


Sistem Pakar Pendeteksi Penyakit Pernapasan Menggunakan *Gradient Boosting* dan Metode CNN

DOI: <http://dx.doi.org/10.35889/jutisi.v15i2.3601>

Creative Commons License 4.0 (CC BY – NC) 

Eva Gultom^{1*}, Tony Darmanto², Jimmy Tjen³

^{1,2} Sistem Informasi, Universitas Widya Dharma Pontianak, Pontianak, Indonesia

³ Teknik Informatika, Universitas Widya Dharma Pontianak, Pontianak, Indonesia

*e-mail *Corresponding Author*: Salfarizi70t@gmail.com

Abstract

Respiratory diseases are among the most common illnesses in the community and are often underestimated. Low public awareness of respiratory diseases has led to a spike in mortality rates due to chronic respiratory diseases and slow treatment. One of the obstacles currently faced is that the manual diagnosis system takes a long time and requires limited specialist expertise. This study provides an expert system that can be used to detect respiratory tract diseases with two different types of input data. The Gradient Boosting algorithm is applied to improve diagnostic accuracy based on clinical data, while the CNN method is used to identify diseases using automatic features by extracting chest X-ray images. This study uses a dataset from Kaggle, which produces a data accuracy rate of 99.7% using Gradient Boosting and 95.93% using the CNN method. The accuracy results from each method show that this system can provide accurate respiratory disease detection results.

Keywords: CNN; Gradient Boosting; Respiratory Disease; Expert System

Abstrak

Penyakit pernapasan merupakan salah satu penyakit yang sering ditemui di kalangan masyarakat dan sering diremehkan. Rendahnya tingkat kesadaran masyarakat terhadap penyakit pernapasan menyebabkan melonjaknya tingkat kematian, dikarenakan penyakit pernapasan kronis dan penanganan yang lambat. Kendala yang dihadapi saat ini salah satunya yaitu sistem diagnosis manual yang digunakan memerlukan waktu yang lama serta keahlian spesialis yang terbatas. Penelitian ini menyediakan sistem pakar yang dapat digunakan untuk mendeteksi penyakit saluran pernapasan dengan dua jenis data input yang berbeda. Algoritma *Gradient Boosting* diterapkan untuk meningkatkan akurasi diagnostik berdasarkan data klinis, sedangkan metode *Convolutional Neural Networks (CNN)* digunakan untuk mengidentifikasi penyakit menggunakan fitur otomatis dengan mengekstrak citra rontgen dada. Penelitian ini menggunakan dataset dari Kaggle, yang menghasilkan tingkat akurasi data sebesar 99,7% menggunakan *Gradient Boosting* dan 95,93% menggunakan metode *CNN*. Tingkat akurasi dari masing-masing metode menunjukkan bahwa sistem ini dapat memberikan hasil deteksi penyakit pernapasan yang akurat.

Kata kunci: CNN; Gradient Boosting; Penyakit Pernapasan; Sistem Pakar

1. Pendahuluan

Penyakit pernapasan sering kali diremehkan meskipun dampaknya secara global sangat signifikan. Beberapa kondisi pernapasan dengan tingkat mortalitas yang tinggi meliputi Penyakit Paru Obstruktif Kronis (PPOK) yang merenggut sekitar 3,2 juta jiwa setiap tahun, dengan 81,7% dari kematian ini disebabkan oleh masalah pernapasan kronis [1]. Selain itu, pneumonia menjadi penyebab utama kematian pada kelompok usia rentan, sementara kanker paru-paru memiliki tingkat kelangsungan hidup yang relatif rendah. Munculnya COVID-19 juga semakin memperparah kondisi ini dengan jumlah kematian yang tinggi dalam waktu singkat [1], [2]. Tingginya angka kejadian dan kematian akibat penyakit pernapasan menunjukkan bahwa

permasalahan ini merupakan tantangan serius dalam dunia kesehatan yang memerlukan perhatian dan penanganan yang lebih efektif [3].

Diagnosis penyakit pernapasan dalam praktik medis masih banyak dilakukan secara manual melalui analisis gejala klinis, tanda vital, serta interpretasi citra rontgen dada. Proses ini membutuhkan waktu yang relatif lama dan sangat bergantung pada keahlian tenaga medis, khususnya dokter spesialis, yang ketersediaannya masih terbatas. Selain itu, proses diagnosis juga berpotensi mengalami kendala subjektivitas serta kesalahan interpretasi, terutama ketika data pasien memiliki karakteristik yang kompleks dan beragam. Di sisi lain, data yang digunakan dalam diagnosis penyakit pernapasan tidak hanya berupa data klinis, tetapi juga mencakup data citra medis yang memerlukan pendekatan analisis yang berbeda. Kondisi ini menunjukkan adanya kebutuhan akan sistem bantu diagnosis yang mampu mengintegrasikan berbagai jenis data secara cepat, objektif, dan akurat [3].

Berbagai penelitian telah dilakukan untuk mengembangkan sistem deteksi penyakit dengan memanfaatkan pendekatan sistem pakar, metode *machine learning* dan *deep learning*. Pendekatan sistem pakar tradisional banyak menggunakan metode berbasis aturan seperti Forward Chaining dan Certainty Factor untuk mendiagnosis penyakit berdasarkan gejala yang diberikan oleh pengguna, serta metode Dempster-Shafer yang mampu menangani ketidakpastian dengan menggabungkan berbagai evidensi gejala dalam proses diagnosis [4], [5], [6]. Pendekatan berbasis *Convolutional Neural Networks* (CNN) banyak digunakan dalam analisis citra medis, termasuk klasifikasi pneumonia pada citra rontgen dada, yang menunjukkan kemampuan dalam mengekstraksi fitur visual secara efektif [7], [8], [9]. Di sisi lain, algoritma *Gradient Boosting* juga banyak digunakan dalam klasifikasi data terstruktur dan menunjukkan performa yang baik dalam menghasilkan prediksi yang akurat [10], [11]. Sejumlah penelitian turut membandingkan kinerja model untuk memperkuat hasil klasifikasi. Penelitian [12] menunjukkan bahwa *Gradient Boosting* memiliki performa yang lebih baik dibandingkan Random Forest dalam klasifikasi penyakit dengan akurasi masing-masing sebesar 81% dan 79%. Selain itu, pendekatan berbasis CNN seperti CNN-ELM juga menunjukkan peningkatan performa dalam klasifikasi pneumonia dengan perbedaan akurasi sebesar 8,81% dan peningkatan skor F1 sebesar 0,0729 [7]. Penelitian lain [13] menggabungkan CNN dengan metode *Gradient Boosting* seperti XGBoost, LightGBM, dan CatBoost untuk klasifikasi data kompleks, dengan hasil bahwa metode berbasis CNN-*Gradient Boosting* mampu menghasilkan akurasi yang lebih tinggi dibandingkan dengan metode lainnya. Metode *Gradient Boosting* dan CNN dimanfaatkan secara adaptif sesuai dengan jenis data yang tersedia untuk mengatasi keterbatasan masing-masing pendekatan. Studi ini secara spesifik mengatasi masalah ketidakseimbangan data yang sering ditemukan pada dataset medis [14], [15] dengan menggunakan data yang telah diimbangi sebelumnya (*pre-balanced*), untuk mengurangi bias dalam klasifikasi penyakit dan meningkatkan perbedaan antara pola visual yang serupa. Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada satu jenis data atau belum mengintegrasikan data klinis dan citra medis secara terpadu dalam satu sistem, sehingga masih terdapat gap dalam pengembangan model diagnosis yang lebih komprehensif.

Berdasarkan celah penelitian sebelumnya, studi ini mengusulkan sistem pakar pendeteksi penyakit pernapasan berbasis pendekatan multimodal dengan memanfaatkan metode *Gradient Boosting* dan CNN secara adaptif sesuai dengan jenis data yang tersedia. *Gradient Boosting* digunakan untuk mengolah data klinis yang terstruktur, seperti gejala dan tanda vital, sedangkan CNN dimanfaatkan untuk menganalisis citra rontgen dada. Pendekatan ini dirancang agar sistem dapat tetap memberikan hasil diagnosis meskipun pengguna hanya memiliki salah satu jenis data, sehingga meningkatkan fleksibilitas dan kegunaan sistem dalam berbagai kondisi. Dengan demikian, sistem tidak bergantung pada satu sumber data tertentu, melainkan mampu menyesuaikan metode yang digunakan berdasarkan input yang tersedia.

2. Metodologi

2.1 Model/Algoritma *Gradient Boosting* dan Metode CNN

Penelitian ini memanfaatkan dua pendekatan berbasis *machine learning* dan *deep learning*, yaitu *Gradient Boosting* dan CNN, untuk mendeteksi penyakit pernapasan berdasarkan jenis data yang tersedia. Pemilihan kedua metode ini didasarkan pada kemampuannya dalam mengolah jenis data yang berbeda, yaitu data klinis terstruktur dan data citra medis. Dalam penelitian ini, *Gradient Boosting* digunakan untuk mengolah data klinis yang bersifat terstruktur, seperti gejala dan tanda vital pasien, karena algoritma ini memiliki

kemampuan yang baik dalam menangani data tabular dan menghasilkan performa klasifikasi yang tinggi. *Gradient Boosting* merupakan algoritma ensemble yang membangun model secara iteratif dengan meminimalkan fungsi loss melalui penambahan weak learner pada setiap tahap. Secara matematis, model pada iterasi ke- m dinyatakan sebagai:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (1)$$

$F_m(x)$ merupakan model pada iterasi ke- m , $h_m(x)$ adalah weak learner yang biasanya berupa decision tree, dan γ_m merupakan learning rate yang mengontrol kontribusi setiap model terhadap model akhir. Proses ini memungkinkan model untuk secara bertahap memperbaiki kesalahan prediksi pada iterasi sebelumnya sehingga meningkatkan akurasi klasifikasi.

Sementara itu, CNN merupakan metode *deep learning* yang dirancang untuk memproses data berbasis citra. CNN bekerja melalui serangkaian proses konvolusi, aktivasi, dan pooling untuk mengekstraksi fitur penting dari citra secara otomatis. Fitur yang dihasilkan kemudian digunakan dalam proses klasifikasi untuk mengidentifikasi pola penyakit pada citra rontgen dada. Dalam penelitian ini, CNN digunakan untuk mengklasifikasikan citra rontgen ke dalam beberapa kategori penyakit pernapasan, seperti pneumonia, tuberkulosis, COVID-19, dan kondisi normal. Operasi utama dalam CNN adalah konvolusi, yang secara matematis dapat dirumuskan sebagai berikut:

$$S(i, j) = \sum_m \sum_n X(i + m, j + n) K(m, n) \quad (2)$$

X merupakan citra input, K adalah kernel atau filter yang digunakan untuk mendeteksi pola tertentu, dan $S(i, j)$ adalah hasil feature map. Proses konvolusi ini memungkinkan sistem untuk mengenali pola visual seperti infiltrasi paru dan opasitas yang berkaitan dengan penyakit pernapasan.

Kedua metode tersebut tidak digunakan secara bersamaan dalam satu proses klasifikasi, melainkan diterapkan secara adaptif berdasarkan jenis data yang dimasukkan oleh pengguna. Jika data yang tersedia berupa data klinis, maka sistem akan menggunakan model *Gradient Boosting* untuk menghasilkan diagnosis. Sebaliknya, jika data berupa citra rontgen dada, maka sistem akan menggunakan model CNN. Pendekatan ini memungkinkan sistem untuk tetap memberikan hasil deteksi meskipun hanya satu jenis data yang tersedia, sehingga meningkatkan fleksibilitas dan kegunaan sistem dalam berbagai kondisi.

2.2 Variabel yang Dianalisis

Variabel yang digunakan dalam penelitian ini dirancang untuk merepresentasikan kondisi klinis dan visual pasien dalam proses diagnosis penyakit pernapasan. Secara umum, variabel yang dianalisis terdiri dari dua kelompok utama, yaitu variabel data klinis (non-visual) dan variabel data citra medis (visual), yang masing-masing diproses menggunakan model yang berbeda sesuai dengan karakteristik datanya. Variabel data klinis merupakan data terstruktur yang menggambarkan kondisi fisiologis dan gejala yang dialami oleh pasien. Data ini meliputi beberapa parameter medis penting, seperti suhu tubuh, detak jantung, saturasi oksigen (SpO_2), tekanan darah (sistolik dan diastolik), serta gejala pernapasan seperti batuk, sesak napas, dan kelelahan. Variabel-variabel tersebut digunakan sebagai input pada model *Gradient Boosting* untuk menghasilkan klasifikasi penyakit berdasarkan pola hubungan antar fitur, sehingga mampu merepresentasikan kondisi klinis pasien secara komprehensif. Selain itu, penelitian ini juga menggunakan variabel data citra medis berupa gambar rontgen dada. Data citra ini digunakan sebagai input pada model CNN, yang berfungsi untuk mengekstraksi fitur visual dari kondisi paru-paru pasien. Proses ini memungkinkan sistem untuk mengidentifikasi pola-pola tertentu yang berkaitan dengan penyakit pernapasan, seperti pneumonia, tuberkulosis, dan COVID-19. Setiap citra diproses melalui model CNN untuk mengekstraksi fitur visual secara otomatis, seperti pola infiltrasi, opasitas paru, serta kelainan struktur jaringan paru. Proses ekstraksi fitur ini memungkinkan sistem untuk mengenali karakteristik visual yang membedakan setiap jenis penyakit pernapasan.

Adapun variabel output dalam sistem ini berupa hasil diagnosis penyakit pernapasan yang diklasifikasikan ke dalam beberapa kelas, yaitu normal, pneumonia, tuberkulosis, dan COVID-19. Selain itu, sistem juga memberikan informasi tambahan berupa tingkat keparahan penyakit yang dikategorikan menjadi ringan, sedang, dan berat sebagai bentuk dukungan terhadap pengambilan keputusan medis. Dengan demikian, variabel-variabel yang digunakan

dalam penelitian ini tidak hanya mencakup data input yang relevan, tetapi juga menghasilkan output yang informatif dan aplikatif dalam membantu proses diagnosis penyakit pernapasan.

2.3 Rule base Sistem Pakar

Sistem pakar yang dikembangkan dalam penelitian ini menggunakan pendekatan berbasis aturan (*rule-based system*) yang terintegrasi dengan model machine learning. Basis pengetahuan disusun berdasarkan variabel yang terdapat pada dataset klinis, meliputi usia, jenis kelamin, gejala (*Symptom_1*, *Symptom_2*, *Symptom_3*), detak jantung (*Heart_Rate_bpm*), suhu tubuh (*Body_Temperature_C*), tekanan darah (*sistolik* dan *diastolik*), serta saturasi oksigen (*Oxygen_Saturation_%*). Aturan disusun untuk merepresentasikan pola hubungan antara kombinasi gejala dan parameter fisiologis terhadap diagnosis penyakit pernapasan. Gejala yang dominan pada dataset antara lain *fever*, *cough*, *sore throat*, *fatigue*, *shortness of breath*, *headache*, *body ache*, dan *runny nose*. Berdasarkan analisis pola data tersebut, aturan dirumuskan dalam bentuk *IF-THEN* sebagai berikut.

Aturan pertama disusun berdasarkan kombinasi gejala dan suhu tubuh sebagai indikator awal kondisi penyakit. Aturan-aturan tersebut adalah sebagai berikut:

Rule 1:

IF (Fever = Ya) AND (Body_Temperature_C \geq 38°C) AND (Cough = Ya)
THEN Diagnosis = Flu

Rule 2:

IF (Fever = Ya) AND (Body_Temperature_C \geq 39°C) AND (Shortness of breath = Ya)
THEN Diagnosis = Pneumonia

Rule 3:

IF (Cough = Ya) AND (Shortness of breath = Ya)
THEN Diagnosis = Bronchitis

Rule 4:

IF (Runny nose = Ya) AND (Sore throat = Ya)
THEN Diagnosis = Cold

Rule 5:

IF (Fatigue = Ya) AND (Body ache = Ya) AND (Fever = Ya)
THEN Diagnosis = Flu

Selanjutnya, aturan juga disusun berdasarkan parameter fisiologis pasien, khususnya nilai saturasi oksigen dan detak jantung yang berpengaruh terhadap kondisi pernapasan. Aturan tersebut dirumuskan sebagai berikut:

Rule 6:

IF (Oxygen_Saturation_% < 92)
THEN Diagnosis = Pneumonia

Rule 7:

IF (Oxygen_Saturation_% \geq 95) AND (Tidak terdapat gejala berat)
THEN Diagnosis = Healthy

Rule 8:

IF (Heart_Rate_bpm > 110) AND (Fever = Ya)
THEN Indikasi = Infeksi Berat

Selain itu, sistem juga memanfaatkan hasil klasifikasi yang dihasilkan oleh model Gradient Boosting sebagai dasar utama dalam penentuan diagnosis. Hasil prediksi model digunakan secara langsung dalam aturan berikut:

Rule 9:

IF (Hasil prediksi model = Flu)
THEN Diagnosis = Flu

Rule 10:

IF (Hasil prediksi model = Bronchitis)
THEN Diagnosis = Bronchitis

Rule 11:

IF (Hasil prediksi model = Pneumonia)
THEN Diagnosis = Pneumonia

Rule 12:

IF (Hasil prediksi model = Cold)
THEN Diagnosis = Cold

Rule 13:

IF (Hasil prediksi model = Healthy)

THEN Diagnosis = Healthy

Selanjutnya, sistem menentukan tingkat keparahan penyakit berdasarkan nilai saturasi oksigen yang diperoleh dari data klinis. Aturan penentuan tingkat keparahan dirumuskan sebagai berikut:

Rule 14:

IF (Oxygen_Saturation_% \geq 95)

THEN Severity = Mild

Rule 15:

IF (Oxygen_Saturation_% antara 92–94)

THEN Severity = Moderate

Rule 16:

IF (Oxygen_Saturation_% $<$ 92)

THEN Severity = Severe

Pada data citra medis, aturan digunakan untuk menginterpretasikan hasil klasifikasi citra rontgen dada yang dihasilkan oleh model CNN. Citra yang telah diproses melalui tahapan resizing dan normalisasi akan dievaluasi oleh model CNN untuk menghasilkan label klasifikasi. Hasil tersebut kemudian dipetakan ke dalam diagnosis menggunakan aturan berikut:

Rule 17:

IF (Output CNN = Pneumonia)

THEN Diagnosis = Pneumonia

Rule 18:

IF (Output CNN = Tuberkulosis)

THEN Diagnosis = Tuberkulosis

Rule 19:

IF (Output CNN = COVID-19)

THEN Diagnosis = COVID-19

Rule 20:

IF (Output CNN = Normal)

THEN Diagnosis = Healthy

Proses inferensi dalam sistem pakar ini menggunakan metode forward chaining, di mana sistem akan mencocokkan data input pengguna dengan aturan yang tersedia untuk menghasilkan diagnosis akhir. Integrasi antara hasil prediksi model dan rule base memungkinkan sistem untuk meningkatkan akurasi sekaligus memberikan interpretasi yang lebih transparan. Dengan demikian, sistem tidak hanya berfungsi sebagai alat klasifikasi berbasis machine learning, tetapi juga sebagai sistem pendukung keputusan yang mampu menjelaskan hasil diagnosis berdasarkan aturan yang telah ditentukan.

3. Hasil dan Pembahasan

3.1. Data Penelitian

Data penelitian yang digunakan dalam sistem pakar ini terdiri dari dua jenis, yaitu data klinis dan data citra medis yang diperoleh dari platform Kaggle. Data klinis digunakan sebagai input untuk model *Gradient Boosting*, sedangkan data citra medis digunakan sebagai input untuk model CNN.

Dataset data klinis terdiri dari 2.000 data pasien yang mencakup berbagai parameter medis, seperti gejala dan tanda vital, termasuk suhu tubuh, detak jantung, saturasi oksigen, serta tekanan darah. Data ini digunakan untuk mengklasifikasikan beberapa kondisi penyakit pernapasan, seperti bronchitis, flu, pneumonia, serta kondisi sehat. Sementara itu, dataset citra medis terdiri dari 7.132 citra rontgen dada yang terbagi ke dalam beberapa kategori, yaitu 572 citra COVID-19, 4.273 citra pneumonia, 700 citra tuberkulosis, dan 1.583 citra paru-paru normal. Dataset ini digunakan untuk melatih dan menguji model CNN dalam mengenali pola visual penyakit pernapasan. Data yang digunakan dalam penelitian ini telah melalui proses pembagian data menjadi data latih dan data uji untuk memastikan bahwa model yang dikembangkan dapat dievaluasi secara objektif terhadap data yang belum pernah dilihat sebelumnya. Sebagai representasi data yang digunakan, Tabel 1 menunjukkan contoh sampel data klinis yang digunakan dalam proses pelatihan dan pengujian model.

Tabel 1. Sampel Data Klinis

Age	Gender	Symptom 1	Symptom 2	Symptom 3	Heart Rate (bpm)	Body Temperature (C)	Systol	Diastol	Oxygen Saturation	Diagnosis	Severity
74	Male	Fatigue	Sore throat	Fever	69	39.4	132	91	94	Flu	Moderate
22	Male	Sore throat	Fever	Cough	90	39.5	107	92	93	Flu	Moderate
21	Male	Sore throat	Fatigue	Cough	71	37.5	126	82	93	Bronchitis	Severe
19	Female	Fatigue	Fever	Runny nose	101	39.8	100	67	91	Pneumonia	Severe
28	Female	Runny nose	Sore throat	Cough	63	36.3	151	96	96	Cold	Mild

3.2. Proses Deteksi Penyakit Pernapasan Menggunakan *Gradient Boosting* dan Metode CNN

Proses deteksi penyakit pernapasan dalam sistem pakar ini dilakukan melalui serangkaian tahapan yang dimulai dari pengolahan data input hingga menghasilkan diagnosis akhir.

Tabel 2. Alur Deteksi Penyakit Menggunakan Gradient Boosting

Langkah	Tahap Proses	Data Masukan	Proses yang Dilakukan	Data Keluaran
1	Input Data	Usia, gender, gejala (fatigue, sore throat, fever)	Pengambilan data pasien	Data klinis mentah
2	Parameter Fisiologis	Heart rate, suhu tubuh, SpO2	Pencatatan kondisi fisiologis pasien	Data fisiologis terstruktur
3	Preprocessing	Data klinis mentah + fisiologis	Transformasi ke format numerik dan kategorikal	Fitur siap diproses model
4	Model Gradient Boosting	Fitur klinis	Analisis hubungan antar fitur untuk klasifikasi	Prediksi awal (Flu)
5	Interpretasi Rule Base	Prediksi model + gejala	Penerapan aturan berbasis gejala dan suhu	Diagnosis sementara (Flu)
6	Penentuan Severity	Nilai SpO2 (94%)	Klasifikasi tingkat keparahan	Moderate
7	Output Akhir	Diagnosis sementara + severity	Penggabungan hasil model dan rule base	Flu (Moderate)

Tabel 2 menunjukkan alur proses deteksi penyakit pernapasan menggunakan data klinis dengan metode Gradient Boosting. Proses dimulai dari input data pasien yang mencakup informasi gejala dan parameter fisiologis, kemudian dilakukan tahap preprocessing untuk mengubah data ke dalam format yang sesuai dengan model. Selanjutnya, model Gradient Boosting melakukan klasifikasi berdasarkan hubungan antar fitur yang terdapat dalam data klinis. Hasil prediksi yang dihasilkan oleh model kemudian diinterpretasikan menggunakan rule base untuk memperkuat diagnosis berdasarkan kondisi gejala pasien. Selain itu, sistem juga menentukan tingkat keparahan penyakit berdasarkan nilai saturasi oksigen (SpO2). Tahap akhir menghasilkan diagnosis yang merupakan kombinasi dari hasil model dan aturan sistem pakar, sehingga memberikan hasil yang lebih informatif dan mudah dipahami.

Sebagai contoh proses diagnosis menggunakan metode Gradient Boosting, sistem menerima data pasien berupa usia, jenis kelamin, gejala, suhu tubuh, detak jantung, tekanan darah, dan saturasi oksigen. Salah satu sampel data memiliki karakteristik berupa usia 35 tahun, jenis kelamin laki-laki, gejala fever dan cough, suhu tubuh 38,7°C, serta saturasi oksigen sebesar 91%. Data tersebut terlebih dahulu melalui tahap preprocessing dan diubah ke dalam bentuk data tabular sebelum diproses oleh model Gradient Boosting. Selanjutnya, model melakukan klasifikasi secara iteratif menggunakan beberapa decision tree untuk menghasilkan prediksi akhir berdasarkan pola data pelatihan. Berdasarkan hasil klasifikasi, sistem menghasilkan diagnosis Pneumonia dengan tingkat keparahan severe karena nilai saturasi oksigen berada di bawah batas normal.

Tabel 3. Alur Deteksi Penyakit Menggunakan CNN

Langkah	Tahap Proses	Data Masukan	Proses yang Dilakukan	Data Keluaran
1	Input Data	Citra rontgen dada	Pengambilan citra pasien	Citra mentah
2	Preprocessing	Citra mentah	Resize dan normalisasi piksel	Citra siap diproses
3	Model CNN	Citra	Ekstraksi fitur visual menggunakan CNN	Fitur visual paru
4	Analisis Citra	Fitur visual	Deteksi pola penyakit (infiltrasi paru)	Pola penyakit teridentifikasi
5	Output Model	Pola penyakit	Klasifikasi citra oleh CNN	Prediksi (Pneumonia)
6	Interpretasi Rule Base	Output CNN	Pemetaan hasil klasifikasi ke diagnosis	Diagnosis Pneumonia
7	Output Akhir	Diagnosis dari CNN	Penetapan hasil akhir sistem	Pneumonia

Tabel 3 menyajikan proses deteksi penyakit pernapasan menggunakan data citra medis dengan metode *Convolutional Neural Network* (CNN). Proses dimulai dari input citra rontgen dada yang kemudian melalui tahap preprocessing untuk menyesuaikan ukuran dan nilai piksel citra. Selanjutnya, model CNN digunakan untuk mengekstraksi fitur visual dari citra, seperti pola infiltrasi paru yang berkaitan dengan penyakit pernapasan. Fitur yang telah diekstraksi kemudian dianalisis untuk mengidentifikasi pola penyakit, yang selanjutnya diklasifikasikan oleh model sebagai jenis penyakit tertentu. Hasil klasifikasi tersebut kemudian diinterpretasikan menggunakan rule base untuk menghasilkan diagnosis akhir. Dengan pendekatan ini, sistem mampu mengenali pola visual pada citra medis secara efektif dan memberikan hasil diagnosis yang akurat.

Pada proses diagnosis menggunakan metode *Convolutional Neural Network* (CNN), sistem menerima input berupa citra rontgen dada pasien yang kemudian diproses untuk mengidentifikasi pola penyakit pernapasan. Salah satu sampel citra yang digunakan dalam penelitian ini merupakan citra rontgen pasien dengan indikasi pneumonia. Sebelum dilakukan klasifikasi, citra terlebih dahulu melalui tahap preprocessing berupa resizing menjadi ukuran 224×224 piksel agar sesuai dengan ukuran input model CNN. Selanjutnya, citra diproses melalui convolution layer untuk mengekstraksi fitur penting pada paru-paru, seperti area infiltrasi dan opasitas, kemudian dilanjutkan dengan pooling layer untuk menyederhanakan hasil ekstraksi fitur. Hasil fitur tersebut diteruskan ke *fully connected layer* untuk proses klasifikasi akhir. Berdasarkan hasil klasifikasi, model CNN mengidentifikasi citra sebagai Pneumonia karena pola citra memiliki karakteristik yang sesuai dengan data pelatihan pada kelas pneumonia.

3.3. Pengujian Performa Akurasi Sistem Pakar

Proses pengujian dilakukan secara terpisah untuk masing-masing metode yang digunakan, yaitu *Gradient Boosting* untuk data klinis dan CNN untuk data citra medis. Pada pengujian menggunakan metode *Gradient Boosting*, evaluasi kinerja dilakukan dengan menggunakan confusion matrix untuk mengukur tingkat ketepatan klasifikasi pada setiap kelas penyakit. *Confusion matrix* digunakan untuk menilai kinerja klasifikasi setiap kelas. Gambar 1 menampilkan *confusion matrix* terdapat lima kelas penyakit pernapasan yang digunakan untuk menguji kinerja dari sistem. Kelas-kelas ini terdiri dari: Bronkitis (*Bronchitis*), Pilek (*Cold*), Flu, Sehat (*Healthy*), dan Pneumonia. Matriks menunjukkan bahwa ada data yang diprediksi dengan benar, yaitu 166 data untuk *Bronchitis*, 84 data untuk *Cold*, 163 data untuk Flu, 563 data untuk *Healthy*, dan 21 data untuk Pneumonia. Matriks pada Gambar 1 menunjukkan terdapat kesalahan prediksi: 2 kasus Flu diprediksi sebagai Pneumonia, 1 kasus Pneumonia diprediksi sebagai *Cold*, dan 1 kasus *Healthy* diprediksi sebagai Pneumonia. Secara keseluruhan, nilai akurasi total data yang telah diproses melalui model *Gradient Boosting* adalah 99,7%. Komponen utama yang digunakan dalam matriks ini adalah *True Positives* (TP), *False Positives* (FP), dan *False Negatives* (FN).

Selain menggunakan *confusion matrix*, evaluasi juga dilakukan dengan menggunakan metrik precision, recall, dan F1-score untuk menilai performa model pada masing-masing kelas. Hasil pengujian menunjukkan bahwa sebagian besar kelas memiliki nilai precision, recall, dan F1-score yang mendekati 1, yang menunjukkan bahwa model memiliki tingkat akurasi dan konsistensi yang tinggi dalam melakukan klasifikasi.

True Class	Bronchitis	166				
	Cold		84			
	Flu			163		2
	Healthy				563	
	Pneumonia				1	21
		Predicted Class				
		Bronchitis	Cold	Flu	Healthy	Pneumonia

Gambar 1. Confusion Matrix

Tabel 4. Hasil Klasifikasi Kelas

Class	Precision	Recall	F1-Score
Bronchitis	1.000	1.000	1.000
Cold	1.000	1.000	1.000
Flu	0.9939	0.9879	0.9909
Healthy	0.9965	1.000	0.9982
Pneumonia	0.913	0.9545	0.9333

Tabel 4 menunjukkan hasil evaluasi kinerja per kelas, menunjukkan nilai presisi (*precision*), *recall*, dan skor F1 (*F1-Score*) untuk setiap kelas. Nilai presisi mengukur frekuensi prediksi positif model yang benar. Nilai *recall* digunakan untuk mengukur jumlah kasus positif yang diidentifikasi dengan benar. Skor F1 digunakan untuk mengukur rata-rata antara presisi dan *recall*. Hasil pengujian menunjukkan bahwa sebagian besar kelas memiliki nilai presisi, *recall*, dan F1-score yang mendekati 1, yang menunjukkan bahwa model memiliki tingkat akurasi dan konsistensi yang tinggi dalam melakukan klasifikasi. Nilai presisi dan *recall* yang tidak mencapai 1.00 menunjukkan adanya kesalahan klasifikasi pada kasus dengan karakteristik yang serupa, seperti antara Flu dan Pneumonia yang memiliki gejala klinis yang mirip. Hal ini mengindikasikan bahwa meskipun model memiliki performa yang tinggi, masih terdapat tantangan dalam membedakan kelas dengan pola gejala yang saling tumpang tindih.

Selain evaluasi berbasis keseluruhan data, dilakukan pula pengujian menggunakan sampel data yang telah disajikan pada Subbab 3.1. Pengujian ini dilakukan dengan membandingkan hasil diagnosis sistem dengan hasil preferensi pakar untuk setiap kasus yang diuji.

Table 5. Perbandingan Diagnosis Sistem dan Pakar

No	Diagnosis Pakar	Diagnosis Sistem
1	Flu	Flu
2	Flu	Pneumonia
3	Bronchitis	Bronchitis
4	Pneumonia	Pneumonia
5	Cold	Cold

Berdasarkan Tabel 5, terdapat satu kasus ketidaksesuaian antara hasil diagnosis sistem dan pakar, yaitu pada data ke-2, di mana sistem mengklasifikasikan Flu sebagai Pneumonia. Secara keseluruhan, sistem berhasil memberikan diagnosis yang sesuai dengan pakar pada 4 dari 5 kasus yang diuji. Berdasarkan hasil tersebut, diperoleh nilai presisi sebesar 0.80 dan *recall* sebesar 0.80. Nilai ini menunjukkan bahwa sistem memiliki kemampuan yang cukup baik dalam mendeteksi penyakit pernapasan berdasarkan data klinis, meskipun masih terdapat kesalahan pada kasus dengan karakteristik gejala yang mirip.

Pada data citra medis, model CNN juga menunjukkan performa yang baik dalam mengklasifikasikan citra rontgen dada. Model dilatih menggunakan parameter pelatihan yang ditunjukkan pada Tabel 6.

Tabel 6. Parameter Pelatihan

Parameter	Value
<i>Epoch</i>	15 of 15
<i>Iteration</i>	2340 of 2340
<i>Iteration per Epoch</i>	156
<i>Maximum Iterations</i>	2340
Parameter	Value
<i>Validation Frequency</i>	30 Iterations
<i>Learning Rate</i>	0.001

Berdasarkan parameter pelatihan tersebut, model CNN menghasilkan tingkat akurasi validasi sebesar 95,93%. Hasil ini menunjukkan bahwa model CNN mampu mengklasifikasikan citra medis dengan baik berdasarkan pola visual yang terdeteksi.

Secara keseluruhan, hasil pengujian menunjukkan bahwa sistem pakar yang dikembangkan memiliki performa yang tinggi dalam mendeteksi penyakit pernapasan, baik pada data klinis maupun data citra medis. Integrasi antara metode Gradient Boosting, CNN, dan rule base memungkinkan sistem untuk memberikan hasil diagnosis yang akurat sekaligus dapat diinterpretasikan secara logis.

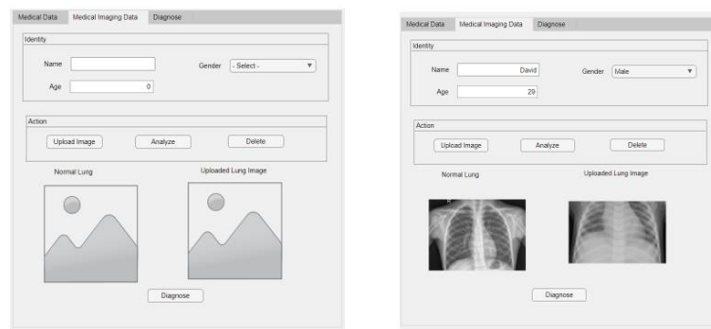
3.4 Implementasi Sistem dan Antarmuka Pengguna

Sistem pakar yang dikembangkan dalam penelitian ini diimplementasikan dalam bentuk aplikasi yang dilengkapi dengan antarmuka pengguna (user interface) untuk memfasilitasi interaksi antara pengguna dan sistem. Antarmuka dirancang dengan pendekatan yang sederhana dan intuitif, sehingga pengguna dapat dengan mudah mengoperasikan sistem tanpa memerlukan keahlian teknis yang mendalam. Secara umum, sistem menyediakan dua fitur utama, yaitu pengolahan data klinis dan analisis citra medis. Fitur pengolahan data klinis pada Gambar 2, pengguna dapat memasukkan informasi pasien yang meliputi usia, jenis kelamin, gejala, serta parameter fisiologis seperti suhu tubuh, detak jantung, tekanan darah, dan saturasi oksigen. Data yang dimasukkan kemudian diproses oleh model *Gradient Boosting* untuk menghasilkan diagnosis penyakit pernapasan.

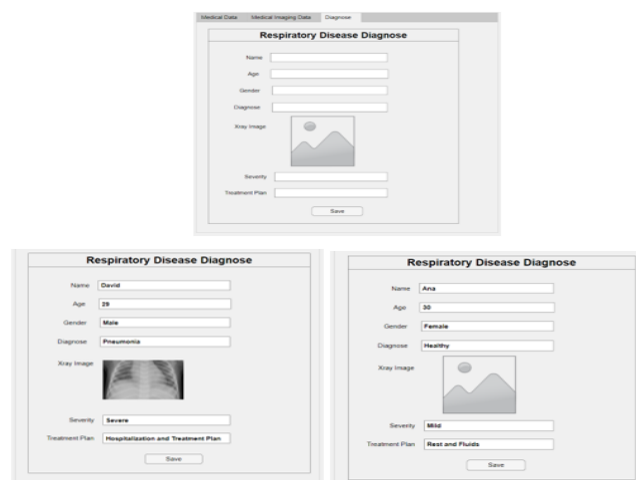
Gambar 2. Medical Data

Sementara itu, Gambar 3 menunjukkan fitur analisis citra medis, pengguna dapat mengunggah citra rontgen dada melalui antarmuka yang telah disediakan. Citra tersebut kemudian diproses oleh model CNN untuk mengidentifikasi pola visual yang berkaitan dengan penyakit pernapasan.

Setelah proses analisis selesai, sistem akan menampilkan hasil diagnosis seperti pada Gambar 4. Mencakup jenis penyakit yang terdeteksi, tingkat keparahan, serta informasi tambahan yang dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan. Dengan adanya antarmuka ini, sistem tidak hanya berfungsi sebagai model klasifikasi berbasis machine learning, tetapi juga sebagai sistem pendukung keputusan yang dapat digunakan secara langsung oleh pengguna. Implementasi antarmuka pengguna ini juga mendukung fleksibilitas sistem dalam menangani berbagai jenis *input* data, sehingga pengguna dapat memilih metode analisis sesuai dengan data yang dimiliki. Hal ini menjadikan sistem lebih adaptif dan aplikatif dalam berbagai kondisi penggunaan.



Gambar 3. Medical Imaging Data



Gambar 4. Respiratory Disease Diagnose

3.5 Pembahasan

Hasil pengujian performa pada penelitian ini menunjukkan bahwa sistem pakar yang dikembangkan memiliki tingkat akurasi yang tinggi dalam mendeteksi penyakit pernapasan, baik pada data klinis maupun data citra medis. Pada data klinis, model Gradient Boosting menghasilkan akurasi sebesar 99,7% dengan nilai precision, recall, dan F1-score yang mendekati 1 pada sebagian besar kelas. Sementara itu, pada data citra medis, model CNN menghasilkan akurasi validasi sebesar 95,93%, yang menunjukkan kemampuan model dalam mengenali pola visual pada citra rontgen dada secara efektif. Temuan ini memperkuat hasil penelitian terdahulu yang menyatakan bahwa metode Gradient Boosting memiliki performa yang sangat baik dalam mengolah data terstruktur dan menghasilkan klasifikasi yang akurat [10], [11]. Selain itu, hasil penelitian ini juga sejalan dengan studi yang menunjukkan bahwa CNN mampu mengekstraksi fitur visual secara otomatis dan efektif dalam klasifikasi citra medis, khususnya pada kasus pneumonia [8], [9]. Penelitian lain juga menunjukkan bahwa kombinasi pendekatan berbasis deep learning dan ensemble learning dapat meningkatkan performa klasifikasi pada data yang kompleks [16], [17].

Namun demikian, penelitian ini memberikan penguatan tambahan dengan menunjukkan bahwa kedua metode tersebut dapat diintegrasikan dalam satu sistem pakar yang adaptif terhadap jenis data yang tersedia. Berbeda dengan penelitian sebelumnya yang umumnya hanya berfokus pada satu jenis data, penelitian ini menggabungkan dua pendekatan berbeda dalam satu sistem yang mampu memilih metode secara otomatis berdasarkan input yang diberikan. Dengan demikian, temuan dalam penelitian ini tidak hanya mendukung hasil penelitian terdahulu, tetapi juga mengintegrasikannya dalam suatu pendekatan yang lebih komprehensif. Meskipun hasil yang diperoleh menunjukkan performa yang tinggi, terdapat beberapa kesalahan klasifikasi, terutama pada kelas dengan karakteristik gejala yang serupa, seperti antara Flu dan Pneumonia. Hal ini menunjukkan bahwa meskipun model memiliki kemampuan yang baik dalam mengenali pola data, masih terdapat tantangan dalam

membedakan kelas dengan fitur yang saling tumpang tindih. Temuan ini konsisten dengan penelitian sebelumnya yang menyatakan bahwa kemiripan fitur antar kelas dapat mempengaruhi performa model klasifikasi [14], [15].

Dari sisi kontribusi ilmiah, penelitian ini memberikan kontribusi dalam pengembangan sistem pakar berbasis kecerdasan buatan dengan pendekatan multimodal, yaitu mengintegrasikan data klinis dan data citra medis dalam satu sistem deteksi penyakit. Pendekatan ini memungkinkan sistem untuk tetap memberikan diagnosis meskipun hanya satu jenis data yang tersedia, sehingga meningkatkan fleksibilitas dan kegunaan sistem dalam praktik nyata. Selain itu, integrasi antara model machine learning dan rule base juga memberikan kontribusi dalam pengembangan sistem pendukung keputusan yang tidak hanya akurat, tetapi juga dapat memberikan interpretasi yang lebih transparan terhadap hasil diagnosis. Lebih lanjut, penelitian ini juga menunjukkan bahwa penggunaan rule base sebagai lapisan interpretasi dapat membantu menjembatani hasil prediksi model dengan penalaran medis yang lebih mudah dipahami. Hal ini menjadi nilai tambah dibandingkan dengan sistem berbasis machine learning murni yang cenderung bersifat black-box. Dengan demikian, penelitian ini berkontribusi tidak hanya pada peningkatan performa klasifikasi, tetapi juga pada aspek interpretabilitas sistem dalam bidang kesehatan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan multimodal yang mengintegrasikan *Gradient Boosting*, CNN, dan rule base dalam satu sistem pakar merupakan solusi yang efektif dalam mendukung proses diagnosis penyakit pernapasan. Pendekatan ini tidak hanya memperkuat temuan-temuan sebelumnya, tetapi juga memberikan kontribusi baru dalam pengembangan sistem diagnosis yang lebih adaptif, akurat, dan mudah digunakan dalam praktik klinis. Hasil yang diperoleh dalam penelitian ini selanjutnya diimplementasikan dalam bentuk sistem pakar yang dilengkapi dengan antarmuka pengguna untuk memudahkan proses interaksi, yang akan dijelaskan pada subbab berikutnya.

4. Kesimpulan

Penelitian ini menunjukkan bahwa sistem pakar yang digunakan berhasil membuktikan kelayakan dan efektivitas dalam mendeteksi penyakit saluran pernapasan secara cepat dan akurat dengan menggunakan dua metode *machine learning* yang berbeda, yaitu *Gradient Boosting* dan CNN. Temuan ini sejalan dengan penelitian-penelitian terdahulu yang menunjukkan bahwa metode *Gradient Boosting* memiliki kemampuan yang baik dalam mengolah data terstruktur dan menghasilkan akurasi klasifikasi yang tinggi [7], [18]. Selain itu, penggunaan *Convolutional Neural Networks* (CNN) dalam penelitian ini juga memperkuat temuan sebelumnya bahwa CNN efektif dalam mengekstraksi fitur visual dari citra medis, khususnya dalam klasifikasi penyakit pernapasan seperti pneumonia [10], [11]. Hasil penelitian ini juga mendukung studi yang mengombinasikan metode berbasis *deep learning* dan ensemble learning untuk meningkatkan performa klasifikasi pada data yang kompleks [17]. Dalam penelitian tersebut, integrasi CNN dengan metode *Gradient Boosting* menunjukkan peningkatan akurasi dibandingkan penggunaan metode tunggal. Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada satu jenis data, baik data klinis maupun data citra, sehingga belum sepenuhnya mengakomodasi kebutuhan sistem diagnosis yang fleksibel. Dalam konteks ini, penelitian yang dilakukan memberikan penguatan terhadap pentingnya pemanfaatan metode yang mampu menangani berbagai jenis data dalam sistem deteksi penyakit. Hasil pengujian menunjukkan bahwa metode *Gradient Boosting* menghasilkan akurasi sebesar 99,7% pada data klinis, sedangkan metode CNN menghasilkan akurasi sebesar 95,93% pada data citra medis. Selain itu, sistem juga menunjukkan kemampuan yang baik dalam menghasilkan diagnosis yang mendekati hasil preferensi pakar.

Kontribusi utama penelitian ini terletak pada pengembangan sistem pakar yang mampu mengintegrasikan dua jenis data, yaitu data non-visual dan data visual, secara adaptif dalam satu sistem diagnosis. Pendekatan ini memungkinkan sistem untuk tetap memberikan hasil diagnosis meskipun hanya satu jenis data yang tersedia, sehingga meningkatkan fleksibilitas dan kegunaan sistem dalam praktik nyata. Meskipun memiliki tingkat akurasi yang tinggi, sistem masih memiliki keterbatasan dalam membedakan kelas dengan karakteristik yang serupa serta belum dapat memproses data klinis dan citra secara bersamaan dalam satu proses. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan metode yang mampu meningkatkan kemampuan klasifikasi pada kelas yang memiliki kemiripan pola serta meningkatkan interpretabilitas hasil model.

Daftar Referensi

- [1] S. M. Levine and D. D. Marciniuk, "Global Impact of Respiratory Disease: What Can We Do, Together, to Make a Difference?," May 01, 2022, *Elsevier Inc.* doi: 10.1016/j.chest.2022.01.014.
- [2] D. K. Ahorsu, C. Y. Lin, V. Imani, M. Saffari, M. D. Griffiths, and A. H. Pakpour, "The Fear of COVID-19 Scale: Development and Initial Validation," *Int. J. Ment. Health Addict.*, vol. 20, no. 3, pp. 1537–1545, 2022, doi: 10.1007/s11469-020-00270-8.
- [3] A. Zolanda, M. Raharjo, and O. Setiani, "Faktor Risiko Kejadian Infeksi Saluran Pernafasan Akut Pada Balita Di Indonesia," *Link*, vol. 17, no. 1, pp. 73–80, 2021, doi: 10.31983/link.v17i1.6828.
- [4] G. E. Saputra, R. F. Simbolon, M. Hanindia, and P. Swari, "Sistem Pakar Untuk Mendiagnosis Penyakit Infeksi Saluran Pernapasan Akut (ISPA) Menggunakan Algoritma Dempster Shafer," *Semin. Nas. Inform. Bela Negara*, vol. 4, pp. 6–13, 2024.
- [5] R. Hasanah And F. Helmi, Muhasshanah, "Sistem Pakar Diagnosa Penyakit Saluran Pernapasan Menggunakan Metode Forward Chaining," vol. 1, no. 1, pp. 33–50, 2022.
- [6] T. F. Ramadhani, I. Fitri, and E. T. E. Handayani, "Sistem Pakar Diagnosa Penyakit ISPA Berbasis Web Dengan Metode Forward Chaining," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 5, no. 2, pp. 81–90, 2020.
- [7] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas," *Techno.Com*, vol. 20, no. 3, pp. 400–410, 2021, doi: 10.33633/tc.v20i3.4576.
- [8] B. Nugroho and E. Y. Puspaningrum, "Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 533–538, 2021, doi: 10.25126/jtiik.2021834515.
- [9] Y. Li, J. Zhao, Z. Lv, and Z. Pan, "Multimodal Medical Supervised Image Fusion Method by CNN," *Front. Neurosci.*, vol. 15, 2021, doi: 10.3389/fnins.2021.638976.
- [10] Sri Diantika, Hiya Nalatissifa, Riki Supriyadi, Nurlaelatul Maulidah, and Ahmad Fauzi, "Implementation of Multi-Class Gradient Boosting To Classify Animal Species in Zoos," *Antivirus J. Ilm. Tek. Inform.*, vol. 17, no. 1, pp. 32–40, 2023, doi: 10.35457/antivirus.v17i1.2812.
- [11] D. Boldini, F. Grisoni, D. Kuhn, L. Friedrich, and S. A. Sieber, "Practical guidelines for the use of gradient boosting for molecular property prediction," *J. Cheminform.*, vol. 15, no. 1, pp. 1–13, 2023, doi: 10.1186/s13321-023-00743-7.
- [12] S. P. Nainggolan and A. Sinaga, "Comparative Analysis Of Accuracy Of Random Forest And Gradient Boosting Classifier Algorithm For Diabetes Classification," *Sebatik*, vol. 27, no. 1, pp. 97-102, 2023, doi: 10.46984/sebatik.v27i1.2157.
- [13] Q. T. Bui *et al.*, "Gradient boosting machine and object-based cnn for land cover classification," *Remote Sens.*, vol. 13, no. 14, pp. 1–15, 2021, doi: 10.3390/rs13142709.
- [14] O. Lyashevskaya, F. Malone, E. MacCarthy, J. Fiehler, J. H. Buhk, and L. Morris, "Class imbalance in gradient boosting classification algorithms: Application to experimental stroke data," *Stat. Methods Med. Res.*, vol. 30, no. 3, pp. 916–925, 2021, doi: 10.1177/0962280220980484.
- [15] P. Septiana Rizky, R. Haiban Hirzi, and U. Hidayaturrohman, "Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 228–236, 2022, doi: 10.36456/jstat.vol15.no2.a5548.
- [16] Q. T. Bui *et al.*, "Gradient boosting machine and object-based CNN for land cover classification," *Remote Sensing*, vol. 13, no. 14, p. 2709, 2021, doi: 10.3390/rs13142709..
- [17] A. Gasmi, "Machine learning techniques for the identification and diagnosis of COVID-19," in *EAI/Springer Innovations in Communication and Computing*, 2021, pp. 231–256. doi: 10.1007/978-3-030-68936-0_12.
- [18] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023, doi: 10.3390/su15075930.