

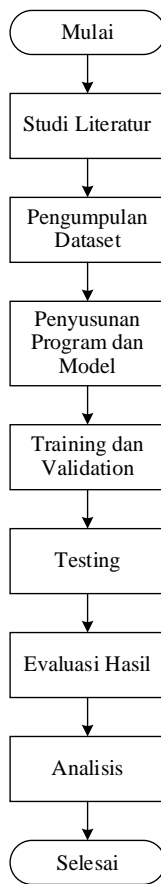
BAB III

METODE PENELITIAN

Pada bab 3 ini akan membahas terkait dengan *tools* yang digunakan untuk menunjang penelitian yaitu alur penelitian, dan rancangan dari sistem.

3.1. ALUR PENELITIAN

Pada penelitian ini dimulai dari studi literatur, pengumpulan data, penyusunan program, proses *training and validation*, proses *testing*, evaluasi, analisis hasil hingga kesimpulan penelitian yang ditunjukkan pada blok diagram dibawah ini.



Gambar 3.1 Tahapan Penelitian

3.2. ALAT YANG DIGUNAKAN

Pada subbab ini membahas tentang perangkat keras dan perangkat lunak yang digunakan pada penulisan penelitian ini.

3.2.1. Perangkat Keras (*Hardware*)

Perangkat keras yang dipakai dalam penelitian ini hanya komputer atau laptop untuk membuat dan menjalankan program dengan spesifikasi sebagai berikut:

1. AMD A4
2. *Operating System* Windows 10
3. RAM 4 GB

3.2.2. Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam penelitian ini diantaranya:

1. Jupyter Notebook

Penelitian ini menggunakan *software* Jupyter Notebook untuk merancang program, dan menjalankan program untuk Klasifikasi Citra X-Ray Untuk Mendeteksi Penyakit Paru-Paru Menggunakan *Convolutional Neural Network*. Beberapa tahapan dalam penelitian ini seperti penyusunan program, *training and validation*, dan proses *testing* dilakukan menggunakan *software* Jupyter Notebook.

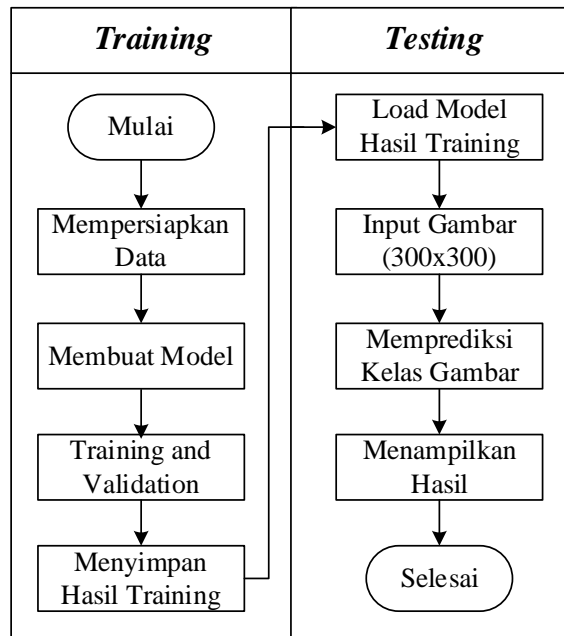
2. Bahasa Pemrograman Python

Bahasa yang digunakan dalam pembuatan program Klasifikasi Citra X-Ray Untuk Mendeteksi Penyakit Paru-Paru Menggunakan *Convolutional Neural Network* ini adalah bahasa pemrograman python. Lebih spesifiknya adalah python versi 3 dengan menggunakan beberapa library untuk membentuk sebuah program yang dapat berjalan dengan baik.

3.3. RANCANGAN SISTEM

Bentuk dari rancangan sistem yang dibuat menggunakan *flowchart* yang dapat dilihat pada gambar 3.2. Pada diagram *flowchart* tersebut menjelaskan rancangan sistem dimulai dari mempersiapkan data hingga menjadi sebuah dataset, kemudian dilakukan proses *training and validation* terhadap dataset, dan menyimpan model hasil dari proses *training and validation*. Selanjutnya masuk ke

tahap *testing* dengan citra yang di *resize* menjadi 300x300 dan diuji menggunakan model dari hasil *training and validation*. Kemudian sistem akan menampilkan hasil dari proses klasifikasi terhadap citra yang dimasukkan.



Gambar 3.2 *Flowchart* Sistem Klasifikasi

3.3.1. Dataset

Dataset yang digunakan dalam penelitian ini merupakan data citra hasil pemeriksaan x-ray paru-paru yang berjumlah 2000 data dan terdiri atas 4 kelas, yaitu Normal, Pneumonia, Tuberkulosis, dan Covid-19. Data tersebut merupakan data sekunder yang didapatkan dari sumber publik, yaitu Kaggle. Kaggle merupakan situs yang menjadi tempat bagi *data scientist* untuk membuat model tertentu dan dapat membagikannya kepada peneliti lainnya. Selain dataset utama, terdapat pula dataset tambahan yang jumlahnya hanya sebanyak 40 citra untuk menguji model yang telah selesai dilatih.

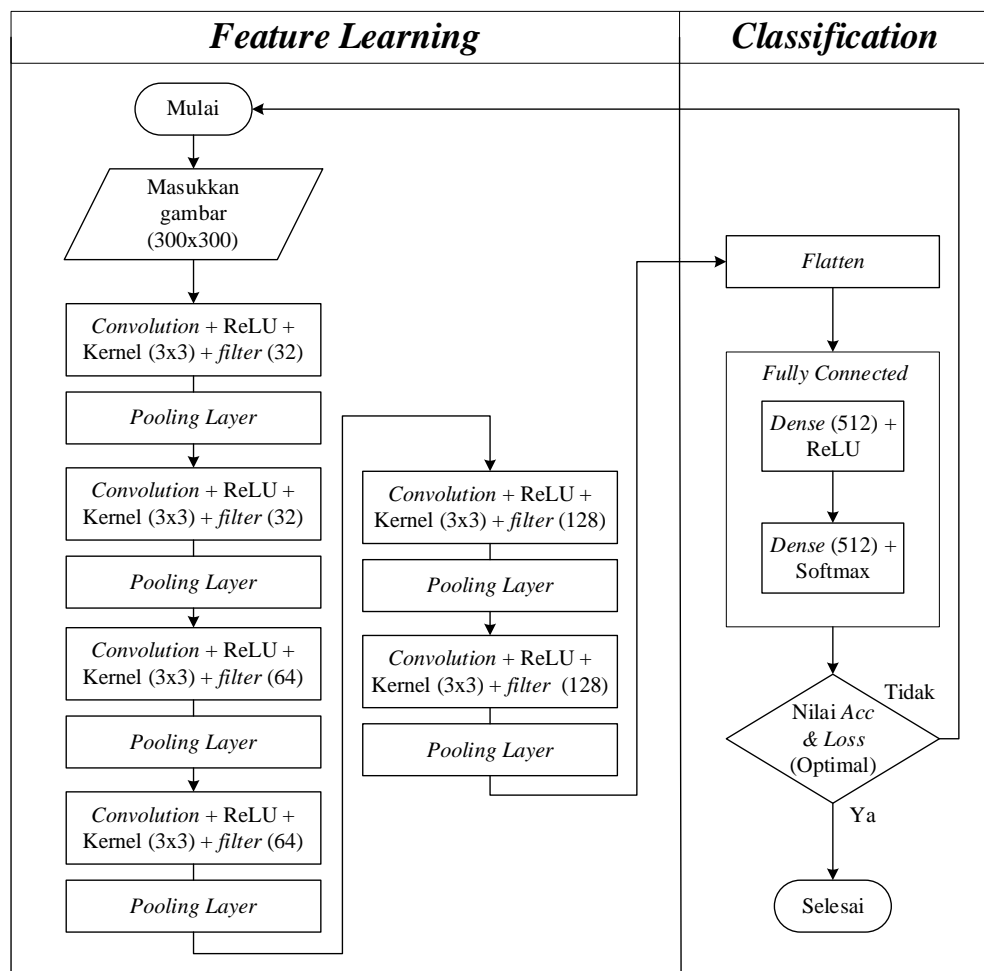
3.3.2. Data Preprocessing

Pada tahapan ini, teknik yang dilakukan adalah augmentasi data dengan cara *shear*, *zoom*, dan *horizontal flip*. Tujuan dilakukannya augmentasi data adalah untuk menambah variasi data yang akan digunakan untuk proses *training and validation*. Pada tahap *preprocessing* ini, keluaran yang dihasilkan adalah

data citra dengan ukuran yang sama, yaitu 300x300 piksel. Selanjutnya, citra yang telah melewati tahapan *preprocessing* akan dilatih menggunakan arsitektur CNN yang akan dibangun.

3.3.3. Arsitektur CNN

Arsitektur CNN yang digunakan pada penelitian ini dapat dilihat pada gambar 3.3 yang terdiri atas 2 tahapan. Tahapan pertama adalah *feature learning* dengan 6 *convolution* dan 6 *pooling layer*. Tahapan kedua adalah *classification* dengan *flatten layer*, dan *fully connected layer* yang terdiri atas 2 *dense layer*.



Gambar 3.3 Rancangan Arsitektur CNN

3.4. RANCANGAN PENELITIAN

Pada sub bab ini akan membahas tentang rancangan yang dilakukan pada penelitian ini, yaitu proses pelatihan model dan pengujian sistem klasifikasi.

3.4.1. Pelatihan Model

Pada penelitian ini hanya dibangun 1 arsitektur CNN yang dapat dilihat pada gambar 3.3, namun akan terdapat 2 skema pengujian yang disiapkan dalam proses pengujian arsitektur atau model tersebut. Masing-masing skema dibedakan oleh persentase pembagian dataset menjadi data latih dan data uji. Skema-skema tersebut dibedakan oleh persentase pembagian dataset untuk proses latih dan proses uji seperti yang terlihat pada Tabel 3.1 dibawah. Pada skema pertama, pembagian dataset adalah 80%:20%, dimana 80% akan digunakan sebagai data latih dan 20% sisanya digunakan sebagai data validasi. Kemudian pada skema kedua, pembagian dataset adalah 60%:40%, dimana 60% digunakan sebagai data latih dan 40% digunakan sebagai data validasi. Kemudian proses pelatihan data dari kedua skema tersebut akan menggunakan *epoch* sebesar 100.

Tabel 3.1 Skema pembagian data

| Skema | Data Latih | Data Validasi | Keterangan |
|-------|------------|---------------|---|
| 1 | 80% | 20% | 1600 data sebagai data latih 400 data sebagai data validasi. |
| 2 | 60% | 40% | 1200 data sebagai data latih 800 data sebagai data validasi. |

Sehingga dalam proses pengujian akan dilakukan 2 kali proses pelatihan untuk mendapatkan perbandingan hasil dari pengujian kedua skema tersebut. Dalam setiap pengujian model akan menghasilkan nilai *loss*, *accuracy*, *validation loss*, dan *validation accuracy* serta grafik dari tiap iterasi yang berjalan selama proses *training and validation*. Grafik yang dihasilkan akan menunjukkan apakah selama proses *training and validation* terjadi kondisi *underfitting*, *optimal*, atau *overfitting*. Semua hasil dari proses ini akan disimpan sebagai model yang akan

digunakan pada proses selanjutnya, yaitu prediksi atau klasifikasi terhadap data baru yang bukan bagian dari *training and validation set*.

3.4.2. Pengujian Data Baru

Pada proses ini akan memanfaatkan model yang telah dilatih untuk melakukan klasifikasi data. Data untuk pengujian ini juga diambil dari sumber yang sama dengan dataset yang dilatih, yaitu Kaggle. Data-data ini juga terdiri atas 4 kelas yang sama, yaitu Normal, Covid-19, Penumoniam, dan Tuberkulosis. Tetapi masing-masing kelas tersebut hanya terdiri dari 10 citra saja, sehingga keseluruhan data ini berjumlah 40 citra saja. Proses pengujian ini akan dilakukan dengan 2 cara, yaitu memasukkan tiap citra saja dan langsung memasukkan 1 folder yang berisi 10 citra. Dari pengujian ini hanya akan menampilkan nilai diantara 0 hingga 1 untuk setiap gambar maupun folder yang dimasukkan. Hasil dari pengujian ini juga dapat digunakan untuk mengevaluasi hasil dari pelatihan yang telah dilakukan.