

Klasifikasi Penyakit Pneumonia Citra Digital X-Ray Menggunakan Metode *Convolutional Neural Network* dan RGB *Equalization*

Muhammad Attiqi Alghozali¹, Johan Rizky Triosaputra², Arwienda Kayan³

^{1,2,3}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: ¹[*¹muhammadattiqialghozali@gmail.com](mailto:muhammadattiqialghozali@gmail.com), ²johanrizkysaputra@gmail.com,

³arwiendakayan@gmail.com

Abstrak – Paru-paru basah atau yang biasa disebut dengan pneumonia, penyakit yang menyebabkan kematian tertinggi terutama pada anak-anak. Penyakit ini dapat disebabkan oleh virus, bakteri, ataupun jamur dan kurangnya dalam menjaga Kebugaran tubuh juga dapat menyebabkan penyakit ini menyerang pada organ paru-paru. Dalam identifikasi penyakit ini dapat dilakukan tes berupa pengambilan X-ray pada bagian dada. Pengambilan X-ray ini merupakan Tindakan cepat yang dilakukan agar penyakit segera teridentifikasi, saat membaca atau mengamati citra X-ray yang didapatkan dokter memerlukan waktu untuk mengatinya. Sehingga dalam penelitian ini dilakukan pembangunan sistem untuk melakukan klasifikasi penyakit pneumonia dengan citra X-ray menggunakan metode Convolution Neural Network dan histogram Equalization. Penggunaan pengolahan citra tersebut guna memperbaiki kualitas citra dan mempertajam kontras. Dengan total dataset 771 Gambar dari 4 kelas X-ray pneumonia, covid-19, normal, dan tuberculosis. Dataset gambar tersebut didapatkan dari website *Kaggle*, lalu dilakukan pengujian model CNN dan pengolahan citra X-ray didapatkan akurasi sebesar 91% memberikan performa dalam melakukan klasifikasi dengan sangat baik.

Kata Kunci — *Pneumonia, Citra Digital, RGB, CNN, X-Ray*

1. PENDAHULUAN

Paru-paru basah atau yang bisa disebut sebagai pneumonia adalah penyakit pada organ paru-paru manusia. Pneumonia masih menjadi penyakit yang serius karena menyerang organ inti manusia, di Indonesia penyakit paru-paru ini sudah ada kasus orang terjangkit penyakit ini. Umumnya penyakit ini menyerang pada anak-anak dengan usia di bawah 2 tahun dan orang tua yang berusia 65 tahun atau lanisa. Kondisi paru-paru ini dikarenakan adanya infeksi seperti bakteri, virus, atau jamur yang berkembang pada organ paru-paru. Tidak hanya pneumonia yang disebabkan oleh infeksi tersebut melainkan terdapat penyakit lain seperti, paru obstruktif kronis (PPOK), asma, pneumonia, tuberculosis (TBC), fibrosis paru, dan kanker paru-paru [1].

Teknologi sering dimanfaatkan oleh manusia sebagai alat atau fasilitas pendamping dalam melakukan aktivitas atau pekerjaan dalam sehari-hari. Tentunya teknologi ini selalu mengalami perkembangan pada setiap harinya yang dapat mempengaruhi beberapa bidang. Salah satu bidang yang mengalami perkembangan digitalisasi ini adalah pada bidang medis [2]. X-ray yang merupakan sebuah mesin pencitraan yang dapat digunakan untuk pengambilan gambar pada bagian dalam tubuh dengan warna hitam putih. Teknologi digital ini digunakan membantu tenaga medis dalam menganalisis bagian dalam tubuh manusia yang sulit dilihat dengan mata secara langsung atau telanjang, untuk melakukan diagnosis, dan menemukan Tindakan yang tepat dalam mengatasinya [2]. X-ray sering digunakan untuk menganalisis tulang atau rongga dada pasien. Citra X-ray dalam pengambilannya tidak selalu menghasilkan citra yang bagus seperti tingkat kontras yang kurang dan kabur, sehingga dapat menghambat dalam melakukan identifikasi penyakit yang diderita yang menyebabkan keterlambatan dalam penanganan.

Pemanfaatan perkembangan teknologi dengan objek penelitian X-Ray *chest* ini terdapat peneliti yang telah menggunakan beberapa metode dan pengolahan citra untuk melakukan klasifikasi penyakit paru-paru menggunakan citra X-ray. Metode yang dilakukan dalam melakukan deteksi gangguan paru-paru citra X-ray dengan *Deep Learning Convolutional Neural Network* dengan arsitektur ResNet-34 mendapatkan akurasi 100% untuk latih dan 95% akurasi uji [3]. Metode Algoritma C4.5 (*decision tree*) dalam mendeteksi Tuberculosis pada fitur GLCM citra *Chest X-ray*, mendapatkan tingkat akurasi 78,09% dengan 5-fold [4]. Metode Jaringan Syaraf Tiruan perambatan balik untuk Klasifikasi Covid-19 berbasis tekstur menggunakan metode orde pertama berdasarkan citra *Chest X-ray*, mendapatkan akurasi sebesar 94,17% untuk kelas normal dan 77,5% untuk kelas Covid-19 [5]. Metode *Convolutional neural network* untuk Identifikasi penyakit Pneumonia berdasarkan citra *Chest X-ray*, mendapatkan akurasi sebesar 97,16% pada latih dan 88,46% pada uji [2].

Dari penelitian yang sudah dilakukan dengan metode yang berbeda dan objek yang sama, setiap metode memiliki tingkat akurasi yang berbeda. Metode *Convolutional Neural Network* yang digunakan oleh peneliti terdahulu dalam penelitian ini akan digunakan dengan penambahan proses pengolahan citra menggunakan RGB Histogram. Dengan menggunakan dataset Citra *X-Ray* yang didapatkan dari *Kaggle*, penelitian ini bertujuan untuk mengetahui tingkat performa metode yang digunakan dan apakah penelitian ini dengan hasil yang didapatkan dapat digunakan dan dimanfaatkan dengan baik.

2. METODE PENELITIAN

2.1 X-Ray

X-Ray atau yang biasa kita kenal dengan rontgen merupakan sebuah hasil pengambilan gambar pada objek dalam tubuh manusia. Teknik ini digunakan oleh para ahli radiologi agar dapat melakukan identifikasi ciri dari hasil gambar yang didapatkan, untuk mengetahui kondisi seperti apa yang terjadi pada pasiennya [6]. X-Ray ini memiliki beberapa kelebihan dalam dunia medis salah satunya mudah digunakan dan memiliki nilai ekonomis yang besar. Dalam penelitian ini didapatkan citra X-ray pada bagian dada manusia untuk dilakukan klasifikasi, data yang digunakan diperoleh dari *Kaggle*.



Gambar 1. X-Ray *Chest* paru-paru Normal



Gambar 2. X-Ray *Chest* paru-paru Pneumonia

Pada gambar 1. Kondisi dimana paru-paru X-ray normal dimana pada bagian *pulmonary* pada objek yang dilingkari, terlihat lebih bersih dan terlihat jelas bahwa bagian tersebut tidak ada berwarna putih diluar areanya. Kondisi ini berbeda dengan penderita pneumonia, pada gambar 2. Terlihat pada bagian *pulmonary* terdapat objek berwarna putih disekitarnya diakibatkan oleh adanya inflamasi virus.

2.2 Citra Digital

Matriks yang terdiri dari baris dan kolom yang menghasilkan sebuah titik bersinggungan disebut dengan citra digital. Pada elemen matriks tersebut dinamakan piksel yang merepresentasikan nilai level keabuan titik yang bersinggungan [1]. Citra ini dapat dilakukan pengolahan citra untuk mendapatkan ciri atau pola yang ingin diketahui untuk dilakukan penelitian. Beberapa contoh dari pengolahan citra antara lain yaitu, RGB histogram, *grayscale*, *noise*, dan banyak metode lain dalam pengolahan citra digital.

2.3 RGB Histogram (*Red, Green, Blue*)

RGB digunakan untuk mengetahui nilai pada gambar dari jumlah piksel yang berbeda pada setiap nilai *Red*, *Green*, dan *Blue*. RGB ini memanipulasi gambar dengan nilai numerik pada tiap nilai RGBnya. Hal ini diperlukan agar dapat dilakukan identifikasi gambar, mengetahui informasi gambar, mengetahui ciri gambar, dan berguna dalam penelitian.

2.4 Histogram Equalization

Histogram Equalization merupakan metode pengolahan citra yang melakukan perataan nilai histogram. Metode ini digunakan sebagai pemrosesan dalam perbaikan struktur pada kualitas citra.

Rumus dalam proses equalization histogram ini terdapat pada persamaan (1) berikut.

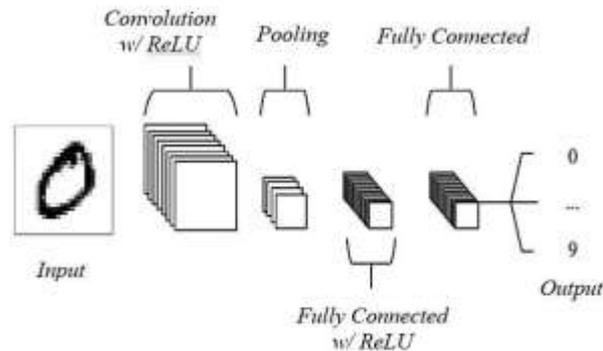
$$h(v) = \text{round} \left(\frac{cdf(v) - cdf_{min}}{2a(MxN) - (cdf_{min})} \times (L - 1) \right) \dots\dots\dots(1)$$

keterangan :

- v : nilai piksel yang ingin dicari penggantinya
- cdf(v) : fungsi distribusi kumulatif untuk nilai v
- cdfmin : nilai minimum dari distribusi kumulatif
- MxN : piksel penyusun citra, dengan M jumlah kolom dan N jumlah baris
- L : caca habu-abu yang dapat digunakan, citra abu-abu 8 bit maka L=256

2.5 Convolutional Neural Network (CNN)

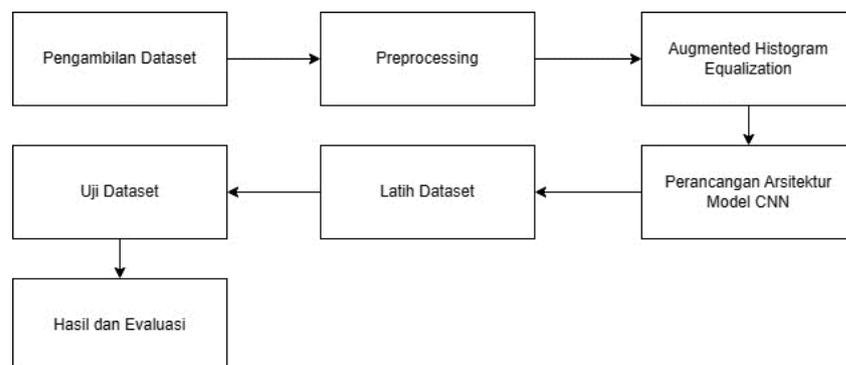
Metode CNN merupakan suatu *feed-forward artificial neural network* yang melakukan proses mendalam pada layernya dimana *neural network* dia mempertahankan struktur hubungannya antara layer dengan mempelajari struktur fitur pada internal dan melakukan generaliasasi fitur pada proses seperti pengenalan objek citra.



Gambar 3. Konfigurasi Convolutional Neural Network (CNN)

Pada gambar 3. Merupakan ilustrasi dari konfigurasi CNN, dimana CNN ini memiliki kelebihan dalam melakukan eliminasi tahapan ekstraksi fitur pola. Dalam prosesnya CNN menggunakan *layer* atau lapisannya yang digunakan melakukan ekstrak pola fitur secara otomatis. Hingga dapat melakukan melakukan eksekusi fitur yang akan dihasilkan sampai proses klasifikasi objek. Pada gambar 3. Tersebut dalam prosesnya CNN terdapat ReLU (*Rectified Linear Unit*) yang membantu dalam proses konvolusi. ReLU sendiri memiliki kelebihan dalam prosesnya dalam mengoptimalkan waktu dalam pelatihan dan pengujian dataset.

2.6 Tahapan Penelitian



Gambar 4. Tahapan Penelitian

Pada gambar 4. Merupakan diagram dari tahapan dalam melakukan penelitian ini. Terdapat 7 tahapan yang dilakukan untuk mendapatkan hasil dari latar belakang yang diselesaikan, sebagai berikut :

1. Pengambilan Dataset

Pengambilan dataset pada tahapan ini dilakukan dengan mengambil data gambar citra X-ray dari website *Kaggle* untuk dilakukan klasifikasi penyakit pneumonia. Dataset yang didapatkan terdapat 4 label berbeda yaitu citra X-ray Covid-19 data test terdapat 106 gambar, Normal 234 gambar, Pneumonia 390 gambar, dan Tuberculosis 41 gambar. Pada gambar dataset yang didapatkan tersebut selanjutnya akan diolah kedalam metode yang telah ditentukan.

2. Preprocessing

Preprocessing merupakan tahapan sebelum masuknya kedalam pengolahan data pada metode yang dipilih, pada tahapan ini data dilakukan pembersihan, transformasi, ekstraksi fitur, dan lain sebagainya. Tahapan ini bertujuan untuk melakukan perbaikan data agar terorganisir dengan baik dan terstruktur sesuai kebutuhan metode yang akan digunakan.

3. Augmented Histogram Equalization

Pada tahapan ini gambar atau data yang akan diolah dilakukan proses pengolahan citra, karena data berupa citra digital jadi dapat diolah untuk dilakukan perbaikan citra. Salah satunya Augmented Histogram Equalization dimana pada proses ini kontras citra pada X-ray akan ditingkatkan dengan menyeimbangkan intensitas piksel dalam gambar. Hal ini diperlukan agar objek yang kurang terlihat dengan jelas dapat ditampilkan dengan lebih jelas dan baik.

4. Perancangan Arsitektur metode CNN

Pada tahapan ini dilakukan penentuan tahapan apa saja yang akan digunakan dalam CNN, seperti fitur, pemilihan layer, penentuan arsitektur CNN, fungsi aktivasi, pelatihan dan evaluasi.

5. Latih Dataset

Pada tahapan ini setelah dilakukannya tahapan dari preprocessing hingga masuk kedalam model metode CNN dilakukan pelatihan dataset untuk melihat tingkat performa dari model yang digunakan apakah memiliki tingkat akurasi yang baik atau tidak.

6. Uji Dataset

Pada tahapan ini dilakukan pengujian model dengan dataset yang berbeda dengan dataset yang digunakan dalam pelatihan dataset.

7. Hasil dan Evaluasi

Pada tahapan ini dilakukan untuk mengevaluasi sejauh mana penelitian yang dilakukan. tahapan ini merupakan tahap penting untuk melakukan penyusunan cerita yang menjelaskan hasil temuan dari penelitian yang dilakukan seperti hasil akurasi, presisi, dan *f1-score*.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Program

Dataset yang digunakan untuk melakukan preprocessing citra Histogram Equalization dan Klasifikasi menggunakan CNN adalah dataset Citra gambar X-Ray *Chest*. Program digunakan untuk melakukan klasifikasi jenis penyakit melalui citra *Chest* X-ray, terdapat data citra X-ray dengan detail 106 Covid-19, 234 Normal, 390 Pneumonia, 41 Tuberculosis. Dataset yang akan digunakan memiliki 4 kelas, data tersebut yang akan digunakan untuk melakukan pelatihan pengenalan dalam CNN dengan *preprocessing* citra Histogram Equalization, Augmented, Original, dan Normalized. Implementasi program ini menggunakan *jupyter notebook google colab*, untuk hasil implementasi dan penjelasannya sebagai berikut :

1. Melakukan Import Library dan Pemanggilan Dataset

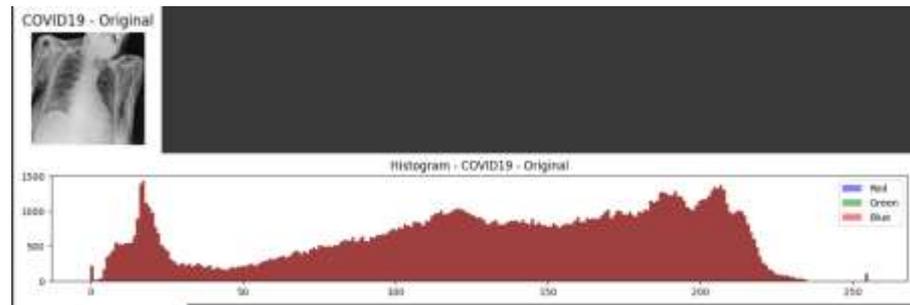


Gambar 5. Import Library

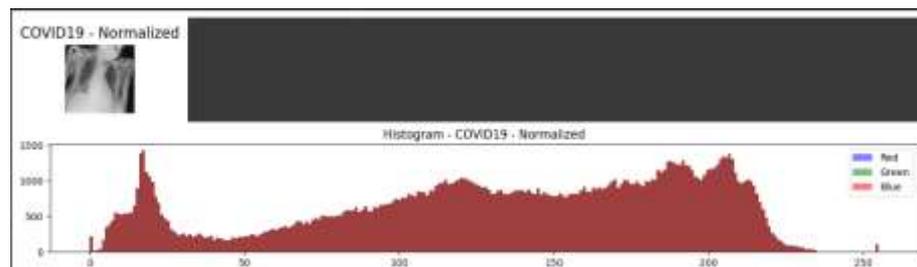
Pada Gambar 5. Melakukan import library yang diperlukan untuk memanggil fungsi-fungsi yang dapat digunakan untuk membantu dalam proses pengolahan data. Library keras digunakan untuk pembangunan dan pelatihan model dari neural network yang digunakan nantinya untuk melakukan klasifikasi gambar. Library OpenCV digunakan untuk melakukan pengolahan citra gambar. Library train test split untuk pembagian dataset latih dan uji. Library Keras memiliki peran yang banyak dalam pemodelan CNN sehingga diperlukan adanya library tersebut. Lalu ada label encoder untuk label kelas menjadi bilangan biner agar data dapat diolah dengan baik.

2. Load dataset dan Preprocessing Dataset

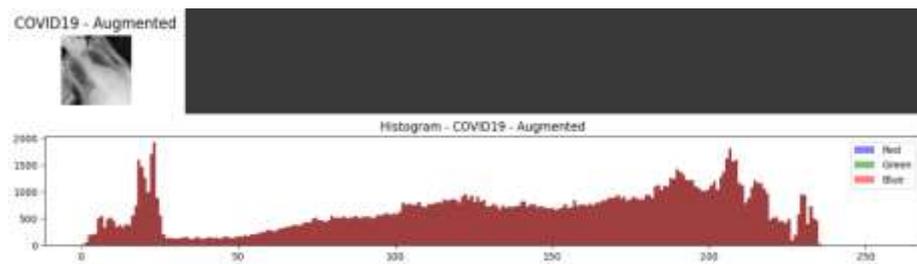
Dataset dalam folder dipanggil dan akan dipanggil sesuai dengan masing-masing labelnya. Dataset terdiri 4 kelas masing-masing akan ditampilkan salah satu untuk sampel, karena data berupa citra jadi dimunculkan salah satu gambar untuk tiap preprocessing nya untuk menunjukkan bahwa data berhasil di panggil dan dilakukan preprocessing atau pengolahan.



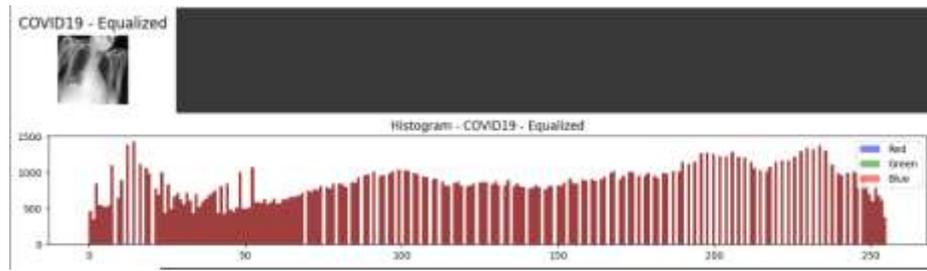
Gambar 6. Covid-19 Original dan Histogram



Gambar 7. Covid-19 Normalized dan Histogram



Gambar 8. Covid-19 Augmented dan Histogram



Gambar 9. Covid-19 Equalized dan Histogram

Pada Gambar 6. Terdapat contoh dari *load* gambar original dari citra X-ray covid-19 dan histogram dari gambar tersebut. Pada Gambar 7. Hasil dari proses pengolahan citra *Normalized*, Gambar 8. Hasil proses pengolahan citra *Augmented*, dan Gambar 9. Hasil dari pengolahan citra *Equalized*. Dari sampel Gambar 6,7,8,9 proses tersebut juga dilakukan pada citra X-ray label Normal, Pneumonia, dan Tuberculosis.

3. Encoder

Melakukan konversi gambar menjadi integer menggunakan Encoder, hal ini dilakukan agar data dapat digunakan dan dibaca oleh model yang digunakan.

4. Membagi dataset

Menggunakan library split data, dataset yang digunakan dibagi menjadi data *training* dan data *testing*. Dengan test_size sebesar 0,2 dengan random state = 42.

5. Membangun model CNN

Model CNN digunakan sebagai metode yang diterapkan untuk melakukan pelatihan atau pengenalan dataset yang diolah.

```
# Bangun model CNN
model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(128, 128, 1)))
model.add(MaxPooling2D((2, 2)))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D((2, 2)))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dense(4, activation='softmax'))
```

Gambar 10. Lapisan CNN

Pada Gambar 10. Merupakan lapisan model dari CNN, lapisan-lapisan dalam model tersebut berperan dalam melakukan pelatihan dan pengenalan objek pada tiap dataset.

6. Pelatihan Model

Pelatihan model ini menggunakan X_train dan y_train yang sudah dibagi datasetnya pada split data, lalu dilakukan pelatihan.

7. Evaluasi

Evaluasi ini menggunakan X_test dan y_test, untuk melihat performa dari model yang telah digunakan. Didapatkan tingkat *Accuracy* sebesar 0.91 atau 91% Pada Gambar 11.

```
Epoch 5/10: 12s 416c/step - loss: 1.4090 - accuracy: 0.4149 - val_loss: 0.8998 - val_accuracy: 0.5011
Epoch 6/10: 12s 416c/step - loss: 0.8492 - accuracy: 0.7552 - val_loss: 0.4431 - val_accuracy: 0.8294
Epoch 7/10: 12s 416c/step - loss: 0.3718 - accuracy: 0.8522 - val_loss: 0.2388 - val_accuracy: 0.8774
Epoch 8/10: 12s 416c/step - loss: 0.2981 - accuracy: 0.8809 - val_loss: 0.1583 - val_accuracy: 0.9384
Epoch 9/10: 12s 416c/step - loss: 0.2478 - accuracy: 0.9008 - val_loss: 0.1079 - val_accuracy: 0.9581
Epoch 10/10: 12s 416c/step - loss: 0.1998 - accuracy: 0.9329 - val_loss: 0.0968 - val_accuracy: 0.9516
Epoch 11/10: 12s 416c/step - loss: 0.1612 - accuracy: 0.9622 - val_loss: 0.1008 - val_accuracy: 0.9612
Epoch 12/10: 12s 416c/step - loss: 0.1267 - accuracy: 0.9763 - val_loss: 0.1065 - val_accuracy: 0.9607
Epoch 13/10: 12s 416c/step - loss: 0.0963 - accuracy: 0.9768 - val_loss: 0.1065 - val_accuracy: 0.9607
Epoch 14/10: 12s 416c/step - loss: 0.0841 - accuracy: 0.9794 - val_loss: 0.2222 - val_accuracy: 0.8228
Epoch 15/10: 12s 416c/step - loss: 0.0809 - accuracy: 0.9819 - val_loss: 0.2708 - val_accuracy: 0.8384
1/1 [12s 416c/step - loss: 0.1118 - accuracy: 0.9081]
loss: 0.21891424012106, accuracy: 0.81611263119696
```

Gambar 11. Hasil Akurasi

8. Save Model

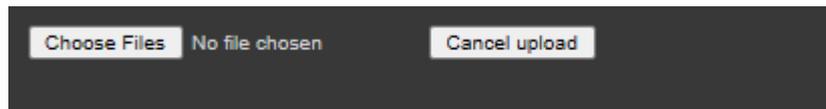
Model yang sudah dibuat disimpan untuk diload agar saat melakukan test pada data baru tidak perlu melakukan load dari awal lagi. Save model dilakukan dengan cara Pada Gambar 12.

```
# Save the model
model.save('/content/drive/MyDrive/Dataset/simsalabim.h5')
```

Gambar 12. Save Model

9. Melakukan Test Data Baru

Untuk membuktikan performa dari model dan sistem program yang sudah dibuat. Sebagai berikut Pada Gambar 13 Melakukan input Gambar baru selain gambar didalam folder test atau data latih.



Gambar 13. Choose file



Gambar 14. Input Gambar Pneumonia

```
1/1 [=====] - 0s 59ms/step
Predicted Class: Pneumonia
```

Gambar 15. Hasil Klasifikasi

Pada Gambar 14. Merupakan data baru Citra X-Ray Pneumonia yang dimasukkan untuk dilakukan klasifikasi menggunakan model yang sudah dibuat. Dan menghasilkan Prediksi kelas sesuai dengan kelas yang seharusnya.

4. SIMPULAN

Dari hasil implementasi program yang sudah dilakukan penelitian ini menghasilkan beberapa simpulan dan manfaat yang didapatkan :

1. Pengolahan citra dengan Augmented, Normalized, dan Equalized dapat membantu dalam pengolahan klasifikasi Gambar, dengan mempertajam citra dapat meningkatkan detail objek citra yang akan diolah.
2. Model CNN yang dibuat dapat melakukan klasifikasi dengan baik dan pengenalan dataset dengan baik.
3. Penggabungan pengolahan citra yang digunakan dan model CNN memperdetail pengolahan pengenalan dan memberikan pelatihan yang baik dalam pengenalan gambar melalui layer dalam CNN.
4. Teknologi dapat dimanfaatkan dengan baik sebagai alat bantu dalam melakukan klasifikasi jenis penyakit pneumonia citra X-ray.

5. Metode yang dilakukan dalam penelitian ini dapat membantu dalam melakukan klasifikasi dengan gambar citra X-Ray

5. SARAN

Berdasarkan hasil implementasi yang telah dilakukan dalam melakukan pengenalan citra X-ray *Chest* penyakit Pneumonia, Normal, Covid-19, dan Tuberculosis dengan dataset total sebanyak 771 secara keseluruhan dapat melakukan klasifikasi dengan akurasi hasil 91%. Tidak menutup kemungkinan jika digunakan arsitektur lain dalam metode CNN yang digunakan dapat memberikan tingkat akurasi yang lebih tinggi dan penggunaan dataset sebagai data latih dan uji yang lebih besar juga dapat meningkatkan performa dari model yang dibuat. Oleh karena itu sebagai penulis untuk penelitian selanjutnya, dapat menambahkan dataset yang lebih banyak dan pemilihan arsitektur yang tepat.

DAFTAR PUSTAKA

- [1] P. D. A. N. Non-pneumonia, M. Ramadhan, D. I. Mulyana, and M. B. Yel, "OPTIMASI ALGORITMA CNN MENGGUNAKAN METODE TRANSFER LEARNING UNTUK KLASIFIKASI CITRA X-RAY PARU-PARU," vol. 6, no. 2, pp. 670–679, 2022.
- [2] N. P. Ekananda and D. Riminalsih, "Identifikasi Penyakit Pneumonia Berdasarkan Citra Chest X-Ray Menggunakan Convolutional Neural Network," *J. Ilm. Inform. Komput.*, vol. 27, no. 1, pp. 79–94, 2022, doi: 10.35760/ik.2022.v27i1.6487.
- [3] O. F. Soedradjat, R. Magdalena, and N. K. C. Pratiwi, "Deteksi Gangguan Paru-Paru Berbasis Citra X-Ray Menggunakan Deep Learning," *e-Proceeding Eng.*, vol. 8, no. 6, pp. 2891–2896, 2022.
- [4] I. Junaedi, "Analisa Performa Algoritma C4.5 dalam Mendeteksi Tuberculosis pada Fitur GLCM Citra Chest X-Ray," *J. Sains dan Inform.*, vol. 9, pp. 46–55, 2023, doi: 10.34128/jsi.v9i1.590.
- [5] M. A. S. Yudono, "Perambatan Balik Untuk Klasifikasi Covid-19 Berbasis Orde Pertama Berdasarkan Citra Chest X-Ray," vol. 9, pp. 799–808, 2023.
- [6] I. M. Dendi Maysanjaya, "Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network (Classification of Pneumonia Based on Lung X-rays Images using Convolutional Neural Network)," *J. Nas. Tek. Elektro dan Teknol. Inf. /*, vol. 9, no. 2, p. 190, 2020, [Online]. Available: <https://garuda.kemdikbud.go.id/documents/detail/2807288>