

PENGARUH DATASET TERHADAP PERFORMA *CONVOLUTIONAL NEURAL NETWORK* PADA KLASIFIKASI *X-RAY* PASIEN COVID-19

Chyntia Raras Ajeng Widiawati*¹

¹Universitas Amikom Purwokerto, Purwokerto

Email: ¹chyntiaraw@mikompurwokerto.ac.id

*Penulis Korespondensi

(Naskah masuk: 07 Oktober 2021, diterima untuk diterbitkan: 08 Desember 2022)

Abstrak

Identifikasi pasien COVID-19 di Indonesia merupakan sebuah permasalahan yang harus diatasi. Identifikasi tersebut sebaiknya bisa lebih mudah dan cepat, sehingga deteksi dini pasien COVID-19 merupakan hal yang harus diperhatikan. Saat ini deteksi pasien COVID-19 bergantung pada *Swab Test* (RT-PCR) dan *Rapid Test* (Tes Antibodi), padahal *Rapid Test* tidak bisa memberikan tingkat akurasi yang tinggi sedangkan *Swab Test* memiliki biaya yang mahal. Salah satu solusi untuk membantu deteksi dini pasien COVID-19 adalah dengan memanfaatkan citra *X-Ray* paru dari pasien. Algoritma *Convolutional Neural Network* (CNN) adalah salah satu algoritma populer dengan performa yang sangat baik pada klasifikasi citra *X-Ray* pasien COVID-19. Walaupun CNN memiliki performa yang baik, keberhasilan suatu algoritma sangat bergantung dengan kualitas dataset yang digunakan. Selain itu citra *X-Ray* sangat bergantung terhadap pencahayaan di proses pengambilan gambar. Untuk itu perlu analisa pengaruh dataset terhadap performa model CNN yang digunakan. Penelitian ini bertujuan untuk melihat pengaruh kualitas dataset dan jumlah dataset terhadap performa CNN pada klasifikasi *X-Ray* pasien COVID-19. Dari eksperimen yang dilakukan terhadap dataset yang ada, CNN dapat mencapai akurasi sebesar 89,83%, sensitivitas sebesar 84,14%, spesifisitas sebesar 92,14%, PPV sebesar 71,35%, NPV sebesar 95,09 dan F1-score sebesar 76,10%. Sehingga dapat disimpulkan bahwa CNN memiliki performa yang baik dalam melakukan klasifikasi citra *X-Ray* pasien COVID-19, meskipun hasil tersebut lebih rendah dibandingkan dengan performa CNN terhadap dataset dengan jumlah dan kualitas citra yang lebih baik. Hal tersebut menunjukkan bahwa kualitas dan jumlah dataset sangat berpengaruh pada performa CNN dalam melakukan proses klasifikasi *X-Ray*.

Kata kunci: pengaruh kualitas dataset, convolutional neural network, COVID-19, citra, *X-Ray*

EFFECT OF DATASET ON CONVOLUTIONAL NEURAL NETWORK PERFORMANCE ON X-RAY CLASSIFICATION OF COVID-19 PATIENTS

Abstract

Identification of COVID-19 patients in Indonesia is a problem that must be solved. This identification can be easier and faster, so early detection of COVID-19 patients must be considered. Currently, the detection of COVID-19 patients depend on the Swab Test (RT-PCR) and Rapid Test (Antibody Test), even though the Rapid Test cannot provide a high level accuracy while the Swab Test is expensive. One of the solutions to help early detection of COVID-19 patients is by utilizing lung X-Ray images from patients. Convolutional Neural Network (CNN) algorithm is one of the popular algorithms with very good performance in X-Ray image classification of COVID-19 patients. Although CNN has a good performance, the success of algorithm is strongly supported by the quality of the dataset used. Besides X-Ray images are very dependent on lighting in shooting process. Therefore, it is necessary to analyze the influence of the dataset on the performance of the CNN model used. This study aims to see the effect of dataset quality and number of datasets on CNN's performance on X-Ray classification of COVID-19 patients. From the experiments conducted on existing datasets, CNN can achieve 89,83% accuracy, 84,14% sensitivity, 92,14% specificity, 71,35% PPV, 95,05% NPV and 76,10 F1-score. So it can be concluded that CNN has a good performance in classifying X-Ray images of COVID-19 patients, although these results are lower than CNN's performance on datasets with better number and image quality. This shows that the quality and number of datasets affect CNN's performance in conducting the X-Ray classification process.

Keywords: influence of dataset quality, convolutional neural network, COVID-19, image, *X-Ray*

1. PENDAHULUAN

Pandemi COVID-19 pertama kali dilaporkan terjadi di Wuhan, Cina pada Desember 2019 (HUSSAIN *et al.*, 2020). Pandemi COVID-19 kemudian meluas menjadi pandemi global yang terjadi hampir diseluruh negara, tak terkecuali Indonesia. Pandemi COVID-19 telah menyebar merata di hampir seluruh wilayah di Indonesia. COVID-19 pertama kali di Indonesia pada 2 Maret 2020 dan per 25 Maret telah terjadi sebanyak 790 kasus terkonfirmasi positif (secara kumulatif), diantaranya 58 meninggal, 31 sembuh, dan 701 dalam perawatan (WATRATAN, PUSPITA and MOEIS, 2020). Kondisi tersebut mengakibatkan Indonesia ada pada status waspada dan harus segera bertindak dalam penurunan angka kasus COVID-19. Beberapa hal telah dilakukan untuk menurunkan tingkat perkembangan COVID-19 di Indonesia, tak terkecuali pada proses identifikasi maupun perawatan pasien positif COVID-19.

Saat ini diagnosis utama ditegakkan melalui uji *polymerase chain reaction* (PCR) dan *swab* tenggorokan ditambah dengan pendekatan konfirmasi untuk memberikan diagnosa yang akurat. Metode tersebut membutuhkan waktu dan biaya yang tidak sedikit. Cara lain yang dapat dipertimbangkan adalah tindakan *non-invasive* melalui pencitraan paru karena mengacu pada gejala yang ditimbulkan yaitu sesak nafas sebagai akibat *pneumonia* akut (HRIYANI, HADIYOSO and SIADARI, 2020). Rontgen paru (*X-Ray*) dilakukan dengan tujuan diantaranya membantu dan menegakkan diagnosa penyakit, serta untuk melihat anggota tubuh bagian dalam yang digunakan sebagai dokumentasi rekam medis, selain itu peralatan yang diperlukan tidak terlalu rumit dan mudah dibawa (TABIK *et al.*, 2020). *Radiografi* dada biasanya menunjukkan kekeruhan ruang udara asimetris yang tidak merata atau *difus*, mirip dengan penyebab lain *pneumonia coronavirus* (CHEN *et al.*, 2020). Teknik pencitraan medis, seperti *X-Ray* atau *computed tomography* dada, biasanya digunakan untuk tujuan ini karena keandalannya untuk diagnosis COVID-19 (DURAN-LOPEZ *et al.*, 2020). Penelitian menggunakan teknik pencitraan radiologi (seperti *X-ray* dan *CT scan*) dapat terbukti membantu untuk mendeteksi COVID-19 karena citra *X-Ray* dan *CT scan* memberikan informasi penting tentang penyakit yang disebabkan oleh virus COVID-19 (HUSSAIN *et al.*, 2020).

Saat ini beberapa penelitian mengenai identifikasi maupun klasifikasi terhadap citra *X-Ray* pasien COVID-19 telah banyak dilakukan, diantaranya adalah penelitian yang dilakukan oleh (TABIK *et al.*, 2020) yang menerapkan algoritma CNN pada tahap klasifikasi. Hasil akurasi yang diperoleh pada penelitian tersebut sebesar 97.72%. Selain itu, penelitian yang dilakukan oleh (KASSANI *et al.*, 2020) juga menerapkan CNN pada proses klasifikasi. Pada penelitian tersebut

menghasilkan akurasi yang sangat tinggi yaitu sebesar 99%. Kedua penelitian tersebut menggunakan algoritma yang cukup populer keandalannya dalam melakukan proses klasifikasi, yaitu algoritma CNN dengan menggunakan model yang berbeda dan cukup rumit.

Transfer learning menggunakan CNN adalah strategi yang umum digunakan untuk menangani tugas klasifikasi citra medis, di mana CNN yang telah dilatih sebelumnya pada kumpulan data yang sangat besar (ALLAOUZI and AHMED, 2019). Pada banyak penelitian seperti klasifikasi citra, deteksi objek, dan analisis citra medis, CNN telah menunjukkan kinerja yang sangat baik (ISLAM, ISLAM and ASRAF, 2020).

Penelitian lain yang menggunakan algoritma CNN dengan model yang lebih sederhana juga dilakukan oleh salah satu peneliti di Indonesia yaitu penelitian yang dilakukan oleh (HARTATO, 2021). Penelitian tersebut memiliki tujuan untuk mengukur performa *Convolutional Neural Network* dalam melakukan klasifikasi *suspect* pasien COVID-19 berdasarkan citra rontgen paru-paru (*X-Ray*) dari pasien. Penelitian tersebut menggunakan dataset COVID-19 *Radiography Database* (RAHMAN, CHOWDHURY and KHANDAKAR, 2020) yang terdiri dari 3 kelas, yaitu kelas COVID-19 dengan jumlah data sebanyak 219 citra, kelas Normal yang terdiri dari 1341 citra, dan kelas *Viral_Pneumonia* yang terdiri dari 1345 citra. Semua citra yang ada pada dataset tersebut memiliki format tipe yang sama yaitu *Portable Network Graphics* (png). Model yang dibangun pada penelitian tersebut memberikan hasil akurasi sebesar 98,69%, sensitivitas sebesar 97,71%, dan spesifisitas sebesar 98,90%, dengan kesimpulan bahwa CNN memiliki performa yang sangat baik dalam melakukan klasifikasi citra rontgen seseorang untuk mengetahui apakah orang tersebut terpapar COVID-19 atau tidak.

Hasil yang diperoleh dengan model yang dirancang pada penelitian (HARTATO, 2021) dapat dikatakan memiliki performa yang cukup tinggi dan sangat baik meskipun model yang dirancang cukup sederhana, terlebih dataset yang digunakan merupakan dataset dengan kualitas citra dan jumlah yang cukup besar sehingga dapat memberikan performa yang sangat baik. Hal tersebut menunjukkan bahwa performa CNN memang tidak dapat diragukan lagi dalam mengenali pola data dan melakukan klasifikasi. Walaupun telah diperoleh hasil yang tinggi dari penelitian tersebut belum bisa membuktikan apakah dapat memberikan hasil yang sama baiknya pada dataset lain, dan apakah kualitas dataset akan berpengaruh pada hasil yang diberikan.

Salah satu penelitian yang dilakukan oleh (CRUZ *et al.*, 2021) telah membahas keterkaitan kualitas dataset dan dampaknya terhadap evaluasi model. Penelitian tersebut memberikan penilaian sistematis kumpulan data rontgen paru COVID-19. Metode yang digunakan pada penelitian tersebut

adalah *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA). PRISMA merupakan kumpulan item minimum berbasis bukti untuk pelaporan dalam tinjauan sistematis dan meta-analisis. PRISMA berfokus pada pelaporan tinjauan, mengevaluasi efek intervensi, tetapi juga dapat digunakan sebagai dasar untuk pelaporan tinjauan sistematis dengan tujuan selain mengevaluasi intervensi (misalnya mengevaluasi etiologi, prevalensi, diagnosis atau prognosis) (MOHER *et al.*, 2009).

Meskipun penelitian tersebut telah melakukan analisa terhadap 210 makalah dengan sekitar 100 kumpulan dataset berbeda, hasil yang diberikan tidak bisa menunjukkan performa model atau algoritma secara kuantitatif. Hal tersebut karena pada penelitian tersebut keterkaitan pengaruh dataset dengan metode tidak dilakukan dengan analisa eksperimental. Sehingga pada penelitian ini dilakukan eksperimen terhadap dataset yang diperoleh dari sumber lain dengan kualitas yang beragam. Analisa eksperimental dilakukan untuk melihat pengaruh kualitas dataset terhadap performa suatu metode atau algoritma, dalam hal ini adalah CNN. Penelitian ini menggunakan model sederhana yang telah dilakukan sebelumnya pada penelitian (HARTATO, 2021) untuk melihat apakah CNN tetap akan memberikan hasil akurasi, sensitivitas, dan spesifisitas yang tinggi apabila dataset yang digunakan memiliki keterbatasan jumlah dan kualitas yang belum seragam. Selain tiga parameter pengukuran tersebut, pada penelitian ini ditambahkan tiga parameter pengukuran lain yaitu *positive predictive values* (PPV), *negative predictive values* (NPV) dan F1-score dengan tujuan untuk melihat tingkat keakuratan sebuah metode secara statistik. Hasil analisa ini nantinya bisa menjadi dasar dalam mengolah dataset sebelum diterapkan kepada algoritma, khususnya bagi para peneliti yang akan melakukan penelitian dengan dataset yang berupa citra (*image*).

2. METODE PENELITIAN

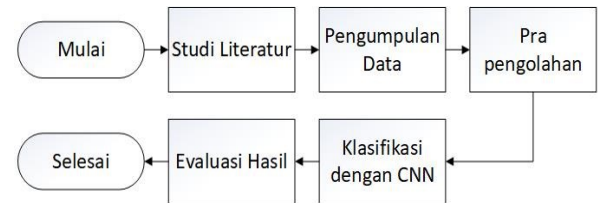
Secara garis besar, alur penelitian ini terbagi menjadi beberapa tahapan yakni studi literatur, pengumpulan data, pra pengolahan, klasifikasi dan evaluasi hasil.

2.1. Studi Literatur

Pada tahap ini, penulis mencari dan mempelajari literatur yang berkaitan dengan topik penelitian. Penulis juga mempelajari beberapa penelitian yang telah dilakukan sebelumnya. Penulis

juga menentukan literatur mana yang modelnya akan diterapkan pada dataset yang mau dianalisis pada penelitian ini, dengan mempertimbangkan kualitas dataset yang berasal dari sumber lain.

Alur penelitian tersebut ditunjukkan pada Gambar 1 dan dijelaskan pada sub bab berikutnya.



Gambar 1. Metode Penelitian

2.2. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dari sumber lain. Data yang akan digunakan berfokus pada data citra *X-Ray* yang bersifat publik dan biasa digunakan pada penelitian terkait klasifikasi COVID-19.

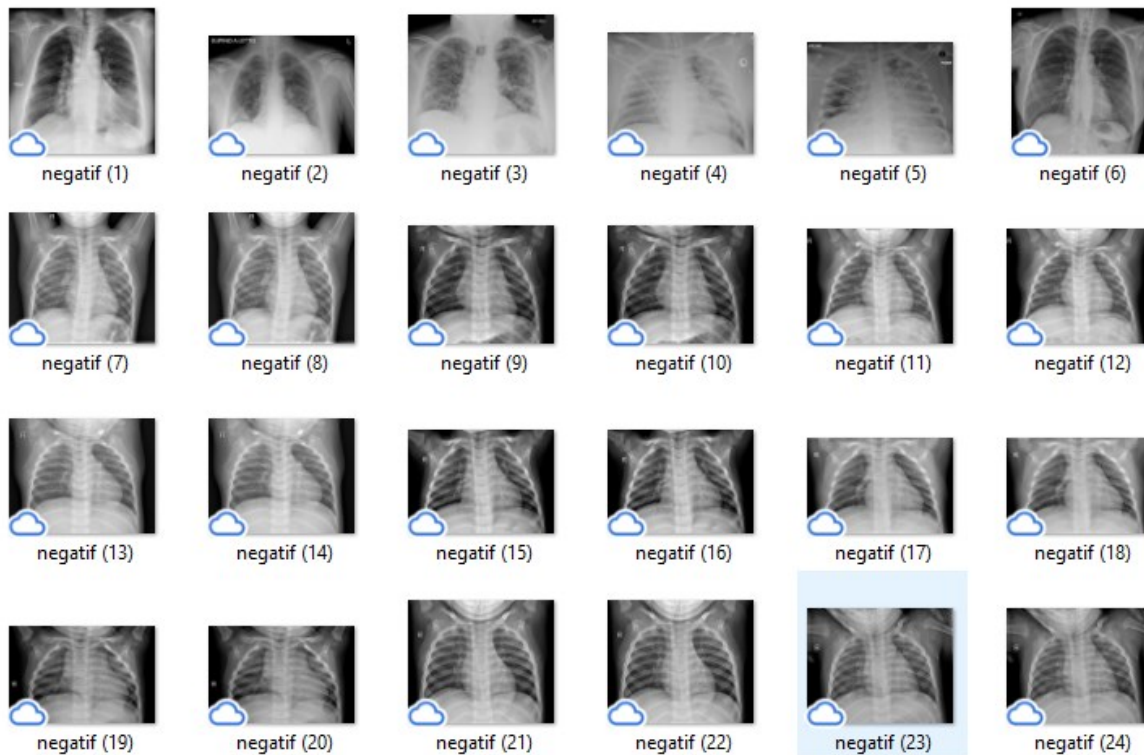
Dataset yang digunakan adalah dataset yang diperoleh dari (COHEN, 2021). Dataset ini terdiri dari 809 citra dengan 271 citra merupakan *citra X-Ray* dan *CT-Scan* pasien COVID-19. Pada dataset ini format dari citra yang ada terdiri dari 2 tipe yaitu *Joint Photographic Experts Group* (jpeg/jpg) dan *Portable Network Graphics* (png). Dataset ini merupakan dataset yang dikumpulkan dari sumber publik serta melalui pengumpulan tidak langsung dari rumah sakit dan dokter. Gambar 2 merupakan sampel citra dari dataset tersebut.



Gambar 2. Sampel Citra X-Ray Pasien COVID-19 (COHEN, 2021)

2.3. Pra pengolahan

Pada tahap pra pengolahan akan dilakukan beberapa hal mulai dari penghapusan citra *CT-Scan*, karena fokus penelitian ini hanya pada citra *X-Ray*. Selanjutnya dilakukan pengelompokan citra ke dalam folder sesuai dengan kelasnya yaitu kelas positif COVID-19 dan kelas negatif COVID-19.



Gambar 3. Contoh penamaan file dalam folder kelas negatif

Gambar 3 merupakan contoh penamaan file pada folder kelas negatif. Dataset yang digunakan pada penelitian ini kemudian diunggah pada *Google Drive*.

2.4. Klasifikasi

Tahap berikutnya pada metode yang diusulkan adalah klasifikasi untuk deteksi pasien COVID-19, citra tersebut diklasifikasikan ke dalam dua kelas yaitu positif COVID-19 dan negatif COVID-19. Pada penelitian ini akan dilakukan penerapan metode CNN dengan model dan arsitektur minimalis (HARTATO, 2021) pada klasifikasi terhadap dataset yang sudah dikumpulkan sebelumnya untuk melihat pengaruh dataset yang digunakan terhadap performa model tersebut.

Pada penelitian ini, implementasi model diawali dengan input berupa citra dengan ukuran yang beragam atau berbeda-beda dari citra yang satu dengan citra yang lainnya. Sebagai contoh terdapat citra inputan dengan ukuran 1904 x 1474 piksel, dan citra lain berukuran berbeda misalnya 1020 x 1024 piksel. Perbedaan ukuran tersebut dikarenakan dataset yang digunakan adalah dataset mentah yang memiliki ukuran dan kualitas citra sangat beragam. Disinilah letak pengujian terhadap performa CNN, apakah CNN mampu memberikan hasil yang optimal dimana dataset yang digunakan merupakan dataset mentah.

Dataset dengan ukuran citra yang beragam tersebut kemudian mengalami *resizing* untuk mengurangi beban komputasi sehingga citra yang

semula memiliki ukuran sangat beragam semuanya diubah menjadi berukuran 150 x 150 piksel.

2.5. Evaluasi Hasil

Untuk mengetahui tingkat kebenaran dari suatu metode klasifikasi dapat dilakukan dengan menghitung jumlah data sesuai dengan kesesuaiannya terhadap kelas. Jumlah data yang benar dikenali pada kelas tertentu dihitung sebagai TP (*true positive*). Jumlah data yang benar dikenali bukan sebagai kelas tersebut dihitung sebagai TN (*true negative*). Jumlah data yang salah dikenali sebagai kelas tertentu dihitung sebagai FP (*false positive*) dan jumlah data yang salah dikenali bukan sebagai kelas tersebut dihitung sebagai FN (*false negative*). Keempat jenis data ini membentuk *confusion matrix* yang ditunjukkan pada tabel 1.

Tabel 1. *Confusion Matrix*

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	<i>true positive</i> (tp)	<i>false negative</i> (fn)
Negatif	<i>false positive</i> (fp)	<i>true negative</i> (tn)

Tiga parameter utama yang umum digunakan pada klasifikasi dalam bidang biomedis yaitu akurasi, sensitivitas dan spesifisitas. Ketiga parameter tersebut digunakan dengan alasan akurasi menunjukkan presentase keberhasilan metode dalam pemilihan kedua kelas (positif COVID-19 dan negatif COVID-19) sesuai dengan keadaan aslinya. Sensitivitas menunjukkan tingkat keberhasilan metode dalam mengenali data yang benar dari kelas yang ingin dikenali (kasus positif). Sedangkan

spesifisitas menunjukkan tingkat keberhasilan metode dalam mengenali kasus negatif. Penghitungan nilai akurasi, sensitivitas dan spesifisitas dapat dirumuskan pada persamaan (1) hingga persamaan (3).

$$\text{akurasi} = \frac{tp+tn}{tp+fn+fp+tn} \quad (1)$$

$$\text{sensitivitas} = \frac{tp}{tp+fn} \quad (2)$$

$$\text{spesifisitas} = \frac{tn}{fp+tn} \quad (3)$$

Selain menggunakan tiga parameter tersebut, untuk mengetahui tingkat keakuratan sebuah metode dari segi statistik perlu dihitung dengan menggunakan parameter *positive predictive values* (PPV) dan *negative predictive values* (NPV) (FLETCHER and FLETCHER, 2005). PPV menggambarkan kinerja tes diagnostik hasil positif, sedangkan NPV menggambarkan kinerja tes diagnostik hasil negatif. Selanjutnya parameter pengukuran *F1-score* ditambahkan karena merupakan parameter pengukuran analisis statistik pada klasifikasi biner. Terutama pada penelitian ini merupakan klasifikasi biner yang hanya terdiri dari dua kelas yaitu positif COVID-19 dan negatif COVID-19. Ketiga parameter tersebut dirumuskan pada persamaan (4) hingga persamaan (6).

$$PPV = \frac{tp}{tp+fp} \quad (4)$$

$$NPV = \frac{tn}{tn+fn} \quad (5)$$

$$F1 - score = \frac{2tp}{2tp+fp+fn} \quad (6)$$

3. HASIL DAN PEMBAHASAN

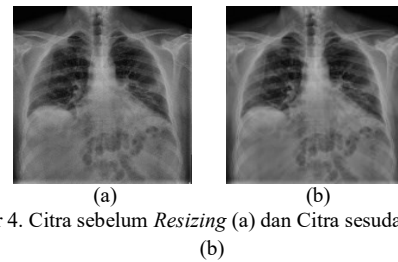
Hasil dari penelitian ini berdasarkan metode yang digunakan akan dijelaskan pada sub-bab berikut ini.

3.1. Pra pengolahan

Pada tahap ini dataset yang semula memiliki jumlah citra sebanyak 809 dipilah sesuai dengan kebutuhan penelitian. Citra *CT-Scan* dihapus sehingga jumlah data tersebut menjadi 578 dengan proporsi 123 citra merupakan citra *X-Ray* negatif COVID-19 dan 456 citra merupakan citra *X-Ray* positif COVID-19. Jika dibandingkan dengan penelitian yang telah dilakukan (HARTATO, 2021), jumlah data ini relatif sedikit dimana pada penelitian sebelumnya dataset yang digunakan berjumlah 2906 citra dengan kualitas yang sudah sangat baik.

Selanjutnya 578 citra tersebut yang sudah dikelompokkan kedalam 2 folder yaitu positif dan negatif mengalami perubahan ukuran atau *resizing*, sehingga semua berukuran 150 x 150 piksel.

Gambar 4 merupakan gambar perbandingan citra sebelum dan sesudah mengalami *resizing*.

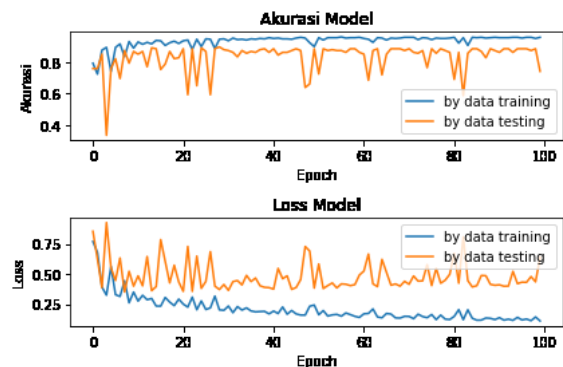


Gambar 4. Citra sebelum *Resizing* (a) dan Citra sesudah *Resizing* (b)

Secara visual kedua citra tersebut seolah tidak ada yang berbeda, namun citra hasil *resizing* sangat membantu meminimalisir beban komputasi, sehingga hal tersebut dilakukan. Selain itu, hal ini dilakukan untuk menyeragamkan ukuran citra yang semula sangat beragam.

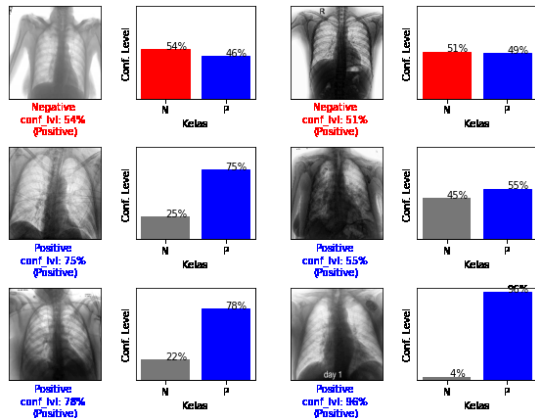
3.2. Klasifikasi

Selanjutnya citra yang sudah mengalami *resizing* menjadi data masukan pada proses klasifikasi menggunakan CNN dengan proporsi yaitu 70 % data *training* dan 30% data *testing*. Proses klasifikasi menggunakan 100 *epoch* dan 30 kali percobaan pengacakan data. Dari masing-masing pengacakan data pada dataset ini memberikan hasil yang sangat variatif, sebagai contoh dapat dilihat pada Gambar 5 berikut ini yang merupakan hasil eksperimen pada percobaan ke 10 sebanyak 100 *epoch*.



Gambar 5. Grafik hasil pelatihan pada percobaan ke 10 untuk setiap *epoch*

Jika diamati dari grafik yang ditunjukkan pada Gambar 5 diatas, hasil pelatihan sangat fluktuatif atau dapat dikatakan tidak stabil. Sangat berbeda dengan penelitian yang dilakukan oleh (HARTATO, 2021) dimana hasil grafik yang diperoleh sangat baik dan relatif stabil. Hasil sampel prediksi pada percobaan ke 10 ditunjukkan pada Gambar 6 berikut ini.



Gambar 6. Sampel prediksi pada percobaan ke 10

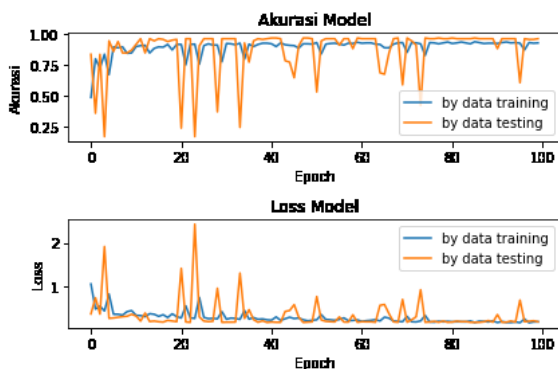
Berdasarkan Gambar 6 dapat dilihat bahwa dari 9 citra yang diprediksi pada percobaan ke 10 terdapat 2 citra yang seharusnya merupakan citra pasien positif terprediksi sebagai pasien negatif COVID-19 dengan tingkat presentase keyakinan mesin sebesar 54% dan 51% yang artinya sekitar 50% mesin salah dalam melakukan prediksi terhadap kedua citra tersebut. Selanjutnya hasil *confusion matrix* dari percobaan ke 10 dapat dilihat pada tabel 2 berikut ini.

Tabel 2. *Confusion Matrix* percobaan ke 10

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	31	34
Negatif	11	98

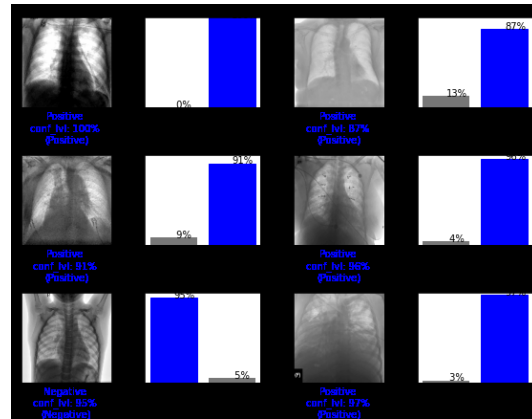
Berdasarkan tabel 2 dapat dilihat bahwa *confusion matrix* pada percobaan ke 10 menghasilkan TP sebanyak 31 data, TN sebanyak 98 data, FP sebanyak 11 data dan FN sebanyak 34 data. Dari *confusion matrix* tersebut diperoleh hasil pada yaitu akurasi sebesar 74,14%, sensitivitas sebesar 47,69%, spesifisitas sebesar 89,91%, PPV sebesar 73,81, NPV sebesar 74,24% dan F1-score sebesar 57,94% pada percobaan ke 10.

Sampel berikutnya yaitu hasil eksperimen pada percobaan ke 15 yang ditunjukkan pada Gambar 7 berikut ini.



Gambar 7. Grafik hasil pelatihan pada percobaan ke 15 untuk setiap *epoch*

Berdasarkan grafik yang ditunjukkan pada Gambar 7, dapat dilihat bahwa ketidak stabilan hasil pada percobaan ke 15 lebih rendah atau lebih baik dibandingkan pada percobaan ke 10. Sedangkan untuk hasil prediksi dari percobaan ke 15 ditunjukkan pada Gambar 8 berikut ini.



Gambar 8. Sampel prediksi pada percobaan ke 15

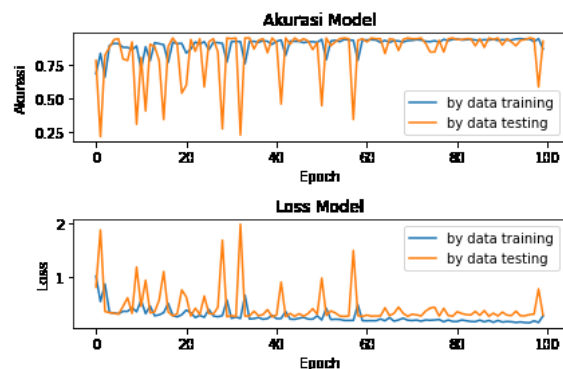
Berdasarkan Gambar 8 dapat dilihat bahwa dari 9 citra yang diprediksi pada percobaan ke 15 tidak ada kesalahan dalam proses klasifikasi citra. Semua sampel citra diprediksi dengan hasil yang sesuai dan tingkat keyakinan mesin yang cukup tinggi. Hasil *confusion matrix* dari percobaan ke 15 dapat dilihat pada tabel 3 berikut ini.

Tabel 3. *Confusion Matrix* percobaan ke 15

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	23	6
Negatif	1	144

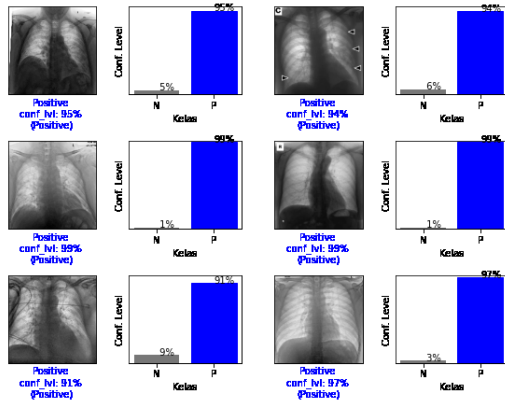
Berdasarkan tabel 3 dapat dilihat bahwa *confusion matrix* pada percobaan ke 15 menghasilkan TP sebanyak 23 data, TN sebanyak 144 data, FP sebanyak 1 data dan FN sebanyak 6 data. Dari *confusion matrix* tersebut diperoleh hasil pada yaitu akurasi sebesar 95,83%, sensitivitas sebesar 95,83%, spesifisitas sebesar 96,00%, PPV sebesar 79,31, NPV sebesar 99,31% dan F1-score sebesar 86,79% pada percobaan ke 15.

Sampel berikutnya yaitu hasil eksperimen pada percobaan ke 18 yang ditunjukkan pada Gambar 9 berikut ini.



Gambar 9. Grafik hasil pelatihan pada percobaan ke 18 untuk setiap *epoch*

Berdasarkan grafik yang ditunjukkan pada Gambar 9, dapat dilihat bahwa ketidak stabilan hasil pada percobaan ke 18 lebih baik jika dibandingkan dengan percobaan sebelumnya. Sedangkan untuk hasil prediksi dari percobaan ke 18 ditunjukkan pada Gambar 10 berikut ini.



Gambar 10. Sampel prediksi pada percobaan ke 18

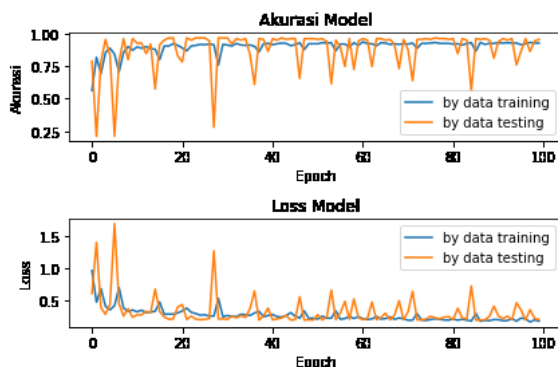
Berdasarkan Gambar 10 dapat dilihat bahwa dari 9 citra yang diprediksi pada percobaan ke 18 tidak ada kesalahan dalam proses klasifikasi citra. Semua sampel citra diprediksi dengan hasil yang sesuai dan tingkat keyakinan mesin hampir 100%. Hasil *confusion matrix* dari percobaan ke 18 dapat dilihat pada tabel 4 berikut ini.

Tabel 4. *Confusion Matrix* percobaan ke 18

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	31	6
Negatif	7	130

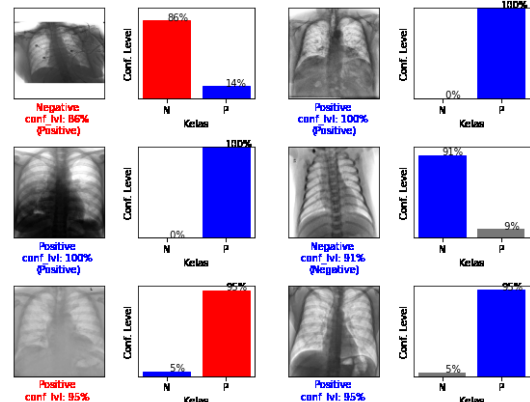
Berdasarkan tabel 4 dapat dilihat bahwa *confusion matrix* pada percobaan ke 18 menghasilkan TP sebanyak 31 data, TN sebanyak 130 data, FP sebanyak 7 data dan FN sebanyak 6 data. Dari *confusion matrix* tersebut diperoleh hasil pada yaitu akurasi sebesar 92,53%, sensitivitas sebesar 83,78%, spesifisitas sebesar 94,89%, PPV sebesar 81,58, NPV sebesar 95,59% dan F1-score sebesar 82,67% pada percobaan ke 15.

Sampel berikutnya yaitu hasil eksperimen pada percobaan ke 24 yang ditunjukkan pada Gambar 11 berikut ini.



Gambar 11. Grafik hasil pelatihan pada percobaan ke 24 untuk setiap *epoch*

Berdasarkan grafik yang ditunjukkan pada Gambar 11, dapat dilihat bahwa ketidak stabilan hasil pada percobaan ke 24 lebih tinggi jika dibandingkan dengan percobaan sebelumnya dan penurunan performa cukup signifikan di beberapa titik *epoch*. Sedangkan untuk hasil prediksi dari percobaan ke 24 ditunjukkan pada Gambar 12 berikut ini.



Gambar 12. Sampel prediksi pada percobaan ke 24

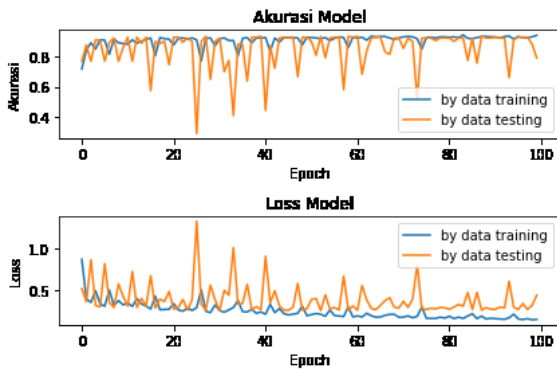
Berdasarkan Gambar 12 dapat dilihat bahwa dari 9 citra yang diprediksi pada percobaan ke 24 terdapat 1 sampel citra yang seharusnya merupakan citra pasien positif COVID-19 terprediksi sebagai pasien dalam kelas negatif COVID-19 dengan tingkat presentase keyakinan mesin sebesar 86%. Serta terdapat 1 sampel citra yang seharusnya merupakan citra pasien negatif COVID-19 terprediksi sebagai pasien dalam kelas positif COVID-19 dengan tingkat presentase keyakinan mesin sebesar 95%. Selanjutnya hasil *confusion matrix* dari percobaan ke 24 dapat dilihat pada tabel 5 berikut ini.

Tabel 5. *Confusion Matrix* percobaan ke 24

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	31	2
Negatif	6	135

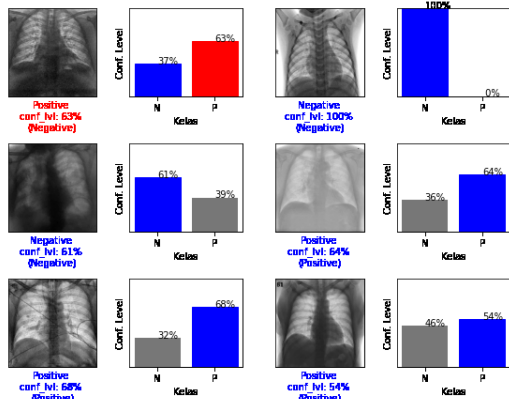
Berdasarkan tabel 5 dapat dilihat bahwa *confusion matrix* pada percobaan ke 24 menghasilkan TP sebanyak 31 data, TN sebanyak 135 data, FP sebanyak 6 data dan FN sebanyak 2 data. Dari *confusion matrix* tersebut diperoleh hasil pada yaitu akurasi sebesar 95,40%, sensitivitas sebesar 93,94%, spesifisitas sebesar 95,74%, PPV sebesar 83,78%, NPV sebesar 98,54% dan F1-score sebesar 88,57% pada percobaan ke 15.

Sampel terakhir yang disajikan pada paper ini yaitu hasil eksperimen pada percobaan ke 30 yang ditunjukkan pada Gambar 13 berikut ini.



Gambar 13. Grafik hasil pelatihan pada percobaan ke 30 untuk setiap epoch

Berdasarkan grafik yang ditunjukkan pada Gambar 13, dapat dilihat bahwa ketidak stabilan hasil pada percobaan ke 30 lebih tinggi jika dibandingkan dengan percobaan sebelumnya dan penurunan performa cukup signifikan di beberapa titik epoch. Sedangkan untuk hasil prediksi dari percobaan ke 30 ditunjukkan pada Gambar 14 berikut ini.



Gambar 14. Sampel prediksi pada percobaan ke 30

Berdasarkan Gambar 14 dapat dilihat bahwa dari 9 citra yang diprediksi pada percobaan ke 30 terdapat 1 sampel citra yang seharusnya merupakan citra pasien negatif COVID-19 terprediksi sebagai pasien dalam kelas positif COVID-19 dengan

tingkat presentase keyakinan mesin sebesar 63%. Selanjutnya hasil confusion matrix dari percobaan ke 30 dapat dilihat pada tabel 6 berikut ini.

Tabel 6. Confusion Matrix percobaan ke 30

Kelas Data	Prediksi Positif	Prediksi Negatif
Positif	34	30
Negatif	6	104

Berdasarkan tabel 5 dapat dilihat bahwa confusion matrix pada percobaan ke 24 menghasilkan TP sebanyak 34 data, TN sebanyak 104 data, FP sebanyak 6 data dan FN sebanyak 30 data. Dari confusion matrix tersebut diperoleh hasil pada yaitu akurasi sebesar 79,31%, sensitivitas sebesar 53,12%, spesifisitas sebesar 94,55%, PPV sebesar 85,00%, NPV sebesar 77,61% dan F1-score sebesar 65,38% pada percobaan ke 30.

Dari kelima sampel percobaan diatas dapat dilihat bahwa kualitas citra dan jumlah dataset yang digunakan cukup mempengaruhi terhadap hasil pelatihan model dan juga mempengaruhi hasil prediksi. Terlebih beberapa kesalahan prediksi terhadap pasien COVID-19 yang seharusnya dinyatakan positif tapi terprediksi negatif menjadi isu tersendiri karena kesalahan prediksi seperti ini akan sangat berbahaya dalam diagnosa medis.

3.3. Evaluasi Hasil

Untuk melihat seberapa besar pengaruh kualitas dan jumlah dataset terhadap performa CNN, pengamatan visual saja tidak disarankan. Sehingga perlu dilakukan evaluasi hasil dengan menggunakan parameter pengukuran seperti yang sudah dijelaskan pada bab sebelumnya yaitu akurasi, sensitivitas, spesifisitas, PPV, NPV dan F1-score.

Dari 30 percobaan yang telah dilakukan terhadap dataset yang ada, dimana dataset tersebut memiliki ukuran dan kualitas yang cukup beragam, serta jumlah dataset yang tidak terlalu banyak diperoleh hasil seperti yang ditunjukkan pada tabel 7 berikut ini.

Tabel 7. Hasil Pengukuran Performa

Percobaan	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)	PPV (%)	NPV (%)	F1-Score (%)
1	89,66	89,66	89,66	63,41	97,74	74,29
2	89,66	86,21	90,34	64,10	97,04	73,53
3	94,83	100,00	93,62	78,57	100,00	88,00
4	90,23	96,15	89,19	60,98	99,25	74,63
5	91,95	93,75	91,55	71,43	98,48	81,08
6	80,46	55,56	91,67	75,00	82,09	63,83
7	90,23	96,30	89,12	61,90	99,24	75,36
8	89,08	90,91	88,65	65,22	97,66	75,95
9	94,25	96,77	93,71	76,92	99,26	85,71
10	74,14	47,69	89,91	73,81	74,24	57,94
11	91,38	85,71	92,47	68,57	97,12	76,19
12	85,06	68,18	90,77	71,43	89,39	69,77
13	90,80	90,00	90,97	67,50	97,76	77,14

Percobaan	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)	PPV (%)	NPV (%)	F1-Score (%)
14	94,25	90,62	95,07	80,56	97,83	85,29
15	95,98	95,83	96,00	79,31	99,31	86,79
16	95,40	89,29	96,58	83,33	97,92	86,21
17	90,80	84,62	91,89	64,71	97,14	73,33
18	92,53	83,78	94,89	81,58	95,59	82,67
19	91,38	82,61	92,72	63,33	97,22	71,70
20	91,95	100,00	90,48	65,85	100,00	79,41
21	90,80	83,87	92,31	70,27	96,35	76,47
22	86,78	55,00	96,27	81,48	87,76	65,67
23	82,76	57,50	90,30	63,89	87,68	60,53
24	95,40	93,94	95,74	83,78	98,54	88,57
25	91,95	95,24	91,50	60,61	99,29	74,07
26	92,53	83,87	94,41	76,47	96,43	80,00
27	89,66	93,10	88,97	62,79	98,47	75,00
28	90,80	93,55	90,21	67,44	98,47	78,38
29	90,80	91,43	90,65	71,11	97,67	80,00
30	79,31	53,12	94,55	85,00	77,61	65,38
Rata-rata	89,83	84,14	92,14	71,35	95,09	76,10

Berdasarkan tabel 7 dapat dilihat bahwa rerata akurasi yang diperoleh setelah 30 percobaan yaitu sebesar 89,83% yang menunjukkan performa model dalam memprediksi kedua kelas yaitu kelas positif COVID-19 dan kelas negatif COVID-19. Sedangkan untuk hasil sensitivitas yang diperoleh yaitu sebesar 84,14% yang artinya performa model dalam melakukan klasifikasi kelas positif COVID-19. Selanjutnya untuk hasil spesifisitas sebesar 92,14% yang merupakan performa model dalam melakukan klasifikasi kelas negatif COVID-19. Untuk hasil rerata PPV sebesar 71,35% yang menunjukkan kinerja tes diagnostik pada klasifikasi kasus positif. Selanjutnya untuk hasil NPV sebesar 95,09% yang menunjukkan kinerja tes diagnostik pada kasus negatif, serta hasil rerata F1-Score yang diperoleh pada penelitian ini sebesar 76,10% yang menunjukkan tingkat keandalan metode dalam melakukan klasifikasi biner.

Sehingga dapat disimpulkan dengan menggunakan model yang sama seperti penelitian (HARTATO, 2021) namun dataset yang digunakan lebih sedikit dan beragam mulai dari format citra, posisi pengambilan citra, serta intensitas cahaya dan kontras masing-masing citra memberikan hasil bahwa performa model tersebut lebih andal dalam mengenali kasus negatif dibanding mengenali kasus positif. Hal tersebut dibuktikan dengan hasil spesifisitas dan NPV yang diperoleh lebih tinggi jika dibandingkan dengan sensitivitas dan PPV.

4. KESIMPULAN

Adapun kesimpulan yang diperoleh dari penelitian ini yaitu kualitas dan jumlah dataset sangat berpengaruh terhadap performa dari model CNN yang diusulkan oleh penelitian (HARTATO, 2021). Meskipun demikian, model tersebut tetap mampu memberikan hasil yang cukup baik yaitu dengan tingkat akurasi sebesar 89,83%, sensitivitas sebesar 84,14%, spesifisitas sebesar 92,14%, PPV

sebesar 71,35%, NPV sebesar 95,09 dan F1-score sebesar 76,10%. Untuk dapat meningkatkan hasil pada penelitian ini perlu adanya proses penyiapan dataset yang lebih baik dari segi jumlah, intensitas cahaya, kontras dan keseragaman posisi pengambilan citra. Selain itu proses segmentasi mungkin diperlukan untuk bisa meningkatkan performa CNN, sehingga CNN bisa mengenali karakteristik utama dari dataset X-Ray pasien COVID-19 (KHOIRO and BUSTOMI, 2014).

Selanjutnya, penelitian ini akan dikembangkan lagi dengan mencoba menerapkan teknik perbaikan kualitas citra dan segmentasi dalam eksperimen model berikutnya terhadap citra *X-Ray* pasien COVID-19. Peneliti berharap hasil eksperimen ini bisa menjadi dasar dalam penelitian sejenis sehingga mampu memberikan hasil yang lebih optimal dan model yang lebih baik.

UCAPAN TERIMAKASIH

Ucapan terima kasih dipersembahkan kepada Universitas Amikom Purwokerto yang telah mendanai penelitian ini. Ucapan terima kasih juga dipersembahkan kepada Bambang Pulu Hartato, S.Kom., M.Eng. yang telah memberikan izin untuk mengujikan model penelitiannya terhadap dataset yang digunakan pada penelitian ini.

DAFTAR PUSTAKA

- ALLAOUZI, I. & AHMED, M. BEN. 2019. A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases. *IEEE Access*, 7, pp. 64279–64288. doi: 10.1109/ACCESS.2019.2916849.
- CHEN, N, dkk. 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395(10223),

- pp. 507–513. doi: 10.1016/S0140-6736(20)30211-7.
- COHEN, J. P. 2021. Covid Chestxray Dataset', *github.com*. Available at: <https://github.com/ieee8023/covid-chestxray-dataset>.
- CRUZ, B. G. S., dkk. 2021. Public Covid-19 X-ray datasets and their impact on model bias – A systematic review of a significant problem. *Medical Image Analysis*, 74, p. 102225. doi: 10.1016/j.media.2021.102225.
- DURAN-LOPEZ, L. dkk. 2020. COVID-XNet: A custom deep learning system to diagnose and locate COVID-19 in chest x-ray images. *Applied Sciences (Switzerland)*, 10(16), pp. 1–13. doi: 10.3390/app10165683.
- FLETCHER, R. H. & FLETCHER, S. W. 2005 *Clinical epidemiology: the essentials*. 4th edn. Philadelphia: Lippincott Williams & Wilkins.
- HARTATO, B. P. 2021. Penerapan Convolutional Neural Network pada Citra Rontgen Paru-Paru untuk Deteksi SARS-Cov-2. *JURNAL RESTI: Rekayasa Sistem dan Teknologi Informasi*, 5(10), pp. 747–759. doi: <https://doi.org/10.29207/resti.v5i4.3153>.
- HRIYANI, Y. S., HADIYOSO, S. & SIADARI, T. S. 2020. Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 8(2), pp. 443–453. doi: <http://dx.doi.org/10.26760/elkomika.v8i2.443>.
- HUSSAIN, E., dkk. 2020. CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons & Fractals*, (xxxx), p. 110495. doi: 10.1016/j.chaos.2020.110495.
- ISLAM, M. Z., ISLAM, M. M. & ASRAF, A. 2020. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, 20, pp. 1–20. doi: 10.1016/j.imu.2020.100412.
- KASSANI, S. H. dkk. 2020. Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: A machine learning based approach', *arXiv*, pp. 1–18. Available at: https://www.researchgate.net/publication/340859631_Automatic_Detection_of_Coronavirus_Disease_COVID-19_in_X-ray_and_CT_Images_A_Machine_Learning-Based_Approach.
- KHOIRO, M. & BUSTOMI, M. A. 2014. Analisa Pengaruh Proses Segmentasi Citra terhadap Klasifikasi Citra Rontgen Paru-Paru dengan JST Backpropagation. *Jurnal Sains dan Seni Pomits*, 3(1), pp. 1–4.
- MOHER, D. dkk. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ (Online)*, 339(7716), pp. 332–336. doi: 10.1136/bmj.b2535.
- RAHMAN, T., CHOWDHURY, D. M. & KHANDAKAR, A. 2020. COVID-19 Radiography Database. Available at: <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>.
- TABIK, S., dkk. 2020. COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-ray Images. *arXiv*, 24(12), pp. 3595–3605. doi: 10.1109/JBHI.2020.3037127.
- WATRATAN, A. F. B., PUSPITA, A. & MOEIS, D. 2020. Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science and Technology (Jacost)*, 1(1), pp. 7–14. doi: <https://doi.org/10.52158/jacost.v1i1.9>.