

Implementasi Ekstraksi Fitur GLCM dengan Klasifikasi Algoritma C5.0 Pada Data Computerized Tomography Scan Covid-19

Muhammad Rofiq¹, Triando Hamonangan Saragih², Dodon Turianto Nugrahadi³

^{1,2,3} Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat
¹muhammadrofiq1405@gmail.com, ²triando.saragih@ulm.ac.id, ³dodonturianto@ulm.ac.id,
Corresponding author: Triando.saragih@ulm.ac.id

Abstrak

Teknologi pencitraan digital telah banyak digunakan dalam bidang medis dalam diagnosis data citra biologis untuk memandu dokter untuk mengetahui kondisi pasien. Salah satu teknik pencitraan medis yang dapat menggambarkan kondisi di dalam tubuh manusia adalah *Computed Tomography* (CT). Penelitian ini menggunakan dataset citra CT scan dada berjumlah 625 data CT. Proses indentifikasi data citra CT *scan* ini dapat di identifikasikan berdasarkan warna, tekstur, dan juga ciri lainnya. Beberapa *paper* telah menunjukkan penelitian dengan menggunakan berbagai metode, diantaranya menggunakan ekstraksi fitur. Ekstraksi fitur merupakan proses untuk menemukan pemetaan dari fitur-fitur asli ke dalam fitur- fitur baru yang diharapkan dapat menghasilkan keterpisahan kelas secara lebih baik. Ekstraksi fitur yang digunakan adalah GLCM (*Gray Level Co-Occurrence Matrix*). GLCM adalah suatu metode yang digunakan untuk analisis tekstur/ekstraksi ciri. GLCM merupakan suatu matriks yang menggambarkan frekuensi munculnya pasangan dua piksel dengan intensitas tertentu dalam jarak dan arah tertentu dalam citra. Dalam GLCM Jarak direpresentasikan sebagai piksel sedangkan orientasi direpresentasikan dalam derajat. Orientasi terbentuk dari empat arah sudut dengan interval 0°, 45°, 90°, dan 135°. Sedangkan jarak antar piksel biasanya ditetapkan sebesar 1 piksel. Setelah dilakukan ekstraksi fitur akan dilakukan klasifikasi dengan menggunakan metode algoritma C5.0 Hasil akurasi dari metode klasifikasi C5.0 menggunakan ekstraksi fitur GLCM mendapatkan hasil akurasi sebesar 87% pada sudut 90°, 84% pada sudut 45°, 83% pada sudut 135°, dan 82% pada sudut 0°.

Kata kunci : *Computed Tomography, GLCM, C5.0, Ekstraksi Fitur, Citra, Covid-19, CT-Scan.*

1. Pendahuluan

Tahun 2020 diawali dengan merebaknya suatu wabah virus atau pandemi yang sekarang disebut dengan Covid-19. Para peneliti sangat gencar dan bersaing untuk menangani virus ini. Selain dari pihak medis, semua peneliti atau bahkan mahasiswa-mahasiswa di semua Universitas saling berlomba untuk ikut serta dalam menangani pandemi (Li et al, 2020). Salah satu cara yang dapat dilakukan untuk mengidentifikasi gejala covid-19 adalah dengan menggunakan citra CT (*Computed Tomography*) scan dada, yang mana dari CT scan ini dapat mengidentifikasi bagaimana bentuk atau infeksi yang ditimbulkan oleh Covid-19. Data citra CT untuk pasien Covid tersedia online dalam paper Yudistira, et all (2020).

Proses indentifikasi data citra CT scan ini dapat diidentifikasi berdasarkan warna, tekstur, dan juga ciri lainnya. Beberapa paper telah menunjukkan penelitian dengan menggunakan berbagai metode, diantaranya menggunakan ekstraksi fitur. Ekstraksi fitur merupakan langkah dalam proses klasifikasi. Ekstraksi fitur adalah metode untuk mendapatkan beberapa fitur statistik tentang citra. Tingkat akurasi dalam klasifikasi tergantung

pada ekstraksi fitur. Untuk mendeteksi virus Covid-19, ada banyak fitur yang dapat digunakan untuk mengklasifikasikannya, antara lain ekstraksi ciri morfologi, tekstur orde pertama dan orde kedua GLCM (*Gray Level Co-Occurrence Matrix*). GLCM (*Gray Level Co-Occurrence Matrix*) merupakan metode ekstraksi fitur yang menggunakan perhitungan tekstur pada orde kedua yaitu memperhitungkan pasangan dua piksel citra asli. GLCM mengekstrak beberapa fitur seperti eksentrisitas, metrik, mean, varriance, skewness, kontras, korelasi, energi, dan homogenitas (Auliasari, 2021). Dalam mengidentifikasi deteksi virus Covid-19 diperlukan juga metode klasifikasi untuk mendeteksi virus Covid-19 pada data citra CT scan tersebut. Seperti yang dilakukan oleh Yudistira, et all (2020). menggunakan metode klasifikasi *deep learning* yaitu CNN (*Convolutional Neural Network*) dan membandingkan kemampuan *ShuffleNet*, *EfficientNet*, dan *ResNet50*. Sharmila & Jemi, Florinabel, (2021) menggunakan metode klasifikasi *deep learning* yaitu CNN (*Convolutional Neural Network*) untuk mengidentifikasi Covid-19 pada citra CT scan. Auliasari (2021) menggunakan metode klasifikasi K-Means serta ekstraksi fitur *Morphology*, *First-order Texture*, dan GLCM untuk mengidentifikasi Covid-19 pada citra CT scan.

Romdhoni et al (2020) menggunakan data 12 Citra Kedelai. Lalu menggunakan metode klasifikasi *Decision Tree* serta ekstraksi fitur *Gray-Level CoOccurrence Matrix (GLCM)* untuk mengidentifikasi Citra Kedelai. Dan mendapatkan akurasi 95%. Bakheet & Hamadi (2021) menggunakan data citra X-Ray sebanyak 4,600 Covid 19. Lalu menggunakan metode *Latent-Dynamic Conditional Random Fields (LDCRFs)* serta ekstraksi fitur *Gray-Level Co-Occurrence Matrix (GLCM)* untuk mengidentifikasi Covid-19 pada citra X-Ray mendapatkan akurasi 95.88%. Nur Sakinah et al (2020) dalam penelitiannya menggunakan data citra CT scan otak. Lalu menggunakan klasifikasi *Naive Bayes, Logistic Regression, Neural Network, Support Vector Machine* serta ekstraksi fitur GLCM untuk mengidentifikasi penyakit stroke pada citra CT. Dan mendapatkan akurasi sebesar 96.78%. Wuryani & Agustiani, Sarifah. (2020) dalam penelitiannya menggunakan metode klasifikasi *Random Forest* serta ekstraksi fitur *Haralick, Color Histogram, dan Hu-Moments*. Penelitian dimulai dengan memasukkan satu fitur ke dalam percobaan, lalu mengkombinasikan dengan fitur yang lain, kemudian membandingkannya menggunakan klasifikasi oleh metode klasifikasi lain seperti *K-Nearest Neighbor (KNN), Decision Tree, Linear Discriminant Analysis (LDA), Logistic Regression, Support Vector Machine (SVM), dan Naive Bayes*. untuk mengidentifikasi Covid-19 pada citra CT scan. Moe (2017) dalam penelitiannya menggunakan data Citra satelit yang dihasilkan dari remote sensing. Lalu menggunakan metode klasifikasi C5.0 untuk mengklasifikasi citra satelit lahan gambut sebelum terbakar, terbakar, dan setelah terbakar. Area yang digunakan adalah Kabupaten Ogan Komering Ilir, Provinsi Sumatera Selatan. Hasil dari penelitian ini menunjukkan bahwa algoritma C5.0 memiliki akurasi sebesar 97.10% dan nilai Kappa sebesar 0.96. Robianto et al (2021) dalam penelitiannya dibuat sistem aplikasi website berupa implementasi metode decision tree untuk mengklasifikasikan mutu buah jeruk siam berdasarkan fitur warna dengan model citra HSV dan ukuran menggunakan hasil citra buah. Metode *decision tree* memiliki kemampuan untuk mempersingkat proses pengambilan keputusan yang kompleks menjadi lebih sederhana dengan struktur berbentuk pohon. Berdasarkan hasil pengambilan 25 sampel citra buah jeruk siam menggunakan kamera telepon genggam, selanjutnya citra tersebut diproses oleh aplikasi klasifikasi. Dari hasil pengklasifikasian citra jeruk siam menggunakan metode *decision tree* didapat tingkat keberhasilan sebesar 92%.

Algoritma klasifikasi C5.0 memiliki kemampuan untuk mengklasifikasikan sebagai pohon keputusan atau seperangkat aturan. Algoritma ini cocok dalam menangani nilai, serta lebih sedikit waktu diperlukan untuk mempelajarinya. Algoritma

klasifikasi C5.0 adalah klasifikasi yang mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan klasifikasi lain. Untuk menghasilkan pohon keputusan penggunaan memori minimum dan juga meningkatkan akurasi. Algoritma klasifikasi C5.0 merupakan terusan dari algoritma C4.5, dimana C5.0 memperbaiki kelemahan dari C4.5 dari segi hasil klasifikasi, kesalahan klasifikasi, prediksi akurasi, waktu, tingkat kesalahan, dan pemakaian memori.

Penelitian yang akan dilakukan adalah berfokus dengan menggunakan dataset citra *Computed Tomography* scan berjumlah 625 data. Yang akan dilakukan ekstraksi fitur menggunakan *Gray Level Co-Occurrence Matrix* dengan sudut 0°, 45°, 90°, dan 135°. Dan menggunakan metode klasifikasi C5.0. Untuk mengetahui tingkat akurasinya akan digunakan *Confusion Matrix* pada dataset citra *Computed Tomography* scan Covid-19 (Utomo et al, 2020)

2. Metode

2.1 Alat Penelitian

Alat-alat yang digunakan dalam penelitian ini terbagi menjadi dua jenis meliputi perangkat keras dan perangkat lunak :

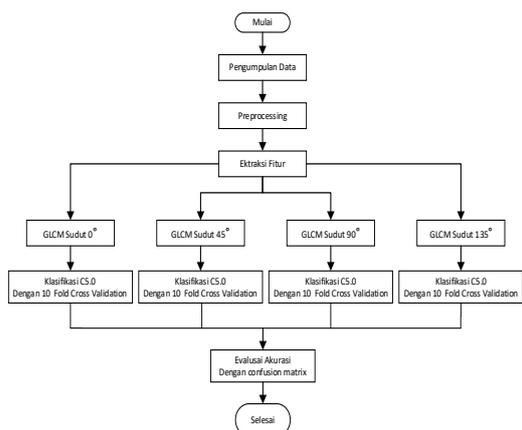
1. Perangkat Keras
 - a. Processor : Intel® Core™ i3-3100U
 - b. RAM : 8.00 GB
 - c. Monitor : 24 Inch
 - d. VGA : Nvidia GTX1050
2. Perangkat Lunak
 - a. Windows 10
 - b. Microsoft Excel
 - c. Microsoft Word
 - d. Microsoft Visio
 - e. R Studio
 - f. Photoshop CC 2013

2.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini adalah Dataset citra CT (Computed Tomography) scan dada yang digunakan oleh Yudistira, et al (2020) terdapat 625 data CT yang terdiri atas 125 citra pasien yang positif Covid-19, 500 citra *No_finding* atau tidak ditemukan gejala Covid-19.

2.3 Prosedur penelitian

Adapun alur penelitian yang dilaksanakan dalam penelitian ini dapat dilihat pada Gambar 1 sebagai berikut:



Gambar 1. Diagram alur prosedur penelitian

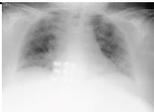
2.4 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah 625 data CT (Computed Tomography) scan dada. Dataset yang digunakan merupakan data CT scan dada yang sudah pernah digunakan dalam penelitian Yudistira, et all (2020). Data set terdiri atas 125 citra pasien yang positif Covid-19 dan 500 citra pasien *No_finding* atau tidak ditemukan gejala Covid-19. Seperti yang dirincikan pada Tabel 1, Tabel 2 dan Tabel 3.

Tabel 1. Rincian Citra CT Scan Dada

Class label	Format Data			Jumlah Data	Sumber data
	JPEG	JPG	PNG		
covid-19	64	39	22	125	(Yudistira, et al, 2020)
No_finding	0	0	500	500	(Yudistira, et al, 2020)

Tabel 2. Rincian Gambar Citra CT Scan Dada Covid-19

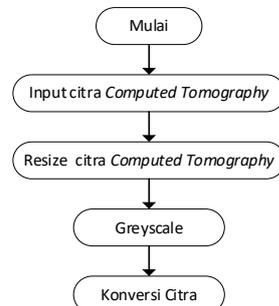
Nama	Gambar	Ukuran	Dimensi
Covid 19 (1).jpeg		208kb	1793 x 1307
Covid 19 (2).jpeg		270kb	1696 x 1403
Covid 19 (3).jpeg		637kb	2337 x 2164

Tabel 3. Rincian Gambar Citra CT Scan Dada No Finding

Nama	Gambar	Ukuran	Dimensi
no finding (1)		347kb	1024 x 1024
no finding (2)		348kb	1024 x 1024
no finding (3)		342kb	1024 x 1024

2.5 Preprocessing Data

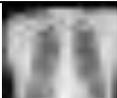
Tahap *preprocessing* dilakukan untuk mempercepat kinerja GLCM dan memperoleh hasil yang lebih akurat pada informasi yang terkandung dalam data citra CT scan dada yang akan digunakan. Tahapan ini juga dilakukan agar nilai data dari citra tersebut menjadi lebih kecil tanpa merubah informasi yang ada didalamnya dan dapat digunakan untuk proses selanjutnya dan dapat dilihat pada Gambar 2 sebagai berikut:



Gambar 2. Tahap Preprocessing Citra

Hasil *preprocessing* citra CT scan dada dapat dilihat pada Tabel 4 sebagai berikut.

Tabel 4. Rincian Gambar Data Citra CT Scan Dada Setelah di Preprocessing

Covid 19		No Finding	
Gambar	Dimensi	Gambar	Dimensi
	18 x 13		20 x 20
	17 x 14		20 x 20
	23 x 22		20 x 20

2.6 Proses GLCM (Gray Level Co-Occurrence Matrix)

Ekstraksi fitur yang dilakukan pada penelitian ini adalah dengan menggunakan ekstraksi fitur GLCM. 625 data CT scan dada yang digunakan nantinya akan dilakukan ekstraksi fitur pada setiap data. Pada pengekstrakan nanti akan digunakan 4 fitur yaitu *Contrast*, *Second Moment*, *Correlation*, dan *Homogeneity*. Di setiap fitur akan digunakan 4 parameter untuk ekstraksi fiturnya yaitu menggunakan sudut 0°, 45°, 90° dan 135°.

Tahap selanjutnya yaitu ekstraksi tekstur menggunakan 4 fitur yang ada pada GLCM, melalui kontras, energi (*Second Moment*), *Correlation*, dan *Homogeneity*.

1. Menghitung nilai Energi (*Second Moment*)

$$E = \sum_{i,j} p(i,j)^2 \quad (1)$$

2. Menghitung nilai Korelasi (*Correlation*)

$$Cor = \frac{\sum_i \sum_j p(i,j) \mu_i \mu_j}{\sigma_i \sigma_j} \quad (2)$$

3. Menghitung nilai Kontras (*Contrast*)

$$CON = \sum_{i,j} i-j^2 p(i,j) \quad (3)$$

4. Homogenitas (*Homogeneity*)

$$Hom = \sum_i \sum_j \frac{p(i-j)}{1+|i-j|} \quad (4)$$

(Pamungkas, 2019).

2.7 Klasifikasi dengan metode C5.0 dan Cross Validation

Hasil ekstraksi fitur GLCM yang sudah terpilih pada 625 data CT scan kemudian dilakukan klasifikasi dengan metode C5.0 dan 10 *Fold Cross Validation*. Algoritma klasifikasi C5.0 memiliki kemampuan untuk mengklasifikasikan sebagai pohon keputusan atau seperangkat aturan. Kekuatan algoritma ini dalam menangani nilai, serta lebih sedikit waktu diperlukan untuk mempelajarinya. Algoritma klasifikasi C5.0 mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan klasifikasi lain. Untuk menghasilkan pohon keputusan penggunaan memori minimum dan juga meningkatkan akurasi. Algoritma C5.0 merupakan terusan dari algoritma C4.5, dimana algoritma C5.0 memperbaiki kelemahan dari C4.5 dari segi hasil klasifikasi, kesalahan klasifikasi, prediksi akurasi, waktu, tingkat kesalahan, dan pemakaian memori (Utomo, et al, 2020).

Cross validation atau dapat disebut estimasi rotasi adalah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen. Teknik ini utamanya digunakan untuk melakukan prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya. Salah satu teknik dari validasi silang

adalah k-fold cross-validation, yang mana memecah data menjadi K bagian set data dengan ukuran yang sama (Azis,.H, 2020).

K-Fold Cross Validation membagi data orisinal menjadi data training dan data testing. *Ten-fold* adalah definisi untuk nilai K, dimana nilai K = 10. Data nantinya akan dibagi menjadi sejumlah dengan nilai K yang ditentukan yaitu 10, maka akan dibagi menjadi 10 set data. Satu set data akan digunakan menjadi data training dan sisanya akan digunakan sebagai data testing dan seterusnya berurutan dan bergantian untuk setiap set nya (Wei et al, 2019).

2.8 Evaluasi dengan Confusion Matrix

Algoritma C5.0 adalah salah satu algoritma data mining yang khususnya diterapkan pada algoritma decision tree. Algoritma C5.0 ini merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut diproses menggunakan gain ratio. Algoritma ini menghasilkan *tree* dengan jumlah cabang per node bervariasi.

Algoritma klasifikasi C5.0 memiliki kemampuan untuk mengklasifikasikan sebagai pohon keputusan atau seperangkat aturan. Kekuatan algoritma ini dalam menangani nilai, serta lebih sedikit waktu diperlukan untuk mempelajarinya. Algoritma klasifikasi C5.0 mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan klasifikasi lain. Untuk menghasilkan pohon keputusan penggunaan memori minimum dan juga meningkatkan akurasi. Algoritma C5.0 merupakan terusan dari algoritma C4.5, dimana algoritma C5.0 memperbaiki kelemahan dari C4.5 dari segi hasil klasifikasi, kesalahan klasifikasi, prediksi akurasi, waktu, tingkat kesalahan, dan pemakaian memori (Utomo, et al, 2020).

2.9 Evaluasi dengan Confusion Matrix

Setelah hasil dari klasifikasi sudah di lakukan kemudian akan di evaluasi hasil klasifikasi metode C5.0 pada data CT dengan menggunakan Confusion Matrix. Confusion matrix dapat diartikan sebagai suatu alat yang memiliki fungsi untuk melakukan analisis apakah classifier tersebut baik dalam mengenali tuple dari kelas yang berbeda. Nilai dari True-Positive dan True-Negative memberikan informasi ketika classifier dalam melakukan klasifikasi data bernilai benar, sedangkan False-Positive dan False-Negative memberikan informasi ketika classifier salah dalam melakukan klasifikasi data (Ghani, et al 2021).

3. Hasil dan Pembahasan

Penelitian ini menggunakan dataset citra CT (*Computed Tomography*) scan dada yang digunakan oleh Yudistira, et al (2020), terdapat 625 data CT yang terdiri atas 125 citra pasien yang positif Covid-19, 500 citra *No_finding* atau tidak ditemukan gejala Covid-19. Proses indentifikasi data citra CT scan ini dapat diidentifikasi berdasarkan warna, tekstur, dan juga ciri lainnya.

Dalam penelitian Yudistira et al (2020), data citra CT yang digunakan sebanyak 1125 gambar untuk mendeteksi 3 kategori yaitu Covid-19, Pneumonia, dan *no_finding* melalui uji *5-fold cross validation*. Lalu menggunakan metode klasifikasi *deep learning* yaitu CNN dan membandingkan kemampuan *ResNet50*, *EfficientNet*, dan *ShuffleNetd*. Dan mendapatkan akurasi 90.8 %, 87.5 %, 86.93 %.

Dalam penelitian ini ekstraksi fitur yang digunakan adalah *Gray Level Co-Occurrence Matrix* (GLCM) yang akan digunakan untuk memberikan nilai untuk masing-masing fitur, dengan sudut 0°, 45°, 90°, dan 135° pada data citra *Computerized Tomography Scan* Covid-19. Dan klasifikasinya menggunakan metode C5.0.

Dalam penelitian Auliasari (2021), pada penelitiannya menggunakan data citra CT sebanyak 3322 gambar. Lalu menggunakan metode klasifikasi *K-Means* serta ekstraksi fitur *Morphology*, *First-order Texture*, dan *GLCM* untuk mengidentifikasi Covid-19 pada citra CT. Dan mendapatkan akurasi 49.23%, 50%, dan 47.69%

Dalam penelitian Romdhoni et al (2020), data penelitiannya menggunakan data 12 Citra Kedelai. Lalu menggunakan metode klasifikasi *Desicion Tree* serta ekstraksi fitur *Gray-Level CoOccurrence Matrix* (*GLCM*) untuk mengidentifikasi Citra Kedelai. Dan mendapatkan akurasi 95%.

Sebelum dilakukan ekstraksi fitur, citra CT (*Computed Tomography*) scan dada akan dilakukan *preprocessing* dilakukan agar kinerja GLCM lebih cepat dan lebih optimal apabila tidak di lakukan *preprocessing* maka kinerja GLCM lebih lama dan ada beberapa hasil Ekstraksi Fitur dari GLCM yang tidak mengeluarkan hasil pada data citra CT scan dada. *Preprocessing* yang dilakukan pada penelitian adalah merubah ukuran (*Resize*) data dari citra tersebut menjadi lebih kecil tanpa merubah informasi yang ada didalamnya.

Tabel 5. Hasil Ekstraksi GLCM sudut 0° pada data citra CT Covid-19

second_moment	Correlation	Contrast	Homogeneity	class
0,0148	0,6969	7,7407	0,2584	covid-19
0,0157	0,1931	0,8519	0,2265	covid-19
0,0133	0,514	29,6667	0,1356	covid-19
0,0136	0,4903	4,2346	0,1471	covid-19

second_moment	Correlation	Contrast	Homogeneity	class
0,0145	0,4378	5,0617	0,1691	covid-19
0,0142	0,4858	3	0,2054	covid-19
0,0151	0,4685	6,2593	0,1914	covid-19
...				
0,013	0,6086	1,1358	0,1375	covid-19
0,0145	0,5532	1,2346	0,1667	covid-19

Tabel 6. Hasil Ekstraksi GLCM sudut 0° pada data citra CT *No_finding*

second_moment	Correlation	Contrast	Homogeneity	class
0,0145	0,3148	10,0617	0,1808	<i>No_finding</i> s
0,0154	0,4526	10,7407	0,1843	<i>No_finding</i> s
0,0139	0,4306	5,6296	0,1786	<i>No_finding</i> s
0,016	0,174	0,4198	0,2909	<i>No_finding</i> s
0,0148	0,2501	9,8272	0,1389	<i>No_finding</i> s
0,0148	0,4741	5,3086	0,2138	<i>No_finding</i> s
0,0172	0,5713	0,5926	0,2087	<i>No_finding</i> s
...				
0,0148	0,307	7,0617	0,2166	<i>No_finding</i> s
0,0145	0,2731	0,8642	0,1926	<i>No_finding</i> s

Tabel 7. Hasil Ekstraksi GLCM sudut 45° pada data citra CT Covid-19

second_moment	Correlation	Contrast	Homogeneity	class
0,0142	0,4781	7,6173	0,1787	covid-19
0,0148	0,114	1,1111	0,1673	covid-19
0,0148	0,4246	41,0988	0,1391	covid-19
0,0145	0,3791	5,4444	0,1766	covid-19
0,0142	0,3227	7,9383	0,1574	covid-19
0,0139	0,2435	3,1852	0,1728	covid-19
0,0136	0,382	7,0617	0,1553	covid-19
...				
0,0136	0,4346	1,9136	0,174	covid-19
0,0139	0,4123	1,5802	0,1656	covid-19

Tabel 8. Hasil Ekstraksi GLCM sudut 45° pada data citra CT *No_finding*

second_moment	Correlation	Contrast	Homogeneity	class
0,0142	0,181	14,0617	0,0936	<i>No_finding</i> s
0,0142	0,2519	14,7654	0,1408	<i>No_finding</i> s
0,0148	0,1258	8,4691	0,1347	<i>No_finding</i> s
0,0166	0,0825	1,1481	0,2089	<i>No_finding</i> s
0,0154	0,158	13,5185	0,1257	<i>No_finding</i> s
0,0154	0,2704	3,7037	0,1945	<i>No_finding</i> s
0,0169	0,4116	1,9012	0,2377	<i>No_finding</i> s
...				
0,0136	0,3019	10,5679	0,1713	<i>No_finding</i> s
0,0136	0,163	3,1975	0,1549	<i>No_finding</i> s

Tabel 9. Hasil Ekstraksi GLCM sudut 90° pada data citra CT Covid-19

second_moment	Correlation	Contrast	Homogeneity	class
0,0157	0,7062	6,6543	0,2637	covid-19
0,0148	0,4643	0,7901	0,255	covid-19
0,013	0,6971	12,4568	0,2035	covid-19
0,0136	0,665	2,679	0,2794	covid-19
0,016	0,5774	3,7531	0,286	covid-19
0,0148	0,5441	2,1235	0,2069	covid-19
0,016	0,549	3,0247	0,2662	covid-19
...				
0,0142	0,7514	0,6049	0,2413	covid-19
0,0145	0,5643	0,5556	0,2506	covid-19

Tabel 10. Hasil Ekstraksi GLCM sudut 90° pada data citra CT *No_finding*

second_moment	Correlation	Contrast	Homogeneity	class
0,016	0,3392	4,9012	0,256	<i>No_finding</i> s
0,0175	0,4363	3,5556	0,3218	<i>No_finding</i> s
0,0148	0,3412	5,1728	0,2461	<i>No_finding</i> s
0,02	0,5918	1,284	0,3399	<i>No_finding</i> s
0,0169	0,4655	4,0617	0,3166	<i>No_finding</i> s
0,0178	0,4696	2,4321	0,3652	<i>No_finding</i> s
0,0178	0,7013	2,5679	0,387	<i>No_finding</i> s
...				
0,0169	0,623	3,7407	0,2581	<i>No_finding</i> s
0,0154	0,6952	3,6173	0,2846	<i>No_finding</i> s

Tabel 11. Hasil Ekstraksi GLCM sudut 135° pada data citra CT Covid-19

second_moment	Correlation	Contrast	Homogeneity	class
0,0172	0,5426	9,3210	0,1987	covid-19
0,0151	0,1173	1,4568	0,1739	covid-19
0,0133	0,4662	26,9259	0,1289	covid-19
0,0130	0,3445	7,6790	0,1019	covid-19
0,0145	0,4161	8,0494	0,1552	covid-19
0,0136	0,3680	3,9753	0,1468	covid-19
0,0145	0,3143	10,8272	0,1475	covid-19
...				
0,0127	0,4681	1,1975	0,1321	covid-19
0,0136	0,4342	1,4938	0,1680	covid-19

Tabel 12. Hasil Ekstraksi GLCM sudut 135° pada data citra CT *No_finding*

second_moment	Correlation	Contrast	Homogeneity	class
0,0139	0,2490	14,6420	0,1511	<i>No_finding</i> s
0,0151	0,2241	15,2099	0,1503	<i>No_finding</i> s
0,0142	0,2020	7,6296	0,1370	<i>No_finding</i> s
0,0136	0,1429	1,8642	0,2372	<i>No_finding</i> s
0,0145	0,2493	13,1358	0,1278	<i>No_finding</i> s
0,0157	0,1950	6,2222	0,2009	<i>No_finding</i> s
0,0157	0,4356	2,7531	0,1749	<i>No_finding</i> s
...				
0,0142	0,2541	11,8765	0,1337	<i>No_finding</i> s
0,0145	0,2065	4,9877	0,1601	<i>No_finding</i> s

Cross-validation digunakan untuk mengevaluasi kinerja algoritma dimana data dipisahkan menjadi dua subset yaitu data proses training dan data testing. Algoritma dilatih oleh subset training dan testing oleh subset validasi. Dalam 10Fold Cross Validation, data dibagi menjadi 10fold berukuran kira-kira sama, sehingga kita memiliki 10 subset data untuk mengevaluasi kinerja model atau algoritma. Untuk masing-masing dari 10 subset data tersebut, Cross Validation akan menggunakan 1 fold untuk data testing dan 9fold untuk data training berulang-ulang hingga ke pengujian 10 seperti diilustrasikan pada Tabel 13.

Tabel 13. Ilustrasi 10Fold Cross Validation

Pengujian Ke	625 Citra CT Scan Dada dibagi menjadi 10 subset data									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10

Pengujian Ke	625 Citra CT Scan Dada dibagi menjadi 10 subset data									
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

Keterangan: Data Testing
 Data Training

Evaluasi hasil akurasi dari algoritma C5.0 pada masing-masing sudut dengan menggunakan *Confusion Matrix* maka didapatkan nilai akurasi. Berikut ini tabel hasil dari *Confusion Matrix* dari masing-masing sudut

Tabel 14. *Confusion Matrix* GLCM Sudut 0° dengan klasifikasi C5.0

<i>Confusion Matrix</i>	<i>Predicted</i>	
	<i>No_finding</i>	<i>Covid-19</i>
<i>Actual No_finding</i>	458	42
<i>Actual Covid-19</i>	72	53

Dari tabel XIV. hasil *Confusion Matrix* diatas menunjukkan bahwa nilai yang terklasifikasi dengan benar yaitu class *No_finding* sebanyak 458 dan class Covid-19 sebanyak 53. Sedangkan untuk nilai yang terklasifikasi dengan salah yaitu class *No_finding* yang terklasifikasikan sebagai Covid-19 sebanyak 72 dan class Covid-19 yang terklasifikasikan sebagai *No_finding* sebanyak 42. Untuk perhitungan akurasi klasifikasi digunakan rumus sebagai berikut.

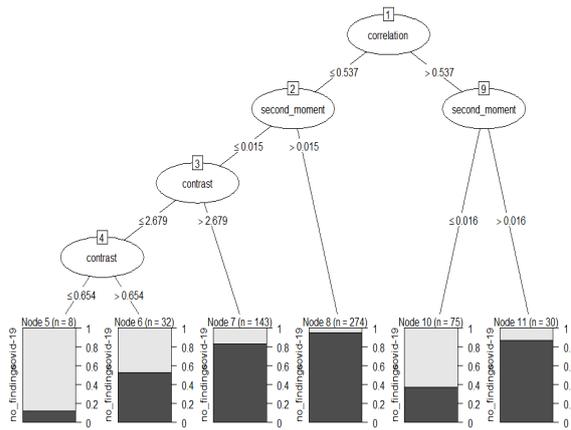
Dari *Confusion Matrix* tersebut, maka dapat dilakukan perhitungan *Accuracy* menggunakan persamaan sebagai berikut:

$$ACC = \frac{TP + TN}{P + N}$$

$$ACC = \frac{458 + 53}{530 + 95}$$

$$= 0,8176 \times 100\% = 81,76\%$$

Dari perhitungan *Accuracy* tersebut, maka hasil akurasi menggunakan algoritma klasifikasi C5.0 dengan ekstraksi fitur GLCM pada sudut 0° adalah 81,76% dan didapatkan *rule model* dan pohon keputusan sebagai berikut.



Gambar 3. Pohon Keputusan GLCM Sudut 0° dengan klasifikasi C5.0

Tabel 15. Confusion Matrix GLCM Sudut 45° dengan klasifikasi C5.0

Confusion Matrix		Predicted	
		No_finding	Covid-19
Actual	No_finding	59	66
	Covid-19	34	466

Dari tabel hasil *Confusion Matrix* diatas menunjukkan bahwa nilai yang terklasifikasi dengan benar yaitu class *No_finding* sebanyak 59 dan class *Covid-19* sebanyak 466. Sedangkan untuk nilai yang terklasifikasi dengan salah yaitu class *No_finding* yang terklasifikasikan sebagai *Covid-19* sebanyak 34 dan class *Covid-19* yang terklasifikasikan sebagai *No_finding* sebanyak 66. Untuk perhitungan akurasi klasifikasi digunakan rumus sebagai berikut.

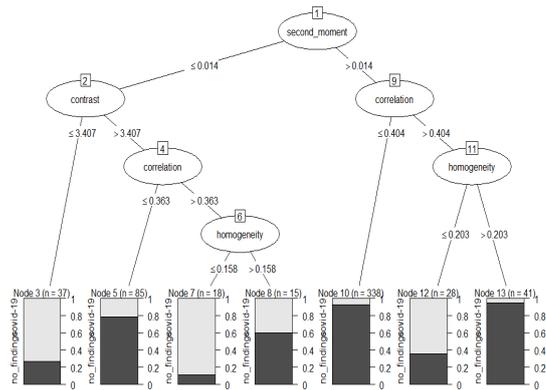
Dari *Confusion Matrix* tersebut, maka dapat dilakukan perhitungan *Accuracy* menggunakan persamaan sebagai berikut:

$$ACC = \frac{TP + TN}{P + N}$$

$$ACC = \frac{59 + 466}{93 + 532}$$

$$= 0,84 \times 100\% = 84\%$$

Dari perhitungan *Accuracy* tersebut, maka hasil akurasi menggunakan algoritma klasifikasi C5.0 dengan ekstraksi fitur GLCM pada sudut 45° adalah 84% dan didapatkan *rule model* dan pohon keputusan sebagai berikut.



Gambar 4. Pohon Keputusan GLCM Sudut 45° dengan klasifikasi C5.0

Tabel 16. Confusion Matrix GLCM Sudut 90° dengan klasifikasi C5.0

Confusion Matrix		Predicted	
		No_finding	Covid-19
Actual	No_finding	466	34
	Covid-19	48	77

Dari tabel hasil *Confusion Matrix* diatas menunjukkan bahwa nilai yang terklasifikasi dengan benar yaitu class *No_finding* sebanyak 466 dan class *Covid-19* sebanyak 77. Sedangkan untuk nilai yang terklasifikasi dengan salah yaitu class *No_finding* yang terklasifikasikan sebagai *Covid-19* sebanyak 48 dan class *Covid-19* yang terklasifikasikan sebagai *No_finding* sebanyak 34. Untuk perhitungan akurasi klasifikasi digunakan rumus sebagai berikut.

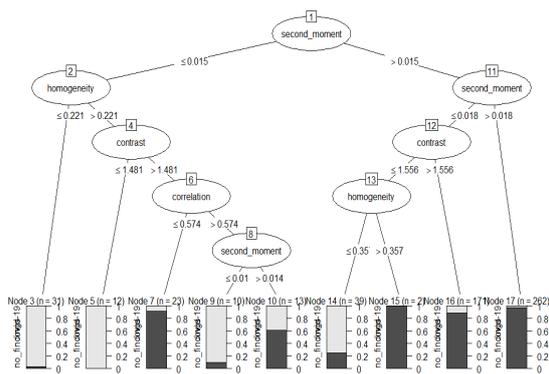
Dari *Confusion Matrix* tersebut, maka dapat dilakukan perhitungan *Accuracy* menggunakan persamaan sebagai berikut:

$$ACC = \frac{TP + TN}{P + N}$$

$$ACC = \frac{466 + 77}{512 + 110}$$

$$= 0,8688 \times 100\% = 86,88\%$$

Dari perhitungan *Accuracy* tersebut, maka hasil akurasi menggunakan C5.0 dengan GLCM pada sudut 90° adalah 86,88% dan didapatkan *rule model* dan pohon keputusan sebagai berikut.



Gambar 5. Pohon Keputusan GLCM Sudut 90° dengan klasifikasi C5.0

Tabel 17. Confusion Matrix GLCM Sudut 135° dengan klasifikasi C5.0

Confusion Matrix		Predicted	
		No_finding	Covid-19
Actual	No_finding	469	31
	Covid-19	76	49

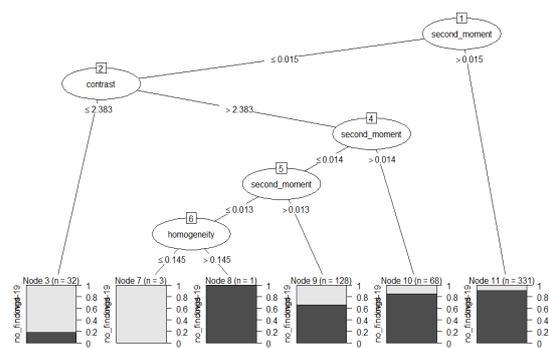
Dari tabel hasil *Confusion Matrix* diatas menunjukkan bahwa nilai yang terklasifikasi dengan benar yaitu class *No_finding* sebanyak 469 dan class *Covid-19* sebanyak 49. Sedangkan untuk nilai yang terklasifikasi dengan salah yaitu class *No_finding* yang terklasifikasikan sebagai *Covid-19* sebanyak 76 dan class *Covid-19* yang terklasifikasikan sebagai *No_finding* sebanyak 31. Untuk perhitungan akurasi klasifikasi digunakan rumus sebagai berikut.

Dari *Confusion Matrix* tersebut, maka dapat dilakukan perhitungan *Accuracy* menggunakan persamaan sebagai berikut:

$$ACC = \frac{TP + TN}{P + N}$$

$$ACC = \frac{469 + 49}{545 + 80} = 0,8288 \times 100\% = 82.88\%$$

Dari perhitungan *Accuracy* tersebut, maka hasil akurasi menggunakan C5.0 dengan GLCM pada sudut 135° adalah 82.88% dan didapatkan *rule model* dan pohon keputusan sebagai berikut.



Gambar 6. Pohon Keputusan GLCM Sudut 135° dengan klasifikasi C5.0

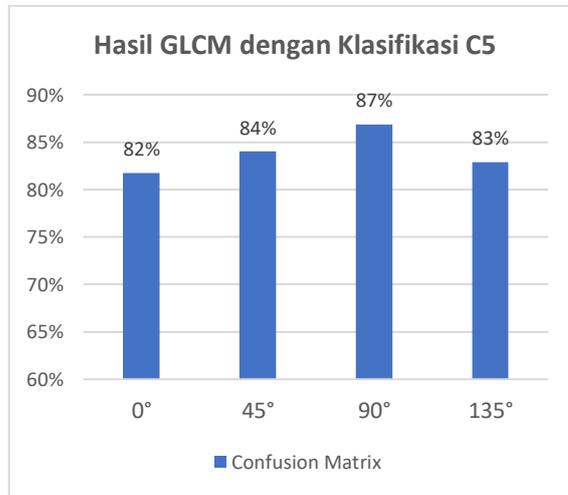
Penelitian ini menggunakan *Confusion matrix* untuk mendapat nilai akurasi dari metode klasifikasi C5.0. Hasil akurasi menggunakan metode klasifikasi C5.0 pada data citra *Computerized Tomography Scan Covid-19* dengan ekstraksi fitur GLCM. Hasil akurasi dapat dilihat pada Tabel 18.

Tabel 18. Hasil Akurasi GLCM semua Sudut dengan klasifikasi C5.0

No.	Hasil Klasifikasi C5.0	
	Sudut GLCM	Confusion Matrix
1	0°	81,76%
2	45°	84%
3	90°	86,88%
4	135°	82,88%

Hasil akurasi dari metode klasifikasi C5.0 pada data citra *Computerized Tomography Scan Covid-19* dengan ekstraksi fitur GLCM menggunakan sudut 0° mendapatkan nilai akurasi 81%, sudut 45° mendapatkan nilai akurasi 84%, sudut 90° mendapatkan nilai akurasi 86%, dan sudut 135° mendapatkan nilai akurasi 82%.

Sedangkan secara grafik *Confusion Matrix* GLCM semua Sudut dengan klasifikasi C5.0 dapat dilihat pada Gambar 7.



Gambar 7. Grafik Hasil *Confusion Matrix* GLCM semua Sudut dengan klasifikasi C5.0

Dari Hasil Pengujian diatas telah dipaparkan hasil akurasi dari metode klasifikasi C5.0 menggunakan ekstraksi fitur GLCM mendapatkan hasil akurasi sebesar 87% pada sudut 90°, 84% pada sudut 45°, 83% pada sudut 135°, dan 82% pada sudut 0°.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa performa yang dihasilkan dari metode klasifikasi C5.0 pada data citra

Computerized Tomography Scan Covid-19 dengan ekstraksi fitur GLCM mendapatkan hasil akurasi tertinggi sebesar 87% pada sudut 90°. Pada sudut 45° mendapatkan nilai akurasi 84%. Pada sudut 135° mendapatkan nilai akurasi 83%. Dan terakhir pada sudut 0° mendapatkan nilai akurasi 82%

DAFTAR PUSTAKA:

- Auliasari. (2021), *Feature Extraction And K-Means Clustering Approach To Classify The Covid-19 Lung Ct-Scan Image*, IJISCS (International Journal of Information System an Computer Science)
- Azis, H., Purnawansyah, P., Fattah, F. & Putri, I.P. (2020), *Performa Klasifikasi K-NN dan Cross Validation pada Data Pasien Pengidap Penyakit Jantung*, ILKOM Jurnal Ilmiah, 12(2), pp.81-86.
- Ghani, M. A., Fahrizal, F., & Lawi, A. (2021), *Implementasi Arsitektur Xception Untuk Klasifikasi Citra Covid-19 Radiography*, 413–419.
- Li, Q, Guan X, Wu P, Wang, X. & Feng, Z. (2020), *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia*, New England Journal of Medicine.
- Meo, Meliana O. (2017), *Penerapan Algoritme Pohon Keputusan C5.0 Untuk Klasifikasi Lahan Gambut Yang Terbakar Di Kabupaten Ogan Komering Ilir*. Jurnal Teknologi Terpadu. Vol. 3, No. 1, Juli 2017.
- Nurchayani, Arissa Aprilia & Saptono, Ristu. (2020), *Identifikasi Kualitas Beras dengan Citra Digital*, Scientific Journal of Informatics. Vol. 2, No.1, Mei 2020.
- Pamungkas, Danar Putra. (2019), *Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek (Orchidaceae)*, Innovation in Research of Informatics (INNOVATICS).
- Romdhoni, Nur Fitriyati, Usman, Koredianto & Hidayat, Bambang. (2020), *Deteksi Kualitas Kacang Kedelai Melalui Pengolahan Citra Digital dengan Metode Gray-Level Co-Occurrence Matrix (GlcM) dan Klasifikasi Decision Tree*, Prosiding Seminar Nasional Riset Dan Information Science (SENARIS) 2020 Vol. 2, (2020), pp. 132-137
- Robianto, Sitorus, Sampe Hotlan & Ristian, Uray. (2021), *Penerapan Metode Decision Tree Untuk Mengklasifikasikan Mutu Buah Jeruk Berdasarkan Fitur Warna Dan Ukuran*, Jurnal Komputer dan Aplikasi Volume 09, No. 01 (2021), hal 76-86.
- Sharmila & Jemi, Florinabel. (2021), *Deep learning Algorithm for COVID-19 Classification Using Chest X-Ray Images*, Computational and Mathematical Methods in Medicine, Volume 2021, Article ID 9269173, 10 pages
- Utomo, Putro., Pahala, Sirait & Roni, Yunis. (2020), *Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0*, Jurnal Media Informatika Budidarma. Vol. 4, No. 4, Oktober 2020, Page 994-1006
- Yudistira, Novanto., Widodo, Agus Wahyu & Rahayudi, Bayu. (2020), *Deteksi Covid-19 Pada Citra Sinar-X Dada Menggunakan Deep learning Yang Efisien*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK). Vol. 7, No. 6, Desember 2020, hlm. 1289-1296.
- Wei, hu, Xue Y, & Zhang Q. (2019), *Establishing a software defect prediction model via effective dimension reduction*, Information Sciences.
- Wuryani, Nanik. & Agustiani, Sarifah. (2020), *Random Forest Classifier untuk Deteksi Penderita COVID-19 Berbasis Citra CT Scan*, Jurnal Teknik Komputer AMIK BSI Volume 7, No.2, Juli 2021

