

DETEKSI COVID-19 PADA CITRA X-RAY DADA MENGUNAKAN MACHINE LEARNING

MALIATUL FITRIYASARI

Prodi Informatika

Universitas Bhinneka PGRI

Jl. Mayor Sujadi No.7 Tulungagung, Jawa Timur 66229

E-mail: maliatul@ubhi.ac.id

ABSTRAK

COVID-19 merupakan penyakit yang mewabah di seluruh dunia mulai tahun 2019. Penyakit ini menyerang saluran pernafasan, menyebabkan sesak nafas dan pneumonia pada gejala berat. Prosedur utama untuk mendeteksi COVID-19 adalah uji *swab* PCR, akan tetapi pemeriksaan X-ray dada dapat dijadikan metode alternatif untuk deteksi COVID-19 yang lebih cepat dan terjangkau. Penelitian ini bertujuan untuk membuat suatu model yang mampu mendeteksi COVID-19 dari citra X-ray dada menggunakan *machine learning* yang sederhana, yaitu jaringan syaraf tiruan yang digabungkan dengan teknik pengolahan citra. Pelatihan dilakukan menggunakan algoritma *backpropagation* yang dimodifikasi. Fitur yang di ekstrak adalah fitur statistik dari citra. Kami menggunakan data latih sebanyak 154 citra, dan data uji sebanyak 41 citra. Akurasi tertinggi yang dicapai dalam penelitian ini sebesar 87,8 %.

Kata Kunci: COVID-19; X-ray dada; Machine learning; Jaringan syaraf tiruan

I. PENDAHULUAN

COVID-19 adalah penyakit yang disebabkan oleh virus SARS-CoV-2 yang menyerang saluran pernafasan manusia. Virus ini pertama kali ditemukan di kota Wuhan, provinsi Hubei, China. Penyakit ini menyebar dengan cepat ke seluruh dunia dan berubah menjadi pandemi global. Lebih dari 219 juta orang dilaporkan telah tertular COVID-19, dan 4,7 juta diantaranya dilaporkan meninggal dunia. Penularan COVID-19 yang cepat ini membuat sistem kesehatan di beberapa negara kolaps. Para Ilmuwan diseluruh dunia berlomba-lomba untuk melakukan penelitian guna mencari solusi untuk membantu sistem kesehatan, baik itu untuk membantu proses diagnosis, pengobatan, vaksinasi, maupun prediksi kenaikan kasus.

Virus SARS-CoV-2 menyerang organ pernafasan, dimana target utamanya adalah organ paru-paru manusia. Prosedur utama untuk mendeteksi COVID-19

adalah *Swab PCR test*, yang bekerja dengan mendeteksi keberadaan material genetik virus SARS-CoV-2 pada cairan yang ada di saluran pernafasan manusia. Sementara itu, mengingat serangan virus ini dapat menyebabkan pneumonia pada paru-paru, yaitu suatu kondisi dimana organ paru-paru terisi cairan akibat infeksi virus, hal ini menjadi sebab munculnya *ground glass opacity*, yaitu area yang lebih terang atau bercak abu-abu pada citra radiologi paru-paru (Schmitt et al.,2020). Selanjutnya, hal inilah yang menjadikan X-ray dapat dijadikan sebagai prosedur alternatif untuk mendeteksi COVID-19 yang lebih cepat dan terjangkau dibandingkan dengan uji *swab PCR*.

Penelitian deteksi COVID-19 pada citra X-ray juga pernah dilakukan oleh beberapa peneliti sebelumnya. Yudistira dkk (2020) menggunakan *deep learning CNN* untuk mendeteksi COVID-19 pada citra X-ray paru-paru dengan memilih teknik pertukaran *channel ResNet 50* agar model yang dihasilkan memiliki beban komputasi lebih ringan, dengan akurasi tertinggi yang dicapai sebesar 90,8%. Selanjutnya, Hariyani dkk (2020) juga melakukan penelitian deteksi COVID-19 pada citra X-ray dada menggunakan *deep learning* berbasis *residual deep network*, dimana model yang dihasilkan memiliki nilai *precision, recall, F1*, dan akurasi masing-masing adalah 0,98, 0,95, 0,97 dan 99%. Selain itu, Fattah dkk (2021) juga membuat model untuk mendeteksi COVID-19 pada citra X-ray menggunakan *deep learning convolutional neural network* dengan arsitektur GoogleNet. Model yang dibuat mampu mencapai akurasi 97,22%, sensitivitas 96,71%, dan spesifisitas 98,63%.

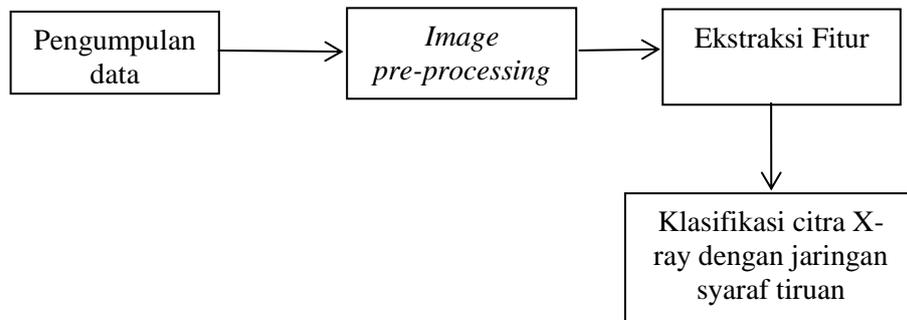
Selain menggunakan *deep learning*, deteksi COVID-19 pada citra X-ray juga dapat dilakukan menggunakan *machine learning* sederhana yang dikombinasikan dengan teknik pengolahan citra digital dan ekstraksi fitur. Wahyudi dkk (2021) membuat model deteksi COVID-19 pada citra X-ray dengan fitur parameter dari GLCM yaitu jarak, sudut, *energy, contrast, correlation, variance, homogeneity* dan *entropy*. Setelah itu, citra diklasifikasikan menggunakan *K-Nearest Neighbour*. Akurasi yang dicapai pada penelitian ini adalah 85,6%. Nugroho (2021) Juga menggunakan KNN untuk mengklasifikasikan citra X-ray kedalam kelas terinfeksi COVID-19 dan tidak terinfeksi COVID-19,

dengan metode ekstraksi fitur *Haralick Features* dan *Histogram oriented gradient* (HOG), model yang dihasilkan dapat mencapai akurasi tertinggi 99%. Selanjutnya, Putra dkk (2021) melakukan deteksi COVID-19 menggunakan jaringan syaraf tiruan *Backpropagation* dengan metode ekstraksi fitur GLCM yang meliputi *Correlation*, *Energy*, *Contrast*, dan *Homogeneity*. Jaringan syaraf tiruan yang dihasilkan mencapai akurasi 73%.

Pada penelitian ini dibuat suatu model untuk mendeteksi COVID-19 pada citra X-ray dada, menggunakan teknik pengolahan citra digital, dengan pemilihan fitur nilai statistik dari citra, serta menggunakan *machine learning* sederhana yaitu jaringan syaraf tiruan *backpropagation* sebagai *classifier*. Jaringan syaraf tiruan dipilih karena memiliki beban komputasi yang lebih ringan, serta waktu training lebih singkat, sehingga lebih efisien.

II. METODE PENELITIAN

Secara garis besar, penelitian ini terdiri dari 4 tahapan yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pada penelitian ini dilakukan pengumpulan data dari dua sumber yang tersedia di Internet. Data dikumpulkan dari *website Italian society of Medical and Interventional Radiology* (<https://sirm.org/covid-19/>), dan GitHub (<https://github.com/muhammedtalo/COVID-19>). Total data yang terkumpul sebanyak 205 citra X-ray, dengan rincian 101 citra terinfeksi COVID-19, dan 104 citra tidak terinfeksi COVID-19. Data citra yang terkumpul diambil dari lokasi dan peralatan yang berbeda-beda, hal ini mengakibatkan terjadinya perbedaan intensitas dan kontras, yang membawa dampak buruk terhadap proses klasifikasi menggunakan jaringan syaraf tiruan.

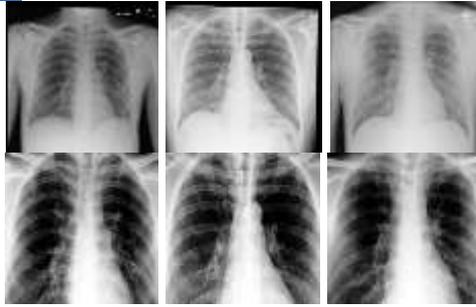


Gambar 1. Citra X-ray dada terinfeksi COVID-19 (dua gambar di kiri) dan citra X-ray dada tidak terinfeksi COVID-19 (dua gambar di kanan)

2.2. Image pre-processing

Tahap *pre-processing* sangat penting dilakukan sebelum ekstraksi fitur. Pada tahapan ini, dilakukan *cropping* untuk mengambil bagian dari citra yang diperlukan dalam proses klasifikasi (bagian paru-paru) dan meninggalkan bagian lain yang tidak diperlukan. *Cropping* dilakukan pada semua citra, dengan cara manual.

Setelah *cropping*, tahapan dalam *pre-processing* ini adalah *image adjustment*. Citra X-ray dada yang dikumpulkan memiliki rentang nilai intensitas yang berbeda-beda, sehingga *image adjustment* diperlukan agar intensitas semua citra seragam. Proses ini dilakukan menggunakan *built in function* yang sudah tersedia di matlab, yaitu *imadjust()*. Fungsi ini secara otomatis melakukan penyesuaian intensitas pada citra secara linear, selain itu kontras gambar juga akan meningkat dan menjadi seragam. Selain itu, dalam tahap *pre-processing* ini, semua citra dirubah ke dalam bentuk *grayscale*.



Gambar 3. Citra asli (atas) citra hasil *pre-processing* (bawah).

2.3. Ekstraksi Fitur

Pada tahap ini, dilakukan ekstraksi atau pengambilan fitur pada citra. Dalam penelitian ini, dipilih fitur statistik dari citra yang meliputi :

1. Prosentase jumlah *pixel* yang nilainya berada pada rentang 0 - 86
2. Prosentase jumlah *pixel* yang nilainya berada pada rentang 87 - 171
3. Prosentase jumlah *pixel* yang nilainya berada pada rentang 172 - 255
4. Entropi

Entropi merupakan ukuran keserampangan (*randomness*) dari suatu data. Entropi dapat digunakan untuk analisis tekstur dari suatu citra. Entropi dapat dihitung menggunakan rumus :

$$Entropi = \sum_{j=1}^k -p_j \log_2 p_j \quad (1)$$

Dimana k merupakan jumlah kelas, dan p_j merupakan proporsi sampel untuk kelas j , keduanya diperoleh dari histogram yang telah dinormalisasi.

5. Standar deviasi

Standar deviasi adalah nilai yang menunjukkan sebaran data terhadap nilai rata-rata. Standar deviasi dihitung menggunakan rumus :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2)$$

6. Rata-rata

Nilai rata-rata merupakan jumlah nilai pada semua *pixel* dibagi dengan

banyaknya *pixel*. Nilai rata rata dari suatu data dapat dikatakan mewakili data tersebut. Nilai rata-rata dapat dihitung menggunakan rumus :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

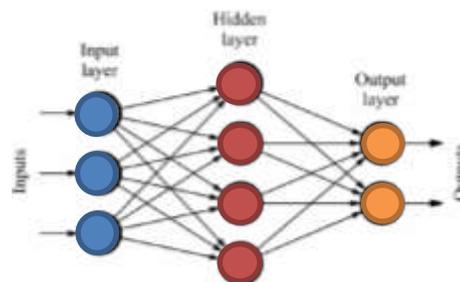
7. Varians

Varians adalah suatu nilai yang menunjukkan seberapa jauh sebuah kumpulan bilangan tersebar nilainya. Varians dapat dihitung menggunakan rumus :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

2. 4. Klasifikasi menggunakan Jaringan syaraf tiruan

Fitur yang telah diekstrak dari citra selanjutnya akan dijadikan sebagai dasar untuk mengklasifikasikan citra ke dalam kelas COVID-19 dan bukan COVID-19 menggunakan jaringan syaraf tiruan. Jaringan syaraf tiruan yang dibuat dalam penelitian ini memiliki arsitektur sederhana, dengan satu lapisan *input*, satu lapisan tersembunyi (*hidden layer*), dan satu lapisan *output*.

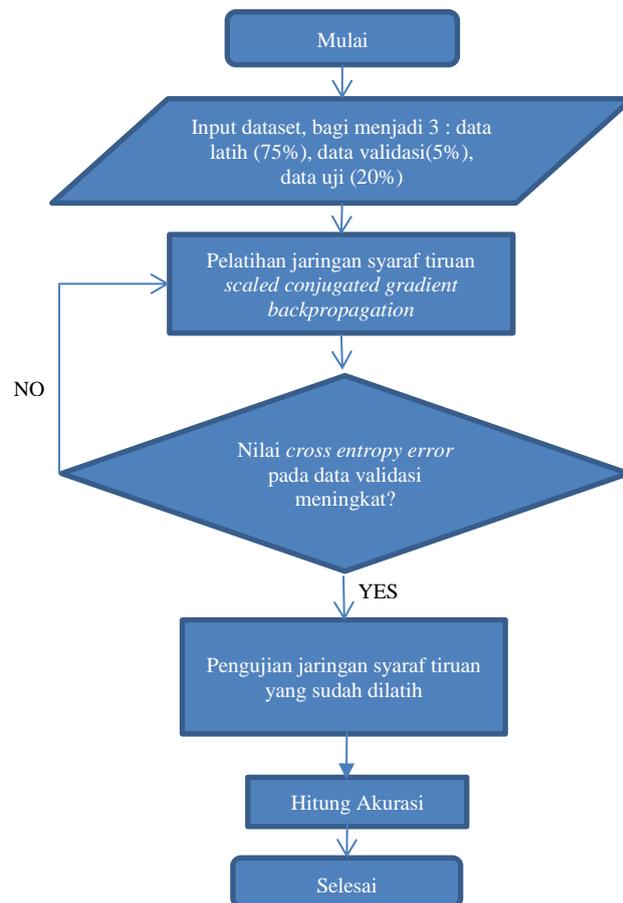


Gambar 4. Arsitektur jaringan syaraf tiruan sederhana

Didalam proses klasifikasi menggunakan jaringan syaraf tiruan, ada 2 tahapan yang dilakukan, yaitu tahap pelatihan dan pengujian. Pada tahap pelatihan, digunakan algoritma *scale conjugate gradient backpropagation* (Moller, 1991). Algoritma ini merupakan modifikasi dari *backpropagation*. Proses pelatihan pada jaringan syaraf tiruan akan dihentikan ketika generalisasi dari jaringan syaraf tiruan sudah tidak lagi meningkat, hal ini ditandai dengan peningkatan *cross entropy error* pada data yang dijadikan sampel validasi.

Setelah proses pelatihan berhenti, dilakukan pengujian untuk mengetahui

kemampuan jaringan syaraf tiruan yang telah dibuat dalam mengenali atau mengklasifikasikan data yang berbeda dengan data pelatihan. Secara umum, alur proses pelatihan dan pengujian jaringan syaraf tiruan pada penelitian ini digambarkan pada *flowchart* yang ada pada gambar 5.



Gambar 5. *Flowchart* klasifikasi citra menggunakan jaringan syaraf tiruan

III. HASIL DAN PEMBAHASAN

Jaringan syaraf tiruan yang dibuat dalam penelitian ini memiliki 1 *input layer*, 1 *hidden layer*, dan 1 *output layer*. Pada *input layer*, terdapat 7 buah *neuron*, sesuai dengan jumlah fitur yang diekstrak dari citra. Selanjutnya, pada *output layer* terdapat 1 *neuron*. *Output* hanya ada 2 kelas, yaitu COVID-19 dan bukan COVID-19. Pada *hidden layer*, jumlah *neuron* dibuat bervariasi yaitu 6, 8,10, dan 12.

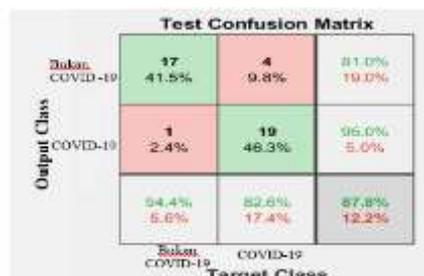
Selanjutnya, pembagian data dalam penelitian ini adalah 75% untuk data

latih, 5% untuk data validasi, dan 20% untuk data uji. Data uji sebanyak 41 citra digunakan untuk menguji performa jaringan syaraf tiruan yang telah dilatih. Hasil pengujian jaringan syaraf tiruan tertera pada tabel 1.

Tabel 1. Hasil akurasi pengujian jaringan syaraf tiruan

Jumlah neuron pada hidden layer	Akurasi
6	82,1 %
8	87,8 %
10	83 %
12	85,4 %

Dari hasil percobaan, diketahui bahwa jaringan syaraf tiruan dengan 6 *neuron* pada *hidden layer* menghasilkan akurasi terendah, sedangkan jaringan syaraf tiruan dengan 8 *neuron* pada *hidden layer* memiliki akurasi yang paling tinggi, yaitu 87,8%. Pelatihan pada jaringan syaraf tiruan dengan 8 *neuron* dilakukan dengan iterasi sebanyak 16 kali, dalam waktu kurang dari 1 detik. Hasil dari pengujian model ini dapat dilihat pada gambar 6 dalam bentuk *confusion matrix*.



Gambar 6. *Confusion matrix* pengujian jaringan syaraf tiruan dengan 8 *neuron* pada *hidden layer*

Berdasarkan tabel *confusion matrix*, jumlah citra X-ray paru-paru COVID-19 yang diklasifikasikan ke dalam kelas bukan COVID-19 (*misclassification*) cukup banyak yaitu 4 citra, sedangkan citra X-ray paru-paru bukan COVID-19 yang diklasifikasikan ke dalam kelas COVID-19 ada 1 citra. Data citra X-ray penderita COVID-19 lebih banyak mengalami *misclassification*, karena fitur statistik dari citra tersebut memiliki nilai yang polanya lebih tidak menentu dibanding citra bukan COVID-19.

Selanjutnya, perbandingan hasil akurasi dari penelitian ini dengan penelitian

lain yang menggunakan metode jaringan syaraf tiruan untuk mendeteksi COVID-19 pada citra X-ray dapat dilihat pada tabel 2.

Tabel 2. Perbandingan akurasi dengan penelitian lain

No.	Metode	Akurasi
1.	Putra & Yuhandri, 2021	73%
2.	Metode yang diusulkan	87,8 %

Dari perbandingan diatas, dapat disimpulkan bahwa hasil akurasi dari penelitian menghasilkan akurasi yang lebih tinggi dari penelitian sebelumnya (Putra & Yuhandri, 2021) dengan keunggulan sebesar 14,8%.

IV. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa deteksi COVID-19 pada citra X- ray dada menggunakan *machine learning* sederhana yaitu jaringan syaraf tiruan yang dikombinasikan dengan teknik pengolahan citra digital, dengan ekstraksi fitur statistik dari citra, dapat menghasilkan akurasi sistem yang cukup tinggi. Akurasi tertinggi yang dicapai pada penelitian ini sebesar 87,8%. Akurasi ini dicapai oleh jaringan syaraf tiruan dengan arsitektur 7 *neuron* pada *input layer*, 1 *hidden layer* dengan 8 *neuron* di dalamnya, dan 1 *neuron* pada *output layer*. Metode yang diusulkan juga mampu menghasilkan akurasi yang lebih tinggi dibanding penelitian sebelumnya yang menggunakan jaringan syaraf tiruan.

DAFTAR PUSTAKA

- Antony, F., Irsyad, H. and Al Rivan, M.E., 2021. KNN Dan Gabor Filter Serta Wiener Filter Untuk Mendiagnosis Penyakit Pneumonia Citra X-RAY Pada Paru-Paru. *Jurnal Algoritme*, vol 1(2) : 147-155
- DARMATASIA, D., PENGENALAN SISTEM ISYARAT BAHASA INDONESIA (SIBI) MENGGUNAKAN GRADIENT-CONVOLUTIONAL NEURAL NETWORK. *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol 6(1) : 56-65.
- Fattah, M.S., 2021. Deteksi penyakit pneumonia dan COVID-19 menggunakan citra X-ray dengan metode Convolutional Neural Network (CNN) model

- GoogleNet [disertasi]. Surabaya : UIN Sunan Ampel Surabaya.
- Fitriyarsari, M., 2016. Deteksi Osteoporosis berdasarkan Citra CT-Scan menggunakan Jaringan Saraf Tiruan [skripsi]. Surabaya: Universitas Airlangga.
- HARIYANI, Y.S., HADIYOSO, S. and 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*, vol 8(2) : 443.
- Kushwaha, S., Bahl, S., Bagha, A.K., Kulwinder Singh, P., Javaid, M., Haleem, A. and Singh, R.P., 2020. Significant applications of machine learning for COVID-19 pandemic. *Journal of Industrial Integration and Management*.
- Nugroho, C.A., 2021. Klasifikasi K-Nearest Neighbor Chest X-ray Pasien COVID-19 Dengan Haralick Features dan Histogram Of Oriented Gradient. *J. Ilm. Mat*, vol 9(1) : 188-195.
- Putra, H.R.W. and Yuhandri, Y., 2021. Identifikasi Penderita COVID-19 Berdasarkan Chest X-ray Menggunakan Algoritma Jaringan Syaraf Tiruan Backpropagation. *Jurnal Sistim Informasi dan Teknologi*, vol 3(4) : 197-202.
- Schmitt, W. and Marchiori, E., 2020. Covid-19: round and oval areas of ground-glass opacity. *Pulmonology*, vol 26(4) : 246.
- Wahyudi, D.S., Patmasari, R. and Fu'adah, R.Y.N., 2021. Sistem Deteksi COVID-19 Berdasarkan Citra X-ray Menggunakan Metode Gray Level Co-occurrence Matrix (glcm) Dan K-nearest Neighbour (k-nn). *eProceedings of Engineering*, vol 8(1). 2021 Feb 1
- Yudistira, N., Widodo, A.W. and Rahayudi, B., 2020. Deteksi COVID-19 pada Citra Sinar-X Dada Menggunakan Deep Learning yang Efisien. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol 7(6)

