

KLASIFIKASI KANDUNGAN NIKOTIN PADA CITRA TEMBAKAU ROKOK MENGGUNAKAN HYBRID GLCM DAN KNN

Kriel Eleson Upa^{1*}, Supatman²

Universitas Mercu Buana Yogyakarta; Jl. Jembatan Merah No.84C Gejayan Yogyakarta 55283, Telp. (0274) 563589, Faks. (0274) 550703

Keywords:

Kandungan nikotin;
Klasifikasi;
Gray Level Co-occurrence
Matrix (GLCM);
K-Nearest Neighbor (KNN).

Correspondent Email:

rielanwar47@gmail.com



Copyright © [JITET](#) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstrak. Nikotin merupakan zat adiktif yang berbahaya bagi kesehatan, sehingga diperlukan cara yang mudah dan tepat untuk mengetahui kandungan nikotin pada tembakau rokok secara praktis dan berbasis teknologi. Penelitian ini bertujuan untuk mengklasifikasikan kandungan nikotin pada citra tembakau rokok menggunakan pendekatan hybrid metode *Gray Level Co-occurrence Matrix* (GLCM) untuk ekstraksi fitur tekstur dan algoritma *K-Nearest Neighbor* (KNN) untuk klasifikasi. Sebanyak 400 citra diklasifikasikan ke dalam dua kelas kandungan nikotin (0,5-1,5) dan (1,6-2,0). Setiap citra diolah melalui preprocessing, ekstraksi fitur statistik, dan klasifikasi menggunakan KNN. Hasil kinerja dengan K=13, diperoleh akurasi 67,5%. Pendekatan hybrid GLCM-KNN berpotensi digunakan sebagai dasar pengembangan sistem identifikasi kandungan nikotin berbasis citra digital.

Abstract. Nicotine is an addictive substance that is harmful to health, so an easy and precise method is needed to determine the nicotine content in cigarette tobacco practically and based on technology. This study aims to classify the nicotine content in cigarette tobacco images using a hybrid approach, namely using the Gray Level Co-occurrence Matrix (GLCM) method for texture feature extraction and the K-Nearest Neighbor (KNN) algorithm for classification. A total of 400 images were classified into two nicotine content classes, namely (0.5-1.5) and (1.6-2.0). Each image was processed through preprocessing, statistical feature extraction, and classification using KNN. The evaluation results with K = 13 obtained an accuracy of 67.5% with a balanced level of precision and detection ability in both classes. The GLCM-KNN approach has the potential to be used as a basis for developing a digital image-based nicotine content identification system.

1. PENDAHULUAN

Tingkat perokok aktif di Indonesia menunjukkan tren peningkatan yang signifikan. Menurut Survei Kesehatan Indonesia (SKI) 2023 oleh Kementerian Kesehatan, jumlah perokok aktif diperkirakan mencapai 70 juta orang, dengan 7,4% di antaranya berusia antara 10 hingga 18 tahun (Kementerian Kesehatan, 2023). Perhatian utama tertuju pada kelompok usia muda, khususnya anak-anak dan remaja, sebagaimana ditunjukkan dalam laporan Global

Youth Tobacco Survey (GYTS) 2023 yang mencatat kenaikan prevalensi merokok di kalangan pelajar usia 13-15 tahun dari 18,3% pada tahun 2016 menjadi 20,6% pada tahun 2023 (WHO,2023)[1].

Rokok merupakan hasil olahan tembakau yang dikonsumsi melalui proses pembakaran, dimana asap yang dihasilkan dihirup oleh pengguna. Produk ini mencakup berbagai bentuk, seperti rokok kretek, rokok putih, cerutu, dan sejenisnya. Individu yang

secara langsung mengkonsumsi rokok, batik secara reguler maupun sesekali, termasuk hanya satu batang perhari, dikategorikan sebagai perokok aktif. Paparan terhadap asap rokok telah diketahui menurunkan kualitas hidup, mengingat kandungannya yang mencakup lebih dari 4.000 senyawa kimia berbahaya, diantaranya karbon monoksida (CO), polycyclic aromatic hydrocarbons (PAHs), tar, dan nikotin. Nikotin, sebagai zat adiktif, memiliki potensi menimbulkan ketergantungan tidak hanya pada perokok aktif, tetapi juga pada individu yang terpapar secara pasif[2].

Nikotin merupakan alkaloid alami yang terdapat pada tanaman tembakau, seperti *Nicotiana tabacum* dan *Nicotiana rustica*, serta dapat diperoleh secara sintetis. Senyawa ini bersifat adiktif dan diketahui dapat memicu ketergantungan. Secara fisiologis, nikotin mempengaruhi sistem saraf pusat dengan merangsang pelepasan dopamin, sehingga menghasilkan sensasi sementara berupa kenyamanan dan relaksasi, namun meningkatkan risiko ketergantungan dalam jangka panjang[3].

Pengawasan terhadap produk tembakau memerlukan klasifikasi nikotin yang akurat sebagai langkah strategis dalam menekan risiko kesehatan masyarakat. Namun, proses identifikasi kadar nikotin masih bergantung pada teknik analisis konvensional seperti kromatografi, yang dikenal memakan biaya tinggi dan waktu yang lama. Kondisi ini mendorong kebutuhan akan metode alternatif yang lebih efisien dan aplikatif. Salah satu solusi potensial adalah penggunaan teknologi pengolahan citra digital yang dikombinasikan dengan metode ekstraksi fitur tekstur dan algoritma klasifikasi.

Penelitian ini menggunakan pendekatan Gray Level Co-occurrence Matrix (GLCM) sebagai metode ekstraksi fitur tekstur pada citra tembakau rokok. GLCM, yang merupakan teknik statistik orde kedua, menganalisis keterkaitan antar pasangan piksel dalam citra guna mengidentifikasi karakteristik tekstur yang berpotensi mencerminkan kandungan nikotin. Fitur tekstur yang dihasilkan memberikan representasi visual terhadap kualitas serta komposisi tembakau[4].

Pada tahap klasifikasi, diterapkan algoritma K-Nearest Neighbor (KNN), yang dikenal sebagai metode yang sederhana namun

memiliki efektivitas tinggi dalam pengenalan pola. KNN mengklasifikasikan data berdasarkan kedekatan jarak fitur antara data uji dan data latih, serta menunjukkan keunggulan dalam efisiensi komputasi dan kemampuan menangani volume data yang besar. Karakteristik tersebut menjadikan KNN metode yang relevan untuk diterapkan dalam sistem klasifikasi berbasis citra[4].

Efektivitas integrasi metode Gray Level Co-occurrence Matrix (GLCM) dan K-Nearest Neighbor (KNN) dalam klasifikasi objek berbasis citra telah didemonstrasikan oleh berbagai studi sebelumnya. Salah satunya adalah penelitian berjudul “Klasifikasi Bunga Mawar Menggunakan KNN dan Ekstraksi Fitur GLCM dan HSV” yang mendapatkan akurasi hingga 96%[5]. Penelitian lainnya juga yang berjudul “Klasifikasi Jenis Jamur Menggunakan Ekstraksi Fitur GLCM dan K-Nearest Neighbor (KNN)” menghasilkan tingkat akurasi sebesar 77%[6]. Hasil-hasil ini menunjukkan bahwa kombinasi algoritma klasifikasi sederhana seperti KNN dengan metode ekstraksi fitur tekstur berpotensi tinggi dalam membangun sistem identifikasi berbasis citra yang akurat.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi kandungan nikotin pada citra tembakau rokok dengan mengintegrasikan metode GLCM sebagai ekstraksi fitur tekstur dan algoritma KNN sebagai kalsifikator. Pendekatan hybrid GLCM dan KNN diharapkan dapat memberikan solusi klasifikasi yang efisien, akurat, dan dapat diimplementasikan sebagai bagian dari sistem monitoring kualitas produk tembakau di Indonesia.

2. TINJAUAN PUSTAKA

2.1 Nikotin

Nikotin merupakan alkaloid adiktif yang secara alami terdapat pada tanaman tembakau seperti *Nicotiana tabacum* dan *Nicotiana rustica*, maupun hasil sintesis secara kimia. Senyawa ini memiliki struktur 3-(1-metilpirolidin-2-il)piridina dan bersifat basa lemah dengan pH sekitar 8,0 dengan sekitar 31% berada dalam bentuk non-ionik yang mampu menembus membran sel. Dalam bentuk murni, nikotin berupa cairan tak berwarna, bersifat higroskopis, dan mudah menguap,

namun akan mengalami perubahan warna menjadi cokelat dan beraroma tembakau saat terpapar udara atau cahaya. Senyawa ini larut dalam berbagai pelarut termasuk air, alkohol, eter, dan minyak serta dapat dimurnikan melalui penyulingan uap dari larutan yang telah dibasakan[3].

Secara fisiologis, nikotin bersifat toksik dan berperan sebagai stimulan yang mempengaruhi sistem saraf pusat melalui stimulasi pelepasan dopamin, yang menimbulkan efek nyaman sementara dan berpotensi menyebabkan ketergantungan. Kandungan nikotin perbatang rokok berkisar antara 0,5 hingga 3 mg, dengan paparan zat ini diketahui menghambat aktivitas silia pada saluran pernapasan dan turut menyebabkan disfungsi respirasi. Karakteristik adiktif dan psikoaktif tersebut menjadikan nikotin sebagai elemen kunci dalam pengawasan produk tembakau, serta mendorong inovasi teknologi deteksi kadar nikotin yang efisien, baik melalui analisis laboratorium maupun pendekatan berbasis citra dengan pemanfaatan kecerdasan buatan[3].

2.2 Rokok

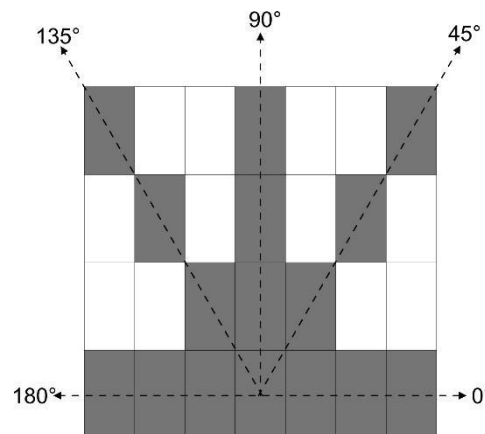
Rokok merupakan produk tembakau yang dibuat dari daun tembakau rajangan yang dibungkus dalam kertas untuk memungkinkan konsumsi melalui pembakaran. Variasinya meliputi rokok kretek buatan tangan atau mesin, rokok putih, cerutu, dan rokok berbahan dasar daun tembakau. Seiring kemajuan teknologi, muncul berbagai inovasi produk tembakau baru, seperti ekstrak tembakau, perisa tembakau, tembakau molase, serta produk yang dikonsumsi dengan cara dihirup atau dikunyah[7].

Rokok elektrik merupakan inovasi terkini dalam konsumsi produk tembakau, yang beroperasi dengan sistem pemanas bertenaga baterai untuk menghasilkan uap dari katrid berisi humektan, seperti propilen glikol atau gliserol, perisa, nikotin, dan dalam beberapa kasus zat lain seperti rimonabant atau amino tadalafil. Karakteristik fisik, ukuran, dan cara penggunaannya, rokok elektrik dirancang menyerupai pengalaman mengkonsumsi rokok tembakau konvensional, termasuk sensasi halusinasi melalui mulut[8].

2.3 GLCM

Gray Level Co-occurrence Matrix (GLCM) merupakan metode ekstraksi fitur tekstur yang diperkenalkan oleh Haralick pada tahun 1973, dengan mengidentifikasi 28 parameter untuk mempresentasikan pola spasial dalam citra[9]. Metode ini diawali dengan pembentukan matriks co-occurrence, yang mencatat frekuensi kemunculan pasangan piksel bernilai keabuan tertentu dalam konfigurasi spasial tertentu. Analisis selanjutnya mengevaluasi hubungan antara piksel referensi dan tetangganya, yang ditentukan berdasarkan jarak dan arah (sudut). Umumnya, jarak antar piksel (d) yang digunakan adalah 1 piksel, dengan orientasi sudut 0° sebagai acuan[9].

Beberapa fitur utama yang dapat diekstraksi dari GLCM meliputi energi, kontras, homogenitas, dan kolerasi. Masing-masing fitur dihitung dengan memperhitungkan jarak satu piksel dalam lima arah berbeda, yaitu pada sudut 0° , 45° , 90° , 135° , dan 180° [4]. Visualisasi orientasi arah sudut yang digunakan dalam proses ini ditampilkan pada Gambar 1.



Gambar 1. Arah Sudut GLCM yang Digunakan[4]

2.4 KNN

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi berbasis jarak yang mengklasifikasikan data baru dengan mempertimbangkan kedekatannya terhadap sejumlah data latih terdekat, sebagaimana ditentukan oleh parameter k [3][10]. Tujuan utamanya adalah menentukan kelas suatu data atau objek berdasarkan kesamaan fitur dengan data latih yang telah diketahui kelasnya[10][11].

Dalam algoritma K-Nearest Neighbor (KNN), proses klasifikasi dilakukan dengan memetakan data latih ke dalam ruang multidimensi, di mana penentuan kelas bergantung pada kedekatan antar data. Kedekatan ini umumnya diukur menggunakan jarak Euclidean, yang mempresentasikan tingkat kesamaan atau perbedaan antara dua objek dalam ruang fitur [11][12], sesuai dengan persamaan 1.

$$d(x - y) = \sqrt{\sum_{j=1}^i (x_j - y_j)^2} \quad (1)$$

Yaitu:

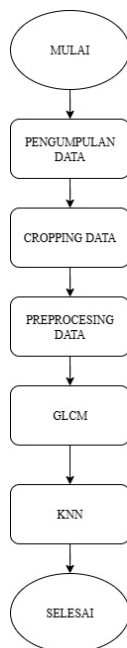
d = jarak data testing ke data training

x_j = data testing ke- j , dengan $j = 1, 2, \dots, n$.

y_j = data training ke- j dengan $j = 1, 2, \dots, n$.

3. METODE PENELITIAN

Penelitian ini dilaksanakan melalui lima langkah terstruktur, dimulai dari pengumpulan data produk rokok hingga tahap evaluasi kinerja GLCM dan KNN dalam mengklasifikasikan, kandungan nikotin yang diagram alurnya ditunjukkan pada Gambar 2.



Gambar 2. Diagram Alur Penelitian

3.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian diambil langsung dari berbagai jenis produk rokok terdaftar resmi yang dijual di pasaran. Pengambilan gambar dilakukan menggunakan kamera ponsel iPhone 7 yang memiliki resolusi 12 MP. Ponsel diletakkan pada tripod dengan jarak pengambilan citra sejauh 20 cm, dan setiap objek rokok difoto di atas kertas HVS putih sebagai latar belakang untuk memastikan konsistensi pencahayaan dan kontras gambar. Proses pengambilan citra dilakukan dengan mengeluarkan setiap batang rokok dari bungkusnya. Jumlah rokok dalam setiap bungkus bervariasi yaitu diantaranya berisi 12, 16, hingga 20 batang setiap bungkusnya. Secara keseluruhan, citra yang dikumpulkan sebanyak 400, yang berasal dari 24 bungkus rokok dari berbagai merek yang berbeda.

3.2 Cropping Data

Setelah proses pengumpulan data sebanyak 400 citra, setelah itu setiap gambar kemudian melalui tahap pemotongan (cropping) secara manual. Setelah itu, data yang telah di crop diklasifikasikan ke dalam dua folder, masing-masing diberi label berdasarkan kandungan nikotin. Folder pertama dilabeli dengan kandungan-nikotin(0,5-1,5), sedangkan folder kedua dilabeli dengan kandungan-nikotin(1,6-2,0). Masing-masing kelas berisi 200 citra, sehingga distribusi data antar kelas seimbang yang ditunjukkan pada Tabel 1.

Tabel 1. Rangkuman Data Citra Tembakau Berdasarkan Kandungan Nikotin

No	Kelas Nikotin	Citra Tembakau	Jumlah Citra
1	Kandungan-Nikotin(0,5-1,5)		200
2	Kandungan-Nikotin(1,6-2,0)		200

3.3 Preprocessing Data

Citra hasil akuisisi kamera umumnya disimpan dalam format .JPEG dengan komposisi warna RGB. Namun, untuk mendukung proses ekstraksi tekstur menggunakan metode Gray Level Co-occurrence Matrix (GLCM), citra perlu dikonversi ke dalam format grayscale dengan resolusi yang lebih rendah. Oleh karena itu, tahap pra-pemrosesan mencakup dua langkah utama, yaitu penyesuaian ukuran citra menjadi 258 x 258 piksel dan transformasi ruang warna dari RGB ke grayscale sebagai persiapan untuk ekstraksi fitur.

3.4 Dataset

Dataset yang telah diproses kemudian dibagi menjadi 2 bagian yaitu, data training dan testing. Pembagian ini dilakukan untuk memastikan model dapat mempelajari pola data dengan baik dan menghasilkan prediksi yang akurat[13]. Proporsi dataset yang digunakan adalah 80% untuk training, dan 20% untuk testing.

3.5 Tahap Ekstraksi Tekstur

Gray Level Co-occurrence Matrix (GLCM) merupakan salah satu metode analisis tekstur citra yang banyak digunakan dan dikembangkan berdasarkan prinsip statistik orde dua. Metode ini memodelkan hubungan kedekatan antar piksel dalam citra melalui berbagai arah orientasi dan jarak spasial tertentu[14]. GLCM menghasilkan matriks persegi berdimensi $N \times N$, dimana N menunjukkan jumlah tingkat keabuan dalam citra tersebut.

Proses ekstraksi fitur tekstur menggunakan metode Gray Level Co-occurrence Matrix (GLCM) dilakukan dengan menganalisis relasi antar piksel pada jarak satu piksel melalui empat orientasi sudut, yakni $0^\circ, 45^\circ, 90^\circ$, dan 135° . Setelah pembentukan matriks GLCM, dilakukan tahap normalisasi, kemudian dihitung empat parameter utama, yaitu: energy, contrast, homogeneity, dan correlation.

3.6 Tahap Klasifikasi Menggunakan KNN

Tahapan akhir dalam penelitian ini mencakup klasifikasi citra tembakau berdasarkan kandungan nikotin dengan

menerapkan algoritma K-Nearest Neighbor (KNN). Algoritma ini termasuk dalam pendekatan lazy learning dan beroperasi dengan cara membandingkan data uji terhadap data latih terdekat yang ditentukan oleh kedekatan fitur, KNN dikategorikan sebagai metode instance-based learning, di mana proses klasifikasi mengacu pada label dari data latih yang menyerupai data uji.

Pemilihan jumlah tetangga terdekat (k) memainkan peran penting dalam menentukan akurasi klasifikasi. Nilai k yang besar berpotensi mereduksi pengaruh noise, namun dapat mengaburkan garis pemisah antar kelas. Oleh karena itu, dalam penelitian ini digunakan parameter $k = 3$ guna mencapai hasil prediksi yang optimal.

Proses perhitungan kedekatan antar data dilakukan menggunakan jarak Euclidean, yang merupakan metode pengukuran jarak linier antara dua titik dalam ruang berdimensi banyak.

Setelah proses klasifikasi dilakukan, sistem akan menghasilkan nilai keluaran berupa probabilitas dari masing-masing kelas, menggunakan pendekatan mirip fungsi softmax. Nilai tertinggi dari hasil ini menentukan label akhir dari citra tembakau. Sistem akan memilih satu dari dua kelas, yaitu: Kandungan Nikotin (0,5-1,5) dan Kandungan Nikotin (1,6-2,0). Keluaran akhir akan ditampilkan dalam bentuk teks yang menyatakan prediksi kandungan nikotin dari citra tembakau rokok tersebut berdasarkan hasil klasifikasi KNN.

3.7 Evaluasi Model

Proses Pengujian dilakukan untuk menilai performa model dalam mengklasifikasikan kandungan nikotin pada rokok berdasarkan dua kategori kandungan nikotin, yaitu 0,5-1,5 dan 1,6-2,0. Model yang telah melalui tahap pelatihan sebelumnya dievaluasi menggunakan data uji yang telah dipisahkan sejak awal. Kinerja klasifikasi dinilai melalui sejumlah metrik evaluasi, antara lain *accuracy*, *precision*, *recall*, dan *F1-score*, guna mengukur tingkat keakuratan hasil prediksi.

Accuracy merupakan metode evaluasi yang digunakan untuk mengukur seberapa tepat model dalam mengklasifikasikan data secara

keseluruhan, baik prediksi yang benar maupun salah[16], ditunjukkan pada persamaan 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision digunakan untuk mengevaluasi tingkat ketepatan model dalam mengidentifikasi data yang termasuk ke dalam suatu kelas tertentu[16], ditunjukkan pada persamaan 3.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall mengukur sejauh mana model mampu mendeteksi data positif yang benar sesuai dengan kenyataan[16], ditunjukkan pada persamaan 4.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1-score adalah metrik evaluasi yang menggabungkan precision dan recall untuk memberikan gambaran menyeluruh mengenai kinerja model dalam klasifikasi[16], ditunjukkan pada persamaan 5.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Keterangan:

TP: True Positive (prediksi benar untuk kelas tertentu)

TN: True Negative (Prediksi benar untuk kelas lain)

FP: False Positive (prediksi salah untuk kelas tertentu)

FN: False Negative (prediksi salah untuk kelas lain)

4. HASIL DAN PEMBAHASAN

4.1 Hasil Ekstraksi Fitur Citra

Setelah dilakukan akuisisi data citra tembakau, tahap klasifikasi dilakukan dengan membagi citra ke dalam dua kategori kelas berdasarkan kandungan nikotin, yaitu kelas dengan kandungan nikotin 0,5-1,5 dan kelas dengan kandungan nikotin 1,6-2,0. Ekstraksi fitur kemudian dilakukan menggunakan metode

Gray Level Co-occurrence Matrix (GLCM), yang menghasilkan data numerik. Hasil ekstraksi tersebut disajikan dalam dua tabel terpisah, masing-masing mempresentasikan fitur tekstur citra dari kedua kelas kandungan nikotin, seperti yang ditampilkan pada Tabel 2 dan Tabel 3.

Tabel 2. Hasil Ekstraksi GLCM kandungan nikotin(0,5-1,5)

contrast	correlation	energy	homogeneity	dissimilarity	ASM	entropy	label
53.6054	0.9462	0.0277	0.2296	4.9571	0.0007	11.0281	Kandut nikotin (0,5-1)
75.0711	0.9536	0.0261	0.2271	5.5787	0.0006	11.3853	Kandut nikotin (0,5-1)
81.7826	0.9266	0.0221	0.1812	6.3379	0.0004	11.5114	Kandut nikotin (0,5-1)
85.1838	0.9252	0.0228	0.1830	6.3421	0.0005	11.4867	Kandut nikotin (0,5-1)
79.0160	0.9290	0.0229	0.1889	6.1957	0.0005	11.4581	Kandut nikotin (0,5-1)

Tabel 3 Hasil Ekstraksi GLCM kandungan nikotin(1,6-2,0)

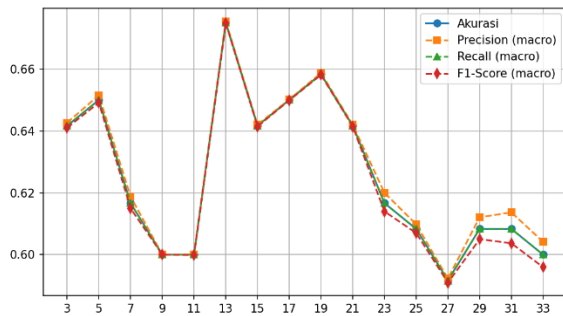
Contrast	Correlation	Energy	Homogeneity	Dissimilarity	ASM	Entropy	Label
46.7816	0.9262	0.0355	0.2489	4.5454	0.0012	10.4674	Kandut n-nikot (1,6-2,)
74.9142	0.9012	0.0250	0.1814	6.1632	0.0006	11.1779	Kandut n-nikot (1,6-2,)
49.1268	0.9444	0.0276	0.2228	4.8869	0.0007	10.9597	Kandut n-nikot (1,6-2,)
58.1285	0.9617	0.0240	0.2307	5.0028	0.0005	11.3232	Kandut n-nikot (1,6-2,)
53.4876	0.9602	0.0273	0.2345	4.8760	0.0007	11.1349	Kandut n-nikot (1,6-2,)

Tabel 2 dan 3 yang ditampilkan mempresentasikan hasil ekstraksi tekstur citra tembakau menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM). Ekstraksi ini menghasilkan sejumlah parameter statistik, antara lain *contrast*, *correlation*, *energy*, *homogeneity*, *dissimilarity*, *angular second moment* (ASM), dan *entropy*. Berdasarkan hasil analisis, fitur *entropy* dan *contrast* menunjukkan variasi antar kelas dengan diskriminatif yang tinggi.

4.2 Optimasi Parameter KNN

Untuk menentukan parameter optimal pada algoritma K-Nearest Neighbors (KNN), dilakukan pengujian terhadap nilai K pada rentang 3 hingga 33. Pemilihan nilai K digunakan untuk sensitivitas model terhadap

data pelatihan dan generalisasi terhadap data uji. Nilai K yang terlalu kecil cenderung menyebabkan overfitting, sedangkan nilai yang terlalu besar dapat menyebabkan underfitting, karena prediksi rata-rata terhadap banyak tetangga yang ditunjukkan pada Gambar 3.

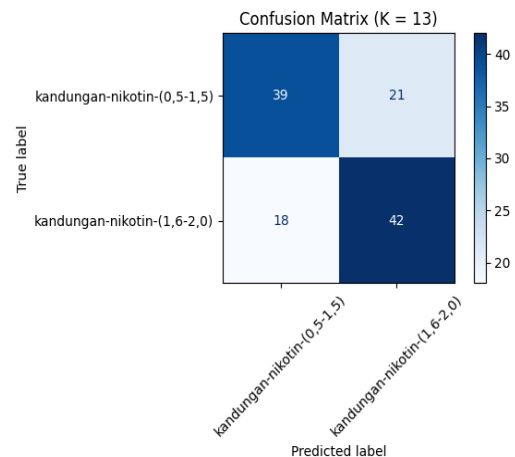


Gambar 3. Akurasi KNN untuk $K=3$ hingga 33

Hasil uji akurasi model pada tiap nilai K ditampilkan pada Gambar 3. Grafik tersebut menunjukkan bahwa akurasi mengalami naik-turun seiring perubahan nilai K , dengan nilai tertinggi tercapai pada $K = 13$, yaitu sebesar 67,5%. Setelah nilai ini, akurasi tidak menunjukkan peningkatan berarti dan cenderung menurun secara bertahap, mengindikasikan bahwa penambahan tetangga justru menyebabkan hilangnya ketajaman klasifikasi. Berdasarkan hasil tersebut, $K = 13$ dipilih sebagai nilai optimal, karena memberikan keseimbangan terbaik antara kompleksitas model dan performa klasifikasi. Nilai ini kemudian digunakan untuk evaluasi lanjutan, termasuk pembentukan confusion matrix dan perhitungan metrik evaluasi lainnya.

4.3 Confusion Matrix

Untuk mengevaluasi performa model secara lebih menyeluruh, dilakukan analisis menggunakan confusion matrix pada nilai parameter optimal, yaitu $K = 13$. Model KNN digunakan untuk mengklasifikasikan citra tembakau ke dalam dua kategori berdasarkan kandungan nikotin yaitu kelas kandungan-nikotin(0,5-1,5) dan kelas kandungan-nikotin(1,6-2,0). Yang ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix Model KNN pada ($K = 13$)

Berdasarkan Gambar 4, model berhasil mengklasifikasikan 39 sampel dari kelas kandungan nikotin (0,5-1,5) dan 42 sampel dari kelas kandungan nikotin(1,6-2,0) secara benar. Sementara itu, terdapat 21 sampel dari kelas kandungan nikotin (0,5-1,5) yang salah diprediksi sebagai kelas kandungan nikotin(1,6-2,0), serta 18 sampel dari kelas kandungan nikotin(1,6-2,0) yang salah diprediksi sebagai kelas(0,5-1,5). Secara keseluruhan, jumlah prediksi benar mencapai 81 dari 120 data uji, sehingga diperoleh akurasi klasifikasi sebesar 67,5%. Distribusi hasil klasifikasi ini relatif seimbang antar kelas, menunjukkan bahwa model tidak mengalami kecenderungan yang berat, terhadap salah satu kategori. Namun demikian, jumlah kesalahan klasifikasi silang masih cukup signifikan. Hal ini menunjukkan bahwa beberapa fitur tekstur yang dihasilkan oleh GLCM mungkin belum sepenuhnya mampu membedakan karakteristik antar kelas, terutama ketika nilai-nilai fitur memiliki overlap yang tinggi. Selain itu, algoritma KNN yang bergantung pada jarak Euclidean memiliki keterbatasan dalam memisahkan data yang rumit. Dengan demikian, meskipun performa model sudah menunjukkan kecenderungan klasifikasi yang cukup baik, masih terdapat ruang untuk peningkatan melalui pendekatan seperti peningkatan kualitas fitur, seleksi fitur yang lebih informatif, atau penggunaan algoritma klasifikasi yang lebih kompleks.

4.4 Evaluasi Kinerja Model (Classification Report)

Untuk mengevaluasi performa model KNN secara kuantitatif, dilakukan analisis menggunakan metrik evaluasi berupa precision, recal, dan f1-score. Evaluasi dilakukan pada nilai $K = 13$, yang sebelumnya diketahui memberikan akurasi terbaik. Hasil evaluasi ditampilkan pada Tabel 4.

Tabel.4 Hasil Evaluasi Model KNN ($K = 13$)

	precision	recall	F1-score	Support
Kandungan-nikotin-(0,5-1,5)	0.6842	0.6500	0.6667	60
Kandungan-nikotin-(1,6-2,0)	0.6667	0.7000	0.6829	60
Accuracy			0.6750	120
Macro avg	0.6754	0.6750	0.6748	120
Weighted avg	0.6754	0.6750	0.6748	120

Berdasarkan hasil evaluasi pada Tabel 4, diketahui bahwa model KNN memberikan akurasi sebesar 67,5% secara keseluruhan. Nilai precision dan recal antar kedua kelas tergolong cukup seimbang, menunjukkan bahwa model tidak terlalu bias terhadap kelas tertentu. Kelas dengan kandungan nikotin(1,6-2,0) menunjukkan recal yang sedikit lebih tinggi (0,7000) dibandingkan kelas (0,5-1,5) dengan hasil (0,6500), yang menunjukkan bahwa model lebih mampu mengenali sampel dari kelas tersebut.

4.5 Pembahasan

Hasil penelitian ini menunjukkan bahwa proses ekstraksi fitur menggunakan metode Gray Level Co-occurrence Matrix (GLCM) mampu menangkap karakteristik tekstur dari citra tembakau rokok secara representatif. Fitur-fitur statistik seperti contrast, correlation, energy, homogeneity, dissimilarity, ASM, dan entropy digunakan untuk merepresentasikan pola distribusi intensitas piksel yang mendasari struktur visual tembakau. Setiap fitur memiliki kontribusi tersendiri dalam membedakan antara dua kelas kandungan nikotin, yaitu (0,5–1,5) dan (1,6–2,0). Meski secara visual perbedaan antar kelas relatif tidak mencolok, pendekatan berbasis tekstur ini memberikan dasar diskriminatif bagi

proses klasifikasi. Penerapan algoritma K-Nearest Neighbors (KNN) sebagai model klasifikasi menghasilkan performa terbaik pada nilai $K = 13$, dengan akurasi klasifikasi sebesar 67,5%. Nilai ini diperoleh dari proses pengujian model terhadap variasi K dari 3 hingga 30 dengan interval ganjil, di mana akurasi mengalami ketidakstabilan sebelum stabil dan mencapai titik maksimum pada K tertentu.

Hal ini menunjukkan pentingnya pemilihan parameter K yang optimal dalam menjaga keseimbangan antara kompleksitas model dan tingkat kesalahan klasifikasi. Akurasi tersebut menunjukkan bahwa pendekatan hybrid GLCM-KNN dapat mengidentifikasi pola klasifikasi dengan tingkat ketepatan yang layak, meskipun masih terdapat ruang untuk perbaikan. Analisis terhadap confusion matrix menunjukkan bahwa model dapat mengklasifikasikan sebagian besar sampel secara benar pada kedua kelas. Sebanyak 39 citra dari kelas kandungan nikotin(0,5-1,5) dan 42 citra dari kelas kandungan nikotin(1,6-2,0) berhasil diklasifikasikan dengan akurasi 67,5%. Meski demikian, jumlah false negative dan false positif yang relatif seimbang (masing-masing 21 dan 18) menunjukkan bahwa model masih mengalami kesulitan dalam membedakan citra dengan nilai tekstur yang saling beririsan. Hal ini dapat mengindikasikan bahwa fitur tekstur yang dihasilkan GLCM belum sepenuhnya mampu menangkap aspek-aspek pembeda yang spesifik antar kelas.

Lebih lanjut, hasil evaluasi kuantitatif melalui metrik precision, recall, dan f1-score menunjukkan bahwa performa model relatif seimbang antar kelas. Kelas kandungan nikotin(1,6-2,0) memiliki nilai recall sebesar 0,7000, sedikit lebih tinggi dibandingkan kelas kandungan nikotin(0,5-1,5) (0,6500), yang menunjukkan kecenderungan model untuk lebih mampu mengenali ciri-ciri tekstur yang berkaitan dengan kadar nikotin yang lebih tinggi. Sedangkan nilai precision pada kedua kelas cukup setara 0,66–0,68. Keseimbangan ini menunjukkan bahwa model tidak mengalami kecenderungan terhadap salah satu kelas, yang turut diperkuat oleh nilai macro average dan weighted average yang konsisten.

Meskipun performa model tergolong moderat, pendekatan ini sudah memberikan kontribusi awal dalam pengembangan sistem

klasifikasi kandungan nikotin berbasis citra. Beberapa keterbatasan yang teridentifikasi, seperti tingkat akurasi yang belum tinggi dan masih adanya mis-klasifikasi, dapat diatasi melalui perluasan dataset, peningkatan resolusi citra, atau integrasi fitur GLCM dengan teknik ekstraksi lain yang bersifat komplementer.

Dengan mempertimbangkan jumlah data yang terbatas dan kompleksitas visual antar kelas yang tinggi, pendekatan hybrid GLCM-KNN ini dapat diposisikan sebagai baseline yang menjanjikan untuk klasifikasi awal citra daun tembakau berdasarkan kandungan nikotin.

5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem klasifikasi kandungan nikotin pada citra tembakau rokok dengan pendekatan hybrid menggunakan metode ekstraksi fitur tekstur Gray Level Co-occurrence Matrix (GLCM) dan algoritma klasifikasi K-Nearest Neighbor (KNN). Dari hasil ekstraksi fitur GLCM, ditemukan bahwa parameter entropy dan contrast mampu membedakan kelas kandungan nikotin yaitu kandungan nikotin (0,5–1,5) dan kandungan nikotin (1,6–2,0). Algoritma KNN menunjukkan performa klasifikasi pada nilai $K = 13$ dengan akurasi klasifikasi sebesar 67,5%.

Meskipun model belum mencapai tingkat akurasi yang tinggi, pendekatan ini telah menunjukkan penelitian awal dalam melakukan klasifikasi berdasarkan tekstur citra tembakau.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] GoodStats. (2024). Persentase perokok muda Indonesia kembali naik pada 2024. [Online]. Tersedia: <https://data.goodstats.id/statistic/persentase-perokok-muda-indonesia-kembali-naik-pada-2024-kh7Yz>
- [2] Lathifah, Q. A., Hermawati, A. H., & Putri, A. Y. (2020). Gambaran Nikotin Pada Perokok Pasif Di Kabupaten Tulungagung. *Borneo Journal of Medical Laboratory Technology*, 3(1), 178–183. <https://doi.org/10.33084/bjmlt.v3i1.1594>
- [3] Hermawati, A. H., Pratiwi, C. D., ST, S., & Lathifah, Q. A. Y. (2023). Nikotin, Tembakau, Dan Rokok. Penerbit Andi.
- [4] Lamasigi, Z. Y., & Lasena, Y. (2022). Identifikasi Tingkat Kesegaran Ikan Tuna Menggunakan Metode GLCM dan KNN. *Jambura Journal of Electrical and Electronics Engineering*, 4(1), 70–76.
- [5] Sari, W. S., & Sari, C. A. (2022). Klasifikasi Bunga Mawar Menggunakan KNN dan Ekstraksi Fitur GLCM dan HSV. *SKANIKA: Sistem Komputer dan Teknik Informatika*, 5(2), 145–156.
- [6] Sarimole, F. M., & Diadi, R. R. (2022). Klasifikasi Jenis Jamur Menggunakan Ekstraksi Fitur Glcm Dan K-Nearest Neighbor (Knn). *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 4(3), 286–290.
- [7] Firdiansyah, A. (2020). Optimalisasi Penerimaan Cukai HPTL Vape di Masa yang Akan Datang. *Optimalisasi Penerimaan Cukai HPTL Vape Di Masa Yang Akan Datang*, 4, 1–10.
- [8] Sriyanto, A., & Pangestu, A. P. (2022). Dampak Konsumsi Rokok Konvensional Dan Rokok Elektrik Terhadap Kesehatan, Penerimaan Negara. *Jurnal Perspektif Bea Dan Cukai*, 6(2), 428–450.
- [9] Lamasigi, Z. Y. (2021). DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN. *Jambura Journal of Electrical and Electronics Engineering*, 3(1), 1–6.
- [10] Situmorang, N. W. R. (2019). Penerapan Metode K-Nearest Neighbor Dalam Identifikasi Kesegaran Ikan (Skripsi, Universitas Sumatera Utara).
- [11] Lamasigi, Z. Y., Hasan, M., & Lasena, Y. (2020). Local Binary Pattern untuk Pengenalan Jenis Daun Tanaman Obat menggunakan K-Nearest Neighbor. *ILKOM Jurnal Ilmiah*, 12(3), 208–218.
- [12] Husdi, H. (2020). Sistem Cerdas dalam Mengidentifikasi Kematangan Buah Naga Berdasarkan Fitur Tekstur dengan Metode K-Nearest Neighbor. *ILKOM Jurnal Ilmiah*, 12(3), 225–232.
- [13] Khairunisa, N., & Jamaludin, A. (2024). Analisis Perbandingan Algoritma CNN dan YOLO dalam Mengidentifikasi Kerusakan Jalan. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3).
- [14] Wibowo, D. W., Erwanto, D., & Kusumastutie, D. A. W. (2021). Klasifikasi Jenis Kayu Menggunakan Ekstraksi Fitur Gray Level Co-Occurrence Matrix dan Multilayer Perceptron. *Jurnal Nasional Teknik Elektro*, 1–10.

- [15] Prabowo, A., Erwanto, D., & Rahayu, P. N. (2021). Klasifikasi Kesegaran Daging Sapi Menggunakan Metode Ekstraksi Tekstur GLCM dan KNN. *Electro Luceat*, 7(1), 74–81.
- [16] Ripa'i, A., Santoso, F., & Lazim, F. (2024). Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1749–1758.