



## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

Florenza Jovana<sup>1</sup>, Imelda Sunario<sup>2</sup>, Wincent Coleus Phan<sup>3</sup>, Ade Maulana<sup>4\*</sup>

<sup>1,2,3,4</sup>Sistem Informasi, Fakultas Teknologi Informasi, Universitas Pelita Harapan, Medan, Indonesia

Email: <sup>1</sup>03081230079@student.uph.edu, <sup>2</sup>03081230055@student.uph.edu,

<sup>3</sup>03081230007@student.uph.edu, <sup>4</sup>ade.maulana@lecturer.uph.edu

### Informasi Artikel

Diterima : 19-04-2025

Disetujui : 05-05-2025

Diterbitkan : 20-05-2025

### ABSTRACT

*This study aims to develop a lung cancer risk prediction system using the K-Nearest Neighbors (KNN) algorithm. The dataset consists of 5,000 patient medical records. During preprocessing, several steps were conducted, including handling missing values, data normalization, label encoding, and selecting relevant features for prediction. A total of 61 outlier entries were removed using the Interquartile Range (IQR) method, resulting in 4,939 clean data entries. The model was trained by tuning the k parameter, with the best performance achieved at k = 20, reaching a training accuracy of 89%. Model evaluation on the test data produced an accuracy of 88%, along with high precision, recall, and F1-score for both classes. After training, the model was integrated into a mobile-based application called LungHealth, which allows users to assess their lung cancer risk. This system is expected to support early detection in a fast, accurate, and efficient manner, enabling individuals—especially those with high-risk factors—to take preventive actions promptly and improve awareness of their lung health.*

**Keyword:** K-Nearest Neighbors, Kanker Paru, Machine Learning

### ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem prediksi risiko kanker paru-paru menggunakan algoritma *K-Nearest Neighbors* (KNN). Dataset yang digunakan terdiri dari 5.000 data rekam medis pasien. Tahapan pra-pemrosesan data mencakup penanganan nilai yang hilang, normalisasi data, label encoding, dan seleksi fitur berdasarkan relevansi terhadap target prediksi. Sebanyak 61 data yang teridentifikasi sebagai outlier dihapus menggunakan metode Interquartile Range (IQR), sehingga diperoleh 4.939 data bersih. Proses pelatihan model dilakukan dengan menyesuaikan parameter k, dan performa terbaik dicapai

# Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

saat  $k = 20$  dengan akurasi pelatihan sebesar 89%. Model kemudian dievaluasi menggunakan data uji, dan menghasilkan akurasi sebesar 88%, disertai nilai precision, recall, dan F1-score yang tinggi untuk masing-masing kelas. Setelah pelatihan, model diintegrasikan ke dalam aplikasi mobile bernama LungHealth untuk memberikan layanan prediksi risiko kanker paru-paru secara cepat dan akurat. Sistem ini diharapkan dapat membantu pengguna dalam mendeteksi risiko secara dini serta mengambil langkah preventif dengan lebih efisien, terutama bagi individu dengan riwayat kesehatan yang berisiko tinggi terhadap kanker paru-paru.

**Kata Kunci:** K-Nearest Neighbors, Lung Cancer, Machine Learning

---

## 1. PENDAHULUAN

Menurut data dari World Health Organization (WHO), kanker paru-paru adalah salah satu jenis kanker yang paling mematikan di dunia. Penyakit ini sering kali sulit dideteksi pada tahap awal karena gejalanya yang tidak spesifik, seperti batuk kronis atau sesak napas. Faktor-faktor seperti kebiasaan merokok, paparan polusi udara, riwayat keluarga, dan paparan bahan kimia tertentu dapat meningkatkan risiko seseorang terkena kanker paru-paru (Sambi Ua et al., 2023; Sugiharto et al., 2021). Karena tingkat keparahan dan dampaknya yang besar, deteksi dini kanker paru-paru menjadi sangat penting untuk meningkatkan peluang kesembuhan pasien.

Kanker paru-paru menyumbang sekitar 20% dari total kematian akibat kanker di seluruh dunia. Di negara-negara berkembang, jumlah kasus kanker paru-paru terus meningkat akibat tingginya tingkat polusi udara dan prevalensi kebiasaan merokok. Sayangnya, metode deteksi tradisional seperti biopsi dan pencitraan medis sering kali mahal, invasif, dan memerlukan waktu yang lama. Oleh karena itu, dibutuhkan pendekatan baru yang lebih cepat dan efisien untuk mendeteksi kanker paru-paru (Bharathy & Pavithra, 2022).

Salah satu solusi yang dapat digunakan adalah penerapan algoritma machine learning. Algoritma K-Nearest Neighbors (KNN) merupakan salah satu metode yang sering digunakan untuk klasifikasi data medis (Pangaribuan et al., 2024). KNN bekerja dengan cara membandingkan data baru dengan data yang sudah ada berdasarkan kedekatan jarak (Maleki et al., 2021). Dengan KNN, hal ini dapat membantu mengidentifikasi pasien dengan risiko tinggi terkena kanker paru-paru berdasarkan data seperti usia, kebiasaan merokok, tingkat stres, dan lainnya.

Penelitian sebelumnya menunjukkan bahwa algoritma KNN sangat menjanjikan untuk mendeteksi dan memprediksi kanker paru-paru. Beberapa studi membuktikan kemampuannya: Moon dan Jetawat (2024) mencapai akurasi 95% menggunakan data pasien umum(Moon & Jetawat, 2024); Liang dan Luo (2025) bahkan menggunakan KNN untuk memprediksi risiko kematian pada pasien kanker paru-paru sel kecil, menunjukkan kemampuannya untuk analisis yang lebih dalam (Liang & Luo, 2025); Ada juga studi oleh Ganesh dan Baskar (2025) yang berhasil mendeteksi kanker paru-paru dengan akurasi 94,409% hanya dari analisis napas, yang

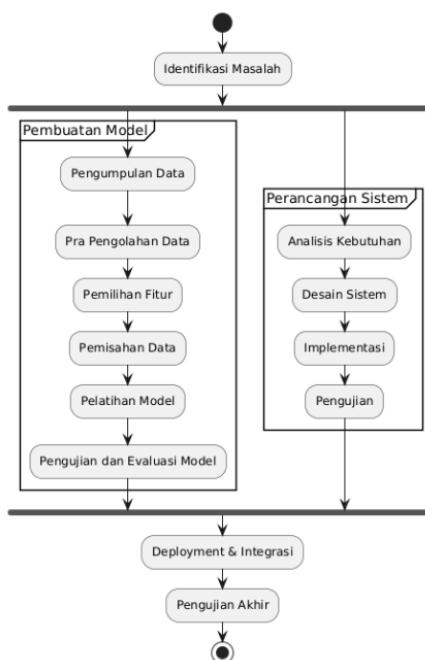
## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

merupakan metode non-invasif (Ganesh & Baskar, 2025); Terakhir, Sachdeva et al. (2024) mengembangkan versi KNN yang lebih baik yang mencapai akurasi luar biasa 98,39% menggunakan data gejala dan riwayat medis, membuktikan bahwa KNN bisa terus ditingkatkan (Sachdeva et al., 2024). Semua penelitian ini secara kolektif menegaskan bahwa KNN adalah algoritma *machine learning* yang serbaguna dan sangat efektif untuk membantu prediksi kanker paru-paru.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk merancang sebuah aplikasi mobile yang dapat memprediksi potensi kanker paru-paru dengan menggunakan algoritma KNN. Dataset yang digunakan mencakup berbagai faktor risiko genetik dan lingkungan secara bersamaan. Diharapkan hasil dari penelitian ini dapat menjadi solusi deteksi dini yang cepat, akurat, dan mudah diakses oleh tenaga medis maupun masyarakat umum.

## 2. METODE

Rangkaian proses yang akan dilakukan dalam penelitian pendekripsi risiko kanker paru-paru menggunakan algoritma KNN ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, tahapan yang dilakukan pada penelitian ini dimulai dari identifikasi masalah, kemudian dilanjutkan dengan perancangan sistem yang mencakup: analisis kebutuhan, desain sistem, implementasi, dan pengujian. Bersamaan dengan itu, dilakukan pula proses pembuatan model yang meliputi: pengumpulan data, pra pengolahan data, pemilihan fitur, pemisahan data menjadi data latih dan data uji, pelatihan model, dan pengujian & evaluasi model. Setelah kedua proses tersebut selesai, dilanjutkan ke tahap deployment & integrasi, dan ditutup dengan testing akhir. Tabel 1 menjelaskan apa saja yang dilakukan di setiap tahap dan fungsinya dalam keseluruhan proses penelitian.

Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan  
Algoritma K-Nearest Neighbors

Tabel 1. Alur Penelitian

| Tahap                  | Subtahap                     | Penjelasan  |
|------------------------|------------------------------|---|
| Identifikasi Masalah   | -                            |   |
| Pembuatan Model        | Pengumpulan Data             | Dataset dari Kaggle, terdiri dari 5.000 baris data dan 18 atribut terkait kondisi kesehatan pasien.           |
|                        | Pra Pengolahan Data          | Normalisasi data dan penghapusan outlier menggunakan metode IQR untuk menjaga kualitas data.                  |
|                        | Pemilihan Fitur              | Memilih fitur yang relevan untuk meningkatkan akurasi dan efisiensi pemrosesan model.                         |
|                        | Pemisahan Data               | Membagi data menjadi 80% untuk pelatihan dan 20% untuk pengujian model.                                       |
|                        | Pelatihan Model              | Melatih model KNN di Google Colab, dengan normalisasi fitur dan penentuan nilai k optimal.                    |
| Perancangan Sistem     | Pengujian dan Evaluasi Model | Evaluasi performa model menggunakan confusion matrix dan metrik akurasi, presisi, recall, dan F1-score.       |
|                        | Analisis Kebutuhan           | Identifikasi kebutuhan pengguna melalui studi dokumen dan diskusi untuk fitur dan fungsi sistem.              |
|                        | Desain Sistem                | Membuat diagram UML ( <i>use case, activity, sequence, class diagram</i> ) sebagai acuan pengembangan sistem. |
|                        | Implementasi                 | Penulisan kode program menggunakan python, php dan java dengan Visual Studio Code dan android studio.         |
|                        | Pengujian                    | Melakukan black box testing untuk memastikan fungsi berjalan sesuai harapan.                                  |
| Deployment & Integrasi | -                            | Mengintegrasikan model ke dalam aplikasi berbasis mobile agar dapat berfungsi menyeluruh.                     |
| Pengujian Akhir        | -                            | Pengujian final untuk memastikan sistem memenuhi kebutuhan dan siap digunakan oleh pengguna.                  |

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Pembuatan Model

##### 3.1.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset yang diperoleh dari situs Kaggle (<https://www.kaggle.com/datasets/shantanugarg274/lung-cancer-prediction-dataset>) format csv. Data tersebut terdiri dari 5.000 baris data dengan 18 atribut yang dapat dilihat pada tabel 2. Tipe atribut dan partisi nilai dari atribut-atribut yang terdapat pada dataset ditunjukkan pada Tabel 1.

Tabel 2. Atribut dan Nilai Atribut

| Nama Atribut                 | Tipe Atribut   | Partisi Nilai              |
|------------------------------|----------------|----------------------------|
| <i>AGE</i>                   | <i>Integer</i> |                            |
| <i>GENDER</i>                | <i>Integer</i> | 0: Perempuan, 1: Laki-laki |
| <i>SMOKING</i>               | <i>Integer</i> | 0: Tidak, 1: Ya            |
| <i>FINGER_DISCOLORATION</i>  | <i>Integer</i> | 0: Tidak, 1: Ya            |
| <i>MENTAL_STRESS</i>         | <i>Integer</i> | 0: Tidak, 1: Ya            |
| <i>EXPOSURE_TO_POLLUTION</i> | <i>Integer</i> | 0: Tidak, 1: Ya            |
| <i>LONG_TERM_ILLNESS</i>     | <i>Integer</i> | 0: Tidak, 1: Ya            |
| <i>ENERGY_LEVEL</i>          | <i>Decimal</i> |                            |

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

|                               |                |                 |
|-------------------------------|----------------|-----------------|
| <i>IMMUNE_WEAKNESS</i>        | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>BREATHING_ISSUE</i>        | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>ALCOHOL_CONSUMPTION</i>    | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>THROAT_DISCOMFORT</i>      | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>OXYGEN_SATURATION</i>      | <i>Decimal</i> |                 |
| <i>CHEST_TIGHTNESS</i>        | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>FAMILY_HISTORY</i>         | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>SMOKING_FAMILY_HISTORY</i> | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>STRESS_IMMUNE</i>          | <i>Integer</i> | 0: Tidak, 1: Ya |
| <i>PULMONARY_DISEASE</i>      | <i>Boolean</i> | YES dan NO      |

Berdasarkan Tabel 2, dari 18 atribut pada dataset prediksi kanker paru-paru dimana 15 atribut bertipe integer, 2 atribut bertipe decimal, dan 1 atribut bertipe boolean. Terdapat dua kelas pada atribut PULMONARY\_DISEASE, yakni YES untuk pasien yang mengidap penyakit paru-paru dan NO untuk pasien yang tidak mengidap penyakit paru-paru.

### 3.1.2. Prapengolahan Data

Sebelum menganalisis data, pra pengolahan data dilakukan untuk mempersiapkan data yang lebih berkualitas dalam pemodelan (Harun et al., 2020). Dalam riset ini, beberapa tahapan pra-pengolahan telah dilakukan terhadap dataset kanker paru-paru. Pra-pengolahan dimulai dengan mengatasi *missing values* pada data menggunakan modus data, atau data yang paling sering muncul. Kemudian, konversi tipe data dilakukan untuk memastikan semua kolom memiliki tipe data yang benar. Kolom *PULMONARY\_DISEASE* yang menggunakan label kategorikan “NO” dan “YES” juga diubah menjadi nilai dengan tipe data numerik 0 dan 1 agar dapat dengan lebih mudah digunakan dalam machine learning melalui proses *encoding*. Setelah itu, penanganan outlier dilakukan pada beberapa kolom numerik termasuk AGE, ENERGY\_LEVEL, dan juga OXYGEN\_SATURATION menggunakan metode interquartile range (IQR). Pada tahap ini, sebanyak 61 baris data dihapus, sehingga jumlah data menjadi sebanyak 4.939 baris. Setelah seluruh data telah konsisten dan bersih, hasil data divisualisasikan dalam bentuk boxplot untuk melihat gambaran data sebelum dan setelah pra-pengolahan.

### 3.1.3. Pemilihan Fitur

Pemilihan fitur bertujuan untuk meningkatkan akurasi prediksi dan efisiensi pemrosesan data dengan mempertahankan fitur yang relevan dan menghilangkan yang tidak diperlukan (Sephya et al., 2023). Pada dataset kanker paru-paru, seluruh fitur dipertahankan karena memiliki keterkaitan langsung dengan risiko atau gejala penyakit tersebut. Usia dan jenis kelamin terbukti signifikan, di mana risiko dan mortalitas meningkat pada pria usia 70–79 tahun dan wanita 60–69 tahun, serta pria memiliki risiko lebih tinggi dengan rasio insidensi 2,46–3,01 dibanding Wanita (Tamási et al., 2021). Faktor gaya hidup dan lingkungan seperti merokok, konsumsi alkohol, riwayat keluarga, penyakit paru, dan paparan PM2.5 juga menjadi faktor risiko signifikan, terutama pada pria perokok yang terpapar polusi tinggi (Li et al., 2023). Stres mental, gangguan imun, dan penyakit jangka panjang meningkatkan insidensi serta menurunkan kelangsungan hidup pasien, sementara paparan asap dari anggota keluarga perokok juga memperbesar risiko (Chida et al., 2008). Gejala fisik seperti *Finger discoloration*, *energy level*, *immune weakness*, *breathing issue*, *throat discomfort*, *oxygen saturation*, dan

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

*chest tightness* membantu mengidentifikasi potensi kanker, sehingga seluruh fitur dinilai penting dalam membangun model prediksi yang akurat.

### 3.1.4. Pemisahan Data

Pemisahan data dilakukan untuk membagi dataset menjadi data latih dan data uji dengan rasio 80:20. Sebanyak 80% data digunakan untuk pelatihan model dan 20% sisanya digunakan untuk pengujian model. Pemisahan ini bertujuan untuk mengevaluasi sejauh mana model dapat bekerja dengan baik terhadap data yang belum pernah dilatih sebelumnya (Shujaaddeen et al., 2024).

### 3.1.5. Pelatihan Model

Model dilatih menggunakan algoritma KNN dari *library scikit-learn* pada bahasa pemrograman Python, sebagaimana ditunjukkan pada persamaan (1).

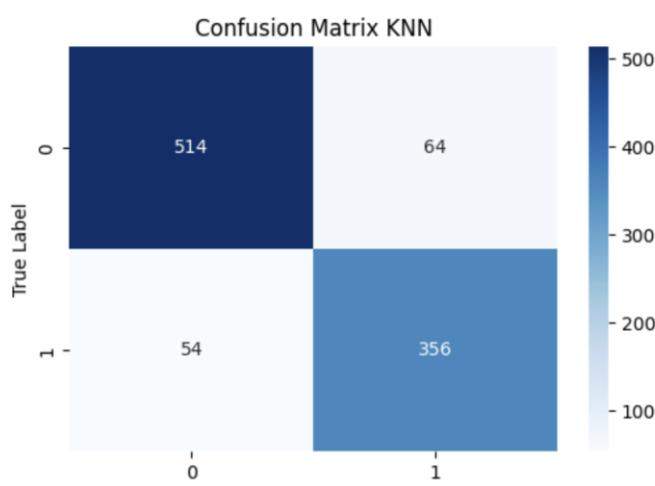
$$knn = KNeighborsClassifier(n_neighbors=20)$$

$$knn.fit(X\_train, y\_train) \quad (1)$$

Pelatihan dilakukan menggunakan data latih sebanyak 3.951 data dengan pendekatan eksperimental terhadap nilai k (jumlah tetangga terdekat). Percobaan dilakukan dengan berbagai nilai k, mulai dari 1 hingga 20 untuk menentukan parameter k yang menghasilkan performa terbaik. Berdasarkan hasil percobaan, ditemukan bahwa penggunaan k = 20 memberikan tingkat akurasi tertinggi, yaitu sebesar 0,89.

### 3.1.6. Pengujian dan Evaluasi Model

Berdasarkan hasil confusion matrix pada Gambar 2 dan Tabel 3, jumlah prediksi benar (*True Negative*) untuk kelas 0 (tidak terkena kanker paru-paru) sebanyak 514, dengan jumlah prediksi salah (*False Positive*) sebanyak 64. Presisi untuk kelas 0 tercatat sebesar 0,90, recall 0,89, dan F1-score 0,90. Untuk kelas 1 (terkena kanker paru-paru), jumlah prediksi benar (*True Positive*) sebanyak 356, dengan jumlah prediksi salah (*False Negative*) sebanyak 54.



Gambar 2. Confusion Matrix

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

Presisi untuk kelas 1 adalah 0,85, recall 0,87, dan F1-score 0,86. Secara keseluruhan, model mencapai akurasi sebesar 88% dari total 988 data yang diuji. Confusion matrix ini menunjukkan bahwa model KNN mampu memprediksi sebagian besar data dengan benar, meskipun terdapat beberapa kesalahan klasifikasi. Nilai presisi, recall, dan F1-score menunjukkan bahwa model bekerja cukup baik dalam menangani prediksi untuk kedua kategori.

Tabel 3. Confusion Matrix

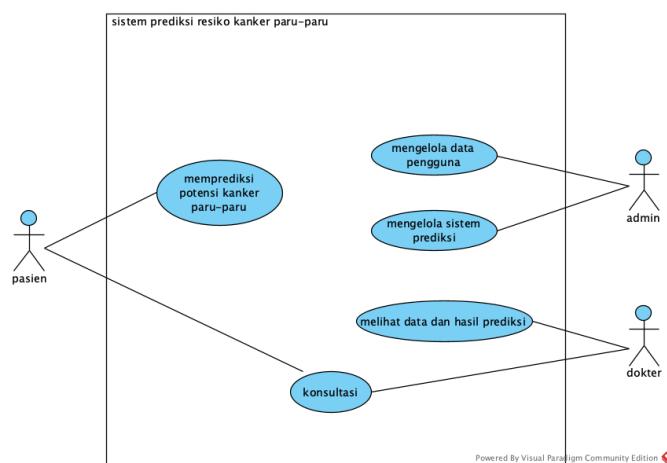
|                                | precision | recall | F1-score | support |
|--------------------------------|-----------|--------|----------|---------|
| Tidak Terkena Kanker Paru-paru | 0.90      | 0.89   | 0.90     | 578     |
| Terkena Kanker Paru-Paru       | 0.85      | 0.87   | 0.86     | 410     |
| Accuracy                       |           |        | 0.88     | 988     |
| Macro Avg                      | 0.88      | 0.88   | 0.88     | 988     |
| Weighted Avg                   | 0.88      | 0.88   | 0.88     | 988     |

### 3.2. Hasil Perancangan Sistem

Perancangan sistem dilakukan untuk menerjemahkan kebutuhan pengguna dan proses prediksi ke dalam bentuk sistem yang terstruktur. Tahapan ini mencakup pembuatan desain konseptual dan teknis sebagai dasar dalam pengembangan aplikasi. Bagian berikut menjelaskan secara rinci desain sistem yang digunakan dalam membangun aplikasi LungHealth.

#### 3.2.1. Desain Sistem

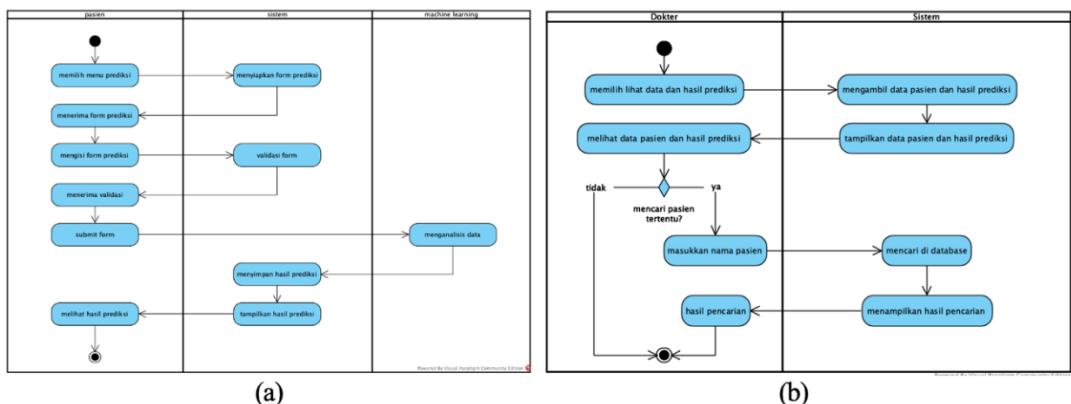
Desain sistem dibuat menggunakan diagram UML yang mencakup use case diagram, activity diagram, sequence diagram, dan class diagram sebagai acuan dalam pembuatan sistem (Kurniawan et al., 2020; Maulana, Heryana, et al., 2023). Tujuannya adalah untuk memberikan gambaran yang jelas mengenai fungsionalitas serta alur kerja sistem yang diinginkan. Use case diagram pada Gambar 3 menunjukkan interaksi antara pasien, dokter dan admin dengan sistem. Pasien dapat melakukan prediksi dan konsultasi, dokter dapat melihat data prediksi dan berkonsultasi dengan pasien, sementara admin mengelola data pengguna dan sistem prediksi.



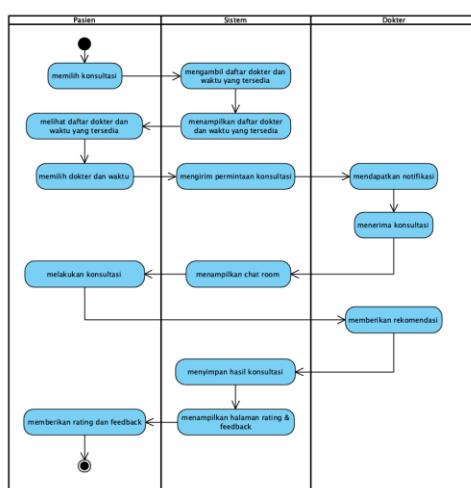
Gambar 3 Use Case Diagram

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

Pada Gambar 4a, pasien dapat mengisi form prediksi dan melihat hasil analisis dari sistem berbasis machine learning. Selain itu, pasien juga dapat melakukan konsultasi dengan dokter melalui proses pemilihan jadwal, komunikasi, dan pemberian feedback (Gambar 5). Gambar 4b menunjukkan bahwa dokter memiliki akses untuk melihat data pasien dan hasil prediksi guna memberikan rekomendasi. Sementara itu, admin bertugas mengelola data pengguna dan sistem prediksi, yang meliputi penambahan, pembaruan, pembacaan, dan penghapusan data pengguna maupun model yang digunakan.



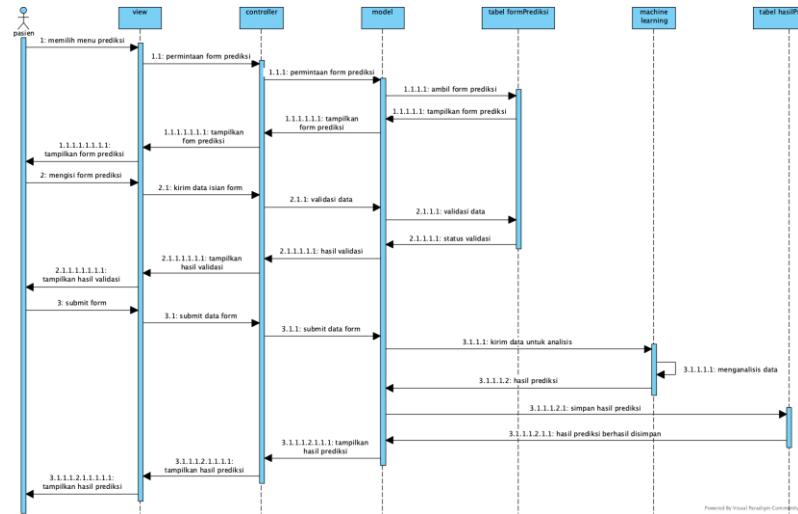
Gambar 4. (a) Activity Diagram Memprediksi Potensi Kanker Paru-Paru dan (b) Activity Diagram Melihat Data dan Hasil Prediksi



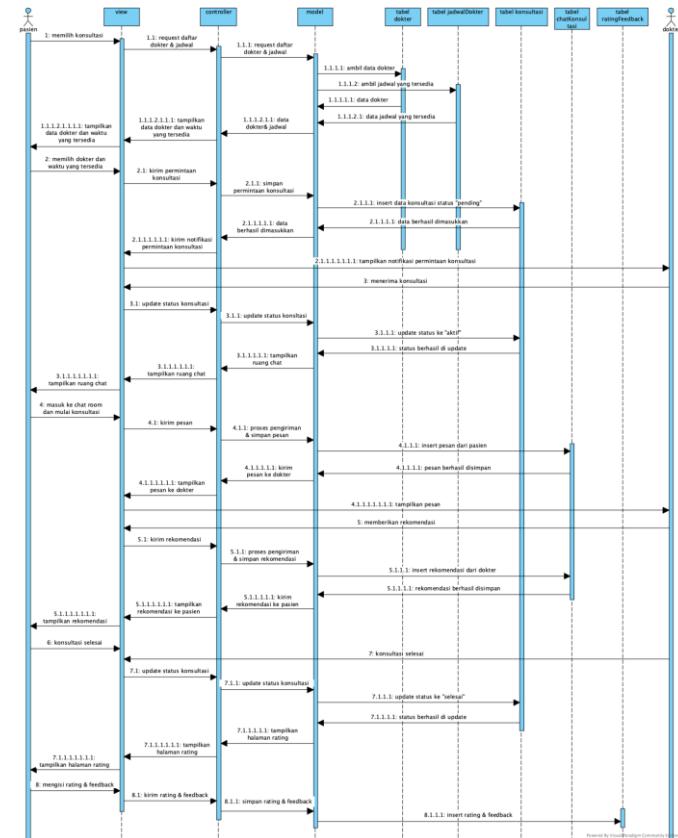
Gambar 5. Activity Diagram Konsultasi

Sequence diagram menggambarkan urutan interaksi antara aktor dan sistem dalam menjalankan fungsionalitas yang tersedia. Pasien dapat mengirimkan data untuk prediksi, kemudian sistem akan memproses dan menampilkan hasilnya, seperti ditunjukkan pada Gambar 6. Dalam proses konsultasi pada Gambar 7, pasien memilih jadwal, berkonsultasi dengan dokter, dan memberikan *feedback*. Dokter, sebagaimana pada Gambar 5, dapat melihat data pasien dan hasil prediksi untuk memberikan rekomendasi. Sementara itu, admin berinteraksi dengan sistem untuk mengelola data pengguna dan model prediksi, termasuk menambah, memperbarui, membaca dan menghapus data.

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors



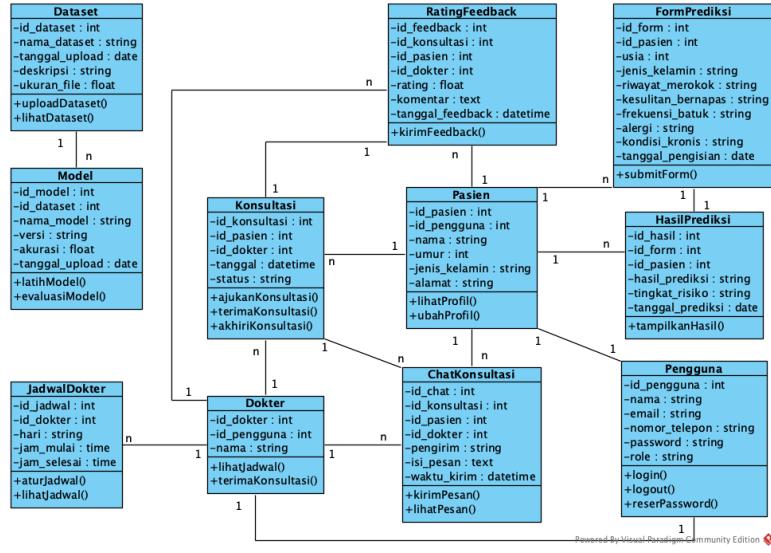
Gambar 6. Sequence Diagram Memprediksi Potensi Kanker Paru-Paru



Gambar 7. Sequence Diagram Konsultasi

Berdasarkan Gambar 8, class diagram menggambarkan struktur statis dari sistem yang terdiri atas berbagai kelas beserta atribut dan metodenya, yang merepresentasikan fungsi utama sistem (Aryani, 2020). Diagram ini juga menunjukkan hubungan antar kelas yang mencerminkan keterkaitan data serta alur interaksi dalam sistem secara keseluruhan.

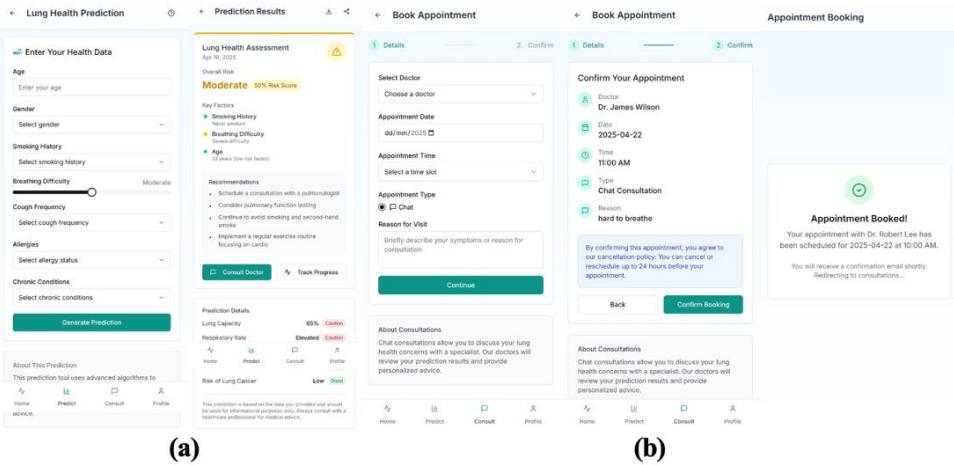
## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors



Gambar 8. Class Diagram

### 3.3. Implementasi

Tahap implementasi adalah proses penulisan kode program untuk membangun sistem berdasarkan desain yang telah dirancang sebelumnya (Maulana, Bau, et al., 2023; Pangaribuan et al., 2022). Dalam aplikasi LungHealth, terdapat tiga peran utama pengguna yaitu pasien, dokter, dan admin, yang masing-masing memiliki fungsi dan akses yang berbeda sesuai kebutuhan sistem.

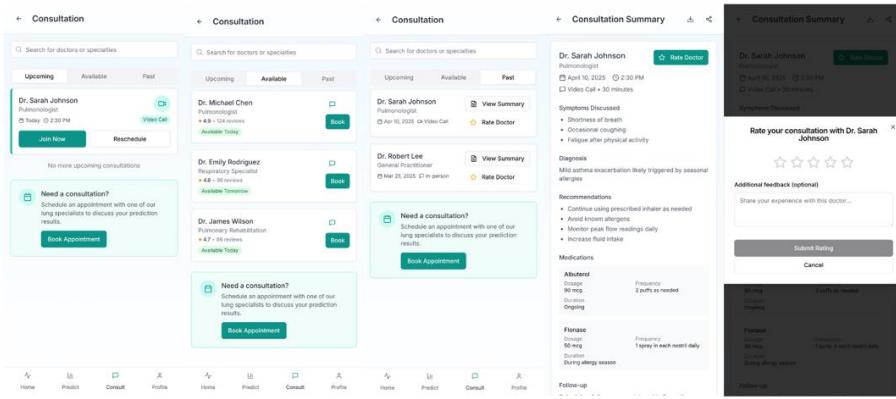


Gambar 9. (a) Halaman Pengisian Form Prediksi dan (b) Halaman Hasil Prediksi

Bagi pasien, sistem mewajibkan autentikasi terlebih dahulu melalui proses login atau pendaftaran dengan mengisi data pribadi seperti nama lengkap, email, jenis kelamin, tanggal lahir, dan password. Setelah berhasil masuk, pasien mengisi data kesehatan seperti usia, riwayat merokok, kesulitan bernapas, batuk, alergi, dan kondisi medis lainnya (Gambar 9a). Data ini digunakan sebagai input model prediksi untuk menentukan tingkat risiko serta memberikan rekomendasi gaya hidup sehat.

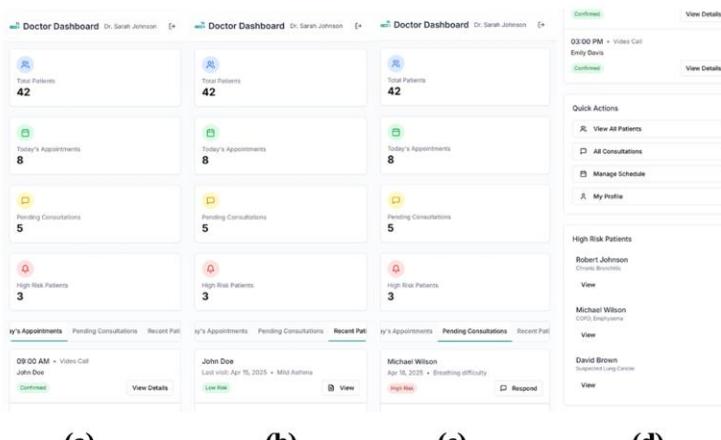
## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

Selanjutnya, pasien dapat memesan konsultasi dengan dokter berdasarkan spesialisasi, memilih waktu dan jenis konsultasi, serta mengisi keluhan yang dirasakan (Gambar 9b). Jadwal konsultasi akan dikonfirmasi, dan notifikasi akan dikirimkan kepada pasien. Konsultasi dilakukan secara langsung melalui aplikasi (Gambar 10). Setelah sesi berakhir, pasien menerima ringkasan hasil dan dapat memberikan penilaian terhadap dokter sebagai umpan balik.



Gambar 10. Halaman Konsultasi dan Pemberian Rating

Untuk dokter, aplikasi menyediakan dashboard khusus yang menampilkan jumlah pasien terdaftar, jadwal konsultasi hari ini, dan status konsultasi yang sedang berlangsung (Gambar 11). Dokter juga dapat mengakses daftar pasien berisiko tinggi untuk penanganan prioritas, merespons konsultasi yang tertunda, serta mengatur jadwal konsultasi secara virtual maupun tatap muka. Fitur-fitur ini dirancang untuk membantu dokter dalam mengelola waktu dan memberikan layanan medis secara lebih efektif dan efisien.

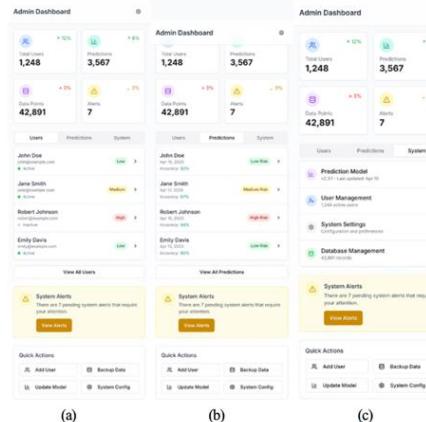


Gambar 11. Halaman Doctor Dashboard (a) Janji Temu, (b) Pasien Terakhir, (c) Konsultasi Tertunda dan (d) Halaman Konfirmasi Janji Temu

Sedangkan untuk admin, tersedia dashboard yang berfungsi untuk memantau keseluruhan aktivitas pengguna aplikasi, mulai dari jumlah pengguna terdaftar hingga hasil prediksi kesehatan yang telah dilakukan (Gambar 12). Admin bertanggung jawab atas

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

pemantauan status konsultasi, pembaruan model prediksi sesuai kebutuhan, serta menjaga keamanan data melalui backup rutin dan penanganan masalah teknis guna memastikan sistem aplikasi berjalan lancar dan aman.



Gambar 12. Halaman Admin Dashboard (a) Pengguna, (b) Riwayat Prediksi dan (c) System

### 3.4. Pengujian

Pada tahap ini, sistem diuji menggunakan metode black box untuk memastikan semua fungsi berjalan sesuai harapan tanpa melihat struktur kode di dalamnya (Lestari et al., 2023). Tabel 4 Hasil Pengujian Black Box menunjukkan bahwa semua skenario pengujian berhasil. Setiap skenario diuji dengan memberikan input sesuai fungsi sistem, lalu dicek apakah outputnya sesuai dengan yang diharapkan. Misalnya, untuk fungsi login, pengujian dilakukan dengan mengisi data login, lalu sistem harus menampilkan halaman utama. Hasil nyata sesuai dengan harapan, sehingga dinyatakan sukses. Hal yang sama juga dilakukan untuk fitur prediksi, konsultasi, manajemen pengguna, hingga tampilan dashboard dokter. Dengan demikian, sistem dinyatakan telah memenuhi fungsi utamanya dan layak digunakan.

Tabel 4. Hasil Pengujian Black Box

| No | Fungsi   | Skenario  | Hasil yang Diharapkan                                   | Hasil Nyata                                       | Hasil  |
|----|--|---|---|---|--------|
| 1. | Melakukan login/sign up                        | Masuk ke halaman login/sign up dan mengisi data         | Aplikasi menampilkan halaman home setelah login/sign up | Berhasil menampilkan halaman home                 | Sukses |
| 2. | Melakukan prediksi dan melihat hasil prediksi  | Pasien login, engisi form prediksi dan mengirimkan data | Aplikasi menampilkan hasil prediksi                     | Berhasil menampilkan hasil prediksi               | Sukses |
| 3. | Melakukan penjadwalan konsultasi dengan dokter | Memilih dokter dan mengisi form jadwal konsultasi       | Aplikasi menampilkan Konfirmasi jadwal konsultasi       | Berhasil menampilkan konfirmasi jadwal konsultasi | Sukses |

## Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan Algoritma K-Nearest Neighbors

|    |   |   |  |   |        |
|----|---|---|--|---|--------|
| 4. | Melakukan konsultasi dan memberikan rating kepada dokter    | Mengikuti konsultasi sesuai jadwal dan mengisi rating setelah selesai           | Aplikasi mencatat hasil konsultasi dan menampilkan form rating                                   | Berhasil mencatat hasil konsultasi dan menyimpan rating dokter                          | Sukses |
| 5. | Mengelola pengguna dan dataset/model                        | Admin login, mengelola akun pengguna, dataset/model                             | Aplikasi berhasil menambahkan, memperbarui, membaca, atau menghapus data pengguna, dataset/model | Berhasil menambahkan, memperbarui, membaca, atau menghapus data pengguna, dataset/model | Sukses |
| 6. | Melihat data pasien, hasil prediksi dan menerima konsultasi | Dokter login, melihat data dan hasil prediksi pasien, melihat jadwal konsultasi | Aplikasi menampilkan dashboard dokter yang memuat data pasien, hasil prediksi, jadwal konsultasi | Berhasil menampilkan informasi pasien, hasil prediksi dan jadwal konsultasi             | Sukses |

## 4. PENUTUP

### 4.1. Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma K-Nearest Neighbors (KNN) mampu memberikan hasil prediksi yang cukup baik dalam mendeteksi potensi kanker paru-paru. Dengan tingkat akurasi sebesar 88% dan performa presisi serta recall yang seimbang, model KNN terbukti efektif setelah dilakukan tahap pra-pengolahan data, pemilihan fitur yang relevan, serta penyetelan nilai k secara eksperimental. Model ini kemudian berhasil diintegrasikan ke dalam sistem aplikasi mobile "LungHealth" yang dilengkapi dengan fitur prediksi, konsultasi dokter, dan manajemen pengguna oleh admin. Sistem ini diharapkan dapat menjadi alat bantu yang bermanfaat dalam proses deteksi dini kanker paru-paru secara cepat, efisien, dan mudah diakses.

### 4.2. Saran

Untuk pengembangan lebih lanjut, disarankan agar penelitian ini memperluas cakupan algoritma dengan membandingkan kinerja KNN dengan metode machine learning lainnya, seperti Random Forest, Support Vector Machine, atau Deep Neural Network, guna memperoleh hasil yang lebih komprehensif. Selain itu, penggunaan data real-time dan pengujian langsung di lingkungan medis sesungguhnya sangat penting dilakukan agar sistem dapat beradaptasi dengan kondisi nyata. Penambahan fitur pendukung seperti riwayat pasien, integrasi dengan wearable device, dan fitur edukasi juga dapat meningkatkan nilai guna aplikasi bagi pasien maupun tenaga medis.

## DAFTAR PUSTAKA

- Aryani, Y. (2020). Sistem informasi penjualan barang dengan metode regresi linear berganda dalam prediksi pendapatan perusahaan. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 2(2), 39–51.
- Bharathy, S., & Pavithra, R. (2022). Lung cancer detection using machine learning. *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 539–543.
- Chida, Y., Hamer, M., Wardle, J., & Steptoe, A. (2008). Do stress-related psychosocial factors contribute to cancer incidence and survival? *Nature Clinical Practice Oncology*, 5(8), 466–475.
- Ganesh, S., & Baskar, R. (2025). Comparative analysis of early detection of lung cancer through breath analysis by comparing K-nearest neighbours (KNN) with Naive Bayes. *AIP Conference Proceedings*, 3270(1), 20169.
- Harun, R., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining untuk Menentukan Potensi Hujan Harian dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *Jurnal Manajemen Informatika Dan Sistem Informasi*, 3(1), 8–15.
- Kurniawan, H., Apriliah, W., Kurniawan, I., & Firmansyah, D. (2020). Penerapan metode waterfall dalam perancangan sistem informasi penggajian pada SMK Bina Karya Karawang. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 14(4), 159–169.
- Lestari, V. A., Ananta, A. Y., & Basudewa, P. (2023). Sistem informasi prediksi persediaan obat di apotek Naylun Farma menggunakan Holt-Winters. *Jurnal Informatika Polinema*, 9(2), 229–236.
- Li, D., Shi, J., Liang, D., Ren, M., & He, Y. (2023). Lung cancer risk and exposure to air pollution: a multicenter North China case-control study involving 14604 subjects. *BMC Pulmonary Medicine*, 23(1), 182.
- Liang, M., & Luo, F. (2025). Machine learning insights into early mortality risks for small cell lung cancer patients post-chemotherapy. *Frontiers in Medicine*, 12, 1483097.
- Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. *Expert Systems with Applications*, 164, 113981.
- Maulana, A., Bau, R. T. R. L., Hermila, A., Munawar, Z., Setiawan, R., Aisa, S., Istiono, W., Irfan, A., Pratama, Y. A., Ramadhany, E. D., & others. (2023). *Pemrograman Web 101 : Memahami Dasar-dasar untuk Mengembangkan Situs Web*. Get Press Indonesia.

Perancangan Aplikasi Mobile untuk Memprediksi Potensi Kanker Paru-Paru Menggunakan  
Algoritma K-Nearest Neighbors

- Maulana, A., Heryana, N., Pasaribu, J. S., Aditya, A., Elisawati, Rudiansyah, Amna, Permana, A. A., Rukmana, A. Y., Abdillah, R., & Wahyono, T. (2023). *Rekayasa Perangkat Lunak: Konsep, Metode, dan Praktik Terbaik*. Global Eksekutif Teknologi.
- Moon, K., & Jetawat, A. (2024). Predicting Lung Cancer with K-Nearest Neighbors (KNN): A Computational Approach. *Indian J. Sci. Technol*, 17(21), 2199–2206.
- Pangaribuan, J. J., Maulana, A., Romindo, R., & others. (2024). UNLEASHING THE POWER OF SVM AND KNN: ENHANCED EARLY DETECTION OF HEART DISEASE. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 10(2), 342–351.
- Pangaribuan, J. J., Maulana, A., Zesty, J., Nadjar, F., & others. (2022). Simpuru: Gamifikasi Pembelajaran Bahasa Jepang dalam Aplikasi Berbasis Web. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 5(2), 60–65.
- Sachdeva, R. K., Bathla, P., Rani, P., Lamba, R., Ghantasala, G. S. P., & Nassar, I. F. (2024). A novel K-nearest neighbor classifier for lung cancer disease diagnosis. *Neural Computing and Applications*, 36(35), 22403–22416.
- Sambi Ua, A., Lestriani H, D., Sonia Kristanty Marpaung, E., Ong, J., Savinka, M., Nurhaliza, P., & Yulia Ningsih, R. (2023). Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru. *Jurnal Publikasi Teknik Informatika*, 2(2), 88–99. <https://doi.org/10.55606/jupti.v2i2.1742>
- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
- Shujaaddeen, A. A., Ba-Alwi, F. M., Zahary, A. T., & Alhegami, A. S. (2024). A Model for Measuring the Effect of Splitting Data Method on the Efficiency of Machine Learning Models: A Comparative Study. *2024 4th International Conference on Emerging Smart Technologies and Applications (ESmarTA)*, 1–13.
- Sugiharto, S., Putri, R. A., Simanjuntak, S., & Larissa, O. (2021). *Kanker Paru, Faktor Risiko Dan Pencegahannya*.
- Tamási, L., Horvath, K., Kiss, Z., Bogos, K., Ostros, G., Müller, V., Urbán, L., Bittner, N., Sárosi, V., Vastag, A., & others. (2021). Age and gender specific lung cancer incidence and mortality in Hungary: Trends from 2011 through 2016. *Pathology and Oncology Research*, 27, 598862.