BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini dilakukan dengan dataset berupa data citra rontgen dada yang memiliki dua kelas, yaitu paru-paru normal dan paru-paru terinfeksi TBC. Data tersebut diambil dari RS *Jogjakarta International Hospital* (JIH) Solo. Dalam dataset ini ditemukan ukuran citra yang berbeda-beda sehingga memerlukan tahapan pra proses data. Untuk menunjang penelitian, dataset yang digunakan sebanyak 406 data dengan sebaran data rontgen dada normal sebanyak 300 data dan rontgen dada terinfeksi TBC sebanyak 106 data. Lalu ditambahkan data yang diambil dari kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset/data berupa data rontgrn dada TBC sebanyak 139 data. Kebutuhan sistem yang digunakan pada penelitian ini berupa:

1. Kebutuhan Hardware berupa laptop Asus ROG Strix G15, RAM 8gb, Processor 6 core AMD Ryzen 5.
2. Kebutuhan Software meliputi: Sistem operasi Microsoft Windows 11, Google Colab menggunakan bahasa Python
3. Kebutuhan Data meliputi: data rontgen dada pasien RS JIS Solo yang terbagi menjadi dua kelas yaitu paru-paru normal dan paru-paru terinfeksi TBC yang discan menggunakan Digital Radiografi dengan merk GE

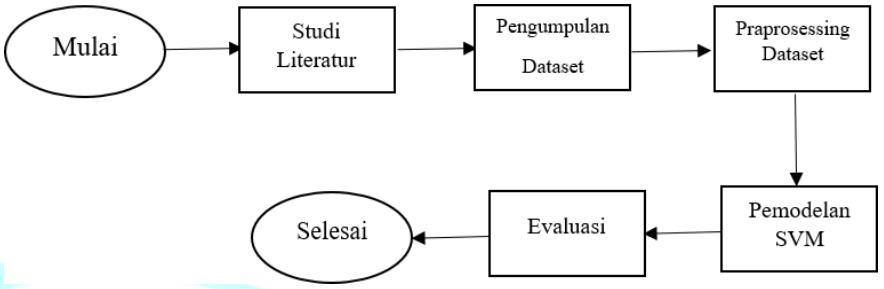
Penelitian ini dilaksanakan selama 6 bulan, dengan tahapan utama sebagai berikut:

Tabel 3. 1 Tabel Rincian Waktu Penelitian

No	Tahapan	November 2024				Desember 2024				Januari 2025				Febuari 2025				Maret 2025				April 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi Literatur																								
2	Pengumpulan Dataset																								
3	Pembuatan Sistem																								
4	Laporan Penelitian																								

3.2 Prosedur Penelitian

Prosedur pada penelitian ini memiliki alur seperti berikut:



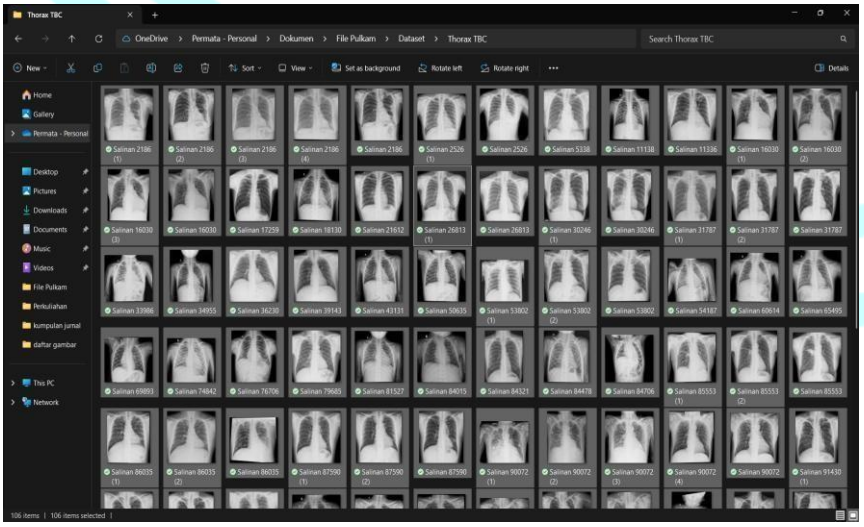
Gambar 3. 1 Flowchart Alur Prosedur Penelitian

3.2.1. Studi Literatur

Teori yang digunakan untuk melakukan penelitian ini berdasarkan dari literatur yang telah dikaji pada penelitian terkait. Meliputi sumber dari buku dan jurnal penelitian sebelumnya yang relevan. Refrensi dari studi literatur yang telah dipelajari dan digunakan dalam penelitian ini dilampirkan pada bagian akhir daftar refrensi.

3.2.2. Pengumpulan Dataset

Dataset yang dikumpulkan berupa data citra rontgen dada yang memiliki dua kelas, yaitu paru-paru normal dan paru-paru terinfeksi TBC. Data tersebut diambil dari RS JIH Solo. Untuk menunjang penelitian, dataset yang digunakan sebanyak 545 data dengan sebaran data rontgen dada normal sebanyak 300 data dan rontgen dada terinfeksi TBC sebanyak 106 data. Dan dari kaggle sebanyak 139 rontgen TBC.

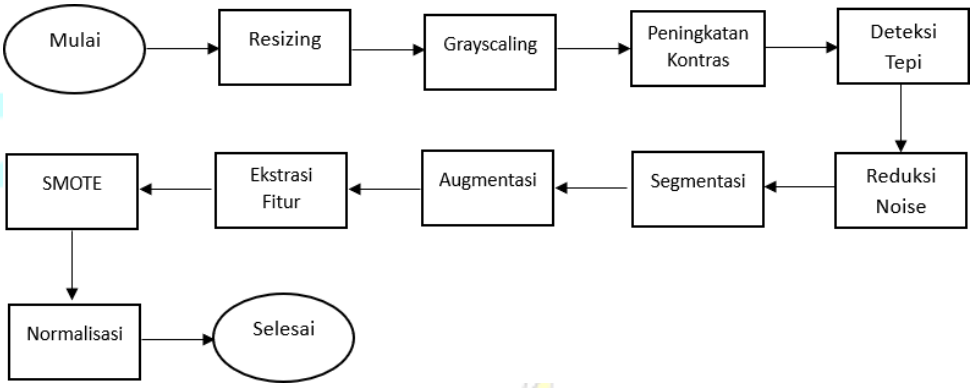


Gambar 3. 2 Kumpulan Dataset

3.2.3. Praprosessing Dataset

Praprosessing dataset bertujuan untuk memperbaiki citra sebelum dianalisa sehingga data mentah siap untuk diolah (Putri, 2023).

Tahapan-tahapan saat praprosessing dataset yang dilakukan pada penelitian ini berlandaskan pada penelitian-penelitian terkait sebelumnya:



Gambar 3. 3 Flowchart praprosessing dataset

Berikut adalah langkah-langkah proses prapengolahan dataset:

1. Proses *resizing* pada citra bertujuan untuk menyeragamkan ukuran resolusi seluruh citra agar memiliki dimensi yang sama. Resolusi citra akan diubah menjadi 256×256 piksel, sehingga setiap citra dapat direpresentasikan dalam bentuk matriks berukuran 256 baris dan 256 kolom
2. *Grayscaling* pada citra bertujuan untuk mengubah gambar menjadi keabuan dengan intensitas 0 hingga 255 (citra 8-bit)
3. Peningkatan kontras dilakukan untuk menonjolkan detail penting pada citra sehingga fitur menjadi lebih mudah dikenali. Langkah ini penting terutama untuk citra dengan kontras rendah yang membuat fitur sulit terlihat. Teknik umum yang digunakan adalah histogram equalization untuk meratakan distribusi intensitas piksel ke seluruh rentang (0-255) atau Contrast Limited Adaptive Histogram Equalization (CLAHE) yang meningkatkan kontras lokal dengan membatasi distribusi intensitas untuk mencegah amplifikasi noise. Hasil peningkatan kontras perlu dievaluasi untuk memastikan detail penting terlihat jelas tanpa meningkatkan noise berlebihan
4. Deteksi tepi citra adalah proses dalam pengolahan citra untuk mengidentifikasi batas atau kontur objek dengan mendeteksi perubahan intensitas piksel yang signifikan. Proses ini bertujuan untuk

menyederhanakan citra, menyoroti struktur penting, dan mempermudah analisis lanjutan seperti segmentasi

5. Noise reduksi dengan Gaussian Blur adalah teknik yang digunakan dalam pengolahan citra untuk mengurangi noise atau gangguan pada citra dengan cara meratakan intensitas piksel menggunakan filter gaussian
6. Segmentasi dengan kontur digunakan untuk mengambil kontur terbesar dalam dambar (misalnya, area paru-paru pada X-ray) dan membuat masker untuk menyoroti bagian tersebut
7. Augmentasi citra adalah teknik yang dignakan dalam pengolahan citra digital untuk meningkatkan ukuran atau jumlah dataset gambar dengan cara memodifikasi gambar-gambar yang ada. Dalam model ini menggunakan ImageDataGenerator, citra asli dirotasi 20 derajat, diperbesar 20%, serta dibalik vertikal (flip vertikal) untuk menghasilkan variasi data. Tujuanya adalah untuk meningkatkan keberagaman data
8. Ekstraksi fitur adalah proses di mana informasi penting dari citra diambil dan direpresentasikan dalam bentuk fitur numerik. Tujuannya adalah menyederhanakan citra sehingga dapat lebih mudah dianalisis oleh algoritma klasifikasi atau model lainnya. Dalam model ini, dua teknik utama yang digunakan:
    1. HOG (Histogram of Oriented Gradients)

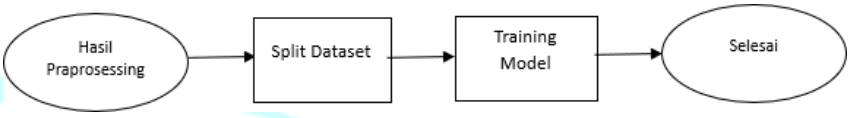
       Menghitung Orientasi gradien pada setiap piksel dalam gambar.Dalam berbagai penelitian HOG cocok dijadikan deskriptor untuk pencarian gambar berbasis sketsa
    2. LBP (Local Binary Pattern)

       Menangkap pola tekstur lokal dengan menghitung perbedaan intensitas piksel terhadap tetangganya. Histogram dari pola ini menjadi representasi fitur
9. *SMOTE (Synthetic Minority Over-sampling Technique)* dilakukan karena citra positif TBC setelah praprosessing lebih sedikit dibandingkan citra normal, *SMOTE* digunakan untuk membuat data sintetik sehingga kelas menjadi seimbang. Dengan cara ini, SMOTE dapat menghasilkan sampel baru yang mewakili variasi dalam kelas minoritas
10. Normalisasi data adalah proses mengubah skala nilai-nilai dalam dataset sehingga berada dalam rentang tertentu atau memiliki distrbusi tertentu,

tanpa mengubah pola atau hubungan antara data tersebut. Tujuan utama dari normalisasi adalah untuk meningkatkan efisiensi dan akurasi algoritma yang digunakan dalam analisis data

3.2.4. Pemodelan SVM

Alur Pemodelan SVM:



Gambar 3. 4 Flowchart Pemodelan SVM

Proses pemodelan dilakukan setelah praprosessing dataset di mana citra rontgen dada telah melalui langkah penting untuk meningkatkan kualitas data dan mempermudah analisis selanjutnya. Berikut adalah penjelasannya:

1. *Split* Dataset

   Setelah proses pra-pemrosesan selesai, dataset dibagi menjadi dua subset, yaitu dataset training dan dataset testing. Pembagian dataset dilakukan dengan hati-hati agar data dalam kedua subset ini mewakili distribusi kelas secara seimbang. Dataset training digunakan untuk melatih model, yaitu proses di mana model belajar mengenali pola dari data yang diberikan. Dataset testing, di sisi lain, digunakan untuk mengevaluasi kemampuan model dalam menggeneralisasi terhadap data baru. Proporsi umum untuk pembagian ini adalah 80% data untuk training dan 20% untuk testing, meskipun proporsi ini dapat disesuaikan berdasarkan ukuran dataset. Pembagian dilakukan menggunakan teknik seperti *train-test split* yang tersedia pada pustaka scikit-learn. Parameter seperti *stratify* digunakan untuk menjaga keseimbangan distribusi kelas (TBC dan normal). Dengan menjaga keseimbangan ini, model dapat dilatih dengan data yang representatif, menghindari bias terhadap salah satu kelas. Pembagian yang baik memastikan proses pelatihan dan evaluasi berjalan dengan optimal, sehingga kinerja model dapat diukur secara objektif pada tahap selanjutnya.

2. *Training* Model

   Pada tahap ini, model Support Vector Machine (SVM) dilatih menggunakan dataset training yang telah dipersiapkan. Model SVM bertujuan

untuk menemukan *hyperplane* optimal yang memisahkan dua kelas (TBC dan normal) dengan margin maksimal. Model dilatih untuk mengenali hubungan antara fitur-fitur yang telah diekstraksi dari citra dengan label kelasnya. Proses pelatihan dimulai dengan memilih linear kernel, yang merupakan pilihan yang tepat ketika data dapat dipisahkan secara linier. Linear kernel bekerja dengan cara mengubah data ke dalam bentuk yang lebih sederhana, di mana pemisahan antara kelas dilakukan menggunakan garis lurus (untuk data 2D) atau bidang datar (untuk data berdimensi lebih tinggi).

3.2.5. Evaluasi

Setelah model *Support Vector Machine* dilatih, langkah selanjutnya adalah melakukan evaluasi kinerja model menggunakan confusion matrix. Confusion matrix memberikan gambaran visual mengenai bagaimana model mengklasifikasikan data, dengan menunjukkan jumlah prediksi yang benar dan salah dalam masing-masing kelas (misalnya TBC dan normal). Matriks ini terdiri dari empat elemen utama: True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN). Dengan demikian, confusion matrix membantu kita untuk memahami kesalahan klasifikasi yang dilakukan oleh model. Selain itu, dari confusion matrix, kita dapat menghitung metrik-metrik evaluasi penting lainnya, seperti precision, recall, dan F1-score. Precision mengukur seberapa tepat prediksi model untuk kelas positif (misalnya, TBC), sementara recall mengukur seberapa baik model dalam mengidentifikasi semua kasus positif yang sebenarnya. F1- score, yang merupakan rata-rata harmonis dari precision dan recall, memberikan gambaran umum tentang keseimbangan model dalam menangani kedua metrik tersebut. Evaluasi ini akan memberikan pemahaman yang lebih jelas tentang kinerja model dalam mengklasifikasikan citra dengan akurat, baik untuk kelas yang lebih sering muncul maupun yang lebih jarang, dan akan membantu dalam menyempurnakan model lebih lanjut jika diperlukan.