

DETEKSI COVID-19 PADA CITRA CT-SCAN MENGGUNAKAN K-NEAREST NEIGHBOR

Maliatul Fitriyasari*¹⁾

1. Universitas Bhinneka PGRI, Indonesia

Article Info

Kata Kunci: K-Nearest Neighbour, COVID-19, CT-Scan, image processing

Keyword: K-Nearest Neighbour, COVID-19, CT-Scan, image processing

Article history:

Received 4 January 2023

Revised 23 January 2023

Accepted 2 February 2023

Available online 1 March 2023

DOI :

<https://doi.org/10.29100/jupi.v8i1.3996>

* Corresponding author.

Corresponding Author

E-mail address:

maliatul@ubhi.ac.id

ABSTRAK

Covid 19 adalah penyakit yang disebabkan oleh virus SARS-CoV-2 yang menyerang organ pada sistem pernafasan manusia, dengan sasaran utama organ paru-paru. Pemeriksaan utama untuk diagnosis COVID 19 adalah melalui tes swab PCR, akan tetapi pemeriksaan tersebut membutuhkan waktu yang relative lama. Sementara itu, pemeriksaan CT Scan dapat menjadi alternatif untuk mendeteksi COVID 19 dengan cepat, karena terdapat ground glass opacity pada citra radiologi paru-paru yang terinfeksi covid 19. Pada penelitian ini dilakukan eksperimen untuk mendeteksi covid 19 pada citra CT Scan dada, menggunakan K-nearest neighbor yang ringan dan efisien, dikombinasikan dengan metode segmentasi paru-paru, serta ekstraksi fitur nilai prosentase area terang dan nilai prosentase area gelap pada citra paru-paru. Akurasi tertinggi yang dicapai pada penelitian adalah 98,44%.

ABSTRACT

Covid 19 is a disease caused by the SARS-CoV-2 virus, which attacks the organs in the human respiratory system, primarily targeting the lungs. The standard examination tool for diagnosing COVID-19 is PCR swab test, but this examination usually needs a long time to get the result. Meanwhile, a CT scan can be an alternative examination tool for detecting COVID-19 more quickly, because there is a different ground glass opacity in radiological images of lungs which infected by covid-19. In this study, experiments were carried out to detect covid-19 on chest CT scan images, using a simple K-nearest neighbor (K-NN) method, combined with the image segmentation method, using 2 features which are the percentage value of the bright area and the percentage value of the dark area in lung images. The highest accuracy achieved in the study was 98.44%.

PENDAHULUAN

Covid 19 merupakan wabah yang muncul di Wuhan pada tahun 2019, dan menyebar ke seluruh dunia dalam waktu yang sangat singkat. Wabah ini disebabkan oleh virus SARS-CoV-2 (*severe acute respiratory syndrome coronavirus 2*). Penyakit ini menyerang sistem pernafasan manusia, terutama paru-paru serta dapat menyebabkan pneumonia. Gejala yang ditimbulkan antara lain demam, batuk, sakit tenggorokan, anosmia, kehilangan kemampuan indra pengecap, dan sesak nafas.

Pemeriksaan standar untuk mendeteksi Covid 19 adalah melalui tes PCR (*polymerase chain reaction*), yaitu dengan mendeteksi keberadaan materi genetik virus SARS-CoV-2 pada lendir atau cairan yang ada di saluran nasofaring. Akan tetapi, hasil dari tes ini seringkali baru bisa diketahui dalam waktu yang cukup lama. Sementara itu, pemeriksaan radiologis seperti CT-Scan dada juga bisa dijadikan alternatif untuk mendeteksi penyakit ini (T. Ai et al, 2020). Pada pemeriksaan radiologis diketahui bahwa ketika penyakit covid-19 menyerang paru-paru, akan menimbulkan *ground glass opacity* (GGO), yaitu munculnya area yang lebih terang atau bercak keabuan pada citra radiologis paru-paru (Schmitt et al, 2020).

Dengan berkembangnya teknologi komputasi saat ini, deteksi covid-19 pada citra radiologis dapat dilakukan menggunakan metode kecerdasan buatan maupun pengolahan citra digital. Beberapa penelitian untuk mendeteksi covid-19 pada citra CT-Scan dada menggunakan kecerdasan buatan *machine learning* dan teknik pengolahan citra digital telah dilakukan. Swastika (2020) melakukan studi deteksi covid-19 melalui citra CT-Scan dada menggunakan *deep learning* CNN (*Convolutional neural network*) dengan arsitektur VGG16 dan optimizer SGD dan Adam, hasil dari penelitian tersebut diperoleh model pendeteksi covid-19 dengan akurasi sebesar 92,86%. Zhao et al.(2021) menggunakan *deep learning* CNN yang dikombinasikan dengan *transfer learning* dan dilatih

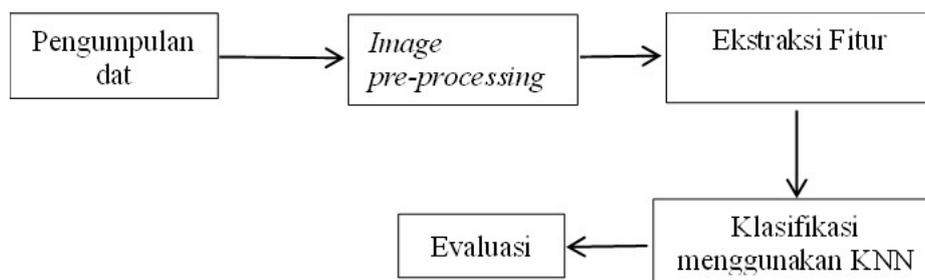
menggunakan dataset yang berjumlah banyak serta dataset *out of field* untuk mendeteksi covid-19. Model pendeteksi covid-19 yang dihasilkan mampu mencapai akurasi sebesar 99,2 %.

Selain menggunakan *deep learning*, deteksi covid-19 pada citra CT-Scan juga dapat dilakukan menggunakan *machine learning* yang lebih sederhana (*shallow learning*). Keunggulan dari *shallow learning* adalah memiliki beban komputasi yang lebih ringan dan proses pelatihan yang relative singkat. Wuryani & Agustiani (2021) memanfaatkan *random forest classifier* untuk mendeteksi covid-19 pada citra CT-Scan paru-paru yang telah tersegmentasi dengan memasukkan fitur *haralick* dan *color histogram* dari citra. Hasil akhir yang diperoleh yaitu akurasi sebesar 96,9%. Selanjutnya, Nugroho dkk. (2021) menggunakan algoritma KNN (*K-nearest neighbor*) yang dikombinasikan dengan *genetic algorithm* untuk pemilihan fitur, dengan pilihan fitur *haralick* dan *local binary pattern*. Akurasi yang dihasilkan sebesar 93,30%. Antony dkk. (2021) menggunakan KNN dengan metode ekstraksi fitur *gabor filter* dan *wiener filter* untuk mendeteksi pneumonia pada citra CT Scan paru-paru, akurasi yang dicapai sebesar 79,62%.

Dalam penelitian ini, diusulkan suatu metode yang menggunakan *machine learning* KNN untuk mendeteksi Covid-19 pada citra CT-Scan dada yang dikombinasikan dengan metode pengolahan citra digital untuk segmentasi citra paru-paru. Pada penelitian ini, *thresholding* akan dilakukan pada citra paru-paru untuk mendeteksi area pada paru yang lebih terang (mengalami *ground glass opacity*). Terakhir, prosentase luas area yang lebih terang dan prosentase luas area yang gelap pada citra paru-paru dijadikan sebagai fitur masukan untuk model KNN yang akan dibuat.

METODE PENELITIAN

Secara garis besar, penelitian ini terdiri dari 5 tahapan yang dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

A. Pengumpulan Data

Pada penelitian ini, dilakukan pengumpulan data berupa citra CT-Scan dada dari beberapa sumber. Data dari pasien yang terinfeksi covid-19 yang berjumlah 105 citra diambil dari *Italian society of Medical and Interventional Radiology*¹. Data tersebut dikumpulkan dari beberapa rumah sakit yang ada di Italia. Sedangkan 132 data citra CT Scan dari pasien yang tidak terinfeksi covid 19 diambil dari *VIA Management System- Cornell University*². Total data ada 237 citra CT Scan digunakan dalam penelitian ini.

Citra CT Scan yang terkumpul diambil dari peralatan yang berbeda, hal ini menjadikan intensitas dan kontras dari citra tersebut tidak seragam. Selain itu, Citra CT Scan dada yang diperoleh masih mengandung gambar organ tubuh selain paru-paru. Untuk itu, perlu dilakukan *pre-processing* untuk menyeragamkan intensitas dan segmentasi citra paru-paru.

B. Image pre-processing

Image pre-processing dilakukan dalam 3 tahapan, yaitu :

Tahap 1. Image Adjustment

Citra CT Scan dada yang terkumpul memiliki intensitas dan kontras yang tidak seragam. *Image Adjustment* diterapkan untuk menyeragamkan intensitas dan kontras citra tersebut. *Image adjustment* dilakukan menggunakan *built-in function* yang tersedia di matlab, yaitu *imadjust()*.

Tahap 2. Median Filtering

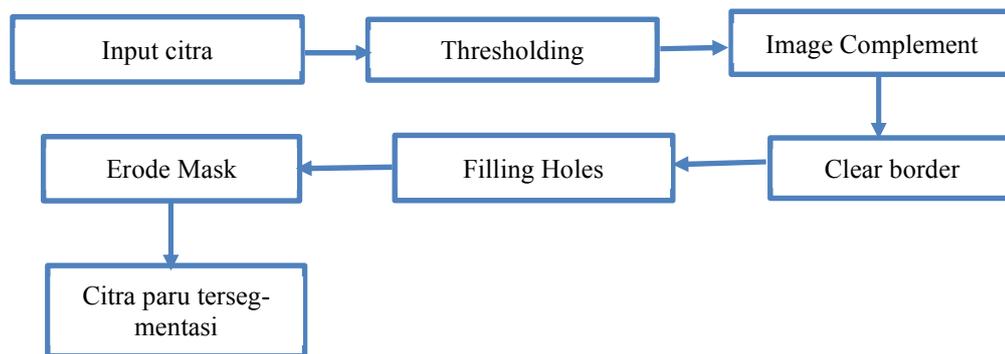
¹ <https://sirm.org/covid-19/>

² <https://veet.via.cornell.edu/cgi-bin/datac/signon.cgi>

Median filtering dilakukan untuk mengurangi *noise* atau derau yang ada pada citra berupa *salt and pepper noise*. Median filtering bekerja dengan mengganti nilai tiap piksel dengan nilai median dari piksel-piksel tetangga. Hasil dari median filtering adalah citra yang lebih halus. *Median filtering* dilakukan menggunakan *built-in function* yang tersedia di matlab, yaitu *medfilt2()*.

Tahap 3. Segmentasi Citra

Citra CT Scan dada berisi gambar organ-organ tubuh yang ada di dalam dada. Segmentasi dilakukan untuk memisahkan gambar organ paru-paru dengan organ lainnya di dalam dada. Hal ini sangat penting dilakukan karena organ tubuh yang diperlukan untuk dianalisis dalam mendeteksi covid-19 hanya organ paru-paru, sehingga jika gambar organ tubuh lainnya tidak dihilangkan dipisahkan dapat mengganggu proses klasifikasi citra ditahap selanjutnya. Proses segmentasi dilakukan dengan metode yang ada pada *website mathworks*³, yang secara umum dapat dituliskan dalam sebuah *flowchart* pada gambar 2. Sebagai tambahan, semua tahapan segmentasi dilakukan menggunakan *built-in functions* yang sudah tersedia di Matlab.



Gambar 2. Flowchart segmentasi citra paru-paru

C. Ekstraksi Fitur

Fitur merupakan ciri khusus yang mewakili citra. Munculnya *ground glass opacity* pada citra CT Scan paru-paru penderita covid-19, yang ditandai dengan adanya area paru-paru yang lebih terang dibanding area normal atau muncul bercak abu-abu terang menjadi dasar pemilihan fitur pada penelitian ini. Proses ekstraksi fitur diawali dengan *thresholding* yang bertujuan untuk membedakan area paru yang lebih terang dengan area paru yang gelap. Nilai *threshold* dihitung secara adaptif menggunakan metode *Otsu thresholding* (Zhu et al, 2009). Selanjutnya, fitur yang diekstrak dari citra hanya dua yaitu prosentase luas area paru yang terang dan prosentase luas area paru yang gelap. Keduanya dihitung melalui rumus :

$$\text{Prosentase area teran} = \frac{\text{Luas area terang (pixel)}}{\text{Luas area paru - paru (pixel)}} \times 100\% \quad (1)$$

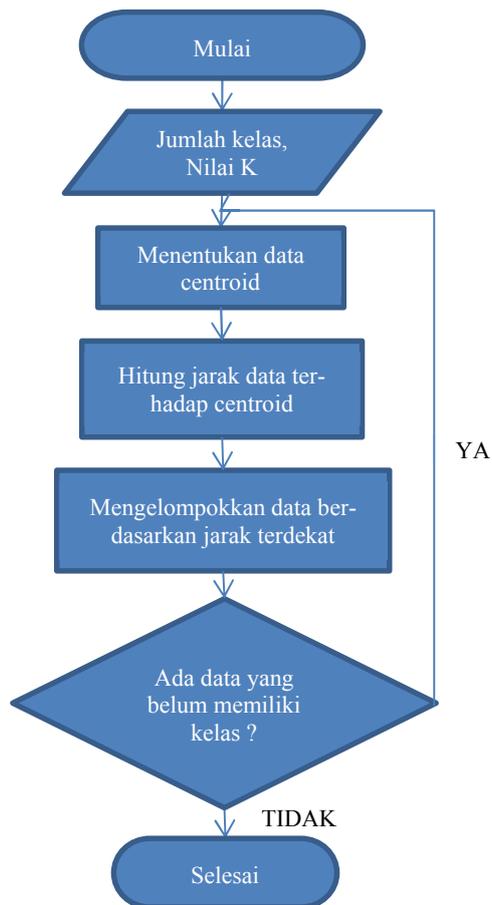
$$\text{Prosentase area gelap} = \frac{\text{Luas area gelap (pixel)}}{\text{Luas area paru - paru (pixel)}} \times 100\% \quad (2)$$

Kedua fitur diatas selanjutnya akan digunakan dalam proses klasifikasi citra menggunakan *K-Nearest Neighbor*.

D. Ekstraksi Fitur

K-nearest neighbour (KNN) merupakan suatu algoritma *machine learning* yang dapat digunakan untuk proses klasifikasi dengan metode pembelajaran tak terbimbing (*unsupervised learning*). Prinsip kerja dari KNN adalah dengan mencari jarak terdekat antara suatu data dengan sejumlah K tetangga dekatnya. Dalam KNN, nilai jarak merupakan pendekatan yang digunakan untuk menentukan kelas dari suatu data. Pengukuran jarak pada KNN dalam penelitian ini dilakukan dengan 4 metode yaitu *Euclidean*, *City-Block*, *Chebyshev*, dan *mikowski*. Flowchart klasifikasi data menggunakan KNN dituliskan pada gambar 3.

³<https://ch.mathworks.com/help/images/segment-lungs-from-3-d-chest-mri-data.html>



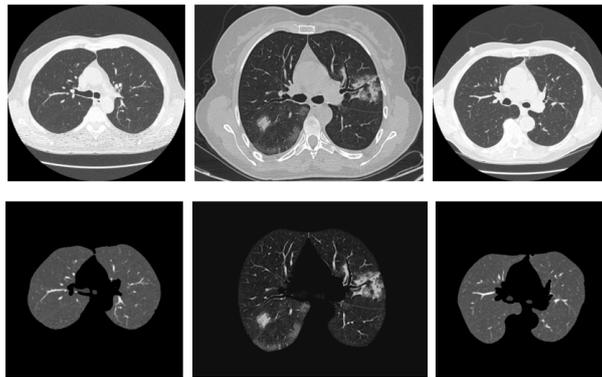
Gambar 3. Flowchart klasifikasi data menggunakan KNN

Pelatihan KNN dilakukan dengan data latih 173 citra, yang terdiri dari 73 citra paru-paru terinfeksi covid 19, dan 100 citra paru-paru tidak terinfeksi covid 19. Sedangkan data uji dalam penelitian ini ada 64 citra, dengan 32 citra paru-paru terinfeksi covid 19 dan 32 citra paru-paru tidak terinfeksi covid 19. Sehingga dalam penelitian ini, 27% data dijadikan data uji, dan 73% data dijadikan data latih. Pelatihan KNN dilakukan berulang kali dengan metode pengukuran jarak dan nilai K yang bervariasi. Setiap kali selesai pelatihan, dilakukan pengujian menggunakan data latih dan data uji untuk mengetahui akurasi dari model KNN yang telah dibuat. Akurasi dihitung menggunakan rumus :

$$Akurasi = \frac{Jumlah\ data\ yang\ terklasifikasi\ dengan\ benar}{Jumlah\ seluruh\ data} \times 100\% \quad (3)$$

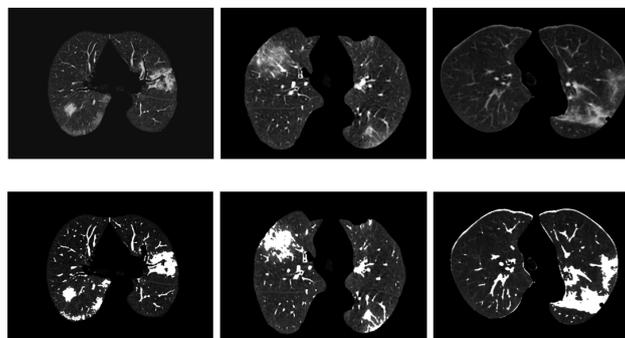
HASIL DAN PEMBAHASAN

Tahapan pertama dalam deteksi covid 19 pada citra CT Scan dalam penelitian ini adalah *image pre-processing*. Hasil dari tahapan *pre-processing* dapat dilihat pada gambar 4. Setelah melalui tahapan *pre-processing*, citra yang diperoleh adalah citra yang berisi paru-paru, tanpa ada gambar organ lain disekitarnya.



Gambar 4. Citra asli (atas) Citra setelah melalui tahap pre-processing (bawah)

Selanjutnya, ekstraksi fitur dilakukan pada citra yang telah melalui tahap pre-processing. Fitur yang digunakan pada penelitian ini hanya 2, yaitu prosentase area terang dan prosentase area gelap. Untuk membedakan area terang dan area gelap, dilakukan *thresholding* terlebih dahulu dengan menggunakan metode Otsu. Hasil dari *thresholding* dapat dilihat pada gambar 4.



Gambar 4. Citra sebelum thresholding (atas), dan setelah thresholding (bawah)

Dari citra hasil thresholding ini, bagian paru-paru yang berwarna putih dihitung sebagai area terang, sedangkan bagian paru-paru yang tidak berwarna putih dihitung sebagai area gelap. Dalam tahap ini terlihat bahwa citra CT Scan paru-paru yang terinfeksi covid 19 cenderung memiliki prosentase area terang yang lebih besar dibanding citra paru-paru yang tidak terinfeksi covid 19. Selanjutnya, kedua fitur yang telah diekstrak dijadikan sebagai input pada proses klasifikasi menggunakan KNN. Dalam membuat model KNN dilakukan eksperimen menggunakan 4 metode pengukuran jarak yaitu *Euclidean*, *Chebyshev*, *minkowski* dan *cityblock*. Selain itu, eksperimen juga dilakukan dengan variasi 3 nilai ketetanggaan (K) yang berbeda yaitu 3,5, dan 7. Hasil dari eksperimen ini tercantum dalam tabel I, tabel II, table III, dan tabel IV.

TABEL I.
 HASIL EKSPERIMEN MENGGUNAKAN JARAK EUCLIDEAN

Nilai ketetanggaan	Akurasi Latih	Akurasi Uji
3	98,84 %	96,87 %
5	98,26 %	98,44 %
7	98,26 %	98,44 %

TABEL II.
 HASIL EKSPERIMEN MENGGUNAKAN JARAK CHEBYSHEV

Nilai ketetanggaan	Akurasi Latih	Akurasi Uji
3	98,84 %	96,87 %
5	98,26 %	98,44 %
7	98,26 %	98,44 %

TABEL III.
 HASIL EKSPERIMEN MENGGUNAKAN JARAK CITYBLOCK

Nilai ketetanggaan	Akurasi Latih	Akurasi Uji
3	98,84 %	96,87 %
5	98,26 %	98,44 %
7	98,26 %	98,44 %

TABEL IV.
 HASIL EKSPERIMEN MENGGUNAKAN JARAK MINKOWSKI

Nilai ketetangaan	Akurasi Latih	Akurasi Uji
3	98,84 %	96,87 %
5	98,26 %	98,44 %
7	98,26 %	98,44 %

Dari keempat tabel diatas, dapat dilihat bahwa akurasi latih dan akurasi uji dengan nilai ketetangaan 3, 5, dan 7 bernilai tetap, meskipun metode pengukuran jarak yang dipakai berbeda-beda. Selanjutnya, dapat dilihat pada tahap pelatihan, akurasi tertinggi yang dapat diperoleh adalah 98,84%. Sedangkan pada tahap pengujian, akurasi tertinggi yang dapat diperoleh pada penelitian ini sebesar 98,44%. Hal ini menunjukkan bahwa pada tahap pengujian, dari 64 data latih ada 1 data yang tidak terklasifikasi dengan benar. Data tersebut berasal dari kelas covid 19, tetapi diklasifikasikan ke dalam kelas bukan covid 19. Hal ini terjadi karena data tersebut memiliki fitur yang hampir sama dengan data yang berada pada kelas bukan covid 19.

Dari penelitian ini, dapat dilihat bahwa metode pengukuran jarak tidak terlalu mempengaruhi akurasi model KNN yang dibuat. Sementara itu, fitur dari citra yang akan menjadi masukan KNN menjadi hal yang sangat mempengaruhi akurasi. Dua fitur yang digunakan dalam penelitian ini, yaitu prosentase area terang dan prosentase area gelap sangat dipengaruhi oleh tingkat kerusakan paru-paru penderita covid 19. Pada penderita dengan tingkat kerusakan paru-paru yang rendah, area yang mengalami *ground glass opacity* atau yang berwarna terang hanya sedikit. Sehingga KNN akan cenderung mengklasifikasikan data tersebut ke dalam kelas bukan covid 19. Sebaliknya, untuk penderita dengan tingkat kerusakan paru - paru sedang atau tinggi, maka area paru-paru yang mengalami *ground glass opacity* akan lebih luas, sehingga KNN akan cenderung mengklasifikasikan data tersebut ke dalam kelas covid 19.

Selanjutnya, perbandingan tingkat akurasi dari hasil penelitian ini dengan penelitian lain yang juga menggunakan Citra CT Scan untuk mendeteksi covid 19 serta menggunakan KNN sebagai metode klasifikasinya, tertera pada tabel V.

TABEL V.
 PERBANDINGAN AKURASI HASIL PENELITIAN

No.	Metode	Akurasi
1.	Nugroho, dkk. (2021)	93,3 %
2.	Antony, dkk. (2021)	79,62 %
3.	Metode yang diusulkan	98,44 %

Dari tabel diatas, dapat disimpulkan bahwa metode deteksi covid 19 melalui citra CT Scan yang diusulkan dalam penelitian ini menghasilkan akurasi yang lebih unggul dibanding 2 penelitian sebelumnya.

KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa deteksi covid 19 pada citra CT Scan menggunakan K-NN yang dikombinasikan dengan teknik pengolahan citra digital, dengan fitur nilai prosentase area terang dan prosentase area gelap pada citra CT Scan paru-paru, dapat menghasilkan akurasi yang cukup tinggi. Akurasi tertinggi yang dicapai pada penelitian ini sebesar 98,84% untuk data latih, dan 98,44% untuk data uji.

DAFTAR PUSTAKA

- [1] B.A. Skourt, A. El Hassani, and A. Majda, "Lung CT Image Segmentation Using Deep Neural Networks," *Procedia Computer Science*, vol. 127, pp. 109–113, Jan. 2018, doi: <https://doi.org/10.1016/j.procs.2018.01.104>.
- [2] F. Antony, H. Irsyad, and M. E. Al Rivan, "KNN Dan Gabor Filter Serta Wiener Filter Untuk Mendiagnosis Penyakit Pneumonia Citra X-RAY Pada Paru-Paru," *Jurnal Algoritme*, vol. 1, no. 2, pp. 147–155, Apr. 2021, doi: <https://doi.org/10.35957/algoritme.v1i2.893>.
- [3] M. FITRIYASARI, "DETEKSI COVID-19 PADA CITRA X-RAY DADA MENGGUNAKAN MACHINE LEARNING," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 7, no. 1, pp. 1–10, Dec. 2021, doi: <https://doi.org/10.24252/instek.v7i1.25712>.
- [4] N. Harsa Pratama, E. Rachmawati, and G. Kosala, "CLASSIFICATION OF DOG BREEDS FROM SPORTING GROUPS USING CONVOLUTIONAL NEURAL NETWORK," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, vol. 07, no. 04, pp. 1080–1087, Dec. 2022, doi: <https://doi.org/10.29100/jipi.v7i4.3208>.
- [5] N. Wuryani and S. Agustiani, "Random Forest Classifier untuk Deteksi Penderita COVID-19 berbasis Citra CT Scan," *Jurnal Teknik Komputer*, vol. 7, no. 2, pp. 187–193, Jul. 2021, doi: <https://doi.org/10.31294/jtk.v7i2.10468>.
- [6] N. Zhu, G. Wang, G. Yang, and W. Dai, "A Fast 2D Otsu Thresholding Algorithm Based on Improved Histogram," in *2009 Chinese Conference on Pattern Recognition*, Dec. 2009, pp. 1–5. doi: 10.1109/CCPR.2009.5344078
- [7] P. A. H. Pratama, R. Teguh, A. S. Sahay, and V. Wilentine, "Deteksi COVID-19 Berdasarkan Hasil Rontgen Dada (Chest Xray) Menggunakan Python," *Journal of Information Technology and Computer Science*, vol. 1, no. 1, pp. 58–67, Jun. 2021, doi: <https://doi.org/10.47111/jointcoms.v1i1.2956>.

- [8] R. A. Nugroho, A. S. Nugraha, A. A. Rasyid, and F. W. Rahayu, "Improvement on KNN using genetic algorithm and combined feature extraction to identify COVID-19 sufferers based on CT scan image," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 5, p. 1581, Oct. 2021, doi: <https://doi.org/10.12928/telkomnika.v19i5.18535>.
- [9] Sadly Syamsuddin, Kalfin Alloto'dang, Risnayanti Andi Djamro, and Ahyuna, "Literature Review Artificial Intelligence Deteksi Hasil Ctscan Paru-Paru Pasien Terjangkit COVID-19," *Jurnal Pendidikan Indonesia*, vol. 2, no. 3, pp. 502-516, Mar. 2021, doi: <https://doi.org/10.36418/japendi.v2i3.120>.
- [10] T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases," *Radiology*, vol. 296, no. 2, p. 200642, Feb. 2020, doi: <https://doi.org/10.1148/radiol.2020200642>.
- [11] V. Rajinikanth, N. Dey, A. Noel Joseph Raj, A. Ella Hassanien, K. C. Santosh, and N. Sri Madhava Raja, "Harmony-Search and Otsu based System for Coronavirus Disease (COVID-19) Detection using Lung CT Scan Images," *arXiv preprint arXiv:2004.03431*, Apr. 2020, Available: <https://arxiv.org/abs/2004.03431>
- [12] W. Schmitt and E. Marchiori, "Covid-19: Round and oval areas of ground-glass opacity," *Pulmonology*, Apr. 2020, doi: <https://doi.org/10.1016/j.pulmoe.2020.04.001>.
- [13] W. Swastika, "Studi Awal Deteksi COVID-19 Menggunakan Citra CT Berbasis Deep Learning," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, p. 629, May 2020, doi: <https://doi.org/10.25126/jtiik.2020733399>.
- [14] W. Zhao, W. Jiang, and X. qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific report*, vol. 11, no. 1, pp. 1-12, Jul. 2021, doi: <https://doi.org/10.1038/s41598-021-93832-2>.
- [15] X. Yang, X. he, J. Zhao, Y. Zhang, S. Zhang, and P. xie, "COVID-CT-Dataset: A CT Scan Dataset about COVID-19," *arXiv preprint arXiv:2003.13865*, Jun. 2020, Available: <https://arxiv.org/abs/2003.13865>