



UNIVERSITAS INDONESIA

**KLASIFIKASI KASUS POSITIF DAN NEGATIF COVID-19 DARI HASIL
RONTGEN PARU-PARU DENGAN CONVOLUTIONAL NEURAL
NETWORK**

TUGAS MAKALAH

**CARLES OCTAVIANUS (20068568613)
AGUSTINUS BRAVY TETUKO OMPUSUNGGU (2006521300)
RENDI HARTADI (2006533383)**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI MATEMATIKA
DEPOK
JUNI 2022**

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa. Atas rahmat dan hidayah-Nya, kami dapat menyelesaikan tugas makalah yang berjudul "Klasifikasi Kasus Positif dan Negatif Covid-19 dari Hasil X-ray Paru-paru dengan Convolutional Neural Network" dengan tepat waktu.

Makalah ini disusun untuk memenuhi tugas Mata Kuliah Sains Data kurikulum 2020. Selain itu, makalah ini bertujuan menambah wawasan kami dan para pembaca mengenai klasifikasi gambar dengan *Convolutional Neural Network (CNN)*.

Penulis mengucapkan terima kasih kepada Bu Bevina selaku dosen pengajar mata kuliah sains data. terima kasih juga kami berikan kepada semua pihak yang telah membantu penyelesaian makalah ini.

Penulis menyadari makalah ini masih jauh dari sempurna. Oleh sebab itu, saran dan kritik yang membangun diharapkan demi kesempurnaan makalah ini.

Depok, 8 Juni 2022

Kelompok 5

ABSTRAK

Pandemi penyakit COVID-19 telah menimbulkan perlunya peningkatan teknik telaah medis dalam hal mengidentifikasi dan membantu dokter mendiagnosis penyakit ini. Umumnya, pasien COVID-19 menunjukkan berbagai gejala, antara lain demam, batuk, dan kelelahan. Sebab gejala ini juga muncul pada pasien pneumonia, hal ini menimbulkan komplikasi dalam deteksi COVID-19, terutama selama musim flu. Studi awal menunjukkan bahwa kelainan pada gambar rontgen dada pasien yang terinfeksi COVID-19 dapat digunakan untuk membantu diagnosis penyakit. Teknik *Convolutional Neural Network* (CNN), pada khususnya, umum digunakan untuk masalah klasifikasi gambar. Dalam proyek akhir ini, kami menggunakan CNN untuk melakukan klasifikasi gambar rontgen pasien COVID-19. Kami menggunakan 13.808 gambar yang terdiri dari hasil rontgen 10.192 pasien yang negatif COVID-19 dan 3.616 pasien yang positif COVID-19. Hasil eksperimen telah menunjukkan akurasi keseluruhan setinggi ...

Kata kunci: Klasifikasi, COVID-19, CNN

DAFTAR ISI

KATA PENGANTAR	ii
ABSTRAK	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	v
DAFTAR TABEL	vi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	1
2 PEMBAHASAN	2
2.1 Dataset	2
2.2 Preprocessing	3
2.3 Pembuatan Model	4
2.4 Analisis Performa	6
3 PENUTUPAN	9
3.1 Kesimpulan	9
3.2 Saran	9
DAFTAR REFERENSI	10
LAMPIRAN	1
Lampiran 1	2

DAFTAR GAMBAR

Gambar 2.1	Contoh Hasil Rontgen Pasien Positif COVID-19	2
Gambar 2.2	Contoh hasil Rontgen Pasien Negatif COVID-19	2
Gambar 2.3	Contoh Variasi Tata Letak	3
Gambar 2.4	Contoh Gambar yang Gelap (kiri) dan Kurang Jelas (kanan)	3
Gambar 2.5	Kiri: Jumlah Tiap Kasus Sebelum Downsampling, Kanan: Jumlah Tiap Kasus Sebelum Downsampling	4
Gambar 2.6	Keluaran yang Dihasilkan Setiap layernya	5
Gambar 2.7	Diagram Model	5
Gambar 2.8	Grafik Akurasi Model dari Setiap Epoch Terhadap Data Train dan Data Validasi	6
Gambar 2.9	Grafik Loss Model dari Setiap Epoch Terhadap Data Train dan Data Validasi	6
Gambar 2.10	<i>Confussion Matrix</i> Terhadap Data Test (0 = Negatif, 1 = Positif)	7
Gambar 2.11	Evaluasi Model dengan metric <i>accuracy</i> , <i>precision</i> , <i>re- call</i> , dan <i>f1-score</i>	7
Gambar 1	Tangkapan Layar Pengerjaan Proyek	2

DAFTAR TABEL

Tabel 2.1	Detail Partisi Dataset	4
Tabel 2.2	Paramater yang Digunakan Dalam Melatih Model	5

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Sejak akhir tahun 2019, dunia terlanda penyakit dari jenis *coronavirus* baru, yaitu COVID-19. COVID-19 adalah penyakit yang dapat menyebabkan gangguan pernapasan, nyeri dada, dan bahkan kematian. Karena penyebaran virus ini sangat cepat, WHO mengklasifikasikan penyakit ini sebagai pandemi. Akibatnya, berbagai pihak dalam dunia kesehatan tengah mencari cara-cara untuk menanganinya. Salah satu komponen yang penting dalam penanganan wabah COVID-19 adalah diagnosis. Karena dampaknya pada paru-paru, umumnya orang yang terjangkit virus COVID-19 akan menunjukkan kelainan pada citra paru-parunya. Metode citra medis paru-paru yang sering digunakan adalah radiografi rontgen. Artinya, seorang radiolog dapat mendeteksi kelainan dan menggunakan hasil rontgen untuk membantu dalam diagnosis COVID-19. Namun, menimbang jumlah kasus yang besar dan kemiripan hasil rontgen pasien pneumonia dari COVID-19 dan virus atau bakteri lainnya, pendeteksian berbasis komputer akan sangat membantu radiolog dalam mencapai sebuah diagnosis. Diperlukan sebuah sistem yang dapat mengklasifikasi citra rontgen dan menerka ada atau tidaknya peran COVID-19 di situ. Atas dasar tujuan ini, kelompok kami telah membangun sebuah model berbasis *Convolutional Neural Network (CNN)* untuk melaksanakan klasifikasi biner dari citra rontgen.

1.2 Tujuan

Tujuan proyek ini adalah untuk mengimplementasikan CNN pada masalah klasifikasi gambar. Dari hasil rontgen, kami bertujuan untuk membuat model yang dapat memprediksi terjangkit COVID-19 atau tidaknya pasien tersebut dengan akurasi besar.

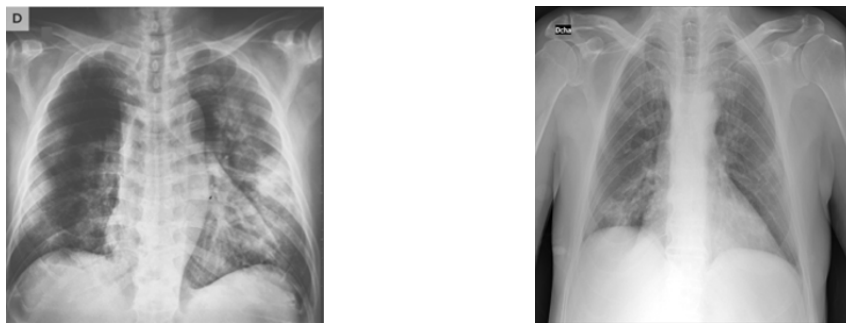
BAB 2

PEMBAHASAN

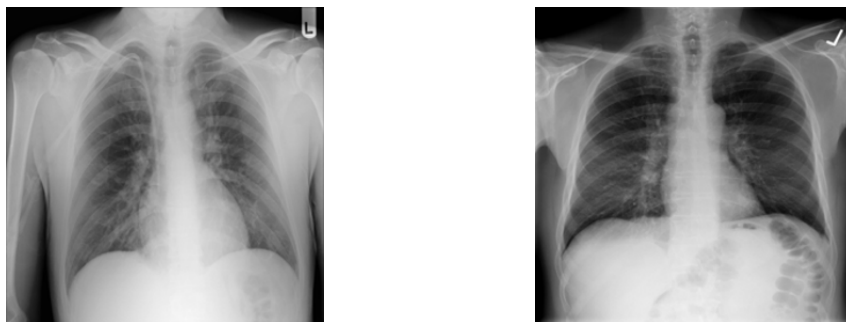
2.1 Dataset

Data yang kami gunakan berupa hasil rontgen 13.808 pasien, yang terdiri dari hasil rontgen 10.192 pasien yang negatif COVID-19 dan 3.616 pasien yang positif COVID-19. Setiap gambar memiliki label "Positive" atau "Negative", yang menentukan apakah pasien tersebut terjangkit penyakit COVID-19 atau tidak.

Gambar 2.1 merupakan contoh hasil rontgen pasien positif COVID-19, sedangkan gambar 2.2 merupakan contoh hasil rontgen pasien negatif COVID-19.

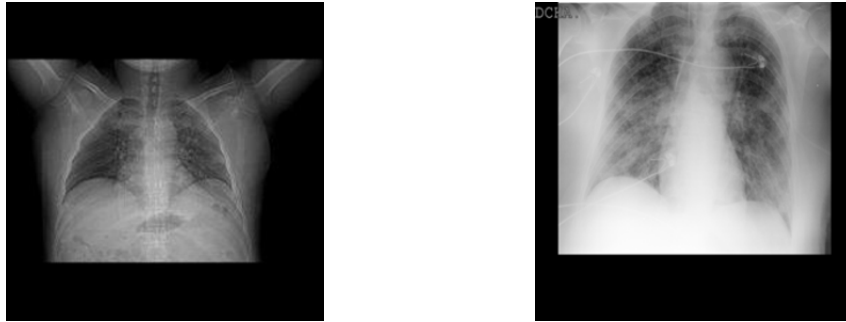


Gambar 2.1: Contoh Hasil Rontgen Pasien Positif COVID-19

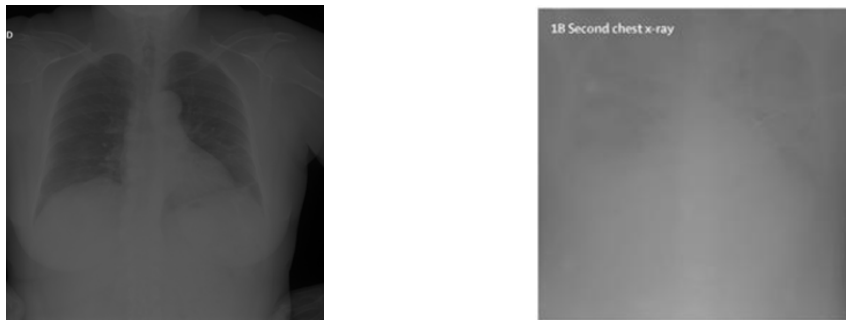


Gambar 2.2: Contoh hasil Rontgen Pasien Negatif COVID-19

Semua gambar kami memiliki pewarnaan *grayscale* dengan ukuran 299×299 piksel. Gambar kami sudah cukup tertata, walaupun ada variasi dalam hal penerangan dan tata letaknya seperti pada gambar 2.3 dan gambar 2.4.



Gambar 2.3: Contoh Variasi Tata Letak



Gambar 2.4: Contoh Gambar yang Gelap (kiri) dan Kurang Jelas (kanan)

2.2 Preprocessing

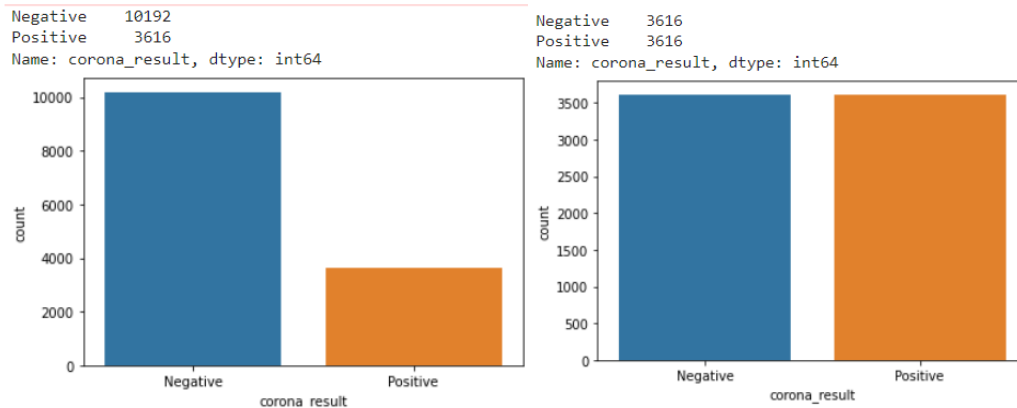
Sebelum menggunakan data untuk membangun model, kami melakukan preprocessing pada dataset. Preprocessing yang kita lakukan adalah mengubah gambar menjadi array yang berisi integer dari 0 sampai 255. Ini adalah nilai yang menandakan tingkat kecerahan tiap piksel dalam gambar, dimana nilai yang besar berkorespondensi dengan piksel yang cerah.

Selain itu, kita juga mengubah label tiap gambar menjadi integer 0 atau 1, dimana label "Negative" dipetakan ke 0 dan "Positive" ke 1.

Kami juga melakukan *down sampling* data set, sehingga jumlah kasus positif dengan kasus negatif memiliki perbandingan 1:1 (lihat gambar 2.5). Setelah itu, kami juga mempartisi dataset yang sudah di-*down sampling* menjadi 3 bagian, yaitu *training data*, *validation data*, dan *test data* dengan rasio 70%, 15%, 15% terhadap data total secara berurutan (lihat tabel 2.1)

Dataset	Jumlah	Rasio (%)
Training data	5062	70%
Validation data	1085	15%
Test data	1085	15%

Tabel 2.1: Detail Partisi Dataset



Gambar 2.5: Kiri: Jumlah Tiap Kasus Sebelum Downsampling, Kanan: Jumlah Tiap Kasus Sebelum Downsampling

2.3 Pembuatan Model

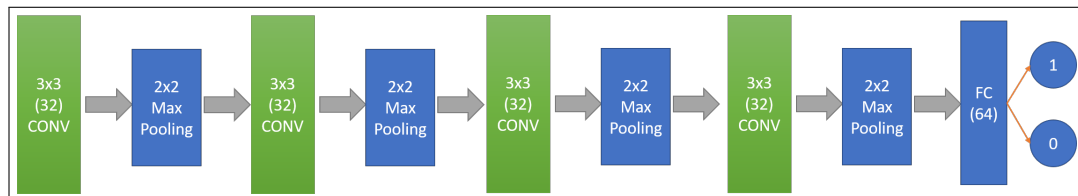
Pada makalah ini, kami gunakan model CNN sederhana untuk mengklasifikasikan kasus positif COVID-19 dan negatif COVID-19 dari kumpulan gambar rontgen. gambar 2.7 merupakan diagram model, dan gambar 2.6 adalah tabel mengenai detail keluaran tiap layer-nya. Model ini memiliki 2 pasang layer konvolusi dan 1 *fully-connected layer*. setiap layer konvolusi terdapat 32 filter dengan ukuran 3×3 . Kami juga menerapkan layer Max Pooling 2×2 untuk setiap layer konvolusinya. pada *fully-connected layer* diterapkan juga regularisasi dropout sebesar 50%. Selain itu, kami juga menerapkan fungsi aktivasi Relu pada setiap layer Konvolusi dan *fully-connected layer*-nya. Pada layer prediksi, kami menggunakan fungsi aktivasi softmax karena model mengklasifikasi 2 kelas, positif COVID-19 dan negatif COVID-19.

Model akan memprediksikan kelas dari data masukan (*input* dengan melihat keluaran tertinggi pada layer prediksinya).

Model ini menggunakan fungsi *Cross-Entropy Loss function* sebagai fungsi *loss*-nya, dan algoritma Adam sebagai *optimizer*. Lebih lanjut lagi, model ini dilatih dengan menggunakan parameter pada tabel 2.2.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 299, 299, 32)	320
max_pooling2d (MaxPooling2D)	(None, 149, 149, 32)	0
conv2d_1 (Conv2D)	(None, 149, 149, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_2 (Conv2D)	(None, 74, 74, 32)	9248
max_pooling2d_2 (MaxPooling2D)	(None, 37, 37, 32)	0
conv2d_3 (Conv2D)	(None, 37, 37, 32)	9248
max_pooling2d_3 (MaxPooling2D)	(None, 18, 18, 32)	0
flatten (Flatten)	(None, 10368)	0
dense (Dense)	(None, 64)	663616
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 2)	130
Total params: 691,810		
Trainable params: 691,810		
Non-trainable params: 0		

Gambar 2.6: Keluaran yang Dihasilkan Setiap layernya



Gambar 2.7: Diagram Model

Parameters	Nilai
Cost Function	Cross-Entropy loss function
Laju pembelajaran (awal)	0.001
Ukuran batch	64
Optimizer	Adam
Max epochs	50
Execution enviroment	GPU

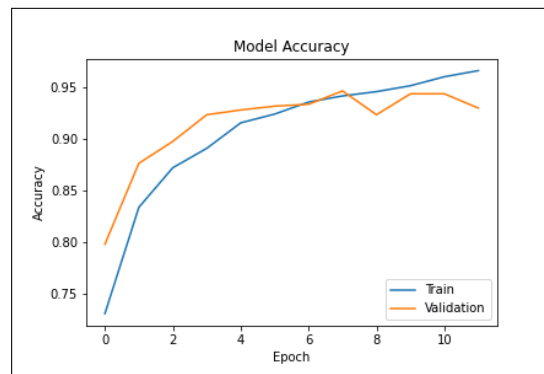
Tabel 2.2: Paramater yang Digunakan Dalam Melatih Model

2.4 Analisis Performa

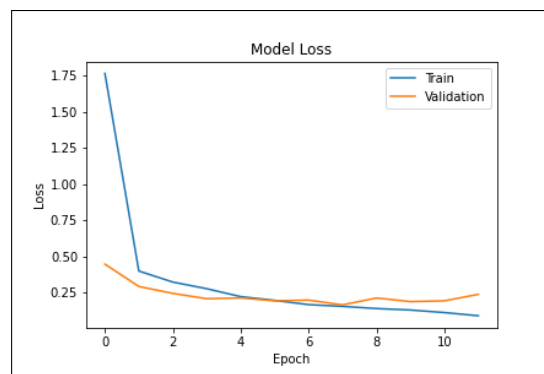
Gambar 2.8 dan gambar 2.9 merupakan grafik akurasi dan loss dari model pada *training data*, dan *validation data* selama proses *training* berlangsung (11 epochs). Berdasarkan gambar 2.8, akurasi (*accuracy*) model kami terus bertambah baik (mendekati 1) seiring bertambahnya epoch baik pada *training data*, maupun *validation data*. hal ini menunjukkan bahwa model tidaklah *overfitting*.

Nilai loss model pada gambar 2.9 dari penggunaan fungsi *Cross-Entropy Loss function* dan nilai loss model kami terus bertambah baik (menurun mendekati 0) seiring bertambahnya epoch.

Kami mengevaluasi model kami dengan berbagai macam metric, seperti *accuracy*, *precision*, *recall*, dan *f1-score*, serta dengan menggunakan *confussion matrix* terhadap data test.



Gambar 2.8: Grafik Akurasi Model dari Setiap Epoch Terhadap Data Train dan Data Validasi

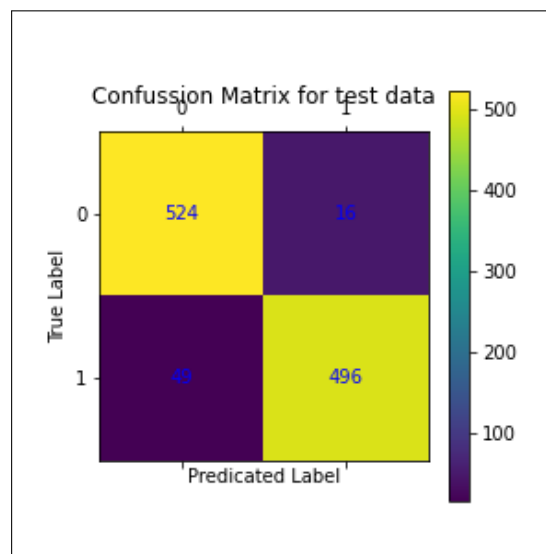


Gambar 2.9: Grafik Loss Model dari Setiap Epoch Terhadap Data Train dan Data Validasi

Confussion matrix terdiri dari empat karakteristik dasar (angka) yang digunakan untuk menentukan metric pengukuran pengklasifikasi, yaitu

- True Negative (TN) merepresentasikan jumlah orang yang diklasifikasikan dengan benar bahwa orang tersebut negatif COVID-19.

- True Positive (TP) merepresentasikan jumlah orang yang diklasifikasikan dengan benar bahwa orang tersebut positif COVID-19.
- False Negative (FN) merepresentasikan jumlah orang yang diklasifikasikan negatif COVID-19 tetapi sebenarnya orang tersebut positif COVID-19.
- False Positive (FP) merepresentasikan jumlah orang yang diklasifikasikan positif COVID-19 tetapi sebenarnya orang tersebut negatif COVID-19.



Gambar 2.10: Confussion Matrix Terhadap Data Test (0 = Negatif, 1 = Positif)

Confussion Matrix pada gambar 2.10 menunjukkan bahwa model kami memprediksi benar 524 orang yang negatif COVID-19 dan 496 orang yang positif COVID-19, sedangkan model kami salah memprediksi 49 orang yang negatif COVID-19 (diprediksi positif oleh model kami) dan 16 orang yang positif COVID-19 (diprediksi negatif oleh model kami).

Selanjutnya nilai TN (524), TP (496), FN (49), dan FP (16) digunakan untuk menghitung metric *accuracy*, *precision*, *recall*, dan *f1-score*.

Classification Report for Test Data					
	precision	recall	f1-score	support	
0	0.91	0.97	0.94	540	
1	0.97	0.91	0.94	545	
accuracy			0.94	1085	
macro avg	0.94	0.94	0.94	1085	
weighted avg	0.94	0.94	0.94	1085	

Gambar 2.11: Evaluasi Model dengan metric *accuracy*, *precision*, *recall*, dan *f1-score*

Dari informasi *accuracy*, *precision*, *recall*, dan *f1-score* pada gambar 2.11, dapat disimpulkan bahwa model kami sudah cukup baik dalam melakukan klasifikasi positif dan negatif COVID-19 dari hasil rontgen paru-paru karena nilai masing-masing metric tersebut mendekati 1.

BAB 3

PENUTUPAN

3.1 Kesimpulan

Pada makalah ini, kami menggunakan model CNN sederhana untuk mengklasifikasi pasien positif COVID-19 dan negatif COVID-19 dengan menggunakan hasil rontgen pasien. Model ini hanya menggunakan 4 layer konvolusi dan max-pooling, serta 1 *fully-connected layer*. Model mendapatkan hasil yang baik pada semua metrik standar selama proses *training*, *validation*, dan *testing*, dengan begitu, model ini tidaklah *overfitting*. Model bisa mendapatkan hasil akurasi sebesar 94% pada data test, oleh karena itu model tersebut sudah cukup baik dalam melakukan klasifikasi positif dan negatif COVID-19 dari hasil rontgen paru-paru.

3.2 Saran

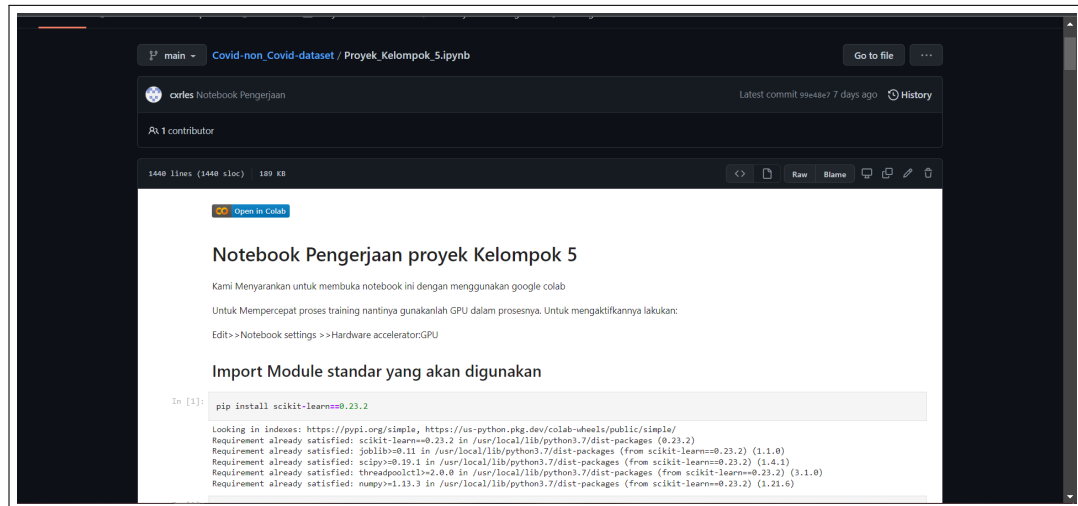
Saran yang bisa kami sampaikan adalah diperlukannya komparasi model dengan model lainnya. Selain itu, diperlukan juga pengujian ulang model terhadap dataset (rontgen paru-paru) yang terbuka lainnya.

DAFTAR REFERENSI

- [1] Zhang, A., Lipton, Z., Li, M. & Smola, A. Dive into Deep Learning. *ArXiv Preprint ArXiv:2106.11342*. (2021)
- [2] Masud, M. A light-weight convolutional Neural Network Architecture for classification of COVID-19 chest X-Ray images. *Multimedia Systems* (2022). <https://doi.org/10.1007/s00530-021-00857-8>
- [3] M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, “Can AI help in screening Viral and COVID-19 pneumonia?” *IEEE Access*, Vol. 8, 2020, pp. 132665 - 132676.
- [4] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2020. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images.

LAMPIRAN

Data, notebook pengerjaan, dan hasil dari proyek ini dapat diakses pada repositori github berikut ini: https://github.com/cxrles/Covid-non_Covid-dataset/tree/main



Gambar 1: Tangkapan Layar Pengerjaan Proyek