



BAB III


METODOLOGI PENELITIAN

3.1 Variabel Penelitian

Macam-macam variabel yang digunakan pada penelitian ini adalah:

Tabel 3. 1 Variabel Penelitian

Data Citra	Variabel	Definisi
	Covid-19	Pada paru-paru penyakit Covid-19 dapat terlihat bercak putih atau abu-abu muda. Bercak-bercak ini merupakan pertanda adanya peradangan atau infeksi.
	Pneumonia	Pada rontgen dada pneumonia biasanya terlihat area buram atau berawan di paru-paru. Biasanya terlihat di lobus bawah paru-paru. Area ini disebut ground-glass opacity (GGO). Selain GGO, pneumonia juga dapat terlihat sebagai area yang padat di paru-paru. Area ini disebut konsolidasi. Konsolidasi pada pneumonia biasanya terlihat di kedua paru-paru, terutama di lobus bawah.

	Tuberkulosis	Gambaran rontgen pada dada pada penderita Tuberkulosis terdapat bercak di area paru-paru dan terlihat buram atau berawan. Bercak pada Tuberkulosis biasanya terlihat di lobus atas paru-paru.
---	--------------	---

3.2 Alat Bahan Penelitian

Secara umum alat yang digunakan pada penelitian ini terdiri dari dua bagian, yaitu :

1. Perangkat Keras (Hardware)

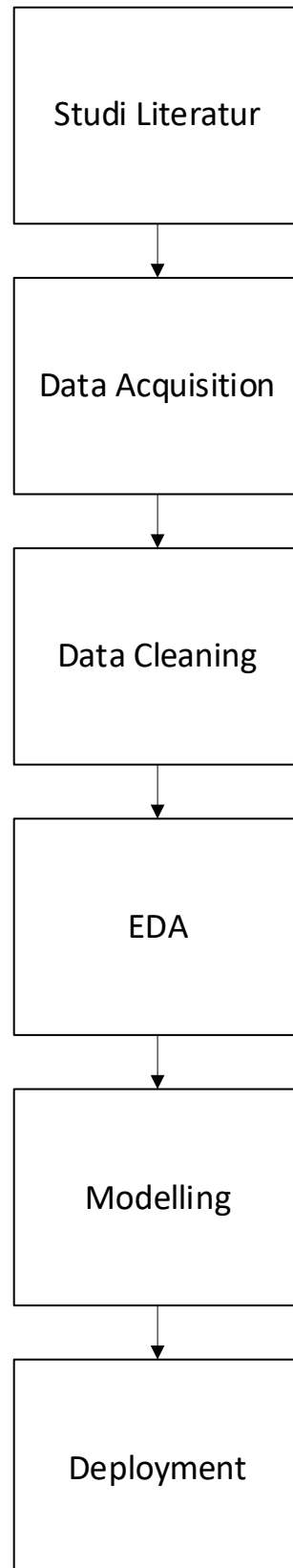
- *Processor* : Intel® Core™ i5-6200U 2.8 GHz
- *Grafis* : Nvidia® GeForce® 940MX 2GB
- *RAM* : 8GB DDR4

2. Perangkat Lunak (Software)

- *Sistem Operasi* : Windows 11
- *Arsitektur* : 64-bit *operating system*
- *Pengolahan Data* : Matlab 2022a

3.3 Tahapan Penelitian

Tahapan penelitian ini dimaksudkan untuk menjelaskan semua tahapan proses yang akan dilakukan selama penelitian. Dalam penelitian ini terdapat beberapa tahapan alur. Berikut tahapan alur nya :



Gambar 3. 1 Tahapan Penelitian

3.3.1 Studi Literatur

Studi literatur merupakan serangkaian kegiatan yang berkaitan dengan identifikasi masalah, pengolahan bahan penelitian dan mencari referensi yang berkaitan dengan penelitian ini, seperti mencari jurnal-jurnal yang berkaitan dengan metode yang sama untuk mengembangkan konsep penelitian yang akan dikerjakan.

3.3.2 Data Acquisition

Data citra rontgen paru-paru dikumpulkan dari sumber *online* dari halaman *Lungs Disease Dataset* di situs <https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types>. Dataset ini mencakup citra untuk tiga kelas penyakit: Pneumonia, Tuberkulosis, dan COVID-19. Selanjutnya untuk pengkategorian menggunakan simbol agar memudahkan memilah jenis rontgen pada penyakit paru-paru.

3.3.3 Data Cleaning

Tahapan kedua dalam penelitian ini adalah Pembersihan dan Persiapan Data (*Preprocessing*). Langkah ini sangat penting untuk memastikan data citra rontgen paru-paru siap diolah dan dianalisis oleh model CNN. Proses ini melibatkan beberapa langkah. Pertama, citra rontgen dipotong (*cropping*) menjadi bentuk persegi. Langkah ini dilakukan untuk menyeragamkan ukuran citra, karena citra rontgen yang diperoleh memiliki ukuran dan dimensi yang

berbeda-beda. Kemudian, ukuran citra diubah (*resize*) menjadi 128x128 piksel. Ukuran ini dipilih karena dianggap optimal untuk pelatihan model CNN. Selanjutnya, citra diubah menjadi skala abu-abu (*grayscale*). Konversi ini menyederhanakan representasi piksel dengan hanya menggunakan satu channel warna, sehingga mengurangi kompleksitas model dan mempercepat pelatihan.

Data terlebih dahulu dilakukan pembagian menjadi 2 folder yaitu data *training* dan data *test*. Pada setiap sub folder dibuat 3 folder dan diberi nama covid-19, pneumonia, tuberkulosis. Pada penelitian ini peneliti menggunakan dataset 90% : 10%, 80% : 20%, 70% : 30%, 60% : 40%. Pembagian dataset dapat dijelaskan pada tabel :

Tabel 3. 2 Pembagian data training dan data test

Pembagian Data 90% : 10%	No	Nama	Jenis	Jumlah
	1.	Citra Covid-19	<i>Data Training</i>	315
	2.	Citra Pneumonia	<i>Data Training</i>	315
	3.	Citra Tuberkulosis	<i>Data Training</i>	315
	4.	Citra Covid-19	<i>Data Test</i>	35
	5.	Citra Pneumonia	<i>Data Test</i>	35
	6.	Citra Tuberkulosis	<i>Data Test</i>	35
Pembagian Data 80% : 20%	No	Nama	Jenis	Jumlah
	1.	Citra Covid-19	<i>Data Training</i>	280
	2.	Citra Pneumonia	<i>Data Training</i>	280
	3.	Citra Tuberkulosis	<i>Data Training</i>	280
	4.	Citra Covid-19	<i>Data Test</i>	70
	5.	Citra Pneumonia	<i>Data Test</i>	70

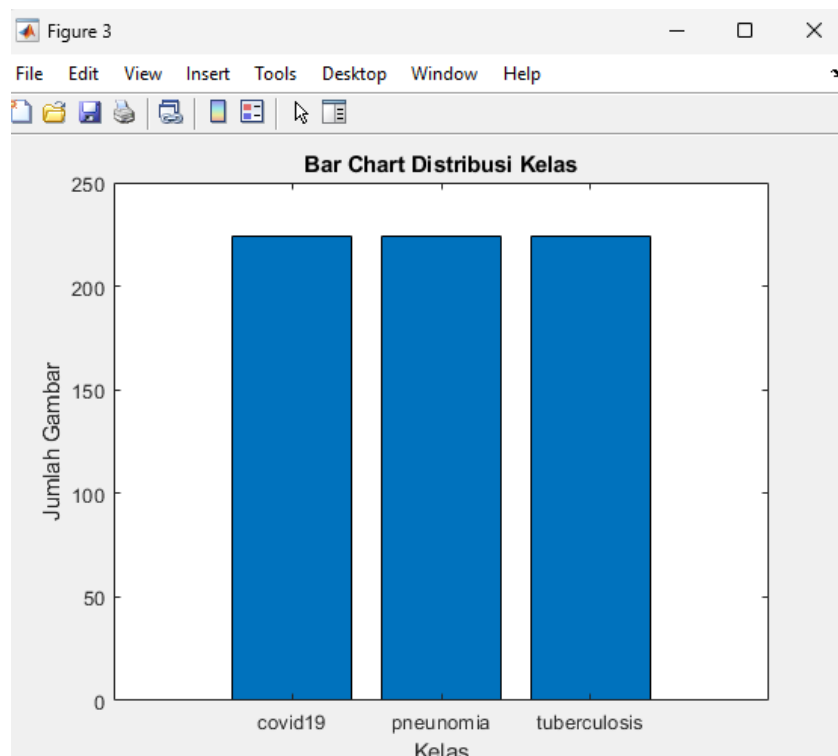
	6.	Citra Tuberkulosis	Data <i>Test</i>	70
Pembagian Data 70% : 30%	No	Nama	Jenis	Jumlah
	1.	Citra Covid-19	Data <i>Training</i>	245
	2.	Citra Pneumonia	Data <i>Training</i>	245
	3.	Citra Tuberkulosis	Data <i>Training</i>	245
	4.	Citra Covid-19	Data <i>Test</i>	105
	5.	Citra Pneumonia	Data <i>Test</i>	105
	6.	Citra Tuberkulosis	Data <i>Test</i>	105
Pembagian Data 60% : 40%	No	Nama	Jenis	Jumlah
	1.	Citra Covid-19	Data <i>Training</i>	210
	2.	Citra Pneumonia	Data <i>Training</i>	210
	3.	Citra Tuberkulosis	Data <i>Training</i>	210
	4.	Citra Covid-19	Data <i>Test</i>	140
	5.	Citra Pneumonia	Data <i>Test</i>	140
	6.	Citra Tuberkulosis	Data <i>Test</i>	140

3.3.4 Exploratory Data Analysis (EDA)

Analisis Data Eksplorasi (EDA) adalah proses analisis data terhadap suatu kumpulan data dengan tujuan untuk memahami karakteristik utama data tersebut. Pendekatan ini sering kali melibatkan penggunaan visualisasi untuk merangkum informasi penting dari data. EDA sangat penting dalam tahap awal analisis data karena membantu mengidentifikasi pola-pola menarik, anomali, dan hubungan antar variabel. Pemahaman yang mendalam tentang data ini akan memandu pemilihan teknik pemodelan dan prediksi yang tepat untuk diterapkan selanjutnya (James et al., 2013). Beberapa tahapan EDA yang dilakukan :

A. Bar Chart

Diagram batang (bar chart) menampilkan distribusi variabel kategori dengan menampilkan jumlah atau proporsi pengamatan yang termasuk dalam setiap kategori. Diagram batang terdiri dari batang-batang persegi panjang, di mana panjang setiap batang sebanding dengan jumlah atau proporsi pengamatan dalam kategori yang diwakilinya.



Gambar 3. 2 Bar Chart

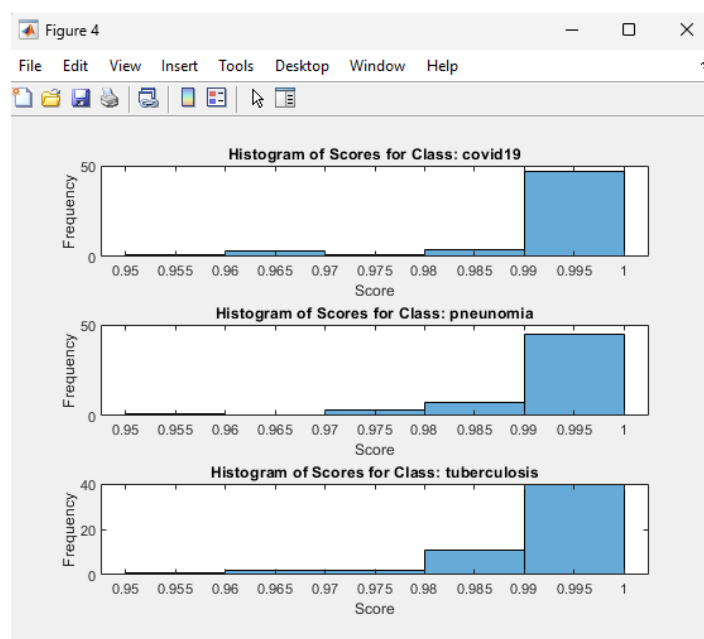
Pada gambar 3.2 menunjukkan diagram batang (bar chart) yang menggambarkan distribusi kelas dalam set pengujian (test set) untuk tiga kelas, yaitu COVID-19, pneumonia, dan tuberkulosis. Diagram

batang ini digunakan untuk memvisualisasikan jumlah sampel yang terdapat dalam setiap kelas pada set pengujian.

Pada sumbu x, terdapat tiga kelas yang diwakili oleh COVID-19, pneumonia, dan tuberkulosis. Sedangkan pada sumbu y, tertera jumlah sampel (count) dalam masing-masing kelas. Dari diagram batang ini, dapat dilihat bahwa distribusi sampel dalam setiap kelas pada set pengujian adalah sama, yaitu 224 sampel per kelas.

B. Histogram

Histogram adalah plot yang menampilkan distribusi skor prediksi untuk setiap kelas penyakit secara terpisah. Histogram ini memberikan informasi tentang seberapa sering skor prediksi tertentu muncul dalam setiap kelas, memungkinkan kita untuk melihat bentuk distribusi skor, pusat distribusi, dan adanya pencilan.



Gambar 3. 3 Histogram

Gambar 3.3 menunjukkan histogram skor prediksi untuk tiga kelas, yaitu COVID-19, pneumonia, dan tuberkulosis. Histogram ini digunakan untuk memvisualisasikan distribusi frekuensi dari skor prediksi yang dihasilkan oleh model untuk masing-masing kelas. Pada sumbu x dari setiap histogram, terdapat nilai skor prediksi yang berkisar dari 0,95 hingga 1, sedangkan pada sumbu y, tertera frekuensi dari skor prediksi tersebut. Setiap histogram merepresentasikan distribusi skor prediksi untuk satu kelas tertentu.

Histogram untuk kelas COVID-19 menunjukkan bahwa sebagian besar skor prediksi berada di kisaran 0,995 hingga 1. Hal ini menunjukkan bahwa model memiliki tingkat keyakinan yang sangat tinggi ketika mengklasifikasikan sampel sebagai COVID-19. Distribusi skor yang sangat terpusat di nilai yang tinggi menunjukkan konsistensi yang baik dalam prediksi model untuk kelas ini.

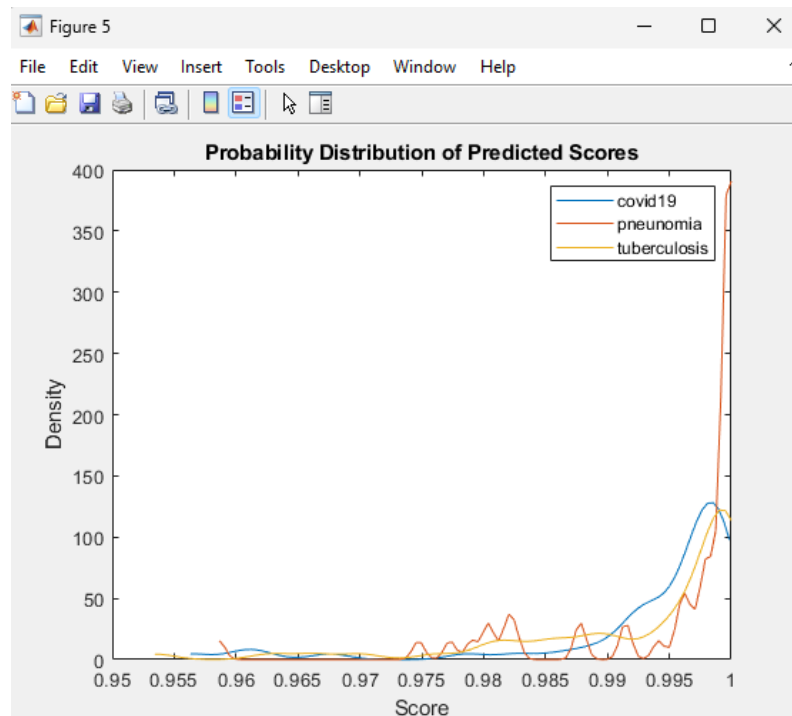
Histogram untuk kelas pneumonia juga menunjukkan sebagian besar skor prediksi berada di kisaran 0,99 hingga 1. Meskipun distribusi ini sedikit lebih lebar dibandingkan dengan kelas COVID-19, skor prediksi yang tinggi menunjukkan bahwa model umumnya cukup yakin ketika mengklasifikasikan sampel sebagai pneumonia. Namun, ada beberapa sampel dengan skor yang lebih rendah, yang mengindikasikan bahwa model memiliki sedikit ketidakpastian dalam beberapa kasus.

Histogram untuk kelas tuberkulosis menunjukkan pola yang serupa dengan dua kelas lainnya, dengan sebagian besar skor prediksi berada

di kisaran 0,99 hingga 1. Meskipun demikian, distribusi skor prediksi untuk kelas ini juga menunjukkan adanya beberapa sampel dengan skor yang lebih rendah, yang menunjukkan bahwa model mungkin memiliki tantangan yang lebih besar dalam mengklasifikasikan sampel sebagai tuberkulosis dibandingkan dengan dua kelas lainnya. Histogram-histogram ini menunjukkan bahwa model memiliki tingkat keyakinan yang tinggi dalam membuat prediksi untuk ketiga kelas, dengan sebagian besar skor prediksi berada di kisaran yang sangat tinggi (0,99 hingga 1).

C. Distribution Plots

Plot distribusi probabilitas skor prediksi menampilkan distribusi probabilitas yang terkait dengan prediksi yang dibuat oleh model klasifikasi. Dalam model klasifikasi, tujuannya adalah untuk memprediksi kelas atau kategori yang dimiliki oleh suatu observasi berdasarkan fitur-fiturnya. Plot distribusi probabilitas skor prediksi menunjukkan seberapa yakin model dalam memprediksi setiap kelas untuk setiap observasi.



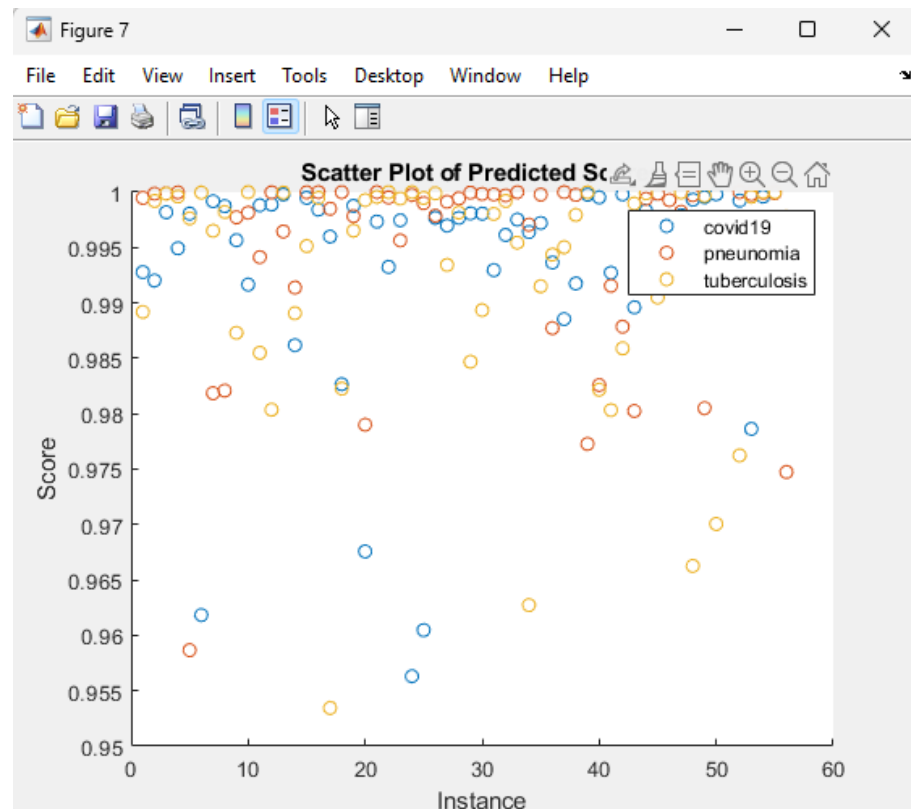
Gambar 3. 4 Distribution Plots

Plot distribusi probabilitas skor prediksi yang ditampilkan menunjukkan bagaimana model jaringan saraf tiruan (JST) memprediksi klasifikasi gambar medis menjadi tiga kategori: COVID-19, pneumonia, dan tuberkulosis. Sumbu X pada plot ini merepresentasikan skor probabilitas prediksi yang berkisar antara 0,95 hingga 1,0, yang mencerminkan tingkat kepercayaan model dalam membuat prediksi untuk setiap kelas. Sementara itu, sumbu Y menunjukkan kepadatan atau frekuensi relatif dari skor prediksi tersebut. Dari plot ini, terlihat bahwa sebagian besar skor prediksi terkonsentrasi sangat dekat dengan nilai 1,0 untuk ketiga kelas, yaitu COVID-19 (biru), pneumonia (oranye), dan tuberkulosis (kuning). Hal ini menunjukkan bahwa model memiliki tingkat kepercayaan yang

sangat tinggi terhadap prediksinya. Puncak kepadatan yang muncul di dekat skor 1,0. Tidak adanya tumpang tindih yang signifikan antara distribusi skor untuk ketiga kelas ini menunjukkan bahwa model dapat membedakan dengan sangat baik antara ketiga kategori penyakit tersebut.

D. Scatter Plot

Scatter plot adalah representasi grafis dari hubungan antara dua variabel kuantitatif. *Scatter plot* terdiri dari titik-titik, di mana setiap titik mewakili sebuah observasi. Posisi setiap titik ditentukan oleh nilai-nilai dari dua variabel yang diukur pada observasi tersebut.



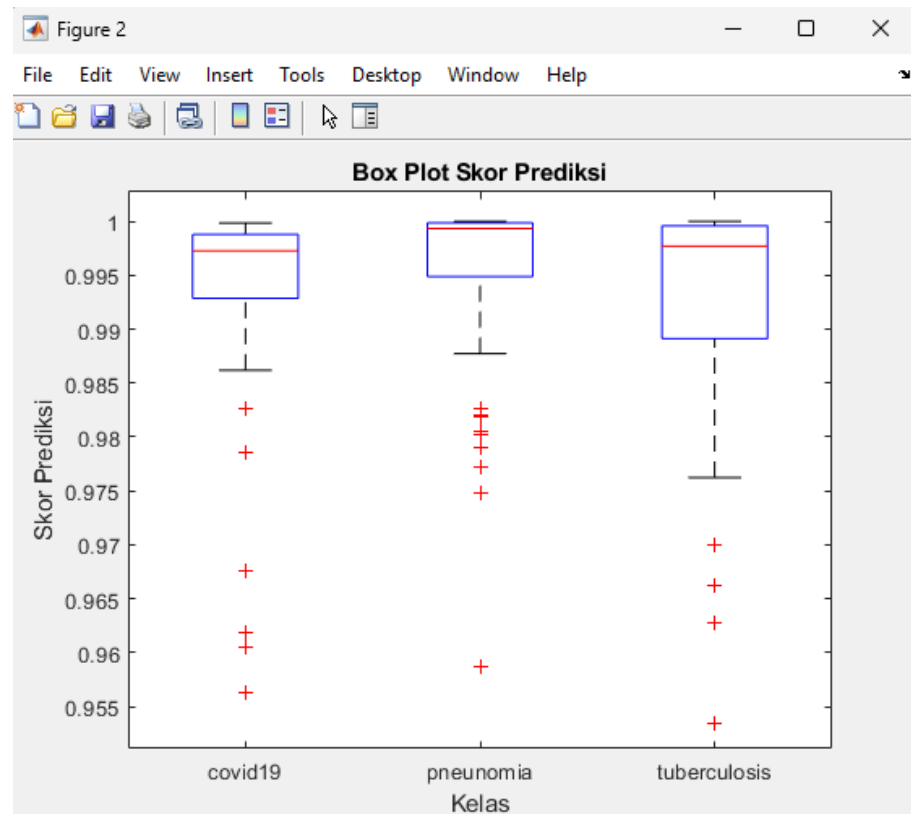
Gambar 3. 5 Scatter Plot

Scatter plot pada Gambar 3.5 menunjukkan distribusi skor prediksi dari model jaringan saraf tiruan (JST) untuk setiap *instance* dalam data uji. Sumbu Y pada plot ini merepresentasikan skor prediksi, yang berkisar dari 0,95 hingga 1,0, sedangkan sumbu X merepresentasikan nomor instance dari data uji. Setiap titik pada *scatter plot* ini menunjukkan skor prediksi untuk suatu instance tertentu, dengan warna yang berbeda untuk masing-masing kelas: biru untuk COVID-19, oranye untuk pneumonia, dan kuning untuk tuberkulosis.

Dari scatter plot ini, dapat diamati bahwa sebagian besar skor prediksi berkumpul di sekitar nilai 1,0 untuk ketiga kelas. Ini menunjukkan bahwa model memiliki tingkat kepercayaan yang sangat tinggi dalam klasifikasinya. Titik-titik yang berada di dekat nilai 1,0 mengindikasikan bahwa model memberikan skor prediksi yang hampir sempurna untuk banyak instance dalam data uji. Namun, terdapat beberapa instance dengan skor prediksi yang sedikit lebih rendah, meskipun masih dalam kisaran yang sangat baik (sekitar 0,95 hingga 0,99).

E. Box Plot

Boxplot menampilkan distribusi dari variabel kuantitatif. *Boxplot* terdiri dari kotak (box) yang membentang dari kuartil pertama hingga kuartil ketiga data.



Gambar 3. 6 Box Plot

Gambar 3.6 menunjukkan plot kotak (box plot) dari skor prediksi untuk tiga kelas: COVID-19, pneumonia, dan tuberkulosis. Plot kotak ini digunakan untuk memvisualisasikan distribusi skor prediksi yang dihasilkan oleh model untuk setiap kelas. Kotak dalam plot ini merepresentasikan rentang interkuartil (IQR) yang mencakup 50% dari data, di mana garis merah di dalam kotak menunjukkan median dari skor prediksi. Garis hitam yang memanjang dari kotak (whiskers) menggambarkan jangkauan skor prediksi yang berada dalam 1,5 kali IQR. Titik-titik merah yang berada di luar whiskers menunjukkan

outlier, yaitu skor prediksi yang jatuh jauh dari rentang yang diharapkan.

Median skor prediksi untuk kelas COVID-19 berada sangat dekat dengan nilai 1, yang menunjukkan tingkat keyakinan yang sangat tinggi dari model dalam membuat prediksi untuk kelas ini. Distribusi skor prediksi untuk kelas COVID-19 terlihat sangat rapat, menunjukkan konsistensi yang tinggi dalam prediksi model.

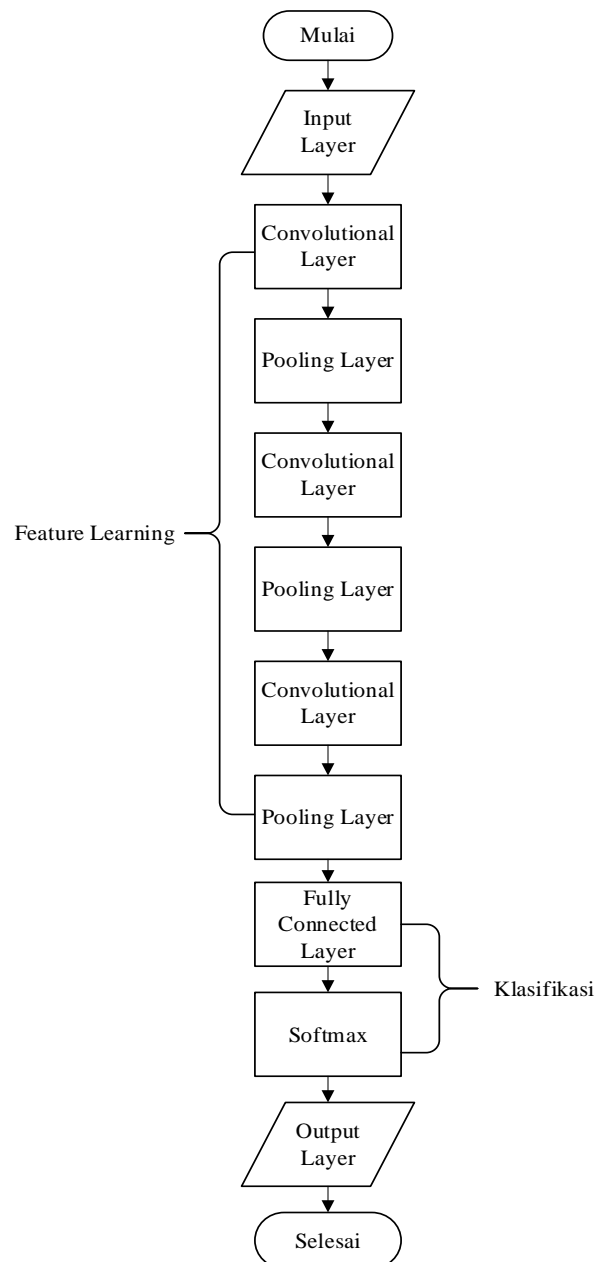
Median skor prediksi untuk kelas pneumonia juga tinggi dan mendekati nilai 1, namun sedikit lebih rendah dibandingkan dengan kelas COVID-19. Rentang interkuartil untuk pneumonia lebih lebar, mengindikasikan bahwa ada lebih banyak variabilitas dalam keyakinan model saat membuat prediksi untuk kelas ini. Jumlah outlier yang lebih banyak dibandingkan dengan kelas COVID-19 menunjukkan bahwa prediksi untuk pneumonia kurang konsisten.

Median skor prediksi untuk kelas tuberkulosis juga tinggi, namun distribusi skor prediksi memiliki rentang yang paling lebar di antara ketiga kelas. Hal ini mengindikasikan bahwa model memiliki tingkat keyakinan yang lebih rendah dan lebih bervariasi dalam prediksi untuk kelas tuberkulosis. Banyaknya outlier juga menunjukkan adanya variabilitas yang signifikan dalam prediksi untuk kelas ini.

3.3.5 Modelling

Perancangan model dengan membuat arsitektur dari model *Convolutional Neural Network* yang akan digunakan dalam proses klasifikasi.

Citra yang dimasukkan akan di proses melalui proses konvulsi dan proses pooling pada tahapan *feature learning*. Setelah melewati tahapan *feature learning*, setelah melewati proses *feature learning* citra akan memasuki tahapan *convulational neural network* yaitu proses klasifikasi. *Flowchart* dari proses perancangan model CNN :



Gambar 3. 7 Flowchart CNN

a. *Input Layer*

Tahap pertama yang dilakukan adalah memasukkan data citra rontgen ke input layer. Pada layer ini citra rontgen dilakukan proses *cropping* untuk mendapatkan ukuran persegi, lalu citra rontgen dikonversi menjadi grayscale dengan ukuran pixel 128x128.

b. *Convolutional Layer*

Pada proses konvolusi menggunakan 3 layer konvolusi dengan 16,32,64 filter dan penambahan *padding* masing-masing 1 serta *stride* sebesar 2. Untuk memperoleh objek dalam suatu gambar, diperlukan proses pemisahan antara objek dan latar belakang. Dalam penelitian ini, fungsi aktivasi ReLU dipilih untuk menentukan apakah suatu neuron dalam jaringan saraf bersifat aktif atau tidak. Hal ini bertujuan agar hanya neuron yang terkait dengan objek yang menjadi fokus perhatian.

c. *Pooling Layer*

Pooling layer berfungsi untuk mengurangi dimensi spasial dari representasi data, memungkinkan invariansi terhadap pergeseran kecil dalam gambar, menonjolkan fitur penting dalam suatu region, dan secara efisien mempercepat proses pelatihan dan inferensi model. Pada penelitian ini *Pooling layer* yang digunakan yaitu *Max Pooling* dengan ukuran kernel 2x2.

d. Fully Connected Layer

Dalam *fully connected layer* ini, setiap neuron di setiap lapisan akan memiliki koneksi penuh dengan neuron di lapisan sebelumnya dan sesudahnya, terutama setelah melalui proses *convolutional layer* dan *pooling layer* terakhir. Pada *Softmax Layer*, hasil perhitungan dari *fully connected layer* dihitung probabilitas untuk setiap kelasnya. Nilai *Softmax* mulai dari 0 hingga 1. Nilai tertinggi pada hasil *softmax* dari semua kelas akan menjadi output dan keputusan sistem.

e. Output Layer

Output layer menandai akhir dari proses klasifikasi dengan menetapkan prediksi kelas berdasarkan nilai probabilitas. Pada tahap ini, prediksi kelas ditentukan oleh nilai probabilitas tertinggi di antara semua kelas yang dihasilkan.

Selain itu, penelitian ini juga melakukan berbagai uji coba untuk mengoptimalkan model CNN. Uji coba tersebut meliputi variasi jumlah *epoch* (10, 15, 20, 25, dan 30), pembagian data antara data latih dan data uji (90% : 10%, 80% : 20%, 70% : 30%, dan 60% : 40%), ukuran mini-batch (8, 16, 32, dan 64), serta penggunaan tiga jenis *Optimizer* yang berbeda (Adam, SGDM, dan RMSprop). Tujuan dari uji coba ini adalah untuk menemukan kombinasi *hyperparameter* terbaik yang menghasilkan prediksi klasifikasi tertinggi.

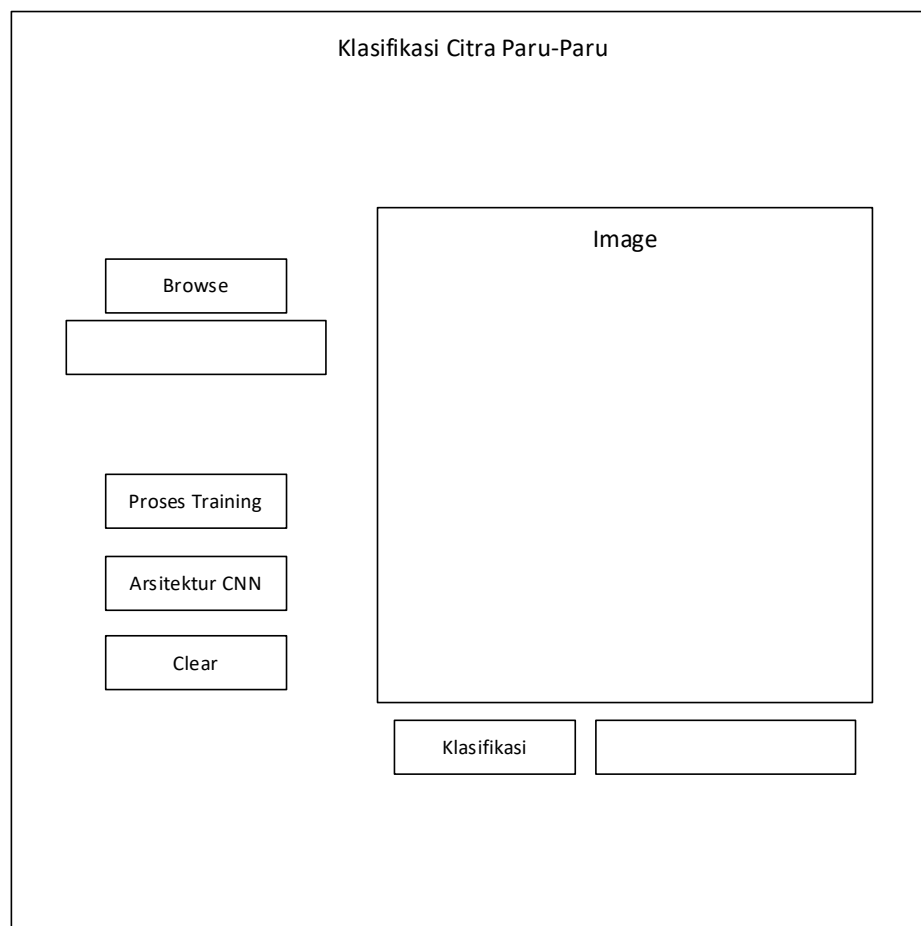
3.3.6 Deployment

Tahapan kelima dalam penelitian ini adalah Evaluasi dan Penyebaran (*Deployment*). Tahap ini terdiri dari dua bagian utama. Pertama, model yang telah dilatih pada tahap sebelumnya dievaluasi kinerjanya menggunakan data uji (*data test*). Data uji ini adalah bagian dari dataset yang tidak digunakan selama pelatihan model, sehingga dapat memberikan gambaran yang lebih objektif tentang seberapa baik model dapat menggeneralisasi pada data baru. Metrik utama yang digunakan untuk evaluasi adalah prediksi, yang mengukur seberapa sering model dapat mengklasifikasikan citra rontgen paru-paru dengan benar ke dalam kelas penyakit yang sesuai.

Bagian kedua dari tahap ini adalah penyebaran (*deployment*). Model terbaik, yaitu model yang mencapai akurasi tertinggi selama proses pengujian, dipilih untuk diintegrasikan ke dalam sebuah antarmuka pengguna grafis (*Graphical User Interface/GUI*). Antarmuka ini dirancang agar mudah digunakan, memungkinkan pengguna untuk mengunggah citra rontgen paru-paru yang ingin diklasifikasikan. Sistem kemudian akan memproses citra tersebut menggunakan model CNN yang telah dilatih dan menampilkan hasil klasifikasi, yaitu jenis penyakit paru-paru yang terdeteksi (Pneumonia, Tuberkulosis, atau COVID-19). Dengan demikian, sistem ini dapat berfungsi sebagai alat bantu diagnosis bagi tenaga medis, memberikan hasil klasifikasi yang cepat dan akurat berdasarkan citra rontgen paru-paru.

3.4 Rancangan Antarmuka

Berdasarkan hasil pengujian yang dilakukan dengan menggunakan metode *Convolutional Neural Network* (CNN), langkah selanjutnya adalah mengimplementasikan algoritma tersebut ke dalam bentuk *Graphical User Interface* (GUI) yang interaktif. Adapun rancangan antarmuka pengguna (user interface) yang akan digunakan dalam penelitian ini akan dirancang sedemikian rupa sehingga mudah digunakan dan dipahami oleh pengguna, serta mampu memvisualisasikan hasil dari metode CNN secara efektif.



Gambar 3. 8 Rancangan Antarmuka

Rancangan antarmuka pengguna (GUI) untuk klasifikasi citra paru-paru ini dibangun menggunakan fitur GUIDE (Graphical User Interface Development Environment) yang tersedia dalam MATLAB. Fitur ini memungkinkan pembuatan antarmuka pengguna secara visual dan interaktif tanpa memerlukan penulisan kode secara manual. GUI ini memungkinkan pengguna untuk:

1. **Memuat Citra Paru-Paru** : Tombol "Browse" di bagian atas memungkinkan pengguna memilih citra paru-paru dari sistem lokal mereka untuk dianalisis.
2. **Menampilkan Citra** : Area "Image" akan menampilkan citra paru-paru yang telah dimuat, memberikan visualisasi kepada pengguna.
3. **Melatih Model CNN** : Tombol "Proses Training" akan memulai proses pelatihan model *Convolutional Neural Network* (CNN) untuk klasifikasi citra paru-paru.
4. **Menampilkan Arsitektur CNN** : Tombol "Arsitektur CNN" akan menampilkan visualisasi dari struktur model CNN yang digunakan dalam proses klasifikasi.
5. **Membersihkan Tampilan** : Tombol "Clear" akan membersihkan area tampilan citra dan hasil lainnya.
6. **Menampilkan Klasifikasi** : Tombol "Klasifikasi" berfungsi untuk memulai proses klasifikasi citra paru-paru yang telah dimuat sebelumnya. Hasil dari proses klasifikasi ini akan ditampilkan di area kosong yang terletak di samping tombol "Klasifikasi".