

BAB III

METODOLOGI PENELITIAN

3.1 Subyek dan Obyek Penelitian

Subyek dari penelitian ini adalah penyakit paru-paru akibat Covid 19. Sedangkan obyek dari penelitian ini adalah citra *x-ray* paru-paru yang terkena Covid 19 dan normal.

3.2 Alat dan Bahan Penelitian

3.2.1 Alat Penelitian

Alat yang digunakan pada penelitian ini terdiri dari *hardware* dan *software*. Berikut spesifikasi *hardware* yang digunakan:

1. *Processor AMD Ryzen 5 4500U 2.3 GHz base clock*
2. *16 GB DDR4-2667 SDRAM*
3. *AMD Radeon™ Graphics*
4. *512 GB PCIe® NVMe™ M.2 SSD*

Adapun *software* yang digunakan antara lain:

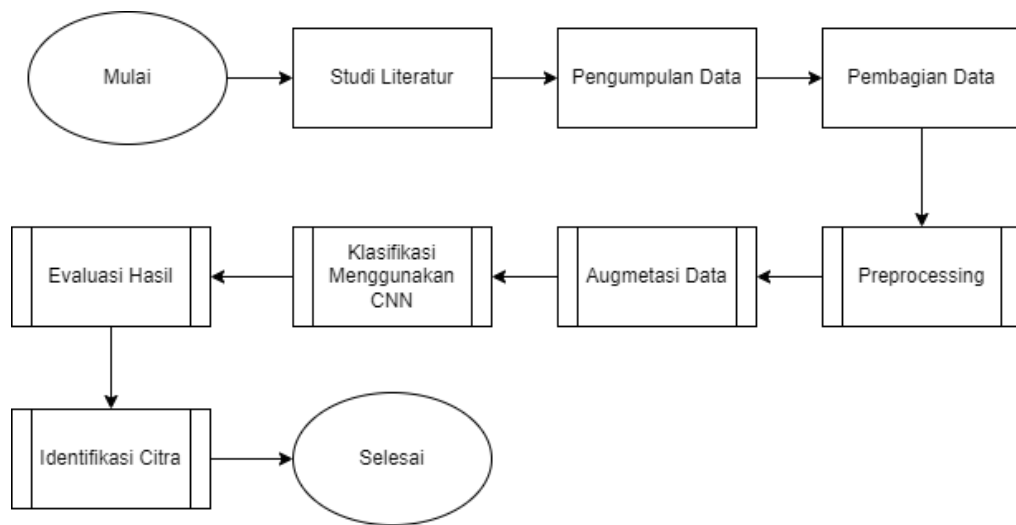
1. *Microsoft Word 2019*
2. *Anaconda Navigator*
3. *Jupyter Notebook*

3.2.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah dataset citra *x-ray* paru-paru. Jumlah dataset yang digunakan pada penelitian ini yaitu 3166 data yang terdiri dari 1583 data citra *x-ray* pneumonia dan 1583 data citra *x-ray* normal. Kumpulan dataset citra *x-ray* diperoleh dari penelitian [29].

3.3 Diagram Alir Penelitian

Alur penelitian yang akan dilalui dalam penelitian ini digambarkan seperti pada Gambar 3.1 dibawah ini:



Gambar 3.1 Diagram Alir Penelitian

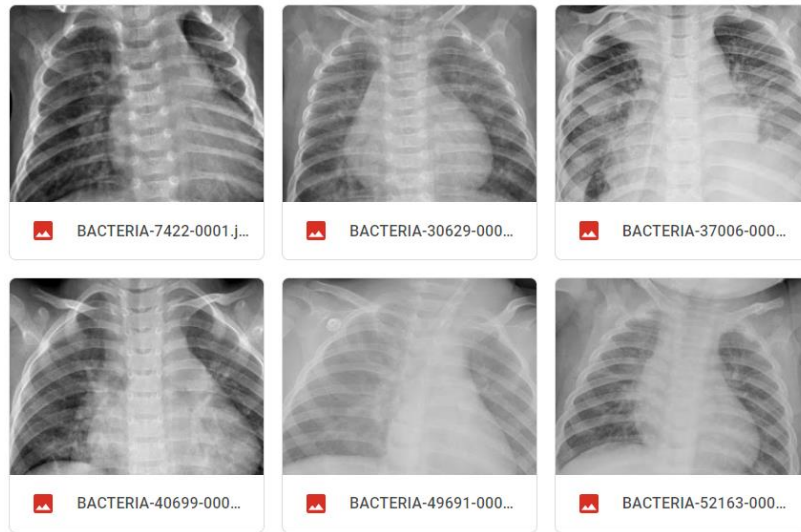
Berdasarkan Gambar 3.1 dapat dijelaskan alur tahapan penelitian yang akan dilakukan yaitu:

3.3.1 Studi Literatur

Studi Literatur dilakukan dengan mencari sumber yang berkaitan dengan penelitian baik dari jurnal, skripsi, buku, *report*, atau artikel terkait yang relevan.

3.3.2 Pengumpulan Data

Pada penelitian ini dataset berupa citra *x-ray* pneumonia dan normal, dengan jumlah total dataset yaitu 3166 data citra. Sampel dataset dapat dilihat pada Gambar 3.2 berikut.



Gambar 3.2 Sampel Dataset

3.3.3 Pembagian Data

Tahap awal yang dilakukan pada penelitian ini yaitu pembagian data. Pada penelitian ini dataset terdiri dari dua jenis data, yaitu data *train* dan data *test*, seluruh dataset dibagi ke dalam dua kelas tersebut dengan komposisi data *train* sebanyak 80 % dan data *test* sebanyak 20 % [7]. Pembagian dataset dapat ditunjukkan pada Tabel 3.1 berikut.

Tabel 3.1 Pembagian Dataset

| No | Kategori | Data train | Data test |
|-------|-----------|------------|-----------|
| 1. | Pneumonia | 1266 | 317 |
| 2. | Normal | 1266 | 317 |
| Total | | 2532 | 634 |

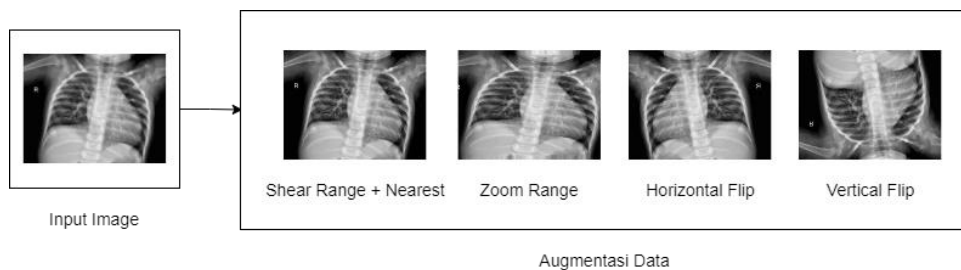
3.3.4 Preprocessing

Proses *preprocessing* pada penelitian ini bertujuan untuk meningkatkan hasil akurasi. Terdapat tiga metode yang dilakukan dalam *preprocessing* pada penelitian ini yaitu *Histogram Equalization* (HE), *Adaptive Histogram Equalization* (AHE), *Contrast Limited Adaptive Histogram Equalization* (CLAHE).

3.3.5 Augmentasi Data

Tahap selanjutnya setelah *preprocessing* yaitu augmentasi data. Augmentasi data adalah proses untuk memperbanyak data pelatihan sehingga terjadinya *overfitting* akan diminimalisir. Augmentasi data yang dilakukan pada penelitian ini yaitu *shear range*, *zoom range*, *fill mode*, *horizontal flip*, *vertical flip*, *rescale*. Tahapan *shear range* dilakukan dengan metode *shear transformation*, cara kerjanya yaitu dengan merotasi citra sesuai dengan derajat nilai yang sudah ditentukan.

Tahapan *horizontal flip* dilakukan dengan memperbanyak data dengan cara memutar gambar secara *horizontal* sebanyak 90 derajat. Sedangkan *vertical flip* dilakukan dengan cara memutar gambar secara flip sebanyak 90 derajat. Tahapan *zoom range* merupakan tahapan untuk memperbesar citra dengan skala yang sudah ditentukan dari citra asli [30]. Adapun *rescale* merupakan tahap merubah nilai *pixel* dari [0,255] menjadi [0,1]. Perubahan nilai *pixel* bertujuan untuk mengurasi *loss* dan meningkatkan akurasi pada saat *training* [7]. Hasil sampel yang didapat pada proses augmentasi dapat dilihat pada gambar 3.3 berikut.



Gambar 3.3 Hasil Sampel Augmentasi Data

3.3.6 Klasifikasi Menggunakan CNN

Tahap selanjutnya setelah augmentasi data yaitu membangun model arsitektur CNN yang akan dirancang untuk mengolah citra kemudian mengkategorikan hasil citra tersebut. Proses pembangunan model arsitektur CNN yang dilakukan antara lain [31]:

1. Penentuan *filter*, *pool size*, *kernel size*, *padding*
Pada penelitian ini menggunakan *filter* 32, 64, dan 128. Menggunakan *pool size* 2x2, dan *kernel size* 3x3. Sedangkan *padding* bernilai *same*.
2. Penentuan fungsi aktivasi
Terdapat dua fungsi aktivasi yang digunakan pada penelitian ini yaitu fungsi aktivasi *relu* dan *sigmoid*.
3. Penentuan *optimizer*
Optimizer merupakan sebuah algoritma untuk menentukan weight yang optimal. Pada penelitian ini menggunakan *optimizer* Adam.
4. Penentuan *batch size*
Pada penelitian ini menggunakan nilai *batch size* sebesar 32. Artinya, seluruh data citra pada proses *train* akan dibagi menjadi beberapa ukuran dengan isinya masing-masing 32 citra.
5. Penentuan *epoch*
Epoch merupakan jumlah iterasi yang akan digunakan untuk mengulang proses pelatihan. Semakin banyak *epoch* maka akan semakin meningkatkan hasil dari proses pelatihan. Pada penelitian ini menggunakan 96, 64, dan 32 *epochs*.
6. Konfigurasi *layer*
Terdapat tiga tahapan pokok dalam metode *Convolutional Neural Network* yaitu *input*, *feature extraction*, dan *classification*. Proses diawali dengan citra *input* yang berukuran 299x299x3 akan diproses ke dalam *Convolutional Layer* pertama + *Relu* dengan *filter* 32, *kernel size* 3x3. Pada *pooling layer* pertama menggunakan *pool size* berukuran 2x2.
Hasil dari *Convolutional Layer* pertama + *Relu* akan menjadi input untuk *Convolutional Layer* kedua + *Relu* dengan *filter* 64, *kernel size* 3x3. Pada *pooling layer* kedua menggunakan *pool size* berukuran 2x2.

Hasil dari *Convolutional Layer* kedua + *Relu* akan menjadi input untuk *Convolutional Layer* ketiga + *Relu* dengan *filter* 64, *kernel size* 3x3. Pada *pooling layer* ketiga menggunakan *pool size* berukuran 2x2.

Hasil dari *Convolutional Layer* ketiga + *Relu* akan menjadi input untuk *Convolutional Layer* keempat + *Relu* dengan *filter* 128, *kernel size* 3x3. Pada *pooling layer* keempat menggunakan *pool size* berukuran 2x2. Setelah semua proses selesai maka akan masuk ke tahap selanjutnya yaitu *classification*, tahap ini merupakan tahap akhir dalam metode *Convolutional Neural Network*.

3.3.7 Evaluasi Hasil

Evaluasi dilakukan untuk menguji presentase keakuratan sebuah model yang sudah dibuat. Evaluasi pada penelitian ini menggunakan metode *ROC Curve* dan *Confusion Matrix*. *ROC Curve* digunakan untuk mengevaluasi suatu model berdasarkan luas kurva. Luas kurva diperoleh dari hasil perbandingan TPR (*True Positive Rate*) dan FPR (*False Positive Rate*). TPR dan FPR dihitung berdasarkan persamaan 3.2 dan 3.3 berikut [32].

$$TPR = \frac{TP}{(TP + FN)} \quad (3.2)$$

$$FPR = \frac{FP}{(FP + TN)} \quad (3.3)$$

Keterangan:

TPR = *True Positive Rate*

FPR = *False Positive Rate*

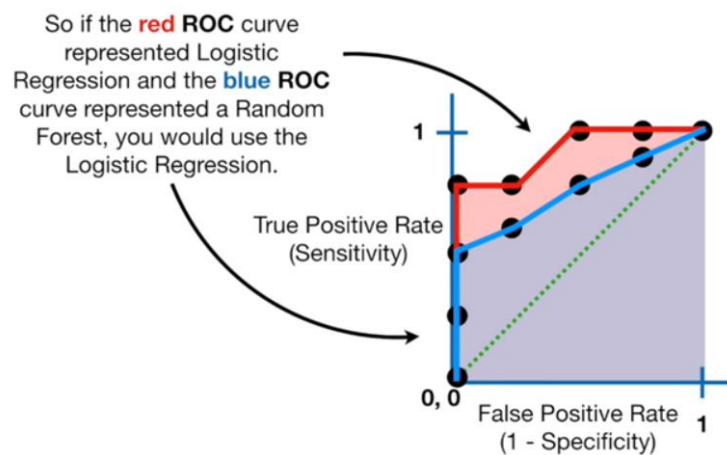
TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

Nilai luas kurva atau AUC (*Area Under Curve*) dapat digunakan untuk mempermudah melihat perbandingan kemampuan model dengan model lainnya. Semakin besar nilai AUC maka semakin bagus kemampuan suatu model. Hal itu dikarenakan model memiliki nilai TPR lebih tinggi dan atau nilai FPR yang lebih rendah [33]. Visualisasi AUC pada ROC Curve dapat dilihat pada gambar 3.4 berikut.



Gambar 3.4 AUC pada ROC Curve [33]

Hasil *Confusion Matrix* yaitu nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Adapun tabel dari *Confusion Matrix* dapat dijelaskan pada Tabel 3.2 berikut.

Tabel 3.2 Confusion Matrix

| Aktual | Prediksi | |
|--------|----------|-------|
| | True | False |
| True | TP | FN |
| False | FP | TN |

Pada tabel 3.2 terdapat empat parameter yaitu, TP (*True Positive*) adalah parameter untuk jumlah prediksi positif dari kelas positif aktual. TN (*True Negative*) adalah parameter untuk jumlah prediksi negatif dari kelas

negatif aktual. FP (*False Positive*) adalah parameter untuk jumlah prediksi positif dari kelas negatif aktual. FN (*False Negative*) adalah parameter untuk jumlah prediksi negatif dari kelas positif aktual. Perhitungan dalam *Confusion Matrix* dapat ditentukan dari persamaan 3.3 sampai dengan persamaan 3.6 berikut [34].

$$(A) = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (3.3)$$

$$(P) = \frac{(TP)}{(TP + FP)} \quad (3.4)$$

$$(R) = \frac{(TP)}{(TP + FN)} \quad (3.5)$$

$$f1\text{-score} = \frac{(2 \times P \times R)}{(P + R)} \quad (3.6)$$

Keterangan:

A = *Accuracy*

P = *Precision*

R = *Recall*

F1-Score = *f1-score*

TP = *True Positive*

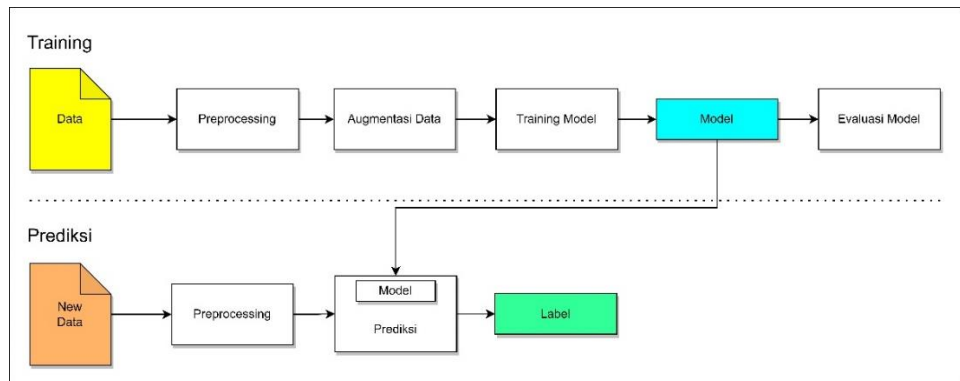
TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

3.3.8 Identifikasi Citra

Pada tahap ini dilakukan identifikasi citra *x-ray* menggunakan untuk mengetahui hasil kategori apakah citra akan dikategorikan sebagai Covid 19 atau normal. Alur proses identifikasi pada penelitian ini dapat dilihat pada gambar 3.5.



Gambar 3.5 Alur Proses Identifikasi

Berdasarkan gambar 3.5 identifikasi dilakukan menggunakan data baru yang tidak digunakan pada saat *training* model. Selanjutnya data baru akan diproses menggunakan *preprocessing*. Hasil *preprocessing* kemudian diidentifikasi menggunakan model yang sudah disimpan pada saat *training* model. Hasil identifikasi berupa label yang menandakan apakah citra termasuk Covid 19 atau normal.