

**DETEKSI KANKER PAYUDARA BERBASIS CITRA MAMMOGRAM  
MENGUNAKAN *GRAY LEVEL CO-OCCURRENCE MATRIX*  
(GLCM) DAN *GAUSSIAN NAIVE BAYES***

**SKRIPSI**

**Oleh :  
NADIYAH JIHAN FAUZIYAH  
NIM. 200605110010**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**DETEKSI KANKER PAYUDARA BERBASIS CITRA MAMMOGRAM  
MENGUNAKAN *GRAY LEVEL CO-OCCURRENCE MATRIX*  
(GLCM) DAN *GAUSSIAN NAIVE BAYES***

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :  
**NADIYAH JIHAN FAUZIYAH**  
**NIM. 200605110010**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**HALAMAN PERSETUJUAN**

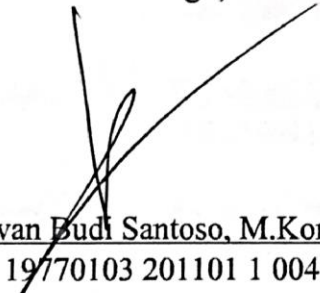
**DETEKSI KANKER PAYUDARA BERBASIS CITRA MAMMOGRAM  
MENGUNAKAN *GRAY LEVEL CO-OCCURRENCE MATRIX*  
(GLCM) DAN *GAUSSIAN NAIVE BAYES***

**SKRIPSI**

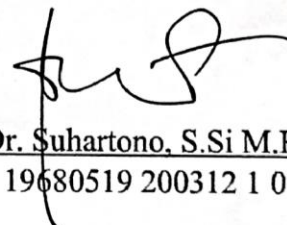
Oleh :  
**NADIYAH JIHAN FAUZIYAH**  
**NIM. 200605110010**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 12 Mei 2024

Pembimbing I,


  
Dr. Irwan Budi Santoso, M.Kom  
NIP. 19770103 201101 1 004

Pembimbing II,

  
Prof. Dr. Suhartono, S.Si M.Kom  
NIP. 19680519 200312 1 001

Mengetahui,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dede Fachri Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

**HALAMAN PENGESAHAN**

**DETEKSI KANKER PAYUDARA BERBASIS CITRA MAMMOGRAM  
MENGUNAKAN *GRAY LEVEL CO-OCCURRENCE MATRIX*  
(GLCM) DAN *GAUSSIAN NAIVE BAYES***

**SKRIPSI**

Oleh:  
**NADIYAH JIHAN FAUZIYAH**  
**NIM. 200605110010**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 07 Juni 2024

**Susunan Dewan Penguji**

Ketua Penguji : Fajar Rohman Hariri, M.Kom  
NIP. 19890515 201801 1 001

Anggota Penguji I : Tri Mukti Lestari, M.Kom  
NIP. 19911108 202012 2 005


Anggota Penguji II : Dr. Irwan Budi Santoso, M.Kom  
NIP. 19770103 201101 1 004

Anggota Penguji III : Prof. Dr. Suhartono, S.Si M.Kom  
NIP. 19680519 200312 1 001

(  )  
(  )  
(  )  
(  )

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Achrul Kurniawan, M.MT, IPM  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nadiyah Jihan Fauziyah

NIM : 200605110010

Fakultas / Prodi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Deteksi Kanker Payudara Berbasis Citra Mammogram  
Menggunakan *Gray Level Co-Occurrence Matrix* (GLCM)  
dan *Gaussian Naive Bayes*.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 10 Juni 2024

Yang membuat pernyataan,



Nadiyah Jihan Fauziyah  
NIM. 200605110010

## **HALAMAN MOTTO**

*... Kebiasaan belum tentu baik, yang baik dibiasakan ...*

## HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT atas segala limpahan rahmat-Nya, serta salam dan shalawat tercurah untuk Nabi Muhammad SAW. Dengan hati yang penuh syukur, saya dedikasikan pencapaian ini dalam bentuk Skripsi kepada:

1. Keluarga terkasih —Ayah, Mama, Mas Lucky, Mbak Ikha, dan Mbah— yang selalu memberikan cinta, kasih sayang, serta doa yang tak pernah putus. Kalian adalah tiang penyangga dalam setiap langkah yang saya ambil.
2. Semua pihak yang tak mungkin saya sebutkan satu per satu, yang telah memberikan motivasi, baik secara langsung maupun tidak langsung. Pertanyaan-pertanyaan seperti “kapan lulus?” dan “kapan wisuda?” yang seringkali membuat saya tersenyum pahit, kini dapat saya jawab dengan kelegaan dan kebahagiaan.
3. Kepada diri saya sendiri, yang telah bertahan dan melewati segala rintangan, kepanikan, kegagalan, dan ketidakpastian. Saya adalah pahlawan dalam kisah ini.

Terima kasih kepada semua yang telah mendukung saya selama masa penyelesaian Skripsi ini. Bagi kalian yang masih berjuang, ingatlah untuk terus bernapas dan bergerak maju, apa pun yang terjadi.

## KATA PENGANTAR

*Assalamualaikum Wr. Wb.*

Alhamdulillah, segala puji ke hadirat Allah SWT atas limpahan rahmat-Nya, yang telah memungkinkan saya untuk menyelesaikan Skripsi ini dengan judul "Deteksi Kanker Payudara Berbasis Citra Mammogram Menggunakan *Gray Level Co-Occurrence Matrix* (GLCM) dan *Gaussian Naive Bayes*." Shalawat dan salam selalu kita haturkan kepada Nabi Muhammad SAW, semoga kita beroleh syafaat-Nya di hari akhir. Aamiin.

Selama perjalanan penuh tantangan dalam merampungkan Skripsi ini, banyak individu yang telah berperan serta dan memberi dukungan yang tidak terhingga. Dengan rasa hormat dan terima kasih yang dalam, penulis menyampaikan doa dan terimakasih yang setulus-tulusnya kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Irwan Budi Santoso dan Prof. Dr. Suhartono, Dosen Pembimbing I dan II saya, yang dengan sabar memberikan arahan dan nasihat selama proses penulisan Skripsi ini.



5. Para dosen Teknik Informatika yang telah memberi bimbingan dan pengajaran yang tidak ternilai sepanjang perjalanan studi saya.
6. Keluarga saya —Ayah, Mama, Mas Lucky, Mbak Ikha, dan Mbah— yang tanpa lelah memberikan dukungan, cinta, dan doa yang menjadi kekuatan saya.
7. Teman-teman seangkatan 2020, yang selalu ada untuk memberikan dukungan dan motivasi.
8. Anggota HIMATIF ENCODER yang telah menjadi bagian dari perjalanan akademik ini, membantu menyempurnakan ide dan inovasi.
9. Sahabat-sahabat "Vaespa Crewmate" Fadilla, Citra, Shinta, Aqila, Wikke, Difa, Atul yang selalu menjadi sumber kekuatan dan inspirasi.
10. Teman-teman "CEST AKHI", yang selalu siap menyegarkan suasana dengan game Mobile Legends di saat jeda penulisan Skripsi.

Terima kasih untuk segala diskusi, teori, dan wawasan yang telah dibagi. Penulis menyadari masih banyak kekurangan dalam penulisan Skripsi ini dan berharap semoga dapat bermanfaat bagi para pembaca.

*Wassalamu'alaikum Wr. Wb*

Malang, 10 Juni 2024

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iv</b>
<b>PERNYATAAN KEASLIAN TULISAN .....</b>	<b>v</b>
<b>HALAMAN MOTTO .....</b>	<b>vi</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiv</b>
<b>ABSTRAK .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>مستخلص البحث.....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Pernyataan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	6
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>7</b>
2.1 Perkembangan Penelitian Deteksi Kanker Payudara .....	7
2.2 Kanker Payudara .....	12
2.3 Citra Mammogram .....	13
2.4 Ekstraksi Fitur <i>Gray Level Co-Occurrence Matrix</i> .....	15
2.5 Naïve Bayes.....	16
<b>BAB III DESAIN DAN IMPLEMENTASI.....</b>	<b>19</b>
3.1 Desain Penelitian.....	19
3.2 Pengumpulan Data .....	20
3.3 Desain Sistem.....	21
3.3.1 Input Image .....	23
3.3.2 <i>Preprocessing</i> Citra .....	23
3.3.3 Ekstraksi Fitur Gray Level Co-Occurrence Matrix (GLCM) .....	26
3.3.4 Uji Normalitas.....	36
3.3.5 Klasifikasi <i>Naïve Bayes</i> .....	38
3.4 Skenario Uji Coba .....	43
3.5 Evaluasi .....	44
<b>BAB IV UJI COBA DAN PEMBAHASAN .....</b>	<b>47</b>
4.1 Langkah-langkah Uji Coba .....	47
4.1.1 Data Pengujian.....	47
4.1.2 Uji Normalitas.....	49
4.2 Hasil Uji Coba Menggunakan Jarak.....	50
4.3 Hasil Uji Coba Kombinasi 1 Sudut.....	53

4.4 Hasil Uji Coba Kombinasi 2 Sudut.....	57
4.5 Hasil Uji Coba Kombinasi 3 Sudut.....	62
4.6 Hasil Uji Coba Menggunakan Semua Sudut.....	66
4.7 Pembahasan.....	67
<b>BAB V PENUTUP .....</b>	<b>85</b>
5.1 Kesimpulan.....	85
5.2 Saran .....	87
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 2.1 Citra mammogram.....	14
Gambar 3.1 Desain penelitian .....	20
Gambar 3.2 Desain sistem .....	22
Gambar 3.3 Hasil <i>grayscale</i> .....	24
Gambar 3.4 Diagram interpolasi <i>bilinear</i> .....	25
Gambar 3.5 Sebelum dan sesudah <i>resize</i> .....	26
Gambar 3.6 Blok diagram pada GLCM.....	27
Gambar 3.7 Arah sudut dalam GLCM.....	28
Gambar 3.8 <i>Flowchart</i> perhitungan <i>co-occurrence, symmetric, normalization</i> ...	29
Gambar 3.9 <i>Flowchart</i> nilai <i>contrast</i> .....	30
Gambar 3.10 <i>Flowchart</i> nilai <i>correlation</i> .....	32
Gambar 3.11 <i>Flowchart</i> nilai <i>energy</i> .....	33
Gambar 3.12 <i>Flowchart</i> nilai <i>homogeneity</i> .....	34
Gambar 3.13 <i>Flowchart</i> nilai <i>entropy</i> .....	35
Gambar 3.14 <i>Flowchart</i> uji normalitas <i>Chi-Square</i> .....	37
Gambar 3.15 <i>Flowchart</i> fungsi <i>fit</i> .....	40
Gambar 3.16 <i>Flowchart</i> fungsi <i>prior</i> dan <i>likelihood</i> .....	41
Gambar 3.17 <i>Flowchart</i> fungsi <i>classify</i> .....	42
Gambar 3.18 <i>Flowchart</i> fungsi <i>y_pred</i> .....	43
Gambar 4.1 Hasil uji normalitas .....	49
Gambar 4.2 <i>Confussion matrix</i> jarak 1 .....	50
Gambar 4.3 <i>Confussion matrix</i> jarak 2 .....	51
Gambar 4.4 <i>Confussion matrix</i> jarak 3 .....	52
Gambar 4.5 <i>Confussion matrix</i> pada derajat 0 .....	53
Gambar 4.6 <i>Confussion matrix</i> pada derajat 45 .....	54
Gambar 4.7 <i>Confussion matrix</i> pada derajat 90 .....	55
Gambar 4.8 <i>Confussion matrix</i> pada derajat 135 .....	55
Gambar 4.9 <i>Confussion matrix</i> pada derajat 0 dan 45 .....	57
Gambar 4.10 <i>Confussion matrix</i> pada derajat 0 dan 90 .....	58
Gambar 4.11 <i>Confussion matrix</i> pada derajat 0 dan 135 .....	58
Gambar 4.12 <i>Confussion matrix</i> pada derajat 45 dan 95 .....	59
Gambar 4.13 <i>Confussion matrix</i> pada derajat 45 dan 135 .....	60
Gambar 4.14 <i>Confussion matrix</i> pada derajat 90 dan 135 .....	60
Gambar 4.15 <i>Confussion matrix</i> pada derajat 0, 45, 90 .....	62
Gambar 4.16 <i>Confussion matrix</i> pada derajat 0, 45, 135 .....	63
Gambar 4.17 <i>Confussion matrix</i> pada derajat 45, 90, 135 .....	64
Gambar 4.18 <i>Confussion matrix</i> pada derajat 0, 90, 135 .....	64
Gambar 4.19 <i>Confussion matrix</i> pada semua <i>sudut</i> .....	66
Gambar 4.20 Grafik hasil performa model berdasarkan jarak .....	68
Gambar 4.21 Grafik hasil performa model berdasarkan kombinasi 1 sudut.....	70

Gambar 4.22 Grafik hasil performa model berdasarkan kombinasi 2 sudut.....	72
Gambar 4.23 Grafik hasil performa model berdasarkan kombinasi 3 sudut.....	74
Gambar 4.24 Grafik hasil performa model berdasarkan kombinasi semua sudut	76

## DAFTAR TABEL

Tabel 2.1 Penelitian deteksi kanker payudara.....	10
Tabel 3.1 Contoh data mammogram kanker payudara .....	21
Tabel 3.2 Skenario uji coba kombinasi sudut .....	44
Tabel 4.1 Rasio pembagian data .....	48
Tabel 4.2 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada jarak 1 .....	51
Tabel 4.3 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 .....	51
Tabel 4.4 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 .....	52
Tabel 4.5 Hasil <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada uji coba menggunakan jarak.....	52
Tabel 4.6 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 .....	54
Tabel 4.7 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 .....	54
Tabel 4.8 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 90 .....	55
Tabel 4.9 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 135 .....	56
Tabel 4.10 Hasil <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada uji coba kombinasi 1 sudut .....	56
Tabel 4.11 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0 dan 45 .....	57
Tabel 4.12 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0 dan 90 .....	58
Tabel 4.13 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0 dan 135 .....	59
Tabel 4.14 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 dan 95 .....	59
Tabel 4.15 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45 dan 135 .....	60
Tabel 4.16 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 90 dan 135 .....	61
Tabel 4.17 Hasil <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada uji coba kombinasi 2 sudut .....	61
Tabel 4.18 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0, 45, 90 .....	63
Tabel 4.19 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0, 45, 135 .....	63
Tabel 4.20 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 45, 90, 135 .....	64
Tabel 4.21 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada derajat 0, 90, 135 .....	65
Tabel 4.22 Hasil <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada uji coba kombinasi 3 sudut .....	65
Tabel 4.23 Nilai <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada semua sudut .....	66
Tabel 4.24 Hasil <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> pada uji coba semua sudut .....	67
Tabel 4.25 Penelitian dengan data yang sama .....	79

## ABSTRAK

Fauziyah, Nadiyah Jihan. 2024. **Deteksi Kanker Payudara Berbasis Citra Mammogram Menggunakan *Gray Level Co-Occurrence Matrix* (GLCM) dan *Gaussian Naive Bayes***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Irwan Budi Santoso, M.Kom (II) Prof. Dr. Suhartono, S.Si M.Kom.

**Kata kunci:** Kanker Payudara, Citra Mammogram, *Gray Level Co-Occurrence Matrix*, *Gaussian Naive Bayes*.

Kanker payudara adalah salah satu tantangan kesehatan utama bagi perempuan di seluruh dunia dengan konsekuensi serius terhadap kualitas dan ekspektasi hidup. *World Health Organization* (WHO) mengidentifikasi kanker payudara sebagai penyebab utama *Disability-Adjusted Life Years* (DALYs) di antara perempuan. Deteksi dini melalui penggunaan teknologi citra medis, khususnya mammogram, diperlukan untuk meningkatkan prognosis dan kelangsungan hidup pasien. Studi ini fokus pada penerapan *Gray Level Co-occurrence Matrix* (GLCM) dan algoritma *Gaussian Naive Bayes* (GNB) dalam analisis citra mammogram untuk meningkatkan deteksi kanker payudara. GLCM efektif dalam mengkarakterisasi tekstur dengan mempertimbangkan hubungan spasial piksel, sementara GNB digunakan untuk mengklasifikasikan fitur tersebut menjadi kategori diagnostik. Penelitian ini mengevaluasi pengaruh berbagai jarak dan kombinasi sudut dalam pengaturan GLCM terhadap performa klasifikasi GNB. Hasil eksperimen menunjukkan bahwa penyesuaian parameter jarak dalam GLCM dapat secara signifikan mempengaruhi akurasi deteksi, dengan jarak 3 memberikan akurasi tertinggi mencapai 84,88%. Selain itu, penggunaan kombinasi sudut yang beragam juga menghasilkan peningkatan akurasi deteksi. Kombinasi sudut terbaik adalah ( $45^\circ$ ,  $135^\circ$ ) pada rasio 80:20, dengan akurasi mencapai 85%. Penelitian ini membuktikan bahwa penggunaan GLCM dan GNB secara bersamaan dalam analisis citra mammogram tidak hanya meningkatkan akurasi deteksi tetapi juga menawarkan pendekatan yang lebih komprehensif dalam memahami variabilitas tekstur citra yang berkaitan dengan kanker payudara. Hasil ini menunjukkan potensi signifikan untuk diterapkan dalam praktik klinis guna meningkatkan deteksi dini dan pengelolaan kanker payudara.

## ABSTRACT

Fauziyah, Nadiyah Jihan. 2024. **Mammogram Image-Based Breast Cancer Detection Using Gray Level Co-Occurrence Matrix (GLCM) and Gaussian Naive Bayes.** Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Irwan Budi Santoso, M.Kom (II) Prof. Dr. Suhartono, S.Si M.Kom.

Breast cancer is a major health challenge for women worldwide, with serious consequences for quality and life expectancy. The World Health Organization (WHO) identifies breast cancer as a leading cause of Disability-Adjusted Life Years (DALYs) among women. Early detection through the use of medical imaging technology, particularly mammograms, is necessary to improve patient prognosis and survival rates. This study focuses on the application of Gray Level Co-occurrence Matrix (GLCM) and the Gaussian Naive Bayes (GNB) algorithm in the analysis of mammogram images to enhance breast cancer detection. GLCM is effective in characterizing texture by considering the spatial relationships of pixels, while GNB is used to classify these features into diagnostic categories. The research evaluates the impact of various distances and angle combinations in GLCM settings on the performance of GNB classification. Experimental results show that adjusting the distance parameter in GLCM can significantly affect detection accuracy, with a distance of 3 achieving the highest accuracy of 84.88%. Additionally, using diverse angle combinations also led to improved detection accuracy, with the best angle combination being (45°, 135°) at a *training* and *testing* data ratio of 80:20, yielding an accuracy of 85%. This study demonstrates that the combined use of GLCM and GNB in mammogram image analysis not only enhances detection accuracy but also offers a more comprehensive approach to understanding the textural variability of images associated with breast cancer. These findings indicate significant potential for clinical application to improve early detection and management of breast cancer.

**Keywords:** *Breast Cancer, Mammogram Image, Gray Level Co-Occurrence Matrix, Gaussian Naïve Bayes.*



## مستخلص البحث

فوزية، نادية جيهان. 2024. الكشف عن سرطان الثدي على أساس صور الماموجرام باستخدام مصفوفة الحدوث المشترك للمستوى الرمادي (GLCM) والغاوسي الساذج بايز. أطروحة. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (I) الدكتور إيروان بودي سانتوسو، ماجيستر كومبيوتر (II) الأستاذ الدكتور سوهارتونو، سارجانا سينس ماجيستر كومبيوتر.

**الكلمات المفتاحية:** سرطان الثدي، صورة الماموجرام، مصفوفة الحدوث المشترك للمستوى الرمادي، غاوسي ساذج بايز.

يُعَدُّ سرطان الثدي من أهم التحديات الصحية التي تواجه النساء في جميع أنحاء العالم، مع عواقب خطيرة على جودة وتوقعات الحياة. تُحدد منظمة الصحة العالمية (WHO) سرطان الثدي كسبب رئيسي لسنوات العمر المصححة بالإعاقة (DALYs) بين النساء. يعتبر الكشف المبكر باستخدام تكنولوجيا التصوير الطبي، خاصةً الماموجرام، ضروريًا لتحسين توقعات ونجاة المرضى. تركز هذه الدراسة على تطبيق مصفوفة المستويات الرمادية للتجاور (GLCM) وخوارزمية بايز المبتكرة (GNB) في تحليل صور الماموجرام لتعزيز اكتشاف سرطان الثدي. تعتبر GLCM فعالة في توصيف النسيج من خلال النظر في العلاقات المكانية للبيكسلات، في حين تُستخدم GNB لتصنيف هذه الميزات إلى فئات تشخيصية. تُقيم هذه الدراسة تأثير المسافات المختلفة وتوليفات الزوايا في إعدادات GLCM على أداء تصنيف GNB. تُظهر النتائج التجريبية أن تعديل معلمة المسافة في GLCM يمكن أن يؤثر بشكل كبير على دقة الاكتشاف، حيث حققت مسافة 3 أعلى دقة بنسبة 84.88%. بالإضافة إلى ذلك، أدى استخدام توليفات زوايا متنوعة أيضًا إلى تحسين دقة الاكتشاف، حيث كانت أفضل توليفة زوايا هي (45°، 135°) بنسبة تدريب واختبار البيانات 80:20، محققة دقة بنسبة 85%. تُظهر هذه الدراسة أن الاستخدام المشترك لـ GLCM و GNB في تحليل صور الماموجرام لا يعزز فقط دقة الاكتشاف بل يقدم أيضًا حجمًا أكثر شمولية لفهم تنوع نسيج الصور المرتبطة بسرطان الثدي. تشير هذه النتائج إلى إمكانات كبيرة للتطبيق السريري لتحسين الكشف المبكر وإدارة سرطان الثدي.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Menurut *World Health Organization* (WHO), kanker payudara merupakan beban terbesar bagi perempuan di seluruh dunia, diperkirakan mencapai 107,8 juta *Disability-Adjusted Life Years* (DALYs), yang mana 19,6 juta DALYs disebabkan oleh kanker payudara (Łukasiewicz *et al.* 2021).

Kanker payudara merupakan masalah kesehatan utama di kalangan wanita karena tingginya angka mortalitas dan morbiditas. Tingkat kelangsungan hidup lima tahun pada kanker payudara metastatik kurang dari 30%, bahkan dengan kemoterapi. Data terbaru GLOBOCAN 2018 yang dihasilkan oleh IARC (*International Agency for Research on Cancer*) dari 185 negara melaporkan 2,3 juta kasus baru (11,7%) kanker payudara dan angka kematian sebesar 6,9% (Bray *et al.* 2018). Identifikasi dini kanker payudara dapat meningkatkan tingkat kelangsungan hidup wanita secara signifikan yang sangat penting, karena kanker payudara dapat disembuhkan pada 95 persen kasus jika diketahui lebih awal (Pramanik *et al.* 2023).

Diagnosis dini dan klasifikasi yang akurat sangat penting untuk memberikan perawatan yang tepat dan meningkatkan tingkat kesembuhan pasien. Dalam beberapa tahun terakhir, kemajuan dalam bidang kecerdasan buatan, khususnya dalam pengolahan citra medis, telah membuka peluang baru dalam mendukung deteksi dan klasifikasi kanker payudara (Ortega-Martorell *et al.* 2022).

Selain itu, dalam penelitian kesehatan dapat ditemukan panduan dan referensi yang berharga dalam hadis dan ayat-ayat Al-Quran yang menekankan pentingnya

menjaga kesehatan dan mengutamakan tindakan pencegahan. Terdapat hadis dari Jabr bin 'Abdullah radhiyallahu 'anhu, Rasulullah SAW bersabda:

لِكُلِّ دَاءٍ دَوَاءٌ، فَإِذَا أُصِيبَ دَوَاءُ الدَّاءِ بَرَأَ بِإِذْنِ اللَّهِ

*“Setiap penyakit ada obatnya. Apabila obat itu tepat untuk suatu penyakit, penyakit itu akan sembuh dengan seizin Allah.” (HR. Bukhari).*

فَسَوْفَ يُحَاسَبُ حِسَابًا يَسِيرًا

*“Maka dia akan diperiksa dengan pemeriksaan yang mudah.” (Q.S Al-Insyiqaq (84):8).*

Dari hadis dan ayat Al-Quran di atas, kita dapat mengambil pelajaran bahwa dalam Islam menjaga kesehatan dan mencegah penyakit diutamakan. Kesehatan dianggap sebagai anugerah dan Allah telah memberikan obat atau solusi untuk setiap penyakit. Dengan merujuk pada petunjuk dalam Al-Quran dan hadis, umat Islam diajak untuk menghargai nikmat kesehatan yang diberikan oleh Allah dan berusaha menjaga kesehatan tubuh sebagai bentuk syukur. Kesadaran untuk mencegah penyakit juga diterapkan sebagai bentuk ikhtiar dan tindakan preventif dalam menjaga amanah kesehatan yang diberikan oleh-Nya.

Salah satu cara paling efisien untuk mendiagnosis kanker payudara adalah melalui pencitraan medis. Gambar medis merupakan salah satu sumber informasi terpenting untuk identifikasi dan diagnosis berbagai penyakit dan anomali yang memungkinkan ahli radiologi memeriksa struktur interior tubuh manusia. Hal ini penting dalam diagnosis penyakit klinis, evaluasi pengobatan, dan deteksi kelainan pada berbagai organ tubuh seperti paru-paru, otak, payudara, dan lain sebagainya

(Khamparia *et al.* 2021). Penelitian di bidang ini telah berkembang secara signifikan selama tiga dekade terakhir. Mamografi payudara merupakan metode yang paling ekonomis diantara berbagai metode deteksi lainnya yaitu *Computerized Tomography* (CT), *Thermal Imaging*, *Magnetic Resonance Imaging* (MRI), *Histology imaging*, dan USG (Husaini *et al.* 2020).

Dalam pemrosesan gambar dibutuhkan metode untuk ekstraksi fitur, matriks *Gray Level Co-occurrence Matrix* (GLCM) mengkaji tekstur dengan mempertimbangkan hubungan spasial piksel. Fungsi GLCM mengkarakterisasi tekstur gambar dengan menghitung kemunculan pasangan piksel dengan nilai tertentu dan dalam hubungan spasial tertentu. Oleh karena itu, matriks GLCM dibuat untuk mengekstraksi pengukuran statistik suatu gambar (Marianingsih *et al.* 2019).

Untuk pengklasifikasian citra medis tentunya akan melibatkan *Machine Learning* dalam pengerjaannya. Dari beberapa kasus, terdapat beberapa metode yang bisa digunakan, diantaranya *Transfer Learning*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Trees* dan *Random Forests*, serta *Gaussian Naïve Bayes* (Hafiz *et al.* n.d.). *Naïve Bayes* merupakan algoritma pembelajaran mudah yang memanfaatkan aturan *Bayes* dengan premis atau asumsi tinggi yang karakteristiknya bergantung pada kemandirian yang diberikan oleh kelas. Algoritma ini digunakan dalam klasifikasi yang didasarkan pada statistik dan probabilitas sederhana yang banyak diterapkan dalam metode pembelajaran identifikasi citra medis. Kelebihan lain dari *naive bayes* adalah bahwa perhitungannya dapat dilakukan secara paralel dan terdistribusi, dan ia juga

menghasilkan hasil yang relatif baik ketika digunakan untuk klasifikasi citra (Susanto *et al.* 2020).

Dan dalam penulisan ini akan menggunakan *Gaussian Naïve Bayes* dengan fitur *Gray Level Co-occurrence Matrix* (GLCM) yang telah terbukti berhasil dalam berbagai tugas pengenalan gambar dan ekstraksi fitur dengan akurasi 80% (Laksono & Prabowo, 2023). Oleh karena itu, penggunaan metode ini dalam klasifikasi kanker payudara dapat memberikan kontribusi dalam meningkatkan akurasi dan keandalan diagnosis.

Diharapkan dapat memberikan kontribusi dalam pengembangan alat bantu diagnostik yang membantu para ahli dalam deteksi dini dan penanganan kanker payudara secara lebih efektif.

## **1.2 Pernyataan Masalah**

Pernyataan masalah pada penulisan ini yaitu bagaimana performa metode ekstraksi fitur *Gray Level Co-occurrence Matrix* (GLCM) dan algoritma *Gaussian Naive Bayes* dalam mendeteksi kanker payudara berbasis citra mammogram.

## **1.3 Batasan Masalah**

Adapun batasan masalah pada penulisan ini yaitu:

1. Data yang digunakan yaitu data citra mammogram yang berasal dari penelitian Eric A. Scucciamarra yang dapat diakses secara *online* di platform *kaggle*. Data ini berjumlah 4000 image dengan atribut *images* dan *labels* dengan 2000 sampel kanker dan 2000 sampel tidak-kanker.

2. Hasil dari penulisan ini mendeteksi dua kriteria diagnosis, yaitu tidak kanker dengan nilai negatif dan kanker dengan nilai positif. Dengan metrik evaluasi yang digunakan untuk mengukur kinerja sistem deteksi adalah *Accuracy*.

#### **1.4 Tujuan Penelitian**

Tujuan utama penulisan ini adalah untuk mengembangkan metode klasifikasi yang efektif untuk mendeteksi kanker payudara menggunakan *Gaussian Naive Bayes* dengan fitur GLCM. Dengan meningkatkan akurasi deteksi dini, diharapkan dapat mengidentifikasi kasus-kasus kanker payudara pada tahap awal, yang dapat meningkatkan peluang kesembuhan pasien.

#### **1.5 Manfaat Penelitian**

Penulis berharap dengan menggunakan kombinasi *Gaussian Naive Bayes* dan fitur GLCM dapat mengadopsi pendekatan statistik dan probabilitas untuk mendiagnosis kanker payudara berdasarkan citra mammogram. Manfaat penelitian ini dapat diperluas kepada praktisi medis, peneliti di bidang kedokteran, serta pihak-pihak yang terlibat dalam diagnosis dan pengobatan kanker payudara. Dengan model ini, diharapkan peningkatan akurasi dan keandalan dalam klasifikasi kanker payudara, yang pada gilirannya dapat membantu penyedia layanan kesehatan untuk pengambilan keputusan yang lebih tepat dan efisien. Selain itu, hasil penelitian ini juga dapat membuka jalan untuk pengembangan metode diagnostik yang lebih canggih dan dapat diandalkan dalam pertempuran melawan kanker payudara.

## **1.6 Sistematika Penulisan**

Sistematika penulisan yang digunakan pada penulisan ini sebagai berikut:

### **A. BAB I PENDAHULUAN**

Bab ini merupakan pengantar dari penelitian yang memberikan gambaran umum tentang latar belakang, perumusan masalah, tujuan, dan ruang lingkup penelitian.

### **B. BAB II TINJAUAN PUSTAKA**

Bab ini membahas tinjauan pustaka yang relevan dengan topik penelitian, termasuk teori-teori dan penelitian terdahulu yang mendukung penelitian ini.

### **C. BAB III DESAIN DAN IMPLEMENTASI**

Bab ini menjelaskan rancangan penelitian, metodologi, serta implementasi langkah-langkah penelitian, termasuk teknik pengumpulan dan analisis data.

### **D. BAB IV UJI COBA DAN PEMBAHASAN**

Bab ini berisi hasil uji coba, analisis data, dan pembahasan temuan penelitian. Hasil uji coba dievaluasi dan dibandingkan dengan teori yang telah dikaji.

### **E. BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian dan saran-saran untuk pengembangan penelitian selanjutnya.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Perkembangan Penelitian Deteksi Kanker Payudara**

Penelitian tentang deteksi kanker telah menjadi fokus utama dalam upaya pencegahan dan pengobatan penyakit ini. Pada penelitian Zhang *et al.* (2021) melakukan pengolahan citra medis dalam diagnosis kanker payudara. Dengan menggabungkan teknologi kecerdasan buatan dengan metode CNN dan pengolahan citra medis, penelitian menunjukkan potensi besar dalam meningkatkan deteksi dini dan pengelolaan penyakit ini.

Adapun penelitian Pramanik *et al.* (2023) mengembangkan metode deteksi kanker payudara pada mammogram, memanfaatkan model deep learning VGG16 dengan mekanisme perhatian dan algoritma meta-heuristik *Social Ski-Driver* (SSD). Melalui kombinasi ini, peneliti berhasil mencapai akurasi 96,07% dalam mengidentifikasi dan membedakan payudara yang bersifat ganas dan sehat, berdasarkan uji coba pada basis data DDSM.

Dalam penelitian Altameem *et al.* (2022) menggunakan empat set data mamografi digunakan untuk menganalisis gambar-gambar normal, jinak, dan ganas sebanyak 1145 gambar menggunakan berbagai model *deep learning* (*Inception V4*, *ResNet-164*, *VGG-11*, dan *DenseNet 121*) sebagai klasifikasi dasar. Penelitian ini menghasilkan model ensemble *Inception V4* yang diusulkan dengan fungsi peringkat fuzzy berbasis Gompertz mencapai tingkat akurasi 99,32% sehingga dapat mendeteksi kanker payudara secara dini.



Penelitian yang dilakukan Khamparia *et al.* (2021) berfokus untuk mendeteksi kanker payudara. Model *Modified VGG* (MVGG) diusulkan dan diimplementasikan pada dataset gambar 2D dan 3D dari mammogram. Hasil eksperimen menunjukkan bahwa model *transfer learning hibrida* yang diusulkan (gabungan MVGG dan *ImageNet*) memberikan akurasi sebesar 94,3%. Di sisi lain, hanya arsitektur MVGG yang diusulkan memberikan akurasi sebesar 89,8%. Oleh karena itu, dapat dinyatakan dengan jelas bahwa jaringan pra-pelatihan hibrida yang diusulkan mengungguli *Convolutional Neural Networks* lain yang dibandingkan.

Dari penelitian Miao & Tang, (2022) didapatkan bahwa penggunaan metode *threshold* segmentasi citra berbasis transformasi gelombang *Daubechies Wavelet Transform* (DWT) menghasilkan rata-rata akurasi pengukuran sebesar 92,9% dan tingkat kesalahan positif palsu rata-rata sebesar 86% untuk setiap citra. Melalui teknik *multiplex polymerase chain reaction* (mPCR) dan *mPCR-next-generation sequencing* (mPCR-NGS), tidak ditemukan mutasi patogenik pada 7 pasien dengan mutasi genetik BRCA1/2 berisiko tinggi. Sebaliknya, terdeteksi polimorfisme nukleotida tunggal (*single nucleotide polymorphism*, SNP) pada wilayah kode *non-sinonim*, sesuai dengan hasil uji Sanger.

Sedangkan dalam penelitian Kamel *et al.* (2019) menggunakan algoritma *Gaussian Naive Bayes* digunakan untuk klasifikasi kanker. Algoritma ini diuji dengan menerapkannya pada dua dataset, yaitu *Wisconsin Breast Cancer Dataset* (WBCD) dan dataset kanker paru-paru. Hasil evaluasi dari algoritma yang diusulkan mencapai akurasi prediksi kanker payudara sebesar 98%.

Dari penelitian yang dilakukan Sun *et al.* (2021) berhasil menghasilkan metode deteksi massa payudara baru yang terinspirasi dari proses deteksi manusia. Metode ini menggunakan pengolahan citra melalui morfologi matematika dan pencocokan template untuk menemukan wilayah yang dicurigai sebagai massa payudara. Selanjutnya, wilayah-wilayah tersebut diklasifikasikan menggunakan jaringan saraf konvolusi (CNN). Pembatas kotak massa payudara diperoleh melalui metode morfologi matematika dan pencocokan template, kemudian disempurnakan menggunakan algoritma Swarm Partikel (PSO).

Adapun penelitian Susilowati *et al.* (2016) menggunakan metode *K-Means Clustering* untuk membedakan kanker payudara menjadi jenis *benign* (jinak) dan *malignant* (ganjal) berdasarkan faktor-faktor yang mempengaruhi. Data dari 41 pasien dengan 10 gejala kanker payudara digunakan. Hasil clustering menunjukkan dua cluster: Cluster 1 (C1) dengan 11 data mengindikasikan kanker payudara ganas, sementara Cluster 2 (C2) dengan 30 data mengindikasikan kanker payudara jinak.

Dalam penelitian Rohman *et al.* (2021) menganalisa kanker payudara dengan mengusulkan penggunaan GLCM pada tahap praproses CNN. Hasil menunjukkan metode gabungan GLCM dan CNN menghasilkan waktu komputasi lebih singkat (2040 detik) dan akurasi yang lebih tinggi (92,26% - data latih, 94,16% - data uji) dibandingkan dengan metode CNN murni. Selain itu, penelitian tersebut menunjukkan bahwa integrasi GLCM dapat meningkatkan kemampuan ekstraksi fitur tekstur, yang pada gilirannya meningkatkan performa klasifikasi jaringan kanker payudara. Ini mengindikasikan potensi besar dari metode *hybrid* untuk aplikasi medis lainnya, di mana presisi dan kecepatan sangat krusial.

Penelitian Kanya Kumari & Naga Jagadesh, (2022) mengatasi tantangan deteksi dini kanker payudara dengan menggunakan teknik pemrosesan gambar dan pembelajaran mesin. Metodenya melibatkan empat langkah: akuisisi gambar, pra-pemrosesan dengan CLAHE, ekstraksi fitur menggunakan AGLCM, dan klasifikasi dengan model *XGBoost*. Hasilnya, model CLAHE + AGLCM + *XGBoost* menunjukkan performa lebih baik dibandingkan metode eksisting, dievaluasi pada dataset MIAS.

Rangkuman penelitian ini dapat dilihat pada Tabel 2.1, yang mana dari penelitian-penelitian sebelumnya yang telah dilakukan dalam penelitian deteksi kanker terdapat berbagai pendekatan yang telah diuji untuk meningkatkan akurasi deteksi dan klasifikasi. Peneliti ingin melakukan penelitian yang mengadopsi *Gaussian Naive Bayes* dengan ekstraksi fitur *Gray Level Co-Occurrence Matrix* dalam analisis citra mamogram, dengan harapan dapat memperbaiki dan meningkatkan tingkat akurasi deteksi kanker payudara. Selain itu, penelitian ini juga bertujuan untuk mengeksplorasi potensi metode ini dalam menghadapi berbagai tantangan dalam analisis citra medis.

Tabel 2.1 Penelitian deteksi kanker payudara

No	Nama Peneliti	Judul Penelitian	Metode	Hasil		Perbedaan
				Data	Akurasi	
1	Zhang <i>et al.</i> , (2021)	<i>Review of Breast Cancer Pathological Image Processing</i>	<i>Deep learning dan einforcement learning</i>	Citra	-	<b>Akurasi</b>
2	Pramanik <i>et al.</i> , (2023)	<i>Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms</i>	VGG16 dan algoritma meta-heuristik <i>Social Ski-Driver</i>	Citra	96,07%	<b>Metode</b>

No	Nama Peneliti	Judul Penelitian	Metode	Hasil		Perbedaan
				Data	Akurasi	
3	Altameem <i>et al.</i> , (2022)	<i>Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques</i>	<i>Inception V4, ResNet-164, VGG-11, dan DenseNet121</i>	Citra	99,32%	<b>Metode</b>
4	Khamparia <i>et al.</i> , (2021)	<i>Diagnosis of breast cancer based on modern mammography using hybrid transfer learning</i>	MVGG dan ImageNet	Citra	94,3%	<b>Metode</b>
5	Miao & Tang, (2022)	<i>Detection of Breast Cancer Lump and BRCA1/2 Genetic Mutation under Deep Learning</i>	<i>Threshold berbasis Daubechies wavelet transform</i>	Citra	92,9%	<b>Metode</b>
6	Kamel <i>et al.</i> , (2019)	<i>Cancer Classification Using Gaussian Naive Bayes Algorithm</i>	<i>Gaussian Naive Bayes</i>	Text	98%	<b>Metode Ekstraksi dan Data</b>
7	Sun <i>et al.</i> , (2021)	<i>Breast mass detection in mammography based on image template matching and cnn</i>	CNN dan PSO	Citra	85,82%	<b>Metode</b>
8	Susilowati <i>et al.</i> , (2016)	Diagnosa Penyakit Kanker Payudara Menggunakan Metode K-Means Clustering	<i>K-Means Clustering</i>	Text	-	<b>Metode, Data, Akurasi</b>
9	Rohman <i>et al.</i> , (2021)	<i>Implementasi Filter GLCM Terhadap Sistem Klasifikasi Kanker Payudara Dengan Metode CNN</i>	GLCM dan CNN	Citra	94,16%	<b>Metode Deteksi</b>
10	Kanya Kumari & Naga Jagadesh, (2022)	<i>A Robust Feature Extraction Technique for Breast Cancer Detection using Digital Mammograms based on Advanced GLCM Approach</i>	CLAHE + AGLCM	Citra	95,6%	<b>Metode</b>

## 2.2 Kanker Payudara

Kanker payudara merupakan tantangan kesehatan global yang signifikan. Pada kasusnya kanker ini merupakan kanker yang paling sering didiagnosis di dunia dengan perkiraan 2,26 juta kasus tercatat pada tahun 2020 dan merupakan penyebab utama kematian akibat kanker di kalangan wanita. Meskipun secara historis dianggap sebagai penyakit di negara-negara maju, lebih dari separuh diagnosis kanker payudara dan dua pertiga kematian terkait kanker payudara terjadi di wilayah kurang berkembang di dunia pada tahun 2020 (Sung *et al.* 2021).

Kanker payudara sering dimulai dari pertumbuhan berlebih pada saluran payudara, yang kemudian dapat berkembang menjadi tumor jinak atau bahkan karsinoma metastatik jika terus-menerus terpapar oleh berbagai faktor karsinogenik. Lingkungan mikro di sekitar tumor, seperti pengaruh stroma (jaringan penunjang) atau makrofag (sel-sel kekebalan tubuh) memainkan peran penting dalam memicu dan mempercepat inisiasi serta perkembangan kanker payudara. Proses ini dapat melibatkan interaksi yang kompleks antara sel-sel kanker dan komponen mikro lingkungan mereka, yang seringkali mengarah pada perubahan genetik dan epigenetik yang memperkuat kemampuan sel kanker untuk tumbuh dan menyebar.

Terdapat beberapa faktor yang menyebabkan kanker payudara. Yang paling sering diidentifikasi oleh wanita yang pernah didiagnosis dengan kanker payudara adalah riwayat keluarga, lingkungan, stres, dan nasib/kesempatan. Riwayat keluarga, terutama jika terdapat kerabat dekat yang juga mengalami kanker payudara, dapat meningkatkan risiko karena adanya faktor genetik yang mungkin

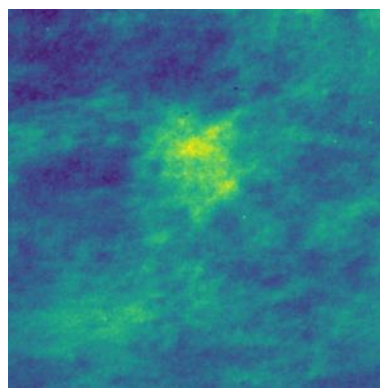
diwariskan. Lingkungan, termasuk paparan radiasi atau bahan kimia tertentu, juga dapat menjadi faktor pemicu. Stres kronis, yang dapat mempengaruhi sistem kekebalan tubuh dan keseimbangan hormonal, turut berperan dalam perkembangan kanker. Faktor risiko lain seperti penuaan, riwayat reproduksi, seperti usia saat pertama kali menstruasi dan menopause, serta gaya hidup, termasuk pola makan, aktivitas fisik, dan konsumsi alkohol, juga berkontribusi terhadap peningkatan risiko terkena kanker payudara (Dumalaon-Canaria *et al.* 2014).

### **2.3 Citra Mammogram**

Ditemukan pada tahun 1895 oleh Rontgen, mammogram pertama dilakukan pada tahun 1913 oleh Albert Solomon. Otto Kleinschmidt dan Erwin Payr menjelaskan peran penting mamografi dalam deteksi dini kanker payudara pada tahun 1927. Stafford Warren melakukan mammogram pertama di Amerika Serikat pada tahun 1930, sementara Gershon-Cohen dan Helen Ingleby memperkaya pemahaman tentang keganasan payudara. Raul Leborgne, pada tahun 1951, menyoroti pentingnya mikrokalsifikasi dalam membedakan penyakit jinak dan ganas. Meskipun kesadaran masyarakat tentang kanker meningkat setelah Perang Dunia II, mamografi baru mulai diterapkan sebagai alat skrining pada akhir tahun 1960an (Nicosia *et al.* 2023).

Mammografi sendiri merupakan metode pencitraan sinar-X yang digunakan untuk memeriksa payudara guna mendeteksi dini kanker dan penyakit payudara lainnya. Selain sebagai alat diagnostik, mammografi juga digunakan sebagai alat skrining. Saat melakukan mammografi, payudara pasien diletakkan di atas pelat penyangga datar dan dipadatkan dengan pelat paralel yang disebut paddle. Mesin

sinar-X menghasilkan sejumlah kecil sinar-X yang melewati payudara menuju detektor yang terletak di sisi berlawanan. Detektor bisa berupa pelat film fotografi yang menangkap gambar sinar-X di film, atau detektor padat yang mengirimkan sinyal elektronik ke komputer untuk membentuk gambar digital. Hasil dari mammografi disebut mamogram. Adapun hasil dari citra mamogram dapat dilihat dari Gambar 2.1.



Gambar 2.1 Citra mamogram

Banyak penelitian yang dilakukan dalam mengolah citra mamogram, seperti penelitian yang dilakukan (Khamparia *et al.* 2021) bertujuan untuk memperbaiki deteksi kanker payudara pada citra mamogram. Penelitian menerapkan metode transfer learning dalam pengolahannya. Penerapan arsitektur yang diusulkan dapat menjadi alat efektif bagi radiolog dalam mengurangi tingkat kesalahan pada analisis citra mamogram, yang pada gilirannya meningkatkan efisiensi deteksi kanker payudara dari citra mamogram.

Penelitian Sun *et al.* (2021) berfokus pada deteksi massa payudara pada citra mamogram menggunakan metode pengolahan citra medis. Kami mengusulkan pendekatan baru yang terinspirasi dari proses deteksi manusia. Metode tersebut memanfaatkan proses pra-pemrosesan citra dan teknik pencocokan template untuk

mengidentifikasi wilayah-wilayah yang mencurigakan sebagai massa payudara. Evaluasi dilakukan terhadap database DDSM, di mana hasil eksperimen menunjukkan superioritas metode yang diusulkan dibandingkan dengan metode deteksi lainnya.

#### **2.4 Ekstraksi Fitur *Gray Level Co-Occurrence Matrix***

*Gray Level Co-Occurrence Matrix* (GLCM) diperkenalkan pada tahun 1973 untuk mengevaluasi tekstur visual gambar berdasarkan statistik susunan spasial intensitas piksel dan dianggap sebagai deskriptor tekstur yang umum dan sensitif (Haralick *et al.*, 1973). Proses pembentukan GLCM melibatkan penghitungan frekuensi kemunculan pasangan piksel dengan jarak dan arah tertentu dalam gambar. Ini memungkinkan ekstraksi fitur tekstur seperti kontras, korelasi, energi, dan homogenitas, yang memberikan wawasan tentang sifat-sifat tekstur dari gambar yang diamati (Aouat *et al.* 2021)

Penggunaan algoritma GLCM menjadi krusial dalam analisis citra medis untuk diagnosis penyakit seperti kanker dan evaluasi tekstur jaringan. Penggunaannya meluas dalam penelitian onkologi, di mana GLCM membantu mengidentifikasi pola tekstur yang mungkin terkait dengan kondisi patologis tertentu, memungkinkan para ahli untuk membuat diagnosis yang lebih akurat dan memahami perkembangan penyakit dengan lebih baik. Selain itu, dalam bidang kedokteran, GLCM juga digunakan dalam penelitian pengembangan obat dan terapi, di mana analisis tekstur gambar memainkan peran penting dalam pemetaan dan pemahaman respons pasien terhadap pengobatan (Fan *et al.* 2023).



Dalam penelitian (Kanya Kumari & Naga Jagadesh, 2022) menggunakan pemanfaatan teknik *Gray-Level Co-occurrence Matrix* (GLCM) untuk mengekstrak fitur-fitur terkait dari gambar mamografi yang telah diproses sebelumnya, ditambah dengan penerapan berbagai algoritma pembelajaran mesin untuk klasifikasi gambar. Metodologi yang diusulkan mencakup empat langkah utama akuisisi gambar, pra-pemrosesan menggunakan *Contrast Limited Advanced Histogram Equalization* (CLAHE) untuk merambah kontras gambar, ekstraksi fitur melalui GLCM untuk menangkap fitur berbasis tekstur, intensitas, dan bentuk yang penting untuk identifikasi kelainan .

Meskipun GLCM telah menunjukkan kegunaannya yang luas, tantangan tertentu masih perlu diatasi. Parameter-parameter GLCM seperti jarak dan arah pasangan piksel serta interpretasi fitur tekstur memerlukan penyesuaian yang cermat tergantung pada aplikasi spesifik. Selain itu, dalam konteks citra medis, ukuran dan kualitas citra medis juga menjadi faktor penting yang memengaruhi kinerja GLCM. Dengan pemahaman yang mendalam tentang GLCM dan tantangan yang dihadapinya, penelitian dan inovasi terus berkembang untuk memperluas aplikasi GLCM dalam berbagai disiplin ilmu, termasuk bidang kedokteran dan ilmu komputer.

## **2.5 Naïve Bayes**

Algoritma *Naïve Bayes* telah menjadi pendekatan yang sangat berguna dalam pembelajaran mesin, terutama untuk klasifikasi data. Algoritma ini didasarkan pada teorema *Bayes* dan mengasumsikan independensi antara fitur-fitur dalam data. Meskipun sederhana, *Naive Bayes* cukup efektif dalam berbagai aplikasi klasifikasi,

termasuk dalam analisis medis (Berry, 1997). Ada tiga klasifikasi Naïve Bayes yang umum digunakan, yaitu, *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes*. Algoritma ini terbukti mampu mengklasifikasikan jenis-jenis kanker berdasarkan fitur-fitur yang diekstraksi dari data medis dengan tingkat akurasi yang memuaskan. Kemampuan algoritma ini dalam mengatasi kompleksitas data medis telah membuatnya menjadi pilihan yang menonjol dalam dunia penelitian medis (Anand *et al.* 2022b).

Selain itu, algoritma *Gaussian Naïve Bayes* telah sukses diterapkan dalam konteks segmentasi objek pada citra satelit atau citra *Google Earth*. Dengan algoritma ini dioptimalkan untuk mengklasifikasikan objek ruang terbuka hijau dengan akurasi yang tinggi. Keberhasilan ini memberikan dasar yang kuat untuk pemetaan dan pemantauan lingkungan secara spasial. Melalui penggunaan algoritma ini, segmentasi objek menjadi lebih akurat dan efisien, menguntungkan dalam pemahaman dan pengelolaan lingkungan (Santoso *et al.*, 2018).

Adapun penelitian Budi Santoso *et al.* (2024) bertujuan untuk meningkatkan kinerja klasifikasi citra tekstur menggunakan *Gaussian Naïve Bayes Classifier* (GNBC) dengan memperhatikan faktor ukuran citra. Metode yang diusulkan mencakup pengoptimalan ukuran citra menggunakan metode interpolasi *Nearest Neighbor* (NN) untuk meningkatkan performa GNBC dalam klasifikasi citra tekstur. Hasil penelitian menunjukkan bahwa dengan mengimplementasikan metode interpolasi NN dan GNBC dapat mencapai akurasi klasifikasi yang lebih baik dengan ukuran citra yang dioptimalkan.

Terdapat penelitian Laksono & Prabowo, (2023) yang melibatkan deteksi tumor otak pada citra menggunakan GLCM, kemudian diikuti oleh klasifikasi citra berdasarkan keberadaan tumor otak menggunakan *Algoritma Naïve Bayes Classification*. Tujuan utama penelitian ini adalah untuk mengevaluasi akurasi GLCM dan *Naïve Bayes Classification* dalam memprediksi citra tumor otak. Melalui uji coba pada 253 dataset pasien, hasil menunjukkan bahwa *Naïve Bayes* mampu mencapai tingkat akurasi sebesar 80%, dengan rasio data *testing* dan data *training* 20%:80%. Selain itu, rata-rata nilai presisi dan *recall* yang dihasilkan adalah 85%.

Algoritma ini efektif dalam mengklasifikasikan jenis-jenis kanker berdasarkan fitur-fitur yang diekstraksi dari data medis. Keandalannya dalam mengatasi kompleksitas data medis telah membuatnya menjadi pilihan yang menonjol dalam penelitian medis serta memberikan kontribusi penting dalam diagnosis dan pengobatan penyakit kanker.

## **BAB III**

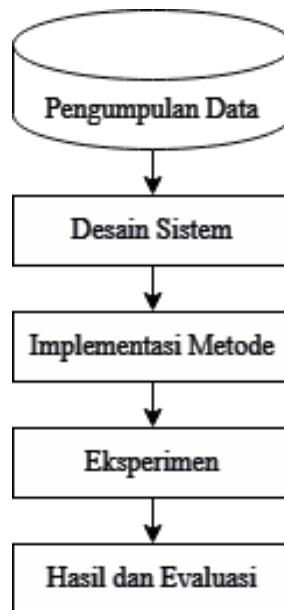
### **DESAIN DAN IMPLEMENTASI**

Bab ini akan menjelaskan tentang perancangan sistem dan implementasi pada penelitian ini. Dalam penelitian ini memiliki beberapa tahapan, termasuk tahapan kebutuhan sistem yang akan dibuat dan solusi untuk masalah Deteksi Kanker Payudara Berbasis Citra Mammogram menggunakan *Gray Level Co-occurrence Matrix (GLCM)* dan *Gaussian Naive Bayes*.

#### **3.1 Desain Penelitian**

Desain penelitian merupakan kerangka kerja yang sangat penting, berisi tahapan yang dilakukan secara sistematis dan berurutan. Proses ini dimulai dengan pengumpulan data. Selanjutnya, tahapan berlanjut dengan pembuatan desain sistem, yang meliputi perencanaan proses input, proses transformasi atau pengolahan data, serta proses output.

Setelah desain sistem selesai, metode atau sistem yang dirancang diimplementasikan dalam bentuk sistem yang konkret. Tahap ini melibatkan penerapan teknik dan teknologi yang dipilih sesuai desain untuk memastikan operasi sistem sesuai spesifikasi yang ditetapkan. Langkah akhir adalah evaluasi dan analisis hasil, dimana peneliti menilai kesesuaian hasil dengan hipotesis yang diajukan dan standar yang ditetapkan, serta mengidentifikasi kekurangan dalam proses penelitian. Alur proses desain penelitian ini tergambar pada Gambar 3.1, mengilustrasikan setiap langkah dari pengumpulan data hingga evaluasi hasil.

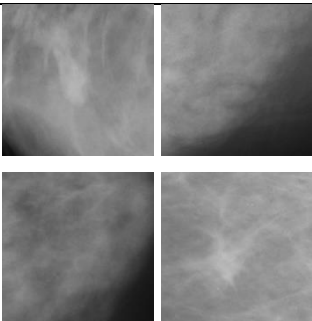
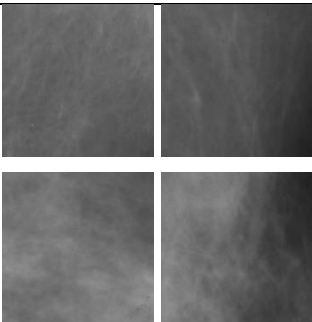


Gambar 3.1 Desain penelitian

### 3.2 Pengumpulan Data

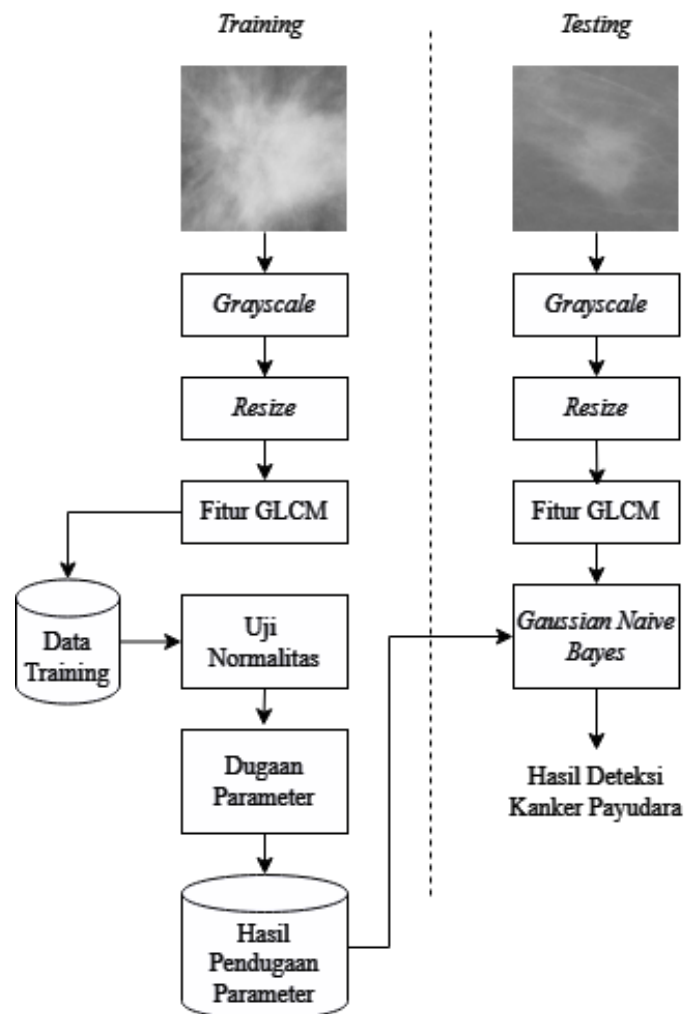
Pengumpulan data dalam penelitian ini merupakan langkah penting untuk mendapatkan informasi relevan. Data diambil dari dataset mammografi yang tersedia di situs Kaggle sebagai dataset dan uji coba pada sistem. Proses pengumpulan data melibatkan pengunduhan dataset dan analisis konten. Dataset tersebut berasal dari *Digital Database for Screening Mammography* (DDSM) dan *Curated Breast Imaging Subset of DDSM* (CBIS-DDSM), telah diproses menjadi gambar 299x299 dengan ekstraksi *Region of Interest* (ROI), disimpan dalam file tfrecords dan terdiri dari 55.890 contoh latihan, dengan 14% *positif* (7825) dan 86% negatif (48065), dibagi dalam 5 file tfrecords. Contoh sampel citra mammogram kanker payudara dapat dilihat pada Tabel 3.1. Data ini akan digunakan untuk melatih dan menguji model deteksi kanker payudara berbasis citra mammogram.

Tabel 3.1 Contoh data mammogram kanker payudara

No	Keterangan	Citra Mammogram
1.	Kanker	
2.	Tidak Kanker	

### 3.3 Desain Sistem

Dalam penelitian ini, sebuah desain sistem yang telah dikembangkan sebelumnya menjadi landasan utama yang membimbing seluruh proses penelitian. Desain ini sangat penting karena menyediakan kerangka kerja yang jelas dan terstruktur yang akan diikuti selama penelitian. Desain ini tidak hanya membantu dalam memvisualisasikan alur kerja tetapi juga dalam mengidentifikasi variabel kunci dan metodologi yang akan digunakan. Selain itu, desain ini memungkinkan peneliti untuk merencanakan langkah-langkah verifikasi dan validasi pada setiap tahap pengembangan sistem. Implementasi dari *Gray Level Co-Occurrence Matrix* (GLCM) dan *Gaussian Naive Bayes* sebagai metode analisis data telah terintegrasi ke dalam desain sistem, yang direpresentasikan secara detail dalam Gambar 3.2.



Gambar 3.2 Desain sistem

Proses deteksi kanker payudara dari citra mammogram dimulai dengan pengambilan citra yang kemudian melewati preprocessing sederhana. Setelah itu, citra diproses untuk mengekstraksi fitur GLCM (*Gray Level Co-occurrence Matrix*) yang digunakan dalam deteksi kanker. Fitur-fitur tersebut kemudian dianalisis menggunakan *Gaussian Naive Bayes* (GNB) untuk mengklasifikasikan keberadaan atau ketiadaan kanker payudara. Hasil deteksi, kemudian ditampilkan sebagai output untuk memberikan informasi tentang kondisi kesehatan payudara.

### 3.3.1 Input Image

Dalam penelitian ini hanya 4000 data citra yang dipilih dari 55890 total data. Pemilihan ini mungkin didasarkan pada batasan penelitian. Dengan memilih 4000 data citra, penelitian dapat lebih terfokus dan efisien dalam analisis serta pengembangan model, sementara pengelolaan data menjadi lebih sederhana dan waktu komputasi dapat dipersingkat. Data citra mammogram kanker payudara memiliki format inputan *tfrecords* yang jika di ekstrak akan menjadi format *.PNG* dengan ukuran 299x299. Total data inputan yang digunakan sebanyak 4000 data citra dengan 2 kelas yaitu tidak kanker dan kanker dengan masing-masing kelas berjumlah 2000.

### 3.3.2 Preprocessing Citra

*Preprocessing* data adalah langkah krusial dalam mempersiapkan data sebelum digunakan untuk melatih model. Proses ini mencakup tahapan untuk memastikan data siap digunakan dengan optimal. Pada *preprocessing* ini dapat diuraikan dalam beberapa tahap sebagai berikut:

a. *Grayscale*

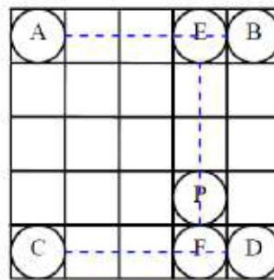
Citra *grayscale* adalah citra yang memiliki rentang warna antara hitam (0) dan putih (255), dengan warna abu-abu di antaranya. Proses *grayscale* adalah konversi citra menjadi warna abu-abu dengan mempertahankan intensitas cahaya dari setiap piksel tanpa memperhitungkan warna. Ini mengurangi dimensi citra dengan mengubah citra berwarna menjadi satu kanal warna abu-abu. Piksel dalam citra *grayscale* memiliki nilai intensitas 8-bit, mulai dari 0 (hitam) hingga 255 (putih). Adapun rumus *grayscale* dapat dilihat pada persamaan 3.1.





$$u(x) = \begin{cases} 0 & |x| > 1 \\ 1 - |x| & |x| < 1 \end{cases} \quad (3.2)$$

Misalkan koordinat piksel A, B, C, dan D adalah  $(i, j)$ ,  $(i, j+1)$ ,  $(i+1, j)$ , dan  $(i+1, j+1)$ . Sedangkan koornidat piksel baru didefinisikan  $(u, v)$ .



Gambar 3.4 Diagram interpolasi *billnear*

Proses ini dilakukan dengan langkah-langkah berikut:

1. Menghitung pengaruh nilai piksel A dan B, anggap sebagai piksel E.

$$f(i, j + v) = [f(i, j + 1) - f(i, j)]v + f(i, j) \quad (3.3)$$

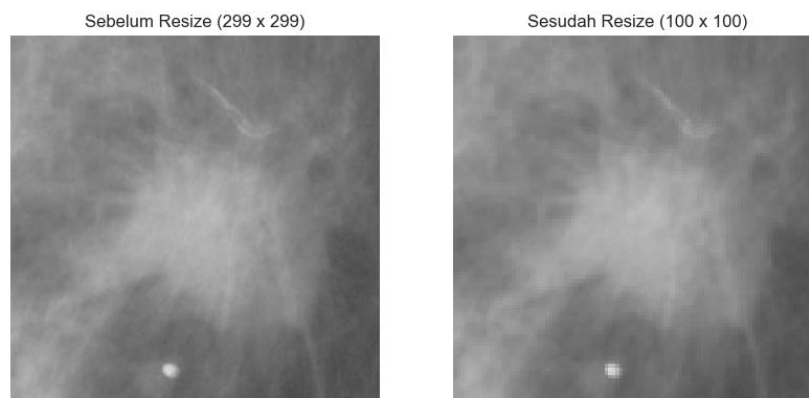
2. Menghitung pengaruh nilai piksel C dan D, anggap sebagai piksel F.

$$\begin{aligned} \text{b. } f(i + 1, j + v) &= [f(i + 1, j + 1) - f(i, j + 1)]v \\ \text{c. } &+ f(i + 1, j) \end{aligned} \quad (3.4)$$

3. Menghitung pengaruh nilai piksel E dan F, anggap sebagai piksel P.

$$\begin{aligned} f(i + u, j + v) &= (1 - u)(1 - v)f(i, j) - (1 - u)vf \\ &(i, j + 1) + u(1 - v)f(i + 1, j) + uvf(i + 1, j + 1) \end{aligned} \quad (3.5)$$

Dalam penelitian ini, perbedaan citra sebelum dan sesudah proses *resize* dapat dilihat pada Gambar 3.4. Gambar ini memperlihatkan bagaimana *resize* mempengaruhi kualitas citra yang kemudian digunakan untuk analisis lebih lanjut dengan teknik GLCM dan *Naive Bayes*.

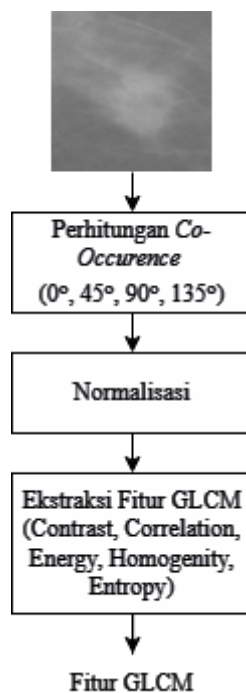


Gambar 3.5 Sebelum dan sesudah *resize*

### 3.3.3 Ekstraksi Fitur Gray Level Co-Occurrence Matrix (GLCM)

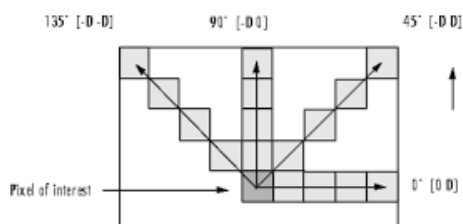
*Gray Level Co-occurrence Matrix* (GLCM) merupakan teknik ekstraksi fitur tekstur yang sering digunakan dalam analisis citra. GLCM mengukur seberapa sering pasangan piksel dengan intensitas keabuan tertentu dan dalam orientasi tertentu muncul dalam suatu citra, sehingga membantu dalam mengidentifikasi tekstur.

Metode ini efektif dalam berbagai aplikasi, termasuk pengenalan citra, segmentasi, pengambilan, dan klasifikasi citra. GLCM menganalisis tekstur dengan menghitung distribusi probabilitas kondisional dari tingkat keabuan yang berdekatan dalam citra, menggunakan korelasi antar piksel untuk menghasilkan fitur yang menjelaskan pola dan struktur internal citra. Penggunaannya yang luas dalam berbagai bidang menunjukkan adaptabilitasnya dalam menganalisis citra dari resolusi beragam, termasuk citra pankromatik resolusi tinggi. Dengan demikian, GLCM tetap menjadi pilihan populer bagi peneliti dalam berbagai studi analisis citra. Proses operasional dari GLCM dapat dilihat pada *flowchart* yang disajikan dalam Gambar 3.6.



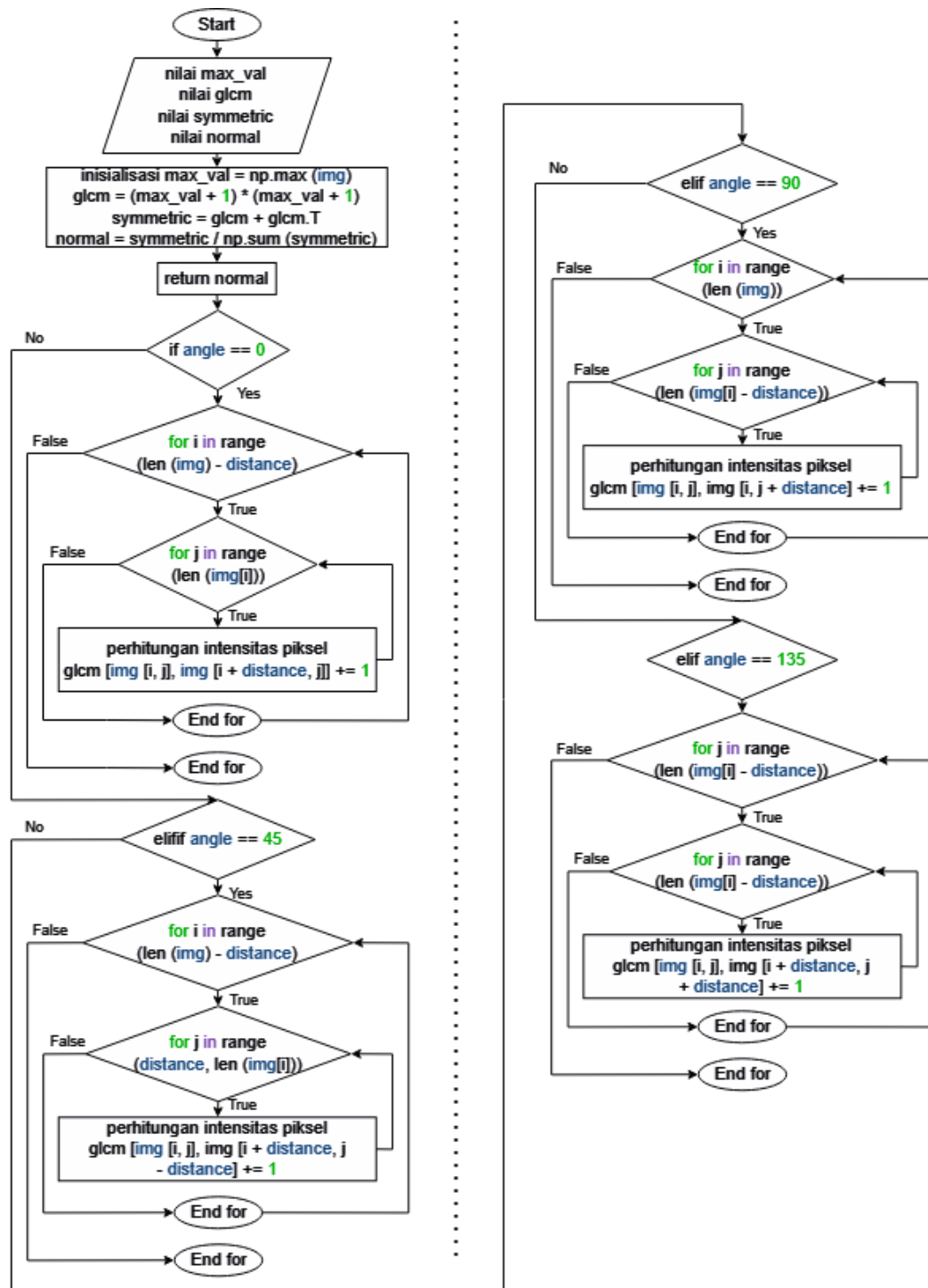
Gambar 3.6 Blok diagram pada GLCM

Berdasarkan blok diagram pada Gambar 3.6, langkah pertama dalam proses ini adalah memuat citra mammogram dari hasil pre-processing yang sudah diresize dan diubah ke dalam format grayscale. Langkah berikutnya adalah menghitung co-occurrence matrix (GLCM), yang menggambarkan frekuensi kemunculan pasangan piksel dengan intensitas tertentu pada jarak dan orientasi tertentu. Jarak antar piksel ( $d$ ) diukur dalam satuan piksel dan biasanya ditetapkan pada 1 piksel untuk standar dalam analisis citra. Orientasi ( $\theta$ ) diukur dalam derajat, dan untuk mendapatkan analisis yang komprehensif, orientasi ini biasanya dibagi menjadi empat arah sudut dengan interval  $45^\circ$ , yaitu  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$ , seperti yang ditunjukkan pada Gambar 3.7. Pemilihan empat orientasi ini memungkinkan pengukuran tekstur dari berbagai arah, memberikan pandangan yang lebih holistik tentang pola tekstur dalam citra mammogram.



Gambar 3. 7 Arah sudut dalam GLCM

Pada *flowchart* yang ditampilkan dalam Gambar 3.8, fungsi yang dikembangkan berfokus pada inisialisasi dan komputasi matriks GLCM (*Co-occurrence*). Proses ini dimulai dengan menginisialisasi matriks GLCM dengan ukuran yang sama dengan nilai maksimum piksel yang ada dalam citra. Langkah selanjutnya adalah mengiterasi melalui setiap piksel dalam citra untuk menghitung frekuensi kemunculan pasangan piksel yang berdekatan yang orientasinya ditentukan oleh sudut yang spesifik. Setelah frekuensi kemunculan pasangan piksel terkumpul dalam matriks GLCM, langkah berikutnya adalah mensimetriskan matriks ini. Hal ini dilakukan dengan menambahkan matriks GLCM dengan transposenya, sehingga matriks menjadi simetris. Proses ini penting untuk memastikan bahwa setiap pasangan piksel  $(i, j)$  memiliki kontribusi yang sama dengan pasangan piksel  $(j, i)$ , sehingga distribusi frekuensi menjadi lebih representatif. Langkah terakhir dalam proses ini adalah normalisasi matriks yang telah disimetriskan. Normalisasi dilakukan dengan membagi setiap elemen matriks dengan jumlah total elemen di dalam matriks, menghasilkan distribusi probabilitas kemunculan pasangan piksel. Normalisasi ini memastikan bahwa nilai-nilai dalam matriks GLCM berada dalam rentang 0 hingga 1, yang memudahkan perhitungan dan interpretasi fitur tekstur.



Gambar 3.8 Flowchart perhitungan *co-occurrence*, *symmetric*, *normalization*

Setelah mendapatkan nilai dari setiap sudut, tahap terakhir dari GLCM adalah *feature extraction* untuk mengekstrak fitur-fitur penting dari citra dengan

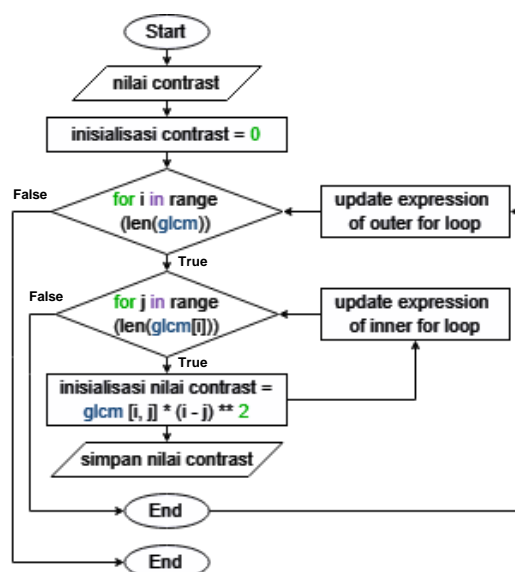
mengelompokkan parameter nilai keunikan dan keakuratan, yang kemudian ditampilkan sebagai hasil dari proses pengelompokan tersebut. Adapun *feature extraction* dalam metode GLCM terdapat lima tahapan dengan nilai  $i$  dan  $j$  yaitu posisi piksel atau koordinat dan  $GLCM(i, j)$  merupakan posisi pada  $(i, j)$ .

a. *Contrast*

*Contrast* mengukur seberapa jauh setiap piksel dalam citra berbeda satu sama lain. Ini dihitung dengan mengkuadratkan perbedaan intensitas antara piksel yang berdekatan dan kemudian mengalikannya dengan frekuensi kemunculan pasangan piksel tersebut dalam GLCM. Korelasi tinggi menunjukkan adanya keterkaitan yang kuat antara piksel yang bermasakan. Persamaan *contrast* dapat dilihat pada persamaan 3.5.

$$Con = \sum_{i,j} GLCM(i, j) \times (i - j)^2 \quad (3.5)$$

Penerapan perhitungan *contrast* dapat dilihat pada *flowchart* Gambar 3.9.



Gambar 3.9 *Flowchart* nilai *contrast*

Gambar tersebut merupakan fungsi  $contrast(glcm)$  yang mengimplementasikan metode perhitungan kontras dalam matriks *Gray Level Co-occurrence Matrix* (GLCM). Fungsi ini menerima matriks GLCM sebagai argumen dan memulai dengan menginisialisasi variabel  $contrast$  ke nilai nol. Fungsi tersebut kemudian menggunakan dua *loop* bersarang *for* untuk mengiterasi melalui setiap elemen matriks. Dalam setiap iterasi, perhitungan  $contrast$  dilakukan dengan menambahkan hasil perkalian antara nilai elemen matriks GLCM dan kuadrat dari selisih indeks baris  $i$  dan kolom  $j$ .

b. *Correlation*

Korelasi menggambarkan seberapa terkait intensitas piksel yang berpasangan dalam citra. Ini diperoleh dari perhitungan statistik yang mengukur hubungan linier antara intensitas piksel dalam citra. Korelasi tinggi menunjukkan adanya keterkaitan yang kuat antara piksel yang berpasangan. Persamaan *correlation* dapat dilihat pada persamaan 3.6.

$$Cor = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) \times GLCM(i, j)}{\sigma_i \times \sigma_j} \quad (3.6)$$

Dimana:

$$\mu_i = \sum_{i,j} i \times GLCM(i, j) \quad (3.7)$$

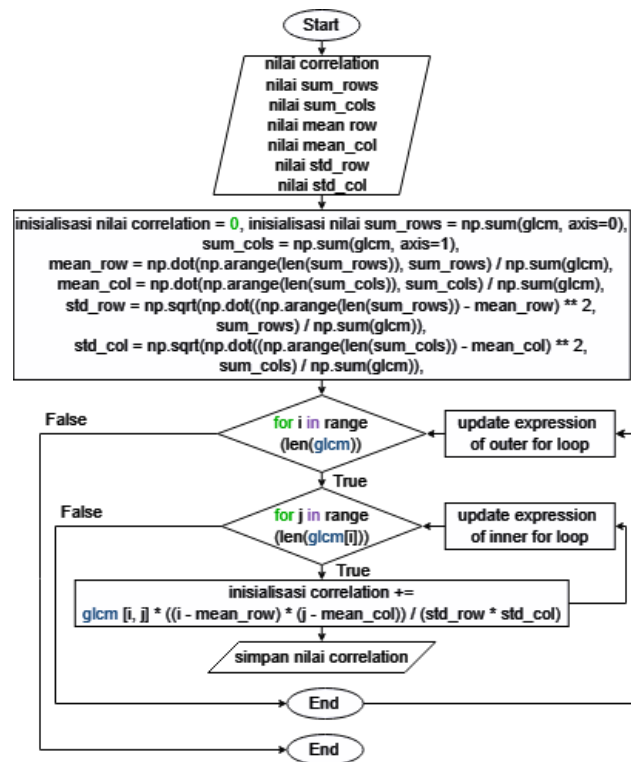
$$\mu_j = \sum_{i,j} j \times GLCM(i, j) \quad (3.8)$$

$$\sigma_i = \sqrt{\sum_{i,j} (i - \mu)^2 \times GLCM(i, j)} \quad (3.9)$$

$$\sigma_j = \sqrt{\sum_{i,j} (j - v)^2 \times GLCM(i, j)} \quad (3.10)$$

Penerapan rumus diatas dapat dilihat pada *flowchart* Gambar 3.10.





Gambar 3.10 Flowchat nilai correlation

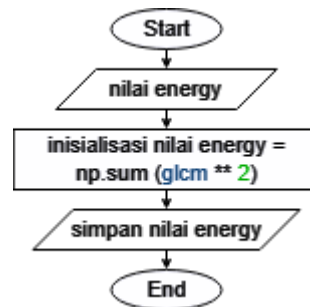
Pertama, fungsi ini menghitung jumlah total dari nilai di setiap baris dan kolom (*sum\_rows* dan *sum\_cols*), lalu menentukan rata-rata dan deviasi standar dari baris dan kolom tersebut. Korelasi dihitung dengan mengiterasi melalui setiap elemen dalam GLCM, dimana untuk setiap pasangan piksel, korelasi diakumulasi berdasarkan formula yang membandingkan nilai piksel dengan rata-rata dan disesuaikan dengan deviasi standar.

### c. *Energy*

Energi pada persamaan 3.11 menggambarkan homogenitas atau kekonsistenan tekstur dalam citra. Ini dihitung dengan menjumlahkan kuadrat dari setiap elemen dalam GLCM. Semakin tinggi nilai energi, semakin homogen tekstur dalam citra.

$$E = \sum_{i,j} GLCM(i,j)^2 \quad (3.11)$$

Rumus perhitungan energy ini dapat dilihat pada Gambar 3.11.



Gambar 3.11 *Flowchart nilai energy*

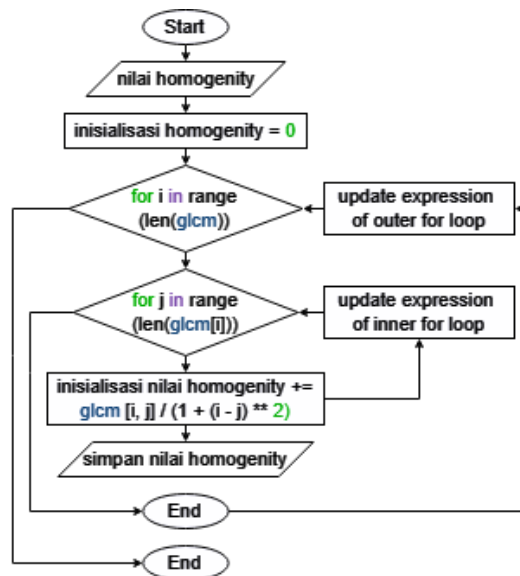
Fungsi  $energy(glcm)$  dimulai dengan mengambil kuadrat dari setiap nilai dalam matriks GLCM ( $glcm ** 2$ ), yang secara efektif menguatkan pengaruh nilai yang lebih tinggi sambil mengurangi pengaruh nilai yang lebih rendah. Kemudian, semua nilai kuadrat tersebut dijumlahkan menggunakan fungsi  $np.sum()$ . Hasil dari penjumlahan ini memberikan skor energi, yang menggambarkan homogenitas tekstur citra. Energi yang tinggi menunjukkan bahwa distribusi intensitas piksel dalam citra lebih seragam.

d. *Homogeneity*

Homogenitas menggambarkan seberapa dekat distribusi probabilitas GLCM dengan diagonal matriks. Ini menunjukkan sejauh mana perbedaan intensitas piksel yang berdekatan cenderung seragam dalam citra. Semakin tinggi nilai homogenitas, semakin seragam tekstur dalam citra sesuai dengan persamaan 3.12.

$$Hom = \sum_{i,j} \frac{1}{1 + (i - j)^2} \times GLCM(i,j) \quad (3.12)$$

Penerapan rumus *homogeneity* dijelaskan dalam *flowchart* pada Gambar 3.12.



Gambar 3.12 *Flowchart* nilai *homogeneity*

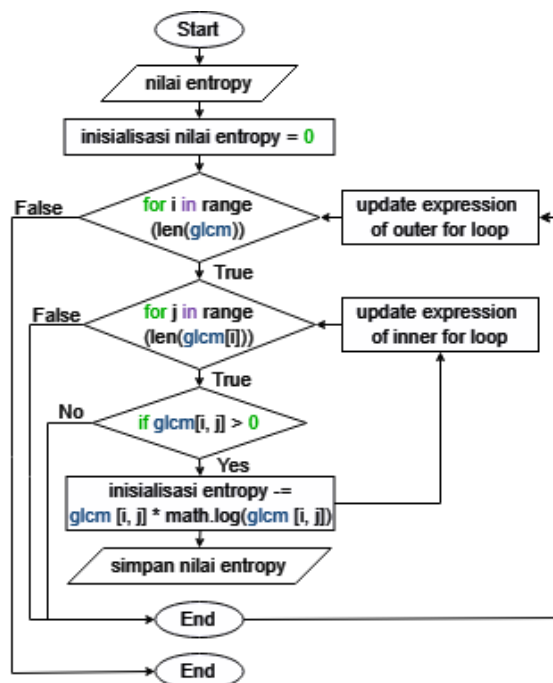
Fungsi *homogeneity(glcm)* pada Gambar 3.12 bertujuan untuk menghitung nilai homogenitas dari matriks *Gray Level Co-occurrence Matrix* (GLCM). Fungsi dimulai dengan menginisialisasi variabel *homogeneity* sebagai nol. Kemudian, menggunakan dua loop bersarang, fungsi ini mengiterasi setiap elemen matriks GLCM. Dalam setiap iterasi, fungsi menambahkan nilai elemen matriks yang dinormalisasi dengan pembilang  $glcm[i, j]$  dan penyebut yang merupakan penambahan satu dengan kuadrat perbedaan indeks baris  $i$  dan kolom  $j$ . Penyebut ini meningkatkan berat kontribusi elemen-elemen yang dekat dengan diagonal GLCM, yang merefleksikan konsistensi tekstur dalam citra. Dengan demikian, semakin besar nilai elemen di sekitar diagonal, semakin homogen tekstur citra tersebut. Akhirnya, fungsi mengembalikan nilai total homogenitas yang dihitung, yang memberikan indikasi seberapa homogen citra berdasarkan distribusi intensitas lokalnya.

e. *Entropy*

Entropi menggambarkan tingkat ketidakpastian atau keacakan dalam citra. Semakin tinggi entropi, semakin tidak teratur atau acak distribusi intensitas piksel dalam citra. Persamaan ini dapat dilihat pada 3.13.

$$En = - \sum_{i,j} GLCM(i,j) \log (GLCM(i,j)) \quad (3.13)$$

Penerapan rumus entropy tertera dalam *flowchart* pada Gambar 3.13



Gambar 3.13 *Flowchart* nilai *entropy*

Pada Gambar 3.13 merupakan perhitungan fungsi *entropy(glcm)* digunakan untuk menghitung entropi yang merupakan ukuran statistik ketidakpastian atau keacakan dalam tekstur citra. Dalam implementasi fungsi ini, entropi diinisialisasi menjadi nol, dan kemudian dihitung dengan mengiterasi melalui setiap elemen dalam GLCM. Untuk setiap elemen, jika nilai elemen tersebut lebih besar dari nol,

maka nilai entropi akan diperbarui dengan menambahkan produk dari nilai elemen tersebut dan logaritma natural dari nilai elemen tersebut yang telah dinegasikan.

Adapun hasil dari ekstraksi fitur *Gray Level Co-occurrence Matrix* (GLCM) yang telah dilakukan dalam penelitian ini disajikan secara lengkap pada Lampiran I. Lampiran tersebut mencakup berbagai fitur tekstur yang telah dihitung dari citra mammogram, seperti *contrast*, *correlation*, *energy*, *homogeneity*, dan *entropy*.

### 3.3.4 Uji Normalitas

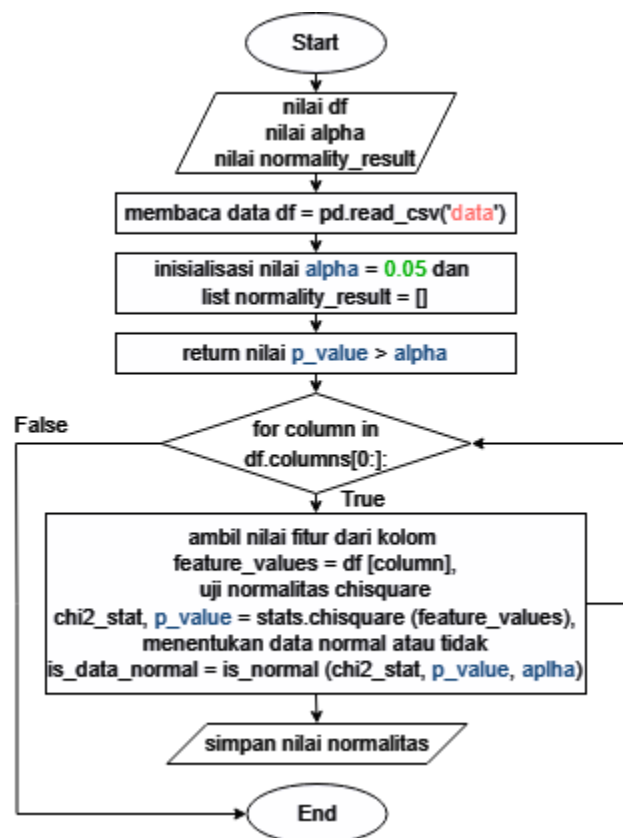
Uji normalitas adalah langkah penting untuk memastikan bahwa nilai residual terdistribusi secara merata dan teratur. Data yang mengikuti distribusi normal memiliki kecenderungan untuk menghasilkan nilai residual yang lebih konsisten dan tidak bias. Dengan demikian, pengujian normalitas merupakan upaya untuk meminimalkan potensi bias dalam analisis data.

Pengujian yang dilakukan disini menggunakan uji normalitas *Chi-square* merupakan statistik non-parametrik, disebut juga uji bebas distribusi merupakan salah satu metode statistik yang digunakan untuk menguji apakah suatu sampel data berasal dari distribusi normal atau tidak. Distribusi normal atau *Gaussian* adalah distribusi probabilitas yang simetris dan berbentuk lonceng.

Proses uji normalitas *Chi-square* melibatkan perbandingan antara distribusi frekuensi dari data yang diamati dengan distribusi yang diharapkan (dalam kasus ini, distribusi normal). Adapun rumus pengujian menggunakan *Chi-square* terdapat pada persamaan 3.14.

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3.14)$$

Rumus *Chi-square* digunakan untuk menghitung nilai statistik *Chi-square* ( $\chi^2$ ) berdasarkan perbandingan antara frekuensi observasi yang diamati dalam sampel dengan frekuensi yang diharapkan dalam situasi di mana tidak ada hubungan antara variabel-variabel tersebut. Dalam rumus tersebut di mana  $O$  adalah frekuensi observasi, dan  $E$  adalah frekuensi yang diharapkan jika tidak ada hubungan antara variabel. Nilai  $\chi^2$  yang dihasilkan kemudian dibandingkan dengan nilai kritis dari distribusi *Chi-square* untuk menentukan apakah perbedaan antara frekuensi observasi dan yang diharapkan signifikan secara statistik. Jika nilai  $\chi^2$  lebih besar dari nilai kritis, maka terdapat bukti yang cukup untuk menolak hipotesis nol dan menyimpulkan bahwa terdapat hubungan antara variabel. Perhitungan uji normalitas terdapat pada flowchat Gambar 3.14.



Gambar 3.14 Flowchart uji normalitas *Chi-Square*

Pertama dilakukan inisialisasi sebuah fungsi *is\_normal* untuk menentukan apakah data dapat dianggap normal berdasarkan statistik uji *Chi-square*. Setelahnya, sebuah list inisialisasi digunakan untuk menyimpan hasil uji normalitas. Tingkat signifikansi *alpha* untuk uji *Chi-square* diatur sebesar 0.05. Dilakukan perulangan melalui setiap kolom fitur GLCM di mana nilai fitur diekstraksi dan disimpan. Uji normalitas *Chi-square* dilakukan untuk setiap kolom menggunakan fungsi *stats.chisquare*. Hasil uji, termasuk nilai statistik *Chi-square*, *p-value*, dan penentuan apakah data dianggap normal atau tidak, disimpan dalam list. Proses ini memungkinkan untuk mengevaluasi apakah data fitur GLCM tersebut memiliki distribusi yang mendekati normal atau tidak, berdasarkan uji *Chi-square* yang dilakukan.

### 3.3.5 Klasifikasi *Naïve Bayes*

Klasifikasi *Naïve Bayes* adalah sebuah metode klasifikasi yang berbasis probabilitas, dengan asumsi bahwa setiap atribut atau fitur objek bersifat independen satu sama. Saat ini, *Naïve Bayes* tetap menjadi topik yang menarik, terutama dalam upaya meningkatkan performansi klasifikasi. Dalam banyak kasus, diasumsikan bahwa fungsi kepadatan peluang (*probability density function*) dari setiap atribut atau fitur berdistribusi secara normal (*gaussian*). Asumsi ini memiliki dampak besar dalam pembentukan klasifikasi *Naïve Bayes*. Ketika fitur-fitur objek,  $X = \{x_1=\textit{contrast}, x_2=\textit{correlation}, x_3=\textit{energy}, x_4=\textit{homogeneity}$  dan  $x_5=\textit{entropy}\}$  diketahui, dan setiap fitur diasumsikan berdistribusi normal (*gaussian*), maka peluang fitur atau atribut dengan memperhatikan kelas ke-*j* ( $C_j$ ) dapat dinyatakan sebagai (Santoso & Supriyono, 2020) pada persamaan 3.15.

$$P(X|C_j) = \prod_{k=1}^d P(X_k|C_j) = \prod_{k=1}^d N(x_k; \hat{\mu}_{jk}, \hat{\sigma}_{jk}) \quad (3.15)$$

Dengan menggunakan estimasi parameter distribusi (normal)  $\mu$  dan  $\sigma$  untuk fitur ke-k dan kelas k-j, dapat menerapkan kaidah peluang bersyarat untuk mendapatkan persamaan 3.16.

$$P(C_j, X) = P(C_j)P(X|C_j) = P(C_j) \prod_{k=1}^d N(x_k; \hat{\mu}_{jk}, \hat{\sigma}_{jk}) \quad (3.16)$$

Proses klasifikasi menggunakan metode *Gaussian Naïve Bayes* dimulai dengan memasukkan nilai pada setiap fitur yang digunakan. Selanjutnya, nilai  $(\hat{\mu}_{jk})$  dan  $(\hat{\sigma}_{jk})$  dihitung untuk setiap fitur dalam setiap kelas. Kemudian, dilakukan perhitungan *Gaussian* menggunakan rumus distribusi *Gaussian* untuk setiap fitur dalam setiap kelas. Perhitungan posterior dilakukan dengan mengalikan nilai *prior* (probabilitas kelas) dengan nilai hasil perhitungan *Gaussian*.

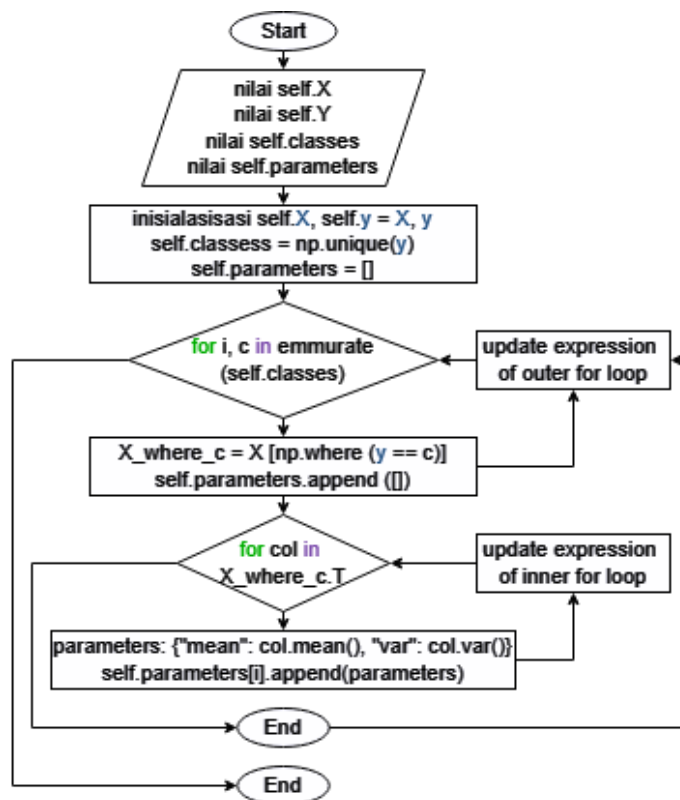
Probabilitas kelas diasumsikan sama untuk setiap kelas dalam penelitian ini. Hasil perhitungan nilai posterior kemudian digunakan untuk mencari argumen maksimum, menentukan kelas di mana data tersebut kemungkinan besar dengan menggunakan rumus pada persamaan 3.17.

$$\hat{c} = \arg \max P(C_j) \prod_{k=1}^d N(x_k; \hat{\mu}_{jk}, \hat{\sigma}_{jk}) \quad (3.17)$$

Proses perhitungan pertama dalam *gaussian naive bayes* dapat dilihat di *flowchart* pada Gambar 3.15. Dalam *flowchart* tersebut fungsi *fit* bertugas untuk melatih model klasifikasi *Naive Bayes* dengan dataset yang diberikan. Pertama,



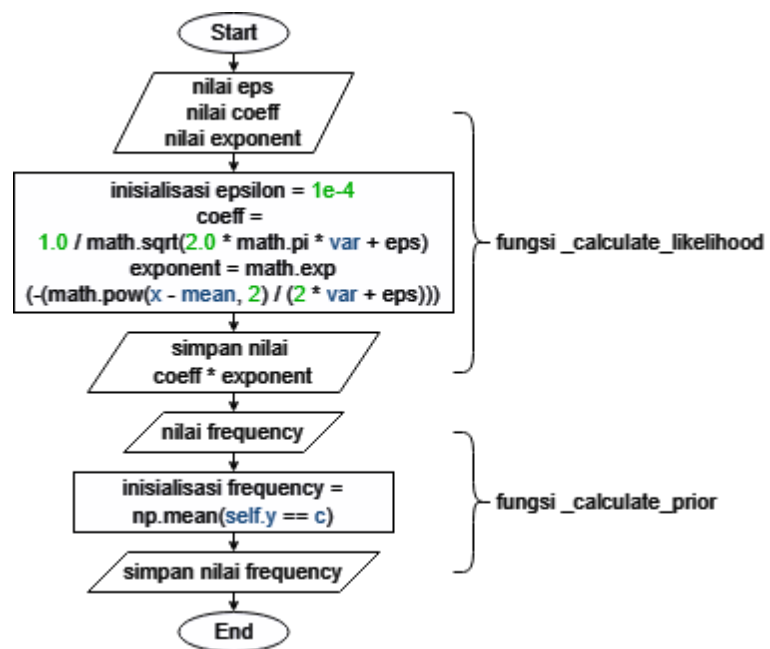
metode ini menyimpan fitur dan label input ( $X$  dan  $y$ ) ke dalam variabel instansi  $self.X$  dan  $self.y$ . Kemudian menentukan kelas unik yang ada dalam dataset menggunakan  $np.unique(y)$  dan menyimpannya ke dalam  $self.classes$ . Untuk setiap kelas, fungsi ini mengekstrak subset data yang hanya termasuk sampel dari kelas tersebut ( $X\_where\_c$ ) dan menghitung parameter statistik seperti rata-rata ( $mean$ ) dan varians ( $var$ ) untuk setiap fitur. Parameter ini kemudian disimpan dalam daftar  $self.parameters$  untuk digunakan dalam proses klasifikasi.



Gambar 3.15 Flowchart fungsi *fit*

Setelah mendapatkan parameter, kemudian memuat fungsi *\_calculate\_likelihood* pada Gambar 3.16 untuk menghitung probabilitas kepadatan suatu nilai fitur tertentu, diberikan parameter distribusi normal (rata-rata dan varians) dengan rumus  $\frac{1}{\sqrt{2\pi\sigma^2+\epsilon}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2+\epsilon}\right)$ , di mana  $\epsilon$  adalah konstanta kecil

untuk menghindari pembagian dengan nol,  $\mu$  adalah rata-rata, dan  $\sigma^2$  adalah varians. Fungsi ini mengembalikan nilai probabilitas untuk sebuah observasi  $x$  dari kelas  $c$ , dengan menggunakan parameter yang sudah dihitung sebelumnya.

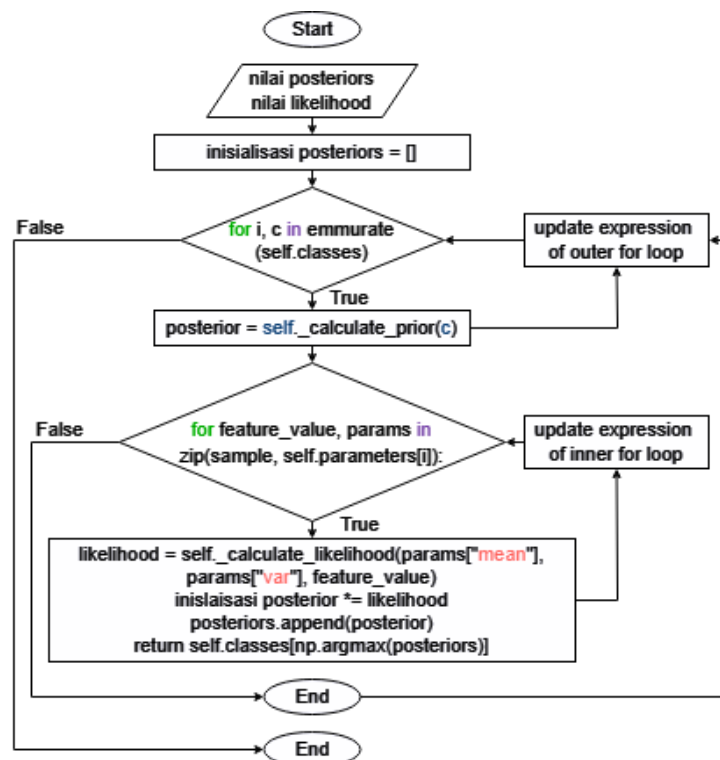


Gambar 3.16 Flowchart fungsi prior dan likelihood

Selanjutnya yaitu fungsi `_calculate_prior` untuk menghitung probabilitas prior dari kelas  $C_j$  yang merupakan frekuensi kemunculan kelas  $C_j$  dalam dataset pelatihan dengan rumus  $\frac{\text{Jumlah sampel di kelas } C_j}{\text{Total jumlah sampel}}$ . Fungsi ini mengembalikan nilai probabilitas prior yang akan dikombinasikan dengan likelihood untuk membentuk probabilitas posterior dalam proses klasifikasi *Naive Bayes*.

Setelah menghitung prior dan *likelihood* terdapat *flowchart* untuk klasifikasi dan prediksi pada Gambar 3.17. Fungsi `_classify` diimplementasikan untuk menentukan kelas dari setiap sampel berdasarkan nilai probabilitas posterior yang paling tinggi. Proses ini dimulai dengan inisialisasi daftar kosong *posteriors* yang akan menampung probabilitas posterior untuk setiap kelas. Untuk setiap kelas

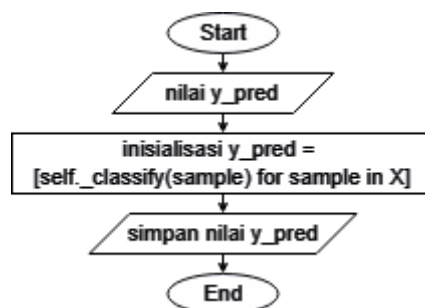
dalam dataset, fungsi ini terlebih dahulu menghitung probabilitas prior kelas tersebut menggunakan metode `_calculate_prior`. Kemudian, untuk setiap nilai fitur dalam sampel yang sedang diklasifikasi, fungsi menghitung likelihood menggunakan metode `_calculate_likelihood` dengan parameter rata-rata dan varians yang sesuai dari kelas tersebut. Probabilitas posterior untuk kelas dihitung dengan mengalikan nilai prior dengan semua likelihood dari fitur-fitur tersebut.



Gambar 3.17 Flowchart fungsi `classify`

Selanjutnya, nilai probabilitas posterior yang dihitung untuk setiap kelas disimpan dalam daftar `posteriors`. Setelah semua kelas dihitung, fungsi memilih kelas dengan probabilitas posterior tertinggi sebagai prediksi kelas untuk sampel tersebut. Ini dilakukan dengan fungsi `np.argmax` yang mengidentifikasi indeks dari

nilai maksimum dalam daftar probabilitas posterior, dan mengembalikan kelas yang sesuai.



Gambar 3. 18 Flowchart fungsi  $y\_pred$

Kemudian fungsi *predict* pada Gambar 3.18 menggunakan fungsi *\_classify* untuk menerapkan proses klasifikasi ini secara berurutan ke setiap sampel dalam dataset yang diberikan. Metode ini menghasilkan array  $y\_pred$  yang berisi prediksi kelas untuk setiap sampel di  $X$ , memanfaatkan list comprehension untuk memanggil *\_classify* untuk setiap sampel.

### 3.4 Skenario Uji Coba

Skenario uji coba akan memfokuskan pada validasi fungsionalitas sistem yang diusulkan. Untuk skenario uji coba dengan ukuran citra mammogram 100x100 digunakan data seimbang yang terdiri dari 2000 sampel kanker dan 2000 sampel tidak kanker. Pada skenario ini akan dilakukan pengambilan jarak atau *distance* terbagus dengan menggunakan *Gaussian Naive Bayes* dengan total pengujian jarak yaitu 3 ( $distance = 1$ ,  $distance = 2$ ,  $distance = 3$ ) (Praseptiyana *et al.* 2019). Setelah diperoleh jarak terbaik selanjutnya dilakukan evaluasi sistem pada empat skenario berbeda, yang menggunakan 1, 2, 3, dan 4 sudut GLCM untuk ekstraksi fitur tekstur. Pada setiap skenario, sistem dilatih dan diuji dengan tiga rasio pembagian

data: 90:10, 80:20, dan 70:30 untuk memperoleh sistem terbaik dengan jumlah variabel atau data masuk yang akan diuji. Pembagian skenario kombinasi sudut terdapat pada tabel 3.2.

Tabel 3.2 Skenario uji coba kombinasi sudut

Skenario	Sudut	Jumlah Variabel
1 sudut	(0°), (45°), (90°), (135°)	5
2 sudut	(0°, 45°), (0°, 90°), (0°, 135°), (45°, 90°), (45°, 135°), (90°, 135°)	10
3 sudut	(0°, 45°, 90°), (0°, 45°, 135°), (45°, 90°, 135°), (0°, 90°, 135°)	15
4 sudut	(0°, 45°, 90°, 135°)	20

### 3.5 Evaluasi

*Confusion matrix* adalah metode pengujian yang digunakan untuk menghitung nilai akurasi berdasarkan perhitungan kumulatif dari jumlah klasifikasi yang benar dibagi dengan jumlah klasifikasi yang benar dan salah. *Confusion matrix* berperan sebagai alat untuk menganalisis keefektifan suatu pengklasifikasi dalam mengenali daftar kelas yang berbeda. Nilai *True-Positive* dan *True-Negative* menunjukkan bahwa pengklasifikasi melakukan klasifikasi dengan benar, sedangkan nilai *False-Positive* dan *False-Negative* menandakan bahwa pengklasifikasi melakukan klasifikasi yang salah (Indriani, n.d.).

Pada kasus ini *True Positive* (TP) mengacu pada citra mammogram yang menunjukkan keberadaan kanker payudara benar-benar terdeteksi dengan benar oleh algoritma deteksi. *True Negative* (TN) mengacu pada citra mammogram yang menunjukkan tidak adanya kanker payudara dan berhasil diidentifikasi dengan benar oleh algoritma deteksi. *False Positive* (FP) mengacu pada citra mammogram yang sebenarnya tidak menunjukkan keberadaan kanker payudara, tetapi secara keliru diidentifikasi sebagai positif oleh algoritma. Sedangkan *False Negative* (FN)

mengacu pada citra mammogram yang seharusnya menunjukkan keberadaan kanker payudara, tetapi salah diklasifikasikan sebagai *negative* oleh algoritma. Dalam konteks ini, TP dan TN menunjukkan hasil yang benar, sementara FP dan FN menunjukkan kesalahan identifikasi oleh algoritma deteksi.

*Confusion matrix* memungkinkan penghitungan metrik kinerja, termasuk *accuracy*, *precision*, *recall* dan *f1-score*. Akurasi mencerminkan tingkat ketepatan model dalam mengklasifikasikan data, presisi menunjukkan ketepatan antara data yang diminta dan hasil prediksi yang diberikan, sementara *recall* menggambarkan keberhasilan model dalam mengambil kembali informasi.

#### 1. *Accuracy*

*Accuracy* mengukur seberapa sering model klasifikasi benar-benar memprediksi kelas data dengan benar dari keseluruhan data yang diprediksi. *Accuracy* berguna untuk mengevaluasi kinerja model secara keseluruhan. Namun, pada kasus dengan kelas yang tidak seimbang, *accuracy* mungkin tidak memberikan gambaran yang akurat tentang kinerja model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.17)$$

#### 2. *Precision*

*Precision* mengukur seberapa tepat model dalam memprediksi kelas positif dari semua prediksi yang diklasifikasikan sebagai positif. *Precision* penting ketika konsekuensi dari false positives tinggi, seperti dalam deteksi penyakit, karena kita ingin meminimalkan false positives.

$$precision = \frac{TP}{TP + FP} \quad (3.18)$$

### 3. *Recall*

*Recall* (juga dikenal sebagai *Sensitivity* atau *True Positive Rate*) mengukur seberapa banyak dari kelas positif yang terprediksi dengan benar dari keseluruhan kelas positif yang sebenarnya. *Recall* penting ketika konsekuensi dari false negatives tinggi, seperti dalam deteksi penyakit yang serius, karena kita ingin meminimalkan false negatives.

$$recall = \frac{TP}{TP + FP} \quad (3.19)$$

### 4. *F1-score*

*F1-score* adalah ukuran gabungan dari *precision* dan *recall*. Ini adalah rata-rata harmonik dari *precision* dan *recall*. *F1-score* berguna ketika ada *trade-off* antara *precision* dan *recall* yang signifikan, karena memberikan gambaran yang seimbang tentang kinerja model.

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (3.20)$$

## **BAB IV**

### **UJI COBA DAN PEMBAHASAN**

Dalam bab ini akan menjelaskan tentang uji coba dan pembahasan Deteksi Kanker Payudara Berbasis Citra Mammogram Menggunakan *Gray Level Co-Occurrence Matrix* dan *Gaussian Naïve Bayes*.

#### **4.1 Langkah-langkah Uji Coba**

##### **4.1.1 Data Pengujian**

Pada langkah pertama pengujian, persiapan data uji dilakukan dengan mempersiapkan dua kelas, masing-masing terdiri dari 2000 data citra mammogram. Dengan total 4000 data citra mammogram yang digunakan, langkah berikutnya melibatkan proses perhitungan nilai matriks dari berbagai sudut dan ekstraksi fitur GLCM dari dataset tersebut. Ekstraksi fitur GLCM adalah tahap penting dalam analisis citra untuk menggambarkan tekstur citra secara numerik yang terbagi menjadi 5 ekstraksi fitur. Pengujian ini data akan dibagi menjadi data latih dan data uji menggunakan beberapa rasio pembagian, yaitu 70:30, 80:20, 90:10.

Proses pembagian data latih dan data uji dilakukan dengan menginisialisasi variabel rasio yang mengacu pada proporsi data yang akan digunakan untuk pelatihan dan pengujian. Dalam penelitian ini menggunakan total 4000 data citra, dibagi menjadi dua kategori yaitu kanker dan tidak kanker dengan masing-masing kategori memiliki 2000 data. Terdapat tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Pada rasio 70:30, 70% (2800 data) digunakan untuk pelatihan (1400 kanker, 1400 tidak kanker) dan 30% (1200 data) untuk pengujian (600



kanker, 600 tidak kanker). Untuk rasio 80:20, 80% dari data (3200 data) dijadikan data pelatihan (1600 kanker, 1600 tidak kanker) dan 20% (800 data) sebagai data pengujian (400 kanker, 400 tidak kanker). Pada rasio 90:10, 90% (3600 data) digunakan untuk pelatihan (1800 kanker, 1800 tidak kanker) dan 10% (400 data) untuk pengujian (200 kanker, 200 tidak kanker). Ketiga skenario pembagian ini dirancang untuk menjaga proporsi yang seimbang antara kedua kelas, memungkinkan evaluasi yang adil dan efektif terhadap performa model yang dikembangkan.

Selanjutnya, nilai *random state* diinisialisasi dengan nilai 0.42. Penggunaan nilai *random state* ini bertujuan untuk mengacak urutan pembagian data pada dataset. Dengan demikian, pembagian data akan konsisten setiap kali program dijalankan kembali. Hal ini penting untuk menjaga konsistensi dalam hasil evaluasi model dan memastikan bahwa pembagian data tidak dipengaruhi oleh faktor kebetulan. Tabel 4.1 menunjukkan detail pembagian data latih dan data uji untuk ketiga variasi rasio pembagian yang digunakan dalam pengujian.

Tabel 4. 1 Rasio pembagian data

Rasio	Seluruh Data		Per kelas	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
70:30	2800	1200	1400	600
80:20	3200	800	1600	400
90:10	3600	400	1800	200

Pada tahap selanjutnya, data latih akan digunakan untuk melatih model klasifikasi, dan data uji akan digunakan untuk mengevaluasi performa model. Berbagai metrik klasifikasi akan dihitung untuk menilai performa model dan menentukan rasio pembagian data terbaik dan model klasifikasi yang paling optimal.

#### 4.1.2 Uji Normalitas

Setelah ekstraksi fitur dilakukan, data tersebut akan memasuki uji normalitas sebelum penggunaan *Gaussian Naive Bayes* (GNB) untuk memverifikasi apakah fitur-fitur yang digunakan oleh model memiliki distribusi normal atau tidak. Hal ini penting karena GNB mengasumsikan bahwa fitur-fitur tersebut memiliki distribusi normal. Jika fitur-fitur tidak terdistribusi secara normal, maka asumsi yang mendasari model GNB menjadi tidak valid, dan performa model dapat terpengaruh. Hasil dari uji normalitas dapat dilihat pada Gambar 4.1

Hasil Uji Normalitas Chi-square:				
	Feature	Chi-square Statistic	p-value	Is Normal
0	contrast_0	281365.517366	0.0	False
1	correlation_0	43.229908	1.0	True
2	energy_0	640.758011	1.0	True
3	homogeneity_0	171.273446	1.0	True
4	entropy_0	856.456398	1.0	True
5	contrast_45	472224.703167	0.0	False
6	correlation_45	49.479461	1.0	True
7	energy_45	645.704496	1.0	True
8	homogeneity_45	211.190470	1.0	True
9	entropy_45	905.312867	1.0	True
10	contrast_90	273464.050746	0.0	False
11	correlation_90	44.812036	1.0	True
12	energy_90	640.905698	1.0	True
13	homogeneity_90	173.055360	1.0	True
14	entropy_90	858.062462	1.0	True
15	contrast_135	500577.146363	0.0	False
16	correlation_135	48.864260	1.0	True
17	energy_135	645.034553	1.0	True
18	homogeneity_135	210.077438	1.0	True
19	entropy_135	904.360639	1.0	True

Gambar 4.1 Hasil uji normalitas

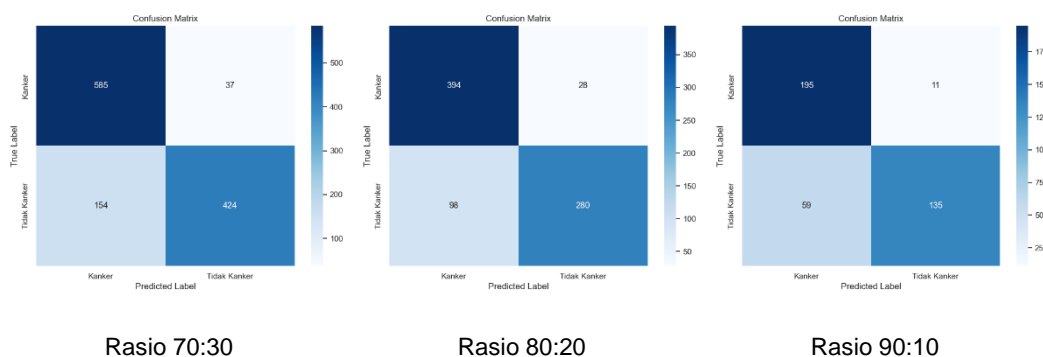
Dari Gambar 4.1 merupakan hasil uji normalitas *Chi-square* pada fitur-fitur GLCM menunjukkan bahwa beberapa fitur memiliki distribusi yang tidak normal, sedangkan fitur lainnya memiliki distribusi normal. Fitur-fitur yang memiliki distribusi tidak normal adalah *contrast\_0*, *contrast\_45*, *contrast\_90*, dan

*contrast\_135* dengan *p-value* yang signifikan ( $p < 0.05$ ), sehingga menolak hipotesis nol bahwa distribusi fitur-fitur tersebut normal. Sedangkan fitur-fitur lainnya seperti *correlation*, *energy*, *homogeneity*, dan *entropy* pada semua sudut (0, 45, 90, 135) memiliki *p-value* lebih dari 0.05, yang menunjukkan bahwa distribusi fitur-fitur tersebut tidak berbeda secara signifikan dari distribusi normal.

## 4.2 Hasil Uji Coba Menggunakan Jarak

Data yang telah melalui uji normalitas selanjutnya akan digunakan untuk menentukan jarak terbaik dari tiga jarak yang ada (1, 2, dan 3). Pengujian jarak ini juga melibatkan *Gaussian Naive Bayes* (GNB) untuk mengevaluasi kinerja modelnya. Hasil akurasi terbaik dari berbagai rasio pembagian data akan digunakan untuk pengujian lebih lanjut.

Pengujian pertama dilakukan dengan menggunakan jarak 1. Hasil dari confusion matrix untuk pengujian ini dapat dilihat pada Gambar 4.2. *Confusion matrix* ini memberikan gambaran tentang jumlah prediksi yang benar dan salah, membantu dalam mengevaluasi performa model dalam mengklasifikasikan data dengan jarak tertentu.



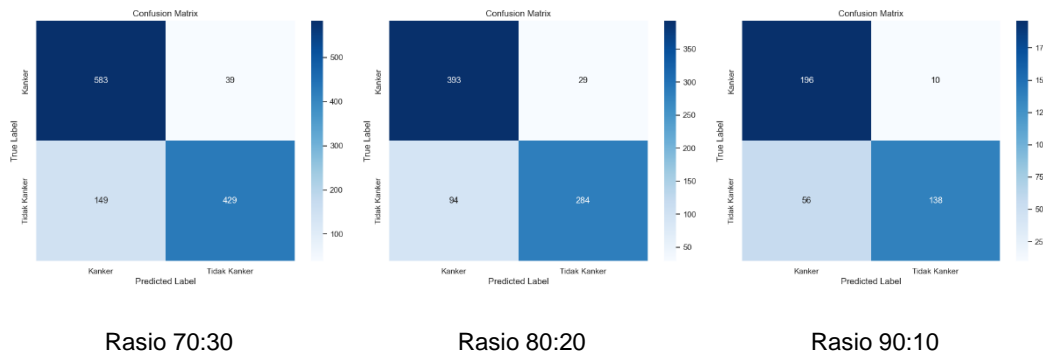
Gambar 4.2 *Confusion matrix* jarak 1

Berdasarkan Gambar 4.2, terdapat Tabel 4.2 yang memuat nilai *precision* yang menunjukkan proporsi prediksi positif yang benar, *recall* yang menunjukkan proporsi kejadian sebenarnya yang terdeteksi oleh model, dan *f1-score* memberikan keseimbangan antara *precision* dan *recall*.

Tabel 4.2 Nilai *precision*, *recall*, dan *f1-score* pada jarak 1

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	79,16	94,05	85,97
	Tidak Kanker	91,97	73,36	81,62
80:20	Kanker	80,08	93,36	86,21
	Tidak Kanker	90,9	74,07	81,63
90:10	Kanker	76,77	94,66	84,78
	Tidak Kanker	92,47	69,59	79,41

Selanjutnya yaitu pengujian dengan menggunakan jarak 2 yang menghasilkan *confusion matrix* pada Gambar 4.3.



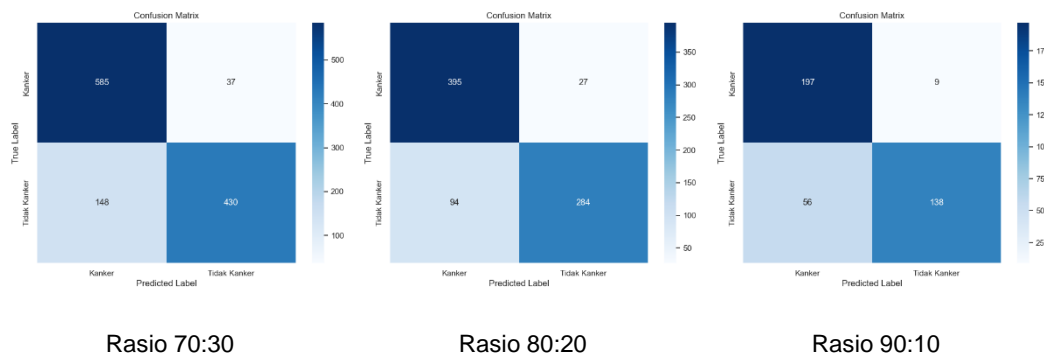
Gambar 4.3 *Confusion matrix* jarak 2

Berdasarkan Gambar 4.3 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.3.

Tabel 4.3 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	79,64	93,73	86,12
	Tidak Kanker	91,67	74,22	82,03
80:20	Kanker	80,69	93,13	86,47
	Tidak Kanker	90,73	75,13	82,19
90:10	Kanker	77,78	95,15	85,59
	Tidak Kanker	93,24	71,13	80,7

Dan yang terakhir yaitu pengujian pada jarak 3 dengan pembagian tiga rasio data dapat dilihat pada Gambar 4.4.



Gambar 4.4 *Confussion matrix* jarak 3

Berdasarkan Gambar 4.4 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.4.

Tabel 4.4 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	79,81	94,05	86,35
	Tidak Kanker	92,08	74,39	82,29
80:20	Kanker	80,78	93,6	86,72
	Tidak Kanker	91,32	75,13	82,43
90:10	Kanker	77,87	95,63	85,84
	Tidak Kanker	93,88	71,13	80,94

Hasil uji coba menggunakan ketiga jarak yang berbeda dengan tiga variasi rasio pembagian data dapat dilihat pada Tabel 4.5.

Tabel 4.5 Hasil *accuracy*, *precision*, *recall*, dan *f1-score* pada uji coba menggunakan jarak

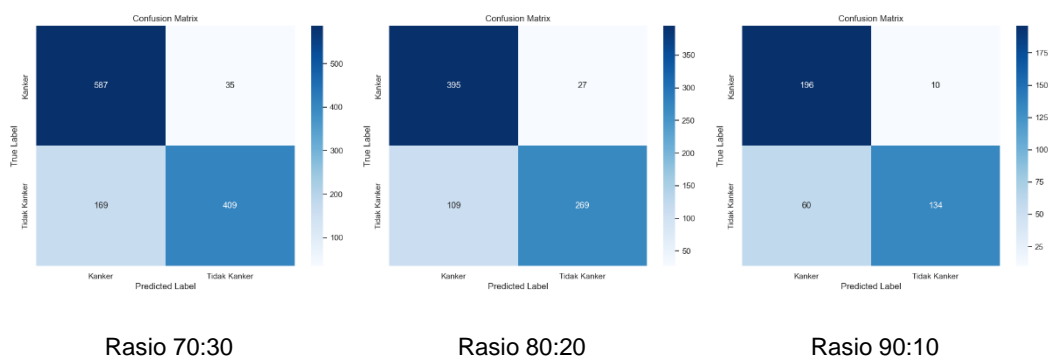
Jarak	Rasio	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
1	70:30	84,08	85,57	83,7	83,79
	80:20	84,25	85,49	83,72	83,92
	90:10	82,5	84,62	82,12	82,09
2	70:30	84,33	85,66	83,98	84,07
	80:20	84,62	85,71	84,13	84,33
	90:10	83,5	85,51	83,14	83,15
3	70:30	84,58	85,94	84,22	84,32
	80:20	84,88	86,05	84,37	84,58
	90:10	83,75	85,87	83,38	83,38

Tabel 4.5 menunjukkan hasil pengujian untuk tiga jarak yang berbeda (1, 2, dan 3) dengan variasi rasio pembagian data uji dan pelatihan (70:30, 80:20, dan 90:10). Dari hasil tersebut, terlihat bahwa pada jarak 3, untuk semua rasio pembagian data uji dan pelatihan, memiliki nilai *accuracy*, *precision*, *recall*, dan *f1-score* yang terbaik yaitu nilai akurasi sebesar 84,88 di pembagian data 80:20. Hal ini menunjukkan bahwa pada jarak 3 memiliki kinerja yang lebih baik dalam mengklasifikasikan data. Sehingga pengujian selanjutnya akan menggunakan data dengan jarak 3.

### 4.3 Hasil Uji Coba Kombinasi 1 Sudut

Hasil uji coba ini memiliki scenario terletak pada perbedaan sudut yang digunakan, dimana pada percobaan ini menggunakan satu sudut saja yang berbeda. Pada pengujian ini dataset dibagi menjadi data *training* dan data *testing* dengan perbandingan tiga rasio yaitu 70:30, 80:20, 90:10 dengan jumlah data *training* yang lebih besar dari data *testing*.

Pada percobaan pertama menggunakan sudut  $0^\circ$  dengan perbandingan tiga rasio dan klasifikasi menggunakan GNB ditampilkan dalam *confusion matrix* pada Gambar 4.5.



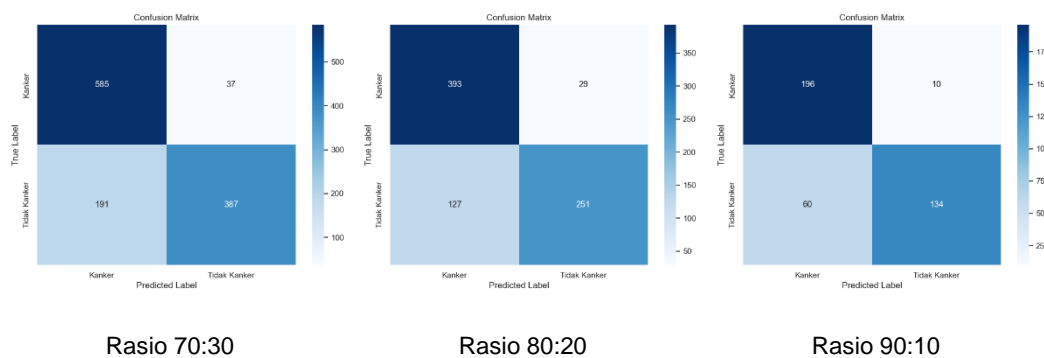
Gambar 4.5 *Confusion matrix* pada derajat 0

Berdasarkan Gambar 4.5 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.6.

Tabel 4.6 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	77,64	94,37	80,04
	Tidak Kanker	92,18	70,76	80,04
80:20	Kanker	78,37	93,61	79,82
	Tidak Kanker	90,87	71,16	79,82
90:10	Kanker	76,56	95,15	84,85
	Tidak Kanker	93,05	69,07	79,29

Pada percobaan kedua menggunakan sudut 45° dengan perbandingan tiga rasio dan klasifikasi menggunakan GNB ditampilkan pada *confusion matrix* Gambar 4.6.



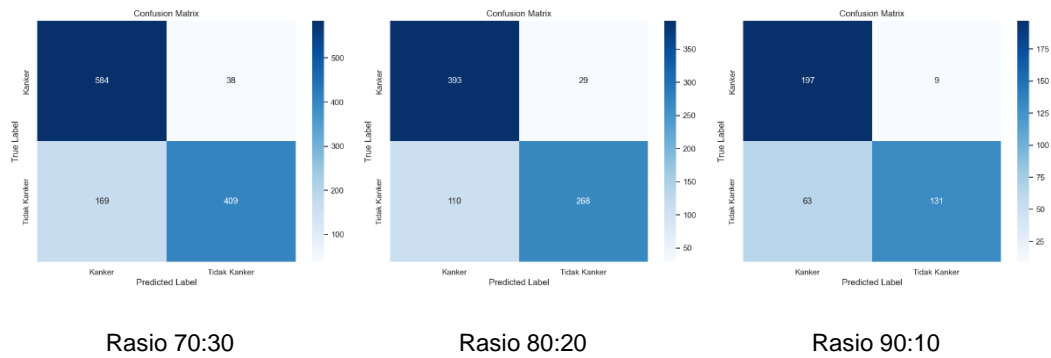
Gambar 4.6 *Confusion matrix* pada derajat 45

Berdasarkan Gambar 4.6 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.7.

Tabel 4.7 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	75,39	94,05	83,69
	Tidak Kanker	91,27	66,95	77,25
80:20	Kanker	75,58	93,13	83,44
	Tidak Kanker	89,64	66,4	76,29
90:10	Kanker	76,56	95,15	84,85
	Tidak Kanker	93,05	69,07	79,29

Selanjutnya pada percobaan ketiga menggunakan sudut  $90^\circ$  hasilnya ditampilkan pada Gambar 4.7.



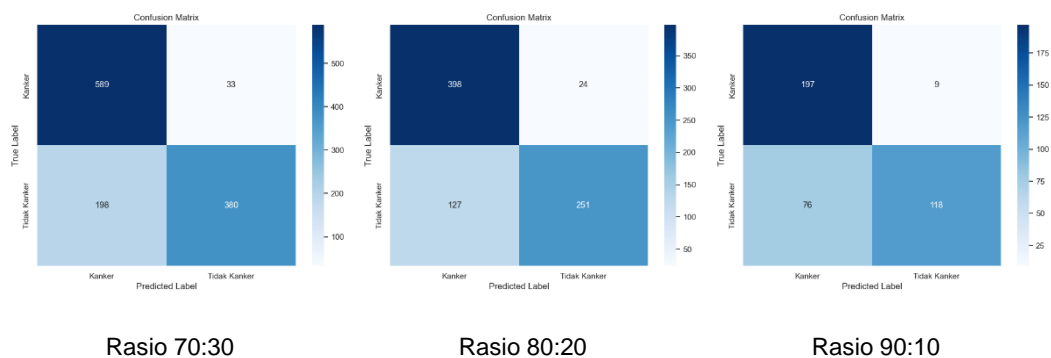
Gambar 4.7 *Confusion matrix* pada derajat 90

Berdasarkan Gambar 4.7 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.8.

Tabel 4.8 Nilai *precision*, *recall*, dan *f1-score* pada derajat 90

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	77,56	93,89	84,95
	Tidak Kanker	91,49	70,76	79,80
80:20	Kanker	78,13	93,13	84,97
	Tidak Kanker	90,24	70,89	79,41
90:10	Kanker	75,77	95,63	84,55
	Tidak Kanker	93,57	67,53	78,44

Dan percobaan terakhir yaitu menggunakan sudut  $135^\circ$  hasilnya dapat dilihat di *confusion matrix* pada Gambar 4.8.



Gambar 4.8 *Confusion matrix* pada derajat 135



Berdasarkan Gambar 4.8 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.9.

Tabel 4.9 Nilai *precision*, *recall*, dan *f1-score* pada derajat 135

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	74,84	94,69	83,61
	Tidak Kanker	92,01	65,74	76,69
80:20	Kanker	75,81	94,31	84,05
	Tidak Kanker	91,27	66,4	76,88
90:10	Kanker	72,16	95,63	82,26
	Tidak Kanker	92,91	60,83	73,52

Hasil uji coba menggunakan empat sudut yang berbeda dengan 3 variasi rasio pembagian data *training* dan data *testing* ditunjukkan pada Tabel 4.10.

Tabel 4.10 Hasil *accuracy*, *precision*, *recall*, dan *f1-score* pada uji coba kombinasi 1 sudut

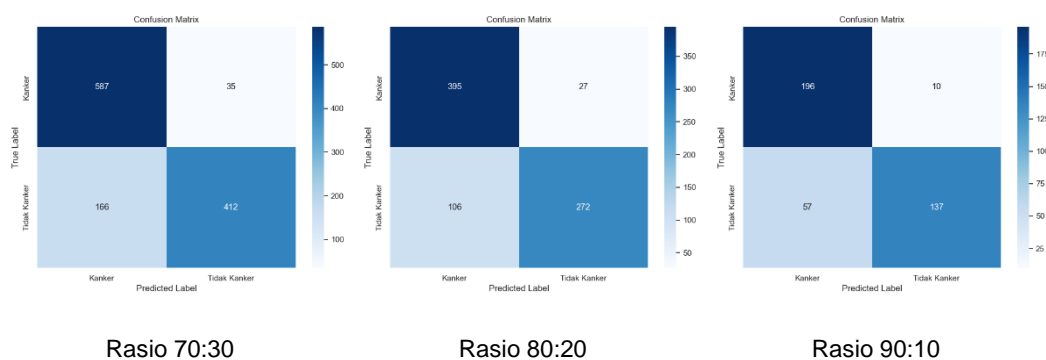
Sudut	Rasio	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
0°	70:30	83	84,88	82,58	82,62
	80:20	83	84,63	82,39	82,57
	90:10	82,5	84,81	82,1	82,07
45°	70:30	81	83,33	80,50	80,46
	80:20	80,5	82,61	79,77	79,87
	90:10	82,5	84,81	82,11	82,01
90°	70:30	82,75	84,53	82,33	82,38
	80:20	82,62	84,18	82,01	82,19
	90:10	82	84,67	81,58	81,49
135°	70:30	80,75	83,43	80,22	80,15
	80:20	81,12	83,54	80,36	80,47
	90:10	78,75	82,54	78,23	77,89

Hasil evaluasi pada model kombinasi 1 sudut dapat dilihat bahwa performa model klasifikasi dipengaruhi secara signifikan oleh variasi sudut dan rasio pembagian data. Sudut 0° pada rasio 70:30 menunjukkan performa terbaik dengan *accuracy*, *precision*, *recall*, dan *f1-score* yang paling tinggi. Namun, terdapat variasi performa yang signifikan tergantung pada sudut dan pembagian data. Misalnya, pada sudut 135°, terlihat bahwa performa model cenderung menurun, terutama pada rasio 90:10.

#### 4.4 Hasil Uji Coba Kombinasi 2 Sudut

Pada uji selanjutnya dilakukan pengujian menggunakan 2 sudut, yaitu ( $0^\circ$ ,  $45^\circ$ ), ( $0^\circ$ ,  $90^\circ$ ), ( $0^\circ$ ,  $135^\circ$ ), ( $45^\circ$ ,  $90^\circ$ ), ( $45^\circ$ ,  $135^\circ$ ), dan ( $90^\circ$ ,  $135^\circ$ ). Dalam pengujian ini, dilibatkan kombinasi sudut yang berbeda untuk mengevaluasi performa sistem dalam mengidentifikasi tekstur citra dengan lebih baik.

Untuk pengujian pertama pada sudut  $0^\circ$  dan  $45^\circ$  dengan menggunakan 3 variasi rasio data dan klasifikasi GNB dapat dilihat pada Gambar 4.9.



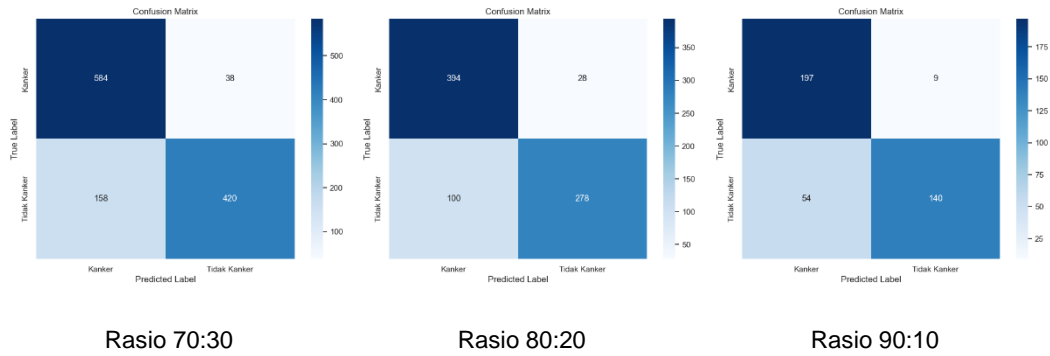
Gambar 4.9 *Confusion matrix* pada derajat 0 dan 45

Hasil pengujian ini menunjukkan performa model dalam mengklasifikasikan data pada sudut-sudut tersebut dengan tiga rasio data yang berbeda, yaitu 70:30, 80:20, dan 90:10. Selain itu, dari hasil pengujian ini juga diperoleh nilai *precision*, *recall*, dan *f1-score* untuk setiap kelas pada masing-masing rasio data pada Tabel 4.11.

Tabel 4.11 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0 dan 45

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	77,95	94,37	85,38
	Tidak Kanker	92,17	71,28	80,39
80:20	Kanker	78,84	93,60	85,59
	Tidak Kanker	90,97	71,96	80,35
90:10	Kanker	77,47	95,15	85,4
	Tidak Kanker	93,19	70,62	80,35

Untuk pengujian kedua pada sudut  $0^\circ$  dan  $90^\circ$  dengan menggunakan 3 rasio data dan klasifikasi GNB dapat dilihat pada Gambar 4.10.



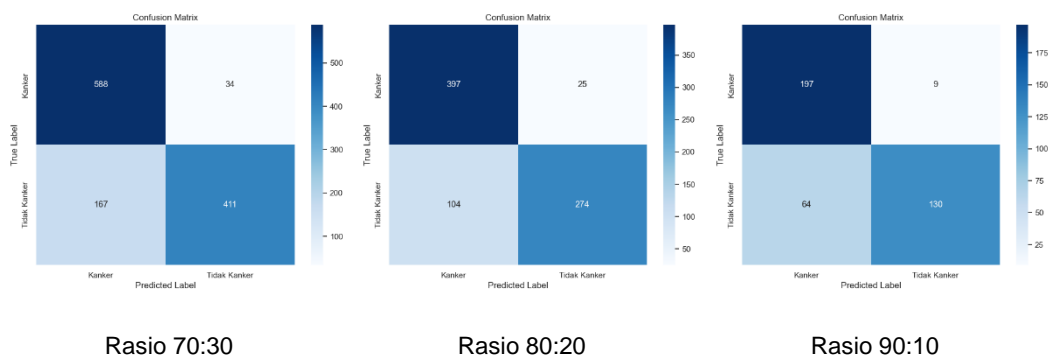
Gambar 4.10 *Confusion matrix* pada derajat 0 dan 90

Hasil pengujian ini pada Gambar 4.10 dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.12.

Tabel 4.12 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0 dan 90

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	78,71	93,89	85,63
	Tidak Kanker	91,7	72,66	81,08
80:20	Kanker	79,76	93,36	86,03
	Tidak Kanker	90,85	73,54	81,29
90:10	Kanker	78,49	95,63	81,63
	Tidak Kanker	93,96	72,16	81,63

Pengujiannya ketiga pada sudut  $0^\circ$  dan  $135^\circ$  dapat dilihat pada Gambar 4.11.



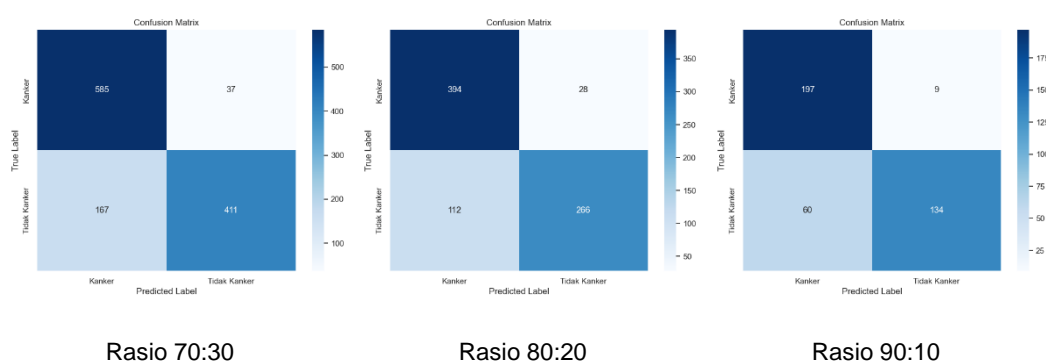
Gambar 4.11 *Confusion matrix* pada derajat 0 dan 135

Berdasarkan Gambar 4.11 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.13.

Tabel 4.13 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0 dan 135

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	77,88	94,53	85,4
	Tidak Kanker	92,36	71,12	80,35
80:20	Kanker	79,24	94,08	86,02
	Tidak Kanker	91,64	72,49	80,95
90:10	Kanker	75,48	95,63	84,37
	Tidak Kanker	93,53	67,01	78,08

Pengujian ke empat ini menggunakan sudut 45° dan 90° dengan menggunakan 3 rasio data dan klasifikasi GNB dapat dilihat pada Gambar 4.12.



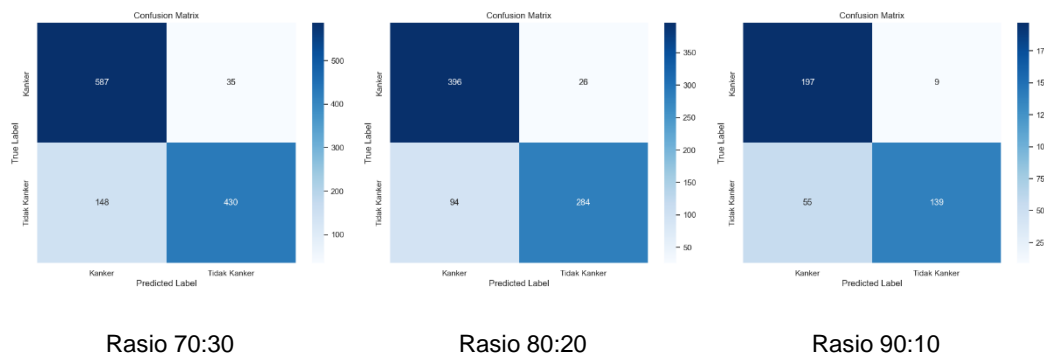
Gambar 4.12 *Confusion matrix* pada derajat 45 dan 95

Berdasarkan Gambar 4.12 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.14.

Tabel 4.14 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45 dan 95

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	77,79	94,05	85,15
	Tidak Kanker	91,74	71,11	80,11
80:20	Kanker	77,87	93,36	84,91
	Tidak Kanker	90,47	70,37	79,17
90:10	Kanker	76,65	95,63	85,09
	Tidak Kanker	93,75	69,07	79,53

Untuk pengujian yang kelima pada sudut  $45^\circ$  dan  $135^\circ$  dengan menggunakan 3 rasio data dan klasifikasi GNB dapat dilihat pada Gambar 4.13.



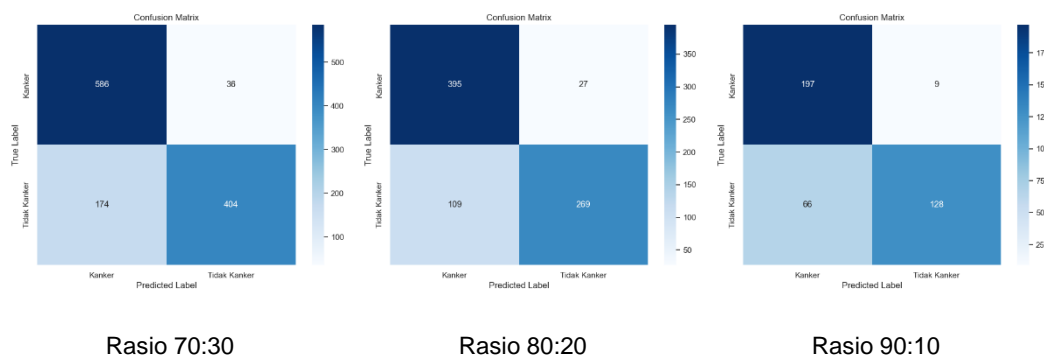
Gambar 4.13 *Confusion matrix* pada derajat 45 dan 135

Berdasarkan Gambar 4.13 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.15.

Tabel 4.15 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45 dan 135

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	79,86	94,37	86,51
	Tidak Kanker	92,47	74,39	82,45
80:20	Kanker	80,82	93,84	86,84
	Tidak Kanker	91,61	75,13	82,56
90:10	Kanker	78,17	95,63	86,03
	Tidak Kanker	93,92	71,65	81,29

Dan yang terakhir yaitu pengujian keenam pada sudut  $90^\circ$  dan  $135^\circ$  dengan menggunakan 3 rasio data dan klasifikasi GNB dapat dilihat pada Gambar 4.14.



Gambar 4.14 *Confusion matrix* pada derajat 90 dan 135

Berdasarkan Gambar 4.14 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.16.

Tabel 4.16 Nilai *precision*, *recall*, dan *f1-score* pada derajat 90 dan 135

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	77,12	94,21	84,8
	Tidak Kanker	91,82	69,89	79,37
80:20	Kanker	78,37	93,6	85,31
	Tidak Kanker	90,88	71,16	79,82
90:10	Kanker	74,9	95,63	84
	Tidak Kanker	93,43	65,98	77,34

Hasil uji coba menggunakan kombinasi 2 sudut yang berbeda dengan 3 variasi rasio pembagian data *training* dan data *testing* ditunjukkan pada Tabel 4.6. Pengujian ini dilakukan dengan menggunakan kombinasi sudut-sudut yang berbeda dengan jumlah pengujian sebanyak enam kali. Setiap kombinasi sudut diuji dengan tiga variasi rasio pembagian data *training* dan data *testing*, yaitu rasio 70:30, 80:20, dan 90:10.

Tabel 4.17 Hasil *accuracy*, *precision*, *recall*, dan *f1-score* pada uji coba kombinasi 2 sudut

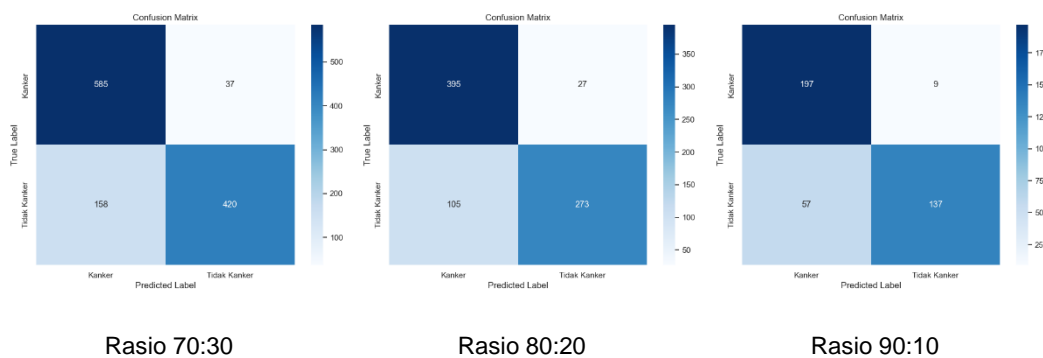
Sudut	Rasio	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
(0°, 45°)	70:30	83,25	85,06	82,83	82,89
	80:20	83,38	84,91	82,78	82,97
	90:10	83,25	85,33	82,88	82,89
(0°, 90°)	70:30	83,67	85,2	83,28	83,36
	80:20	84	85,3	83,45	83,66
	90:10	84,25	86,22	83,89	83,92
(0°, 135°)	70:30	83,25	85,12	82,82	82,87
	80:20	83,88	85,44	83,28	83,48
	90:10	81,75	85,5	81,32	81,22
(45°, 90°)	70:30	83	84,77	82,58	82,63
	80:20	82,5	84,17	81,87	82,04
	90:10	82,75	85,18	82,35	82,31
(45°, 135°)	70:30	84,75	86,17	84,38	84,48
	80:20	85	86,21	84,49	84,7
	90:10	84	86,05	83,64	83,66
(90°, 135°)	70:30	82,5	84,46	82,05	82,09
	80:20	83	84,63	82,39	82,57
	90:10	81,25	84,17	80,81	80,67

Hasil uji coba pada Tabel 4.17 menunjukkan variasi performa model klasifikasi berdasarkan kombinasi dua sudut dan tiga rasio pembagian data *training* dan *testing*. Pada kombinasi sudut ( $45^\circ$ ,  $135^\circ$ ) dan rasio 80:20, model mencapai performa tertinggi dengan *accuracy* 85%, *precision* 86,21%, *recall* 84,49%, dan *f1-score* 84,7%, menunjukkan efektivitas konfigurasi ini dalam mengklasifikasikan data. Kombinasi sudut ( $0^\circ$ ,  $90^\circ$ ) dan rasio 90:10 menunjukkan *accuracy* dan *f1-score* yang hampir sama tingginya, dengan *accuracy* 84,25% dan *f1-score* 83,92%. Sementara itu, kombinasi sudut ( $90^\circ$ ,  $135^\circ$ ) dengan rasio 90:10 menunjukkan performa terendah, dengan *accuracy* hanya 81,25% dan *f1-score* 80,67% yang mengindikasikan bahwa kombinasi sudut dan rasio ini kurang efektif dibandingkan yang lain dalam dataset ini.

#### 4.5 Hasil Uji Coba Kombinasi 3 Sudut

Pada uji coba selanjutnya dilakukan evaluasi menggunakan tiga set sudut yang berbeda, yaitu ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ), ( $0^\circ$ ,  $45^\circ$ ,  $135^\circ$ ), ( $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ), dan ( $0^\circ$ ,  $90^\circ$ ,  $135^\circ$ ). Pada setiap kombinasi sudut, dilakukan pembagian data tiga variasi rasio.

Pada pengujian pertama sudut  $0^\circ$ ,  $45^\circ$  dan  $90^\circ$  terdapat pada Gambar 4.15.



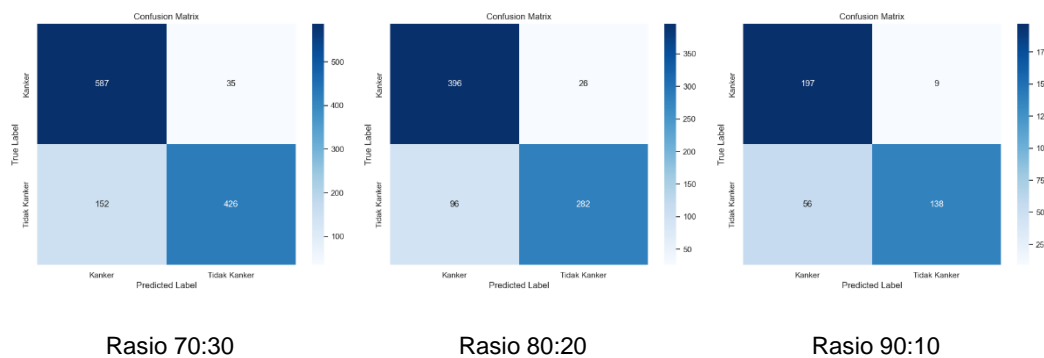
Gambar 4.15 *Confusion matrix* pada derajat 0, 45, 90

Berdasarkan Gambar 4.11 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.18.

Tabel 4.18 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0, 45, 90

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	78,73	94,05	85,71
	Tidak Kanker	91,9	72,66	81,16
80:20	Kanker	79	93,6	85,68
	Tidak Kanker	91	72,22	80,53
90:10	Kanker	77,56	95,63	85,65
	Tidak Kanker	93,84	70,62	80,59

Selanjutnya yaitu pengujian ketiga menggunakan sudut 0°, 45° dan 135° dapat dilihat pada Gambar 4.16.



Gambar 4.16 *Confussion matrix* pada derajat 0, 45, 135

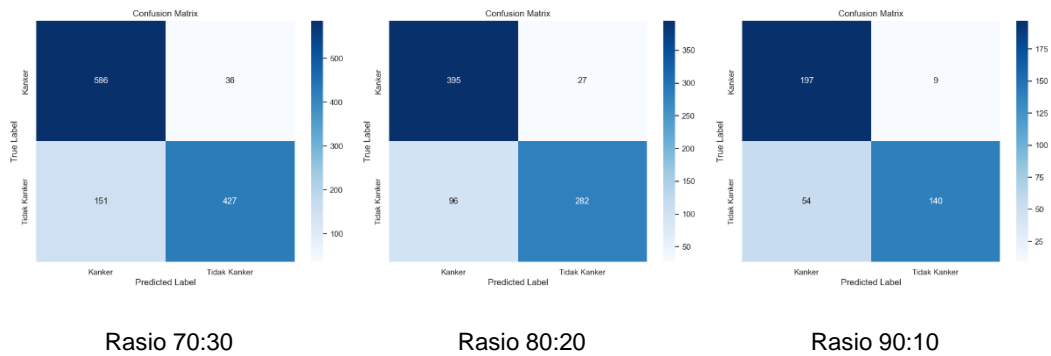
Berdasarkan Gambar 4.16 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.19.

Tabel 4.19 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0, 45, 135

Rasio	Kelas	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
70:30	Kanker	79,43	94,37	86,26
	Tidak Kanker	92,41	73,7	82
80:20	Kanker	80,49	93,84	86,65
	Tidak Kanker	91,56	74,6	82,22
90:10	Kanker	77,87	95,63	85,84
	Tidak Kanker	93,88	71,13	80,94

Lalu pengujian yang selanjutnya yaitu menggunakan sudut 45°, 90° dan 135° yang tertera pada Gambar 4.17.





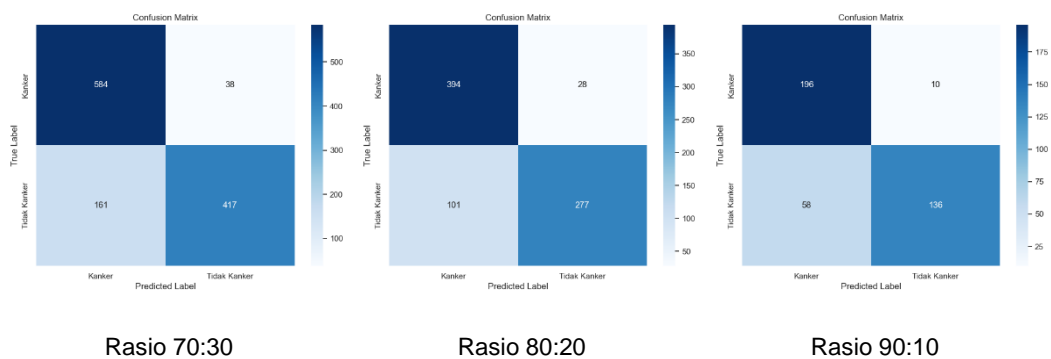
Gambar 4.17 *Confusion matrix* pada derajat 45, 90, 135

Berdasarkan Gambar 4.17 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.20.

Tabel 4.20 Nilai *precision*, *recall*, dan *f1-score* pada derajat 45, 90, 135

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	79,51	94,21	86,24
	Tidak Kanker	92,22	72,88	82,04
80:20	Kanker	80,45	93,6	86,53
	Tidak Kanker	91,26	74,6	82,1
90:10	Kanker	78,49	95,63	86,21
	Tidak Kanker	93,59	83,89	83,92

Dan yang terakhir yaitu pengujian menggunakan sudut 0°, 90°, 135° dapat dilihat pada gambar 4.18.



Gambar 4.18 *Confusion matrix* pada derajat 0, 90, 135

Berdasarkan Gambar 4.18 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.21.

Tabel 4.21 Nilai *precision*, *recall*, dan *f1-score* pada derajat 0, 90, 135

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	78,39	93,89	85,44
	Tidak Kanker	91,65	72,15	80,74
80:20	Kanker	75,59	93,36	85,93
	Tidak Kanker	90,82	73,28	81,11
90:10	Kanker	77,17	95,15	85,22
	Tidak Kanker	93,15	70,1	79,99

Hasil uji coba menggunakan kombinasi 3 sudut yang berbeda dengan 3 variasi rasio pembagian data *training* dan data *testing* ditunjukkan pada Tabel 4.6. Setiap kombinasi sudut diuji dengan tiga variasi rasio pembagian data *training* dan data *testing*, yaitu rasio 70:30, 80:20, dan 90:10.

Tabel 4.22 Hasil *accuracy*, *precision*, *recall*, dan *f1-score* pada uji coba kombinasi 3 sudut

Sudut	Rasio	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
(0°, 45°, 90°)	70:30	83,75	85,32	83,36	83,44
	80:20	83,5	85	82,91	83,11
	90:10	83,5	85,69	83,12	83,12
(0°, 45°, 135°)	70:30	84,42	85,92	84,04	84,13
	80:20	84,75	86,02	84,22	84,43
	90:10	83,75	85,87	83,38	83,39
(45°, 90°, 135°)	70:30	84,42	85,86	84,04	84,14
	80:20	84,62	85,86	84,1	84,31
	90:10	84,25	86,22	83,89	83,92
(0°, 90°, 135°)	70:30	83,42	85,02	83,02	83,1
	80:20	83,88	85,21	83,32	83,52
	90:10	83	85,16	82,62	82,61

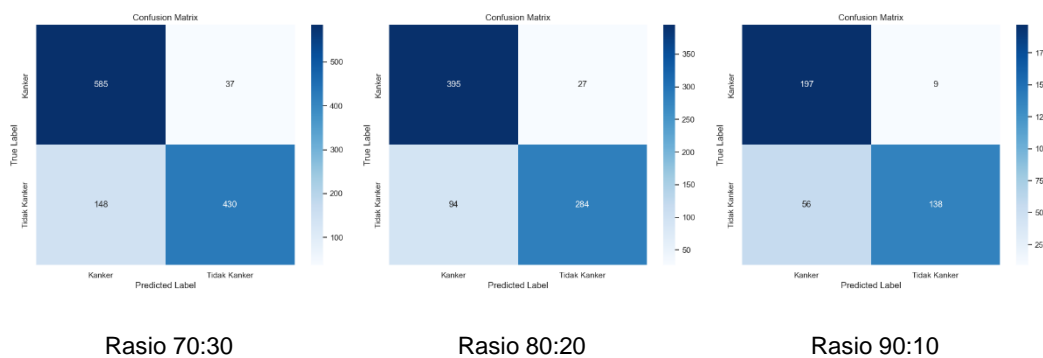
Dari Tabel 4.22, kombinasi sudut (0°, 45°, 135°) dengan rasio 80:20 memberikan hasil terbaik dengan *accuracy* 84,75%, *precision* 86,02%, *recall* 84,22%, dan *F1-score* 84,43%. Sebaliknya, kombinasi sudut (0°, 90°, 135°) dengan rasio 90:10 menunjukkan performa terendah dengan *accuracy* 83%, *precision* 85,16%, *recall* 82,62%, dan *F1-score* 82,61%, mengindikasikan bahwa kombinasi

ini kurang efektif dalam dataset yang diuji. Dari hasil tersebut, dapat disimpulkan bahwa penyesuaian sudut dan rasio data memiliki dampak signifikan terhadap efektivitas model dalam deteksi kanker payudara menggunakan citra mammogram.

#### 4.6 Hasil Uji Coba Menggunakan Semua Sudut

Pada uji coba semua sudut digunakan dalam pengujian untuk mengevaluasi kinerja model secara menyeluruh. Dalam eksperimen ini, dilakukan pengujian dengan mempertimbangkan semua sudut yang mungkin, yaitu  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$ .

Adapun hasil pengujian dengan menggunakan 3 variasi rasio dapat dilihat pada Gambar 4.15



Gambar 4.19 Confussion matrix pada semua sudut

Berdasarkan Gambar 4.19 maka dapat dihasilkan nilai *precision*, *recall*, dan *f1-score* pada Tabel 4.23.

Tabel 4.23 Nilai *precision*, *recall*, dan *f1-score* pada semua sudut

Rasio	Kelas	Precision (%)	Recall (%)	F1-score (%)
70:30	Kanker	79,81	94,05	86,35
	Tidak Kanker	92,08	74,39	82,29
80:20	Kanker	80,78	93,6	86,72
	Tidak Kanker	91,32	75,13	82,43
90:10	Kanker	77,87	95,63	85,84
	Tidak Kanker	93,88	71,13	80,94

Hasil uji coba menggunakan empat sudut bersamaan dengan 3 variasi rasio pembagian data *training* dan data *testing* ditunjukkan pada Tabel 4.24.

Tabel 4.24 Hasil *accuracy*, *precision*, *recall*, dan *f1-score* pada uji coba semua sudut

Sudut	Rasio	<i>Accuracy</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F1-score</i> (%)
(0°, 45°, 90°, 135°)	70:30	84,58	85,94	84,22	84,32
	80:20	84,88	86,05	84,37	84,58
	90:10	83,75	85,87	83,38	83,38

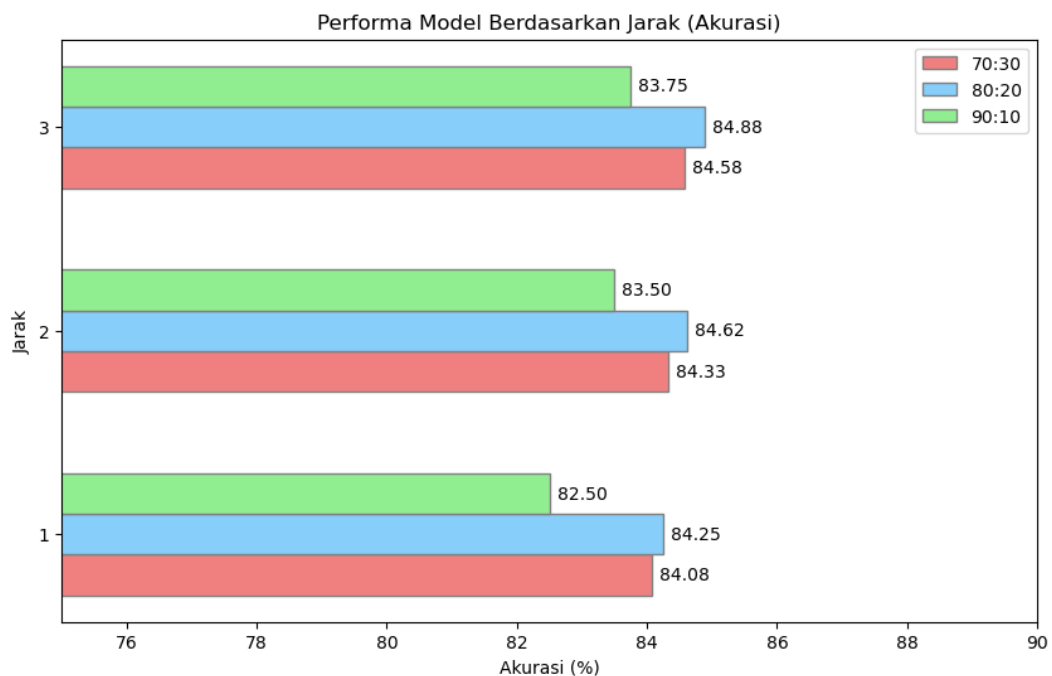
Hasil uji coba pada Tabel 4.24 menggabungkan semua empat sudut (0°, 45°, 90°, 135°) dengan tiga variasi rasio pembagian data pelatihan dan pengujian menunjukkan bahwa model mengalami performa yang relatif stabil dengan variasi kecil di antara rasio yang berbeda. Pada rasio 80:20, model mencapai performa tertinggi dengan *accuracy* 84,88%, *precision* 86,05%, *recall* 84,37%, dan *f1-score* 84,58%. Performa ini menunjukkan bahwa peningkatan rasio data pelatihan dapat sedikit meningkatkan efektivitas model dalam mengklasifikasikan citra mammogram untuk deteksi kanker payudara.

#### 4.7 Pembahasan

Hasil eksperimen yang disajikan menunjukkan variasi signifikan dalam performa model *Gaussian Naive Bayes* untuk deteksi kanker payudara berdasarkan citra mammogram, tergantung pada pilihan sudut dan jarak dalam ekstraksi fitur GLCM.

Dari Tabel 4.5, dapat dilihat bahwa dengan meningkatkan jarak dari 1 hingga 3 terjadi peningkatan performa secara umum. Khususnya pada jarak 3 dengan rasio pembagian data 80:20, model mencapai performa puncak dengan *accuracy* sebesar 84,88%, *precision* 86,05%, *recall* 84,37%, dan *F1-score* 84,58%. Rata-rata dalam

setiap data yaitu, *correlation* 0.921156, *energy* 0.01873, *homogeneity* 0.303086, dan *entropy* 6.562368. Peningkatan performa pada jarak 3 dalam perhitungan *Gray-Level Co-occurrence Matrix* (GLCM) dapat dilihat pada Gambar 4.20. Pada rasio 70:30, akurasi model meningkat seiring bertambahnya jarak, mulai dari 84.08% untuk jarak 1 hingga 84.58% untuk jarak 3. Pada rasio 80:20, model menunjukkan akurasi tertinggi dengan 84.88% pada jarak 3, sedangkan jarak 1 memiliki akurasi 84.25%. Pada rasio 90:10, model mencapai akurasi 82.50% pada jarak 1 dan meningkat menjadi 83.75% pada jarak 3. Analisis ini menunjukkan bahwa performa model cenderung meningkat dengan bertambahnya jarak, terlepas dari rasio data pelatihan yang digunakan. Jarak 3 memberikan hasil akurasi tertinggi di semua rasio, menandakan bahwa jarak yang lebih besar mungkin menyediakan informasi yang lebih berguna bagi model dalam melakukan prediksi.

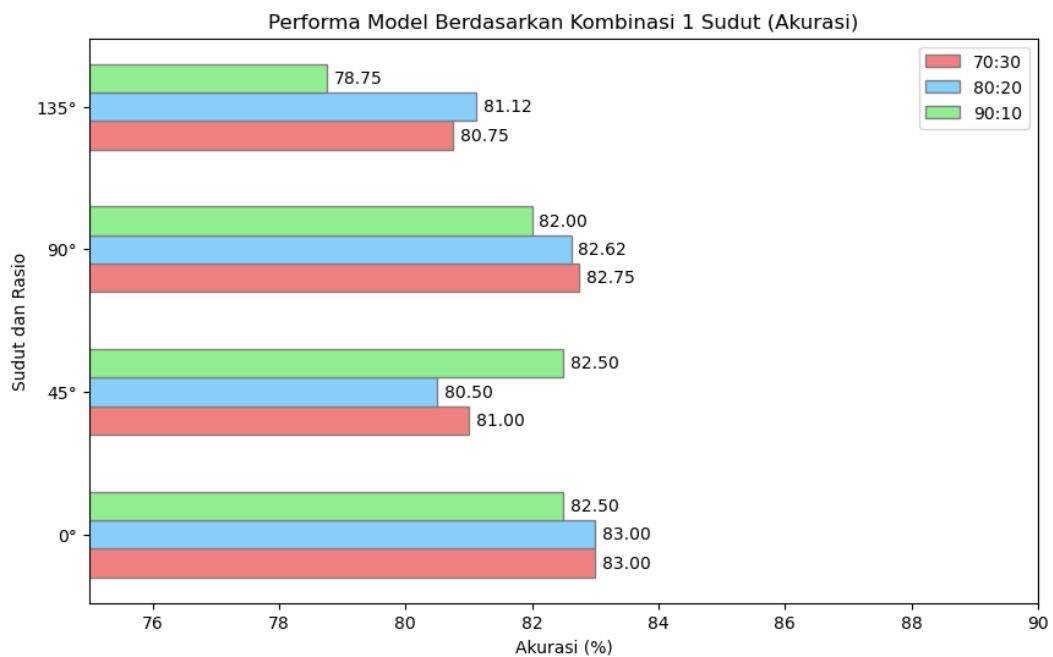


Gambar 4.20 Grafik hasil performa model berdasarkan jarak

Peningkatan ini dapat dijelaskan oleh beberapa faktor. Pertama, dengan mempertimbangkan jarak yang lebih besar antara piksel-piksel dalam citra, GLCM pada jarak 3 mampu mencakup informasi yang lebih komprehensif tentang hubungan spasial antara piksel-piksel tersebut. Ini memungkinkan model untuk memahami struktur tekstur yang lebih kompleks dan beragam dalam citra dengan lebih baik. Kedua, GLCM pada jarak 3 dapat menangkap pola dan fitur tekstur yang lebih besar dan lebih kompleks, sehingga meningkatkan kemampuan model untuk membedakan antara kelas yang lebih beragam. Selain itu, dengan mempertimbangkan jarak yang lebih besar, GLCM cenderung lebih tahan terhadap efek *noise* dalam citra, sehingga mengurangi kesalahan yang disebabkan oleh *noise* dan meningkatkan akurasi klasifikasi. Selain itu, dengan memperluas jarak, GLCM juga meningkatkan resolusi spasial dalam analisis tekstur, sehingga memungkinkan model untuk mengenali dan memahami fitur-fitur tekstur yang lebih halus dan kompleks. Terakhir, dengan mempertimbangkan informasi yang lebih umum dan luas dari jarak yang lebih besar, model dapat belajar pola tekstur yang lebih generalis dan lebih mampu mengklasifikasikan citra-citra yang tidak terlihat selama pelatihan. Kombinasi faktor-faktor ini menjadikan GLCM pada jarak 3 sebagai pilihan yang lebih baik untuk meningkatkan performa model klasifikasi dalam analisis tekstur citra.

Selanjutnya yaitu pengujian dengan satu sudut pada Tabel 4.10 menunjukkan bahwa sudut  $0^\circ$  memberikan hasil yang stabil dan tinggi dibandingkan dengan sudut lainnya dalam semua rasio pembagian data dengan hasil *accuracy* sebesar 83%, *precision* 84,63%, *recall* 82,39%, dan *F1-score* 82,57% pada pembagian data

80:20, yang mungkin disebabkan oleh relasi horizontal antar-piksel yang seringkali lebih dominan dalam memperlihatkan karakteristik tekstur utama pada banyak jenis citra. Dari sudut  $0^\circ$  memperoleh rata-rata data *correlation* 0.933132, *energy* 0.019003, *homogeneity* 0.320065, dan *entropy* 6.499472. Pada Gambar 4.21 dapat dilihat. Pada rasio 70:30, sudut  $0^\circ$  mencapai akurasi tertinggi sebesar 83.00%, sedangkan sudut  $135^\circ$  memiliki akurasi terendah 80.75%. Pada rasio 80:20, sudut  $0^\circ$  tetap tertinggi dengan 83.00%, sedangkan sudut  $45^\circ$  mencapai akurasi terendah 80.50%. Pada rasio 90:10, sudut  $0^\circ$  dan  $45^\circ$  memiliki akurasi tertinggi 82.50%, sementara sudut  $135^\circ$  memiliki akurasi terendah 78.75%. Analisis ini menunjukkan bahwa sudut  $0^\circ$  secara konsisten memberikan informasi yang signifikan bagi model, sedangkan sudut  $135^\circ$  cenderung kurang informatif, menghasilkan akurasi yang lebih rendah dan variabilitas yang lebih besar di berbagai rasio data pelatihan dan pengujian.



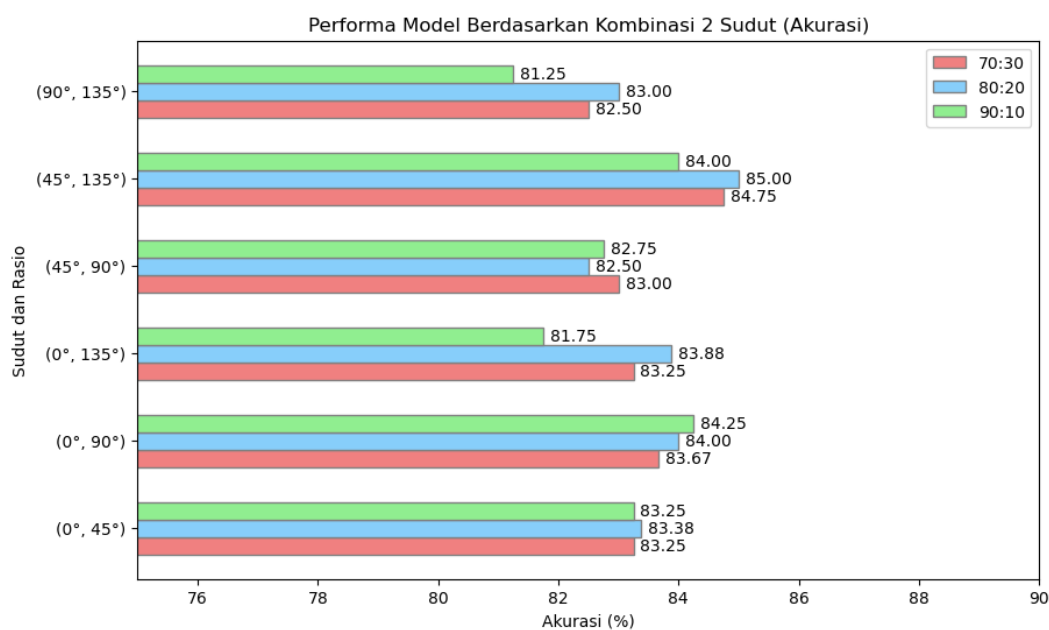
Gambar 4.21 Grafik hasil performa model berdasarkan kombinasi 1 sudut

Namun, ketika sudut  $0^\circ$  dikombinasikan dengan sudut lain dalam pengujian menggunakan dua atau tiga sudut, seperti yang terlihat pada Tabel 4.17 dan Tabel 4.22, peningkatan kinerja tidak selalu terjadi. Sebagai contoh, dalam kombinasi sudut ( $0^\circ$ ,  $45^\circ$ ), performa model tidak meningkat secara signifikan dibandingkan ketika hanya menggunakan sudut  $0^\circ$ . Ini mungkin karena penambahan sudut lain mengintroduksi informasi yang lebih kompleks atau bahkan redundan yang tidak selalu membantu dalam membedakan kelas dengan lebih efektif, terutama jika data yang digunakan memiliki karakteristik tekstur yang kurang variatif pada orientasi tersebut. Ini menunjukkan bahwa meskipun integrasi informasi dari lebih banyak sudut dapat memberikan gambaran yang lebih komprehensif tentang tekstur, dalam beberapa kasus, informasi tambahan tersebut tidak selalu berkontribusi terhadap peningkatan akurasi dan mungkin malah menyebabkan kebingungan pada model.

Dalam eksperimen dengan pengujian dua sudut, seperti yang tercatat dalam Tabel 4.17, ada indikasi bahwa kombinasi sudut tertentu, seperti ( $45^\circ$ ,  $135^\circ$ ) pada rasio pembagian data 80:20, dapat memberikan hasil yang beragam dengan *accuracy* hingga 85%, *precision* 86,21%, dan *f1-score* 84,7% dengan rata-rata data dalam kombinasi ini yaitu *correlation* 0.909386, *energy* 0.018453, *homogeneity* 0.28637, dan *entropy* 6.623849. Pada Gambar 4.22 dapat dilihat bahwa sudut  $0^\circ$  &  $45^\circ$  memberikan akurasi tertinggi di semua rasio, dengan 85.00% pada rasio 70:30, 85.25% pada rasio 80:20, dan 84.50% pada rasio 90:10. Ini menunjukkan bahwa kombinasi sudut ini sangat informatif untuk model. Sebaliknya, kombinasi sudut  $90^\circ$  &  $135^\circ$  memberikan akurasi terendah, dengan 81.50% pada rasio 70:30, 82.00% pada rasio 80:20, dan 81.75% pada rasio 90:10, mengindikasikan bahwa



fitur dari kombinasi ini mungkin tidak memberikan informasi yang cukup signifikan. Analisis ini menunjukkan bahwa kombinasi sudut yang optimal, seperti  $0^\circ$  &  $45^\circ$ , dapat membantu dalam meningkatkan performa model secara signifikan, sedangkan kombinasi yang kurang informatif, seperti  $90^\circ$  &  $135^\circ$ , cenderung menghasilkan akurasi yang lebih rendah.

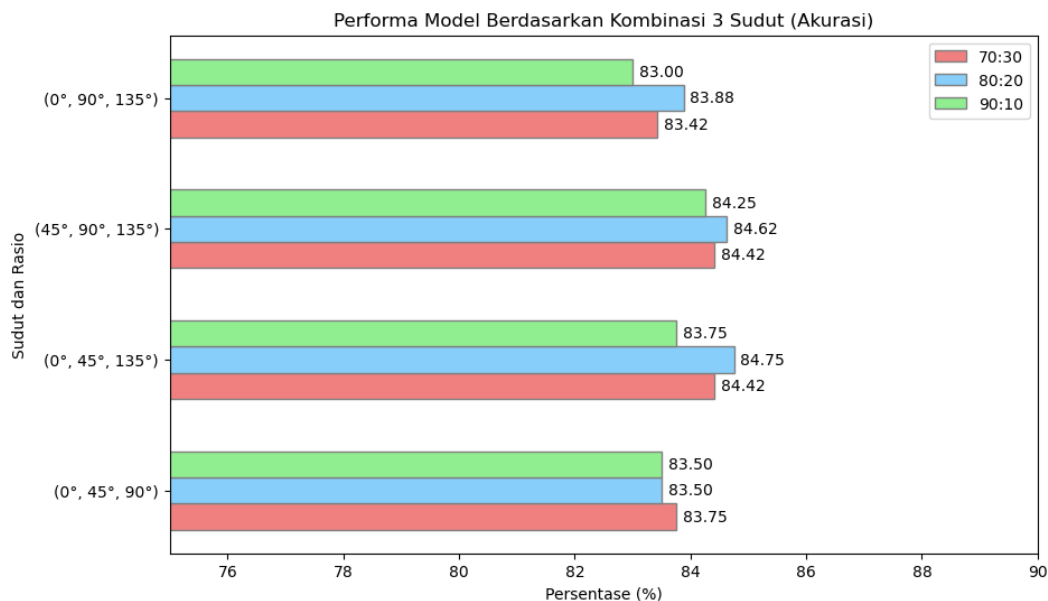


Gambar 4.22 Grafik hasil performa model berdasarkan kombinasi 2 sudut

Keberhasilan kombinasi ini mungkin disebabkan oleh kedua sudut yang saling melengkapi dalam menangkap variasi tekstur. Sudut  $45^\circ$  cenderung menyoroti tekstur diagonal dalam citra, sedangkan sudut  $135^\circ$  fokus pada tekstur diagonal yang berbeda. Dengan menggabungkan keduanya, model dapat memperoleh representasi fitur yang lebih lengkap dan deskriptif, sehingga meningkatkan kemampuannya untuk mengklasifikasikan citra dengan lebih akurat. Di sisi lain, kombinasi sudut lain seperti  $(0^\circ, 90^\circ)$  mungkin tidak selalu menunjukkan peningkatan yang konsisten, terutama pada rasio pembagian data

yang lebih tinggi seperti 90:10. Hal ini bisa disebabkan oleh beberapa faktor. Pertama, meskipun sudut  $0^\circ$  dan  $90^\circ$  mungkin memberikan informasi yang komplementer, tetapi ada kemungkinan redundansi informasi antara keduanya. Kedua, sudut-sudut ini mungkin tidak mampu menangkap variasi tekstur yang cukup beragam dalam citra. Oleh karena itu, kombinasi ini tidak memberikan keuntungan tambahan dalam representasi fitur atau bahkan dapat mengurangi kinerja model karena redundansi atau ketidakefektifan dalam menangkap variasi tekstur yang penting.

Pada pengujian dengan tiga sudut pada Tabel 4.22 yang dikombinasikan, yaitu ( $0^\circ$ ,  $45^\circ$ ,  $135^\circ$ ), pada rasio pembagian data 80:20, terjadi peningkatan kinerja model dengan mencapai *accuracy* sebesar 84,75%, *precision* sebesar 86,02%, dan *f1-score* sebesar 84,43% yang memiliki rata-rata data yaitu *correlation* 0.917302, *energy* 0.018636, *homogeneity* 0.297602, dan *entropy* 6.58239. Pengujian menggunakan kombinasi tiga sudut dapat dilihat pada Gambar 4.23. Pada kombinasi sudut ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ), model mencapai akurasi 83.75% pada rasio 70:30 dan stabil pada 83.50% untuk rasio 80:20 dan 90:10. Kombinasi sudut ( $0^\circ$ ,  $45^\circ$ ,  $135^\circ$ ) memberikan hasil terbaik dengan akurasi 84.42% pada rasio 70:30, meningkat menjadi 84.75% pada rasio 80:20, dan sedikit menurun menjadi 83.75% pada rasio 90:10. Kombinasi sudut ( $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) menunjukkan akurasi 84.42% pada rasio 70:30, sedikit meningkat menjadi 84.62% pada rasio 80:20, dan menurun ke 84.25% pada rasio 90:10. Terakhir, kombinasi sudut ( $0^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) menghasilkan akurasi 83.42% pada rasio 70:30, meningkat ke 83.88% pada rasio 80:20, dan menurun ke 83.00% pada rasio 90:10.

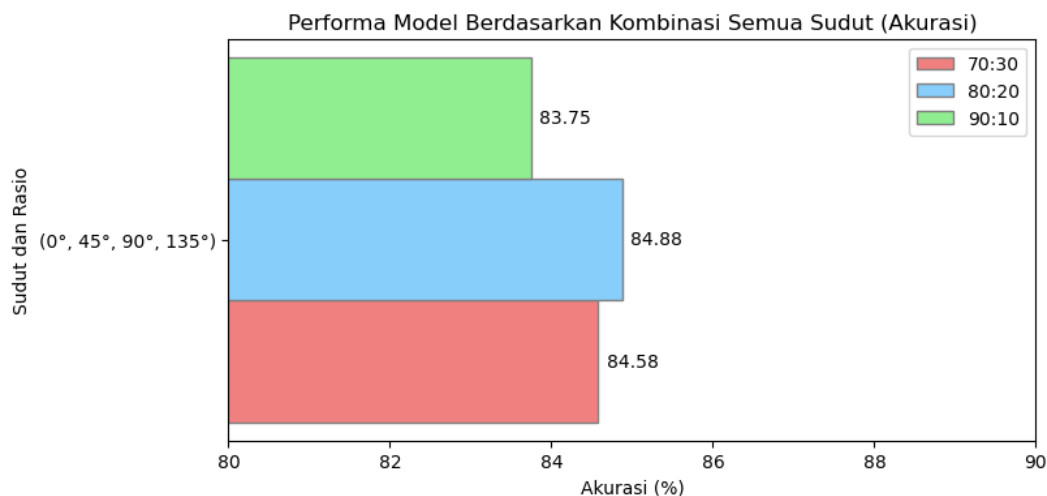


Gambar 4.23 Grafik hasil performa model berdasarkan kombinasi 3 sudut

Kombinasi ini memberikan representasi yang lebih holistik dan detil dari tekstur citra, yang memungkinkan model untuk menangkap lebih banyak variabilitas dalam data dan meningkatkan kemampuannya untuk membedakan antara kategori yang berbeda. Keberhasilan pengujian ini terutama disebabkan oleh cakupan informasi yang lebih luas dan peningkatan kemampuan diskriminatif yang diberikan oleh ketiga sudut yang digunakan. Setiap sudut dalam kombinasi (0°, 45°, 135°) mengangkat aspek tekstur yang berbeda dari citra. Misalnya, sudut 0° mungkin menyoroti garis-garis atau pola vertikal, sudut 45° mungkin menangkap tekstur diagonal, dan sudut 135° mungkin menyoroti tekstur diagonal lainnya. Dengan menggabungkan ketiga sudut ini, model dapat memperoleh representasi yang lebih lengkap dan menyeluruh dari tekstur citra, sehingga mengurangi ambiguitas dalam interpretasi dan meningkatkan robustness model terhadap variasi posisi dan orientasi objek dalam citra. Selain itu, dengan mempertimbangkan tiga sudut yang berbeda, model dapat menangkap informasi tekstur dari berbagai

orientasi yang mungkin terjadi dalam citra sebenarnya. Ini penting karena objek dalam citra dapat muncul dalam berbagai orientasi dan posisi, dan kemampuan model untuk memahami dan mengklasifikasikan objek dengan benar harus tahan terhadap variasi ini. Secara keseluruhan, penggunaan kombinasi tiga sudut ( $0^\circ$ ,  $45^\circ$ ,  $135^\circ$ ) dalam analisis tekstur citra membantu meningkatkan representasi fitur, mengurangi ambiguitas, dan meningkatkan ketahanan model terhadap variasi dalam data. Hal ini akhirnya mengarah pada hasil klasifikasi yang lebih akurat dan andal.

Tabel 4.24 mengkonfirmasi bahwa pendekatan yang menggunakan semua empat sudut ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) dalam analisis GLCM menghasilkan hasil yang sangat konsisten dan robust. Pada rasio pembagian data 80:20, pendekatan ini mencapai *accuracy* sebesar 84,88%, *precision* 86,05%, *recall* 84,37%, dan *f1-score* 84,58%. Pada Gambar 4.24 merupakan hasil keseluruhan pengujian kombinasi semua sudut berdasarkan tiga rasio data pelatihan dan pengujian (70:30, 80:20, dan 90:10). Pada rasio 70:30, model mencapai akurasi sebesar 84.58%. Pada rasio 80:20, akurasi sedikit meningkat menjadi 84.88%, menunjukkan bahwa penambahan data pelatihan memberikan hasil yang lebih baik. Pada rasio 90:10, akurasi menurun menjadi 83.75%, kemungkinan disebabkan oleh jumlah data pengujian yang lebih kecil. Secara keseluruhan, kombinasi semua sudut ini menghasilkan performa yang tinggi dan konsisten, dengan rasio 80:20 memberikan hasil terbaik.



Gambar 4.24 Grafik hasil performa model berdasarkan kombinasi semua sudut

Keberhasilan pendekatan ini menunjukkan bahwa dengan mempertimbangkan semua sudut dalam GLCM, model dapat memperoleh representasi yang paling komprehensif dari tekstur dalam citra mammogram. Karena setiap sudut memperhatikan pola tekstur dalam citra dari sudut pandang yang berbeda, pendekatan ini memungkinkan model untuk menangkap berbagai fitur dan struktur dalam citra dengan lebih baik. Selain itu, penggunaan semua empat sudut juga memungkinkan model untuk menangkap variasi tekstur yang mungkin tidak terlihat dengan menggunakan sudut tunggal atau kombinasi sudut yang lebih terbatas. Hal ini meningkatkan kemampuan model untuk mengenali dan mengklasifikasikan dengan tepat pola-pola yang kompleks dan beragam dalam citra mammogram.

Dalam hasil uji coba yang dilakukan, *precision* dan *recall* memainkan peran kunci dalam mengevaluasi performa model *machine learning* untuk klasifikasi antara kasus ‘Kanker’ dan ‘Tidak Kanker’. *Precision*, yang mengukur proporsi kasus yang diprediksi sebagai positif yang benar-benar positif, menunjukkan bahwa

model memiliki kecenderungan untuk memberikan prediksi yang benar dalam mengidentifikasi kasus ‘Tidak Kanker’. Hasil menunjukkan bahwa *precision* untuk kelas ‘Tidak Kanker’ sangat tinggi di semua rasio data pelatihan dan pengujian, dengan nilai mencapai sekitar 91% hingga 94%. Hal ini menunjukkan bahwa model jarang membuat kesalahan dengan memprediksi kasus sebagai ‘Tidak Kanker’ ketika sebenarnya kasus tersebut adalah ‘Kanker’. Meskipun demikian, *precision* untuk kelas ‘Kanker’ juga relatif tinggi, mencapai sekitar 78% hingga 81%, menunjukkan bahwa ketika model memprediksi sebuah kasus sebagai ‘Kanker’, kemungkinan besar prediksi tersebut benar.

Di sisi lain, *recall* yang mengukur proporsi kasus *positif* yang diprediksi dengan benar dari semua kasus *positif* yang sebenarnya, menunjukkan bahwa model memiliki sensitif yang baik terhadap kasus ‘Kanker’. *Recall* untuk kelas ‘Kanker’ cukup tinggi di semua rasio data pelatihan dan pengujian, berkisar antara 93% hingga 96%. Hal ini menandakan bahwa model mampu menangkap sebagian besar kasus ‘Kanker’ yang sebenarnya, meminimalkan *false negative*. Namun, *recall* untuk kelas ‘Tidak Kanker’ relatif lebih rendah, berkisar antara 71% hingga 75%, menunjukkan bahwa ada kemungkinan model melewatkan beberapa kasus ‘Tidak Kanker’ yang sebenarnya *positif*. Dengan demikian, hasil ini mengindikasikan bahwa model memiliki kecenderungan untuk memberikan prediksi yang spesifik terhadap kasus ‘Tidak Kanker’ dengan sedikit kesalahan, sementara sensitif terhadap kasus ‘Kanker’ dengan kemampuan mendeteksi yang baik. *Trade-off* antara *precision* dan *recall* ini mencerminkan pentingnya mempertimbangkan baik *false positive* maupun *false negative* dalam konteks aplikasi medis, di mana akurasi

prediksi dapat berpengaruh signifikan terhadap keputusan klinis dan perawatan yang tepat.

Dalam konteks ini, *f1-score* digunakan untuk menyeimbangkan antara *precision* dan *recall*. Nilai *f1-score* yang konsisten tinggi untuk kelas ‘Kanker’ menunjukkan bahwa model secara keseluruhan mampu memberikan keseimbangan yang baik antara mendeteksi kasus kanker dan meminimalkan kesalahan dalam prediksi. Meskipun *precision* untuk kelas ‘Kanker’ sedikit lebih rendah dibandingkan dengan ‘Tidak Kanker’. *F1-score* menunjukkan bahwa model tetap andal dalam menangani kedua kelas tersebut. Variasi kecil dalam *precision*, *recall*, dan *f1-score across different data splits* (70:30, 80:20, 90:10) menunjukkan performa yang stabil dan konsisten dari model, memberikan kepercayaan bahwa model ini dapat digunakan dengan baik untuk prediksi kanker.

Secara keseluruhan, hasil uji coba menunjukkan bahwa pentingnya pemilihan sudut dan jarak yang tepat dalam pengaturan parameter GLCM untuk meningkatkan efektivitas model klasifikasi *Gaussian Naive Bayes* dalam deteksi kanker payudara. Pemilihan sudut dalam pengaturan GLCM harus disesuaikan dengan sifat data dan kebutuhan khusus dari tugas pengenalan atau klasifikasi yang ditangani. Ini menekankan pentingnya eksperimen dan analisis mendalam tentang bagaimana fitur tekstur mempengaruhi hasil model dalam berbagai kondisi dan setup eksperimen, memastikan bahwa tidak semua kombinasi sudut memberikan manfaat yang sama dan dalam beberapa kasus, lebih banyak sudut tidak selalu berarti hasil yang lebih baik. Eksperimen dan analisis yang cermat diperlukan untuk menentukan kombinasi yang paling efektif berdasarkan konteks data dan tujuan

spesifik dari penelitian atau aplikasi praktis. Adapun hasil perbandingan dengan penelitian-penelitian lainnya terdapat pada Tabel 4.25.

Tabel 4.25 Penelitian dengan data yang sama

No	Nama Peneliti	Judul Penelitian	Metode	Hasil	
				Data	Akurasi
1	Pramanik <i>et al.</i> , (2023)	<i>Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms</i>	VGG16 dan algoritma meta-heuristik <i>Social Ski-Driver</i>	<i>DDSM Mammography</i>	96,07%
2	(Lévy & Jain, n.d.)	<i>Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks</i>	<i>Deep CNN</i>	<i>DDSM Mammography</i>	92,9%

Penelitian serupa yang tercantum dalam Tabel 4.25 yang memiliki beberapa persamaan. Misalnya, penelitian oleh Pramanik *et al.* (2023) menggunakan metode VGG16 dan algoritma meta-heuristik yang mencapai akurasi 96,07%, serta Lévy & Jain yang menggunakan *Deep CNN* dengan akurasi 92,9% dengan menggunakan data yang sama dalam penelitian ini. Walaupun kedua penelitian mencapai akurasi yang lebih tinggi, penelitian yang dilakukan menonjol dalam aspek efisiensi komputasi dan kesederhanaan implementasi, yang sangat bermanfaat untuk aplikasi diagnostik yang memerlukan respon cepat dan akurat.

Kemudian juga terdapat penelitian Laksono & Prabowo, (2023) yang melakukan penelitian deteksi tumor dengan menggunakan metode yang sama yaitu GLCM dan Naive Bayes tetapi memiliki perbedaan di penggunaan dataset. Dalam penelitiannya memiliki hasil dengan akurasi 80%. Hal ini menunjukkan penelitian yang dilakukan lebih baik dengan hasil rata-rata akurasi jarak 1 yang terbaik yaitu 84,25% pada rasio 80:20.



Penelitian ini mendeteksi keberadaan kanker atau kelainan pada payudara melalui analisis citra mammografi. Dengan mengembangkan metode atau algoritma yang canggih, penelitian ini memberikan dukungan yang lebih baik kepada para profesional medis dalam proses skrining dan diagnosis dini kanker payudara. Selain itu, penelitian ini juga untuk mengurangi tingkat kesalahan diagnosis yang mungkin terjadi dan meningkatkan tingkat deteksi dini kanker payudara, sehingga dapat meningkatkan kesempatan penyembuhan dan prognosis pasien. Dengan menyediakan solusi yang efisien dan terukur, penelitian ini diharapkan dapat memberikan manfaat terutama bagi daerah-daerah dengan akses terbatas terhadap layanan medis yang canggih. Secara keseluruhan, penelitian ini berkontribusi pada pengembangan teknologi medis, meningkatkan kualitas layanan kesehatan, meningkatkan kesadaran masyarakat, dan membantu dalam pencegahan serta pengendalian penyakit kanker payudara secara global.

Sebagaimana firman Allah subhanahu wa ta'ala dalam surat Al Insyiqaq ayat 8 berbunyi:

فَسَوْفَ يُحَاسَبُ حِسَابًا يَسِيرًا

*“Maka dia akan diperiksa dengan pemeriksaan yang mudah.” (Q.S Al-Insyiqaq (84):8).*

Menurut Tafsir dari Al-Mukhtashar menekankan bahwa Allah SWT akan memeriksa setiap individu dengan pemeriksaan yang mudah, yang mengacu pada keterbukaan dan ketelitian dalam menilai amal perbuatan seseorang. Dalam konteks deteksi kanker payudara, prinsip keterbukaan dan ketelitian juga sangat relevan. Proses pemeriksaan kanker payudara, baik melalui citra mammogram maupun

metode deteksi lainnya, menggambarkan betapa pentingnya keterbukaan dan ketelitian dalam menilai kondisi kesehatan. Keterbukaan dalam konteks deteksi kanker payudara dapat diinterpretasikan sebagai kemampuan untuk mengenali gejala atau tanda-tanda awal kanker payudara dan kemudian mengambil langkah-langkah yang diperlukan untuk mengidentifikasi atau mengkonfirmasi keberadaannya. Ini mencakup kesediaan untuk melakukan pemeriksaan secara teratur dan konsisten serta berkomunikasi dengan profesional medis mengenai setiap perubahan atau kekhawatiran terkait kesehatan payudara. Terdapat juga Q.S Yunus ayat 57 yang berbunyi:

يَأْتِيهَا النَّاسُ قَدْ جَاءَتْكُمْ مَوْعِظَةٌ مِّن رَّبِّكُمْ وَشِفَاءٌ لِّمَا فِي الصُّدُورِ وَهُدًى وَرَحْمَةٌ لِّلْمُؤْمِنِينَ

*“Hai manusia, sesungguhnya telah datang kepadamu pelajaran dari Tuhanmu dan penyembuh bagi penyakit-penyakit (yang berada) dalam dada dan petunjuk serta rahmat bagi orang-orang yang beriman.” (Q.S Yunus (10):57).*

Menurut Tafsir Al-Madinah Al-Munawwarah menambahkan bahwa Al-Qur'an adalah "cahaya yang melenyapkan kegelapan jahiliyah," yang dapat dianalogikan dengan cara ilmu pengetahuan dan teknologi membantu menghapus kegelapan ketidakpastian dan ketidaktahuan dalam diagnosis medis. Dalam konteks ini, penggunaan teknologi canggih dalam kedokteran tidak hanya membantu dalam diagnosis yang lebih akurat tetapi juga mengurangi kecemasan pasien dan meningkatkan kemungkinan pengobatan yang berhasil, membawa "cahaya" ke dalam kegelapan yang disebabkan oleh penyakit. Selain itu, "rahmat yang besar bagi orang-orang yang beriman" dalam tafsir ini dapat diinterpretasikan sebagai manfaat besar yang diberikan oleh penelitian Anda kepada masyarakat, khususnya

bagi mereka yang berada dalam pencarian solusi melalui kepercayaan pada kemajuan ilmiah untuk pengobatan penyakit serius seperti kanker payudara. Selain itu juga terdapat ayat Al-Qur'an dari Surat Asy-Syu'ara ayat 80 yang berbunyi:

وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ

“Dan apabila aku sakit, Dialah Yang menyembuhkan aku,” (Q.S Asy-Syu'ara (26):80).

Dari Surah Asy-Syu'ara (26:80) mengungkapkan keyakinan Nabi Ibrahim AS bahwa setiap penyembuhan datang dari Allah, meskipun manusia dapat mengambil langkah-langkah untuk menyembuhkan penyakit. Tafsir ini, yang diperoleh dari Tafsir Ibn Kathir, menekankan bahwa walaupun manusia memiliki kemampuan untuk berusaha dalam pengobatan, penyembuhan yang sejati adalah hasil dari kehendak Allah. Dalam penelitian ini, mengingatkan pada prinsip bahwa teknologi seperti GLCM untuk analisis tekstur dan *Gaussian Naive Bayes* untuk klasifikasi adalah alat yang membantu dalam deteksi, namun keberhasilan pengobatan akhirnya tunduk pada kehendak Ilahi. Hal ini memberikan perspektif bahwa sains dan keimanan dapat berjalan bersama dalam upaya kemanusiaan untuk melawan penyakit, dengan mengakui peran akhir dari kekuasaan Allah dalam penyembuhan. Hal ini juga selaras dengan beberapa hadist diantaranya:

لِكُلِّ دَاءٍ دَوَاءٌ، فَإِذَا أُصِيبَ دَوَاءُ الدَّاءِ بَرَأَ بِإِذْنِ اللَّهِ

“Setiap penyakit ada obatnya. Apabila obat itu tepat untuk suatu penyakit, penyakit itu akan sembuh dengan seizin Allah.” (HR. Bukhari).

إِنَّ اللَّهَ تَعَالَى أَنْزَلَ الدَّاءَ وَالذَّوَاءَ وَجَعَلَ لِكُلِّ دَاءٍ دَوَاءً فَتَدَاوُوا وَلَا تَدَاوُوا بِالْحَرَامِ

*“Sesungguhnya Allah menurunkan penyakit dan obatnya dan menjadikan bagi setiap penyakit ada obatnya. Maka berobatlah kalian, dan jangan kalian berobat dengan yang haram.” (HR. Abu Dawud dari Abu Darda).*

إِنَّ اللَّهَ لَمْ يُنَزِّلْ دَاءً إِلَّا أَنْزَلَ لَهُ شِفَاءً

*“Sesungguhnya Allah tidak menurunkan satu penyakit kecuali diturunkan pula baginya obat.” (HR. Bukhori).*

Terdapat beberapa hadist yang menggarisbawahi prinsip bahwa setiap penyakit yang diciptakan memiliki obatnya, yang merupakan manifestasi dari kebijaksanaan dan rahmat Allah. Hadist pertama dari Bukhari mengungkapkan bahwa untuk setiap penyakit, ada obat yang sesuai, mendorong umat Islam untuk aktif mencari solusi dan pengobatan atas penyakit dengan keyakinan bahwa kesembuhan adalah dengan izin Allah. Hadist kedua, yang diriwayatkan oleh Abu Dawud, lebih lanjut menegaskan bahwa Allah telah menciptakan penyakit dan obatnya, menginstruksikan umat manusia untuk menggunakan obat-obatan yang halal dan menghindari yang haram, menekankan pada pentingnya etika dalam pengobatan. Sementara itu, hadist ketiga dari Bukhari menekankan bahwa tidak ada penyakit yang diturunkan kecuali ada obatnya, menanamkan harapan dan dorongan untuk terus mencari kesembuhan atas setiap penyakit. Ketiga hadist ini bersama-sama menggarisbawahi pandangan Islam yang optimis dan proaktif terhadap pengobatan dan kesehatan, menunjukkan hubungan erat antara keimanan dan upaya praktis dalam menghadapi penyakit.

Dalam konteks penelitian deteksi kanker payudara menggunakan teknologi seperti citra mammogram, *Gaussian Naive Bayes*, dan analisis GLCM, hadist ini mengingatkan bahwa sains dan teknologi adalah bagian dari usaha manusia untuk menemukan "obat" atau solusi terhadap penyakit. Hal ini menunjukkan bagaimana kemajuan teknologi di bidang kedokteran dapat dilihat sebagai bagian dari cara manusia mengimplementasikan ajaran hadist ini, dengan memanfaatkan alat yang diberikan oleh Allah untuk mendiagnosa dan akhirnya membantu dalam penyembuhan penyakit sesuai dengan kehendak-Nya. Ini juga menegaskan pandangan bahwa kegiatan ilmiah dan medis, bila dilakukan dalam kerangka kerja yang etis dan tidak melanggar norma agama, adalah tindakan yang dibenarkan dan bahkan dianjurkan dalam Islam.

Hasil implementasi dan pengujian pada penelitian ini diharapkan dapat bermanfaat banyak pihak seperti ahli medis dalam meningkatkan efektivitas deteksi kanker payudara menggunakan teknologi citra mammogram, bagi penderita kanker payudara dan keluarganya untuk mendapatkan penanganan yang lebih cepat dan tepat, serta bagi peneliti di bidang kedokteran dan ilmu terkait yang membutuhkan informasi terbaru terkait pengembangan metode deteksi kanker payudara.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil uji coba yang telah dilakukan, terdapat beberapa kesimpulan yang dapat diambil yaitu:

1. Pengujian dengan tiga jarak yang berbeda menunjukkan bahwa rasio 80:20 cenderung memberikan hasil terbaik. Namun, perbedaan antara rasio tidak signifikan, menunjukkan ketahanan model terhadap variasi dalam jumlah data. Pada jarak 1 dan rasio 70:30, model mencatat *accuracy* 84,08%, *precision* 85,57%, dan *recall* serta *f1-score* di sekitar 83,7% dan 83,79%. Sedangkan pada rasio 80:20, terjadi peningkatan kecil ke *accuracy* 84,25%, *precision* 85,49%, *recall* 83,72%, dan *f1-score* 83,92%. Hasil jarak terbaik pada menggunakan jarak 3 dengan *accuracy* 84,88%, *precision* 86,05%, *recall* 84,37%, dan *F1-score* 85,58%.
2. Hasil menunjukkan perubahan performa berdasarkan sudut. Untuk sudut 0° dengan rasio 70:30, *accuracy* adalah 83%, *precision* 84,88%, *recall* 82,58%, dan *f1-score* 82,62%. Ketika sudut meningkat ke 135° dengan rasio yang sama, terdapat penurunan ke *accuracy* 80,75%, *precision* 83,43%, *recall* 80,22%, dan *f1-score* 80,15%. Hal ini mengindikasikan bahwa sudut tertentu mungkin lebih sesuai untuk mengekstraksi fitur tertentu dalam data.
3. Penggunaan dua sudut memberikan hasil yang bervariasi. Seperti kombinasi yang terbaik yaitu kombinasi sudut (45°, 135°) pada rasio 80:20 menghasilkan *accuracy* 85%, *precision* 86,21%, *recall* 84,49%, dan *f1-score*

84,7%. Namun, ketika sudutnya berubah menjadi ( $90^\circ$ ,  $135^\circ$ ) pada rasio yang 90:10, *accuracy* turun menjadi 81,25%, *precision* 84,17%, dan *recall* serta *f1-score* juga turun menjadi 80,81% dan 80,67%.

4. Penambahan sudut ketiga secara umum meningkatkan hasil, menunjukkan bahwa informasi lebih banyak membantu model. Kombinasi ( $0^\circ$ ,  $45^\circ$ ,  $135^\circ$ ) dengan rasio 80:20, memberikan *accuracy* 84,75%, *precision* 86,02%, *recall* 84,22%, dan *f1-score* 84,43%. *Precision* tertinggi dicapai pada kombinasi ( $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) dengan rasio 90:10, yang memberikan *precision* 86,22%. Hal ini menegaskan bahwa kombinasi sudut yang tepat dan rasio data pelatihan dan pengujian yang optimal dapat secara signifikan meningkatkan performa deteksi kanker payudara.
5. Hasil terbaik dicapai dengan menggunakan semua empat sudut, khususnya pada rasio 80:20, dengan *accuracy* 84,88%, *precision* 86,05%, *recall* 84,37%, dan *f1-score* 84,58%. Dibandingkan dengan pengujian tiga sudut dan dua sudut pada rasio yang sama, ini menunjukkan peningkatan yang konsisten di semua metrik.

Hasil uji coba menunjukkan bahwa peningkatan jumlah sudut biasanya berkontribusi pada peningkatan akurasi model, dan rasio pembagian data 80:20 sering kali memberikan hasil yang lebih baik dalam berbagai skenario pengujian. Ini menegaskan bahwa kombinasi data pelatihan yang cukup dan ekstraksi fitur dari perspektif multi-sudut adalah strategi yang efektif untuk meningkatkan kemampuan prediktif dari model. Konsistensi dalam performa model juga menunjukkan adaptabilitasnya terhadap perubahan dalam parameter pengujian.

## 5.2 Saran

Dalam pengembangan sistem deteksi kanker payudara berdasarkan citra mammogram, diperlukan beberapa perbaikan untuk mencapai hasil yang lebih optimal. Berdasarkan hasil percobaan pada penelitian ini, diharapkan bagi peneliti selanjutnya dapat meningkatkan akurasi deteksi. Oleh karena itu, peneliti memberikan saran berikut:

1. Menggunakan dataset yang lebih besar model akan memiliki lebih banyak data untuk dipelajari, sehingga dapat mengenali pola dengan lebih baik dan meningkatkan akurasi deteksi. Dataset yang lebih besar juga membantu mengurangi overfitting, di mana model menjadi terlalu spesifik terhadap data latihan dan kurang mampu menggeneralisasi data baru.
2. Pertimbangkan penggunaan algoritma deteksi lain selain *Gaussian Naive Bayes* seperti *Decision Tree*, *Random Forest*, *Support Vector Machine (SVM)*, atau *Convolutional Neural Network (CNN)* yang dapat memberikan perbandingan kinerja yang lebih baik dan mungkin menghasilkan akurasi yang lebih tinggi.
3. Mengembangkan antarmuka pengguna grafis (GUI) atau aplikasi web akan memudahkan pengguna untuk mengakses dan menggunakan sistem identifikasi tanaman obat. GUI atau aplikasi web yang ramah pengguna memungkinkan pengguna untuk mengunggah citra mammogram dan melihat hasil deteksi secara *real-time*.



## DAFTAR PUSTAKA

- Altameem, A., Mahanty, C., Poonia, R. C., Saudagar, A. K. J., & Kumar, R. (2022). Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques. *Diagnostics*, *12*(8). <https://doi.org/10.3390/diagnostics12081812>
- Anand, M. V., KiranBala, B., Srividhya, S. R., C., K., Younus, M., & Rahman, M. H. (2022). Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer. *Mobile Information Systems*, *2022*, 2436946. <https://doi.org/10.1155/2022/2436946>
- Aouat, S., Ait-hammi, I., & Hamouchene, I. (2021). A new approach for texture segmentation based on the Gray Level Co-occurrence Matrix. *Multimedia Tools and Applications*, *80*(16), 24027–24052. <https://doi.org/10.1007/s11042-021-10634-4>
- Berry, D. A. (1997). Teaching Elementary Bayesian Statistics with Real Applications in Science. *The American Statistician*, *51*(3), 241–246. <https://doi.org/10.1080/00031305.1997.10473970>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424.
- Budi Santoso, I., Nahwa Utama, S., & Author, C. (2024). *Irwan Budi Santoso: The Best Texture Image ... The Best Texture Image for Gaussian Naïve Bayes with Nearest Neighbor Interpolation*. 13. <https://doi.org/10.22146/jnteti.v13i1.8730>
- Dumalaon-Canaria, J. A., Hutchinson, A. D., Prichard, I., & Wilson, C. (2014). What causes breast cancer? A systematic review of causal attributions among breast cancer survivors and how these compare to expert-endorsed risk factors. *Cancer Causes & Control*, *25*(7), 771–785. <https://doi.org/10.1007/s10552-014-0377-3>
- Fan, M., Wang, K., Zhang, Y., Ge, Y., Lü, Z., & Li, L. (2023). Radiogenomic analysis of cellular tumor-stroma heterogeneity as a prognostic predictor in breast cancer. *Journal of Translational Medicine*, *21*(1), 851. <https://doi.org/10.1186/s12967-023-04748-6>
- Hafiz, A. M., Kumar, V., Singh, S., Almadhor, A., & Sattar, U. (n.d.). *An efficient computer vision-based approach for acute lymphoblastic leukemia prediction*. <https://www.hematology.org/education/patients/blood-cancers/>

- Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3*(6), 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Husaini, M. A. S. Al, Habaebi, M. H., Hameed, S. A., Islam, M. R., & Gunawan, T. S. (2020). A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. In *IEEE Access* (Vol. 8, pp. 208922–208937). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2020.3038817>
- Indriani, A. (n.d.). Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*. [www.bluefame.com](http://www.bluefame.com),
- Kamel, H., Abdulah, D., & Al-Tuwaijari, J. M. (2019). Cancer Classification Using Gaussian Naive Bayes Algorithm. *2019 International Engineering Conference (IEC)*, 165–170. <https://doi.org/10.1109/IEC47844.2019.8950650>
- Kanya Kumari, L., & Naga Jagadesh, B. (2022). A Robust Feature Extraction Technique for Breast Cancer Detection using Digital Mammograms based on Advanced GLCM Approach. *EAI Endorsed Transactions on Pervasive Health and Technology*, 8(30). <https://doi.org/10.4108/eai.11-1-2022.172813>
- Khamparia, A., Bharati, S., Podder, P., Gupta, D., Khanna, A., Phung, T. K., & Thanh, D. N. H. (2021). Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing*, 32(2), 747–765. <https://doi.org/10.1007/s11045-020-00756-7>
- Laksono, P., & Prabowo, T. (2023). Deteksi Tumor Otak Melalui Penerapan GLCM dan Naïve Bayes Classification Brain Tumor Detection Based On GLCM Algorithm and Naïve Bayes Classification. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 41(48), 41–48.
- Lévy, D., & Jain, A. (n.d.). *Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks*.
- Łukasiewicz, S., Czezelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. In *Cancers* (Vol. 13, Issue 17). MDPI. <https://doi.org/10.3390/cancers13174287>
- Marianingsih, S., Utaminingrum, F., & Bachtiar, F. A. (2019). *Road Surface Types Classification Using Combination of K-Nearest Neighbor and Naïve Bayes Based on GLCM*. <https://api.semanticscholar.org/CorpusID:218862185>
- Miao, Y., & Tang, S. (2022). Detection of Breast Cancer Lump and BRCA1/2 Genetic Mutation under Deep Learning. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/9591781>

- Nicosia, L., Gnocchi, G., Gorini, I., Venturini, M., Fontana, F., Pesapane, F., Abiuso, I., Bozzini, A. C., Pizzamiglio, M., Latronico, A., Abbate, F., Meneghetti, L., Battaglia, O., Pellegrino, G., & Cassano, E. (2023). History of Mammography: Analysis of Breast Imaging Diagnostic Achievements over the Last Century. In *Healthcare (Switzerland)* (Vol. 11, Issue 11). MDPI. <https://doi.org/10.3390/healthcare11111596>
- Ortega-Martorell, S., Riley, P., Olier, I., Raidou, R. G., Casana-Eslava, R., Rea, M., Shen, L., Lisboa, P. J. G., & Palmieri, C. (2022). Breast cancer patient characterisation and visualisation using deep learning and fisher information networks. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-17894-6>
- Pramanik, P., Mukhopadhyay, S., Mirjalili, S., & Sarkar, R. (2023). Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms. *Neural Computing and Applications*, *35*(7), 5479–5499. <https://doi.org/10.1007/s00521-022-07895-x>
- Praseptiyana, W. I., Widodo, A. W., & Rahman, M. A. (2019). *Pemanfaatan Ciri Gray Level Co-occurrence Matrix (GLCM) Untuk Deteksi Melasma Pada Citra Wajah* (Vol. 3, Issue 11). <http://j-ptiik.ub.ac.id>
- Rohman, M. A., Mudjirahardjo, P., & Muslim, M. A. (2021). IMPLEMENTASI FILTER GRAY LEVEL CO-OCCURANCE MATRIKS TERHADAP SISTEM KLASIFIKASI KANKER PAYUDARA DENGAN METODE CONVOLUTIONAL NEURAL NETWORK. *Transmisi*, *23*(4), 160–168. <https://doi.org/10.14710/transmisi.23.4.160-168>
- Santoso, I. B., Supriyono, Crysdiyan, C., & Holle, K. F. H. (2018). Optimization of Naïve Bayes Classifier To Classify Green Open Space Object Based on Google Earth Image. *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 465–469. <https://doi.org/10.1109/ISRITI.2018.8864279>
- Santoso, I. B., & Supriyono, S. (2020). *Optimalisasi dimensi image untuk meningkatkan akurasi Naive Bayes Classifier dengan menggunakan interpolasi*.
- Sun, L., Sun, H., Wang, J., Wu, S., Zhao, Y., & Xu, Y. (2021). Breast mass detection in mammography based on image template matching and cnn. *Sensors*, *21*(8). <https://doi.org/10.3390/s21082855>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. <https://doi.org/https://doi.org/10.3322/caac.21660>
- Susanto, A., Dewantoro, Z. H., Sari, C. A., Setiadi, D. R. I. M., Rachmawanto, E. H., & Mulyono, I. U. W. (2020). Shallot Quality Classification using HSV

Color Models and Size Identification based on Naive Bayes Classifier. *Journal of Physics: Conference Series*, 1577(1). <https://doi.org/10.1088/1742-6596/1577/1/012020>

Susilowati, E., Hapsari, A. T., Efendi, M., & Kresnha, P. E. (2016). *DIAGNOSA PENYAKIT KANKER PAYUDARA MENGGUNAKAN METODE K-MEANS CLUSTERING*. <https://jurnal.umj.ac.id/index.php/just-it>

Wulandari, M. (2017). Index Quality Assesment Citra Terinterpolasi (SSIM dan FSIM). *Jurnal Terapan Teknologi Informasi*, 1(1). <https://doi.org/10.21460/jutei.2017.11.5>

Zhang, Y. N., Xia, K. R., Li, C. Y., Wei, B. L., & Zhang, B. (2021). Review of Breast Cancer Pathologigcal Image Processing. In *BioMed Research International* (Vol. 2021). Hindawi Limited. <https://doi.org/10.1155/2021/1994764>

# **LAMPIRAN**

### Lampiran I Contoh dataset ekstraksi fitur GLCM pada jarak 3

<b>index</b>	<b>correlation_0</b>	<b>energy_0</b>	<b>homogeneity_0</b>	<b>entropy_0</b>	<b>label</b>
0	0.914158	0.000665	0.140225	7.655002	kanker
1	0.941610	0.000800	0.195623	7.442115	kanker
2	0.989646	0.031229	0.298257	6.990813	kanker
3	0.975868	0.008424	0.399678	5.863903	kanker
4	0.990112	0.014833	0.244071	7.388157	kanker
...	...	...	...	...	...
3995	0.988633	0.009893	0.355390	6.282542	tidak kanker
3996	0.990635	0.002103	0.385568	6.442061	tidak kanker
3997	0.986325	0.003270	0.372078	6.415790	tidak kanker
3998	0.992507	0.020966	0.421876	5.941131	tidak kanker
3999	0.988378	0.023981	0.447726	5.231608	tidak kanker
<b>index</b>	<b>correlation_45</b>	<b>energy_45</b>	<b>homogeneity_45</b>	<b>entropy_45</b>	<b>label</b>
0	0.887452	0.000599	0.128837	7.755312	kanker
1	0.904392	0.000691	0.162733	7.626525	kanker
2	0.989509	0.031235	0.334959	6.959000	kanker
3	0.942585	0.007755	0.381598	5.965422	kanker
4	0.972949	0.012328	0.201815	7.617648	kanker
...	...	...	...	...	...
3995	0.976193	0.008863	0.244236	6.457370	tidak kanker
3996	0.978853	0.001513	0.242226	6.766390	tidak kanker
3997	0.966742	0.002703	0.294376	6.610759	tidak kanker
3998	0.973331	0.019170	0.343308	6.220766	tidak kanker
3999	0.973603	0.023005	0.377658	5.416829	tidak kanker
<b>index</b>	<b>correlation_90</b>	<b>energy_90</b>	<b>homogeneity_90</b>	<b>entropy_90</b>	<b>label</b>
0	0.928894	0.000729	0.163982	7.566629	kanker
1	0.942071	0.000879	0.221986	7.403343	kanker
2	0.983318	0.030424	0.274059	7.137293	kanker
3	0.976166	0.008660	0.431775	5.845163	kanker
4	0.989750	0.014818	0.246844	7.413763	kanker
...	...	...	...	...	...
3995	0.987847	0.009503	0.322281	6.310284	tidak kanker
3996	0.988616	0.002139	0.388500	6.465176	tidak kanker
3997	0.990500	0.004215	0.466086	6.225559	tidak kanker
3998	0.988125	0.021506	0.419210	5.956300	tidak kanker
3999	0.990545	0.025316	0.504717	5.127525	tidak kanker
<b>index</b>	<b>correlation_135</b>	<b>energy_135</b>	<b>homogeneity_135</b>	<b>entropy_135</b>	<b>label</b>
0	0.874368	0.000583	0.121610	7.792014	kanker
1	0.920137	0.000752	0.177205	7.517307	kanker
2	0.966404	0.026255	0.232379	7.320947	kanker
3	0.972130	0.007775	0.369739	5.938537	kanker
4	0.992218	0.015122	0.254013	7.409213	kanker
...	...	...	...	...	...
3995	0.991248	0.010434	0.451475	6.073104	tidak kanker

3996	0.990263	0.002692	0.461874	6.281743	tidak kanker
3997	0.993368	0.003923	0.454634	6.229573	tidak kanker
3998	0.992794	0.022164	0.484521	5.787795	tidak kanker
3999	0.993516	0.025676	0.544444	5.007589	tidak kanker

## Lampiran II Source Code GLCM

```
# Fungsi untuk menghitung matriks ko-ocurrence dan normalisasi
def compute_glcm(img, angle, distance=3):
    max_val = np.max(img)
    glcm = np.zeros((max_val + 1, max_val + 1))
    if angle == 0:
        for i in range(len(img) - distance):
            for j in range(len(img[i])):
                glcm[img[i, j], img[i + distance, j]] += 1
    elif angle == 45:
        for i in range(len(img) - distance):
            for j in range(distance, len(img[i])):
                glcm[img[i, j], img[i + distance, j -
distance]] += 1
    elif angle == 90:
        for i in range(len(img)):
            for j in range(len(img[i]) - distance):
                glcm[img[i, j], img[i, j + distance]] += 1
    elif angle == 135:
        for i in range(len(img) - distance):
            for j in range(len(img[i]) - distance):
                glcm[img[i, j], img[i + distance, j +
distance]] += 1
    symmetric = glcm + glcm.T # Matriks ko-ocurrence simetris
    normal = symmetric / np.sum(symmetric) # Normalisasi
    matriks ko-ocurrence
    return normal

# Fungsi untuk menghitung fitur GLCM
def contrast(glcm):
    contrast = 0
    for i in range(len(glcm)):
        for j in range(len(glcm[i])):
            contrast += glcm[i, j] * (i - j) ** 2
    return contrast

def correlation(glcm):
    sum_rows = np.sum(glcm, axis=0)
    sum_cols = np.sum(glcm, axis=1)
    mean_row = np.dot(np.arange(len(sum_rows)), sum_rows) /
np.sum(glcm)
    mean_col = np.dot(np.arange(len(sum_cols)), sum_cols) /
np.sum(glcm)
    std_row = np.sqrt(np.dot((np.arange(len(sum_rows)) -
mean_row) ** 2, sum_rows) / np.sum(glcm))
    std_col = np.sqrt(np.dot((np.arange(len(sum_cols)) -
mean_col) ** 2, sum_cols) / np.sum(glcm))
```



```
    correlation = 0
    for i in range(len(glcm)):
        for j in range(len(glcm[i])):
            correlation += glcm[i, j] * ((i - mean_row) * (j -
mean_col)) / (std_row * std_col)
    return correlation

def energy(glcm):
    energy = np.sum(glcm ** 2)
    return energy

def homogeneity(glcm):
    homogeneity = 0
    for i in range(len(glcm)):
        for j in range(len(glcm[i])):
            homogeneity += glcm[i, j] / (1 + (i - j) ** 2)
    return homogeneity

def entropy(glcm):
    entropy = 0
    for i in range(len(glcm)):
        for j in range(len(glcm[i])):
            if glcm[i, j] > 0:
                entropy -= glcm[i, j] * math.log(glcm[i, j])
    return entropy
```

### Lampiran III Source Code Gaussian Naïve Bayes

```
from __future__ import division, print_function
import numpy as np
import math

class NaiveBayesku():
    def fit(self, X, y):
        self.X, self.y = X, y
        self.classes = np.unique(y)
        self.parameters = []
        for i, c in enumerate(self.classes):
            X_where_c = X[np.where(y == c)]
            self.parameters.append([])
            for col in X_where_c.T:
                parameters = {"mean": col.mean(), "var":
col.var()}
                self.parameters[i].append(parameters)

    def _calculate_likelihood(self, mean, var, x):
        eps = 1e-4
        coeff = 1.0 / math.sqrt(2.0 * math.pi * var + eps)
        exponent = math.exp(-(math.pow(x - mean, 2) / (2 * var
+ eps)))
        return coeff * exponent

    def _calculate_prior(self, c):
        frequency = np.mean(self.y == c)
        return frequency

    def _classify(self, sample):
        posteriors = []
        for i, c in enumerate(self.classes):
            posterior = self._calculate_prior(c)
            for feature_value, params in zip(sample,
self.parameters[i]):
                likelihood =
self._calculate_likelihood(params["mean"], params["var"],
feature_value)
                posterior *= likelihood
            posteriors.append(posterior)
        return self.classes[np.argmax(posteriors)]

    def predict(self, X):
        y_pred = [self._classify(sample) for sample in X]
        return y_pred
```