

Ekstraksi Fitur dengan Classifier Random Forest untuk Memprediksi Covid 19 Berdasarkan Hasil Rontgen Thorax

Feature Extraction with Random Forest Classifier to Predict Covid 19 Based on Results Thorax X-ray

¹Ali Mustopa, ²Hendri Mahmud Nawawi*, ³Sarifah Agustiani, ⁴Siti Khotimatul Wildah
¹Sistem Informasi Kampus Kota Pontianak, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

²Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

^{3,4}Teknologi Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat, DKI Jakarta

*e-mail: hendri.hiw@nusamandiri.ac.id

(received: 22 Februari 2022, revised: 5 April 2022, accepted: 7 April 2022)

Abstrak

Coronavirus 19 (COVID-19) adalah infeksi yang sangat menular yang disebabkan oleh sindrom pernapasan akut coronavirus 2 (SARS-CoV-2). COVID-19 menjadi virus baru yang belum ditemukan obatnya dengan ditandai dengan meningkatnya angka kematian di seluruh dunia. Penyakit coronavirus yang dapat menyebabkan pneumonia yang menyerang pada kantung udara paru-paru dengan gejala batuk kering, sakit tenggorokan hingga gangguan pernapasan akut (ARDS) yang terjadi pada pasien COVID-19. Salah satu untuk mendeteksi virus tersebut dilakukan dengan mendeteksi hasil rontgen toraks pada pasien. Teknologi machine learning selama satu dekade ini telah berkembang pesat dan terintegrasi ke dalam sistem CAD untuk memberikan akurasi akurat. Penelitian ini dilakukan dengan mendeteksi rontgen toraks dengan menggunakan feature extraction Hu-Moments, Harralick dan Histogram serta mendeteksi akurasi terbaik dengan algoritma klasifikasi untuk mendeteksi hasil COVID-19. Penelitian dilakukan dengan menguji dataset yang didapatkan dari repositori kaggle yang memiliki citra yaitu 1281 Sinar-X COVID-19, Sinar-X Normal 3270, Sinar-X virus pneumonia 1656, dan Sinar-X bakteri-pneumonia 3001. Secara umum penelitian ini termasuk ke dalam kategori baik karena menghasilkan akurasi tertinggi yaitu oleh algoritma klasifikasi Random forest dimana hasil accuracy tersebut sebesar 84% dan Standar Deviasi 0.015847. Selain itu penelitian juga menghasilkan Kappa sebesar 0.713. Hasil akurasi tersebut dilakukan dengan beberapa tahapan yaitu dengan feature extraction berupa hu-moments, Harralick dan histogram. Pada penelitian ini hasil terbaik diberikan oleh algoritma Random forest dengan feature extraction Histogram dan Hu-Moment.

Kata kunci: Coronavirus 19, Toraks, Machine Learning, Random forest, Ekstraksi Fitur

Abstract

Coronavirus 19 (COVID-19) is a highly contagious infection caused by the acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 is a new virus for which no cure has been found, marked by the increasing death rate worldwide. Coronavirus disease which can cause pneumonia which attacks the air sacs of the lungs with symptoms of dry cough, sore throat to acute respiratory distress (ARDS) that occurs in COVID-19 patients. One of the ways to detect the virus is by detecting chest X-rays in the patient. Over the past decade's machine learning technology has developed rapidly and is integrated into CAD systems to provide accurate accuracy. This research was conducted by detecting thoracic radiographs using feature extraction Hu-Moments, Harralick and Histogram and detecting the best accuracy with a classification algorithm to detect the results of COVID-19. The study was conducted by testing the dataset obtained from the Kaggle repository which has images, namely 1281 X-rays of COVID-19, 3270 X-rays Normal, 1656 X-rays of pneumonia, and X-rays of bacteria-pneumonia 3001. In general, this research is included in the Good category because it produces the highest accuracy by the Random forest classification algorithm where the accuracy result is 84% and the standard deviation is 0.015847. In addition, the research also produced Kappa

<http://sistemasi.ftik.unisi.ac.id>

of 0.713. The results of this accuracy are carried out in several stages, namely by feature extraction in the form of hu-moments, Harrallic and histogram. In this study, the best results were given by the Random forest algorithm with feature extraction Histogram and Hu-Moment.

Keywords: Coronavirus 19, Thorax, Mechine Learning, Random forest, Feature Extraction

1 Pendahuluan

Coronavirus 19 (COVID-19) adalah infeksi virus yang sangat menular yang disebabkan oleh sindrom pernapasan akut coronavirus 2 (SARS-CoV-2) [1] yang mulai menyebar luas di Wuhan, Provinsi Hubei, Cina. Sebagian besar pasien secara epidemiologis terkait dengan Pasar Grosir Makanan Laut Huanan [2]. Angka kematian meningkat secara mengkhawatirkan di seluruh dunia yang menuntut tanggapan dini untuk mendiagnosis dan mencegah penyebaran penyakit ini dengan cepat. Karena tidak memiliki obat dan perawatan khusus, situasinya menjadi menakutkan bagi milyaran orang [3].

Coronavirus 19 merupakan penyakit yang dapat menyebabkan peneumonia [4], yaitu infeksi yang mengobarkan kantung udara paru-paru [4]. Gejala mulai dari batuk kering, sakit tenggorokan, dan demam hingga kegagalan organ, syok septik, pneumonia berat, dan Sindrom Gangguan Pernafasan Akut (ARDS) dideteksi dari pasien COVID-19 [5]. Salah satu cara untuk mendeteksi peradangan tersebut adalah dengan menggunakan sinar-X untuk dada. Di sinilah peran *Artificial Intelligence* dan teknik *Mechine learning* akan membantu dokter mendeteksi pneumonia secara akurat dan cepat [6].

Seiring dengan perkembangan ilmu pengetahuan dan teknologi informasi, kehadiran machine learning di bidang komputer telah menarik banyak perhatian. Machine learning menjadi sebuah tren dalam teknologi informasi. Saat ini perkembangan teknologi informasi telah merambah ke berbagai sektor termasuk Kesehatan [7].

Machine learning yang digunakan pada bidang kesehatan bertujuan untuk memprediksi sebuah penyakit. Dengan hasil prediksi yang lebih cepat diterima oleh dokter/pakar kesehatan, semakin cepat pula tindakan atau penanganannya terhadap sebuah penyakit [8]

Selama satu dekade, teknologi *mechine learning* telah berkembang pesat dan diintegrasikan ke dalam sistem CAD untuk memberikan diagnosis yang akurat dan cepat. Keberhasilan luar biasa dari Artificial Intelligence (AI) membawa lebih banyak tanda kemajuan dalam analisis citra medis. Kemampuan model AI yang efektif sangat bergantung pada pembelajaran dari jumlah sampel pelatihan yang memadai [9]. Salah satu yang bisa dilakukan dengan analisis citra medis dalam hal ini dapat dilakukan dengan tambahan *feature extraction* pada citra yang diteliti.

Feature extraction menjawab masalah model template seperti bentuk, ukuran, dan orientasi yang ada pada algoritma template matching dengan cara memetekan ciri-ciri objek citra yang akan dikenali. Ciri-ciri yang dipetakan dari citra yang akan dikenali dan diklasifikasikan terhadap ciri-ciri citra template yang disimpan pada basis data [10]. Salah satu Algoritma *feature extraction* yaitu Hu-Moments, *Harrallic* dan Histogram.

Penelitian lain yang dilakukan dalam mendeteksi coronavirus-19 dengan metode yang diusulkan (CoroNet) adalah jaringan saraf konvolusional yang dirancang untuk mengidentifikasi kasus COVID-19 menggunakan sinar-X pada dada [11]. Model ini telah dilatih dan diuji pada kumpulan data kecil yang terdiri dari 4 kelas yaitu normal, pneumonia bacterial, pneumonia viral dan COVID-19 dengan total keseluruhan citra berjumlah 1.251 citra rontgen dada [11].

Pada paper ini kami memaparkan penelitian lain dengan menggunakan hasil dari rotgen dada dimana penelitian dilakukan dengan menggunakan *feature extraction* Hu-Moments, *Harrallic* dan *Histogram* serta mendeteksi akurasi terbaik dengan algoritma klasifikasi untuk mendeteksi hasil COVID-19. Penelitian dilakukan dengan menguji dataset yang didapatkan dari repositori kaggle yang memiliki citra yaitu 1281 Sinar-X COVID-19, Sinar-X Normal 3270, Sinar-X viruspneumonia 1656, dan Sinar-X bakteri-pneumonia 3001[12].

2 Tinjauan Literatur

Penelitian terkait penentuan deteksi covid-19 dilakukan oleh Muhammad Ghozali dan Heni Suarti [13] yang berjudul Deteksi Tepi pada Citra Rotgen Penyakit Covid-19 Menggunakan metode

Sobel dimana pendeteksian Covid-19 dengan menggunakan metode ini citra yang digunakan menggunakan sinar-X atau proses rontgen melalui proses *clean stage* dengan *high pass filtering* kemudian disegmentasi berdasarkan teknik tresholding. Proses pendeteksian dilakukan untuk melihat bahwa metode sobel dengan teknik tresholding dapat melihat penyebaran virus corona sehingga dapat dijadikan diagnosis Covid-19. Pada penelitian ini belum menyampaikan seberapa pengaruh terhadap persentase keberhasilan penggunaan model untuk mendeteksi covid-19 namun dari hasil penelitian menunjukan bahwa filter *high-pass* menggunakan sobel dapat memberikan informasi tentang daerah yang terinfeksi covid-19.

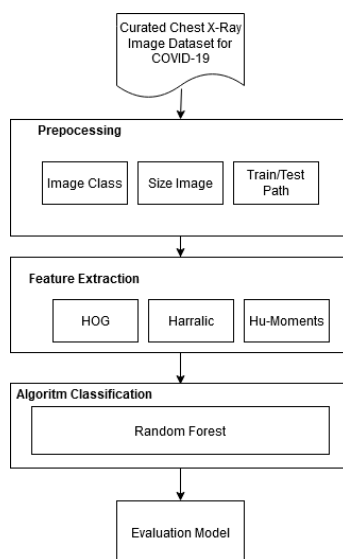
Selanjutnya penelitian yang dilakukan oleh Nisa Nafisah, Riza Ibnu Adam dan Carudin [14] yang berjudul Klasifikasi K-NN dalam identifikasi Penyakit Covid-19 menggunakan Ekstraksi Fitur GLCM penelitian ini menjelaskan proses pendeteksian Covid-19 menggunakan Ekstraksi Fitur GLCM dimana pengklasifikasiannya menggunakan K-NN. Hasil dari deteksi Covid-19 dengan melakukan 3 pemodelan nilai inputan k . Hasil penelitian yang diperoleh dari proses pengklasifikasian yaitu sebesar 80% pada model 3 dengan nilai $k=5$ dan pada model 1 dan 2 diperoleh hasil akurasi sebesar 90% dengan nilai $k=1$ dan $k=3$.

Penelitian terkait ekstraksi fitur juga dapat diterapkan dengan menggunakan jaringan syaraf tiruan dengan model algoritma propagasi balik. Metode ekstraksi fitur yang digunakan menggunakan harralick dimana keakuratan sebesar 66%. Penelitian tersebut dilakukan oleh Sarah Yuli Evangelista Simarmata, Yunita Arum Sari dan Sigit Adinugroho [15]. Sebagai analisis lebih lanjut proses pengklasifikasian ditambahkan berupa ekstraksi fitur GLCM dan Color Histogram pada klasifikasi citra makanan menggunakan Learning Vector Quantization (LVQ) meskipun hasil yang didapatkan masih rendah yaitu 53,33%.

Berdasarkan hasil penelitian terkait, kami mengusulkan beberapa ekstraksi fitur seperti Harralick, Color Histogram dan Hu-moments dimana metode klasifikasi menggunakan algoritma random forest. Hasil eksperimen akurasi dilakukan dengan beberapa tahap mengkomparasikan beberapa ekstraksi fitur sehingga dapat melihat mana yang cocok dalam proses pendeteksian citra rontgen thorax dalam mendeteksi covid-19.

3 Metode Penelitian

Pada metode penelitian ini dilakukan dengan berbagai langkah penelitian yang dapat menghasilkan akurasi algoritma terbaik dalam menentukan hasil *rontgen troax*. Hasil yang diberikan nantinya akan menjadi pilihan untuk menentukan deteksi Covid 19 pada pasien. Adapun langkah penelitian ini adalah sebagaimana Gambar 1 berikut.



Gambar 1. Metode Penelitian

3.1. Dataset Penelitian

Pada tahapan ini dilakukan dengan memilih dataset yang akan kita teliti dengan mengumpulkan, memahami dan mengidentifikasi data yang dimiliki. Kemudian data tersebut juga harus diverifikasi kebenaran dan reliabilitasnya. Dataset yang diperoleh ini berasal dari *kaggle repository*.

3.2. Preprocessing

Tahapan ini dilakukan dengan persiapan data yang sudah didapatkan. Pada tahapan ini meliputi proses Image class, size image dan train/test path untuk menyiapkan data yang akan diuji sesuai dengan label class yang telah ditentukan. Adapun jumlah keseluruhan data yang sudah dikumpulkan berdasarkan class yaitu covid, normal, pneumonia bacterial dan pneumonia viral. Jumlah data yang didapatkan dari keseluruhan data yaitu 9208 citra.

3.3. Feature Extraction

Tahapan *feature extraction* merupakan tahapan yang dilakukan untuk teknik penting dalam *computer vision* yang banyak digunakan untuk warna, tekstur dan bentuk. Penelitian ini dilakukan dengan melakukan dengan *feature extraction* sebagai berikut.

a. *Histogram* (HOG)

Histogram merupakan jenis *feature extraction* yang berfungsi sebagai representasi grafis dari distribusi nada dalam citra digital. Ini memplot jumlah piksel untuk setiap nilai nada. Dengan melihat *Histogram* untuk gambar tertentu, penonton akan dapat menilai seluruh distribusi nada warna secara sekilas.

b. *Haralick*

Haralick digunakan untuk mengkuantifikasi gambar berdasarkan tekstur. Itu ditemukan oleh Haralick pada tahun 1973 dan Anda dapat membacanya secara rinci di sini. Konsep dasar yang terlibat dalam komputasi fitur Tekstur Haralick adalah Matriks Co-kejadian Tingkat Abu-abu atau GLCM.

c. Hu-Moments

Hu-Moments merupakan *feature extraction* yang sangat umum yang diekstrak dari gambar untuk digunakan dalam tugas pengenalan pola, seperti pengenalan bentuk dan pengambilan bentuk.

3.4. Algoritma Klasifikasi

Pada tahapan ini proses pemilihan algoritma dengan parameter nilai yang optimal. Hal ini bertujuan untuk representasi komputasi dari hasil pengamatan yang meruakan hasil dari pencarian pola-pola yang terkandung pada data. Penelitian dilakukan dengan melakukan pengujian beberapa algoritma klasifikasi yaitu algoritma klasifikasi *Random forest Classifier (RF)*.

3.5. Evaluation Model

Pada tahapan ini dilakukan evaluasi model dari klasifikasi algoritma yang dilakukan dengan *feature extraction* yang digunakan untuk melihat hasil dari *accuracy*, *confusion matrix*, *kappa statistic*, *precision* dan *recall*.

4 Hasil dan Pembahasan

Hasil pembahasan merupakan bentuk penjelasan tentang penelitian yang dilakukan dari tahapan awal hingga akhir untuk mendapatkan hasil klasifikasi algoritma terbaik sesuai dengan citra yang dipilih yaitu rotgen torax. Adapun hasil penelitian ini adalah sebagai berikut.

4.1 Dataset

Dataset *Curated Chest X-Ray Image Dataset for COVID-19* adalah kumpulan data yang berasal dari *kaggle repository* yang merupakan gabungan gambar rontgen dada COVID-19 yang diperoleh dengan menyusun 15 kumpulan data yang tersedia untuk umum seperti yang tercantum di bagian referensi. Dataset saat ini berisi 1281 Sinar-X COVID-19, Sinar-X Normal 3270, Sinar-X virus-pneumonia 1656, dan Sinar-X bakteri-pneumonia 3001.

Kumpulan data yang dikumpulkan seperti yang dikutip oleh kumpulan data ini digabungkan untuk membentuk repositori terintegrasi. Repositori terintegrasi ini berisi total 4558 Sinar-X COVID-19, 5403 Sinar-X Normal, 4497 Sinar-X Pneumonia Viral, dan 5768 Sinar-X pneumonia bakteri. Dari 1379 Sinar-X COVID-19, 1476 Sinar-X normal, 2690 Sinar-X pneumonia virus, dan 2588 Sinar-X

<http://sistemasi.ftik.unisi.ac.id>

pneumonia bakterial ditemukan sebagai duplikat berdasarkan kemiripan gambar dan karenanya dihapus. Arsitektur Inception V3 digunakan untuk mendapatkan image embeddings, yang dilanjutkan dengan penggunaan algoritma pembelajaran.





4.2 Preprocessing

Hasil preprocessing dilakukan dengan melakukan perhitungan yang dilakukan dengan tools python dengan memasukan citra terlebih dahulu untuk dilakukan penelitian. Kemudian dari hasil gambar yang dipilih dilakukan dengan menentukan pixel dan data sesuai dengan label yang ingin kita teliti. Penelitian ini menggunakan citra yang berukuran 1024x1024 Pixel sesuai dengan citra yang dihasilkan.

Pada tahapan ini meliputi proses Image class, size image dan train/test path untuk menyiapkan data yang akan diuji sesuai dengan label class yang telah ditentukan. Adapun pembagian data setiap class adalah sebagaimana Tabel 1 berikut.

Nama Class	Jumlah Citra
Covid	1281
Normal	3270
Pneumonia Bacterial	3001
Pneumonia Viral	1656

Jenis citra yang didapatkan dari repositori kaggle dijadikan sebagai bahan eksperimen pada penelitian ini sebagaimana Tabel 2 berikut.

Nama Class	Hasil Citra
Covid	
Normal	
Pneumonia Bacterial	
Pneumonia Viral	

4.3 Feature Extraction

Hasil dari *feature extraction* dilakukan dengan memberikan beberapa uji untuk menampilkan hasil dari eksperimen yang dilakukan. Sebagai bahan eksperimen setiap class wajib di masukkan saat melakukan *feature extraction* agar didapatkan hasil akurasi yang baik sesuai dengan yang diinginkan. Adapun *feature extraction* ini dilakukan adalah sebagai berikut.

1. Hu Moments

Ekstraksi menggunakan Hu-Moments dilakukan dengan mengubah warna RGB ke bentuk grayscale kemudian hasil tersebut dikonversikan kedalam feature *Hu Moments* dengan memberikan hasil array setiap citra untuk dijadikan data yang digunakan sebagai hasil ekstraksi yang diinginkan

2. *Harralic*

Ekstraksi menggunakan *Harralic* merupakan bentuk ekstraksi jenis tekstur dimana gambar RGB diubah kedalam grayscale kemudian dilakukan dengan code feature *Harralic* untuk mengubah gambar dalam bentuk array sebagai hasil ekstraksi yang diinginkan.

3. *Histogram* (HOG)

Ekstraksi menggunakan *Histogram* merupakan bentuk ekstraksi yang dilakukan dengan mengubah RGB kedalam HSV kemudian dilakukan feature *Histogram* dengan menghitung tiap seleksi yang diinginkan.

Pada tahapan tersebut dapat dilakukan dengan melakukan training data dari ketiga *feature extraction* tersebut untuk dijadikan global feature sebagai acuan pada saat melakukan train classifier.

3.4. Algoritma Klasifikasi

Pada proses *cross validation* dilakukan dengan membaca hasil dari *feature extraction* yang dibentuk dengan memasukkan hasil dari *global feature* yang diterapkan pada proses *feature extraction*.

Tahapan proses evaluasi model dilakukan dengan memberikan hasil dari pemrosesan data yang dilakukan dengan *feature extraction* dan *training classifier* sehingga dari hasil olah data tersebut menghasilkan akurasi tiap algoritma sebagai berikut.

Algoritma yang diuji pada penelitian ini diantaranya algoritma *Logistic Regression* (LR), *Linear Discriminant Analysis* (LDA), *Kneighbors Classifier* (KNN), *Decision Tree Classifier* (CART), *Random forest Classifier* (RF), *Gaussian NB* (NB) dan *Support Vector Machine* (SVM).

1. Hasil Pengujian dengan fitur ekstraksi *Hu Moment*, *Harralic* dan *Histogram*.

Penerapan fitur ekstraksi *Hu Moment*, *Harralic* dan *Histogram* terhadap beberapa algoritma klasifikasi menghasilkan nilai akurasi sebagaimana Tabel 3 berikut:

Tabel 3. Hasil Accuracy

Feature Extraction	<i>Hu Moment</i>		<i>Harralic</i>		Histogram	
	Accuracy	Standar Deviasi	Accuracy	Standar Deviasi	Accuracy	Standar Deviasi
RF	0.784449	0.014198	0.74997	0.010789	0.798913	0.017892
CART	0.704999	0.016626	0.699456	0.014916	0.686436	0.013291
KNN	0.611453	0.019418	0.671127	0.012159	0.675107	0.017023
SVM	0.550572	0.01987	0.65413	0.017595	0.669919	0.016377
LR	0.532969	0.019232	0.649305	0.011992	0.628333	0.019498
LDA	0.474622	0.014489	0.618928	0.012358	0.604222	0.019539
NB	0.373716	0.014968	0.572635	0.014248	0.56769	0.011321

Pada Tabel 3 dapat dijelaskan bahwa nilai akurasi yang dihasilkan dari penggunaan *Hu Moment*, *Harralic* dan *Histogram* terhadap beberapa algoritma klasifikasi *random forrest* (RF) menghasilkan nilai akurasi yang lebih baik dibandingkan nilai akurasi pada algoritma lainnya yang dijadikan *sample* pada penelitian ini dengan dengan nilai akurasi 78% untuk fitur *Hu Moment*, 74% pada fitur *Harralic* dan 79% pada fitur *Histogram*.

Kemudian pada pengujian selanjutnya pada masing-masing fitur terhadap algoritma *Random forest* diuji dan menghasilkan nilai sebagaimana Tabel 4 berikut.

Tabel 4. Hasil Pengujian Feature Extraction

	precision			recall			f1-score			support		
	HM	HR	HG	HM	HR	HG	HM	HR	HG	HM	HR	HG
0	0.89	0.74	0.87	0.88	0.65	0.73	0.89	0.69	0.79	381	381	381
1	0.73	0.86	0.82	0.87	0.9	0.93	0.8	0.88	0.87	959	959	959
2	0.7	0.68	0.73	0.71	0.76	0.79	0.71	0.72	0.76	930	930	930

<http://sistemasi.ftik.unisi.ac.id>

3	0.72	0.56	0.66	0.42	0.44	0.45	0.53	0.49	0.54	468	468	468
accuracy							0.74	0.74	0.77	2738	2738	2738
macro avg	0.76	0.71	0.77	0.72	0.69	0.72	0.73	0.7	0.74	2738	2738	2738
weighted avg	0.74	0.73	0.77	0.74	0.74	0.77	0.73	0.73	0.76	2738	2738	2738

HM : Hu Moment, HR : Harrallic, HG : Histogram

2. Hasil Pengujian Hu Moment dan Harrallic, Hu Moment dan Histogram dengan Algoritma Klasifikasi

Pengujian selanjutnya dengan penerapan dengan kombinasi dua fitur ekstraksi Hu Moment dan Harrallic dan Hu Moment dan Histogram menghasilkan nilai akurasi sebagaimana Tabel 5 berikut:

Tabel 5. Hasil Accuracy Kombinasi Feature Hu Moment dengan Harrallic dan Hu Moment dengan Histogram

Feature Extraction	Hu Moment & Harrallic		Hu Moment & Histogram	
Algoritma	Accuracy	Standar Deviasi	Accuracy	Standar Deviasi
RF	0.794214	0.014168	0.845542	0.015847
CART	0.726582	0.014388	0.742123	0.015886
KNN	0.72176	0.017875	0.696634	0.018721
SVM	0.719954	0.012839	0.662363	0.015274
LR	0.707054	0.01233	0.660676	0.017326
LDA	0.684755	0.019575	0.482315	0.018883
NB	0.557205	0.011799	0.355247	0.01837

Berdasarkan Tabel 5 dapat dijelaskan bahwa nilai akurasi yang dihasilkan dari penggunaan Hu Moment dan Harrallic terhadap beberapa algoritma klasifikasi *random forrest* (RF) ini menghasilkan nilai akurasi yang lebih baik daripada algoritma lainnya. pada pengujian dengan dua model *feature extraction* Hu Moment dan Harrallic RF memperoleh akurasi 79% dan pada *feature* Hu Moment dan Histogram pengujian dengan model algoritma random forest memiliki nilai sebesar 84%.

Selanjutnya pada penerapan fitur Hu Moment dengan Harrallic dan Hu Moment dengan Histogram terhadap algoritma *Random forest* ini menghasilkan nilai sebagaimana Tabel 6 berikut.

Tabel 6. Hasil Pengujian Hu Moment dengan Harrallic dan Hu Moment dengan Histogram

	precision		recall		f1-score		support	
	HM & HR	HM & HG	HM & HR	HM & HG	HM & HR	HM & HG	HM & HR	HM & HG
0	0.9	0.93	0.88	0.91	0.89	0.92	381	380
1	0.88	0.85	0.92	0.93	0.9	0.89	959	1006
2	0.72	0.74	0.8	0.82	0.76	0.78	930	878
3	0.61	0.72	0.42	0.45	0.5	0.55	468	471
accuracy		0.81		0.81	0.79	0.81	2738	2735
macro avg	0.78	0.81	0.76	0.78	0.76	0.79	2738	2735
weighted avg	0.78	0.8	0.79	0.81	0.78	0.8	2738	2735

3. Hasil Pengujian Hu Moment, Harrallic, Histogram dengan Algoritma Klasifikasi

Penerapan fitur ekstraksi dengan ketiga *feature extraction* Hu Moment, Harrallic dan Histogram terhadap beberapa algoritma klasifikasi menghasilkan nilai sebagaimana Tabel 7 berikut:

Tabel 7. Hasil Accuracy Hu Moment, Harralic dan Histogram

Algoritma	Accuracy	Standar Deviasi
RF	0.814166	0.016489
LDA	0.738505	0.015158
KNN	0.734279	0.015763
CART	0.722213	0.01668
LR	0.70568	0.018812
NB	0.588509	0.012652
SVM	0.540843	0.010227

Kemudian pada penerapan fitur *Hu Moment*, *Harralic* dan *Histogram* terhadap algoritma *Random forest* ini menghasilkan nilai sebagaimana Tabel 8 berikut.

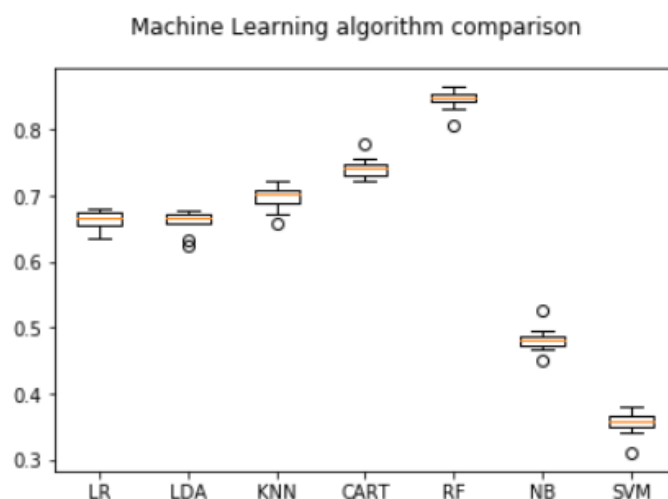
Tabel 8. Hasil Pengujian Hu Moment, Harralic dan Histogram

	precision	recall	f1-score	support
0	0.9	0.86	0.88	380
1	0.86	0.92	0.89	1006
2	0.71	0.78	0.74	878
3	0.59	0.43	0.5	471
micro avg	0.78	0.78	0.78	2735
macro avg	0.77	0.75	0.75	2735
weighted avg	0.77	0.78	0.77	2735

Pada Tabel 7 dapat dijelaskan bahwa nilai akurasi yang dihasilkan dari penggunaan ketiga *feature Extraction Hu Moment, Harralic* dan *Histogram* terhadap beberapa algoritma klasifikasi *random forrest* (RF) menghasilkan nilai akurasi yang lebih baik pada semua algoritma yang memiliki akurasi tertinggi dengan nilai akurasi 81% dibandingkan nilai akurasi algoritma lainnya.

3.5. Evaluation Model

Pada tahapan model evaluasi peneliti juga melakukan berbagai eksperimen untuk melihat akurasi setiap *feature extraction* yang dilakukan ketika menggunakan algoritma klasifikasi *Random forest* untuk membuktikan pengaruh dari *feature extraction* yang diberikan pada algoritma tersebut.



Gambar 2. Machine Learning Algorithm Comparison

Berdasarkan Gambar 2, dari semua percobaan dengan model algoritma klasifikasi model algoritma yang paling baik adalah *Random Forest* (RF) dengan *feature extraction* Hu-Moments +

Histogram dengan accuracy sebesar 84% dan Standard Deviasi sebesar 0.015847. Hasil daripada *confussion matrix* yang dihasilkan dapat dijelaskan pada Tabel 9 berikut.

Tabel 9. Hasil Confussion Matrix

		Prediction Value			
		Covid	Normal	P-Bac	P-Vir
Actual Value	Covid	343	8	25	4
	Normal	11	938	41	16
	P-Bac	13	91	703	71
	P-Vir	5	56	211	199

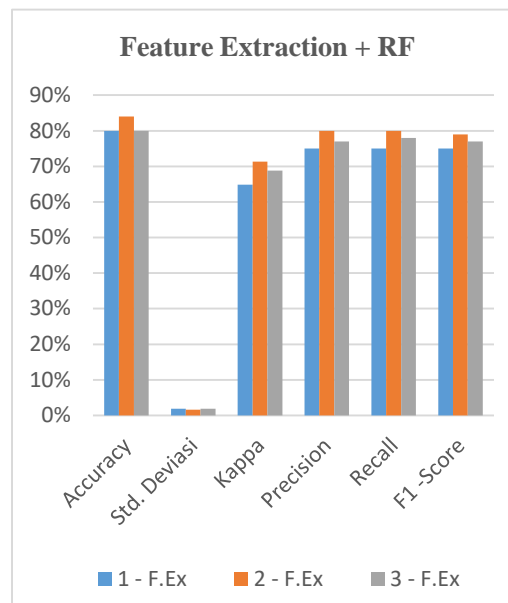
Berdasarkan tabel *confussion matrix*s tersebut menunjukkan bahawa:

- Kolom Warna Hijau menunjukkan bahwa nilai setiap class prediksi terbukti sesuai dengan class yang sebenarnya, artinya sebanyak 343 subject yang diprediksi sebagai Covid terbukti sebagai Covid, 938 subject yang diprediksi sebagai Normal terbukti sebagai Normal dan 703 subject yang diprediksi sebagai Pneumonia Bacterial terbukti sebagai Pneumonia Bacterial, 199 subject diprediksi Pneumonia Viral terbukti sebagai Pneumonia Viral.
- Kolom Warna Kuning menunjukkan salah satu nilai yang diprediksi tidak terbukti sesuai dengan class yang sebenarnya, artinya 41 subject yang diprediksi sebagai Pneumonia Bacterial ternyata hasil sebenarnya adalah Normal

Tabel 10. Jumlah Featrue Extraction dengan *Random forest*

	One Feature Extraction	Two Feature Extraction	Three Feature Extraction
Accuracy	79%	84%	80%
Std. Deviasi	0.01886	0.015847	0.01886
Kappa	0.648	0.713	0.68811
Precision	0.75	0.8	0.77
Recall	0.75	0.8	0.78
F1 -Score	0.75	0.79	0.77

Pada Tabel 10 terlihat bahwa proses *feature extraction* memiliki perbandingan dari tiap ekstrasi yang diberikan sehingga pemilihan *feature extraction* pada citra yang dimiliki memiliki peran penting untuk mendapatkan hasil akurasi terbaik pada proses klasifikasi citra rontgen torax untuk meentukan rontgen torax tersebut menghasilkan deteksi COVID-19 yang lebih baik. Berikut Gambar 3 perbandingan dari tiap *feature extraction* yang dihasilkan dari hasil terbaik tiap feature extraction.



Gambar 3 Hasil Perbandingan Feature Extraction

5 Kesimpulan

Secara umum penelitian ini termasuk kedalam kategori baik karena menghasilkan akurasi tertinggi yaitu oleh algoritma klasifikasi *Random forest* dimana hasil accuracy tersebut sebesar 84% dan Standar Deviasi 0.015847. Selain itu penelitian juga menghasilkan Kappa sebesar 0.713. Hasil akurasi tersebut dilakukan dengan beberapa tahapan yaitu dengan *feature extraction* berupa *h-moments*, *Harrallic* dan *histogram*. Pada penelitian ini hasil terbaik diberikan oleh dua *feature extraction*.

Berdasarkan penelitian yang sudah dilakukan, untuk menentukan hasil rontgen torax covid pada penelitian ini akan lebih baik jika dalam pentuannya dilakukan dengan algoritma *Random forest* dengan *feature extraction Histogram* dan *Hu-Moment* melihat bahwa akurasi yang dihasilkan dari beberapa algoritma klasifikasi serta membandingkan dengan *feature extraction* yang lain memberikan hasil yang lebih unggul dalam proses klasifikasi. Pada tahap selanjutnya penelitian akan dilakukan dengan menggunakan model *deep learning* untuk pengembangan dari *feature extraction* dengan model *machine learning*.

Ucapan Terima Kasih

Terimakasih kepada kedua orang tua tercinta dan kepada semua tim yang terlibat pada penyusunan penelitian ini.

Referensi

- [1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses," *J. Adv. Res.*, vol. 24, pp. 91–98, 2020, doi: 10.1016/j.jare.2020.03.005.
- [2] L. Yang *et al.*, "Epidemiological and clinical features of 200 hospitalized patients with corona virus disease 2019 outside Wuhan, China: A descriptive study.," *J. Clin. Virol.*, vol. 129, no. March, p. 104475, 2020, doi: 10.1016/j.jcv.2020.104475.
- [3] H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," *Journal of Autoimmunity*. 2020, doi: 10.1016/j.jaut.2020.102433.
- [4] A. Torres and C. Cillóniz, *Clinical management of bacterial pneumonia*. 2015.
- [5] C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrob. Agents*, vol. 55, no. 3, p. 105924, 2020, doi:

- 10.1016/j.ijantimicag.2020.105924.
- [6] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset," *arXiv*. 2020.
- [7] R. Munawarah, O. Soesanto, and M. R. Faisal, "Penerapan Metode Support Vector Machine," *Kumpul. J. Ilmu Komput.*, 2016.
- [8] B. Sugara and A. Subekti, "Penerapan Support Vector Machine (Svm) Pada Small Dataset Untuk Deteksi Dini Gangguan Autisme," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 177–182, 2019, doi: 10.33480/pilar.v15i2.649.
- [9] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*. 2018, doi: 10.1038/s41568-018-0016-5.
- [10] R. S. Bahri and I. Maliki, "PERBANDINGAN ALGORITMA TEMPLATE MATCHING DAN FEATURE EXTRACTION PADA OPTICAL CHARACTER RECOGNITION," *J. Komput. dan Inform. Edisi. I Vol. 1, Maret 2012*, 2012.
- [11] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Comput. Methods Programs Biomed.*, 2020, doi: 10.1016/j.cmpb.2020.105581.
- [12] U. Sait, "Curated Chest X-Ray Image Dataset for COVID-19," *Kaggle Repository*, 2020. <https://www.kaggle.com/unaisait/curated-chest-xray-image-dataset-for-covid19>.
- [13] M. Ghozali and H. Sumarti, "Deteksi Tepi pada Citra Rontgen Penyakit COVID-19 Menggunakan Metode Sobel," *J. Imejing Diagnostik*, vol. 6, no. 2, pp. 51–59, 2020, doi: 10.31983/jimed.v6i2.5840.
- [14] N. Nafisah, R. I. Adam, and Carudin, "Klasifikasi K-NN dalam Identifikasi Penyakit COVID-19 Menggunakan Ekstraksi Fitur GLCM," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 128–132, 2021.
- [15] S. Y. E. Simarmata, Y. A. Sari, and S. Adinugroho, "Klasifikasi Citra Makanan Menggunakan Algoritme Learning Vector Quantization Berdasarkan Ekstraksi Fitur Color Histogram dan Gray Level Co-occurrence Matrix," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2369–2378, 2019.