

**Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada
Region of Interest Citra Mammogram Menggunakan *Gray Level*
*Co-Occurrence Matrix***

TUGAS AKHIR



**Ika Ayu Pratiwi
1242912002**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
JAKARTA
2026**

**Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada
Region of Interest Citra Mammogram Menggunakan *Gray Level*
*Co-Occurrence Matrix***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer**



**Ika Ayu Pratiwi
1242912002**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
JAKARTA
2026**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Ika Ayu Pratiwi

NIM : 1242912002

Tanda Tangan : 

Tanggal : 02 September 2026

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Ika Ayu Pratiwi
NIM : 1242912002
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik dan Ilmu Komputer
Judul Skripsi : Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada *Region of Interest* Citra Mammogram Menggunakan *Gray Level Co-Occurrence Matrix*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

Dewan Penguji

Pembimbing 1 : Dita Nurmadewi, S.Kom., M.Kom. ()
Penguji 1 : Prof. Dr. Siti Rohajawati, S.Kom., M.Kom. ()
Penguji 2 : Haris Rafi, S.Kom., M.Kom. ()

Ditetapkan di : Jakarta
Tanggal : 02 September 2026

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkah rahmat- Nya serta segala limpahan karunia-Nya yang telah memberikan kemudahan dan kelancaran dalam penulisan dan pengerjaan tugas akhir ini dengan judul “Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada *Region of Interest* Menggunakan *Gray Level Co-Occurrence Matrix*”. Tujuan penelitian ini dilaksanakan dalam rangka memenuhi syarat dalam mendapatkan gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer di Universitas Bakrie.

Penulis menyadari bahwa penyusunan penelitian ini disertai bantuan, doa dan bimbingan dari berbagai pihak. Dalam kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Ibu Dita Nurmadewi, S.Kom., M.Kom. selaku Dosen Pembimbing 1 yang telah bersedia meluangkan waktu dan tenaga untuk membimbing dan mendukung penulis dalam penyusunan tugas akhir ini.
2. Para Dosen Sistem Informasi Universitas Bakrie yang telah memberikan ilmu bermanfaat selama penulis melaksanakan kegiatan perkuliahan.
3. Ayah, Suami, Kakak dan keluarga saya yang selalu mendoakan, menyemangati , serta mendukung saya baik secara moral maupun material. Terimakasih atas doa yang kalian berikan untuk keberhasilan dan kesuksesan saya.
4. Istia Enza Rendiani, Qoeronesya Meycyana, Sehat Ari Rezekinta Tambunan, Rizki Anugrah Vidian, Zefanya Marcel Kurnia Jaya selaku teman seperjuangan penulis dalam masa perkuliahan, dan senantiasa memberikan semangat dan dukungan.

Dengan adanya penyusunan tugas akhir ini dapat memberikan manfaat dan wawasan kepada pembaca. Penulis menyadari bahwa dalam penyusunan memiliki kekurangan sehingga penulis sangat menerima kritik dan saran yang membangun untuk penyusunan tugas akhir.

Jakarta, 02 September 2026



Ika Ayu Pratiwi

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan dibawah ini:

Nama : Ika Ayu Pratiwi
NIM : 1242912002
Program Studi : Sistem Informasi
Fakultas : Fakultas Teknik dan Ilmu Komputer
Jenis Tugas Akhir : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti-Free Right*) atas karya ilmiah saya yang berjudul:

Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada *Region of Interest* Menggunakan *Gray Level Co-Occurrence Matrix*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 02 September 2026
Yang menyatakan,



Ika Ayu Pratiwi

Perbandingan *K-Nearest Neighbors* dan *Gaussian Naive Bayes* pada *Region of Interest* Citra Mammogram Menggunakan *Gray Level Co-Occurrence Matrix*

Ika Ayu Pratiwi

ABSTRAK

Kanker payudara merupakan salah satu penyakit yang banyak ditemukan pada perempuan dan dapat dideteksi melalui pemeriksaan mammografi. Citra mammogram memiliki karakteristik tekstur yang kompleks sehingga diperlukan metode ekstraksi fitur yang dapat merepresentasikan informasi tekstur pada citra. Gray Level Co-Occurrence Matrix (GLCM) merupakan salah satu metode ekstraksi fitur tekstur yang dapat merepresentasikan hubungan spasial antar piksel. Setelah proses ekstraksi fitur, pemilihan metode klasifikasi yang sesuai diperlukan karena setiap algoritma dapat menghasilkan performa yang berbeda pada representasi fitur yang sama. Penelitian ini bertujuan untuk membandingkan performa metode *K-Nearest Neighbors* (KNN) dan *Gaussian Naive Bayes* (GNB) menggunakan fitur GLCM pada klasifikasi citra mammogram.

Dataset yang digunakan terdiri dari 322 citra mammogram unik dari Mammographic Image Analysis Society (MIAS), yang terdiri atas 207 citra kelas negatif dan 115 citra kelas positif. Tahapan preprocessing meliputi konversi grayscale, penentuan *Region of Interest* (ROI), *Contrast Limited Adaptive Histogram Equalization* (CLAHE), dan kuantisasi tingkat keabuan. Pada citra abnormal, ROI ditentukan berdasarkan koordinat kelainan yang tersedia pada anotasi dataset, sedangkan pada citra normal digunakan pendekatan *center-crop*. Ekstraksi fitur dilakukan menggunakan GLCM dengan jarak 3 piksel, empat arah yaitu 0° , 45° , 90° , dan 135° , serta 64 level kuantisasi. Enam fitur tekstur yang digunakan adalah contrast, dissimilarity, homogeneity, energy, correlation, dan Angular Second Moment (ASM), sehingga diperoleh 24 fitur untuk setiap citra. Pengujian KNN dilakukan menggunakan variasi nilai k yaitu 3, 5, 7, 9, 11, 13, dan 15. Berdasarkan hasil pengujian, $k = 5$ dipilih sebagai konfigurasi akhir karena menghasilkan akurasi dan F1 Macro tertinggi, masing-masing sebesar 0,6677 dan 0,5916. Evaluasi kedua metode dilakukan menggunakan *Stratified 5-Fold Cross Validation* dengan pendekatan *out-of-fold* berdasarkan metrik akurasi, presisi, recall, F1-score, F1 Macro, dan confusion matrix.

Hasil evaluasi menunjukkan bahwa GLCM + KNN dengan $k = 5$ memperoleh akurasi sebesar 0,6677, presisi positif sebesar 0,5588, recall positif sebesar 0,3304, F1-score positif sebesar 0,4153, dan F1 Macro sebesar 0,5916. Sementara itu, GLCM + GNB memperoleh akurasi sebesar 0,6739, presisi positif sebesar 0,6562, recall positif sebesar 0,1826, F1-score positif sebesar 0,2857, dan F1 Macro sebesar 0,5372. Berdasarkan hasil tersebut, GNB memiliki performa lebih tinggi pada akurasi dan presisi positif, sedangkan KNN memiliki performa lebih tinggi pada recall positif, F1-score positif, dan F1 Macro. Dengan demikian, tidak terdapat satu metode yang unggul

pada seluruh metrik evaluasi, sehingga pemilihan metode bergantung pada metrik dan tujuan klasifikasi yang menjadi prioritas.

Kata kunci: Citra Mammogram, Gaussian Naive Bayes, GLCM, K-Nearest Neighbors, Klasifikasi.

Comparison of K-Nearest Neighbors and Gaussian Naïve Bayes on Mammogram Image Regions of Interest Using Gray Level Co-Occurrence Matrix

Ika Ayu Pratiwi

ABSTRACT

Breast cancer is one of the diseases commonly found in women and can be detected through mammography examination. Mammogram images have complex texture characteristics; therefore, a feature extraction method is required to represent the texture information contained in the images. Gray Level Co-Occurrence Matrix (GLCM) is one of the texture feature extraction methods that can represent the spatial relationships between pixels. After feature extraction, selecting an appropriate classification method is necessary because different algorithms may produce different performance on the same feature representation. This study aims to compare the performance of K-Nearest Neighbors (KNN) and Gaussian Naïve Bayes (GNB) using GLCM features for mammogram image classification.

The dataset consists of 322 unique mammogram images from the Mammographic Image Analysis Society (MIAS), consisting of 207 negative-class images and 115 positive-class images. The preprocessing stages include grayscale conversion, Region of Interest (ROI) determination, Contrast Limited Adaptive Histogram Equalization (CLAHE), and grayscale quantization. For abnormal images, the ROI was determined based on the lesion coordinates provided in the dataset annotations, while normal images used a center-crop approach. Feature extraction was performed using GLCM with a pixel distance of 3, four directions of 0°, 45°, 90°, and 135°, and 64 quantization levels. Six texture features were used, namely contrast, dissimilarity, homogeneity, energy, correlation, and Angular Second Moment (ASM), resulting in 24 features for each image. KNN was evaluated using different k values of 3, 5, 7, 9, 11, 13, and 15. Based on the experimental results, $k = 5$ was selected as the final configuration because it produced the highest akurasi and F1 Macro, with values of 0.6677 and 0.5916, respectively. Both methods were evaluated using Stratified 5-Fold Cross Validation with an out-of-fold approach based on akurasi, presisi, recall, F1-score, F1 Macro, and confusion matrix.

The evaluation results show that GLCM + KNN with $k = 5$ achieved an akurasi of 0.6677, positive-class presisi of 0.5588, positive-class recall of 0.3304, positive-class F1-score of 0.4153, and F1 Macro of 0.5916. Meanwhile, GLCM + GNB achieved an akurasi of 0.6739, positive-class presisi of 0.6562, positive-class recall of 0.1826, positive-class F1-score of 0.2857, and F1 Macro of 0.5372. These results indicate that GNB achieved higher performance in terms of akurasi and positive-class presisi, whereas KNN achieved higher performance in terms of positive-class recall, positive-class F1-score, and F1 Macro. Therefore, neither method consistently outperformed the other across all evaluation metrics, and the selection of the classification method depends on the prioritized evaluation metric and classification objective.

Keywords: Mammogram Image, GLCM, K-Nearest Neighbors, Gaussian Naive Bayes, Image Classification.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vi
ABSTRAK.....	vii
<i>ABSTRACT</i>	ix
DAFTAR ISI	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
DAFTAR SKRIP PROGRAM	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Citra Mammogram.....	5
2.2 <i>Gray Level Co-Occurrence Matrix</i> (GLCM)	5
2.3 <i>K-Nearest Neighbors</i> (KNN).....	6
2.4 <i>Gaussian Naive Bayes</i> (GNB)	7
2.5 Standarisasi Data	8
2.6 Ekstraksi Fitur Tekstur pada Citra Medis	9
2.7 <i>Region of Interest</i> (ROI)	9
2.8 <i>Contrast Limited Adaptive Histogram Equalization</i> (CLAHE)	10
2.9 <i>Machine Learning</i>	10

2.10 Penelitian Terdahulu	12
BAB III METODE PENELITIAN	15
3.1 Kerangka Penelitian	15
3.1.1 Identifikasi Masalah	17
3.1.2 Studi Literatur	17
3.1.3 Pengumpulan Data dan Input Data	18
3.1.4 <i>Preprocessing</i> Citra	19
3.1.5 Kuantisasi Citra	20
3.1.6 Ekstraksi Fitur GLCM	20
3.1.7 Fitur Tekstur GLCM	20
3.1.8 Klasifikasi KNN	21
3.1.9 Klasifikasi <i>Gaussian Naive Bayes</i>	22
3.1.10 Evaluasi Model	22
3.1.11 Perbandingan Performa	24
3.1.12 Kesimpulan dan Saran	25
3.2 Metode Pengumpulan Data	25
3.3 Objek Penelitian	26
3.4 Alokasi Waktu Penelitian	27
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Pengumpulan dan Input Data	28
4.2 Hasil <i>Preprocessing</i> Citra	29
4.2.1 Konversi Grayscale	29
4.2.2 Hasil Region of Interest	30
4.2.3 Hasil CLAHE	31
4.3 Hasil Ekstraksi Fitur GLCM	32
4.4 Hasil Klasifikasi GLCM + KNN	34
4.5 Hasil Klasifikasi GLCM + <i>Gaussian Naive Bayes</i>	36
4.6 Perbandingan Hasil GLCM + KNN dan GLCM + <i>Gaussian Naive Bayes</i> ..	39
4.7 Pembahasan	41
4.8 Ringkasan Hasil	43

BAB V KESIMPULAN DAN SARAN	45
5.1 Kesimpulan.....	45
5.2 Saran	46
DAFTAR PUSTAKA.....	47
LAMPIRAN	49

DAFTAR GAMBAR

Gambar 3.1 Kerangka Penelitian	16
Gambar 3.2 Dataset Yang Digunakan Dalam Penelitian	25
Gambar 4.1 Contoh Citra Asli dengan Kotak ROI	31
Gambar 4.2 Contoh Tahapan Preprocessing Citra	32
Gambar 4.3 <i>Confusion Matrix</i> GLCM + KNN	35
Gambar 4.4 <i>Confusion Matrix</i> GLCM + Gaussian Naive Bayes	38

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	12
Tabel 3.1 Alokasi waktu penelitian.....	27
Tabel 4.1 Distribusi Dataset Berdasarkan Kelas	28
Tabel 4.2 Distribusi Dataset Berdasarkan Severity	28
Tabel 4.3 Parameter CLAHE	31
Tabel 4.4 Parameter Ekstraksi Fitur GLCM	33
Tabel 4.5 Parameter KNN.....	34
Tabel 4.6 Hasil Pengujian Variasi Nilai k pada KNN.....	35
Tabel 4.7 Hasil Evaluasi GLCM + Gaussian Naive Bayes	37
Tabel 4.8 Perbandingan Performa Kedua Metode	40

DAFTAR SKRIP PROGRAM

Skrip 4.1 Pembentukan Label Biner Citra.....	29
Skrip 4.2 Proses Crop ROI	31
Skrip 4.3 Implementasi CLAHE	32
Skrip 4.4 Ekstraksi Fitur GLCM	33
Skrip 4.5 Menunjukkan Inti Proses Evaluasi	39

DAFTAR LAMPIRAN

Lampiran 1 <i>Data Set</i> dan Sumber Data	49
Lampiran 2 Contoh Tahapan <i>Preprocessing</i> ROI Citra Mammogram	49
Lampiran 3 Hasil Pengujian KNN dengan $k = 3$	49
Lampiran 4 Hasil Pengujian KNN dengan $k = 5$	50
Lampiran 5 Hasil Pengujian KNN dengan $k = 7$	50
Lampiran 6 Hasil Pengujian KNN dengan $k = 9$	51
Lampiran 7 Hasil Pengujian KNN dengan $k = 9$	51
Lampiran 8 Hasil Pengujian KNN dengan $k = 11$	52
Lampiran 9 Hasil Pengujian KNN dengan $k = 13$	52
Lampiran 10 Hasil Pengujian KNN dengan $k = 15$	52
Lampiran 11 <i>Confusion Matrix</i> KNN	53
Lampiran 12 Hasil Stratified 5-Fold GNB	53
Lampiran 13 <i>Confusion Matrix</i> GNB	53
Lampiran 14 Hasil Perbandingan KNN dan GNB	54