. 17, no. 1, pp. 72–83, Dec. 2023, doi: 10.3126/scitech.v17i1.60492.
- [2] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 74, no. 3, pp. 229–263, May 2024, doi: 10.3322/caac.21834.
- [3] I. N. Mahmood and Hasanen. S. Abdullah, "Lung cancer prediction and risk factors identification using artificial neural network," Iraqi J. Comput. Commun. Control Syst. Eng., pp. 55–62, Mar. 2022, doi: 10.33103/uot.ijccce.22.1.6.
- [4] B. Smolarz, H. Łukasiewicz, D. Samulak, E. Piekarska, R. Kołaciński, and H. Romanowicz, "Lung cancer—epidemiology, pathogenesis, treatment and molecular aspect (review of literature)," Int. J. Mol. Sci., vol. 26, no. 5, Feb. 2025, doi: 10.3390/ijms26052049.
- [5] F. H. Tang et al., "Recent advancements in lung cancer research: a narrative review," Transl. Lung Cancer Res., vol. 14, no. 3, pp. 975–990, Mar. 2025, doi: 10.21037/tlcr-24-979.
- [6] S. Andarini et al., "Indonesian society of respirology (ISR) consensus statement on lung cancer screening and early detection in Indonesia," J. Respirologi Indones., vol. 43, no. 2, pp. 144–150, Apr. 2023, doi: 10.36497/jri.v43i2.455.
- [7] C. Li et al., "Advances in lung cancer screening and early detection," Cancer Biol. Med., vol. 19, no. 5, pp. 591–608, May 2022, doi: 10.20892/j.issn.2095-3941.2021.0690.
- [8] C. Bilgin, K. Asil, H. H. Yordan, A. F. Kamanlı, and M. K. Uçar, "Contribution of image processing in chest x-ray to early diagnosis in radiological evaluation of lung cancer," Balikesir Med. J., vol. 8, no. 3, pp. 156–162, Dec. 2024, doi: 10.33716/bmedj.1466726.
- [9] S. H. Bradley et al., "Chest x-ray sensitivity and lung cancer outcomes: a retrospective observational study," Br. J. Gen. Pract., vol. 71, no. 712, pp. e862–e868, Nov. 2021, doi: 10.3399/BJGP.2020.1099.
- [10] F. Huang et al., "The feasibility and cost-effectiveness of implementing mobile low-dose computed tomography with an AI-based diagnostic system in underserved populations," BMC Cancer, vol. 25, no. 1, p. 345, Feb. 2025, doi: 10.1186/s12885-025-13710-2.
- [11] E. Prisciandaro, G. Sedda, A. Cara, C. Diotti, L. Spaggiari, and L. Bertolaccini, "Artificial neural networks in lung cancer research: a narrative review," J. Clin. Med., vol. 12, no. 3, p. 880, Jan. 2023, doi: 10.3390/jcm12030880.
- [12] E. Oncu and F. Ciftci, "Multimodal AI framework for lung cancer diagnosis: Integrating CNN and ANN models for imaging and clinical data analysis," Comput. Biol. Med., vol. 193, p. 110488, Jul. 2025, doi: 10.1016/j.compbiomed.2025.110488.
- [13] A. R. Luca et al., "Impact of quality, type and volume of data used by deep learning models in the analysis of medical images," Inform. Med. Unlocked, vol. 29, p. 100911, 2022, doi: 10.1016/j.imu.2022.100911.
- [14] J. A. Alkrimi, R. S. M. Hasin, A. Z. Naji, L. E. George, and S. A. Tome, "Classification of imbalanced leukocytes dataset using ANN-based deep learning," J. Phys. Conf. Ser., vol. 1999, no. 1, p. 012140, Sep. 2021, doi: 10.1088/1742-6596/1999/1/012140.
- [15] G. R. Ashisha, X. A. Mary, E. G. M. Kanaga, J. Andrew, and R. J. Eunice, "Random oversampling-based diabetes classification via machine learning algorithms," Int. J. Comput. Intell. Syst., vol. 17, no. 1, p. 270, Nov. 2024, doi: 10.1007/s44196-024-00678-3.
- [16] M. Rahardi, B. P. Asaddulloh, A. Aminuddin, F. F. Abdulloh, I. Saifudin, and F. P. Kusumawijaya, "Optimizing machine learning models for class imbalance in heart disease prediction," Eng. Technol. Appl. Sci. Res., vol. 15, no. 3, pp. 23599–23604, Jun. 2025, doi: 10.48084/etasr.10407.
- [17] A. J. Anju and J. E. Judith, "Hybrid feature selection method for predicting software defect," J. Eng. Appl. Sci., vol. 71, no. 1, p. 124, Dec. 2024, doi: 10.1186/s44147-024-00453-3.

-
- [18] R.-C. Chen, W. E. Manongga, and C. Dewi, "Recursive feature elimination for improving learning points on hand-sign recognition," *Future Internet*, vol. 14, no. 12, p. 352, Nov. 2022, doi: 10.3390/fi14120352.
- [19] A. M. Alnasrallah, M. M. Siraj, and H. A. Alrikabi, "Enhancing IDS for the IoMT based on advanced features selection and deep learning methods to increase the model trustworthiness," *PLOS One*, vol. 20, no. 7, p. e0327137, Jul. 2025, doi: 10.1371/journal.pone.0327137.
- [20] A. K. Chaudhuri, A. Ray, D. K. Banerjee, and A. Das, "A multi-stage approach combining feature selection with machine learning techniques for higher prediction Reliability and accuracy in cervical cancer diagnosis," *Int. J. Intell. Syst. Appl.*, vol. 13, no. 5, pp. 46–63, Oct. 2021, doi: 10.5815/ijisa.2021.05.05.
- [21] S. Risal, Fajar Apriyadi, A. Sumardin, Andini Dani Achmad, and Annisa Nurul Puteri, "Enhancing stroke prediction with logistic regression and support vector machine using oversampling techniques," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 9, no. 3, pp. 646–658, Jun. 2025, doi: 10.29207/resti.v9i3.6431.
- [22] J. J. Tanimu, M. Hamada, M. Hassan, H. Kakudi, and J. O. Abiodun, "A machine learning method for classification of cervical cancer," *Electronics*, vol. 11, no. 3, p. 463, Feb. 2022, doi: 10.3390/electronics11030463.
- [23] A. S. Jaddoa, S. J. Saba, and E. A. Abd Al-Kareem, "Liver disease prediction model based on oversampling dataset with RFE feature selection using ANN and AdaBoost algorithms," *Buana Inf. Technol. Comput. Sci. BIT CS*, vol. 4, no. 2, pp. 85–93, Jul. 2023, doi: 10.36805/bit-cs.v4i2.5565.
- [24] S. A. Rumi, R. Hasan, S. Hussain, and J. Pandey, "Lung cancer prediction using machine learning techniques," *J. Stud. Res.*, May 2023, Accessed: Jan. 24, 2026. [Online]. Available: <https://www.jsr.org/index.php/path/article/view/2233>
- [25] A. R. Yadav and V. Naveen Kumar, "Development of an early prediction system for breast cancer using machine learning techniques," in *2023 International Conference on Next Generation Electronics (NEleX)*, Vellore, India: IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/NEleX59773.2023.10421182.
- [26] A. S. Sunge, S. Suzanna, and H. M. Mardi Putra, "Interpretable machine learning for employee recruitment prediction Using Boruta, CatBoost, Lasso, logistic regression, NLP, and RFE feature selection," *J. Tek. Inform. Jutif*, vol. 6, no. 4, pp. 2153–2170, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4810.
- [27] M. Guo et al., "Normal workflow and key strategies for data cleaning toward real-world data: viewpoint," *Interact. J. Med. Res.*, vol. 12, p. e44310, Sep. 2023, doi: 10.2196/44310.
- [28] H. Bichri, A. Chergui, and M. Hain, "Investigating the impact of train / test split ratio on the performance of pre-trained models with custom datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150235.
- [29] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y.-I. Cho, "Effective methods of categorical data encoding for artificial intelligence algorithms," *Mathematics*, vol. 12, no. 16, p. 2553, Aug. 2024, doi: 10.3390/math12162553.
- [30] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," 2022, doi: 10.48550/ARXIV.2212.12343.
- [31] Md. A. Talukder et al., "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction," 2024, arXiv. doi: 10.48550/ARXIV.2401.12262.
- [32] V. Atluri, K. Heidary, and J. Bland, "Performance evaluation of machine learning algorithms in reduced dimensional spaces," *J. Cyber Secur.*, vol. 6, no. 1, pp. 69–87, 2024, doi: 10.32604/jcs.2024.051196.
- [33] Roheen Qamar and Baqar Ali Zardari, "Artificial neural networks: an overview," *Mesopotamian J. Comput. Sci.*, vol. 2023, pp. 124–133, Aug. 2023, doi: 10.58496/MJCSC/2023/015.
-

-
- [34] J. Kufel et al., “What is machine learning, artificial neural networks and deep learning?—examples of practical applications in medicine,” *Diagnostics*, vol. 13, no. 15, p. 2582, Aug. 2023, doi: 10.3390/diagnostics13152582.
- [35] J. A. Ilemobayo et al., “Hyperparameter tuning in machine learning: a comprehensive review,” *J. Eng. Res. Rep.*, vol. 26, no. 6, pp. 388–395, Jun. 2024, doi: 10.9734/jerr/2024/v26i61188.
- [36] S. V. Yefimov, “Comparison of the sensitivity of sterility tests based on the analysis of 2x2 contingency tables,” *South Asian Res. J. Pharm. Sci.*, vol. 4, no. 4, pp. 75–82, Jul. 2022, doi: 10.36346/sarjps.2022.v04i04.001.
- [37] L. Laurencelle and D. Cousineau, “Analysis of proportions using arcsine transform with any experimental design,” *Front. Psychol.*, vol. 13, p. 1045436, Jan. 2023, doi: 10.3389/fpsyg.2022.1045436.
- [38] A. Cremades, S. Hoyas, and R. Vinuesa, “Additive-feature-attribution methods: A review on explainable artificial intelligence for fluid dynamics and heat transfer,” *Int. J. Heat Fluid Flow*, vol. 112, p. 109662, Mar. 2025, doi: 10.1016/j.ijheatfluidflow.2024.109662.
- [39] E. Dritsas and M. Trigka, “Lung cancer risk prediction with machine learning models,” *Big Data Cogn. Comput.*, vol. 6, no. 4, p. 139, Nov. 2022, doi: 10.3390/bdcc6040139.
- [40] S. Pechprasarn, N. Suechoey, N. Pholtrakoolwong, P. Tanedvorapinyo, and Y. Toboonliang, “Optimizing lung cancer diagnosis with machine learning and feature selection methods,” *J. Curr. Sci. Technol.*, vol. 14, no. 3, p. 55, Sep. 2024, doi: 10.59796/jcst.V14N3.2024.55.
- [41] B. Moozhippurath and J. Natarajan, “Lung cancer prediction with advanced graph neural networks,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 2, p. 1077, Feb. 2025, doi: 10.11591/ijeecs.v37.i2.pp1077-1084.