

# PENERAPAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN K-NEAREST NEIGHBORS (KNN) UNTUK KLASIFIKASI DATA KANKER PARU-PARU

*(Implementation of Principal Component Analysis and K-Nearest Neighbors Methods  
for Classification of Lung Cancer Data)*

Jaelani Muhammad Akbar, Muhammad Sabirin, Gibran Satya Nugraha\* , Noor Alamsyah

Program Studi Teknik Informatika, Universitas Mataram

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: jaelanima411@gmail.com, msabirin1123@gmail.com, [gibransn, nooralamsyah]@unram.ac.id

## Abstract

*Diseases related to breathing can be identified from the throat or the main airways, namely the lungs. It is caused by the spread of certain cells in the lungs and uncontrolled cell growth. Because lung nodules using chest X-ray imaging cannot be detected quickly, interpreting these diagnostic photographs becomes a repetitive and very complicated task. Image pre-processing, detection of malignancy classification, and candidate nodules can be used as parameters in classification. By analyzing CT scan images using an artificial intelligence approach, machine learning techniques help in early diagnosis and evaluation of lung nodules. This system is referred to as a decision support system that examines the image through the process of pre-processing, segmentation, feature extraction, and classification. Processing of lung cancer datasets to classify cancer and non-cancer samples using the Principal Component Analysis (PCA) feature extraction method and the K-Nearest Neighbor (KNN) classification method with values of  $K = 1, 3, 5, 7, 9$ . From the research that has been done, the best accuracy results are 98%, namely at  $K = 9$ . It can be concluded that the PCA feature extraction method and K-NN classification are suitable for processing lung cancer datasets.*

**Keywords:** kanker, paru-paru, citra, KNN, PCA

\*Penulis Korespondensi

## 1. PENDAHULUAN

Penyakit yang berkaitan dengan pernapasan dapat diidentifikasi dari tenggorokan atau saluran napas utama, yaitu paru-paru [1]. Hal ini disebabkan oleh penyebaran sel-sel tertentu di paru-paru dan pertumbuhan sel yang tidak terkendali. Orang yang menderita penyakit sesak nafas atau emfisema memiliki kemungkinan yang lebih besar untuk terkena kanker paru-paru [2]. *Small-cell lung carcinoma* (SCLC) dan *non-small-cell lung carcinoma* adalah dua jenis utama kanker paru-paru. SCLC hampir pasti disebabkan karena merokok dan perkembangannya juga lebih cepat. Sedangkan, *non-small-cell lung carcinoma* lebih umum terjadi dan penyeberannya lebih lambat [3].

Menurut statistik terbaru pada tahun 2018, sekitar 234.030 kasus kanker paru-paru sekitar 85 persen mengarah ke *non-small-cell lung carcinoma* [4]. Faktor yang paling signifikan dalam membuat penyakit ini begitu mematikan adalah perkembangan kanker paru-paru tanpa gejala. Hampir seperempat orang tidak

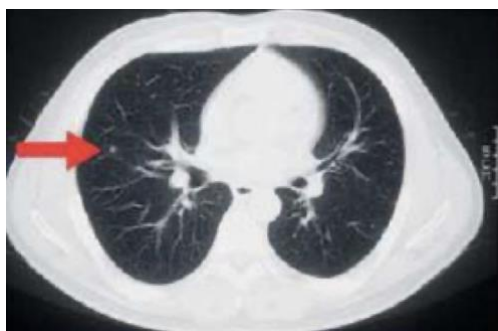
memiliki tanda-tanda kanker. Banyak orang tahu bahwa kanker paru-paru memiliki kondisi lain yang menyebabkan rontgen paru-paru menjadi kurang akurat. Diagnosis dini sangat penting, karena kanker paru-paru juga dapat menyebar dengan cepat. Saat ini, kanker paru-paru dapat didiagnosis pada tahap awal dengan perkembangan teknologi pencitraan seperti *computed tomography* dosis rendah [5].

Sebuah tumor yang disebut *nodule* dari sel-sel di saluran udara dari sistem pernapasan menyebabkan kanker paru-paru. Sel-sel ini apabila dicek melalui rontgen dada terlihat kontras dan berbentuk benda bulat. Jika *nodule* pada paru-paru dapat dideteksi secara akurat pada tahap awal, tingkat kelangsungan hidup pasien dapat ditingkatkan secara signifikan. Karena *nodule* paru-paru menggunakan pencitraan rontgen dada tidak dapat dideteksi dengan cepat, maka dalam menafsirkan foto-foto diagnostik ini menjadi tugas yang berulang dan sangat rumit [6]. Kehadiran *nodule* kecil dari CT scan paru-paru 3D harus dideteksi oleh sistem diagnosis berbantuan komputer (*computer aided system*) [7], [8].

Gambar 1 memberikan contoh *nodule* kanker paru stadium awal yang terlihat pada irisan 2D CT scan. *Noise* dari jaringan di sekitarnya, udara, dan tulang memenuhi CT scan, jadi *noise* ini pertama-tama akan untuk diproses terlebih dahulu agar sistem CAD dapat mengidentifikasi secara efektif. Pre-processing citra, deteksi klasifikasi keganasan, dan kandidat *nodule* dapat digunakan sebagai parameter dalam klasifikasi [9] Dengan menganalisis citra CT scan yang menggunakan pendekatan kecerdasan buatan, teknik *machine learning* membantu dalam diagnosis dini dan evaluasi *nodule* paru-paru [10]. Sistem ini disebut sebagai sistem pendukung keputusan yang memeriksa citra melalui proses *pre-processing*, segmentasi, ekstraksi ciri, dan klasifikasi seperti yang ditunjukkan pada Gambar 2 [11]

Data mining adalah serangkaian aktivitas menggunakan data masa lalu untuk memprediksi kejadian masa depan. Salah satu metode pembelajaran mesin yang dapat menangani data yang kompleks yaitu menggunakan teknik klasifikasi. Terdapat banyak metode yang ada dalam klasifikasi diantaranya adalah Naïve Bayes Classifier, Support Vector Machine, dan lain-lain. Pada penelitian ini metode klasifikasi yang digunakan adalah metode K-Nearest Neighbour (KNN). Untuk mempermudah dan mempersingkat waktu pada proses komputasi, sebelum dilakukannya klasifikasi maka dilakukan ekstraksi fitur. Pada penelitian ini metode ekstraksi fitur yang digunakan adalah metode Principal Component Analysis (PCA).

Metode Principal Component Analysis (PCA) mempunyai kelebihan yaitu kemampuan untuk mempercepat proses komputasi dikarenakan pada metode PCA melakukan reduksi fitur tanpa menghilangkan karakteristik aslinya. Sedangkan untuk metode K-Nearest Neighbour (KNN) mempunyai kelebihan yaitu handal digunakan terhadap *training* data yang mengandung *noise* dan efektif apabila data latih nya besar.



Gambar 1. Citra CT scan kanker paru-paru [9]

## 2. TINJAUAN PUSTAKA

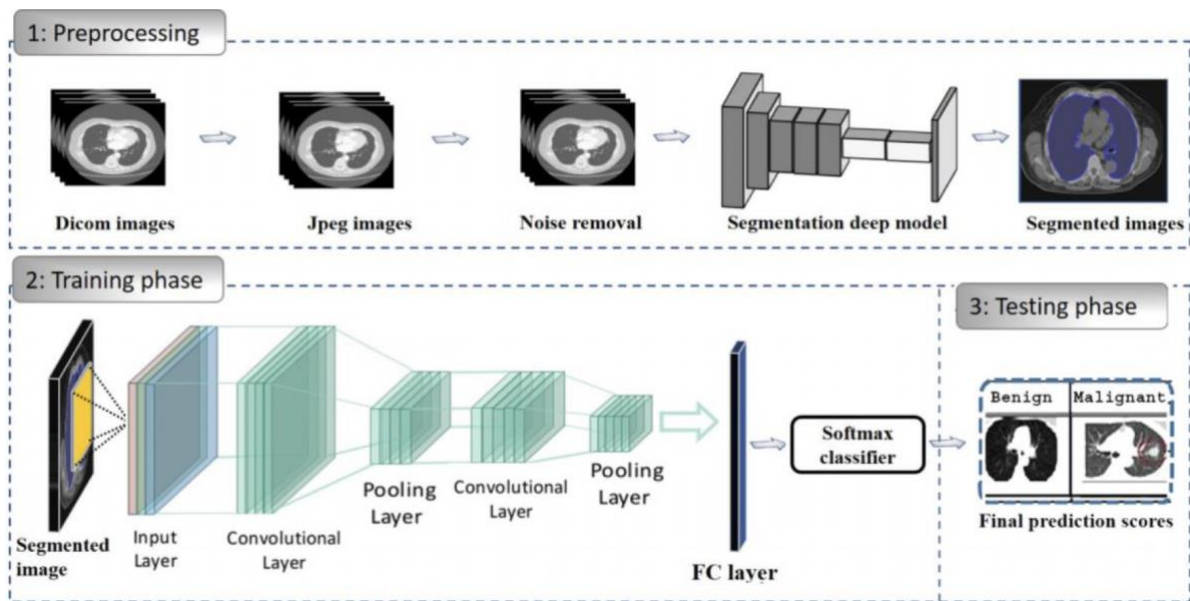
### 2.1 Penelitian Terkait

Seperti di bidang lain, para peneliti secara efektif menerapkan teknik statistik dan pembelajaran mesin untuk mendeteksi penyakit seperti kanker paru-paru, serta membangun model prediksinya. Deteksi kanker paru-paru sebelumnya mampu diidentifikasi menggunakan teknik pengolahan citra oleh [12], dengan menambahkan pendekatan *deep learning* serta jaringan saraf tiruan konvensional. Beberapa peneliti [13], [14] melakukan penelitian dengan mengklasifikasikan citra kanker paru-paru dengan menggunakan teknik jaringan saraf konvensional dan pembelajaran mesin. Baru-baru ini, beberapa peneliti telah mencoba menggunakan beberapa arsitektur dari *deep learning* untuk deteksi kanker paru-paru [15], [16].

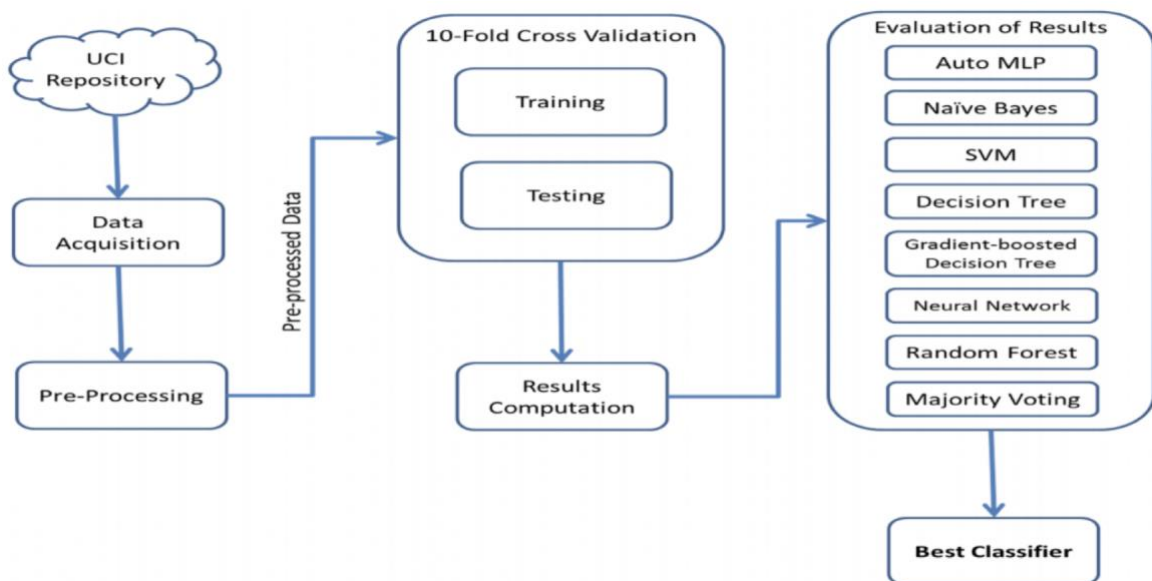
Menggunakan teknik pembelajaran mesin dan susunan sensor kimia juga telah dilakukan [16], penelitian tersebut mengembangkan tes pernapasan untuk mendeteksi kanker paru-paru. Sebuah penelitian prospektif untuk mencatat kasus kanker paru-paru antara tahun 2016 dan 2018 dilakukan, dan satu sampel udara alveolar dianalisis menggunakan *array* sensor karbon *nanotube*. Untuk derivasi model dan validasi digunakan data pasien pada tahun 2016 dan 2017 digunakan. Pada subjek yang direkrut pada tahun 2018, model divalidasi secara eksternal. Menggunakan catatan patologis sebagai standar referensi, akurasi diagnostik dievaluasi. Akurasi yang didapatkan adalah 95%. Para penulis menyimpulkan bahwa kanker paru-paru dapat diidentifikasi dengan presisi tinggi dengan menggabungkan teknik susunan sensor dan pembelajaran mesin.

Chauhan dan Jaiswal merekomendasikan klasifikasi otomatis penyakit dengan pembelajaran mesin berdasarkan pendekatan praktis untuk mendeteksi prinsip kanker paru-paru. Selain itu, beberapa set *benchmark* menunjukkan bahwa model kerja yang diusulkan dapat dibandingkan dengan pendekatan konvensional lainnya [17].

Nasser dan Abu-Naser telah meningkatkan model Jaringan Syaraf Tiruan (JST) untuk mengkonfirmasi deteksi kanker paru-paru, Gejala seperti kecemasan, jari kuning, penyakit pernapasan, kelelahan, alergi, mengi, batuk, sesak napas, nyeri dada dan kesulitan menelan telah digunakan diagnosis kanker paru-paru. Model JST yang diusulkan dikembangkan, dilatih, dan divalidasi menggunakan kumpulan data yang disebut "survei kanker paru-paru" sebagai variabel *input* untuk model JST yang diusulkan dan informasi lain tentang individu sebagai variabel *input* lainnya. Evaluasi model



Gambar 2. Alur proses deteksi kanker paru-paru [11]



Gambar 3. Diagram deteksi kanker paru-paru [18]

mengungkapkan bahwa, dengan akurasi 96,67 persen, model JST yang diusulkan mampu mendeteksi ada tidaknya kanker paru-paru [13].

Prediksi pasien yang rentan terhadap kanker paru-paru akan membantu dokter dalam pengambilan keputusan pengobatan mereka. Dalam hal ini, penelitian yang dilakukan oleh [18] mencoba mengukur kapasitas diskriminatif beberapa prediktor untuk meningkatkan keakuratan diagnosis kanker paru

berdasarkan gejalanya. Berbagai metode klasifikasi diujikan pada dataset yang diambil dari repositori UCI, termasuk SVM, Multi-Layer Perceptron, Naïve Bayes, C4.5 Decision tree, Neural Network, seperti yang terlihat pada Gambar 3. *Ensemble* yang familiar seperti Majority Voting dan Random Forest juga dibandingkan dengan *output*, akurasi yang didapatkan sebesar 90%.

## 2.2 Landasan Teori

Pembelajaran mesin adalah sebuah seni dalam mempelajari komputer, di mana mesin dibuat dengan algoritme yang dapat membuatnya sendiri dan memberikan hasilnya kepada pengguna [19].

Pembelajaran mesin pada dasarnya adalah bagian dari kecerdasan buatan [20], [21]. Untuk klasifikasi data dan pengambilan keputusan yang rumit, *machine learning* adalah pilihan yang dapat digunakan saat ini. Secara umum, pengembangan algoritme membantu mesin untuk mempelajari dan membuat keputusan yang diperlukan [22]–[24].

Teknik dan tugas *machine learning* secara luas diklasifikasikan ke dalam tiga kategori:

- a. Pembelajaran terawasi (*supervised learning*): tipe ini memecahkan masalah regresi, seperti peramalan cuaca, prediksi pertumbuhan populasi dan prediksi pengalaman hidup menggunakan algoritma Regresi Linier atau *Random Forest* [25], [26]. Selain itu, pembelajaran terawasi dapat memecahkan masalah klasifikasi seperti pengenalan suara, pengenalan angka, diagnostik, dan deteksi penipuan identitas dengan menggunakan algoritma di banyak bidang, seperti Support Vector Machines, Random Forest, Nearest Neighbor, dan lain-lain yang digunakan di banyak bidang [27], [28]. Dalam pembelajaran terawasi, ada dua tingkatan, fase pelatihan dan fase pengujian. Harus ada label yang diketahui dalam kumpulan data yang digunakan untuk proses pelatihan. Algoritme yang digunakan mampu mempelajari hubungan antara nilai *input* dan label dan mencoba memprediksi hasil dari data pengujian [29], [30].
- b. Pembelajaran tanpa pengawasan (*unsupervised learning*): model ini dapat digunakan untuk mengurangi dimensi dan visualisasi data yang besar, elaborasi fitur, atau menemukan struktur tersembunyi. Ini juga digunakan untuk masalah pengelompokan seperti kerangka rekomendasi, segmentasi pelanggan, dan pemasaran yang ditargetkan [31]. Dibandingkan dengan pembelajaran yang diawasi, tidak ada label yang tersedia dalam metode ini. Pada tipe ini, algoritma bertujuan untuk mengenali pola pada data pengujian dan memprediksi nilai masa depan atau pengelompokan data [29], [30]

- c. *Reinforcement learning*: jenis pembelajaran ini dilakukan berdasarkan kumpulan parameter dalam algoritma untuk memprediksi luaran dari suatu masalah. Kemudian *output* menjadi parameter *input*, dan kemudian *output* baru ditemukan ketika *output* optimal telah ditemukan. *Deep Learning and Artificial Neural Networks (ANN)* menggunakan model seperti ini [32]. Aplikasi yang menggunakan *reinforcement learning* seperti navigasi robot, *skill acquisition*, *real-time decision*, dan permainan berbasis *artificial intelligence* [29], [30].

Beberapa tahapan atau proses dalam penelitian ini, dijelaskan sebagai berikut:

- a. Ekstraksi Fitur  
Ekstraksi fitur adalah sebuah proses mengambil atau mengekstrak ciri khusus dari sebuah objek. Ciri atau karakteristik yang diambil merupakan karakteristik yang menjadi ciri khas dari objek tersebut [26].
- b. Klasifikasi  
Klasifikasi adalah sebuah proses pengelompokan beberapa data menjadi kelas-kelas tertentu berdasarkan pola-pola yang sudah dianalisis sebelumnya. Proses klasifikasi dimulai dengan pelatihan dari sejumlah data untuk mencari pola yang dicari. Hasil yang didapatkan dari proses ini adalah sebuah model yang digunakan untuk pengujian [33]
- c. Principal Component Analysis (PCA)  
Principal Component Analysis (PCA) adalah suatu metode untuk menseleksi fitur yang sudah diekstrak. Tujuan dari proses ini adalah untuk mengurangi dimensi dari *dataset* yang digunakan [34]. Berikut algoritma dari metode PCA [35]

1. Membentuk matriks parameter klasifikasi dari data yang akan dilatih, seperti yang terlihat pada Persamaan (1).

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & & x_{2n} \\ x_{31} & x_{32} & x_{33} & & x_{3n} \\ \dots & & & & \dots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix} \quad (1)$$

2. Menghitung rata-rata nilai dari seluruh parameter terhadap semua data yang

digunakan,  $m$  adalah jumlah data. Dapat dilihat pada Persamaan (2)

$$\tau = \frac{(x_{12} + x_{21} + x_{31} + \dots + x_{m1})}{m}$$

$$= \frac{\sum_{j=1}^m x_{ji}}{m} = [\tau_1, \tau_2, \dots, \dots, \tau_n] \quad (2)$$

3. Penggandaan rata-rata atau *mean*. Proses ini dilakukan untuk membuat dimensi nilai rata-rata sama dengan dimensi nilai pada citra sampel.  $a$  adalah matriks kolom yang memiliki dimensi  $m \times 1$ , dengan Rumus yang digunakan seperti pada Persamaan (3)

$$\mu = \alpha x \tau \quad (3)$$

4. Melakukan perhitungan terhadap nilai yang memiliki rata-rata nol. Hal itu dilakukan agar hasil perhitungan PCA menjadi lebih akurat. Rumus yang digunakan seperti pada Persamaan (4).

$$\phi = x - \mu \quad (4)$$

5. Memperoleh nilai matriks kovarian berdasarkan hasil perhitungan rata-rata nol. Dapat dihitung menggunakan Persamaan (5)

$$c = \frac{1}{m-1} \phi_{ji} * \phi^T_{ji} \quad (5)$$

6. Menentukan matriks *eigenvalue* dan *eigenvector*. Dapat dihitung dengan Persamaan (6). Dimana  $C$  adalah matriks kovarian, sedangkan  $Z$  adalah matriks *eigenvalue* dengan  $\lambda$  adalah matriks skalarnya
- $$C - Z = |C - \lambda I| \quad (6)$$

- d. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma *supervised learning* yang digunakan untuk melakukan klasifikasi berdasarkan jumlah  $k$  (tetangga) terdekatnya [36] Rumus yang digunakan untuk perhitungan KNN dapat dilihat pada Persamaan (7) [37]. Dengan  $d(i,j)$  adalah nilai jarak,  $x_i$  adalah nilai pada fitur 1, dan  $x_j$  nilai pada fitur 2

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (7)$$

### 3. METODE PENELITIAN

Alur kerja dalam penelitian ini dimulai dari studi pustaka sebagai langkah awal untuk mengumpulkan data dari penelitian terdahulu yang relevan dengan penelitian yang akan di lakukan yaitu tentang penerapan metode principal component analysis (PCA) dan k-nearest neighbors (KNN) untuk klasifikasi data kanker paru-paru.

- a. Data Penelitian

Data yang digunakan merupakan data sekunder dari kaggle yang berjudul "Lung Cancer DataSet" dengan <https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset>

- b. Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini yaitu ekstraksi fitur menggunakan metode Principal Component Analysis (PCA). Adapun proses PCA yang dilakukan yaitu:

1. Normalisasi data
2. Menghitung covariance
3. Analisis eigen
4. Analisis data input menjadi data vector.

Kemudian melakukan klasifikasi menggunakan metode K-Nearest Neighbour (KNN) dengan langkah:

1. Split data
2. Training KNN Model untuk  $k=1,3,5,7$  dan 9.
3. Testing dan bandingkan hasilnya

Analisis ini digunakan untuk pengklasifikasian kanker paru paru dengan dua kelas yaitu jaringan kanker dan tidak kanker. Perangkat lunak yang digunakan pada penelitian ini adalah software Jupiter Notebook. Variabel Penelitian dan Definisi Operasional Variabel. Adapun macam-macam variabel yang digunakan pada penelitian terdapat pada Tabel 1:

TABEL 1. VARIABEL YANG DIGUNAKAN DALAM PENELITIAN

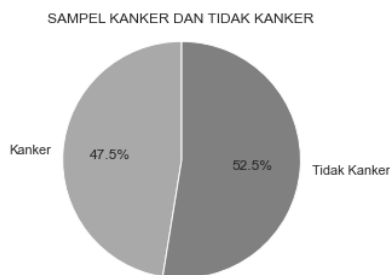
Nama variabel	Definisi Operasional variabel
Age	Merupakan umur dari pasien dengan rentang usia dari 18 tahun sampai 70 tahun
Smokes	merupakan jumlah konsumsi

	rokok pasien dengan rentang konsumsi rokok dari 0 sampai 32
AreaQ	Daerah terkena kanker paru paru
Alkohol	merupakan jumlah konsumsi alkohol pasien dengan rentang konsumsi alkohol dari 0 sampai 8
Result	merupakan kelas dari dataset yang dimana penderita kanker ditandai dengan 1 dan tidak kanker ditandai dengan 0

#### 4. HASIL DAN PEMBAHASAN

##### a. Analisis Deskriptif

Dataset kanker paru paru merupakan kumpulan data pasien yang terdiri dari 59 sampel yang dibedakan menjadi dua kelas yaitu kanker dan tidak kanker.



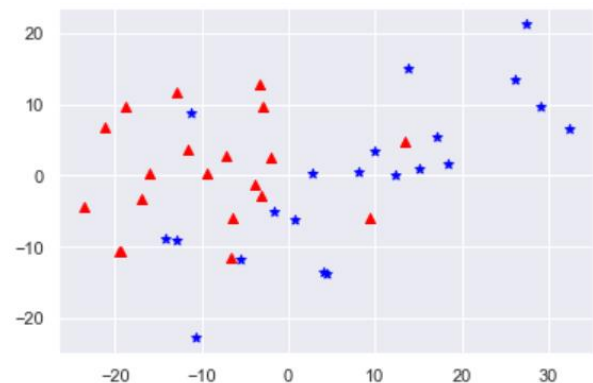
Gambar 4. Pieplot persebaran data kanker dan tidak kanker

Pada Gambar 4 memperlihatkan persebaran dataset dari 59 sampel 47,5% diantaranya adalah sampel kanker dan 52,5% adalah sampel tidak kanker.

##### b. Ekstraksi Fitur

Melakukan ekstraksi fitur dengan menggunakan metode PCA. Hal ini bertujuan untuk mereduksi size menjadi dua dengan memilih dua eigen dominan dari input feature vektor yang berdimensi empat.

Setelah dilakukan metode ekstraksi fitur PCA didapatkan hasil plotting seperti pada Gambar 5.



Gambar 5. Hasil plotting data setelah proses PCA

##### Pembagian data testing dan training

Data yang telah diubah menjadi 2 dimensi menggunakan metode PCA kemudian dilakukan pembagian data testing dan data training dengan perbandingan 70% vs 30%. Dimana data training 70% dan data testing 30%, detail dapat dilihat pada Tabel 2

TABEL 2. JUMLAH DATA TRAINING DAN TESTING

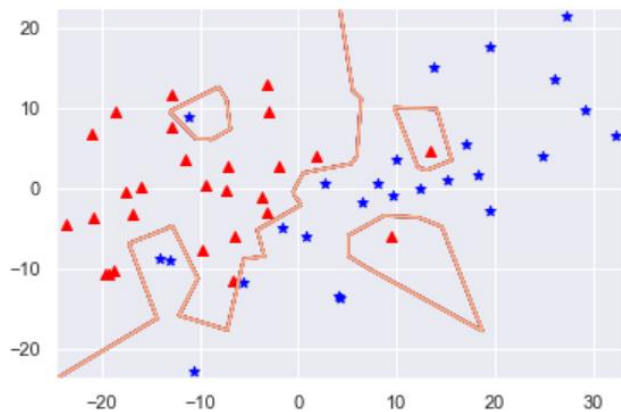
training	testing
45	14

Proses klasifikasi menggunakan metode K-NN dari dataset kanker paru paru dengan data training sebanyak 45 sampel dan data testing sebanyak 14 sampel mendapatkan hasil akurasi sebagai berikut.

TABEL 3. HASIL KLASIFIKASI DENGAN KNN

Nilai K	Akurasi
1	77.77%
3	83.33%
5	94.44%
7	97%
9	98%





Gambar 6. Hasil plotting klasifikasi garis KNN

## 5. KESIMPULAN DAN SARAN

Pengolahan dataset kanker paru paru untuk mengklasifikasikan sampel kanker dan tidak kanker menggunakan metode ekstraksi fitur PCA dan metode klasifikasi KNN dengan nilai  $K = 1, 3, 5, 7, 9$ . Dari penelitian yang telah dilakukan didapatkan hasil akurasi terbaik sebesar 98% yaitu pada  $K = 9$ . Dapat disimpulkan bahwa metode ekstraksi fitur PCA dan klasifikasi K-NN cocok digunakan untuk melakukan Pengolahan dataset kanker paru paru.

## DAFTAR PUSTAKA

- [1] K.-H. Yu *et al.*, "Reproducible machine learning methods for lung cancer detection using computed tomography images: Algorithm development and validation," *J Med Internet Res*, vol. 22, no. 8, p. e16709, 2020.
- [2] P. R. Radhika, R. A. S. Nair, and G. Veena, "A comparative study of lung cancer detection using machine learning algorithms," in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2019, pp. 1–4.
- [3] L. Hussain, S. Rathore, A. A. Abbasi, and S. Saeed, "Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques," in *Medical Imaging 2019: Physics of Medical Imaging*, 2019, vol. 10948, pp. 919–925.
- [4] L. A. Torre *et al.*, "Ovarian cancer statistics, 2018," *CA Cancer J Clin*, vol. 68, no. 4, pp. 284–296, 2018.
- [5] Ö. Günaydin, M. Günay, and Ö. Şengel, "Comparison of lung cancer detection algorithms," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, 2019, pp. 1–4.
- [6] A. Elnakib, H. M. Amer, and F. E. Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," 2020.
- [7] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable model based on new uniform LBP feature to identify the risk of the breast cancer," in *2019 international conference on advanced science and engineering (ICOASE)*, 2019, pp. 106–111.
- [8] D. A. Zebari, D. Q. Zeebaree, A. M. Abdulazeez, H. Haron, and H. N. A. Hamed, "Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images," *Ieee Access*, vol. 8, pp. 203097–203116, 2020.
- [9] W. Alakwaa, M. Nassef, and A. Badr, "Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, 2017.
- [10] D. Q. Zeebaree, A. M. Abdulazeez, D. A. Zebari, H. Haron, and H. N. A. Hamed, "Multi-level fusion in ultrasound for cancer detection based on uniform LBP features," *Computers, Materials & Continua*, vol. 66, no. 3, pp. 3363–3382, 2021.
- [11] T. Saba, "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges," *J Infect Public Health*, vol. 13, no. 9, pp. 1274–1289, 2020.
- [12] B. Abdillah, A. Bustamam, and D. Sarwinda, "Image processing based detection of lung cancer on CT scan images," in *Journal of Physics: Conference Series*, 2017, vol. 893, no. 1, p. 012063.
- [13] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 3, no. 3, pp. 17–23, 2019.
- [14] M. I. Sharif, J. P. Li, J. Naz, and I. Rashid, "A comprehensive review on multi-organs tumor detection based on machine learning," *Pattern Recognit Lett*, vol. 131, pp. 30–37, 2020.
- [15] A. Bhandary *et al.*, "Deep-learning framework to detect lung abnormality—A study with chest X-Ray and lung CT scan images," *Pattern Recognit Lett*, vol. 129, pp. 271–278, 2020.
- [16] X. Li *et al.*, "Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection," *Artif Intell Med*, vol. 103, p. 101744, 2020.
- [17] A. Varshney, A. Katiyar, A. K. Singh, and S. S. Chauhan, "Dog Breed Classification Using Deep Learning," in *2021 International Conference on*

- Intelligent Technologies (CONIT)*, 2021, pp. 1–5. doi: 10.1109/CONIT51480.2021.9498338.
- [18] M. I. Faisal, S. Bashir, Z. S. Khan, and F. H. Khan, "An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer," in *2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST)*, 2018, pp. 1–4.
- [19] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, and W. M. Abdulllah, "Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods," *IEEE Access*, vol. 7, pp. 51691–51713, 2019.
- [20] L. F. Ramos-Lima, V. Waikamp, T. Antonelli-Salgado, I. C. Passos, and L. H. M. Freitas, "The use of machine learning techniques in trauma-related disorders: a systematic review," *J Psychiatr Res*, vol. 121, pp. 159–172, 2020.
- [21] Z. E. Abou El Assad, H. Mousannif, H. al Moatassime, and A. Karkouch, "The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review," *Eng Appl Artif Intell*, vol. 87, p. 103312, 2020.
- [22] M. Somvanshi, P. Chavan, S. Tambade, and S. v Shinde, "A review of machine learning techniques using decision tree and support vector machine," in *2016 international conference on computing communication control and automation (ICCUBEA)*, 2016, pp. 1–7.
- [23] C. Maione, F. Barbosa Jr, and R. M. Barbosa, "Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review," *Comput Electron Agric*, vol. 157, pp. 436–446, 2019.
- [24] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Machine learning and region growing for breast cancer segmentation," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 88–93.
- [25] A. Moujahid et al., "Machine learning techniques in ADAS: a review," in *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2018, pp. 235–242.
- [26] D. M. Abdulqader, A. M. Abdulazeez, and D. Q. Zeebaree, "Machine learning supervised algorithms of gene selection: A review," *Mach Learn*, vol. 62, no. 03, pp. 233–244, 2020.
- [27] N. S. Ahmed and M. H. Sadiq, "Clarify of the random forest algorithm in an educational field," in *2018 international conference on advanced science and engineering (ICOASE)*, 2018, pp. 179–184.
- [28] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 145–150.
- [29] M. Kubat and Kubat, *An introduction to machine learning*, vol. 2. Springer, 2017.
- [30] F. Zantalis, G. Koulouras, S. Karabetsos, and D. Kandris, "A review of machine learning and IoT in smart transportation," *Future Internet*, vol. 11, no. 4, p. 94, 2019.
- [31] D. M. Sulaiman, A. M. Abdulazeez, H. Haron, and S. S. Sadiq, "Unsupervised learning approach-based new optimization K-means clustering for finger vein image localization," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 82–87.
- [32] S. S. Al-jaboriy, N. N. A. Sjarif, S. Chuprat, and W. M. Abdulllah, "Acute lymphoblastic leukemia segmentation using local pixel information," *Pattern Recognit Lett*, vol. 125, pp. 85–90, 2019.
- [33] M. I. Faisal, S. Bashir, Z. S. Khan, and F. H. Khan, "An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer," in *2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST)*, 2018, pp. 1–4.
- [34] D. Kumar, "Feature extraction and selection of kidney ultrasound images using GLCM and PCA," *Procedia Comput Sci*, vol. 167, pp. 1722–1731, 2020.
- [35] P. Rosyani, "Pengenalan Wajah Menggunakan Metode Principal Component Analysis (PCA) dan Canberra Distance," *Jurnal Informatika Universitas Pamulang*, vol. 2, no. 2, pp. 118–121, 2017.
- [36] J. Xiao, "SVM and KNN ensemble learning for traffic incident detection," *Physica A: Statistical Mechanics and its Applications*, vol. 517, pp. 29–35, 2019.
- [37] D. Syahid, J. Jumadi, and D. Nursantika, "Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV)," *Jurnal Online Informatika*, vol. 1, no. 1, pp. 20–23, 2016.