

CONVOLUTIONAL NEURAL NETWORK UNTUK DETEKSI COVID-19 MELALUI CITRA CHEST X-RAY

Risal Ferdinand, Eko Setyawan dan L. M. Rasdi Rere
STMIK Jakarta STI&K
Jl. BRI No.17, Radio Dalam, Kebayoran Baru, Jakarta Selatan 12140
{ferdinandrisal, rasdirere267}@gmail.com

ABSTRAK

Pandemi Covid-19 telah menyebar keseluruh dunia dan secara umum menyebabkan dampak buruk pada kesejahteraan manusia. Salah satu cara untuk mengurangi penyebarannya adalah dengan deteksi dini terhadap SARS CoV-2 yang bertanggung jawab terhadap penyakit Covid-19 ini, melalui citra chest X-ray. Dalam penelitian ini diusulkan penggunaan Convolutional Neural Network untuk dapat mengenali gejalanya menggunakan Covid-19 Radiography Dataset dari Kaggle. Dari empat kelas yang tersedia, pada penelitian ini hanya dipergunakan dua kelas yaitu data citra chest X-ray untuk pasien normal dan pasien terdeteksi Covid-19. Berdasarkan beberapa eksperimen yang dilakukan, untuk mendapatkan parameter optimal, diperoleh hasil accuracy sebesar 0.9675 dengan precision 0.9465, recall 0.9833 dan f1-score 0.9645, pada ukuran citra 100x100, penggunaan jaringan dropout 0.1, validation split 0.1, dengan jenis optimizer adam.

Kata Kunci: Covid-19 Citra Chest X-Ray, Convolutional Neural Network

PENDAHULUAN

Coronavirus 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh strain coronavirus yang disebut *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2). Pertama kali penyakit ini terlihat di kota Wuhan, propinsi Hubei, Cina pada akhir Desember 2019. Dari kota tersebut wabah ini kemudian menyebar ke hampir seluruh dunia. Secara resmi wabah ini diakui pandemi oleh Organisasi Kesehatan Dunia (WHO) pada 11 Maret 2020 [1].

Penyebaran virus yang cepat secara internasional memiliki dampak sosial, ekonomi dan kesehatan yang sangat besar serta memaksa penerapan langkah-langkah sosial dan penataan kesehatan untuk mencegah penyebarannya. Pada awal Agustus 2021, jumlah kasus terkonfirmasi Covid-19 ini diseluruh dunia mencapai 204 juta dengan jumlah kematian melebihi 4 juta orang [2].

Virus SARS-CoV-2 yang menyebabkan Covid-19 menyebar melalui *droplet* atau percikan air liur yang dihasilkan jika orang yang terinfeksi menghembuskan nafas, batuk atau bersin. Karena *Droplet* ini cukup berat, sehingga tidak dapat bertahan lama di udara dan akan

jatuh serta menempel pada permukaan lantai dinding dan lainnya.

Seseorang yang terlalu dekat dengan orang lain yang sudah terinfeksi Covid-19 dapat tertular, karena udara disekitarnya yang mengandung virus kemungkinan terhirup olehnya. Selain itu penularan juga dapat terjadi jika seseorang menyentuh permukaan benda yang telah terkontaminasi virus, kemudian ia mengusap bagian mata, hidung atau mulutnya.

Orang yang terinfeksi COVID-19 mungkin mengalami demam, sakit kepala, nyeri otot, batuk kering, sakit tenggorokan hingga sakit pernafasan ringan sampai sedang. Akan tetapi pada orang tua atau seseorang yang memiliki kondisi medis seperti kardiovaskular, diabetes, penyakit pernapasan kronis dan kanker akan dapat menyebabkan gejala yang lebih serius.

Covid-19 dapat dideteksi dengan *test* atau pengujian *real-time polymerase chain reaction* (RT-PCR) dari SARS-CoV-2. Meskipun hasil pengujian untuk spesifisitasnya cukup tinggi, akan tetapi sensitivitasnya relatif rendah. Hal ini mengakibatkan hasil tesnya memakan waktu 4 jam atau bahkan seharian. Selain itu terkadang memberikan hasil positif palsu (*false positive*) atau negatif palsu (*false*

negative). Permasalahan ini dapat diatasi dengan foto CT (*computed topography*) atau rontgen dari individu yang terduga Covid-19.

Chest-CT dapat mendeteksi temuan abnormal pneumonia Covid-19. Secara umum CT dapat membedakan Covid-19 dengan pneumonia karena virus dan bakteri. Meskipun penggunaan CT untuk mendeteksi pneumonia Covid-19 telah banyak dibuktikan pada beberapa penelitian, akan tetapi pemakaian CT tidak cocok untuk skrining Covid-19 karena biaya dan paparan radiasinya. Di sisi lain pencitraan sinar x dada (*chest x-ray*: CXR) hemat biaya, meskipun sensitivitasnya secara umum lebih rendah dari CT untuk penyakit paru. Akan tetapi diagnosis pneumonia Covid-19 yang akurat menjadi lebih menantang pada CXR dibandingkan dengan *chest-CT* [3].

Saat ini penelitian mengenai citra CXR untuk mendeteksi Covid-19 menggunakan metode *machine learning* dan *deep learning* telah banyak dilakukan. Penelitian yang dilakukan oleh Rafid Mostafiz dkk [4] misalnya, mengusulkan hibridisasi *deep convolutional neural network* (CNN), *discrete wavelet transforms* (DTW) dan *random forest* untuk dapat mendeteksi Covid-19 dari citra *chest x-ray*. Hasil akurasi keseluruhan yang didapatkan dari penelitiannya adalah lebih 98.5%.

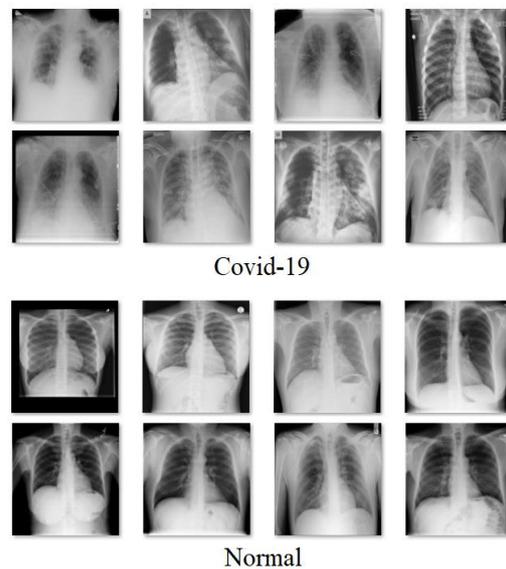
Penelitian lainnya yang dilakukan Boran S. dan Ilker O. [5] bertujuan untuk menghasilkan nilai akurasi yang tinggi pada deteksi citra *chest x-ray*. Mereka melakukan 38 percobaan dengan menggunakan *convolutional neural network*, 10 percobaan dengan lima model *machine learning* serta 14 percobaan jaringan *pre-trained* untuk *transfer learning*. Citra dan data statistik dipertimbangkan secara terpisah dalam eksperimen untuk mengevaluasi kinerja model. Hasil yang diperoleh adalah sensitivitas rata-rata 93,84%, spesifisitas rata-rata 99,18%, akurasi rata-rata 98,50%, dan area karakteristik operasi penerima rata-rata di bawah skor kurva 96,51% tercapai.

Secara umum penelitian untuk deteksi Covid-19 melalui citra *chest x-ray* banyak menggunakan arsitektur CNN

maupun turunannya seperti Interception-V3, VGG, ResNet dan lainnya [4] – [11]. Karena itu dalam penelitian ini juga dipergunakan arsitektur CNN pada *Covid-19 Radiography Dataset* dari *Kaggle*, dengan tuning parameter untuk beberapa eksperimen yang dilakukan.

METODE PENELITIAN

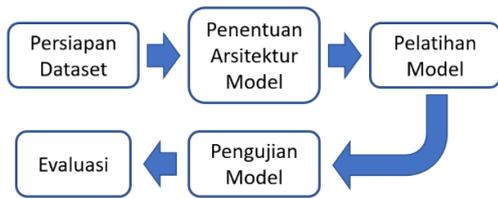
Penelitian ini bertujuan untuk dapat mendeteksi Covid-19 melalui data citra rontgen paru-paru (*chest x-ray*), dengan menggunakan metode *convolutional neural network*. Secara umum tahapan proses penelitian yang dilakukan diperlihatkan seperti pada Gambar 1.



Gambar 1. Proses Penelitian

Persiapan Dataset

Proses pertama dalam penelitian yang dilakukan adalah mempersiapkan dataset yang akan dipergunakan. Dalam penelitian ini dataset yang dipergunakan adalah data citra *Covid-19 Radiography Dataset* dari *Kaggle*. Beberapa contoh dataset yang dipergunakan diperlihatkan seperti pada Gambar 2.



Gambar 2. Contoh citra chest x-ray

Covid-19 Radiography Dataset berukuran 299×299 pixels serta dibagi dalam empat kelas dengan jumlah yang berbeda, yaitu citra chest x-ray untuk pasien normal, Covid-19, Viral Pneumonia dan Lung Opacity. Dalam penelitian ini hanya digunakan 4000 dataset untuk pasien normal dan pasien terindikasi Covid-19, dengan setiap kelas masing-masing berjumlah 2.000 citra chest-x-ray.

Penentuan Arsitektur Model

Arsitektur model CNN yang dipergunakan dalam penelitian ini menggunakan tiga layer dan fully connected layer seperti diperlihatkan Tabel 1. Pada ketiga layer dipergunakan conv2d (convolutional layer 2D) dengan fungsi aktivasi ReLu dan max pooling layer dengan region (2, 2). Setelah proses flatten, pada fully connected layer digunakan dense dan dense_1 dengan fungsi aktivasi sigmoid.

Tabel 1. Model CNN

Layer (type)	Output Shape	Param #
conv2d_3	(None, 94, 94, 64)	3200
activation_5	(None, 94, 94, 64)	0
max_pooling2d_3	(None, 47, 47, 64)	0
conv2d_4	(None, 41, 41, 64)	200768
activation_6	(None, 41, 41, 64)	0
max_pooling2d_4	(None, 20, 20, 64)	0
conv2d_5	(None, 14, 14, 64)	200768
activation_7	(None, 14, 14, 64)	0
max_pooling2d_5	(None, 7, 7, 64)	0
dropout_1	(None, 7, 7, 64)	0
flatten_1	(None, 3136)	0
dense_2	(None, 64)	200768
activation_8	(None, 64)	0
dense_3	(None, 1)	65
activation_9	(None, 1)	0

Total params: 605,569		
Trainable params: 605,569		
Non-trainable params: 0		

Pelatihan Model

Dalam penelitian ini, untuk mendapatkan pelatihan model terbaik, dilakukan sejumlah pelatihan atau training dengan beberapa perbedaan parameter yang ada pada arsitektur CNN yang ditentukan. Seperti diperlihatkan pada Tabel 2, percobaan pertama dan kedua berbeda pada ukuran citra (pixels) yang dipergunakan yaitu 200×200 dan 100×100 . Percobaan ketiga dan keempat berbeda pada penggunaan optimizer yaitu SGD dan adam. Percobaan lima dan enam berbeda pada validation split yaitu 0,1 dan 0,2. Sedangkan percobaan tujuh dan delapan berbeda pada pemakaian dropout pada jaringan dan tanpa dropout.

Pada arsitektur dengan dropout, nilainya adalah 0,1 dan diletakkan pada lapisan ke tiga. Semua pelatihan yang dilakukan, dijalankan dengan batch size 32 dan 50 epoch.

Tabel 2. Desain Eksperimen

No	Percobaan	Parameter
1	P1-100	Citra 100×100 piksel
2	P2-200	Citra 200×200 piksel
3	P3-SGD	Optimizer SGD
4	P4-adam	Optimizer adam
5	P5-VS1	Validation split 0,1
6	P6-VS2	Validation split 0,2
7	P7-DO1	Dropout
8	P8-NDO	Tanpa dropout

Pengujian Model

Setelah model didapatkan berdasarkan pelatihan model yang dilakukan, proses selanjutnya adalah menguji model dengan data uji yang telah ditentukan. Banyaknya data uji dibandingkan dengan data pelatihan atau training bergantung pada setting parameter dalam pelatihan model.

Pada validation split 0.2 misalnya, hal ini berarti bahwa jumlah perbandingan data training dengan data pengujian atau validasi adalah 80% berbanding 20% dari keseluruhan data yang tersedia. Jika total data yang ada adalah 4000, maka data yang dipergunakan untuk pelatihan model sebanyak $4000 \times 80\% = 3200$ data,

sedangkan data pengujian adalah berjumlah $4000 \times 20\% = 800$ data.

Evaluasi

Proses akhir dari tahapan penelitian yang dilakukan adalah evaluasi model yang sudah didapatkan dengan metrik kinerja (*performance metrics*). Evaluasi pada model yang dilakukan dengan metrik kinerja adalah menghitung nilai dari *accuracy*, *recall*, *precision* dan *f1-score*.

Metrik kinerja diperlihatkan pada persamaan (1), (2), (3) dan (4). Persamaan ini diturunkan dari *confusion matrix*, yang terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dalam hal ini TP adalah prediksi kelas positif sebagai positif, TN adalah prediksi kelas negatif sebagai negative, FP adalah prediksi kelas negatif sebagai positif dan FN adalah prediksi kelas positif sebagai negative [5].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

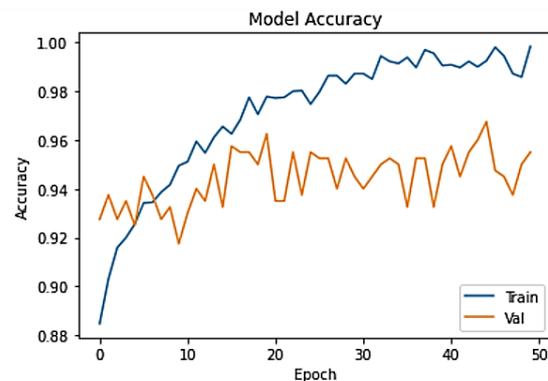
HASIL DAN PEMBAHASAN

Hasil akurasi validasi untuk setiap percobaan yang dilakukan diperlihatkan pada Tabel 3. Berdasarkan tabel, yang terbaik nilai akurasinya didapatkan pada percobaan P4-adam.

Tabel 3. Akurasi Hasil Penelitian

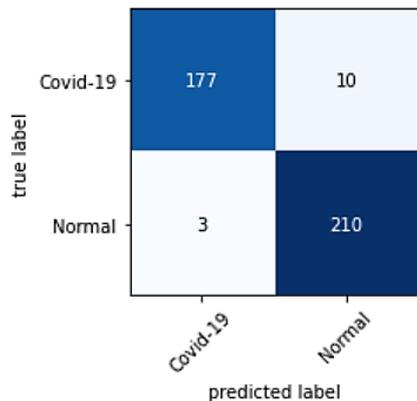
No	Percobaan	Akurasi
1	P1-100	0.9275
2	P2-200	0.9025
3	P3-SGD	0.9525
4	P4-adam	0.9675
5	P5-VS1	0.9450
6	P6-VS2	0.9425
7	P7-DO1	0.9275
8	P8-NDO	0.9662

Parameter selengkapnya yang digunakan pada percobaan model P4-adam adalah *image size* = 100×100 , *epoch* = 50, *batch_size* = 32, *validation_split* = 0.1, *dropout rate* = 0.1, optimizer yaitu adam. Hasil training dan validasi yang dilakukan sampai dengan *epoch* 50 diperlihatkan pada Gambar 3. Berdasarkan gambar *Model Accuracy* ini dapat dilihat bahwa model dapat melakukan generalisasi data dengan cukup baik. Pada akurasi training maupun validasi, grafik yang dihasilkan terlihat fluktuatif, yang menunjukkan bahwa model belum mencapai konvergen. Hal ini dapat terjadi karena parameter yang dipergunakan belum sesuai dengan data yang dilatih.



Gambar 3. Akurasi Train dan Val

Evaluasi model P4-adam dengan metrik kinerja secara keseluruhan adalah untuk *accuracy* = 0.9675, *precision* = 0.9465, *recall* = 0.9833 dan *f1-score* = 0.9645. Hasil akurasi validasi 96,75% ini menunjukkan bahwa model dapat mengenali data yang belum diketahui dengan sangat baik. Selain itu untuk nilai *precision*, *recall* dan *f1-score* yang tinggi lebih dari 94% memperlihatkan bahwa model dapat mengenali dengan cukup berimbang citra *chest x-ray* untuk data Normal maupun data Covid-19.



Gambar 4. Confusion Matrix data validasi

Pada *confusion matrix* hasil prediksi data validasi seperti diperlihatkan pada Gambar 4, prediksi benar untuk kelas Covid-19 adalah 177, sedangkan yang salah prediksi sebanyak 10. Untuk kelas Normal, hasil prediksi benar adalah 210 sementara yang salah prediksi adalah 9.

PENUTUP

Penelitian ini menggunakan metode *Convolutional Neural Network* untuk mendeteksi pasien terdeteksi Covid-19 dan pasien Normal melalui *Covid-19 Radiography Dataset* dari *Kaggle*. Melalui *tuning* parameter pada ukuran citra 100x100, menggunakan jaringan *dropout* 0.1, *validation split* 0.1, dengan jenis optimizer adam, diperoleh nilai *accuracy* sebesar 0.9675 dengan *precision* 0.9465, *recall* 0.9833 dan *f1-score* 0.9645. Hasil ini secara umum menunjukkan bahwa meskipun model belum cukup konvergen, tetapi dapat mengenali data yang belum dikenal dengan sangat baik.

Rencana penelitian selanjutnya terkait dengan deteksi Covid-19 melalui citra *chest x-ray* adalah untuk klasifikasi *multiclass*, yaitu data yang dipergunakan tidak saja untuk pasien normal dan terdeteksi Covid-19, tetapi juga untuk data pasien *lung opacity* dan *viral pneumonia*. Metode yang dipergunakan dengan cara melakukan *fine tuning* parameter pada arsitektur ImageNet, VGG 16 atau ResNet 50. Selain itu untuk mendapatkan *performance metrics* yang terbaik, dimungkinkan untuk melakukan *preprocessing* pada citra yang digunakan,

melalui proses *contrast enhancement* maupun *augmentation*.

DAFTAR PUSTAKA

- [1] Martínez Chamorro, E. et al. "Radiologic diagnosis of patients with COVID-19" [Diagnóstico radiológico del paciente con COVID-19]. *Radiología* vol. 63,1 (2021): 56–73. doi:10.1016/j.rxeng.2020.11.001
- [2] Google Berita. "Virus corona (COVID-19)". Diakses pada 14 Agustus 2021, dari <https://news.google.com/covid19/map?hl=id&gl=ID&ceid=ID%3Aid>
- [3] Mizuho N., Shunjiro N., Hidetoshi M. & Takamichi M., Automatic classification between Covid-19 pneumonia, non-Covid-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods, *Scientific Reports*, (2020), 10:17532, <https://doi.org/10.1038/s41598-020-74539-2>.
- [4] R. Mostafiz, Mohammad Shorif Uddin, Nur-A- Alam et al., Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features, *Journal of King Saud University – Computer and Information Sciences*, <https://doi.org/10.1016/j.jksuci.2020.12.010>.
- [5] Boran S. and Ilker O., Detection of Covid-19 from Chest X-Ray Images Using Convolutional Neural Networks, *SLAS Technology* 2020, Vol. 25(6) 553 – 565. DOI: 10.1177/2472630320958376.
- [6] Fatima A. S. and Inigo B., Covid-19 Detection in Chest X-ray Images using in a Deep Learning Approach, *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 6, No. 2, 30 April 2020.
- [7] Kiran P., Abhishek K., Dakshina R. K., Mamata D., COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image

- Augmented Deep Learning Model,
Cold Spring Harbor Laboratory,
2020.
- [8] Luca Brunese et al., Machine learning for coronavirus covid-19 detection from chest x-rays, 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Procedia Computer Science 176 (2020), 2212-2221.
- [9] Vina Ayumi dan Ida Nurhaida, JSAI: Journal Science and Applied Informatics, Vol.4 No. 2, Juni 2021, hal. 147 – 153. E-ISSN: 2614-3054, DOI: 10.36085.
- [10] Yuli S.H., Sugondo H., Thomhert S. S., Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network, ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi & Teknik Elektronika, Vol. 8, No. 2, hl. 443 – 453, Mei 2020.
- [11] Duc Minh L. N., Detect Covid-19 from Chest X-Ray images using Deep Learning, Bachelor's Thesis, Metropolia University of Applied Science, Bachelor of Engineering Information Technology, May 2020.