

BAB I PENDAHULUAN

1.1 Latar Belakang

Di tengah upaya global untuk meningkatkan kesehatan masyarakat, *pneumonia* tetap menjadi salah satu penyakit infeksi yang memiliki dampak signifikan terhadap angka kematian, khususnya di kelompok usia rentan seperti balita. Pada tahun 2010 diperkirakan 120 juta kasus *pneumonia* di seluruh dunia, dan pada tahun 2011 terdapat 1,3 juta kasus kematian pada anak-anak yang disebabkan oleh penyakit *pneumonia*. Orang dewasa yang memiliki masalah kesehatan sebelumnya seperti penyakit jantung, paru-paru dan *Human Immunodeficiency Virus* (HIV) juga mudah terkena penyakit *pneumonia* (Azmi. et al., 2016).

Pada tahun 2020 di Indonesia, tercatat sebanyak 4.972.553 kunjungan balita dengan keluhan batuk atau kesulitan bernapas, yang kemudian didiagnosis sebagai *pneumonia* pada balita. Penyakit ini disebabkan oleh berbagai jenis mikroorganisme, seperti virus, bakteri, dan jamur, yang menyerang sistem pernapasan. Gejala yang umum dialami oleh penderita *pneumonia* meliputi batuk, kesulitan bernapas seperti napas cepat (Kementerian Kesehatan RI, 2020). Gejala lainnya dapat meliputi nyeri dada, kelelahan, demam, mual, muntah dan diare. *Pneumonia* dapat menular ke orang lain melalui berbagai cara, tergantung pada sumber infeksi. *Pneumonia* dapat diperoleh dari masyarakat (*Community-acquired Pneumonia*), rumah sakit (*Hospital-acquired Pneumonia*), fasilitas kesehatan lainnya (*Healthcare-acquired Pneumonia*), maupun melalui aspirasi. *Pneumonia* aspirasi terjadi ketika makanan, minuman, muntahan, atau air liur yang mengandung bakteri terhirup ke dalam paru-paru, sehingga menyebabkan infeksi (Zaeri. et al., 2024).

Dalam upaya mendukung diagnosis yang akurat dan penanganan yang tepat, salah satu metode diagnostik yang banyak digunakan adalah citra *rontgen* paru-paru. Citra *rontgen* paru-paru merupakan salah satu metode diagnostik yang efektif dan banyak digunakan dalam mendeteksi *pneumonia*. Teknologi ini memungkinkan visualisasi yang jelas terhadap struktur paru-paru, sehingga memudahkan identifikasi adanya penyebaran bakteri, infeksi dan jamur (Sharma. et al., 2023). Pencitraan sinar-x menawarkan keuntungan berupa biaya yang lebih terjangkau dibandingkan dengan metode pencitraan lainnya. Hal ini menjadikan sinar-x sebagai opsi yang lebih mudah diakses, terutama di wilayah dengan sumber daya terbatas atau daerah terpencil, di mana peralatan pencitraan canggih sulit untuk tersedia (Khattab. et al., 2024).



an kemajuan teknologi, pemanfaatan pemrograman komputer ng memungkinkan pengolahan data dan citra digital secara tisasi. GNU Octave merupakan salah satu perangkat lunak memiliki kemampuan kuat dalam perhitungan numerik dan hingga sangat mendukung proses ekstraksi fitur dari data atis dan terstruktur. GNU Octave diluncurkan oleh Richard

Stallman pada tahun 1983 dibawah organisai nirlaba *Free Software Foundation* (FSF) (Sanggala. et al., 2022). Penelitian ini berfokus pada pengembangan metode klasifikasi menggunakan ekstraksi fitur tekstur *Gray-Level Co-occurrence Matrix* (GLCM) dan algoritma pengklasifikasi *Naïve Bayes*. GLCM dipilih karena mampu menangkap karakteristik tekstur citra secara detail yang penting dalam membedakan kondisi paru-paru normal dan terinfeksi *pneumonia* (Ariansyah. et al., 2024). Ukuran yang digunakan dalam GLCM untuk ekstraksi fitur agar seragam adalah 224×224 piksel karena ukuran ini merupakan standar umum dalam pemrosesan citra. Ukuran ini dinilai cukup untuk mempertahankan informasi penting dalam citra, termasuk tekstur, namun tetap efisien dalam proses komputasi dan pelatihan model (Nyasulu. et al., 2023). Sementara itu, *RapidMiner* adalah *platform* analisis data yang mengintegrasikan metode statistik, kecerdasan buatan, dan teknik data mining untuk mengolah dan menganalisis kumpulan data berukuran besar. Perusahaan yang mengembangkan *RapidMiner* adalah *RapidMiner, Inc.* Awalnya, perusahaan ini dikenal dengan nama *Rapid-I* sebelum melakukan rebranding pada tahun 2013 menjadi *RapidMiner*. Perusahaan ini didirikan oleh Dr. Ingo Mierswa dan Ralf Klinkenberg pada tahun 2006 di Jerman. *RapidMiner, Inc.* berkantor pusat di Boston, Massachusetts, dengan kantor tambahan di Jerman, Singapura, dan Inggris. Dengan sistem berbasis operator visual, *RapidMiner* memudahkan pengguna dalam membangun alur proses analisis data tanpa perlu menulis kode secara manual. Salah satu algoritma klasifikasi yang diterapkan di *RapidMiner* adalah *Naïve Bayes*, yang dikenal karena kesederhanaannya dalam perhitungan probabilistik serta kemampuannya dalam menghasilkan performa yang baik, terutama pada data berukuran kecil hingga menengah. Hal ini menjadikan *RapidMiner* sebagai alat yang efektif dalam mendukung proses klasifikasi dan evaluasi performa model *machine learning*, termasuk penerapan algoritma *Naïve Bayes* (Kirana. et al., 2024).

Berdasarkan uraian di atas, penelitian ini dilakukan untuk mendeteksi *pneumonia* dari citra *rontgen* yang diolah menggunakan GLCM dan pembelajaran mesin, penggunaan GNU *Octave* dan analisis data *RapidMiner*, membuat program pengolahan citra serta menganalisis keakuratan program yang telah dibuat.

1.2 Tujuan Penelitian

Tujuan penelitian ini adalah membuat program pengolahan citra untuk mendeteksi *pneumonia* dan menganalisis keakuratan program yang dibuat.

1.3 Manfaat Penelitian



Programan dan analisis data memungkinkan proses lebih cepat.

Keakuratan dalam mendeteksi *pneumonia* melalui analisis citra

menjadi alat bantu bagi dokter dalam mendiagnosis *pneumonia*, sehingga mereka dalam memberikan penanganan yang tepat

kepada pasien.

4. Pengembangan pengetahuan dalam bidang fisika medik dan teknologi kesehatan, serta dapat digunakan sebagai referensi untuk penelitian lebih lanjut di masa depan.



BAB II METODE PENELITIAN

2.1 Gray-Level Co-occurrence (GLCM)

Seiring dengan perkembangan teknologi, metode berbasis analisis citra semakin banyak diterapkan untuk mendukung diagnosis penyakit, termasuk *pneumonia*. Salah satu metode yang digunakan adalah *Gray-Level Co-occurrence Matrix* (GLCM) untuk mengekstraksi detail lokal dan fitur tekstur dalam gambar dengan menganalisis hubungan spasial serta karakteristik statistik antar piksel berdasarkan tingkat keabuan (Jiao. et al., 2024). Teknik ini menghasilkan matriks GLCM dan kemudian mengekstraksi parameter statistik dari matriks tersebut untuk mendeskripsikan tekstur gambar. Matriks ini berbentuk dua dimensi, setiap elemen (i, j) dalam GLCM akhir dihitung dengan menjumlahkan frekuensi kemunculan piksel pada gambar asli yang memiliki nilai tertentu dalam hubungan spasial tertentu dengan piksel bernilai tertentu (Elaraby. et al., 2024).

Dalam penelitian ini, GNU Octave dipilih sebagai alat bantu untuk proses pengolahan data citra, karena kemampuannya dalam melakukan visualisasi serta mendukung komputasi numerik secara efisien di berbagai sistem operasi (Purnama et al., 2021). Analisis tekstur pada citra dilakukan menggunakan metode GLCM, dengan memanfaatkan sejumlah fitur statistik. Fitur-fitur tersebut digunakan dalam proses ekstraksi informasi tekstur dan dijelaskan melalui beberapa persamaan sebagai berikut (Suharyana et al., 2023):

1. *Contrast* merupakan perbedaan intensitas antara piksel-piksel yang berdekatan dalam suatu gambar

$$C = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i - j)^2 P(i, j) \quad (1)$$

dimana C adalah *contrast*, i adalah kolom, j adalah baris, Ng adalah jumlah total level abu-abu, dan $P(i, j)$ adalah elemen dari matriks GLCM pada posisi (i, j) .

2. *Energy* merepresentasikan total nilai elemen yang telah dikuadratkan dalam matriks GLCM

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

dimana E adalah *energy*, i adalah kolom, j adalah baris, dan $P(i, j)$ adalah elemen dari matriks GLCM pada posisi (i, j) .

3. *Homogeneity* menunjukkan kesamaan nilai piksel yang berdekatan

$$H = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad (3)$$

dimana H adalah *homogeneity*, i adalah kolom, j adalah baris, dan $P(i, j)$ adalah elemen dari matriks GLCM pada posisi (i, j) .



ukur hubungan antara piksel dalam gambar

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

correlation, i adalah kolom, j adalah baris, $P(i, j)$ adalah elemen pada posisi (i, j) , μ_i adalah nilai rata-rata baris, μ_j adalah nilai rata-rata kolom, σ_i adalah standar deviasi dihitung berdasarkan baris, dan σ_j adalah standar deviasi dihitung berdasarkan kolom.

standar deviasi dihitung berdasarkan kolom.

5. *Dissimilarity* merupakan ketidaksamaan yang digunakan untuk menunjukkan perbedaan dalam gambar

$$D = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} |i - j| P(i, j) \quad (5)$$

dimana D adalah *dissimilarity*, Ng adalah jumlah total level abu-abu, i adalah kolom, j adalah baris, dan $P(i, j)$ adalah elemen dari matriks GLCM pada posisi (i, j) .

2.2 Naïve Bayes

RapidMiner adalah *platform* perangkat lunak yang canggih untuk ilmu data dan pembelajaran mesin. *Software* ini menawarkan berbagai alat untuk persiapan data, pembuatan model, evaluasi, dan penerapan. *RapidMiner* dipilih untuk analisis data karena kemudahan penggunaannya dalam membangun model pembelajaran mesin, visualisasi data, dan validasi hasil. Dirancang dengan kemudahan penggunaan (Nahjan. et al., 2023). *Naïve Bayes* adalah metode klasifikasi yang digunakan pada penelitian ini. *Naïve Bayes* merupakan salah satu penerapan dari teorema *Bayes* yang dikenal luas, dan sering dimanfaatkan dalam klasifikasi karena sifatnya yang sederhana dan efisien. Teorema *Bayes* sendiri menggabungkan konsep probabilitas dengan asumsi independensi antar variabel. *Naïve Bayes* bekerja dengan mengestimasi kemungkinan suatu kategori berdasarkan data yang tersedia. Keunggulan utama dari metode ini terletak pada kemudahannya dalam implementasi serta kecepatan prosesnya. (Ariansyah. et al., 2024). Rumus dasar yang digunakan yaitu (Damanik. et al., 2021):

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (6)$$

dengan C_i sebagai kelas (*pneumonia* atau paru-paru sehat), X sebagai vektor fitur citra, $P(C_i|X)$ adalah probabilitas bahwa gambar (dengan fitur X) termasuk ke dalam kelas C_i , dan $P(X|C_i)$ adalah probabilitas munculnya fitur X jika diketahui gambar berasal dari kelas C_i . Algoritma ini mengasumsikan independensi antar fitur, maka perhitungannya disederhanakan menjadi:

$$P(C_i|X) \propto P(C_i) \cdot \prod_{j=1}^n P(x_j|C_i) \quad (7)$$

dimana $P(x_j|C_i)$ adalah probabilitas bahwa fitur ke- j bernilai seperti pada x_j jika gambar berasal dari kelas C_i . Setelah probabilitas dihitung untuk masing-masing kelas, data diklasifikasikan ke dalam kelas dengan probabilitas tertinggi yang akan menentukan kelas baik *pneumonia* dan paru-paru sehat:

$$C = \arg \max P(C_i|X) \quad (8)$$

as hasil prediksi, *arg max* adalah operator yang menghasilkan berikan nilai maksimum dari fungsi yang dihitung, dan $P(C_i|X)$ bahwa gambar (dengan fitur X) termasuk ke dalam kelas C_i .

si

erforma model klasifikasi dilakukan menggunakan beberapa *accuracy*, *precision*, dan *recall*. Metrik-metrik ini digunakan



untuk menilai sejauh mana model mampu melakukan klasifikasi secara tepat terhadap data citra *rontgen* paru-paru, khususnya dalam membedakan antara kelas sehat dan *pneumonia* (Chandra. et al., 2021).

Tabel 1. Performance Vector

	True Paru-paru Sehat	True Pneumonia
Predicted True Paru-paru Sehat	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Predicted Pneumonia	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Keterangan:

Benar positif (*True Positive* – TP): jumlah data paru-paru sehat yang berhasil diklasifikasikan dengan benar,

Benar negatif (*True Negative* – TN): jumlah data *pneumonia* yang diklasifikasikan dengan benar,

Salah positif (*False Positive* – FP): jumlah data paru-paru sehat yang salah diklasifikasikan sebagai *pneumonia*,

Salah negatif (*False Negative* – FN): jumlah data *pneumonia* yang salah diklasifikasikan sebagai paru-paru sehat.

1. *Accuracy* merupakan jumlah prediksi yang benar terhadap total jumlah data yang diuji

$$Accuracy: \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (9)$$

2. *Precision* merupakan jumlah data yang benar-benar positif dan diprediksi positif (*True Positive*) terhadap seluruh data yang diprediksi positif (*True Positive* + *False Positive*)

$$Precision: \frac{TP}{TP+FP} \times 100\% \quad (10)$$

3. *Recall* merupakan jumlah data positif yang berhasil dikenali dengan benar (*True Positive*) terhadap total data aktual positif (*True Positive* + *False Negative*)

$$Recall: \frac{TP}{TP+FN} \times 100\% \quad (11)$$

2.4 Alat dan Bahan Penelitian

2.2.1 Alat Penelitian

Alat yang digunakan pada penelitian ini yaitu :

1. Labtop dengan spesifikasi :



eron N4000

DD 500 GB

)

2.2.2 Bahan Penelitian

Bahan yang digunakan pada penelitian ini yaitu :

1. Dataset Citra Rontgen

Berjumlah 90 data yang dibagi menjadi dua kelas, yaitu 45 data pasien sehat dan 45 data pasien *pneumonia*, dengan format JPEG.

2. Label Data

Informasi klasifikasi untuk setiap citra data pasien sehat dan data pasien *pneumonia*.

3. Sumber Data

Website Kaggle (<https://www.kaggle.com/>)

2.5 Prosedur Penelitian

Prosedur pada penelitian ini yaitu :

1. Persiapan dan Pengumpulan Data :

- Pengumpulan Dataset : Mengumpulkan citra *rontgen* paru-paru yang mencakup gambar paru-paru sehat dan yang terinfeksi *pneumonia* dari *website Kaggle*
- Memastikan data dalam format yang sesuai untuk analisis, seperti JPEG
- Melakukan *preprocessing* data :
 - Penyesuaian ukuran citra agar seragam (224x224 piksel)
 - *Levels* (abu-abu) 256
 - Jarak pasangan piksel 1

2. Ekstraksi Fitur dengan GLCM

- Mengubah citra menjadi *grayscale*
- Perhitungan GLCM dengan variasi sudut (0° , 45° , 90° , 135°).
- Fitur yang diambil mencakup *contrast*, *homogeneity*, *energy*, *correlation* dan *dissimilarity*
- Penggunaan GNU Octave untuk menghitung GLCM dan menghasilkan fitur.

3. Pemrosesan Data dengan RapidMiner

- Memasukkan dataset fitur yang telah diekstraksi ke dalam *RapidMiner* untuk proses analisis data.

4. Penerapan Metode Naïve Bayes

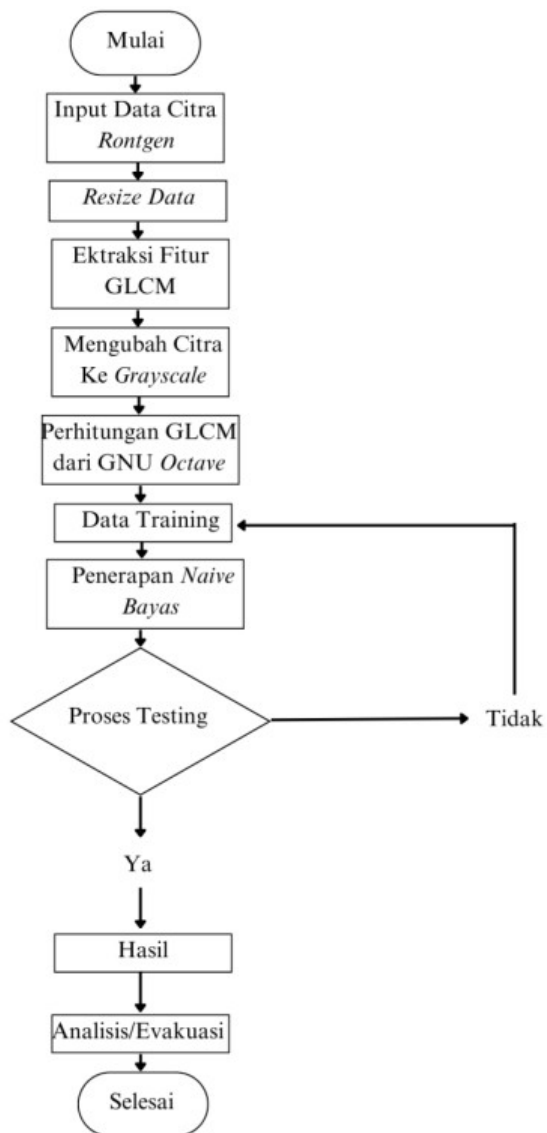
- Pemilihan metode *Naïve Bayes*.
- Pelatihan model menggunakan *RapidMiner* dengan dataset pelatihan.

5. Evaluasi Model

- Hasil evaluasi model dianalisis untuk mengidentifikasi tingkat akurasi deteksi *pneumonia*.



2.6 Bagan Alir Penelitian



Gambar 1. Bagan Alir Penelitian





Optimized using
trial version
www.balesio.com