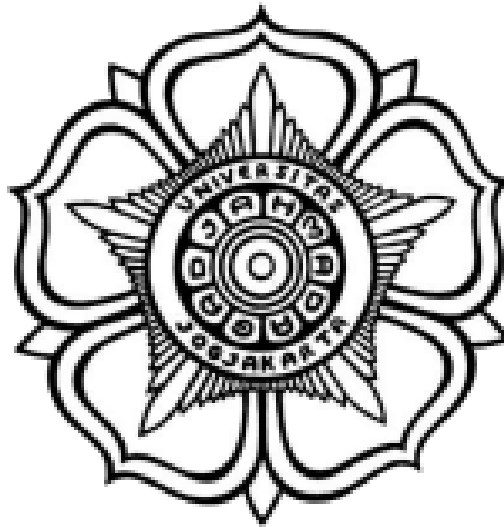


IMPLEMENTASI ENHANCEMENT, SEGMENTASI, DAN EKSTRAKSI FITUR CITRA X-RAY PARU UNTUK MENDUKUNG MODEL KLASIFIKASI TUBERKULOSIS



Disusun Oleh:

1. Widad Muhammad Rafi (24/545635/PA/23190)
2. Keyne Elvaretta (24/533391/PA/22572)
3. Gracella Wiendy Koesnadi (24/541739/PA/22991)
4. Erdizah Ghodi Al Haidar (24/537670/PA/22787)

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2022

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa karena atas rahmat dan karunia-Nya, proposal penelitian ini dapat diselesaikan. Proposal ini disusun sebagai bagian dari upaya untuk mengembangkan sistem deteksi Tuberkulosis (TBC) berbasis pengolahan citra digital dan Machine Learning.

Penulis menyadari bahwa TBC masih menjadi masalah kesehatan serius di Indonesia dan dunia, sehingga diperlukan metode diagnosis yang cepat, akurat, dan efisien. Dengan memanfaatkan teknologi Digital Image Processing (DIP) dan Machine Learning, penelitian ini diharapkan mampu mendukung tenaga medis dalam skrining TBC, terutama di daerah yang memiliki keterbatasan tenaga radiolog.

Penulis menyadari bahwa penyusunan proposal ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, saran, dan masukan sehingga proposal ini dapat terselesaikan.

Penulis berharap proposal ini dapat bermanfaat sebagai referensi ilmiah dan memberikan kontribusi nyata dalam upaya mendeteksi TBC.

Yogyakarta, 3 Oktober 2025

Penulis

DAFTAR ISI

KATA PENGANTAR.....	2
DAFTAR ISI.....	3
BAB I	
PENDAHULUAN.....	4
1.1 Latar Belakang.....	4
1.2 Rumusan Masalah.....	5
1.3 Tujuan.....	5
1.4 Manfaat.....	6
BAB II	
DATASET.....	6
2.1 Sumber Dataset.....	7
2.2 Deskripsi Dataset.....	7
2.3 Ukuran Dataset.....	9
BAB III	
USULAN METODE.....	11
3.1 Acquisition.....	11
3.2 Preprocessing.....	11
3.2.1 Mengonversi data citra menjadi grayscale.....	11
3.2.2 Resizing Citra.....	11
3.3 Enhancement.....	11
3.2.1 Histogram Equalization.....	12
3.2.2 Median Filters.....	12
3.2.3 Laplace Filters.....	12
3.4 Segmentation.....	13
3.4.1 Thresholding Based Image Segmentation.....	13
3.4.2 Segmentation by Region and Feature.....	13
3.5 Feature Extraction.....	14
3.6 Modelling.....	14
3.6.1 Support Vector Machine (SVM).....	14
3.7 Evaluation.....	15
3.7.1 Macro F1.....	15
3.7.3 Confusion Matrix.....	15
3.7.2 Accuracy.....	15
3.7.4 Precision.....	16
DAFTAR PUSTAKA.....	17
LAMPIRAN.....	19

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis (TBC) adalah penyakit menular serius yang disebabkan oleh infeksi bakteri *Mycobacterium tuberculosis* dan telah menjadi masalah kesehatan global selama ribuan tahun (Smeltzer, 2016). Penularannya terjadi melalui udara (*airborne disease*), di mana bakteri tersebar melalui percikan dahak saat penderita TBC aktif batuk atau bersin. Karena mudah menular dan berpotensi fatal jika tidak ditangani, deteksi dini yang cepat dan tepat menjadi sangat penting untuk mengendalikan penyebaran penyakit dan menyelamatkan nyawa.

Urgensi penanganan TBC semakin tinggi mengingat penyakit ini termasuk salah satu penyebab kematian terbesar di dunia. Menurut World Health Organization (WHO), TBC menewaskan lebih dari 1,3 juta orang setiap tahun secara global. Indonesia menghadapi beban yang sangat besar, menempati peringkat kedua negara dengan jumlah kasus TBC tertinggi di dunia. Kondisi ini menunjukkan bahwa TBC bukan hanya masalah kesehatan global, tetapi juga krisis nasional yang membutuhkan solusi inovatif dan efektif.

Metode diagnosis konvensional untuk TBC sering kali mengandalkan interpretasi citra medis seperti X-ray paru oleh dokter spesialis radiologi. Namun, metode manual ini memiliki beberapa keterbatasan, antara lain proses yang memakan waktu, potensi subjektivitas antar para ahli, serta ketergantungan pada ketersediaan tenaga medis terampil yang jumlahnya terbatas, terutama di daerah-daerah terpencil.

Untuk mengatasi tantangan ini, teknologi Digital Image Processing (DIP) dan Machine Learning (ML) menawarkan solusi yang potensial. DIP adalah teknik komputasi yang digunakan untuk menganalisis dan memanipulasi gambar digital dengan tujuan meningkatkan kualitas atau mengekstrak informasi penting. Sejak tahun 1970-an, DIP telah menjadi alat penting dalam dunia medis, mulai dari pengembangan tomografi untuk pencitraan 3D hingga sistem Computer-Aided Diagnosis (CAD) yang membantu dokter dalam pengambilan keputusan klinis. Tahapan DIP yang sistematis, yaitu akuisisi citra,

peningkatan kualitas, segmentasi, dan ekstraksi fitur, memungkinkan analisis citra yang lebih objektif dan terukur.

Dengan menggabungkan DIP dan Machine Learning, sistem dapat secara otomatis mengenali pola-pola penyakit pada citra medis. Pendekatan ini diharapkan menghasilkan alat bantu skrining yang cepat, konsisten, dan akurat untuk mendukung tenaga medis dalam deteksi TBC.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana memanfaatkan Digital Image Processing (DIP) untuk meningkatkan kualitas citra X-ray paru-paru sehingga fitur penting terkait TBC dapat diekstraksi dengan akurat?
2. Bagaimana teknik klasifikasi berbasis Machine Learning dapat digunakan untuk membedakan citra X-ray paru-paru normal dengan yang terindikasi TBC secara otomatis?
3. Seberapa efektif kombinasi DIP dan Machine Learning dalam mempercepat serta meningkatkan akurasi deteksi TBC dibandingkan metode diagnosis manual?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Mengembangkan sistem pengolahan citra X-ray menggunakan DIP untuk mengekstraksi fitur yang relevan dengan TBC.
2. Menerapkan teknik klasifikasi berbasis Machine Learning untuk membedakan citra X-ray paru-paru sehat dan terindikasi TBC.
3. Menciptakan alat bantu skrining TBC yang cepat, akurat, dan konsisten untuk mendukung tenaga medis dalam diagnosis.

1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini meliputi:

1. Memberikan alat bantu diagnosis TBC yang cepat, akurat, dan efisien.
2. Mengurangi ketergantungan pada pemeriksaan manual yang memakan waktu dan bergantung pada keterampilan ahli.
3. Mendukung tenaga medis dalam pengambilan keputusan klinis dan perencanaan perawatan pasien.

BAB II

DATASET

2.1 Sumber Dataset

Dataset yang digunakan dalam penelitian ini adalah “**Tuberculosis (TB) Chest X-ray Database**” yang tersedia melalui **IEEE DataPort**. Dataset ini dipublikasikan sebagai repositori terbuka untuk penelitian medis dan telah digunakan dalam sejumlah studi terkait deteksi TBC pada citra X-ray.

Dataset ini mencakup citra X-Ray dada dari pasien dengan kondisi *normal* (sehat) dan *abnormal* (terdapat infeksi TBC). Selain itu, ketersediaannya dalam platform IEEE DataPort menjamin dataset ini memiliki dokumentasi yang jelas, akses yang terstandar, serta dapat dipertanggungjawabkan secara akademis.

2.2 Deskripsi Dataset

Dataset ini dipilih karena memiliki beberapa keunggulan, yaitu: (1) jumlah data relatif seimbang antara kelas normal dan TBC, sehingga mengurangi bias dalam pelatihan model; (2) ukuran dataset yang moderat (7000 citra, 720 MB), sehingga masih dapat diproses dengan sumber daya komputasi terbatas seperti Google Colab atau Kaggle; serta (3) dataset ini bersifat terbuka dan telah digunakan dalam berbagai penelitian terdahulu, sehingga validitas dan reliabilitasnya lumayan terjamin.

Dataset “Tuberculosis (TB) Chest X-ray Database” berisi citra X-ray dada pasien yang terbagi menjadi dua kategori utama, yaitu **normal (sehat)** dan **abnormal (terdapat indikasi TBC)**. Setiap citra dalam dataset telah melalui proses verifikasi medis sehingga dapat dijadikan acuan dalam pengembangan sistem pendeteksian otomatis berbasis pembelajaran mesin.

Seluruh citra disimpan dalam format PNG dengan resolusi seragam, sehingga memudahkan proses pra-pengolahan (preprocessing) dan pelatihan model. Dataset ini dirancang agar dapat digunakan tidak hanya untuk klasifikasi biner (TB vs non-TB), tetapi juga untuk mendukung penelitian lebih lanjut dalam deteksi pola pada citra medis.

Untuk meningkatkan reliabilitas, dataset ini telah digunakan dalam beberapa penelitian terdahulu mengenai deteksi TBC berbasis citra X-ray. Hal ini memastikan bahwa data yang tersedia memiliki kualitas baik serta relevan bagi pengembangan model kecerdasan buatan di bidang medis.

Secara ringkas, karakteristik utama dataset ini dapat dilihat pada Tabel 2.2 berikut:

Tabel 2.2 Karakteristik utama dataset

Karakteristik	Keterangan
Jenis data	Citra X-ray dada
Kategori	Normal (Tidak terjangkit TBC), Abnormal (TBC)
Format File	JPG
Resolusi	1024 x 1024 piksel
Jumlah total citra	7000 (normal dan TBC)
Tujuan Utama	Penelitian klasifikasi citra medis

Karakteristik ini menunjukkan bahwa dataset cukup representatif untuk penelitian deteksi TBC berbasis citra medis.

Sebagai ilustrasi, berikut ditampilkan contoh citra X-ray dari dataset yang digunakan. Gambar di bawah memperlihatkan perbedaan visual antara citra dada pasien dengan kondisi **abnormal (positif TBC)** dan pasien dengan kondisi **normal (sehat)**. Penyertaan sampel ini bertujuan untuk memberikan gambaran nyata mengenai jenis data yang diproses dalam penelitian.



(a)



(b)

Gambar 2.1 Contoh citra X-ray dari dataset: (a) pasien dengan TBC, (b) pasien sehat/normal.

2.3 Ukuran Dataset

Dataset “Tuberculosis (TB) Chest X-ray Database” memiliki total sekitar 7.000 citra X-ray dengan format JPG dan resolusi seragam 1024×1024 piksel. Jumlah ini dianggap cukup moderat: tidak terlalu kecil sehingga tetap mampu mewakili variasi data medis, namun juga tidak terlalu besar sehingga masih dapat diproses menggunakan sumber daya terbatas seperti Google Colab maupun Kaggle.

Untuk mendukung proses pelatihan dan evaluasi model, dataset ini dibagi menjadi tiga subset, yaitu training set, validation set, dan test set. Proporsi pembagian mengikuti praktik umum dalam machine learning, yaitu 70% : 15% : 15%. Pembagian ini dilakukan secara stratified, artinya distribusi kelas normal dan TBC tetap dijaga seimbang di setiap subset, sehingga model tidak mengalami bias terhadap salah satu kelas.

Distribusi dataset ditunjukkan pada Tabel 2.3 berikut.

Tabel 2.3 Distribusi dataset berdasarkan subset dan kategori

Subset	Normal	TBC	Total
--------	--------	-----	-------

Training (70%)	2.450	2.450	4.900
Validation (15%)	525	525	1.050
Testing (15%)	525	525	1.050
Total	3.500	3.500	7.000

Dengan pembagian data seperti di atas, model diharapkan dapat dilatih secara seimbang dan diuji secara adil terhadap kedua kategori (normal dan TBC).

BAB III

USULAN METODE

Metode yang diusulkan didasarkan pada teknik AESFERM (*Acquisition, Enhancement, Segmentation, Feature Extraction, Evaluation*).

3.1 Acquisition

Akuisisi (acquisition) adalah teknik pengambilan data citra. Data citra yang digunakan dalam proyek akhir adalah X-ray dada. Pengambilan data citra dilakukan dengan mencari kumpulan dataset dari IEEE port, seperti yang tertera pada Bab II. Judul dataset yang digunakan adalah “Tuberculosis (TB) Chest X-ray Database”.

Setelah mencari data citra, langkah selanjutnya yang dilakukan adalah menentukan kategori dari data citra. Data citra X-ray merupakan **citra non spasial**. Citra non spasial adalah citra yang tidak memiliki informasi geografis [4].

3.2 Preprocessing

3.2.1 Mengonversi data citra menjadi grayscale

Teknik preprocessing ini bertujuan agar semua data citra yang masuk diseragamkan menjadi grayscale, antisipasi apabila terdapat data citra yang bukan grayscale. Hal ini akan mempermudah dalam pemrosesan citra selanjutnya.

3.2.2 Resizing Citra

Resizing citra dilakukan agar seluruh resolusi data citra yang masuk menjadi seragam.

3.3 Enhancement

Enhancement adalah teknik untuk meningkatkan kualitas citra. Tujuan dari teknik ini adalah agar citra menjadi lebih jelas, menghilangkan noise dan dapat meng-highlight detail yang nantinya akan digunakan untuk dibaca dan dipelajari model [3]. Teknik enhancement yang dapat digunakan adalah sebagai berikut:

3.2.1 Histogram Equalization

Histogram Equalization adalah teknik meningkatkan citra dengan mendistribusi ulang intensitas pixel agar lebih merata. Histogram equalization dapat membantu memperjelas detail-detail yang ada pada citra.

Langkah-langkah histogram equalization adalah sebagai berikut:

1. Hitung histogram citra asli
2. Hitung fungsi kumulatif (CDF) dari histogram citra
3. Terapkan histogram equalization dengan setiap pixel dipetakan sebagai berikut:

$$S_k = (L - 1) \times CDF(r_k)$$

dimana

S_k = intensitas baru setelah equalization

r_k = intensitas pixel citra asli

L = jumlah level intensitas (256)

3.2.2 Median Filters

Median filters merupakan teknik peningkatan citra dimana pixel tengah dari sebuah matriks $N \times N$ digantikan dengan nilai median dari tetangga-tetangganya. Median filters digunakan untuk menghilangkan noise yang ada [8].

3.2.3 Laplace Filters

Laplace Filters akan digunakan untuk mendeteksi tepian dari paru-paru sehingga model mengetahui batas dari paru-paru. Selain itu, laplace filters akan digunakan untuk mempertegas detail seperti bagian-bagian yang menunjukkan adanya TBC.

Laplace filter merupakan turunan kedua dari citra. Laplace filter bekerja dengan persamaan sebagai berikut:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Gambar 3.1 Persamaan Fungsi Laplace Filters

3.4 Segmentation

Segmentation adalah proses membagi citra menjadi beberapa daerah (region) yang lebih bermakna, biasanya memisahkan objek utama dari background. Metode yang digunakan:

3.4.1 Thresholding Based Image Segmentation

Thresholding bertujuan untuk mengisolasi area umum yang mencurigakan pada X-Ray atau untuk memisahkan bakteri dari latar belakang pada citra mikroskopik.

Thresholding Based Image Segmentation dilakukan dengan memisahkan antara objek/foreground dengan background menggunakan intensitas pixel. Nilai yang memisahkan antara foreground dengan background disebut dengan Threshold.

Salah satu cara mencari Threshold yang akan digunakan dalam proyek ini adalah dengan menggunakan Otsu's Thresholding. Otsu's Thresholding bekerja dengan cara mencari nilai optimal dengan mencari nilai Thresholding yang memiliki nilai varians intra-kelas paling minimum di antara nilai Thresholding yang lain.

3.4.2 Segmentation by Region and Feature

a. Region Growing

Region Growing digunakan untuk memperjelas bentuk lesi TBC atau area paru - paru yang mengalami konsolidasi pada citra X-Ray, di mana pixel-pixel yang berdekatan memiliki properti intensitas yang serupa. Cara kerja Region Growing adalah dengan memulai segmentasi dari satu pixel awal, kemudian tambahkan pixel tetangga ke dalam region jika perbedaan intensitas keduanya tidak lebih dari nilai Threshold.

b. Segmentation as Clustering

Segmenting as clustering digunakan untuk membagi citra menjadi beberapa region/kelompok berdasarkan kemiripan fitur. Cara kerja *clustering segmentation* yaitu di mana setiap pixel dapat direpresentasikan sebagai vektor fitur. Vektor-vektor pixel ini kemudian dikelompokkan menjadi segmen-segmen menggunakan algoritma clustering seperti *K-Means Clustering*.

3.5 Feature Extraction

Feature extraction adalah proses mengambil informasi numerik yang mewakili karakteristik penting dalam citra.

3.6 Modelling

3.6.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) biasanya digunakan untuk masalah klasifikasi. SVM bekerja dengan menggunakan vektor. SVM membedakan dua kelas dengan cara menemukan hyperplane yang optimal. Sebuah hyperplane dikatakan optimal ketika berhasil memaksimalkan margin antara titik data terdekat dari kelas yang berlawanan [9].

Cara kerja SVM:

1. Mencari Hyperplane

$$\omega \cdot x + b = 0$$

Dimana

ω = vektor bobot

x = vektor input (fitur)

b = bias intercept

Kelas dipisahkan oleh tanda fungsi:

$$f(x) = \text{sign}(\omega \cdot x + b)$$

- Jika $f(x) = +1$, kelas positif
- Jika $f(x) = -1$, kelas negatif

2. Margin

$\text{Margin} = \frac{2}{|\omega|}$, dimana semakin besar margin, semakin baik pemisahannya.

Secara umum, SVM akan mencari hyperlane yang memaksimalkan margin, dengan optimasi dan constraint sesuai sebuah label.

Pemilihan model SVM ini didasarkan pada kasus deteksi TBC, dimana terdapat 2 kelas yaitu “normal” atau “TBC”. Dengan cara kerja SVM yang dapat menangani klasifikasi linear dan non linear, serta optimasi margin yang dapat meningkatkan akurasi klasifikasi, SVM menjadi pilihan relevan untuk pembangunan model deteksi TBC.

3.7 Evaluation

3.7.1 Macro F1

Digunakan untuk menguji harmonik rata - rata Precision & Recall. Rumus:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.7.3 Confusion Matrix

Matrix yang digunakan untuk mengetahui jumlah dari TP, TN, FP, FN. Confusion matriks dapat dilihat pada Gambar 3.2.

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Gambar 3.2 Confusion Matrix

3.7.2 Accuracy

Digunakan untuk mengetahui seberapa banyak prediksi yang benar secara keseluruhan. Rumus accuracy dinyatakan sebagai berikut.

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

Dimana:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

F = *False Negative*

3.7.4 Precision

Digunakan untuk mengetahui seberapa akurat model saat mengatakan positif

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

DAFTAR PUSTAKA

- [1] Tawsifur Rahman, Amith Khandakar, Muhammad Enamul Hoque Chowdhury, "Tuberculosis (TB) Chest X-ray Database", IEEE Dataport, October 20, 2020, doi:10.21227/mps8-kb56
- [2] T. Rahman, A. Khandakar, M. Abdul Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. Islam, Z. B. Mahbub, M. A. Ayari, and M. E. H. Chowdhury, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586-191601, 2020, doi:10.1109/[ACCESS.2020.xxx](#).
- [3] Gökçenaz Akyol, "What is Image Enhancement? | Image Processing #3," Medium, Jan. 14, 2023.
<https://medium.com/@gokcenazakyol/what-is-image-enhancement-image-processing-3-32a813087e0a> (accessed Oct. 02, 2025).
- [4] N. Lee, "What is spatial data and non-spatial data? - FME by Safe Software," FME by Safe Software, Oct. 25, 2021. <https://fme.safe.com/blog/2021/10/non-spatial-data-difference-fme/> (accessed Oct. 02, 2025).
- [5] Smeltzer, Fb. &. (2016). Textbook of Medical-surgical Nursing. Lippincott Williams & Wilkins Pty, Limited. <https://books.google.co.id/books?id=FobljwEACAAJ>
- [6] M. Alsaffar, G. Alshammari, A. Alshammari, S. Aljaloud, T.S. Almurayziq, A.A. Hamad, V. Kumar, and A. Belay, "Detection of Tuberculosis Disease Using Image Processing Technique," *Journal of Healthcare Engineering*, Dec. 03, 2021. <https://doi.org/10.1155/2021/7424836> (accessed Oct. 03, 2025).
- [7] Kementerian Kesehatan Republik Indonesia. (2024, 23 Juli). Hari Anak Nasional 2024: Masyarakat harus pahami karakteristik TBC pada anak. Kementerian Kesehatan RI.
<https://kemkes.go.id/id/hari-anak-nasional-2024-masyarakat-harus-pahami-karakteristik-tbc>

[8] İmren Dinç, Semih Dinç, Madhav Sigdel, M. S. Sigdel, R. S. Aygün, and M. L. Pusey, “DT-Binarize,” Elsevier eBooks, pp. 183–199, Dec. 2014, doi: <https://doi.org/10.1016/b978-0-12-802045-6.00012-0>.

[9] IBM, “Support Vector Machine,” *Ibm.com*, Dec. 12, 2023. <https://www.ibm.com/id-id/think/topics/support-vector-machine> (accessed Oct. 03, 2025).