

Implementasi Ekstraksi Fitur dalam Deteksi Pneumonia

Azila Lailannafisa¹, Fetty Tri Anggraeny², Hendra Maulana³

^{1,2,3}Informatika, FAKultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur

E-mail: ¹21081010286@student.upnjatim.ac.id, ²fettyanggraeny.if@upnjatim.ac.id,
³hendra.maulana.if@upnjatim.ac.id

Abstrak

Pneumonia adalah infeksi paru-paru yang menjadi salah satu penyebab utama kematian di seluruh dunia. Deteksi dini yang cepat dan akurat sangat penting dalam penanganan penyakit ini. Penelitian ini bertujuan mengembangkan sistem klasifikasi pneumonia berbasis citra X-ray dengan menggabungkan ekstraksi fitur Local Binary Pattern (LBP) dan Gray Level Co-occurrence Matrix (GLCM), serta menggunakan algoritma XGBoost untuk proses klasifikasi. Metode yang digunakan mencakup pemrosesan citra, yaitu konversi citra ke grayscale, median filtering, dan adaptive histogram equalization. Fitur tekstur diekstraksi menggunakan LBP dan GLCM untuk memperoleh informasi yang lebih mendalam. Selanjutnya, model XGBoost dilatih dengan mengoptimalkan hyperparameter melalui teknik random search. Hasil eksperimen menunjukkan bahwa model yang dikembangkan memperoleh akurasi sebesar 95.72%, presisi sebesar 94.67%, recall sebesar 99.34% dan F1-Score sebesar 96.95%. Penelitian ini membuktikan bahwa kombinasi LBP, GLCM, dan XGBoost dapat menjadi alternatif yang efisien dalam deteksi pneumonia, dengan keunggulan akurasi tinggi dan efisiensi komputasi dibandingkan dengan teknik deep learning yang memerlukan sumber daya lebih besar.

Kata Kunci— Pneumonia, Citra X-ray, Local Binary Pattern, Gray Level Co-occurrence Matrix, XGBoost.

Abstract

Pneumonia is a lung infection that remains one of the leading causes of death worldwide. Early and accurate detection is crucial for effective treatment. This study aims to develop a pneumonia classification system based on X-ray images by combining feature extraction techniques, namely Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM), and using the XGBoost algorithm for classification. The methods used in this research include image preprocessing, such as converting images to grayscale, applying median filtering, and adaptive histogram equalization. Texture features are extracted using LBP and GLCM to obtain more detailed information. The XGBoost model is then trained by optimizing hyperparameters through random search techniques. Experimental results show that the developed model achieves an accuracy of 95.72%, precision of 94.67%, recall of 99.34%, and an F1-score of 96.95%. This research demonstrates that the combination of LBP, GLCM, and XGBoost provides an efficient alternative for pneumonia detection, with high accuracy and computational efficiency compared to deep learning techniques, which require more resources.

Keywords— Pneumonia, X-ray images, Local Binary Pattern, Gray Level Co-occurrence Matrix, XGBoost.

1. PENDAHULUAN

Pneumonia adalah infeksi pernapasan yang menyerang paru-paru, menyebabkan peradangan dan pengisian kantung udara dengan nanah serta cairan, yang menghambat aliran oksigen dan menyebabkan rasa nyeri saat bernapas. Infeksi ini terus menjadi salah satu penyebab utama kematian akibat penyakit di seluruh dunia, dengan sekitar 2,2 juta kematian tercatat pada tahun 2019, termasuk 502.000 di antaranya adalah anak-anak di bawah lima tahun (*Pneumonia in Children*, n.d.). Penyebab pneumonia bervariasi, mencakup bakteri seperti *Streptococcus pneumoniae* dan *Haemophilus influenzae*, serta patogen virus dan jamur (*Pneumonia Prevention and Control | Pneumonia | CDC*, n.d.). Dalam konteks medis, pneumonia menjadi salah satu masalah kesehatan terbesar, bersama dengan penyakit pernapasan lainnya seperti asma dan Penyakit Paru Obstruktif Kronik (PPOK), yang menyebabkan perhatian khusus terkait dampaknya terhadap kesehatan masyarakat.

Secara tradisional, diagnosis pneumonia dilakukan dengan memanfaatkan foto rontgen dada. Namun, pendekatan ini memiliki beberapa kelemahan, termasuk ketergantungan pada interpretasi subjektif dan keterbatasan tenaga ahli di daerah-daerah yang sulit dijangkau. Komputer, yang telah berkembang sebagai alat pengolah data dan penghasil informasi, kini turut berperan dalam pengambilan keputusan. Tidak puas hanya dengan fungsi tersebut, para ahli komputer terus mengembangkan kecanggihan komputer agar dapat memiliki kemampuan serupa dengan manusia. Kesehatan adalah sesuatu yang sangat berharga; tanpa kesehatan, manusia tidak dapat menikmati hidup sepenuhnya. Mengingat hal tersebut, teknologi pengambilan keputusan otomatis mulai banyak diterapkan, terutama dengan menggunakan metode pemrosesan citra medis yang didukung oleh algoritma pembelajaran mesin. Teknologi ini memungkinkan analisis citra dengan tingkat akurasi tinggi yang sebelumnya hanya dapat dicapai oleh tenaga medis berpengalaman (Litjens et al., 2017). Penggunaan algoritma pembelajaran mesin untuk deteksi dini pneumonia berbasis citra X-ray menunjukkan potensi besar untuk meningkatkan kecepatan dan ketepatan diagnosis, terutama pada kasus-kasus dengan gejala yang sulit dibedakan tanpa teknologi bantuan.

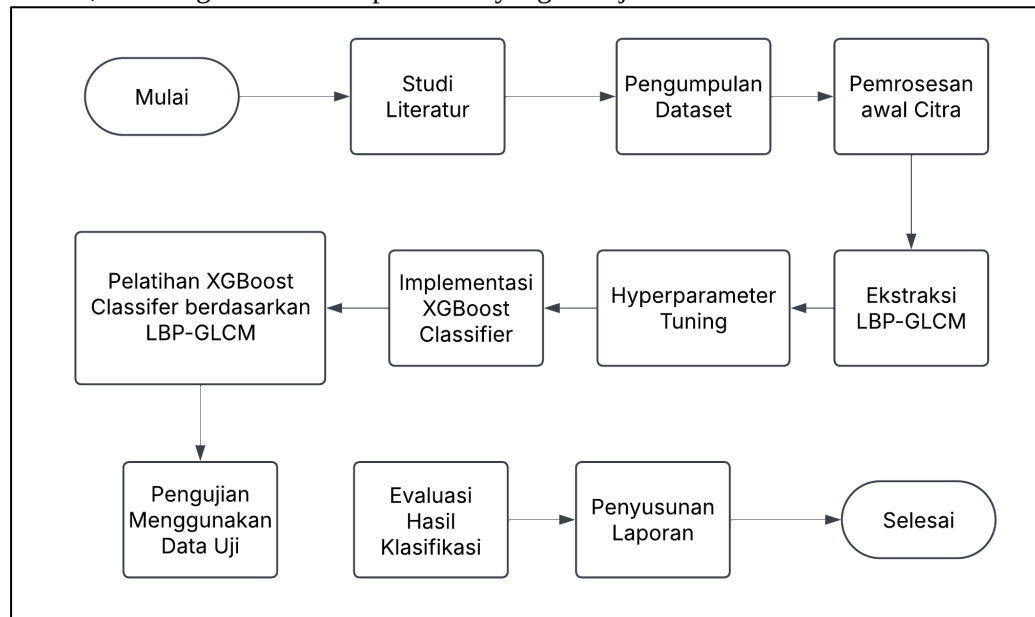
Beberapa penelitian sebelumnya telah memanfaatkan berbagai metode untuk ekstraksi fitur dari citra X-ray guna mendeteksi pneumonia. Misalnya, Mustika Rani (2019) mengaplikasikan Gray Level Co-occurrence Matrix (GLCM) untuk menggali hubungan spasial antar piksel dalam citra X-ray paru-paru. Metode ini terbukti memberikan hasil yang bermanfaat dalam klasifikasi, meskipun memiliki kelemahan berupa sensitivitas terhadap perubahan kecil dalam citra, seperti rotasi atau pergeseran posisi (Mustika Rani et al., 2024). Sebagai alternatif, Anzwar (2020) menyarankan penggunaan Local Binary Pattern (LBP), yang efektif dalam menggambarkan tekstur citra dengan ketahanan terhadap perubahan cahaya dan perhitungan yang lebih sederhana (Anzwar, 2017). Meskipun demikian, GLCM dianggap lebih unggul dalam mengekstraksi informasi tekstur yang lebih mendalam dan terperinci (Ding, 2017).

Sementara itu, algoritma klasifikasi berbasis pembelajaran mesin seperti XGBoost telah menunjukkan hasil yang sangat menjanjikan dalam klasifikasi penyakit berbasis citra medis. Penelitian oleh Chen & Guestrin (2016) mengungkapkan bahwa XGBoost dapat mencapai akurasi hingga 97,8% dalam klasifikasi penyakit paru menggunakan citra CT scan (Chen & Guestrin, 2016a). Di sisi lain, tinjauan sistematis oleh Baltruschat et al. (2019) menunjukkan bahwa meskipun banyak penelitian mengenai pneumonia berfokus pada deep learning, yang membutuhkan sumber daya komputasi besar, metode seperti XGBoost tetap sangat relevan karena lebih efisien dalam hal komputasi dan dapat bekerja dengan dataset yang lebih kecil (Baltruschat et al., 2019).

Penelitian ini bertujuan untuk mengembangkan sistem diagnosis pneumonia yang lebih cepat dan akurat dengan menggabungkan ekstraksi fitur berbasis LBP dan GLCM, yang kemudian akan diklasifikasikan menggunakan algoritma XGBoost. Pendekatan ini diharapkan dapat memberikan solusi alternatif yang lebih efisien dibandingkan dengan metode tradisional dalam mendiagnosis pneumonia melalui citra X-ray.

2. METODE PENELITIAN

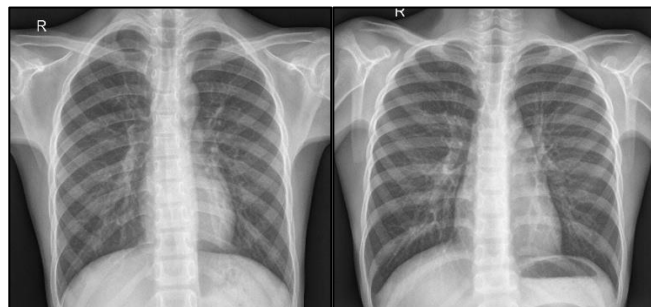
Dalam penelitian ini, kerangka penelitian diperlukan untuk menggambarkan penelitian yang dilakukan, berikut gambaran alur penelitian yang ditunjukkan oleh Gambar. 1



Gambar 1 Alur Penelitian

2.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diambil dari RSU DR. Wahidin Sudiro Husodo dengan total 5000 citra X-ray dada manusia. Data tersebut dibagi menjadi dua kelas, yaitu citra paru-paru yang terinfeksi pneumonia dan citra paru-paru yang normal (tanpa infeksi). Dengan pembagian kelas yang jelas ini, dataset ini memberikan dasar yang kuat untuk pengembangan sistem klasifikasi yang dapat membedakan antara paru-paru yang terinfeksi dan yang sehat, sehingga sangat relevan untuk penelitian terkait deteksi dini pneumonia berbasis citra X-ray



Gambar 2 Contoh Dataset

2.2. Preprocessing

Pemrosesan awal gambar dilakukan dengan tujuan untuk meningkatkan kualitas citra dan menonjolkan informasi yang terdapat di dalamnya, sehingga proses analisis selanjutnya dapat berjalan lebih efektif.

2.2.1. Pembagian data

Pada penelitian ini, dataset yang telah dikumpulkan dibagi menjadi tiga bagian, yaitu data latih (training set) sebesar 70% data validasi sebesar 15%, dan data uji sebesar 15%. Pembagian ini bertujuan untuk memastikan pelatihan model berlangsung optimal serta memperoleh evaluasi performa yang objektif terhadap data yang belum pernah dilihat oleh model.

2.2.3 Citra Grayscale

Citra grayscale adalah gambar yang hanya memiliki satu kanal warna dengan rentang intensitas dari hitam ke putih, berbeda dengan citra berwarna yang menggunakan tiga kanal (RGB). Pada citra grayscale, setiap piksel direpresentasikan dengan tingkat kecerahan yang memiliki nilai intensitas antara 0 (hitam) hingga 255 (putih) dalam representasi 8-bit (Susanto, 2019). Proses konversi citra berwarna menjadi grayscale biasanya dilakukan dengan menghitung rata-rata dari kanal warna atau menggunakan metode pembobotan yang mempertimbangkan sensitivitas mata manusia terhadap warna tertentu. Hal ini menyederhanakan citra dan memudahkan pemrosesan lebih lanjut, karena hanya fokus pada intensitas piksel tanpa mempertimbangkan informasi warna yang lebih kompleks.

2.2.4 Median Filtering

Median filtering digunakan untuk mengurangi noise pada citra X-ray dengan mengganti nilai setiap piksel dengan nilai median dari tetangganya dalam jendela filter. Teknik ini efektif menghilangkan outlier tanpa mengaburkan detail penting, sehingga struktur citra tetap terjaga (Zahra Salsa Azizah et al., 2024). Dalam konteks deteksi pneumonia, metode ini membantu meningkatkan kualitas citra X-ray dada, mempermudah identifikasi area yang terinfeksi, dan mendukung akurasi model dalam membedakan paru-paru sehat dan terinfeksi.

2.2.5 Adaptive Histogram Equalization

Langkah berikutnya dalam pemrosesan citra adalah ekualisasi histogram adaptif, yang bertujuan untuk meningkatkan kontras gambar. Teknik ini bekerja dengan menyesuaikan distribusi intensitas piksel secara lokal, memungkinkan detail yang penting dalam citra menjadi lebih jelas. Dengan meningkatkan kontras, elemen-elemen penting seperti perbedaan antara jaringan sehat dan terinfeksi menjadi lebih mudah dikenali oleh sistem analisis (Bhavadharini et al., 2021). Peningkatan kontras ini memudahkan proses analisis selanjutnya, sehingga model lebih efektif dalam mengenali pola dan elemen relevan dalam citra X-ray. Gabungan dari median filtering dan ekualisasi histogram adaptif memberikan citra yang lebih bersih dan kontras, yang sangat penting untuk meningkatkan akurasi deteksi pneumonia.

2.3. Ekstraksi Local Binary Pattern

Local Binary Pattern (LBP) adalah sebuah deskriptor tekstur yang digunakan untuk menandai piksel citra berdasarkan thresholding terhadap nilai intensitas piksel di sekitarnya. LBP memanfaatkan pola lokal yang terbentuk dari hubungan intensitas piksel-piksel sekitar dengan piksel pusat. Proses ini menghasilkan representasi biner yang menggambarkan pola tekstur citra, dengan tujuan untuk mengekstraksi informasi penting yang dapat digunakan dalam analisis citra lebih lanjut (Md Abdur Rahim et al., 2013).

Pada dasarnya, LBP beroperasi dengan mengubah nilai intensitas piksel pada citra menjadi biner berdasarkan perbandingan antara nilai piksel tetangga dan nilai piksel pusat. Nilai biner ini kemudian disusun dalam urutan melingkar, yang dapat diubah menjadi nilai desimal untuk mewakili pola tekstur lokal. LBP ini didefinisikan dengan persamaan matematika berikut:

$$LBP_{P,R}(X,Y) = \sum_{p=0}^{P-1} S(g_p - g_c) \times 2^p \quad (1)$$

Keterangan:

- $LBP_{p,R}(X,Y)$ adalah nilai LBP pada koordinat (X,Y) dengan jari-jari R jumlah tetangga P
- g_p adalah nilai intensitas piksel pada tetangga ke -p
- g_c adalah nilai intensitas piksel pada piksel pusat

Sedangkan $s(x)$ adalah fungsi ambang batas yang dinyatakan dengan definisi berikut:

$$S(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (2)$$

Fungsi ini digunakan untuk menentukan nilai biner dari setiap bit dalam representasi LBP. Nilai biner ini kemudian digabungkan untuk membentuk string biner yang menggambarkan pola tekstur lokal citra. P adalah jumlah tetangga yang mengelilingi piksel pusat dan RRR adalah jarak radial yang mengindikasikan seberapa jauh tetangga berada dari piksel pusat dalam suatu lingkaran dengan radius RRR.

LBP dapat dianggap seragam jika mengandung paling banyak dua perubahan bit antara tetangga-tetangga tersebut. Dengan kata lain, sebuah LBP seragam adalah pola yang hanya mengandung sedikit variasi dalam nilai intensitasnya, yang sangat berguna dalam pengenalan pola dan klasifikasi tekstur citra.

Proses ini memberikan representasi yang kuat dan efisien dari tekstur citra, yang dapat digunakan dalam berbagai aplikasi seperti pengenalan wajah, klasifikasi citra medis, dan deteksi objek. Dengan mengkombinasikan LBP dengan metode lain seperti Gray Level Co-occurrence Matrix (GLCM), kita dapat mengekstraksi fitur tekstur yang lebih mendalam dan meningkatkan akurasi analisis citra.

2.4. Gray Level Co-occurrence Matrix

Ekstraksi fitur menggunakan Gray Level Co-occurrence Matrix (GLCM) merupakan langkah lanjutan dari ekstraksi fitur dengan menggunakan metode Local Binary Pattern (LBP). GLCM bertujuan untuk mengekstraksi informasi lebih lanjut mengenai tekstur citra yang telah dihasilkan melalui LBP. Proses ini dimulai dengan menentukan arah dan jarak yang digunakan untuk menghitung pasangan piksel dalam citra. Dalam penelitian ini, jarak (d) yang digunakan adalah 1 piksel, dan orientasi sudut (θ) adalah 0 derajat. Selanjutnya, dilakukan perhitungan jumlah pasangan piksel yang memiliki intensitas tertentu dalam arah dan jarak yang telah ditentukan (Yunianto et al., 2021).

Matriks GLCM yang terbentuk kemudian dinormalisasi dengan membagi setiap elemen matriks dengan jumlah total elemen dalam matriks. Normalisasi ini bertujuan untuk mengurangi ketergantungan pada citra dan menghasilkan probabilitas untuk setiap pasangan piksel. Beberapa fitur tekstur penting yang diekstrak dari GLCM meliputi:

1. Kontras: Mengukur perbedaan intensitas antara piksel dalam citra. Semakin besar kontras, semakin besar variasi intensitas antara piksel yang berdekatan.

$$Contrast = \sum_{i,j} |i - j|^2 P(i,j) \quad (3)$$

2. Korelasi: Mengukur hubungan linier antara derajat intensitas citra grayscale. Korelasi yang mendekati 1 menunjukkan keterkaitan kuat antara piksel, sedangkan nilai negatif mengindikasikan ketidakhubungan.

$$Korelasi = \frac{\sum_{i,j} (i - \mu_i)(j - \mu_j)P(i,j)}{\sigma_i \sigma_j} \quad (4)$$

3. Energi: Mengukur keberagaman intensitas citra. Energi yang lebih tinggi menunjukkan bahwa citra memiliki pola yang lebih seragam.

$$Energy = \sum_{i,j} P(i,j)^2 \quad (5)$$

4. Homogenitas: Mengukur kesamaan intensitas antara piksel-piksel dalam citra. Semakin seragam intensitasnya, semakin tinggi nilai homogenitas.

$$Homogeneity = \sum_i \sum_j \frac{P(i,j)}{1 + |i,j|} \quad (6)$$

2.5. Extrem Gradient Boosting

Setelah melakukan ekstraksi fitur, langkah selanjutnya yaitu hasil ekstraksi fitur menjadi inputan untuk melakukan klasifikasi. XGBoost (Extreme Gradient Boosting) adalah metode machine learning yang digunakan untuk prediksi dan klasifikasi dengan memanfaatkan struktur pohon keputusan (Chen & Guestrin, 2016b). Berbeda dengan pohon keputusan konvensional, XGBoost menggunakan gradient boosting, di mana setiap pohon baru memperbaiki kesalahan pohon sebelumnya. Teknik ini mengoptimalkan algoritma GBDT, memungkinkan komputasi paralel, serta efisien dalam menangani data sparse dan penggunaan CPU/memori (Suparyogo, 2023).

XGBoost menambahkan regularisasi pada fungsi objektif untuk mengurangi kompleksitas model dan menghindari overfitting, menggunakan L1 dan L2 regularisasi. Regularisasi ini membantu membuat model lebih sederhana dan meningkatkan kemampuannya dalam menggeneralisasi data. Dengan demikian, XGBoost sangat efektif dalam menangani masalah klasifikasi dan regresi, terutama dalam dataset yang besar dan kompleks. Berdasarkan GBDT, XGBoost menambahkan istilah regularisasi pada fungsi objektif untuk mengurangi kompleksitas model dan menghindari overfitting, seperti yang diberikan pada Persamaan (7) dan Persamaan (8),

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \Omega(f_k) \quad (7)$$

$$\text{dimana } \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

Dimana $l(y_i, \hat{y}_i)$ mempresentasikan loss function yang menghitung selisih atau kesalahan antara nilai aktual (y_i), dan nilai prediksi ($\hat{y}_i$). Kemudian $\Omega(f_k)$ adalah fungsi regulasi yang mempunyai tugas untuk mengontrol kompleksitas model guna mencegah overfitting. T adalah jumlah daun dalam pohon keputusan f_k , w_j adalah skor output pada daun ke- j , γ adalah parameter penalti untuk jumlah daun, dan λ adalah koefisien regulasi L2. Komponen ini berfungsi untuk membatasi ukuran dan kekuatan setiap pohon, sehingga mendorong model yang lebih sederhana dan generalis.

Untuk mengoptimalkan fungsi objektif, XGBoost menggunakan pendekatan Taylor deret orde dua. Oleh karena itu, diperlukan perhitungan turunan pertama (gradient) dan turunan kedua (hessian) dari fungsi loss:

$$g_i = \hat{y}_i - y_i, \quad h_i = \hat{y}_i(1 - \hat{y}_i) \quad (9)$$

Saat membangun pohon baru, XGBoost mengevaluasi keuntungan dari setiap split menggunakan rumus Gain, yaitu:

$$GAIN = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \quad (10)$$

Jika nilai gain dari pemisahan kurang dari nilai tetap yang telah ditentukan, atau jumlah waktu pembagian mencapai kedalaman maksimum yang telah ditetapkan, maka proses pembagian, maka proses pembagian akan berhenti dan model klasifikasi akhir akan diperoleh. Berikut adalah parameter dari XGBoost

Nama	Deskripsi	Nilai Default
------	-----------	---------------

n_estimators	Jumlah pohon (trees) yang digunakan dalam model XGBoost. Semakin banyak pohon yang digunakan, model akan memiliki kapasitas yang lebih besar untuk mempelajari pola dalam data, tetapi berisiko menyebabkan overfitting jika terlalu banyak.	100
learning_rate	Laju pembelajaran (learning rate) mengontrol seberapa besar perubahan pada model di setiap iterasi. Learning rate yang rendah memperlambat proses pelatihan, memberikan kesempatan bagi model untuk belajar lebih rinci, tetapi membutuhkan lebih banyak pohon untuk konvergen.	0.1
max_depth	Kedalaman maksimum pohon keputusan dalam model. Menentukan seberapa banyak lapisan dalam pohon yang akan dibangun. Kedalaman yang terlalu besar bisa menyebabkan model menjadi terlalu kompleks dan overfitting.	3
min_child_weight	Menentukan jumlah sampel minimum yang diperlukan untuk membagi sebuah simpul. Semakin tinggi nilai ini, model akan lebih konservatif dalam membangun pohon, mengurangi kompleksitas model.	1.0
lambda	Parameter regularisasi L2 yang digunakan untuk mengurangi overfitting dengan menambahkan penalti terhadap bobot yang besar dalam pohon keputusan.	0.1
early_stopping_rounds	Menentukan jumlah iterasi berturut-turut yang model akan jalani tanpa perbaikan pada loss validasi sebelum proses pelatihan dihentikan. Teknik ini membantu menghindari overfitting dengan menghentikan pelatihan lebih awal jika tidak ada peningkatan yang signifikan.	20

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing

Sebelum melakukan ekstraksi fitur, ada beberapa tahapan yang harus dilakukan dalam pemrosesan awal citra. Dalam penelitian ini dilakukan beberapa langkah untuk meningkatkan kualitas citra dan memperjelas detail yang relevan. Tahapan pertama yaitu mengubah citra dalam skala keabuan, penerapan filter median, dan ekualisasi histogram adaptif. Hasil pemrosesan awal menampilkan citra array 3 dimensi gambar asli, matriks dari hasil grayscale, matriks dari median filtering, dan matriks hasil adaptive histogram equalization.

```

Original Image Matrix:
[[[ 34 34 34]
  [ 33 33 33]
  [ 33 33 33]
  ...
  [ 23 23 23]
  [ 18 18 18]
  [ 14 14 14]]

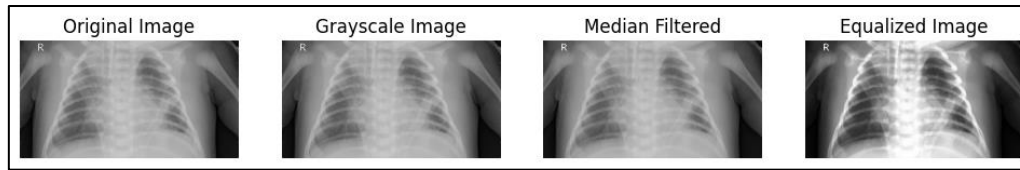
Grayscale Image Matrix:
[[[ 34 33 33 ... 23 18 14]
  [ 33 33 32 ... 24 20 17]
  [ 32 32 32 ... 25 23 21]
  ...
  [ 36 21 27 ... 24 20 36]
  [ 17 0 46 ... 46 0 13]
  [255 247 247 ... 249 245 255]]

Median Filtered Image Matrix:
[[[ 33 33 33 ... 23 21 18]
  [ 33 33 33 ... 24 23 21]
  [ 32 33 33 ... 24 24 23]
  ...
  [ 25 25 24 ... 22 24 24]
  [ 36 31 24 ... 22 33 36]
  [247 46 31 ... 33 46 245]]

Equalized Image Matrix:
[[[ 14 14 14 ... 12 12 10]
  [ 14 14 14 ... 13 12 12]
  [ 14 14 14 ... 13 13 12]
  ...
  [ 13 13 13 ... 12 13 13]
  [ 15 13 13 ... 12 14 15]
  [255 19 13 ... 14 19 255]]

```

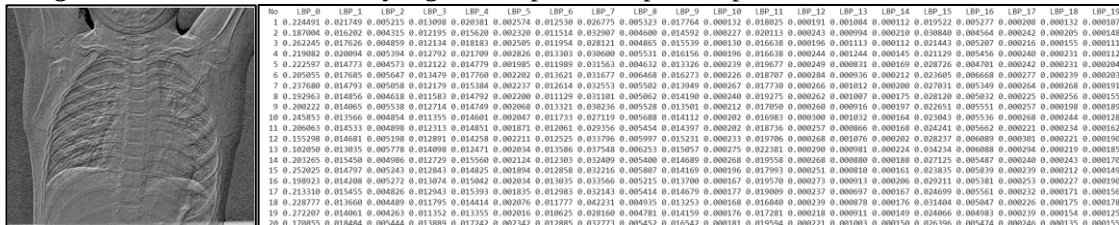
Gambar 3 Matrix Saat Preprocessing



Gambar 4 Citra Saat Preprocessing

3.2 Ekstraksi Fitur LBP

Setelah dilakukan preprocessing kemudian dilanjutkan dengan ekstraksi fitur menggunakan *Local Binary Pattern*. Tahap ini bertujuan untuk merepresentasikan tekstur dari citra x-ray. Langkah pertama adalah memproses citra yang sudah dilakukan *preprocessing* dan memisahkan citra tersebut menjadi region matrix berukuran 3x3. Selanjutnya, setiap nilai piksel dalam region matrix dibandingkan dengan nilai threshold atau nilai tengah dari matrix. Jika nilai piksel lebih besar daripada threshold, nilai tersebut diubah menjadi 1; jika tidak, nilai menjadi 0. Nilai biner yang diperoleh dari setiap region matrix kemudian digabungkan menjadi string biner. Proses ini diulang untuk setiap region matrix, dan string biner dari semua region digabungkan untuk membentuk representasi keseluruhan tekstur citra. Akhirnya, hasil ekstraksi ini menghasilkan tekstur citra LBP yang selesai pada tahap akhir proses.



Gambar 5 Tekstur Hasil Ekstraksi Local Binary Pattern

3.3 Ekstraksi Fitur GLCM

Tahapan selanjutnya yaitu melakukan ekstraksi fitur GLCM dengan menentukan arah dan jarak pada matrix GLCM. Setelah itu, matrix GLCM dibentuk untuk mencatat hubungan spasial antar-piksel dengan jarak dan arah yang telah ditentukan. Setelah matrix GLCM terbentuk, langkah berikutnya adalah menghitung jumlah piksel berpasangan yang muncul dalam matrix tersebut. Proses normalisasi dilakukan pada matrix GLCM untuk memastikan nilai-nilai dalam matrix berada pada rentang yang sesuai. Selanjutnya, dihitunglah fitur-fitur statistik dari matrix GLCM, yang meliputi kontras, korelasi, energi, dan homogenitas, yang berfungsi untuk menggambarkan karakteristik tekstur citra. Hasil dari proses ini berupa fitur-fitur GLCM yang siap digunakan dalam klasifikasi menggunakan algoritma yang telah ditentukan.

Tabel 1 Output Ektstraksi Fitur Gray Level Co-Occurance Matrix

No	Contrast	Correlation	Energy	Homogeneity	Label
1	6853.648087	1.168617799	0.014284154	0.25818641	0
2	5830.383922	1.26248557	0.019073325	0.312054679	0
3	6152.123805	1.327160894	0.034019556	0.321787844	0
4	6735.132012	1.19901674	0.014926128	0.26304887	0
5	5669.271009	1.324356477	0.029389435	0.339318119	0
6	6728.690961	1.218135723	0.018727495	0.272313361	0
7	5999.029645	1.325332318	0.032941644	0.327690148	0
8	5981.111271	1.26841916	0.022882572	0.317214592	0
...					...

617	5284.677002	1.369224009	0.024510701	0.346081495	1
618	5511.940711	1.319561834	0.027387824	0.350526205	1
619	5754.994203	1.284928702	0.021177831	0.327305641	1
620	5421.767044	1.395555179	0.037659216	0.370073881	1
621	5511.85103	1.364977876	0.031043753	0.355515121	1
622	5167.848646	1.485066672	0.058725917	0.409519318	1
623	5819.97267	1.244677108	0.020225275	0.327103022	1

3.4 Implementasi XGBOOST

Proses dimulai dengan inputan hasil dari ekstraksi fitur LBP dan GLCM yang diparalelkan dengan tujuan untuk memperoleh informasi relevan yang akan digunakan dalam proses pelatihan model. Setelah fitur diekstraksi, tahap selanjutnya adalah inisialisasi model XGBoost, di mana model ini dikonfigurasi dengan parameter awal yang akan diuji dan disesuaikan lebih lanjut. Salah satu tahap penting adalah tuning hyperparameter, di mana nilai optimal dari hyperparameter seperti jumlah pohon (n-estimators) dan kedalaman pohon (max_depth) dicari untuk meningkatkan performa model. Setelah hyperparameter tuning dilakukan, hasil terbaik dari tuning ini diimplementasikan untuk memperbaiki parameter model. Untuk kombinasi hyperparameter yang digunakan sebagai berikut:

Tabel 2 Kombinasi Hyperparameter

Nama	Nilai yang dicoba
n_estimators	100, 150
learning_rate	0.01, 0.1
max_depth	3, 5
min_child_weight	3.0, 5.0
lambda	0.2, 0.5

Selanjutnya, model yang telah disesuaikan digunakan untuk melakukan prediksi probabilitas pada data uji. Pada titik ini, perlu dievaluasi apakah jumlah pohon dalam model telah mencapai batas yang ditentukan, baik untuk jumlah estimators maupun kedalaman pohon. Jika belum, model akan terus diperbaiki dengan menghitung gradient dan Hessian, yang digunakan untuk memperbaiki kesalahan prediksi pada setiap iterasi. Berdasarkan perhitungan ini, model membangun pohon keputusan baru untuk meningkatkan akurasi prediksi. Setelah pohon keputusan baru dibangun, model akan mengupdate prediksi berdasarkan pembaruan yang terjadi. Hasil dari pelatihan dan pengujian model menggunakan random search didapat kombinasi parameter terbaik:

Tabel 3 Hasil Parameter Terbaik

No	F1 Score	Accuracy	Parameter	Nilai
1	97.68%	96.75%	n_estimator	100
			learning_rate	0.1
			max_depth	3
			min_child_weight	5.0
			lambda	0.2

Setelah meneumkan parameter yang sesuai kemudian dilakukan prediksi terhadap dataset serta dilakukan perhitungan metrik evaluasi yang berisi Akurasi, Presisi, Recall, dan F1-Score.

Tabel 4 Hasil Evaluasi

Metode	Akurasi	Presisi	Recall	F1-Score
LBP-GLCM dan XGBOOST	96.11%	95.02%	99.53%	97.22%

Tabel Menunjukkan hasil eksperimen kombinasi model ekstraksi fitur, dan metode Gradient Boosting. Metode evaluasi model menggunakan teknik split data 70:15:15. Berdasarkan dari hasil eksperimen, menunjukkan kombinasi antara model Local Binary Pattern, Gray Level Co-occurrence Matrix dan XGBoost mampu memberikan performa terbaik yaitu dengan nilai akurasi sebesar 96.11%, presisi sebesar 95.02%, recall sebesar 99.34% dan F1-Score sebesar 96.95%.

Penelitian ini mengimplementasikan kombinasi teknik ekstraksi fitur Local Binary Pattern (LBP) dan Gray Level Co-occurrence Matrix (GLCM) yang diproses menggunakan model XGBoost untuk klasifikasi pneumonia pada citra X-ray. Berdasarkan hasil eksperimen, model yang diajukan mencapai performa yang baik, dengan nilai akurasi sebesar 96.11%, presisi 95.02%, recall 99.53%, dan F1-Score 97.22%. Sejumlah penelitian sebelumnya telah menggunakan CNN untuk klasifikasi pneumonia pada citra X-ray, yang menunjukkan hasil yang sangat baik. Sebagai contoh, Rajpurkar et al. (2017) menggunakan CNN untuk klasifikasi pneumonia dan memperoleh akurasi lebih dari 95%, yang serupa dengan hasil yang dicapai oleh model dalam penelitian ini. Meskipun demikian, terdapat perbedaan signifikan antara metode yang digunakan dalam penelitian ini dan CNN terkait dengan kompleksitas model dan kebutuhan sumber daya komputasi.

Meskipun CNN sangat efektif dalam klasifikasi citra medis, cenderung memerlukan data yang lebih besar untuk pelatihan serta waktu komputasi yang lebih lama. Model CNN juga sangat bergantung pada perangkat keras yang kuat, seperti GPU, untuk melakukan pelatihan dan optimasi model secara efisien. Sebaliknya, pendekatan yang digunakan dalam penelitian ini, yang menggabungkan ekstraksi fitur LBP dan GLCM dengan XGBoost, menunjukkan efisiensi yang lebih tinggi, terutama dalam hal waktu pelatihan dan penggunaan sumber daya. Pendekatan ini mampu menghasilkan model yang baik meskipun hanya menggunakan data citra yang terbatas, serta mengurangi kebutuhan akan perangkat keras yang memadai.

Selain itu, XGBoost sebagai model boosting yang efisien memiliki kemampuan untuk mengatasi masalah overfitting lebih baik dibandingkan CNN, terutama ketika digunakan pada dataset dengan ukuran terbatas. Dengan mengandalkan fitur tekstur yang diekstraksi melalui LBP dan GLCM, model ini dapat menghasilkan klasifikasi yang akurat tanpa perlu mengandalkan jaringan saraf yang besar dan kompleks, sebagaimana yang diperlukan oleh CNN untuk mempelajari fitur dari citra mentah.

Berdasarkan hasil eksperimen, penggunaan kombinasi teknik ekstraksi fitur LBP-GLCM dan model XGBoost terbukti memberikan hasil yang sangat baik, dengan akurasi mencapai 96.11%, meskipun tidak membutuhkan sumber daya komputasi yang tinggi. Pendekatan ini memiliki keuntungan signifikan dalam hal efisiensi, terutama pada aplikasi yang mengutamakan kecepatan pemrosesan dan penggunaan sumber daya yang terbatas. Dengan demikian, meskipun model CNN memberikan performa yang sangat baik, pendekatan yang diajukan dalam penelitian ini lebih efisien dalam hal waktu pelatihan dan penggunaan sumber daya, menjadikannya pilihan yang lebih praktis untuk aplikasi klasifikasi pneumonia berbasis citra X-ray, terutama dalam konteks lingkungan dengan sumber daya komputasi yang terbatas.

4 KESIMPULAN

- 1) Penelitian ini berhasil mengembangkan sebuah sistem klasifikasi pneumonia berbasis citra X-ray dengan menggabungkan dua teknik ekstraksi fitur, yaitu Local Binary Pattern (LBP) dan Gray Level Co-occurrence Matrix (GLCM), yang kemudian diklasifikasikan menggunakan algoritma XGBoost. Penggunaan dua teknik ekstraksi ini memungkinkan sistem untuk menangkap informasi tekstur citra yang lebih mendalam, sehingga meningkatkan akurasi dalam klasifikasi.
- 2) Hasil eksperimen menunjukkan bahwa model yang dikembangkan memberikan performa yang sangat baik, dengan nilai akurasi 96.11%, presisi 95.02%, recall 99.34% dan F1-score 96.95%. Nilai-nilai ini menegaskan bahwa model mampu mengklasifikasikan citra X-ray pneumonia dengan tingkat keandalan yang tinggi, serta memberikan keseimbangan yang baik antara deteksi positif dan negatif.
- 3) Proses pengoptimalan hyperparameter menggunakan teknik random search, dengan mencari kombinasi terbaik dari parameter seperti `n_estimators`, `learning_rate`, dan `max_depth`, terbukti berkontribusi signifikan terhadap peningkatan performa model. Pengaturan hyperparameter yang tepat memungkinkan model untuk belajar dengan lebih efektif, meningkatkan akurasi prediksi, dan mengurangi overfitting.
- 4) Kombinasi teknik ekstraksi fitur berbasis tekstur (LBP dan GLCM) dan algoritma XGBoost menunjukkan potensi besar dalam meningkatkan akurasi dan efisiensi sistem dalam mendeteksi pneumonia melalui citra X-ray. Metode ini menawarkan alternatif yang lebih efisien dibandingkan dengan teknik deep learning yang memerlukan sumber daya komputasi besar, tanpa mengorbankan kualitas prediksi.

5 SARAN

Penggunaan dataset yang lebih besar dan beragam dapat meningkatkan generalisasi model, mengurangi bias, dan meningkatkan akurasi. Integrasi dengan teknik deep learning seperti Convolutional Neural Networks (CNN) dapat memperbaiki deteksi pneumonia yang lebih kompleks. Pengoptimalan hyperparameter dengan teknik seperti Bayesian optimization atau Grid search juga dapat meningkatkan kinerja model. Pengujian pada platform medis real-time akan menguji keandalan dan efektivitas sistem dalam situasi praktis. Dengan saran-saran ini, sistem dapat lebih ditingkatkan dan diterapkan dalam deteksi pneumonia.

DAFTAR PUSTAKA

- Anzwar. (2017). *Integrasi Ekstraksi Fitur Local Binary Pattern Dan Gray-Level Cooccurrence Metrix Untuk Pengenalan Ekspresi Mulut Pembelajar*. 9, 17–24.
- Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., & Saalbach, A. (2019). Comparison Of Deep Learning Approaches For Multi-Label Chest X-Ray Classification. *Scientific Reports*, 9(1). <https://doi.org/10.1038/S41598-019-42294-8>
- Bhavadharini, K., Ruhan Bevi, A., & Hari, U. (2021). Performance Investigation Of Digital Filters For Image Enhancement. *Journal Of Physics: Conference Series*, 1964(6). <https://doi.org/10.1088/1742-6596/1964/6/062077>
- Chen, T., & Guestrin, C. (2016a). Xgboost: A Scalable Tree Boosting System. *Proceedings Of The Acm Sigkdd International Conference On Knowledge Discovery And Data Mining*, 13-17-August-2016, 785–794. <https://doi.org/10.1145/2939672.2939785>

- Chen, T., & Guestrin, C. (2016b). Xgboost: A Scalable Tree Boosting System. *Proceedings Of The Acm Sigkdd International Conference On Knowledge Discovery And Data Mining*, 13-17-August-2016, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Ding, X. (2017). *Texture Feature Extraction Research Based On Glcm-Clbp Algorithm*.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A. W. M., Van Ginneken, B., & Sánchez, C. I. (2017). *A Survey On Deep Learning In Medical Image Analysis*. <https://doi.org/10.1016/J.Media.2017.07.005>
- Md Abdur Rahim, B., Shafiul Azam, M., Hossain, N., Rashedul Islam, M., Abdur Rahim A, M., Shafiul Azam S, M., & Rashedul Islam G, M. (2013). Face Recognition Using Local Binary Patterns (Lbp). *Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc*, 13.
- Mustika Rani, D., Deswardani, F., & Fendriani, Y. (2024). Classification Of Lung Disease On X-Ray Images Based On Gray Level Co-Occurrence Matrix (Glcm) Feature Extraction And Backpropagation Neural Network Using Python Gui. *Jop*, 9(2), 80–86. <https://www.kaggle.com>
- Pneumonia In Children*. (N.D.). Retrieved May 16, 2025, From <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- Pneumonia Prevention And Control | Pneumonia | Cdc*. (N.D.). Retrieved May 16, 2025, From <https://www.cdc.gov/pneumonia/prevention/index.html>
- Suparyogo, R. (2023). *Analisis Boosting Algorithm Dan Hog Dalam*.
- Susanto, A. (2019). Penerapan Operasi Morfologi Matematika Citra Digital Untuk Ekstraksi Area Plat Nomor Kendaraan Bermotor. In *Jurnal Pseudocode* (Vol. 1). www.ejournal.unib.ac.id/index.php/pseudocode
- Yunianto, M., Anwar, F., Nur Septianingsih, D., Dwi Ardyanto, T., & Farits Pradana, R. (2021). Klasifikasi Kanker Paru Paru Menggunakan Naïve Bayes Dengan Variasi Filter Dan Ekstraksi Ciri Gray Level Co-Occurance Matrix (Glcm). *Indonesian Journal Of Applied Physics*, 11(2).
- Zahra Salsa Azizah, A., Nazar To Yalis, F., Shiva Narayana, I., Azelia Syalom, K., & Rosyani, P. (2024). Pengolahan Citra Digital Dengan Penerapan Teknik Ambang Batas: Studi Kasus Menggunakan Opencv. In *Jurnal Artificial Intelligent Dan Sistem Penunjang Keputusan* (Vol. 1, Issue 4). <https://jurnalmahasiswa.com/index.php/aidanspk>