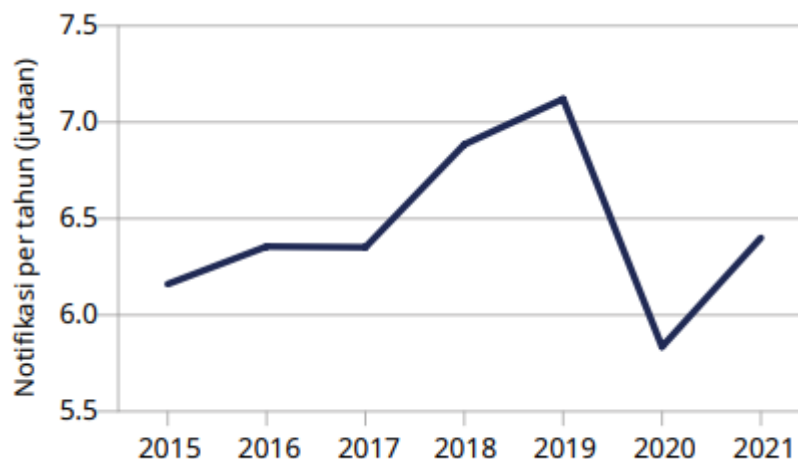


BAB I

PENDAHULUAN

1.1. Latar Belakang

Tuberculosis (TBC) atau (TB) adalah penyakit menular yang merupakan penyebab utama kesehatan yang buruk dan salah satu penyebab utama kematian di seluruh dunia, penyakit TBC menular disebabkan oleh bakteri *Mycobacterium Tuberculosis* dan menyerang organ paru [1]. Hingga pandemi virus corona (COVID-19), TBC merupakan penyebab utama kematian dari agen infeksi tunggal, dengan peringkat di atas *HIV/AIDS*. Menurut data *World Health Organization* (WHO), sekitar 6,4 juta orang terinfeksi TBC pada tahun 2021 dan sekitar 1,6 juta orang meninggal akibat penyakit ini di seluruh dunia [2]. Berdasarkan *Global Tuberculosis Report 2022* dampak yang paling jelas dan langsung terhadap gangguan TB yang disebabkan oleh pandemi COVID-19 adalah penurunan global yang besar dalam jumlah orang yang baru didiagnosis dengan TB dan dilaporkan (yaitu secara resmi diberitahukan) pada tahun 2020, dibandingkan dengan tahun 2019 (Gambar 1.1). Menyusul peningkatan besar antara 2017 dan 2019, terjadi penurunan sebesar 18% antara 2019 dan 2020, dari 7,1 juta menjadi 5,8 juta. Ada peningkatan sebesar 9,375% pada tahun 2021, menjadi 6,4 juta.



Gambar 1. 1 Tren Global Kasus Tuberculosis [2]

Di Indonesia, angka kejadian TBC masih cukup tinggi, bahkan menjadi masalah kesehatan yang serius. Deteksi dini dan pengobatan yang tepat merupakan diagnostik kunci dalam pengendalian TBC [3]. Namun, diagnostik TBC yang tepat dan akurat seringkali sulit karena gejala-gejala yang mirip dengan penyakit lain dan ketidaktersediaan metode diagnostik yang murah dan mudah diakses. Salah satu metode yang dapat digunakan untuk mendeteksi TB adalah citra *chest X-ray* [4]. Citra *chest X-ray* dapat memberikan informasi visual tentang kondisi paru-paru pasien. Namun, interpretasi citra *chest X-ray* untuk diagnosis TB membutuhkan keahlian khusus dari radiolog atau dokter spesialis patologi anatomi. Oleh karena itu, pengembangan sistem komputer yang dapat membantu dalam interpretasi citra *chest X-ray* dapat membantu meningkatkan kecepatan dan akurasi dalam diagnosis TB.

Dalam beberapa tahun terakhir, beragam studi di bidang pencitraan medis telah menunjukkan kemajuan yang signifikan dalam klasifikasi penyakit tuberkulosis [5]. Analisis citra dapat digunakan untuk mendeteksi penyakit tuberkulosis dan bahkan membantu dalam diagnosa penyakit. Dengan menggunakan teknik *deep learning*, citra dapat dianalisis untuk mengetahui sedari awal apakah pasien tersebut menderita penyakit tuberkulosis atau tidak. Data citra *chest X-ray* penyakit tuberkulosis tersedia dalam jumlah yang besar dan mudah diakses. Penelitian ini akan menganalisis citra dalam bentuk asli dan citra yang telah tersegmentasi dimana sumber data citra berasal dari *website Kaggle*. Beberapa metode yang sering diimplementasikan untuk mengklasifikasikan data citra, salah satunya adalah metode *Convolutional Neural Network* (CNN). Sebelumnya penelitian telah dilakukan pada *deep learning* untuk mendeteksi dan mendiagnosis penyakit tuberkulosis [6].

Convolutional Neural Network (CNN) adalah jenis algoritma pembelajaran mesin yang dapat memproses dan memahami citra secara otomatis [7]. CNN telah terbukti berhasil dalam berbagai aplikasi pengolahan citra, termasuk dalam pengklasifikasian citra medis. CNN dapat digunakan untuk mengklasifikasikan citra *chest X-ray* menjadi kategori TB dan non-TB

dengan akurasi yang tinggi [8]. Sebelumnya penelitian telah dilakukan menggunakan *deep learning* untuk mendeteksi dan mendiagnosis penyakit TBC [9]. *Convolutional Neural Network* menerapkan struktur standar yang ditumpuk lapisan konvolusional (opsional diikuti dengan normalisasi dan *max-pooling*) ditambahkan oleh satu atau lebih yang terhubung sepenuhnya lapisan. Varian desain dasar ini umum pada gambar literatur klasifikasi, menghasilkan hasil terbaik dibandingkan untuk *MNIST*, data *CIFAR* dan yang paling menonjol diklasifikasi *ImageNet* [10]. Untuk *dataset* yang lebih besar seperti *ImageNet*, tren saat ini adalah menambah jumlah *layer* [11] dan ukuran lapisan untuk mengatasi masalah *overfitting* [12]. Meskipun ada kekhawatiran bahwa penyatuan maksimal lapisan mungkin terjadi menghasilkan akurasi yang lebih rendah [13], model CNN serupa berhasil terlibat untuk lokalisasi deteksi objek dan estimasi pose manusia [14].

Pada penelitian terdahulu berpendapat bahwa model arsitektur *VGG16* memiliki performa yang bagus untuk *Shenzhen chest X-ray dataset* dengan akurasi 86,74%, namun penelitian tersebut hanya melakukan klasifikasi terhadap citra asli yang tidak dilakukan segmentasi terlebih dahulu [15]. Dalam penelitian terdahulu, penulis menyajikan klasifikasi *deep learning* untuk mendeteksi manifestasi TB dalam *chest X-ray* (CXR), berdasarkan penelitian sebelumnya dalam segmentasi paru-paru dan klasifikasi penyakit paru-paru [16]. Arsitektur CNN yang digunakan dalam penelitian ini adalah *VGG16* dengan mengubah beberapa parameter untuk mendapatkan performa terbaik. Citra X-ray terdiri dari kelas normal dan manifestasi tuberkulosis nantinya akan diklasifikasikan menggunakan citra asli dan citra yang telah tersegmentasi dari penelitian sebelumnya [17], untuk melihat ketepatan tingkat akurasi. Parameter metode pengujian yang diusulkan meliputi *accuracy*, *precision*, dan *recall*. Kontribusi penelitian ini adalah membangun model CNN untuk deteksi TB pada rontgen dada dan optimalisasi parameter model untuk mendapatkan akurasi tertinggi. Dengan metode yang diusulkan ini diharapkan dapat mendukung pencapaian penelitian sebelumnya sehingga dapat memperkuat diagnosis TB menggunakan analisis digital.

Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaan seperti itu, maka penulis tertarik untuk melakukan klasifikasi penyakit tuberkulosis berdasarkan citra rontgen dada dengan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan suatu pendekatan metodologi yang digunakan dalam mengelola proyek data mining atau analisis data. Metodologi ini terdiri dari serangkaian tahapan yang terstruktur untuk mengidentifikasi dan menyelesaikan permasalahan yang terkait dengan analisis data [18]. Untuk mengimplementasikan ketertarikan penulis terhadap masalah tersebut, maka penulis menuangkan penelitiannya dalam bentuk penulisan skripsi yang berjudul **“Klasifikasi Penyakit Tuberculosis Berdasarkan Citra Chest X-Ray Menggunakan Convolutional Neural Network Model Arsitektur VGG16”**.

1.2. Rumusan Masalah

1.2.1. Permasalahan

Terlambatnya penanganan pasien tuberkulosis akan menurunkan fungsi tubuh dari waktu ke waktu. Citra *chest X-ray* dapat memberikan informasi visual tentang kondisi paru-paru pasien. Oleh karena itu dibutuhkan metode diagnosa yang lebih efisien untuk penanganan gejala tuberkulosis sedari awal. *Deep learning* dapat digunakan dalam mengklasifikasi dengan analisis dari citra yang ada.

1.2.2. Metode Usulan

Penelitian ini mengusulkan metode CNN menggunakan model arsitektur *VGG16* untuk melakukan diagnosa penyakit tuberkulosis.

1.2.3. Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan maka dapat diketahui rumusan masalah sebagai berikut:

1. Bagaimana kinerja yang dihasilkan dari *Convolutional Neural Network* model arsitektur *VGG16* dalam melakukan klasifikasi penyakit Tuberculosis pada citra *Chest X-ray*?
2. Bagaimana perbandingan kinerja *Convolutional Neural Network* model arsitektur *VGG16* dari citra asli dan yang telah tersegmentasi dalam melakukan klasifikasi penyakit Tuberculosis dengan menggunakan *Transfer learning* maupun tidak pada citra *Chest X-ray*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan maka dapat diketahui tujuan sebagai berikut:

1. Untuk menguji kinerja dari *Convolutional Neural Network* (CNN) model dengan arsitektur *VGG16* dalam melakukan klasifikasi penyakit Tuberculosis pada citra *Chest X-ray*
2. Untuk membandingkan kinerja *Convolutional Neural Network* (CNN) model dengan arsitektur *VGG16* dalam melakukan klasifikasi penyakit Tuberculosis pada citra *Chest X-ray* antara citra asli dan citra yang telah tersegmentasi dengan menggunakan *Transfer learning* maupun tidak

1.4. Manfaat Penelitian

Berdasarkan tujuan yang telah diuraikan maka dapat diketahui manfaat penelitian sebagai berikut:

1. Manfaat bagi pihak medis
Manfaat penelitian bagi pihak medis yaitu untuk mempermudah proses dan meminimalisir kesalahan diagnosa pada penderita penyakit tuberkulosis.
2. Manfaat bagi pembaca
Adapun beberapa manfaat penelitian bagi pembaca sebagai berikut:

- a. Sebagai bahan studi dan tambahan wawasan pada topik *deep learning* terutama menggunakan model CNN dengan arsitektur *VGG16*
 - b. Sebagai referensi untuk mengetahui keakurasian dari model CNN dengan arsitektur *VGG16* berdasarkan citra *Chest X-ray*
3. Manfaat bagi penulis
- Manfaat penelitian bagi penulis yaitu sebagai penerapan dari ilmu yang telah didapatkan saat menempuh kuliah di Universitas Trunojoyo Madura

1.5. Batasan Masalah

Berdasarkan rumusan masalah yang telah diuraikan maka dapat diketahui beberapa Batasan penelitian sebagai berikut:

1. Penelitian ini menggunakan *Shenzhen Chest X-ray Dataset* dan *Lung Mask for Shenzhen Chest X-ray Dataset*. Dua *dataset* yang digunakan dalam penelitian ini diambil dari situs *Kaggle* yang dimana bisa didownload gratis di link (<https://www.kaggle.com/daddar/tuberculosis-chest-xrays-shenzhen>) dan (<https://www.kaggle.com/yoctoman/shcxr-lung-mask>).
2. *Shenzhen Chest X-ray Dataset* tersebut terdiri dari 662 citra rontgen dada, dimana 326 citra normal dan 336 citra dengan manifestasi tuberkulosis yang meliputi 460 pria dan 202 wanita dengan umur rata-rata 1 hingga 89 tahun. Sedangkan untuk *Lung Mask for Shenzhen Chest X-ray Dataset* berisikan hasil segmentasi *lung mask* kanan dan kiri dari *Shenzhen Chest X-ray Dataset* dengan jumlah sama yakni 662 citra. Label pada *dataset* ini menunjukkan apakah subjek yang diuji menderita penyakit tuberkulosis (1) atau tidak (0).

3. Penerapan data yang digunakan untuk menghasilkan akurasi terbaik dibagi menjadi 4 skenario pengujian yaitu *dataset* asli, *dataset* hasil segmentasi, *dataset* asli menggunakan *Transfer learning* maupun tidak, dan *dataset* tersegmentasi menggunakan *Transfer learning* maupun tidak
4. Metode penelitian yang digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*)
5. Algoritma klasifikasi yang digunakan adalah CNN dengan arsitektur VGG16
6. Evaluasi kinerja *deep learning* dilakukan menggunakan *confusion matrix* yang menunjukkan nilai *accuracy*, *precision*, *recall* dan kurva AUC

1.6. Sistematika Penelitian

Penulisan proposal skripsi ini disusun secara sistematis dan dibagi menjadi beberapa BAB, seperti berikut:

BAB I PENDAHULUAN

BAB ini menjelaskan tentang latar belakang penelitian apa yang akan dilakukan, rumusan masalah, tujuan, batasan masalah, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

BAB ini akan membahas teori tentang penyakit *Tuberculosis*, *Machine Learning*, *Deep Learning*, *Adaptive Histogram Equalization*, *Transfer Learning*, *CNN*, *Convolution Layer*, *Stride*, *Padding*, *Pooling Layer*, *Fully-Connected Layer*, *VGG16*, *Data Segmentation* dan *Data Augmentation*, *Confusion Matrix*, Kurva ROC dan AUC, *K-Fold Cross Validation*, dan penelitian terdahulu

BAB III METODE PENELITIAN

BAB ini akan membahas tentang metode penelitian menggunakan CRISP-DM yang akan digunakan dalam proses pengumpulan dan pengolahan data menggunakan metode yang sudah ditetapkan

BAB IV PEMBAHASAN

BAB ini akan membahas lebih rinci mengenai penerapan dan hasil klasifikasi penyakit tuberculosis menggunakan metode *VGG16*

BAB V PENUTUP

BAB ini berisi kesimpulan yang didapatkan dari penelitian dan saran untuk pengembangan penelitian selanjutnya

DAFTAR PUSTAKA