

# Klasifikasi Kanker Paru-Paru Berdasarkan Diagnosis Menggunakan Ensemble Learning: Analisis Support Vector Machine, Decision Tree, Light Gradient Boosting Machine, Dan Extreme Gradient Boosting

Ahmad Ubai Dullah<sup>1</sup>, Alamsyah<sup>2</sup>

<sup>1,2</sup>Jurusan Ilmu Komputer, FMIPA, Universitas Negeri Semarang

Email: <sup>1</sup>[ubaid@students.unnes.ac.id](mailto:ubaid@students.unnes.ac.id), <sup>2</sup>[alamsyah@mail.unnes.ac.id](mailto:alamsyah@mail.unnes.ac.id)

## Abstrak

Kanker Paru-paru merupakan salah satu jenis penyakit yang menyebabkan kematian Sebagian besar manusia di dunia. Kanker paru-paru terjadi akibat tidak normalnya pertumbuhan sel pada organ paru-paru. Sel yang tidak normal akan tumbuh menjadi sel kanker dan akan menyerang organ yang ditumbuhi sel kanker tersebut. Gejala yang sering dialami oleh penderita kanker paru-paru antara lain seperti, sesak napas, nyeri dada, batuk, mengi dan jari menguning. Kanker paru-paru juga bisa terjadi karena gaya hidup yang buruk, seperti merokok, konsumsi alkohol berlebihan, serta tekanan sosial, seringkali menjadi faktor risiko terjadinya kanker paru-paru. Oleh karena itu, diagnosis kanker paru-paru yang cepat dan akurat sangat penting untuk menangani pasien secara tepat waktu. Penelitian ini bertujuan untuk menemukan model yang akurat dan tepat dalam klasifikasi kanker paru-paru berdasarkan gejala dan gaya hidup pasien. Metode penelitian yang digunakan dalam penelitian ini menggunakan model Machine Learning yaitu Ensemble Learning. Model yang digunakan yaitu VotingClassifier, menggabungkan empat algoritma klasifikasi, antara lain SVM, Decision Tree, LightGBM, dan XGBoost. Hasil akurasi yang didapat dalam penelitian ini menggunakan model yang diusulkan mencapai akurasi 97.78%. berdasarkan hasil akurasi yang didapatkan, model usulan menunjukkan hasil akurasi yang lebih tinggi dan lebih optimal dari penelitian sebelumnya.

**Kata Kunci:** Kanker Paru-paru, klasifikasi, SVM, LightGBM, XGBoost, Ensemble Learning

## Abstract

*Lung Cancer is one type of disease that causes the death of most humans in the world. Lung cancer occurs due to abnormal cell growth in the lung organ. Abnormal cells will grow into cancer cells and will attack the organs that are overgrown with cancer cells. Symptoms that are often experienced by people with lung cancer include shortness of breath, chest pain, coughing, wheezing and yellowing of the fingers. Lung cancer can also occur due to poor lifestyles, such as smoking, excessive alcohol consumption, and social pressure, are often risk factors for lung cancer. Therefore, a quick and accurate diagnosis of lung cancer is essential to treat patients in a timely manner. This study aims to find an accurate and precise model in the classification of lung cancer based on the symptoms and lifestyle of patients. The research method used in this study uses the Machine Learning model, namely Ensemble Learning. The model used is votingclassifier, combining four classification algorithms, including SVM, Decision Tree, LightGBM, and XGBoos. The accuracy results obtained in this study using the proposed model reached 97.78% accuracy. based on the accuracy results obtained, the proposed model shows higher and more optimal accuracy results than previous research.*

**Keyword:** Lung Cancer, Classification, Ensemble Learning, SVM, LGBM, XGBoost

## 1. PENDAHULUAN

Kanker merupakan salah satu jenis penyakit yang mematikan yang disebabkan oleh pertumbuhan sel yang tidak normal pada bagian tubuh manusia. Sel tubuh manusia akan berganti setiap waktu dengan membelah diri untuk menghasilkan sel baru menggantikan sel yang sudah tua. Penyakit kanker memiliki banyak jenis sesuai area ketidaknormalan pertumbuhan sel [1], [2]. Jenis jenis sel antara lain, kanker paru-paru, kanker payudara [3], kanker pankreas [4], dan lain sebagainya. Salah satu jenis penyakit kanker yang menyebabkan banyak kematian di dunia yaitu kanker paru-paru. Kanker ini menyerang sistem pernapasan manusia sehingga penderita akan mengalami kesulitan dalam bernapas [5]. Gejala yang ditimbulkan saat mengidap kanker paru-paru antara lain, seperti mengi, sesak napas, dan batuk [6]–[8]. Ada kebiasaan yang bisa meningkatkan risiko untuk terkena kanker paru-paru seperti mengonsumsi alkohol dan merokok [9], [10]. Oleh karena itu, dalam diagnosis pengidap kanker paru-paru berdasarkan gejala [11] yang muncul perlu dilakukan agar bisa mencegah pertumbuhan kanker menjadi lebih ganas dan terlambat untuk diobati sesuai dengan penanganannya [12], [13].

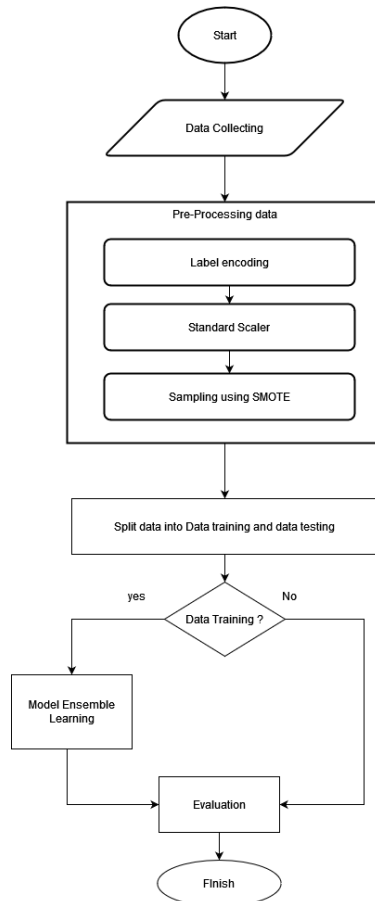
Penelitian tentang klasifikasi kanker paru-paru telah dilakukan oleh beberapa peneliti, antara lain penelitian oleh Haznedar & Simsek [14] pada tahun 2022. Penelitian tersebut menggunakan metode penelitian *Deep Neural Network* dengan optimasi *adadelta*. Penelitian ini menunjukkan keunggulan dari *deep learning* untuk melakukan identifikasi dan klasifikasi data kanker paru-paru dengan tingkat akurasi yang tinggi. DNN digunakan untuk menangani hubungan non-linier yang kompleks karena DNN merupakan bentuk versi lebih lanjut dari ANN yang memiliki tambahan lapisan pada *hidden layer*. Dalam penelitian tersebut menggunakan fungsi aktivasi *softmax* dan *cross entropy* serta dioptimalkan dengan algoritma *Adadelta* yang telah dikembangkan sebagai perpanjangan dari algoritma *adagard*. Hasil dari penelitian ini adalah akurasi tertinggi mencapai 96,15%.

Namun model ini perlu untuk ditingkat kembali ke tingkat yang lebih tinggi demi menaikkan kemungkinan penanganan yang lebih akurat. Oleh karena itu, dalam penelitian ini mengusulkan metode untuk melakukan klasifikasi dengan menggunakan model penelitian *machine learning* yaitu *ensemble learning* dengan *voting classifier* [15]. Model ini akan menggabungkan model penelitian untuk dapat menghasilkan akurasi yang lebih akurat [16]. Model yang akan digabungkan dalam ensemble learning antara lain, *Support Vector Machine (SVM)* [17], *Decision Tree* [18], *Light Gradient Boosting Machine (LightGBM)* [19] dan *Extreme Gradient Boosting* [20]. model di atas memiliki kemampuan untuk klasifikasi dengan baik. Untuk meningkatkan pengolahan datanya dilengkapi dengan oversampling untuk menyeimbangkan data menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)* [21], [22].

## 2. METODE

Dalam penelitian ini dilakukan pendekatan dalam melakukan klasifikasi kanker paru-paru menggunakan metode algoritma *ensemble learning* dengan model *voting classifier*. Menggabungkan model algoritma *support vector machine*, *decision tree*,

*light gradient boosting machine*, dan *extreme gradient boosting*. tahapan dalam penelitian ini dapat dilihat dalam *flowchart* pada gambar 1.



Gambar 1. *Flowchart* Model Usulan

### 2.1. Data Collecting

Data yang akan digunakan dalam penelitian ini, merupakan data yang berasal dari dataset yang disediakan oleh platform data publik kaggle. Data set tersebut dapat diunduh dengan mengunjungi tautan : <https://www.kaggle.com/datasets/nancyalaswad90/lung-cancer>

Dataset tersebut berisi data tentang kondisi pasien yang didiagnosa mengidap kanker paru-paru. Berisi data sebanyak 309 data text dengan 16 kolom, antara lain kolom *Age*, *Gender*, *Smoking*, *Yellow Fingers*, *Anxiety*, *Peer Pressure*, *Chronic Disease*, *Fatigue*, *Allergy*, *Wheezing*, *Alcohol Consuming*, *Coughing*, *Shortness Breath*, *Swallowing Dificulty*, *Chest Pain*, dan *Lung Cancer*. Detail kolom data dilihat pada tabel 1.

Tabel 1. Kolom Data

| <i>Column</i>                | Total data | Tipe data |
|------------------------------|------------|-----------|
| <i>Age</i>                   | 309        | Integer   |
| <i>Gender</i>                | 309        | Object    |
| <i>Smoking</i>               | 309        | Integer   |
| <i>Yellow Fingers</i>        | 309        | Integer   |
| <i>Anxiety</i>               | 309        | Integer   |
| <i>Peer Pressure</i>         | 309        | Integer   |
| <i>Chronic Disease</i>       | 309        | Integer   |
| <i>Fatigue</i>               | 309        | Integer   |
| <i>Allergy</i>               | 309        | Integer   |
| <i>Wheezing</i>              | 309        | Integer   |
| <i>Alcohol Consuming</i>     | 309        | Integer   |
| <i>Coughing</i>              | 309        | Integer   |
| <i>Shortness Breath</i>      | 309        | Integer   |
| <i>Swallowing Difficulty</i> | 309        | Integer   |
| <i>Chest Pain</i>            | 309        | Integer   |
| <i>Lung Cancer</i>           | 309        | Object    |

## 2.2. Pre-processing

Pada tahapan pre-processing ini data diolah dan dibersihkan. Diawali dengan pengecekan data, karena ada kemungkinan isi data yang *missing* dan *NULL* jadi harus dicek. Jika ada data yang *missing* dan *NULL* maka perlu untuk dibersihkan. Jika dirasa data sudah bersih, tahap selanjutnya masuk ke *label encoding* data. Mengubah data menjadi bentuk biner dengan nilai 0 dan 1. Melakukan *label encoding* akan mempermudah dalam pengolahan data selanjutnya karena kebanyakan model *machine learning* akan mengolah data dengan bentuk numerik atau biner. Semua kolom diubah menjadi bentuk biner kecuali kolom *Age*. Karena akan sudah untuk mengubah kolom yang berisi variasi nilai data.

Tahap selanjutnya melakukan standarisasi fitur untuk kolom *age*. Standarisasi diperlukan karena banyak algoritma lebih efektif ketika fitur memiliki skala yang sama, terutama ketika data terdiri dari berbagai fitur yang memiliki rentang nilai yang sangat berbeda. Standarisasi fitur menggunakan *standard scaler* dengan mengubah data sehingga memiliki nilai rata-rata (mean) sama dengan 0 dan standar deviasi sama dengan 1. Nilai *mean* dan standar deviasi digunakan untuk menghitung *Z-score*. Rumus yang digunakan sebagai berikut :

$$Z = \frac{(X - \mu)}{\sigma} \quad \#(1)$$

Penjelasan :

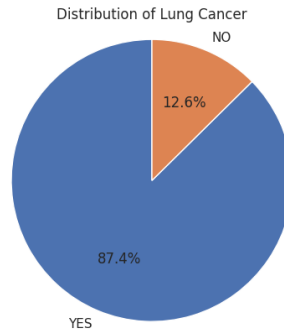
X = nilai asli

$\mu$  = rata-rata dari data

$\sigma$  = standar deviasi

Setelah melakukan standarisasi masuk ke tahap selanjutnya yaitu sampling dengan menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) [23]. Proses dimulai dengan memisahkan data menjadi fitur variabel terikat (y) dan variabel bebas

(X). Kemudian, metode SMOTE di inisialisasi untuk mengatasi masalah ketidakseimbangan kelas dengan menambahkan sampel sintetis pada kelas minoritas [24]. SMOTE bekerja dengan membuat sampel baru berdasarkan jarak antar sampel dalam fitur yang ada untuk menyeimbangkan jumlah sampel antara kelas mayoritas dan minoritas [25]. Setelah proses ini selesai, hasilnya adalah data yang telah di sampel ulang, di mana jumlah sampel pada setiap kelas menjadi seimbang, sehingga model pembelajaran mesin dapat dilatih tanpa bias terhadap kelas mayoritas. Distribusi data pada kolom Lung Cancer dapat dilihat pada gambar 2.

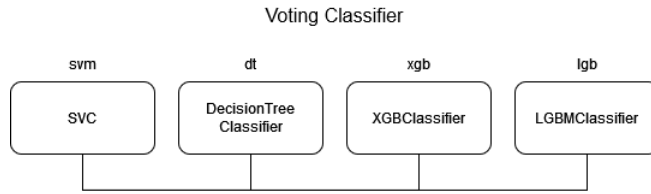


Gambar 2. Distribusi Data kolom Lung Cancer

### 2.3. Building Model

Tahapan selanjutnya adalah membuat model klasifikasi *ensemble learning* dengan model voting classifier. Pada tahap ini perlu diidentifikasi dulu algoritma yang akan digabungkan ke dalam voting *classifier*. Ada empat algoritma yang akan digunakan, antara lain *Support Vector Machine*, *Decision Tree*, *Light Gradient Boosting Machine*, dan *Extreme Gradient Boosting*. Tiap model yang digunakan memiliki parameternya tersendiri untuk melatih model. Parameter yang digunakan oleh SVM antara lain, menggunakan *kernel rbf*, menggunakan probabilitas dengan nilai *true* untuk mengontrol probabilitas, parameter  $c = 1.0$  berguna untuk mengontrol margin keras dan lunak. *Random\_state = 42* berguna untuk menjaga konsistensi ketika model dilatih. Parameter yang digunakan oleh *decision tree* adalah *random\_state* untuk menjaga konsistensi ketika dilatih dan maksimal cabang sebanyak 5. Parameter yang digunakan pada *LightGBM* dan *XGBoost* adalah *random\_state = 42* menjaga konsistensi hasil.

Kemudian setelah model di atas dilatih, model tersebut akan menjadi *estimator* dari model *ensemble learning*. Lalu dengan parameter voting yang *soft* model akan dilatih untuk mengklasifikasikan kanker paru-paru. Untuk *summary* model dari *ensemble learning*-nya ditunjukkan pada gambar 3.



Gambar 3. Summary model Voting Classifier

## 2.4. Evaluation

Pada tahap ini model akan dievaluasi, model untuk evaluasinya adalah *confusion matrix*. Model evaluasi menggunakan *confusion matrix*, dengan komposisi sebagai berikut, *F1-score*, *accuracy*, *recall*, dan *precision*. Fokus dari analisis kinerja ini pada akurasi yang dihasilkan oleh model usulan dibandingkan penelitian sebelumnya. Dalam melakukan analisis hasil menggunakan *confusion matrix*, terdapat rumus matematikanya sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \%$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1\ Score = 2 \times \frac{precision \times recall}{precision+recall}$$

Keterangan :

- TN (*True Negatif*) : Data benar yang dikategorikan negatif
- TP (*True Positif*) : Data benar yang dikategorikan Positif.
- FP (*False Positif*) : Data salah yang dikategorikan positif
- FN (*False Negatif*) : Data salah yang dikategorikan negatif

## 3. HASIL DAN PEMBAHASAN

Pada subbab ini akan dibahas mengenai hasil dari penelitian yang telah dilakukan mulai dari pemrosesan dan pengolahan data hingga hasil analisis dan evaluasi model penelitian. Kemudian akan dibahas perbandingan dengan model penelitian yang sudah ada.

### 3.1. Hasil

Hasil yang akan dibahas dalam penelitian ini dimulai dari label *encoding*, normalisasi data, *oversampling* data, hasil evaluasi model berupa *confusion matrix* dan *classification report*.

#### a. Hasil Label *Encoding*

Hasil dari label *encoding* pada penelitian ini dapat dilihat pada tabel 2.

Tabel 2. Hasil Label *Encoding*

| Sebelum <i>label encoding</i> | Setelah <i>label encoding</i> |
|-------------------------------|-------------------------------|
|-------------------------------|-------------------------------|

| <i>Gender</i> | <i>Lung Cancer</i> | <i>Gender</i> | <i>Lung Cancer</i> |
|---------------|--------------------|---------------|--------------------|
| <i>M</i>      | <i>YES</i>         | 1             | 1                  |
| <i>F</i>      | <i>NO</i>          | 0             | 0                  |
| <i>M</i>      | <i>NO</i>          | 1             | 0                  |
| <i>F</i>      | <i>YES</i>         | 0             | 1                  |

Pada hasil yang didapatkan *Gender* M diubah menjadi nilai numerik 1 dan *gender* F menjadi nilai 0. Kemudian pada *Lung Cancer* *Yes* diubah menjadi 1 dan *No* menjadi nilai 0.

b. Hasil Normalisasi Data

Hasil normalisasi data pada penelitian ini dapat dilihat pada tabel 3.

Tabel 3. Hasil Normalisasi Data

| Sebelum <i>normalisasi</i> | Sesudah <i>normalisasi</i> |
|----------------------------|----------------------------|
| Age                        | Age                        |
| 69                         | 0.771850                   |
| 74                         | 1.381829                   |
| 59                         | -0.448107                  |

Pada proses normalisasi menggunakan *library scikit-learn preprocessing* dengan *standardscaler*. Rumus yang digunakan merupakan *z-score*, sehingga hasil yang didapat seperti pada tabel 3. Hasil tersebut sangat memudahkan model penelitian dalam melakukan klasifikasi.

c. Hasil *Oversampling* Data

*Oversampling* digunakan untuk mengatasi tidak ketidakseimbangan data. Hasil yang didapatkan setelah *oversampling* adalah persebaran nilai pada fitur secara merata. Hasilnya dapat dilihat pada tabel 4.

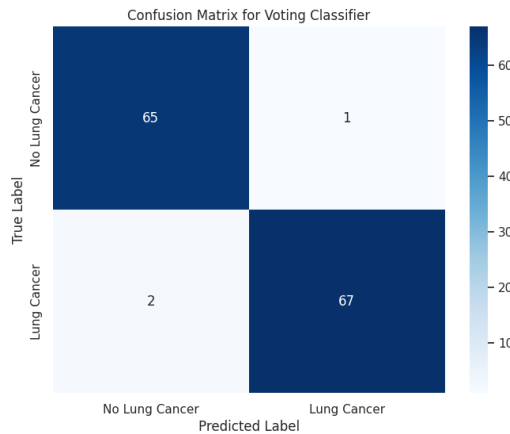
Tabel 4. Hasil *Oversampling*

| Sebelum <i>Oversampling SMOTE</i> | Setelah <i>Oversampling SMOTE</i> |
|-----------------------------------|-----------------------------------|
| 0                                 | 39                                |
| 1                                 | 270                               |

Dari hasil di atas menunjukkan bahwa *oversampling* menyeimbangkan data pada data *y* menjadi *y\_resampled*. Pada *y\_resampled* menunjukkan nilai 0 dan 1 memiliki data yang merata sehingga pada klasifikasi akan lebih baik.

d. *Confusion Matrix*

Hasil dari uji coba model yang telah dilatih akan digambarkan pada *confusion matrix*. Hasilnya dapat dilihat pada gambar 3.



Gambar 4. *Confusion Matrix*

Pada *confusion matrix* ini hasil klasifikasi dari kanker paru-paru diprediksikan. Terdapat 2 data kanker paru-paru yang diprediksi sebagai bukan kanker paru-paru. Kemudian terdapat 1 data yang bukan kanker paru-paru diprediksikan sebagai kanker paru-paru.

e. *Classification Report*

Hasil evaluasi kinerja dari model penelitian dalam melakukan klasifikasi kanker paru-paru dapat dilihat pada tabel 5.

Tabel 5. *Classification Report*

|              | Precision | recall | F1-score | Support |
|--------------|-----------|--------|----------|---------|
| 0            | 0,97      | 0,98   | 0,98     | 66      |
| 1            | 0,99      | 0,97   | 0,98     | 69      |
| Accuracy     |           |        | 0,98     | 135     |
| Macro avg    | 0,98      | 0,98   | 0,98     | 135     |
| Weighted avg | 0,98      | 0,98   | 0,98     | 135     |

Berdasarkan *confusion matrix* di atas didapatkan akurasi dari model penelitian ini mendapatkan 0,98 atau 0,9775 (97,75%). Ini membuktikan bahwa model ini bekerja dengan baik pada klasifikasi kanker paru-paru dengan menghasilkan prediksi yang akurat dan optimal.

### 3.2. Pembahasan

Pada pembahasan ini akan dibandingkan model penelitian yang diusulkan dengan penelitian yang telah ada sebelumnya. Hasil perbandingan dapat dilihat pada tabel 6.

Tabel 6. Perbandingan Model Sebelumnya

| Author                    | Tahun | Model Penelitian    | Hasil               |
|---------------------------|-------|---------------------|---------------------|
| Haznedar & Simsek<br>[14] | 2022  | <i>DNN-Adadelta</i> | Akurasi :<br>96,15% |

|                           |      |                                                                        |                  |
|---------------------------|------|------------------------------------------------------------------------|------------------|
| Shanchen Pang et al. [26] | 2020 | <i>DNN, DenseNet dan AdaBoos (DenseNet, ResNet, VGG16 dan AlexNet)</i> | Akurasi : 89,85% |
| Firdaus et al. [13]       | 2020 | <i>SVM with Gray Level Co-Occurrence Matrix(GLCM)</i>                  | Akurasi L 83,33% |
| Proposed Model            | 2024 | <i>Ensemble Learning ( SVM +DT +LightGBM +XGBoost)</i>                 | Akurasi : 97,75% |

Berdasarkan hasil perbandingan dengan penelitian yang telah dilakukan sebelumnya, hasil yang didapatkan oleh model usulan lebih tinggi dan lebih akurat. Hal ini menunjukkan kinerja model usulan lebih baik.

#### 4. SIMPULAN

Klasifikasi kanker paru-paru menggunakan algoritma *ensemble learning* dengan model voting *classifier* menghasilkan akurasi klasifikasi yang tinggi dan akurat jika dibandingkan dengan model penelitian sebelumnya. Menggabungkan model *SVM*, *DT*, *LightGBM*, dan *XGBoost* memberikan hasil yang baik dan tidak *overfitting*. Oleh karena itu, model ini merupakan hasil pengembangan model klasifikasi kanker paru-paru yang lebih akurat dan tepat.

#### 5. REFERENSI

- [1] L. Sandhya, M. S. Khan, dan R. Chinnaiyan, "Lung Cancer Prediction using Deep Learning Techniques," *2023 Int. Conf. New Front. Commun. Autom. Manag. Secur. ICCAMS 2023*, no. December, 2023, doi: 10.1109/ICCAMS60113.2023.10526146.
- [2] N. Banerjee dan S. Das, "Prediction Lung Cancer- in Machine Learning Perspective," *2020 Int. Conf. Comput. Sci. Eng. Appl. ICCSEA 2020*, 2020, doi: 10.1109/ICCSEA49143.2020.9132913.
- [3] A. U. Dullah, F. N. Apsari, dan J. Jumanto, "Ensemble learning technique to improve breast cancer classification model," *J. Soft Comput. Explor.*, vol. 4, no. 2, hal. 107–113, 2023, doi: 10.52465/joscex.v4i2.166.
- [4] S. Zanini, S. Renzi, A. R. Limongi, P. Bellavite, F. Giovinazzo, dan G. Bermanno, "A review of lifestyle and environment risk factors for pancreatic cancer," *Eur. J. Cancer*, vol. 145, hal. 53–70, 2021, doi: 10.1016/j.ejca.2020.11.040.
- [5] M. Singh, C. Shah, dan P. Patel, "Lung Cancer Prediction Using Machine Learning Models," *Lect. Notes Networks Syst.*, vol. 765 LNNS, no. Icmia, hal. 613–618, 2023, doi: 10.1007/978-981-99-5652-4\_54.
- [6] H. Yu, Z. Zhou, dan Q. Wang, "Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model," *IEEE Access*, vol. 8, hal. 86400–86410, 2020, doi: 10.1109/ACCESS.2020.2992645.
- [7] R. Raza et al., "Lung-EffNet: Lung cancer classification using EfficientNet from

- CT-scan images,” *Eng. Appl. Artif. Intell.*, vol. 126, no. PB, hal. 106902, 2023, doi: 10.1016/j.engappai.2023.106902.
- [8] M. Ragab, I. Katib, S. A. Sharaf, F. Y. Assiri, D. Hamed, dan A. A. M. Al-Ghamdi, “Self-Upgraded Cat Mouse Optimizer with Machine Learning Driven Lung Cancer Classification on Computed Tomography Imaging,” *IEEE Access*, vol. 11, no. July, hal. 107972–107981, 2023, doi: 10.1109/ACCESS.2023.3313508.
- [9] B. Senthilkumar *et al.*, “Ensemble Modelling for Early Breast Cancer Prediction from Diet and Lifestyle,” *IFAC-PapersOnLine*, vol. 55, no. 1, hal. 429–435, 2022, doi: 10.1016/j.ifacol.2022.04.071.
- [10] C. Thallam, A. Peruboyina, S. S. T. Raju, dan N. Sampath, “Early Stage Lung Cancer Prediction Using Various Machine Learning Techniques,” *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, hal. 1285–1292, 2020, doi: 10.1109/ICECA49313.2020.9297576.
- [11] T. I. A. Mohamed dan A. E. S. Ezugwu, “Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data,” *IEEE Access*, vol. 12, no. April, hal. 59880–59892, 2024, doi: 10.1109/ACCESS.2024.3394030.
- [12] M. Mamun, A. Farjana, M. Al Mamun, dan M. S. Ahammed, “Lung cancer prediction model using ensemble learning techniques and a systematic review analysis,” *2022 IEEE World AI IoT Congr. AIIoT 2022*, hal. 187–193, 2022, doi: 10.1109/AIIoT54504.2022.9817326.
- [13] Q. Firdaus, R. Sigit, T. Harsono, dan A. Anwar, “Lung cancer detection based on ct-scan images with detection features using gray level co-occurrence matrix (glcm) and support vector machine (svm) methods,” *IES 2020 - Int. Electron. Symp. Role Auton. Intell. Syst. Hum. Life Conf.*, hal. 643–648, 2020, doi: 10.1109/IES50839.2020.9231663.
- [14] B. Haznedar dan N. Y. Simsek, “A Comparative Study on Classification Methods for Renal Cell and Lung Cancers Using RNA-Seq Data,” *IEEE Access*, vol. 10, no. October, hal. 105412–105420, 2022, doi: 10.1109/ACCESS.2022.3211505.
- [15] S. M. Ganie dan M. B. Malik, “An ensemble Machine Learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators,” *Healthc. Anal.*, vol. 2, no. August, hal. 100092, 2022, doi: 10.1016/j.health.2022.100092.
- [16] M. Rahil, B. N. Anoop, G. N. Girish, A. R. Kothari, S. G. Koolagudi, dan J. Rajan, “A Deep Ensemble Learning-Based CNN Architecture for Multiclass Retinal Fluid Segmentation in OCT Images,” *IEEE Access*, vol. 11, no. December 2022, hal. 17241–17251, 2023, doi: 10.1109/ACCESS.2023.3244922.
- [17] N. S. Haider dan A. K. Behera, “Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD),” *Biocybern. Biomed. Eng.*, vol. 42, no. 1, hal. 42–59, 2022, doi: 10.1016/j.bbe.2021.12.004.
- [18] S. M. Abd-Elsalam, M. M. Ezz, S. Gamalel-Din, G. Esmat, A. Salama, dan M. ElHefnawi, “Early diagnosis of esophageal varices using Boosted-Naïve Bayes Tree: A multicenter cross-sectional study on chronic hepatitis C patients,” *Informatics Med. Unlocked*, vol. 20, hal. 100421, 2020, doi: 10.1016/j.imu.2020.100421.
- [19] J. Yan *et al.*, “LightGBM: accelerated genomically designed crop breeding through ensemble learning,” *Genome Biol.*, vol. 22, no. 1, hal. 1–24, 2021, doi: 10.1186/s13059-021-02492-y.
- [20] H. Chen, N. Wang, X. Du, K. Mei, Y. Zhou, dan G. Cai, “Classification

- Prediction of Breast Cancer Based on Machine Learning,” *Comput. Intell. Neurosci.*, vol. 2023, hal. 1–9, 2023, doi: 10.1155/2023/6530719.
- [21] M. A. Muslim *et al.*, “New model combination meta-learner to improve accuracy prediction P2P lending with stacking ensemble learning,” *Intell. Syst. with Appl.*, vol. 18, no. December 2022, hal. 200204, 2023, doi: 10.1016/j.iswa.2023.200204.
- [22] R. Obiedat *et al.*, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” *IEEE Access*, vol. 10, hal. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
- [23] H. F. El-Sofany, “Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques,” *IEEE Access*, vol. 12, no. August, hal. 106146–106160, 2024, doi: 10.1109/ACCESS.2024.3437181.
- [24] A. S. Almajid, “Multilayer Perceptron Optimization on Imbalanced Data Using SVM-SMOTE and One-Hot Encoding for Credit Card Default Prediction,” *J. Adv. Inf. Syst. Technol.*, vol. 3, no. 2, hal. 67–74, 2022, doi: 10.15294/jaist.v3i2.57061.
- [25] R. Nithya, T. Kokilavani, dan T. L. A. Beena, “Balancing cerebrovascular disease data with integrated ensemble learning and SVM-SMOTE,” *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 13, no. 1, 2024, doi: 10.1007/s13721-024-00447-4.
- [26] S. Pang, Y. Zhang, M. Ding, X. Wang, dan X. Xie, “A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting,” *IEEE Access*, vol. 8, hal. 4799–4805, 2020, doi: 10.1109/ACCESS.2019.2962862.