

**TESIS**

**IDENTIFIKASI DAN KLASIFIKASI COVID-19 PADA CITRA *CHEST*  
CT-SCAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS GLCM  
DENGAN METODE K-NN DAN NAÏVE BAYES**

**REZKY RACHMADANY RACHMAN  
H032211001**



**PROGRAM STUDI MAGISTER FISIKA  
DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2022**

**IDENTIFIKASI DAN KLASIFIKASI SARS-COV-2 PADA CITRA *CHEST*  
CT-SCAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS GLCM  
DENGAN METODE K-NN DAN NAÏVE BAYES**

**TESIS**

*Diajukan sebagai Salah Satu Syarat  
Memperoleh Gelar Magister Sains  
pada Program Studi Magister Fisika Departemen Fisika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin*

**REZKY RACHMADANY RACHMAN**

**H032211001**

**PROGRAM STUDI MAGISTER FISIKA  
DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2022**

**LEMBAR PENGESAHAN TESIS**

**IDENTIFIKASI DAN KLASIFIKASI COVID-19 PADA CITRA CHEST  
CT-SCAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS GLCM  
DENGAN METODE K-NN DAN NAÏVE BAYES**

Disusun dan diajukan oleh

**REZKY RACHMADANY RACHMAN  
H032211001**

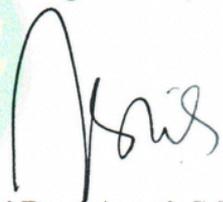
Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka penyelesaian studi Program Magister Program Studi Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada Tanggal 21 September 2022 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui

Pembimbing Utama,

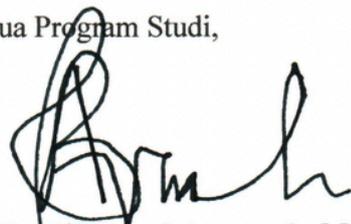
Pembimbing Pendamping,

  
**Prof. Dr. Syamsir Dewang, M.Eng.Sc.**  
NIP. 19670520 199403 1 002

  
**Dr. Sri Dewi Astuti, S.Si., M.Si.**  
NIP. 19750513 199903 2 001

Ketua Program Studi,

Dekan Fakultas,

  
**Dr. Ir. Bidayatul Armynah, M.T.**  
NIP. 19630830 198903 2 001

  
  
**Dr. Eng. Amiruddin, M.Si.**  
NIP. 19720515 199702 1 002

## PERNYATAAN KEASLIAN

Nama : Rezky Rachmadany Rachman  
NIM : H032211001  
Program Studi : Fisika  
Jenjang : S2

Menyatakan dengan ini bahwa karya tulisan saya berjudul

**IDENTIFIKASI DAN KLASIFIKASI COVID-19 PADA CITRA *CHEST*  
CT-SCAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS GLCM  
DENGAN METODE K-NN DAN NAÏVE BAYES**

Adalah karya tulisan saya sendiri, bukan merupakan pengambilan alihan tulisan orang lain dan bahwa tesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi tesis ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 21 September 2022

Yang menyatakan



Rezky Rachmadany Rachman

## ABSTRAK

*Corona virus disease 2019 (Covid-19)* adalah penyakit menular akibat infeksi virus SARS-CoV-2 dan telah menyebar menjadi pandemi global. Virus ini meandering melalui organ vital manusia, yaitu paru-paru. Oleh karena itu, penelitian ini mengidentifikasi covid-19 dan non-covid-19 berdasarkan citra *chest CT-Scan* menggunakan klasifikasi K-NN dan Naïve Bayes. Sistem dibangun melalui *pre-processing*, segmentasi, ekstraksi fitur berbasis GLCM, dan membagi data uji dan latih menggunakan *k-fold cross validation* dengan nilai 5, 7, dan 10, kemudian dievaluasi dengan *confusion matrix*. Nilai akurasi dari pengklasifikasian diperoleh sebesar 99,6% untuk K-NN dan Naïve Bayes mencapai nilai 93,5%, sedangkan tingkat sensitivitas tertinggi diperoleh dengan metode K-NN bernilai 100% dan nilai spesifisitas 98,4% untuk kedua metode yang digunakan. Pada pengujian ini, klasifikasi dengan metode K-NN lebih tepat dibandingkan Naïve Bayes karena beberapa fitur dari GLCM lebih akomodatif terhadap pengklasifikasi K-NN.

Kata Kunci: Covid-19, GLCM, K-NN, Naïve Bayes

## ABSTRACT

Corona virus disease 2019 (Covid-19) is an infectious disease caused by infection with the SARS-CoV-2 virus and has spread to become a global pandemic. This virus attacks through human vital organs, namely the lungs. Therefore, this study identifies covid-19 and non-covid-19 based on chest *CT-Scan* images using the K-NN and Naïve Bayes classification. The system is constructed through pre-processing, segmentation, GLCM-based feature extraction, and dividing the testing and training data with K-fold cross-validation with the value of 5, 7, and 10, then evaluated using Confusion Matrix. The accuracy value of classification is 99.6% for K-NN and Naïve Bayes is 93.5%. In comparison, the K-NN method obtained the highest sensitivity level with a value of 100% and a specificity value of 98.4% for the two methods used. In this test, classification using the K-NN method is more appropriate than Naïve Bayes because some features of GLCM are more accommodating to the K-NN classifier.

Keywords: Covid-19, GLCM, K-NN, Naïve Bayes

## KATA PENGANTAR



Assalamualaikum Warahmatullahi Wabarakatuh. Rasa syukur penulis panjatkan kehadiran ALLAH SWT atas segala limpahan rahmat dan karunia-Nya, Tuhan Semesta Alam, Maha Adil, dan Maha Bijaksana. Selawat serta salam juga penulis haturkan kepada Nabi Muhammad SAW, yang menjadi *role model*, idola, dan kebanggaan bagi hidup penulis, nabi akhir zaman yang membawa misi kedamaian, penyebar syari'at Islam kepada seluruh umat manusia di dunia. Begitu pula salam sejahtera selalu tercurahkan untuk keluarganya, para sahabat, dan ummatnya yang mengikuti ajaran dan petunjuk-Nya hingga akhir zaman.

Alhamdulillah, atas hidayah dan inayah-Nya yang begitu besar dan sempurna sehingga penulis dapat menyelesaikan penyusunan tesis dengan judul “Identifikasi dan Klasifikasi Covid-19 pada Citra *Chest CT-Scan* Menggunakan Ekstraksi Fitur Berbasis GLCM dengan Metode K-NN dan Naïve Bayes” yang merupakan syarat dalam menyelesaikan studi pada Program Studi Magister Fisika. Dalam menyelesaikan tesis ini, penulis banyak mendapat bimbingan, petunjuk, dan bantuan dari berbagai pihak yang sangat berarti dan berharga bagi penulis.

Teruntuk yang teristimewa, orang paling berharga dalam hidup, Almh. Hj. Nur'aeni Umar dan H. Abd. Rachman, penulis persembahkan sebuah tulisan dari didikan kalian, yang teraplikasi dengan ketikan hingga menjadi barisan kata dengan beribu kesatuan, berjuta makna kehidupan, tidak bermaksud yang lain hanya ucapan TERIMA KASIH yang setulusnya tersirat di hati atas doa, usaha, dan jerih payah pengorbanan untuk penulis, terima kasih banyak telah menjadi orang tua yang sempurna. Terkasih, kakak kandung panutan penulis, Rachmayani Rachman, S.Pd., dan Resminayani Rachman, S.Si.T., Gr., CWCCA., terima kasih atas semua kebaikanmu dan menjadi sosok yang ada untuk adikmu. Terima kasih kepada keponakan tersayang Khafid Dzaky Baisa, Ince Saqinah Ayu, Adeeva Qanita Aisa, dan Khalid Sidqi Baisa yang selalu memberikan hiburan serta menjadi teman berkelahi. Tak lupa, Syafreza Fachlevi, S.S., yang siap siaga membantu penulis dalam kondisi apa pun. Terima kasih pula pada Ibu Syahriah yang ikhlas mengurus penulis selama berada di rumah.

Penulis juga ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Eko Juarlin, S.Si., M.Si., selaku Dosen Pembimbing Teknis yang telah begitu banyak meluangkan waktu, tenaga, serta pemikirannya untuk membantu penulis menyelesaikan penelitian ini.
2. Bapak Prof. Dr. Syamsir Dewang M.Eng.Sc selaku Dosen Pembimbing Utama dan Ibu Dr. Sri Dewi Astuty, M.Si selaku Pembimbing Pendamping yang telah banyak membimbing penulis sehingga tesis ini dapat terselesaikan.
3. Ibu Dr. Ir. Bidayatul Armynah, M.T. selaku Ketua Program Studi Magister Fisika sekaligus salah satu anggota Tim Penguji bersama Bapak Prof. Dr. Bualkar Abdullah, M.Eng.Sc., dan Bapak Prof. Dr. Tasrief Surungan, M.Sc., yang telah banyak meluangkan waktu dan tenaga untuk memberikan ilmu dan saran dalam penyelesaian tesis ini.
4. Bapak Prof. Dr. Arifin, M.T selaku Ketua Departemen Fisika serta Bapak dan Ibu Dosen Pengajar Departemen Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin, terima kasih untuk dedikasinya yang sedemikian rupa bagi kampus dan dunia pendidikan. Semoga semangat pengabdianya akan terus menyala hingga ujung usia.
5. Seluruh staf yang ada di Instalasi Radiologi Rumah Sakit Universitas Hadanuddin telah membantu penulis dalam pengambilan data penelitian.
6. Bapak/Ibu Staf Pegawai FMIPA UNHAS, terutama Staf Departemen Fisika; Pak Syukur, Ka Evi, dan Ka Rana yang selalu membantu penulis selama berada di kampus.
7. Kak Maya sekeluarga (Ina, Ka Cimman, Khalif, Mutia), terima kasih sudah mengizinkan numpang di Makassar selama beberapa bulan ini dengan murah hati dan penuh kepedulian.
8. Grup Karibku: Bonte3000, Lambelkon, Asec2ah, serta teman-teman yang tak bisa kusebutkan satu-persatu, sanjyuu atas motivasinya.
9. Lembagaku, Kerukunan Pelajar Mahasiswa Tanah Bumbu Kalimantan Selatan Makassar (KPMTBKSM); Himpunan Mahasiswa Fisika (Himafi) FMIPA Unhas; Unit Tennis Meja Universitas Hasanuddin (UTMUH).

Terimakasih, wahai manusia hebat di dalam sana (kakak, letting, adik), telah menerimaku menjadi keluarga untuk bersama-sama melewati masa-masa penuh duka dan suka dan mempercayakan Sekretaris Umum di tanganku, dan akhirnya menjadi Kanda alumni dan senior wkk.

10. Kakak-Kakak S2 Fisika 2021-1, 2020-2, 2020-1, 2019-2, 2019-1 terima kasih untuk kenangan belajar dan pengurusan berkas yang telah dilalui bersama, terima kasih atas segala semangat, cerita, dan hiburan yang kita lewati selama ini, kalian akan menjadi hal teristimewa dalam perjalanan hidupku.
11. Semua pihak yang tidak dapat penulis sebutkan satu persatu, yang telah berkontribusi, berdoa, perhatian, kepo, *stalking*, dan menyayangi secara implisit maupun eksplisit. Untuk semuanya, terima kasih telah menyisihkan sebagian dari kehidupan berharga kalian buat seonggok daging sepertiku.
12. Akhir dari yang terakhir, teruntuk diri ini. Terima kasih sudah bangkit walaupun rasanya sulit. berani mencoba lagi ketika kegagalan menghampiri, Terima kasih untuk terus bertahan sekuat tenaga. tidak pernah menyerah walau sering kali merasa lelah, meredakan ego dan menikmati segala proses, sabar dari segala hal yang mengejar. Terima kasih telah berpikir positif di kala keadaan sempat tak bepihak dan menyadari bahwa dalam hidup, tidak semua hal berjalan sesuai kehendak,. Terima kasih sekali lagi karena tetap menjadi diri sendiri.

## DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
I PENDAHULUAN .....	1
I.1 Latar Belakang .....	1
I.2 Rumusan Masalah .....	4
I.3 Tujuan Penelitian .....	4
I.4 Manfaat Penelitian .....	4
I.5 Batasan Masalah .....	5
I.6 Sistematika Penulisan .....	5
II TINJAUAN PUSTAKA .....	6
II.1 Aspek Fisika Pencitraan <i>Computed Tomography (CT)-Scan</i> .....	6
II.2 Karakteristik Citra <i>Coronavirus Disease (Covid-19)</i> pada <i>CT-Scan</i> .....	10
II.3 <i>Software Image Processing</i> dan <i>DICOM Viewer</i> .....	12
II.4 Pemrosesan Digital .....	14
II.5 Ekstraksi Fitur <i>Gray Level Co-occurrence Matrix (GLCM)</i> .....	16
II.6 Metode Klasifikasi .....	19
II.7 <i>K-Fold Cross Validation</i> .....	22
III METODE PENELITIAN .....	23
III.1 Jenis Penelitian .....	23
III.2 Waktu dan Tempat Penelitian .....	23

III.3 Dataset .....	23
III.4 <i>Pre-Processing</i> .....	24
III.5 Segmentasi Citra .....	25
III.6 Proses Ekstraksi Fitur .....	26
III.7 Klasifikasi .....	27
III.8 Metode Evaluasi .....	28
III.9 Alur Penelitian .....	30
<b>IV HASIL DAN PEMBAHASAN</b>	
IV.1 Ekstraksi Fitur .....	33
IV.2 Hasil Klasifikasi dan Evaluasi Performa .....	42
IV.3 Perbandingan Performa Metode Ekstraksi .....	45
<b>V KESIMPULAN DAN SARAN</b>	
V.1 Kesimpulan .....	50
V.2 Saran .....	50
<b>DAFTAR PUSTAKA .....</b>	<b>51</b>

## DAFTAR TABEL

<b>Tabel 3.1</b> <i>Confusion Matrix</i> .....	<b>29</b>
<b>Tabel 4.1</b> Hasil Ekstraksi GLCM 0° .....	<b>36</b>
<b>Tabel 4.2</b> Hasil Ekstraksi GLCM 45° .....	<b>37</b>
<b>Tabel 4.3</b> Hasil Ekstraksi GLCM 90° .....	<b>38</b>
<b>Tabel 4.4</b> Hasil Ekstraksi GLCM 135° .....	<b>39</b>
<b>Tabel 4.5</b> Performa dari Klasifikasi Fitur GLCM .....	<b>44</b>
<b>Tabel 4.6</b> Perbandingan Performa Metode Ekstraksi .....	<b>45</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b> <i>Image, picture, dan photo</i> .....	6
<b>Gambar 2.2</b> <i>CT-Scan Machine</i> .....	7
<b>Gambar 2.3</b> Hasil Citra <i>CT-Scan</i> .....	8
<b>Gambar 2.4</b> Skema dasar akuisisi data pada <i>CT-Scan</i> .....	9
<b>Gambar 2.5</b> Arah dan vektor perpindahan GLCM .....	17
<b>Gambar 2.6</b> Pembentukan matriks GLCM .....	17
<b>Gambar 2.7</b> Pembagian Data dengan <i>K-fold cross validation</i> .....	22
<b>Gambar 3.1</b> (a) Citra <i>chest CT-Scan</i> non covid dan (b) covid .....	24
<b>Gambar 3.2</b> (a) Citra asli dan (b) citra setelah di- <i>pre-processing</i> .....	24
<b>Gambar 3.3</b> <i>Active contour</i> sebagai titik terkendali .....	25
<b>Gambar 3.4</b> (a) Hasil segmentasi <i>Active contour</i> non covid dan (b) covid (b) .....	26
<b>Gambar 3.5</b> Bagan Alir Penelitian .....	30
<b>Gambar 4.1</b> (a) Sampel Citra non-Covid-19 (b) Sampel Citra Covid-19 .....	31
<b>Gambar 4.2</b> Grafik Perbedaan Hasil Fitur GLCM 0° pada Covid dan non-Covid .....	36
<b>Gambar 4.3</b> Grafik Perbedaan Hasil Fitur GLCM 45° pada Covid dan non-Covid .....	37
<b>Gambar 4.4</b> Grafik Perbedaan Hasil Fitur GLCM 90° pada Covid dan non-Covid .....	38
<b>Gambar 4.5</b> Grafik Perbedaan Hasil Fitur GLCM 135° pada Covid dan non-Covid .....	39
<b>Gambar 4.6</b> Persentase Perbandingan Rata Ciri pada Covid-19 dan non-Covid-19 .....	41

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Hasil Ekstraksi Fitur GLCM .....	<b>55</b>
<b>Lampiran 2</b> Algoritma Ekstraksi GLCM dan Klasifikasi .....	<b>87</b>
<b>Lampiran 3</b> Perhitungan .....	<b>110</b>

## BAB I

### PENDAHULUAN

#### I.1 Latar Belakang

Penyakit yang baru-baru ini menjadi perbincangan masyarakat di seluruh dunia ditemukan pertama kali pada kelompok masyarakat Wuhan, Provinsi Hubei, China yang dilaporkan ke *World Health Organization* (WHO) pada 31 Desember 2019. Wabah ini menginfeksi saluran pernapasan yang diakibatkan oleh virus dengan istilah *Severe Acute Respiratory Syndrome Corona Virus 2* (SARS-CoV-2) [1,2,3]. Pada 12 Februari 2020, WHO resmi menyatakan penyakit ini dengan sebutan *Corona Virus Disease 2019* (Covid-19) [4].

Data statistik dari *Center for Systems Science and Engineering* (CSSE) di *Johns Hopkins University* (JHU) menunjukkan total kasus terkonfirmasi di seluruh dunia sebanyak 388.048.849 dan 5.712.849 kasus meninggal, sementara di Indonesia ditetapkan 4.353.370 kasus dengan spesimen diperiksa, sedangkan kasus meninggal ialah 144.320. Kasus baru terus terjadi sebab Covid-19 dapat ditularkan dari manusia ke manusia melalui tetesan berukuran mikron dari hidung dan/atau mulut. Penularan dapat terjadi melalui bersin, batuk, dan berbicara [5] Gejala umum yang dialami penderita adalah suhu tubuh di atas 38° C, batuk, dan kesulitan bernapas. Virus ini mempengaruhi saluran pernafasan dan menimbulkan lapisan luka pada *thorax* yang mempengaruhi fungsi dari *thorax* itu sendiri. Tingkat keparahan yang tinggi dapat menyebabkan pneumonia, sindrom pernafasan akut, bahkan kematian [6].

WHO menganjurkan *swab test* dengan *Polymerase Chain Reaction* (PCR) ketika covid mulai menyebar, yang bertujuan untuk *screening* pasien dalam mengetahui adanya materi genetik virus covid-19 [7]. Hasil deteksi keberadaan virus yang diperoleh dari reaksi senyawa kimia perlu divalidasi untuk mengklasifikasikan akut pasien yang terinfeksi berdasarkan status suspek, probabel, ataupun positif Covid-19. Status tersebut didasarkan pada tingkat keparahan gejala dan lamanya pasien telah terinfeksi, maka diperlukan metode tambahan untuk penegakan diagnosa.

Per Juni 2020, WHO menerbitkan panduan penggunaan teknik pencitraan dada dalam penanganan Covid-19. Panduan itu memeriksa bukti dan membuat rekomendasi untuk penggunaan pencitraan dada. Teknik pencitraan itu mencakup radiografi, CT-Scan, dan ultrasonografi (USG) [8]. Para pakar di Amerika Serikat pun menemukan pemeriksaan CT-Scan sebagai metode terbaik untuk mendeteksi pasien Covid-19 pada tahap awal. Mereka menyatakan hasil CT-Scan sangat akurat karena organ dalam bisa terlihat dalam format tiga dimensi [9]. Pencitraan dada dengan CT-Scan juga memainkan peran penting dalam evaluasi keparahan dan tindak lanjut perjalanan penyakit, serta menunjukkan perbedaan yang signifikan antara gambar paru-paru sehat dan penderita Covid-19 [10].

Salah satu indikasi kerusakan *thorax* akibat Covid-19 adalah munculnya objek *Ground-Glass Opacity* (GGO) [11]. Gambar radiografi menunjukkan bahwa objek GGO lebih terang dari objek jaringan dada normal karena tidak mengandung udara lagi [12]. Beberapa penelitian yang dipublikasikan dalam database PubMed, Embase (Elsevier), Google Scholar, dan *World Health Organization* (WHO) memuat 88% penderita yang mengalami GGO. Sebanyak 62 pasien pneumonia akibat Covid-19 di Wuhan, 40,3% *thorax*-nya mengandung GGO. Dari 101 kasus penderita Covid-19 berusia 21-50 tahun di Hunan Cina, 86,1% *thorax*-nya mengandung GGO. Berdasarkan penelitian tersebut, terdapat perubahan fisiologis sebagai pengamatan organ *thorax* pada Covid-19 [13, 14, 15].

Ahli radiologi dalam mengamati citra *thorax* pada penderita Covid-19 masih bersifat subjektif, sehingga hasil analisisnya masih bias karena dipengaruhi oleh pengalaman pengamat itu sendiri [16]. Dengan demikian, pemahaman GGO masih banyak diminati dan menantang karena tingginya variasi tekstur, ukuran, dan akurasi posisi yang disajikan dalam citra CT-Scan. Batas GGO masih sulit dideteksi karena kontrasnya yang rendah dan tampilannya yang kabur. Untuk itu, diperlukan model identifikasi otomatis saat menggunakan pencitraan CT-Scan penderita Covid-19 yang akan mengurangi keterlibatan manual [17].

Proses deteksi citra CT-Scan ini dapat diidentifikasi berdasarkan warna, tekstur, dan juga ciri lainnya. Beberapa *paper* telah menunjukkan penelitian dengan menggunakan berbagai metode, beberapa diantaranya adalah menerapkan *machine learning* untuk *pre-processing*, segmentasi, ekstraksi ciri fitur, dan klasifikasi

Covid-19 pada citra CT-Scan [18]. Seperti yang dilakukan oleh Tri Deviasari Wulan dan rekannya [19], menggunakan 60 citra *chest CT-scan* dengan fitur GLCM dan menggunakan klasifikasi *Probabilistic Neural Network* (PNN) untuk membedakan *nodule* paru-paru dan bukan nodul paru-paru. Dari proses penelitian ini didapatkan hasil dari sampel yang diuji, bahwa akurasinya sebesar 85%.

Penelitian Nilanjan Dey, dkk [20], menggunakan *SGO-K thresholding*, *K-means clustering*, dan *morphological segmentation* untuk proses sebelum data diekstraksi serta dengan beberapa fitur, yaitu *Hu Moments*, CWT, DWT, EWT, dan *haralick* fitur (GLCM) untuk mencari informasi data Covid-19. Klasifikasi yang digunakan dalam penelitian, yakni *Random-Forest* (RF), *Support Vector Machine-radial basis Function* (SVM-RBF), *K-Nearest Neighbors* (K-NN), dan *Decision Tree* (DT). Performa yang dihasilkan dalam penelitian ini sebesar 87%.

Wahyudi, dkk [21] melakukan deteksi Covid-19 berdasarkan citra X-ray menggunakan metode GLCM dan K-NN. Dari hasil pengujian melalui 180 citra, sistem mampu melakukan deteksi covid-19 melalui citra X-ray dengan tingkat nilai akurasi mencapai 85,6%. Thepade, dkk [22] menggunakan ekstraksi fitur GLCM dan mengkombinasikannya dengan *wavelet transform*, serta menggunakan K-NN dan SVM sebagai *classifier* untuk mengidentifikasi Covid-19 pada citra X-ray paru. Penelitian ini menghasilkan nilai akurasi paling tinggi pada saat menggunakan K-NN *Classifier*, yaitu sebesar 92,6%. Selanjutnya, oleh Bakheet san Al-Hamadi [23] menggunakan ekstraksi fitur GLCM dan klasifier LDCRF, menghasilkan akurasi rata-rata sebesar 95,88%. Sedangkan pada *paper* Ameer dan Mohammed [24] mendeteksi Covid-19 menggunakan CT-Scan berdasarkan GLCM, dan mendapatkan akurasi 94%.

Penelitian yang telah dilakukan sebelumnya menjelaskan bahwa proses *chest CT-scan* memerlukan proses tambahan, yaitu pemrosesan citra digital agar informasi terkait deteksi Covid-19 terhadap paru-paru yang terpapar lebih mudah dideteksi dan pastinya pemilihan ekstraksi fitur dan ciri yang tepat akan meningkatkan performa dari deteksi Covid-19. Berdasarkan paparan yang telah dijelaskan di atas, maka peneliti melakukan pengklasifikasian data citra chest CT-Scan untuk mengidentifikasi penyakit Covid-19 dengan menggunakan ekstraksi

fitur *Gray Level Co-occurrence Matrix* (GLCM) dengan metode klasifikasi K-NN dan Naïve Bayes. Penelitian ini dilakukan dengan harapan mampu mengklasifikasi Covid-19 berdasarkan data *chest CT-Scan*, sehingga hasil dari klasifikasi dapat bermanfaat dan membantu tenaga medis untuk mengklasifikasi antara pasien yang terinfeksi Covid-19 dan non-Covid-19.

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil di antaranya adalah sebagai berikut:

1. Bagaimana hasil ciri dari data *chest CT-Scan* menggunakan ekstraksi fitur berbasis *Gray Level Co-occurrence Matrix* (GLCM)?
2. Bagaimana hasil klasifikasi Covid-19 berdasarkan citra *chest CT-Scan* dengan metode *K-Nearest Neighbors* (K-NN) dan Naïve Bayes untuk dua kategori, yaitu non-Covid-19 dan Covid-19?

## **I.3 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini:

1. Memperoleh ciri/fitur dari citra *chest CT-scan* dengan menggunakan ekstraksi *Gray Level Co-occurrence Matrix* (GLCM).
2. Menganalisis dan mengevaluasi hasil klasifikasi dua kategori, yakni non-Covid-19 dan Covid-19 dari metode *K-Nearest Neighbors* (K-NN) dan Naïve Bayes.

## **I.4 Manfaat Penelitian**

Manfaat yang dapat diambil dari latar belakang yang telah dipaparkan di atas adalah:

1. Manfaat Teoritis

Manfaat dari penelitian ini adalah menambah pengetahuan klasifikasi Covid-19 melalui citra *chest CT-Scan* menggunakan ekstraksi fitur berbasis GLCM dengan metode K-NN dan Naïve Bayes.

2. Manfaat Praktis

Penelitian ini diharapkan dapat memudahkan dokter, epidemiolog, serta virology untuk melakukan deteksi Covid-19 yang sedang mewabah di seluruh dunia dengan cepat.

### **I.5 Batasan Masalah**

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian merupakan data sekunder diperoleh dari Instalasi Radiologi Rumah Sakit Umum Pendidikan Universitas Hasanuddin.
2. Deteksi Covid-19 berdasarkan citra *chest CT-Scan* dibagi ke dalam dua kelas, yaitu normal dan Covid-19.
3. Hasil dari klasifikasi ini adalah berupa informasi hasil klasifikasi terinfeksi Covid-19.

### **I.6 Sistematika Penulisan**

Isian dari penulisan tesis ini dapat dijelaskan sebagai berikut:

#### **BAB I: PENDAHULUAN**

Pada bab I ini, menjelaskan tentang pendahuluan dari tesis ini yang berisikan tentang latar belakang dari penelitian dan juga terdapat penelitian terdahulu.

#### **BAB II: TINJAUAN PUSTAKA**

Bab II berisikan tentang dasar teori yang terkait dengan penelitian ini, yakni Aspek Fisika Pencitraan *Computed Tomography (CT) Scan*, Pengenalan dan Klasifikasi Citra, Proses Pembentukan Citra pada CT-Scan, Karakteristik Citra *Coronavirus Disease (Covid-19)* pada CT-Scan, Software *Image Processing* dan *DICOM Viewer*, Pemrosesan Citra Digital, Ekstraksi Fitur GLCM, serta Metode Klasifikasi K-NN dan Naïve Bayes.

#### **BAB III: METODE PENELITIAN**

Pada bab III dijelaskan tentang alur dari penelitian ini mulai dari *pre-processing* sampai dengan tahap klasifikasi yang kemudian dievaluasi.

#### **BAB IV: HASIL DAN PEMBAHASAN**

Pada bab IV menjelaskan hasil dari penelitian ini, yaitu berisikan hasil ekstraksi fitur dengan GLCM, hasil klasifikasi K-NN dan Naïve Bayes, serta evaluasi performa *classifier*.

#### **BAB V: KESIMPULAN DAN SARAN**

Bab V berisikan kesimpulan dari semua hasil yang ada pada bab IV dan saran yang diharapkan dapat dilakukan pada penelitian selanjutnya.

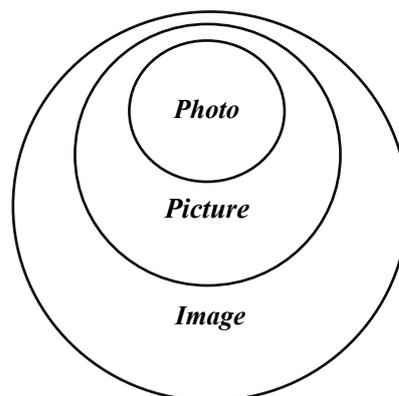
## BAB II

### TINJAUAN PUSTAKA

#### II.1 Aspek Fisika Pencitraan *Computed Tomography (CT)-Scan*

##### II.1.1 Pengenalan dan Klasifikasi Citra

Pencitraan didefinisikan sebagai fungsi dua dimensi dari pasangan koordinat yang menyatakan posisi dari titik citra dan nilai intensitas atau *gray level* dari citra di titik tersebut. Jika nilai intensitas atau derajat keabuan citra dilakukan proses kuantisasi dan pada pasangan koordinat dilakukan proses sampling, maka hasil dari kedua proses tersebut akan menghasilkan citra digital. Citra digital terdiri dari elemen-elemen gambar atau *pixel (picture elemen)* yang mempunyai lokasi dan intensitas tertentu. Elemen itu akan diproses dalam sistem pengolahan citra [25].



**Gambar 2.1** *Image, picture, dan photo* [26].

**Gambar 2.1** menjelaskan perbedaan antara *image* (citra), *picture* (gambar), dan *photo* (foto). *Image* merupakan representasi yang lebih umum daripada foto dan gambar, sedangkan *picture* didefinisikan sebagai representasi dari orang atau suatu objek sebagai hasil melukis, dan foto adalah representasi dari orang atau suatu objek sebagai hasil pengkapturan kamera [26]. Berdasarkan aplikasinya, citra dapat dikelompokkan menjadi dua bagian, yaitu [27, 28]:

1. Citra secara umum adalah citra hasil dari *capture* kamera digital pada umumnya. Format citranya berbentuk BMP, JPEG, GIF, PNG, atau TIFF.

2. Citra medis adalah citra yang dihasilkan oleh peralatan medis seperti *X-ray*, *CT-Scan*, dll. Format khusus untuk citra medis, yaitu DICOM.

Sebuah citra akan memuat informasi format citra dan juga informasi citra itu sendiri yang ditampilkan oleh seluruh *pixel* yang ada dalam citra tersebut. Pada citra medis (DICOM), selain terdapat informasi tentang citranya itu sendiri, terkandung pula informasi tentang pasiennya seperti nama, tempat/tanggal lahir (TTL), jenis kelamin, tanggal diagnosa, tempat diagnosa, dan lain-lain [26].

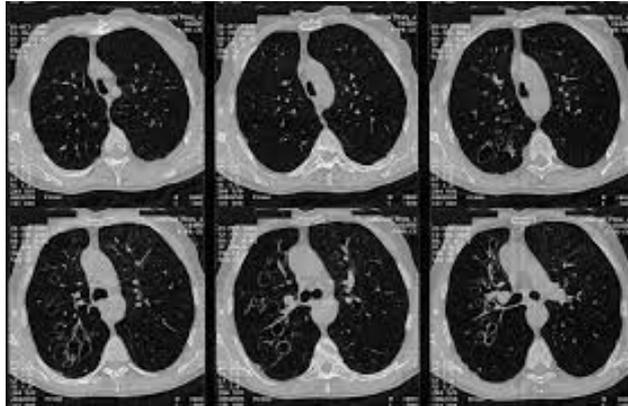
### II.1.2 Proses Pembentukan Citra pada *CT-Scan*

*Computerized tomography* (CT) adalah bentuk ideal dari tomografi yang menghasilkan gambar berurutan dari irisan tipis pasien secara berurutan dan memberikan kesempatan untuk melokalisasi dalam tiga dimensi [29]. Pada **Gambar 2.2** menunjukkan mesin *CT-Scan* berbentuk pipa dengan tempat pasien berbaring di tengahnya. Pemroses citra (*scanner*) terdapat dalam *frame* pipa tersebut. Saat mesin bekerja, pipa pemroses citra berputar sambil menembakkan sinar rontgen ke arah pasien dari berbagai sudut. Untuk setiap putaran, sekitar 1.000 gambar bagian dalam pasien diambil [30].



**Gambar 2.2** *CT-Scan Machine* [29].

*Computerized Tomography* (CT) *Scan* berbasiskan *X-ray* yang dipancarkan ke arah tubuh pasien dengan merotasikannya pada sumbu *x*, *y*, dan *z* sehingga dihasilkan citra arah lintang (*cross-section*) yang berupa potongan (*slice*) citra seperti yang ditunjukkan dalam **Gambar 2.3**. Dalam satu kali pengambilan citra dapat dihasilkan banyak potongan citra tergantung dari panjang atau tinggi area yang akan didiagnosa dan juga tinggi tiap *slice* citra [31].



**Gambar 2.3** Hasil Citra CT-Scan [30].

Pembentukan gambar oleh CT-Scan didasarkan pada prinsip fisika meliputi akuisisi data, pengolahan data, rekonstruksi citra, representasi citra, penyimpanan, dan dokumentasi [32, 34]. Setelah melalui tahap tersebut hasil gambar dapat dianalisis ulang.

#### 1. Akuisisi Data

Akuisisi data berarti kumpulan hasil penghitungan transmisi *X-ray* setelah melalui tubuh pasien. Selama *scanning* tabung *X-ray* dan detektor berputar mengelilingi pasien untuk mendapatkan data atenuasi pasien. Detektor menangkap radiasi yang diteruskan melalui pasien dari beberapa lokasi dan sudut. Sebagai hasilnya, nilai transmisi relatif atau pengukuran atenuasi dapat dihitung pada persamaan 2.1 [30, 32]:

$$I_t = I_0 e^{-\int_0^L \mu(x) dx} \quad (2.1)$$

di mana  $I_0$  dan  $I_t$  masing-masing adalah intensitas sinar datang dan ditransmisikan;  $L$  adalah panjang jalur *X-ray*; dan  $\mu(x)$  merupakan koefisien atenuasi linier *X-ray*, yang bervariasi menurut jenis jaringan dan karenanya merupakan fungsi dari jarak  $x$  melalui pasien.

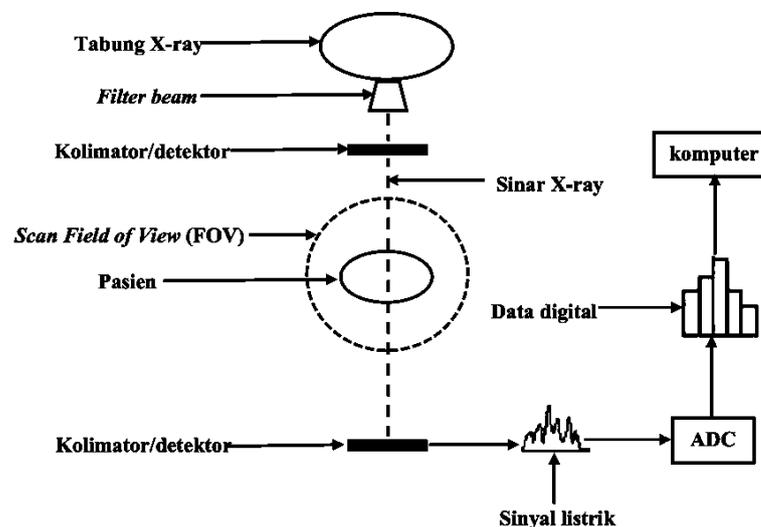
Persamaan di atas menggambarkan intensitas *X-ray* yang datang akan diserap oleh jaringan berdasarkan kepadatan atom, sehingga intensitas *X-ray* yang ditransmisikan akan tervisualisasikan melalui citra CT-Scan. Jaringan yang memiliki koefisien atenuasi linier besar mengindikasikan banyaknya intensitas radiasi *X-ray* yang datang akan dilemahkan sebelum diteruskan.

Oleh karena itu, integral dari koefisien atenuasi diberikan oleh [30]:

$$\int_0^L \mu(x) dx = -\frac{1}{L} \ln \left( \frac{I_t}{I_0} \right) \quad (2.2)$$

Algoritma rekonstruksi memerlukan pengukuran integral ini di sepanjang banyak jalur dalam berkas kipas di setiap sudut di sekitar *isocenter*. Nilai  $L$  diketahui, dan  $I_0$  ditentukan oleh kalibrasi sistem. Oleh karena itu, nilai integral sepanjang setiap jalur dapat ditentukan dari pengukuran  $I_t$  [30].

Skema dasar akuisisi data dapat dilihat pada **Gambar 2.4**, X-ray dikeluarkan dari tabung X-ray dalam bentuk berkas yang menembus irisan melintang tubuh objek yang diamati terletak di antara *field of view* (FOV). Struktur anatomi objek yang diamati akan mengatenuasi X-ray yang melewati objek, sebelum akhirnya ditangkap oleh serangkaian detektor. Setiap elemen detektor ini mengatur intensitas X-ray yang teratenuasi yang melewati suatu objek. Intensitas sinyal elektrik dari detektor kemudian diubah oleh *analog-digital conversion* (ADC), menjadi suatu data masukan untuk komputer [33].



**Gambar 2.4** Skema dasar akuisisi data pada CT-Scan [31].

Menurut Bushberg [34], metode akuisisi data CT-Scan ada dua, yaitu:

- Metode konvensional *slice by slice* atau metode aksial. Prinsipnya, tabung X-ray dan detektor bergerak mengelilingi pasien dan mengumpulkan data proyeksi pasien. Saat pengambilan data proyeksi, posisi meja berhenti,

kemudian meja pasien bergerak untuk menuju posisi kedua dan dilakukan proses *scanning* berikutnya, begitu seterusnya.

- b. Metode *spiral* atau *helical*. pada metode ini tabung *X-ray* bergerak mengelilingi pasien yang juga bergerak. Pada metode ini, berkas *X-ray* membentuk pola *spiral* atau *helical*. Data untuk rekonstruksi citra pada setiap *slice* diperoleh dengan interpolasi. Teknik ini memiliki kelebihan dalam waktu yang relatif cepat.

## 2. Pengolahan Data

Suatu sinar sempit (*narrow beam*) yang dihasilkan oleh *X-ray* didapatkan dari perubahan posisi dari tabung *X-ray*, hal ini juga dipengaruhi oleh collimator dan detector. Sinar *X-ray* yang telah dideteksi oleh detektor kemudian dikonversi menjadi arus listrik yang kemudian ditransmisikan ke komputer [33].

## 3. Rekonstruksi Citra, Representasi Citra, dan Penyimpanan

Setelah detektor mendapatkan perhitungan transmisi yang cukup, data dikirim ke komputer untuk proses selanjutnya. Proses rekonstruksi citra pada *CT-Scan* adalah proses melibatkan beberapa jenis algoritma yang mengolah data nilai koefisien attenuasi dari jaringan tubuh pasien (objek). Gambar *CT-Scan*, yaitu gambar yang berasal dari *cathode ray tube* (CRT) yang dibuat oleh ribuan piksel (*picture element*) kecil. Setiap gambar yang direkonstruksi oleh komputer akan memberi nilai bilangan CT spesifik untuk setiap piksel dari CRT, operator dapat merubah sesuai keinginan untuk merubah skala yang ditampilkan. Bilangan CT yang ditampilkan oleh CRT mendeteksi gambar yang sebenarnya [35].

Setelah komputer melakukan proses rekonstruksi gambar, hasil gambar tersebut bisa ditampilkan dan disimpan untuk nantinya dianalisis ulang. Gambar CT dapat disimpan dalam pita magnetik dan cakram magnetik [33].

## II.2 Karakteristik Citra *Coronavirus Disease 2019* (Covid-19) pada *CT-Scan*

Penelitian oleh Xia, dkk [36], dilakukan di Rumah Sakit Anak, Wuhan, terhadap pasien laki-laki berusia 1-10 tahun dan perempuan berusia 8-14 tahun. Semuanya menunjukkan hasil *CT-scan* toraks berupa GGO. Pada pasien perempuan berusia 8 tahun didapatkan gambaran GGO di lobus inferior paru kiri

dan pasien perempuan berusia 14 tahun didapatkan GGO yang tersebar di lobus inferior paru kanan, terletak subpleural atau meluas dari lesi subpleural.

Wang, dkk [37] melakukan penelitian pada pasien laki-laki berusia 22-78 tahun dan perempuan berusia 21 tahun. Semuanya memperlihatkan adanya gambaran GGO dengan beberapa variasi, berupa: GGO soliter; GGO di bagian bawah bilateral lobus paru; GGO dengan dilatasi vaskular; GGO dengan bronkodilatasi vaskular; GGO yang tidak merata; GGO dengan vasodilatasi dan konsolidasi; serta GGO dengan *halo sign* di lobus kanan bawah.

Pada penelitian oleh Cui, dkk [38] terhadap pasien laki-laki berusia 28-67 tahun dan perempuan berusia 45-69 tahun didapatkan juga gambaran GGO yang bervariasi, berupa: GGO tidak merata (setelah 3 hari masuk rumah sakit/MRS) berlanjut peningkatan area GGO subpleural secara bilateral disertai dengan konsolidasi dan bayangan seperti kabel serta penebalan septum interlobular (setelah 6 hari MRS); GGO di sepanjang berkas vaskular di daerah subpleural segmen lingula lobus kiri atas (setelah 2 hari MRS) berlanjut menjadi GGO yang meluas disertai konsolidasi, kemudian peningkatan konsolidasi yang tidak merata pada area lesi (setelah 9 hari /MRS), dan CT toraks setelah 15 hari MRS menunjukkan kepadatan GGO menurun; GGO tersebar di kedua paru; beberapa GGO subpleural bilateral dan setelah 7 hari MRS gambaran *CT-scan* toraks menunjukkan GGO difus bilateral dan subpleural yang melibatkan beberapa lobus paru; GGO yang tidak merata dan perubahan seperti jaringan secara bilateral di area subpleural dari pinggiran paru, *CT-scan* toraks diambil setelah 3 hari MRS menunjukkan peningkatan GGO yang tidak merata secara bilateral di pinggiran paru disertai dengan konsolidasi, GGO dan bayangan seperti jaringan di dalam paru disertai penebalan septum interlobular.

Zhu, dkk [39] melakukan penelitian pada pasien laki-laki berusia 41-67 tahun dan perempuan berusia 43-57 tahun didapatkan gambaran GGO yang bervariasi, berupa: GGO dengan konsolidasi di lobus kanan bawah, disertai *air bronchogram*; GGO dengan konsolidasi *streak* dan garis subpleural di paru kiri bawah; GGO yang terdistribusi di sekitar pinggiran dengan pola retikuler atau gambaran sarang lebah; GGO dengan konsolidasi beruntun yang terdistribusi di daerah perifer dan garis

subpleural di lobus kanan atas; GGO perifer berbentuk bulat dengan sedikit konsolidasi di lobus kanan bawah; serta beberapa GGO yang tersebar di area perifer, GGO dengan konsolidasi di lobus bawah, GGO dengan gambaran vakuola di segmen atas paru kiri, dan GGO dengan penebalan septum interlobular di lobus inferior kanan.

Penelitian oleh Gu, dkk [40] terhadap pasien laki-laki berusia 46-70 tahun menunjukkan gambaran GGO pada lobus kanan disertai nodul; distribusi campuran bilateral dari GGO dengan kelainan yang terjadi secara bersamaan meliputi pola sarang lebah dan penebalan interlobular.

Penelitian oleh Li, dkk [41] dilakukan pada pasien laki-laki berusia 1-4 tahun. Pada pasien laki-laki usia 1 tahun 5 bulan, *CT-scan* toraks setelah 4 hari menunjukkan adanya GGO yang tidak merata pada lobus kanan atas, namun setelah 5 hari didapatkan penampilan menjadi normal. Demikian pula pada seorang pasien laki-laki berusia 3 tahun, *CT-scan* toraks setelah 9 hari gejala menunjukkan adanya GGO yang tidak merata di lobus kiri bawah, namun setelah 7 hari kemudian didapatkan tampilan normal. Hal yang serupa didapatkan juga pada seorang pasien laki-laki berusia 4 tahun, *CT-scan* toraks setelah 2 hari MRS memperlihatkan adanya GGO yang tidak merata di lobus kiri bawah, namun *CT-scan* toraks setelah 5 hari kemudian memperlihatkan tampilan normal.

Penelitian oleh Steinberger, dkk [42] dilakukan pada pasien laki-laki berusia 8-18 tahun dan perempuan berusia 14 tahun. Hasil penelitiannya mendapatkan gambaran GGO pada semua kasus dengan variasi untuk masing-masing kasus. Pasien laki-laki usia 8 tahun, *CT-scan* toraks menunjukkan GGO minimal pada lobus kanan bawah dan kiri dan tidak didapatkan konsolidasi. Pada seorang pasien laki-laki berusia 15 tahun, *CT-scan* toraks menunjukkan GGO bilateral dan konsolidasi dalam pola predominan perifer dan basilar, yaitu pola yang dapat dilihat pada cedera paru dengan pengorganisasian akut. Pada seorang pasien laki-laki berusia 18 tahun, *CT-scan* toraks menunjukkan distribusi perifer dan basilar yang dominan dari GGO disertai area GGO di lobus kiri atas dengan konsolidasi perifer. Pada seorang pasien perempuan berusia 14 tahun, *CT-scan* toraks menunjukkan

distribusi bilateral dan perifer dari GGO dengan tanda *halo sign* di lobus kiri bawah, sedangkan lobus atas dan lobus tengah kanan tidak menunjukkan kelainan.

Pada penelitian yang dilakukan oleh Zhang, dkk [43] terhadap pasien laki-laki berusia 20-35 tahun. Hasil penelitiannya mendapatkan gambaran GGO pada semua kasus. Pada pasien laki-laki usia 20 tahun, *CT-scan* toraks menunjukkan GGO di paru kanan bawah, kemudian setelah 4 hari *CT-scan* toraks menunjukkan peningkatan GGO. Pada pasien laki-laki berusia 35 tahun dengan *CT-scan* toraks yang menunjukkan lesi GGO dengan satu nodul padat kecil di dalamnya di lobus paru kiri bawah, dan setelah 5 hari *CT-scan* toraks menunjukkan peningkatan lesi dan tanda pembuluh darah membesar serta tanda bronkogram.

Jiang, dkk [44] melakukan penelitian terhadap pasien laki-laki berusia 66 tahun dan pasien perempuan berusia 57-86 tahun. Didapatkan seorang pasien laki-laki usia 66 tahun *CT-scan* toraks pada hari ke-5 menunjukkan GGO dan setelah hari ke 11, 15, 19 dan 23 menunjukkan ukuran dan kepadatan lesi pada kedua paru secara bertahap dan bermakna menurun. Pada seorang pasien perempuan berusia 57 tahun, *CT-scan* toraks awal dilakukan pada hari pertama menunjukkan GGO multifokus dan konsolidasi kemudian setelah hari ke-7 dan 14 hari menunjukkan lesi paru sembuh secara bermakna. Pada seorang pasien perempuan berusia 57 tahun, *CT-scan* toraks awal dilakukan pada hari ke-4 menunjukkan GGO multifokus dan konsolidasi kemudian setelah hari ke-10 dan 20 menunjukkan lesi paru sembuh secara bermakna. Demikian pula pada seorang pasien perempuan berusia 65 tahun, *CT-scan* toraks dilakukan hari ke- 7 setelah onset penyakit menunjukkan GGO yang tidak merata terlihat di area subpleural lobus kanan bawah paru dan pada hari ke-15, *CT-scan* toraks menunjukkan lesi telah sembuh dengan jelas dan ukurannya berkurang. Pada seorang pasien laki-laki berusia 67 tahun, *CT-scan* toraks dilakukan hari ke-8 setelah onset penyakit menunjukkan GGO yang tidak merata terlihat di area subpleural lobus kanan bawah paru dan pada hari ke-15, *CT-scan* toraks menunjukkan lesi telah sembuh dengan jelas dan ukurannya berkurang.

Pada penelitian oleh Linjun, dkk [45] terhadap 49 pasien, 35 pasien tipe ringan tanpa komorbid, dan 14 pasien tipe parah dengan komorbid. Hasil *CT-scan*

toraks memperlihatkan bahwa pada pasien dengan tipe ringan ditemukan GGO murni (34 pasien), konsolidasi (9 pasien), GGO dengan konsolidasi (19 pasien), dan fibrosis (15 pasien). Pada tipe parah ditemukan GGO murni (14 pasien), konsolidasi (4 pasien), GGO dengan konsolidasi (12 pasien), dan fibrosis (11 pasien). Pasien dengan tipe parah lebih cenderung menyandang penyakit hipertensi. *CT-scan* toraks menunjukkan tidak terdapat perbedaan bermakna antara tipe ringan dan tipe parah dengan adanya GGO murni pada 48 pasien. Pada pasien tipe parah, *CT-scan* toraks menunjukkan lebih banyak GGO dan konsolidasi dibandingkan pada pasien tipe ringan; selain itu pasien tipe parah memiliki lebih banyak fibrosis dibandingkan pasien tipe ringan.

Literatur yang dikaji di atas, menunjukkan hasil *CT-scan* toraks secara akurat mengevaluasi fitur dan luas lesi paru dan *CT-scan* yang khas ialah *ground-glass opacities* (GGO), GGO dengan konsolidasi campuran, penebalan pleura yang berdekatan, penebalan septum interlobular, dan bronkogram udara. GGO, *air bronchogram*, dan konsolidasi pada *CT-scan* toraks pasien COVID- 19 sedangkan berdasarkan usia, jenis kelamin, dan komorbid tidak ditemukan perbedaan bermakna.

*Ground-glass opacity* (GGO) merupakan gambaran berupa opasitas yang ada pada ruang alveoli di rongga paru, yang memiliki ciri tepi yang kabur. Diakibatkan oleh peningkatan atenuasi di paru-paru pada saat dilakukan *thorax CT-Scan*. Gambaran ini merupakan tanda nonspesifik dengan etiologi yang luas seperti infeksi paru-paru, penyakit interstitial kronis dan penyakit alveolar akut. Namun, pada pasien yang terkonfirmasi positif Covid-19, gambaran *thorax CT-Scan* memiliki ciri khas berupa GGO yang umumnya ditemukan pada sisi perifer dan basal paru, serta pada kasus Covid-19 yang berat, dapat ditemukan juga GGO yang disertai dengan konsolidasi paru [46, 47].

### **II.3 Software Image Processing dan DICOM Viewer**

Pengolahan citra digital merupakan salah satu bidang ilmu komputer yang sedang berkembang dengan pesat. Secara umum, pengolahan citra digital dapat diartikan sebagai proses manipulasi citra digital menggunakan bantuan perangkat komputer untuk keperluan tertentu. Pengolahan citra digital memiliki manfaat yang

cukup luas dalam bidang sehari-hari, seperti dalam hal robotika, *computer vision*, pengenalan pola, dan penginderaan jarak jauh [48, 49].

#### 1. Matlab

Matlab merupakan bahasa pemrograman tingkat tinggi dengan kemampuan yang sangat baik untuk menyelesaikan suatu komputasi numerik baik sederhana maupun kompleks. Matlab memiliki kemampuan mengintegrasikan komputasi, visualisasi, dan pemrograman. Oleh karenanya, matlab banyak digunakan dalam bidang riset-riset yang memerlukan komputasi numerik yang kompleks. Penggunaan Matlab meliputi bidang matematika dan komputasi, pembentukan algoritma, akuisisi data, pemodelan, simulasi, dan pembuatan *prototype*, analisa data, eksplorasi, dan visualisasi, serta grafik keilmuan, dan bidang rekayasa pengolahan citra digital [48].

#### 2. Scilab

Scilab merupakan perangkat lunak berlisensi terbuka yang dapat digunakan untuk keperluan proses komputasi numerik. Selain berlisensi terbuka (*open source*), Scilab juga dapat dijalankan di berbagai sistem operasi (*cross-platform*). Perintah-perintah yang digunakan juga mudah dipahami dan menggunakan istilah yang umumnya digunakan di bidang komputasi numerik dan teknik. Scilab memiliki beberapa fitur utama antara lain ribuan fungsi matematika, bahasa pemrograman tingkat tinggi, visualisasi 2-D dan 3-D, komputasi numerik, analisis data, serta dapat berinteraksi dengan bahasa Fortran, C, C++ dan Java [48].

#### 3. GNU Octave

Octave merupakan perangkat lunak berlisensi terbuka (GNU) yang sangat bermanfaat untuk menyelesaikan berbagai masalah komputasi numerik. Perangkat lunak ini menawarkan kemudahan dan kesederhanaan dalam menyelesaikan permasalahan yang berhubungan dengan vektor dan matriks. Hal yang menarik dari perangkat komputasi Octave adalah perintah-perintahnya dibuat semirip mungkin dengan perintah Matlab [48].

#### 4. Python

Python adalah salah satu bahasa pemrograman yang dapat melakukan eksekusi sejumlah instruksi multi guna secara langsung (*interpretatif*) dengan metode

orientasi objek (*Object Oriented Programming*) serta menggunakan semantik dinamis untuk memberikan tingkat keterbacaan *syntax*. Sebagian lain mengartikan Python sebagai bahasa yang berkemampuan menggabungkan kapabilitas dan sintaksis kode yang sangat jelas, juga dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif [48].

Ada banyak sekali perangkat lunak yang dapat digunakan untuk membaca *file* DICOM. Berikut adalah daftarnya [49]:

1. Sante DICOM Viewer Free
2. Micro DICOM
3. RadiAnt
4. VISUS
5. Osirix

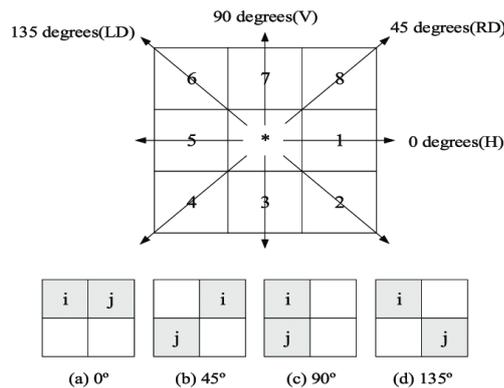
## II.4 Pemrosesan Citra Digital

Pemrosesan citra digital merupakan pemrosesan untuk meningkatkan kualitas dari citra asli. Pada dasarnya, terdapat tiga bidang yang menangani data berupa citra digital, yaitu grafika komputer, visi computer, dan pengolahan citra. Pemrosesan citra terdapat enam jenis operasi, yakni rekonstruksi citra, analisis citra, peningkatan kualitas citra, kompresi citra, segmentasi citra, dan restorasi citra. Pada umumnya informasi pada citra terletak pada strukturnya, agar struktur citra dapat dipahami maka diperlukan penyederhanaan dengan melakukan proses segmentasi citra. Segmentasi citra merupakan proses mempartisi data citra menjadi beberapa objek atau daerah, berdasarkan sifat-sifat tertentu yang bertujuan untuk memperoleh informasi yang ada dalam citra. Terdapat beberapa metode dalam segmentasi citra, yaitu *Thresholding* (*global thresholding* dan *local adaptif thresholding*), *Connected Labelling*, dan *Clustering* (Iterasi, SOM, *fuzzy C-means*, *K-means*) [50].

## II.5 Ekstraksi Fitur *Gray Level Co-occurrence Matrix* (GLCM)

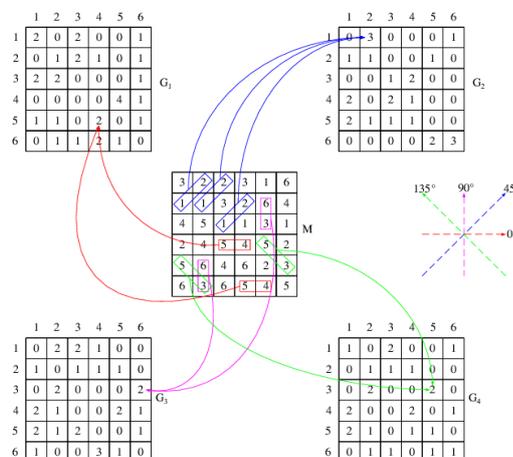
Ekstraksi fitur adalah proses pencarian ciri agar sistem dapat mengenali karakteristik unik dari citra covid dan non-covid-19 yang diwujudkan dalam metode pengolahan citra. *Gray Level Co-occurrence Matrix* (GLCM) merupakan

matriks yang merepresentasikan hubungan ketetanggaan antar piksel dalam citra pada berbagai arah orientasi (sudut  $\theta$ ) dan jarak spasial (d) [51]. Orientasi dinyatakan dalam derajat dan jarak dipresentasikan sebagai piksel. Orientasi dibentuk dalam empat arah sudut, yaitu  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$ , dan jarak antar piksel sebesar 1 piksel [52]. Keempat arah tersebut dapat dilihat pada **Gambar 2.5**.



**Gambar 2.5** Arah dan vector perpindahan GLCM [53]

Ciri GLCM diperoleh dengan menghitung probabilitas hubungan ketetanggaan antar dua pixel. Pendekatan ini bekerja sebagai pembentukan sebuah matriks kookurensi—pada **Gambar 2.6**, yang merupakan matriks bujur sangkar dengan jumlah elemen sebanyak kuadrat jumlah level intensitas piksel pada citra. Kookurensi diartikan sebagai kejadian bersama, berarti setiap titik (i, j) pada matriks kookurensi berorientasi  $\theta$  berisi peluang kejadian piksel bernilai *i* bertetangga dengan piksel bernilai *j* pada jarak *d* [52].



**Gambar 2.6** Pembentukan matriks GLCM [54]

Metode ini memiliki beberapa ciri tekstur yang dapat diekstraksi dengan matriks kookurensi, beberapa di antaranya adalah sebagai berikut [55]:

1. Energi (*Energy*)/Angular Second Moment/*Uniformity*

Energi digunakan untuk menentukan intensitas keabuan dengan ukuran konsentrasi pasangan tertentu. Nilai ASM dan *energy* dapat dihitung menggunakan persamaan 2.3 dan 2.4 [55].

$$ASM = \sum_{i,j=0}^{levels-1} P_{i,j}^2 \quad (2.3)$$

$$Energy = \sqrt{ASM} \quad (2.4)$$

2. Kontras (*Contrast*)

Kontras merupakan perhitungan yang berhubungan dengan jumlah keberagaman intensitas yang ada pada citra keabuan. Nilai kontras dapat dihitung menggunakan persamaan 2.5 [55].

$$Contrast = \sum_{i,j=0}^{levels-1} P_{i,j} (i - j)^2 \quad (2.5)$$

3. Homogenitas (*Homogeneity*)/Inverse difference moment

Homogenitas digunakan untuk menentukan jumlah level keabuan yang semakin tinggi. Nilai homogenitas dihitung menggunakan persamaan 2.6 [55].

$$Homogeneity = \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (2.6)$$

4. Korelasi (*Correlation*)

Korelasi merupakan perhitungan untuk memberikan petunjuk struktur linier dalam citra dengan menunjukkan ketergantungan linier derajat keabuan. Nilai korelasi dapat dihitung dengan persamaan 2.7 [55].

$$Correlation = \sum_{i,j=0}^{levels-1} P_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \quad (2.7)$$

5. Dissimilarity

Mengukur ketidakmiripan suatu tekstur, yang akan bernilai besar bila acak dan sebaliknya akan bernilai kecil bila seragam [55].

$$Dissimilarity = \sum_{i,j=0}^{levels-1} P_{i,j} |i - j| \quad (2.8)$$

Dimana:

$i$  = kolom

$j$  = baris

$P(i, j)$  = baris elemen ke  $i$  dan kolom ke  $j$  dari matriks

$\mu_x$  = nilai rata-rata baris

- $\mu_y$  = nilai rata-rata kolom  
 $\sigma_x$  = standar deviasi dihitung berdasarkan baris  
 $\sigma_y$  = standar deviasi dihitung berdasarkan kolom

## II.6 Metode Klasifikasi

### II.6.1 K-Nearest Neighbors (K-NN)

K-Nearest Neighbors adalah salah satu metode klasifikasi dengan pendekatan *supervised learning* [56]. K-Nearest Neighbors merupakan pengklasifikasi berdasarkan kedekatan lokasi atau jarak suatu data dengan data lainnya yang didasarkan dengan parameter nilai K. Dengan menggunakan metode ini parameter nilai K sangat dibutuhkan. Nilai K umumnya menggunakan nilai ganjil sehingga tidak ada tetangga terdekat (*nearest neighbors*) yang memiliki jumlah yang sama saat menentukan hasil akhir dari klasifikasi K-Nearest Neighbors (K-NN) [14]. Jauh atau dekatnya jarak dapat dihitung melalui besaran jarak yang telah ditentukan dengan menggunakan jarak *Euclidean distance*, jarak *Minkowski*, dan jarak *Mahalanobis*. Namun, jarak *Euclidean distance* seringkali digunakan karena memiliki tingkat akurasi yang tinggi. Konsep jarak *Euclidean distance* melakukan semua perubahan adalah bebas (tidak berkolerasi). Jarak *Euclidean distance* adalah besarnya jarak suatu garis lurus antara objek satu dengan objek lainnya. *Euclidean distance* dijelaskan pada persamaan 2.9 [14, 56].

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^p (x_{ip} - x_{jp})^2} \quad (2.9)$$

Dimana,

- $x_{ip}$  = data *testing* ke-i pada *variable* ke-p  
 $x_{jp}$  = data *training* ke-j pada *variable* ke-p  
 $d(x_i, x_j)$  = jarak *euclidean*  
 $p$  = dimensi data *variable* bebas

### II.6.2 Naïve Bayes

*Naïve Bayes* merupakan pengklasifikasian probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi data tertentu. Algoritma menggunakan teorema Bayes yang mengasumsikan bahwa semua *variable* independen mempertimbangkan *variable* kelas [57]. Teorema Bayes adalah rumus matematika yang digunakan untuk menentukan probabilitas bersyarat seperti persamaan 2.10.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (2.10)$$

Di mana:

A : Hipotesis data merupakan suatu class spesifik

B : Data dengan class yang belum diketahui

P(A|B) : Probabilitas hipotesis A berdasar kondisi B (posteriori probabilitas)

P(A) : Probabilitas hipotesis A (prior probabilitas)

P(B|A) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(B) : Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut [58]:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \quad (2.11)$$

Di mana Variabel *C* merepresentasikan kelas, sementara variabel *F1 ... Fn* merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (*Posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut [58]:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (2.12)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* dilakukan dengan menjabarkan  $(C|F_1, \dots, F_n)$  menggunakan aturan perkalian dapat dilihat pada lampiran 3 [58].

Penjabaran menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk  $(F_1, F_2, \dots, F_n)$  saling bebas (*independent*) satu sama lain, maka berlaku suatu kesamaan pada persamaan 2.13 [58]:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \quad (2.13)$$

Untuk  $i \neq j$ , sehingga:

$$P(F_i|C, F_j) = P(F_i|C) \quad (2.14)$$

Persamaan 2.14 merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* [58]:

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_{ij}^2}} \quad (2.15)$$

Di mana:

$P$  : Peluang

$X_i$  : Atribut ke  $i$

$x_i$  : Nilai atribut ke  $i$

$Y$  : Kelas yang dicari

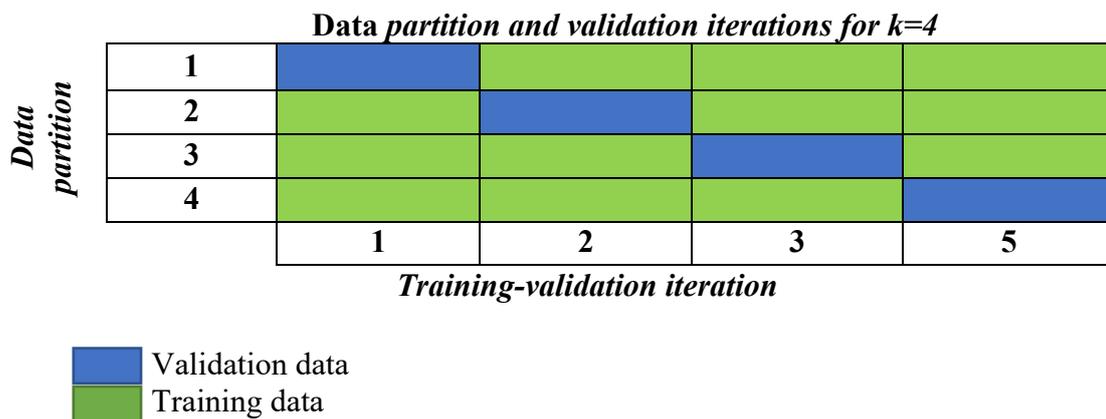
$y_i$  : Sub kelas  $Y$  yang dicari

$\mu$  : *mean*, menyatakan rata-rata dari seluruh atribut

$\sigma$  : *Deviiasi standar*, menyatakan varian dari seluruh atribut

## II.7 K-fold Cross Validation

*K-fold cross validation* atau estimasi rotasi merupakan suatu teknik meminimalisasi bias dengan pengambilan sampel acak dari data training dan testing. Guna memperkirakan seberapa akurat suatu model ketika dijalankan. Selain itu, teknik ini memecah data menjadi  $K$  bagian dengan ukuran yang sama. Training dan testing dilakukan sebanyak  $K$  kali. Untuk percobaan yang pertama adalah subset data  $D_1$  menjadi data testing dan subset data yang lainnya menjadi data training, percobaan kedua subset data  $D_2$  menjadi data testing dan subset data  $D_1$ ,  $D_3$  sampai dengan subset data  $D_k$  merupakan data training [59]. Lebih mudahnya dapat dilihat pada ilustrasi dari *k-fold cross validation* berikut ini.



**Gambar 2.7** Pembagian Data dengan *K-fold cross validation*