PERBANDINGAN KINERJA KNN DAN SVM DALAM KLASIFIKASI SAMPAH ORGANIK DAN ANORGANIK MENGGUNAKAN EKSTRAKSI FITUR HOG DAN LBP

Muhammad Hidayatul Arifin, Ruth Amelia Vega S. Meliala,
Kristin Impana Manik, Aqilah Defiyanti, Hermawan Syahputra
Ilmu Komputer, Universitas Negeri Medan
Jl. William Iskandar, Ps. V, Medan Estate, Kab. Deli Serdang, Indonesia
arif.4233250005@mhs.unimed.ac.id

ABSTRAK

Pertumbuhan limbah yang semakin meningkat menimbulkan tantangan yang signifikan bagi upaya pelestarian lingkungan. Proses pemisahan sampah masih dilakukan secara manual dan sering kali tidak konsisten, yang merupakan kendala utama. Penelitian ini bertujuan untuk mengembangkan sistem yang menggunakan pengolahan citra untuk mengklasifikasikan sampah organik dan anorganik secara otomatis. Penelitian ini menggunakan dataset yang digunakan terdiri dari 1.800 citra dimana 900 organik dan 900 anorganik yang diekstraksi melalui *Histogram of Oriented Gradients* (HOG) dan *Local Binary Pattern* (LBP). Tahap *preprocessing*, yang mencakup pengubahan ukuran dan konversi ke *grayscale*. Selanjutnya, *Principal Component Analysis* (PCA) digunakan untuk mengurangi dimensi fitur HOG, kemudian digabungkan dengan fitur LBP dan diklasifikasikan menggunakan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN). Hasil pengujian menunjukkan bahwa SVM dengan kernel RBF memiliki akurasi tertinggi sebesar 88,89%, sementara KNN dengan nilai k=5 memiliki akurasi sebesar 83,61%. Keunggulan SVM terletak pada kemampuan mereka untuk memaksimalkan margin pemisahan. Hasilnya menunjukkan bahwa metode penggabungan HOG dan LBP dengan klasifikasi berbasis SVM dapat meningkatkan akurasi pemisahan sampah secara otomatis. Hasil ini dapat mendorong upaya untuk mengurangi beban di TPA serta meningkatkan praktik daur ulang yang berkelanjutan.

Kata kunci : *Pengolahan citra, Klasifikasi sampah, Histogram of Oriented Gradients, Local Binary Pattern, K-Nearest Neighbor, Support Vector Machine*

1. PENDAHULUAN

Sampah menjadi masalah lingkungan yang semakin penting untuk ditangani secara menyeluruh. Produksi sapah terus meningkat seiring pertumbuhan populasi dan perubahan gaya hidup masyrakat. Data dari berbagai lembaga lingkungan menunjukkan bahwa jumlah sampah yang tidak terkelola dengan baik dapat dapat menyebabkan pencemaran udara, air, dan tanah serta kontribusi terhadap pemanasan global. Oleh karena itu, pengelolaan sampah yang efisien dan efektif menjadi prioritas utama bagi pemerintah dan masyarakat.

Pemisahan sampah ke dalam kategori yang sesuai adalah langkah awal dalam pengelolaan sampah yang berkelanjutan. Ini termasuk sampah organik dan anorganik. Sampah organik, seperti sisa makanan dan daun, dapat didaur ulang dan dibuat kompos, sementara sampah anorganik, seperti kaca, plastik, dan logam, membutuhkan proses daur ulang atau penanganan khusus agar tidak mencemari lingkungan. Namun, pemilahan sampah masih sering dilakukan secara manual dan bergantung pada kesadaran masyarakat, sehingga prosesnya tidak konsisten.

Dengan kemajuan dalam pengolahan gambar dan pembelajaran mesin, proses pemilahan sampah dapat dioptimatisasi. Sistem dapat mengidentifikasi karakteristik visual sampah organik dan anorganik. Seperti pada penelitian ini menggunakan ekstraksi fitur HOG dan LBP dengan algoritma klasifikasi seperti KNN dan SVM untuk mengukur efektivitas masing-masing algoritma dalam mengklasifikasikan sampah berdasarkan skor F1, akurasi, *precision*, dan *recall*.

Studi sebelumnya telah menyelidiki penerapan algoritma SVM dan KNN pada berbagai klasifikasi citra. Misalnya, penelitian menemukan bahwa algoritma SVM mencapai akurasi 87,50%, presisi 87,94%, recall 87,50%, dan skor F1 87,50% pada gambar infeksi telinga, sedangkan algoritma KNN hanya mencapai akurasi 72,50% dengan metrik evaluasi yang lebih rendah [1]. Pada penelitian deteksi dini penyakit jantung pada penderita hipertensi, menemukan bahwa SVM memiliki akurasi sebesar 81,9% dibandingkan dengan 63% yang diperoleh oleh KNN[2]. Di sisi lain, suatu penelitian mengembangkan sistem identifikasi jenis tanaman berbasis daun dengan KNN dengan akurasi rata-rata 94,44%, dengan hasil 100% pada beberapa jenis daun dan 83% pada jenis lain [3]. Hasilnya menunjukkan bahwa algoritma tertentu memiliki keunggulan dalam aplikasi tertentu di berbagai bidang.

Tujuan penelitian adalah untuk membuat sistem klasifikasi sampah yang efisien yang menggunakan metode HOG dan LBP, serta untuk menganalisis dan membandingkan kinerja algoritma KNN dan SVM untuk mendukung pengelolaan sampah yang lebih berkelanjutan dan ramah lingkungan.

2. TINJAUAN PUSTAKA

2.1. Sampah

Salah satu masalah terbesar yang dihadapi umat manusia adalah sampah, yang semakin memburuk

seiring dengan bertambahnya populasi dan pergeseran kebiasaan konsumsi. Sampah dapat menimbulkan sejumlah dampak yang merugikan jika tidak dikelola dengan baik, termasuk menurunkan daya tarik estetika suatu daerah, menyebarkan penyakit, dan mencemari ekosistem. Meskipun pengelolaan sampah sering kali masih di bawah standar, peningkatan jumlah sampah juga meningkatkan kebutuhan akan lokasi pembuangan yang sesuai. Ketika sampah dibiarkan menumpuk, sampah dapat melepaskan gas beracun seperti metana (CH4), yang mencemari udara dan menyebabkan pemanasan global. Oleh karena itu, untuk mengurangi timbulan sampah dan meningkatkan sistem pengelolaan sampah, diperlukan kesadaran dan langkah-langkah praktis [4].

Memilah sampah berdasarkan jenisnya yaitu sampah organic dari sampah anorganik dan sampah B3 (Bahan Berbahaya dan Beracun) merupakan tahap yang krusial dalam pengelolaan sampah. Sampah anorganik, seperti plastik dan logam, harus didaur ulang untuk mencegah pencemaran lingkungan, sedangkan sampah organik, seperti sisa makanan dan daun, dapat terurai secara alami dan dimanfaatkan sebagai kompos. Untuk mencegah dampak negatif terhadap kesehatan, sampah berbahaya, seperti baterai bekas dan sampah elektronik, perlu dikelola dengan hati-hati. Sangat penting bagi setiap orang untuk memahami dan mendapatkan edukasi mengenai pengelolaan sampah yang efektif agar dapat menggunakan prinsip 3R (Reduce, Reuse, Recycle), yang akan meminimalisir jumlah sampah yang berakhir di tempat pembuangan akhir dan mengurangi dampak buruknya terhadap lingkungan [5].

2.2. Pengolahan Citra (*Image Processing)*

Tujuan dari pengolahan citra digital yang berkembang pesat dalam ilmu komputer dan teknologi informasi adalah untuk meningkatkan kualitas gambar, mengekstrak informasi berharga, dan melakukan analisis rumit yang tidak mungkin dilakukan secara manual. Metode-metode ini sangat penting untuk banyak aplikasi, termasuk ekstraksi fitur, deteksi objek, dan segmentasi gambar. Dengan mengubah gambar menjadi biner berdasarkan nilai *threshold* tertentu, teknik *thresholding* merupakan salah satu teknik mendasar dalam pemrosesan gambar digunakan untuk membedakan objek dari latar belakang. *Thresholding* ini bisa bersifat adaptif, yang memodifikasi *threshold* menurut intensitas piksel lokal, atau global yang menerapkan nilai *threshold* tunggal ke seluruh gambar. Teknik ini bekerja dengan baik dalam berbagai situasi, khususnya apabila latar belakang gambar sangat kontras dengan benda-benda dalam gambar [6].

Selain *thresholding*, metode lain yang digunakan dalam pemrosesan gambar termasuk pengelompokkan warna dengan menggunakan metode *Hierarchical Agglomerative Clustering* (HAC). Teknik ini dapat digunakan untuk mengidentifikasi dan mengkategorikan item dalam foto digital dengan mengelompokkan piksel menurut kedekatan warna. Algoritma tertentu digunakan dalam pemrosesan gambar digital untuk mengekstrak informasi penting dari foto, termasuk warna, tekstur, dan bentuk. Metode ini sering digunakan dalam berbagai domain, termasuk sistem pengawasan dan keamanan, penginderaan jarak jauh, dan pencitraan medis. Untuk meningkatkan akurasi pengenalan item, diperlukan teknik lebih lanjut karena pemrosesan gambar ditantang oleh ketidakmampuannya untuk membedakan objek dengan warna dominan yang sama [7].

2.3. *Histogram of Oriented Gradients* (HOG)

Teknik ekstraksi fitur yang disebut *Histogram of Oriented Gradients* (HOG) digunakan untuk mengidentifikasi fitur bentuk objek dalam gambar digital. Nilai gradien setiap piksel ditentukan dengan menggunakan metode ini, dan hasilnya kemudian ditransformasikan ke dalam histogram sel kecil orientasi gradien. Histogram ini secara lebih tepat menggambarkan bentuk dan struktur objek dengan merefleksikan distribusi arah tepi dalam gambar. Manfaat utama HOG adalah kemampuannya untuk mengidentifikasi pola dalam gambar, bahkan ketika terjadi perubahan pencahayaan dan sudut pandang. Dengan menormalkan karakteristik dalam blok yang lebih besar, gambar menjadi lebih tangguh terhadap perubahan pencahayaan dan faktor lainnya [8].

Pendekatan HOG juga memiliki keuntungan karena mampu mengekstrak fitur dari objek dengan perbedaan intensitas yang rumit dan tangguh terhadap perubahan pencahayaan. Prosedur pembobotan digunakan untuk membuat perbedaan fitur antara berbagai area gambar menjadi lebih jelas setelah langkah normalisasi, yang melibatkan penggabungan nilai histogram dari beberapa sel di dekatnya ke dalam blok. Prosedur ini memungkinkan pendekatan HOG untuk mengekstrak fitur-fitur penting dari gambar, yang kemudian dapat dimasukkan ke dalam algoritma klasifikasi seperti *K-Nearest Neighbor* (KNN) atau *Support Vector Machine* (SVM). Kombinasi ini menjadikan HOG sebagai salah satu teknik ekstraksi fitur yang paling efisien untuk berbagai aplikasi visi komputer dan pemrosesan gambar digital [9].

2.4. *Local Binary Patterns* (LBP)

Dalam pemrosesan gambar digital, teknik ekstraksi fitur tekstur Local Binary Pattern (LBP) digunakan untuk mengidentifikasi pola berdasarkan variasi intensitas piksel dalam suatu wilayah terbatas. Metode ini memberikan label biner berdasarkan hasil perbandingan antara setiap piksel dalam sebuah gambar dengan piksel-piksel di sekitarnya. Histogram tekstur yang menggambarkan gambar kemudian dibuat menggunakan pola biner yang telah dibuat dan ditransformasikan ke dalam bentuk desimal. Menangkap informasi tekstur lokal secara efektif adalah salah satu manfaat utama LBP, yang membuatnya sesuai untuk berbagai aplikasi pengenalan pola dan klasifikasi gambar [10].

LBP terus menjadi salah satu teknik ekstraksi fitur tekstur yang paling banyak digunakan dalam pemrosesan citra digital karena kombinasi ini, yang membuat sistem lebih mudah beradaptasi dengan situasi pencahayaan yang berbeda dan variasi bentuk [11]. Berikut adalah rumus Local Binary Pattern (LBP) yang digunakan untuk mengekstrak fitur tekstur dalam pengolahan citra digital, berdasarkan penelitian yang dilakukan:

$$LBP_{\{P,R\}(x,y)} = \sum_{i=0}^{\{P-1\}} s(Gi - Gc) \cdot 2^i \quad (1)$$

Keterangan :
$LBP_{\{P,R\}(x,y)}$ = Nilai LBP pada Koordinar piksel (x, y)
P = Jumlah piksel tetangga
R = Radius dari piksel pusat
Gc = Intensitas piksel pusat
Gi = Intensitas piksel tetangga ke - i
$s(x)$ = Fungsi pembanding yang menghasilkan nilai biner 0 atau 1

2.5. *Principal Component Analysis* (PCA)

Sebuah metode statistik untuk mengurangi maulanadimensi data multivariat dengan tetap mempertahankan informasi yang paling penting disebut analisis komponen utama, atau PCA. Kumpulan variabel yang berpotensi terkait diubah menjadi satu set variabel baru yang tidak berkorelasi, yang dikenal sebagai komponen utama, dengan PCA. Variasi terbesar dari data ditemukan pada komponen utama pertama (PC1), yang diikuti oleh komponen kedua (PC2), dan seterusnya. Teknik ini sering digunakan dalam analisis data berskala besar, termasuk pemrosesan gambar dan pengelompokan data berdasarkan karakteristik tertentu, karena teknik ini berusaha menyederhanakan variabel yang diamati tanpa mengubah variabel aslinya secara signifikan [12].

Ketika PCA diterapkan, PCA tidak hanya membantu menyederhanakan data, tetapi juga meningkatkan efektivitas analisis, terutama ketika dikombinasikan dengan teknik tambahan seperti pengelompokan berbasis partisi. Dengan menyimpan informasi yang paling penting dalam sejumlah kecil komponen utama, pengurangan dimensi PCA dapat mengoptimalkan pendekatan pengelompokan dan memungkinkan analisis yang lebih cepat dan tepat. Normalisasi data, perhitungan nilai eigen, perhitungan matriks kovarian, dan transformasi ke dalam sistem koordinat baru dengan varians maksimum merupakan langkah-langkah dalam proses tersebut. Hasilnya, PCA dapat menemukan pola dalam data dan meningkatkan ketepatan pemetaan atribut fenomena [13]. PCA digunakan untuk mereduksi dimensi dengan memilih komponen utama yang memiliki varians terbesar, sehingga informasi dari data awal tetap terjaga dengan jumlah variabel yang lebih sedikit [13].

2.6. *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor (KNN) adalah salah satu metode klasifikasi dalam data mining yang digunakan untuk mengelompokkan data baru berdasarkan kedekatannya dengan data sebelumnya. Metode ini termasuk dalam kategori *supervised learning* dan menentukan kelas suatu data berdasarkan mayoritas dari sejumlah K tetangga terdekat, yang dihitung menggunakan ukuran jarak seperti *Euclidean Distance*. KNN dikenal karena kesederhanaannya, kemampuannya menangani data berjumlah besar, serta cukup tahan terhadap data yang mengandung noise. Meskipun begitu, KNN juga memiliki keterbatasan, antara lain pemilihan nilai K yang tepat sangat memengaruhi hasil klasifikasi, dan perhitungan jarak yang banyak dapat menyebabkan beban komputasi yang tinggi [14].

Sementara itu, secara teknis, algoritma KNN bekerja dengan cara mengklasifikasikan suatu objek berdasarkan sejumlah data latih yang memiliki jarak paling dekat dengannya. Nilai K dalam algoritma ini harus dipilih secara cermat, yaitu lebih dari satu, bernilai ganjil, serta tidak melebihi jumlah data latih yang tersedia. Untuk mengukur kedekatan antara objek satu dengan lainnya, biasanya digunakan rumus jarak *Euclidean*, yang menjadi dasar dalam menentukan klasifikasi akhir suatu data baru berdasarkan tetangga terdekatnya [15]. Berikut adalah rumus KNN untuk menghitung jarak menggunakan *Euclidean Distance*:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (2)$$

Keterangan:
x: Data uji (data baru yang ingin diklasifikasikan).
y: Data latih (data yang sudah diketahui kelasnya).
$x1, x2, \ldots, xn$: Fitur dari data uji.
$y1, y2, \ldots, yn$: Fitur dari data latih.
$d(x, y)$: Jarak antara data uji dan data latih.

2.7. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah algoritma dalam machine learning yang digunakan untuk menyelesaikan masalah klasifikasi dan regresi. SVM merupakan metode klasifikasi yang semi-eager learner, Support Vector Machine tidak hanya membutuhkan proses pelatihan, tetapi juga menyimpan beberapa data latih untuk digunakan kembali selama proses prediksi [16]. Tujuan utama dari SVM adalah menemukan sebuah hyperplane atau garis pemisah terbaik yang dapat membedakan dua kelas data secara maksimal. SVM bekerja dengan memaksimalkan margin atau jarak antara data yang paling dekat dengan garis pemisah, yang disebut sebagai *support vector*. Untuk memisahkan data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel yang memetakan data ke ruang berdimensi lebih tinggi. SVM dikenal memiliki performa yang baik, mampu menangani data dalam jumlah besar maupun kecil, serta efektif pada data dengan banyak atribut [17].

Dalam algoritma *Support Vector Machine* (SVM), proses klasifikasi bertujuan untuk mencari *hyperplane* optimal yang dapat memisahkan dua kelas data secara maksimal. *Hyperplane* tersebut dinyatakan dengan persamaan garis pemisah [18]. Berikut merupakan rumus dari SVM yang menggunakan persamaan *hyperlane:*

$$f\ (x) = w\ \times x + b \tag{3}$$

Keterangan:
$f\ (x)$: Fungsi klasifikasi untuk menentukan kelas data.
w : Vektor bobot (parameter yang menentukan arah pemisah/garis batas antar kelas).
x: Data input (fitur-fitur dari data yang diuji).
b : Bias (intersep, penggeser posisi hyperplane).

2.8. *Confusion Matrix*

Confusion matrix merupakan tabel yang menyajikan jumlah data uji yang berhasil diklasifikasikan dengan benar maupun yang salah. Dengan bantuan tabel ini, evaluasi terhadap tingkat akurasi sistem klasifikasi menjadi lebih mudah dan terperinci. Melalui *confusion matrix*, kita dapat memahami secara menyeluruh performa sistem serta mengetahui letak kesalahan klasifikasi yang terjadi. Secara umum, *confusion matrix* adalah metode yang sederhana namun efektif untuk mengukur dan mengevaluasi kualitas hasil klasifikasi suatu sistem [19]. Rumus perhitungan *Confusion matrix* adalah sebagai berikut:

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \tag{4}$$

Keterangan:
a. *True Positive* (TP) menunjukkan jumlah data yang diklasifikasikan dengan benar sebagai kelas positif.
b. *True Negative* (TN) menunjukkan jumlah data yang diklasifikasikan dengan benar sebagai kelas negatif.
c. *False Positive* (FP) adalah jumlah data yang sebenarnya merupakan kelas negatif, namun salah diklasifikasikan sebagai kelas positif.
d. *False Negative* (FN) adalah jumlah data yang sebenarnya termasuk kelas positif, namun salah diklasifikasikan sebagai kelas negatif.

Confusion Matrix digunakan sebagai dasar untuk menghitung nilai *accuracy*, *precision*, *recall*, dan F1-*score*.

Akurasi merupakan salah satu ukuran evaluasi yang digunakan untuk mengetahui tingkat keberhasilan suatu model dalam mengklasifikasikan data secara tepat. *Precision* adalah salah satu metrik evaluasi yang digunakan untuk mengukur seberapa besar proporsi prediksi positif yang benar dibandingkan dengan seluruh data yang diprediksi sebagai positif. *Recall* merupakan metrik evaluasi yang digunakan untuk mengukur seberapa besar jumlah data positif yang berhasil diprediksi dengan benar dibandingkan dengan keseluruhan data yang sebenarnya termasuk dalam kelas positif. [20]. F1-*score* adalah metrik yang diperoleh dari kombinasi antara nilai *precision* dan *recall*. Berikut adalah rumus untuk melakukan evaluasi:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

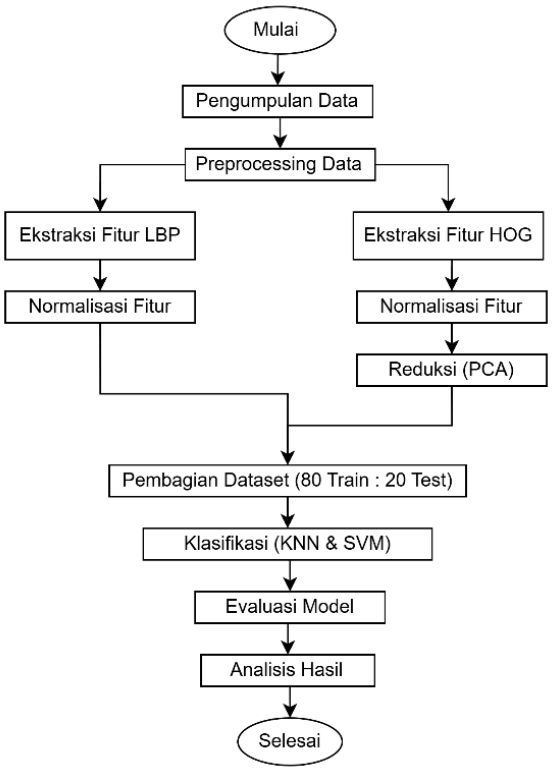
$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision\times Recall} \tag{8}$$

3. METODE PENELITIAN

3.1. Flowchart



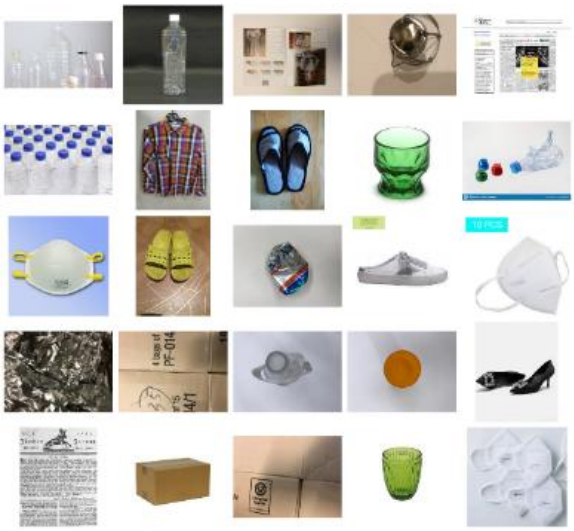
Gambar 1. Flowchart Penelitian

Pada Gambar 1 menunjukkan proses penelitian dalam pengolahan citra yang terstruktur, dimulai dengan pengumpulan data mentah yang kemudian menjalani tahap preprocessing untuk mengubah ukuran dan mengkonversi citra menjadi grayscale, sehingga data tersebut siap untuk dianalisis. Selanjutnya, fitur diekstraksi menggunakan metode HOG untuk mendeteksi orientasi gradien dan LBP untuk menggambarkan tekstur lokal, di mana hasil dari masing-masing fitur dinormalisasi agar tetap konsisten pada skala yang sama. Dimensi pada HOG dikurangi menggunakan PCA untuk menghindari kemungkinan overfitting dan ketidakseimbangan dengan dimensi kecil dari fitur LBP, setelah itu kedua fitur tersebut

digabungkan. Data kemudian dibagi menjadi 80% untuk data latih dan 20% untuk data uji, bertujuan untuk melatih serta menguji model klasifikasi yang dibuat dengan algoritma KNN dan SVM. Pada tahap akhir, kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk menjamin efektivitas dalam pengenalan pola citra.

3.2. Pengumpulan Data

Gambar 2. Citra Organik



Gambar 3. Citra Anorganik

Data yang digunakan saat ini berasal dari salah satu website open souce yaitu Kaggle yang menawarkan berbagai gambar sampah organik dan anorganik. 1.800 gambar digunakan, dengan 900 gambar untuk sampah organik dan 900 gambar untuk sampah anorganik. Sampah organik termasuk sisa buah-buahan, sayuran, roti, kulit pisang, dan bahan makanan lain yang mudah terurai. Sampah anorganik termasuk berbagai jenis material seperti plastik, kertas, kaca, logam, dan barang lain yang tidak mudah terurai. Untuk memberikan gambaran tentang keragaman dan fitur visual masing-masing kelas sampah, semua contoh gambar yang digunakan dalam penelitian ini dilampirkan.

3.3. *Preprocessing* Data

Sebelum dilakukan ekstraksi fitur, citra pada dataset akan melewati proses preprocessing. Pertama, gambar diresize menjadi ukuran 256 × 256 piksel untuk menyamakan dimensi dan memudahkan proses komputasi. Kemudian, informasi warna diubah menjadi intensitas piksel saja dengan mengubah gambar menjadi grayscale. Selain itu, langkah ini membantu proses memfokuskan pada pola tekstur dan bentuk objek di dalam gambar, yang umumnya lebih relevan untuk pemodelan klasifikasi. Tujuan dari proses preprocessing yang konsisten di seluruh gambar adalah untuk memastikan bahwa data masukan memiliki format dan karakteristik yang seragam, yang memudahkan tahap ekstraksi fitur dan pelatihan model.

3.4. Ekstraksi Fitur (HOG dan LBP)

Histogram Orientasi Gradients (HOG) digunakan untuk ekstraksi fitur untuk mengumpulkan data tentang orientasi tepi dan bentuk. Dihitung gradien citra grayscale, dibagi menjadi sel-sel kecil, dan kemudian dibuat histogram yang menunjukkan arah gradien. Untuk meningkatkan ketahanan terhadap perubahan pencahayaan, histogram ini dinormalisasi per blok. Hasil akhir adalah vektor fitur, yang menunjukkan struktur bentuk pada gambar.

Local Binary Patterns (LBP) juga digunakan untuk meningkatkan informasi tekstur. Metode ini menghasilkan pola biner yang dirangkum ke dalam histogram setelah nilai intensitas piksel pusat dan piksel di sekitarnya dibandingkan. Sebelum algoritma klasifikasi KNN dan SVM dapat digunakan, kombinasi HOG untuk bentuk dan LBP untuk tekstur memungkinkan penampilan lebih lengkap dari atribut citra yang penting, baik pola lokal maupun struktur global.

3.5. Normalisasi Fitur

Pada tahap ini, fitur HOG dan LBP telah distandardisasi dengan menggunakan StandardScaler. Rumus $z = \frac{x - \mu}{\sigma}$ digunakan untuk mengubah semua nilai fitur sehingga menghasilkan distribusi dengan mean 0 dan standar deviasi 1. Proses ini menjamin bahwa skala fitur seragam dan mencegah dominasi fitur dengan nilai yang signifikan dalam analisis tambahan, seperti reduksi dimensi dan klasifikasi. Teknik normalisasi diterapkan secara terpisah pada kedua jenis fitur untuk mencapai keseimbangan kontribusi informasi.

3.6. Reduksi Dimensi HOG dengan Principal Component Analysis (PCA)

Ketika *Histogram Gradients Oriented* (HOG) dan *Local Binary Patterns* (LBP) digabungkan, risiko overfitting meningkat dan ketidakseimbangan LBP yang kecil menjadi kurang berpengaruh karena dimensi fitur HOG yang lebih besar. Vektor HOG yang sangat panjang juga membutuhkan lebih banyak ruang penyimpanan dan memperlambat proses

komputasi. Akibatnya, untuk mengidentifikasi variansi yang paling signifikan sambil menghilangkan informasi yang tidak penting, dimensi vektor HOG dikurangi. Dengan menggunakan metode Principal Component Analysis (PCA), fitur HOG dibagi menjadi lebih sedikit komponen. Akibatnya, skala dimensi menjadi lebih seimbang dengan LBP dan kinerja klasifikasi meningkat. Dengan cara ini, model dapat fokus pada informasi bentuk HOG yang paling penting sambil menghindari dimensi yang berlebihan.

3.7. Pembagian Dataset

Pada tahap ini, dataset yang telah melewati ekstraksi fitur dipisahkan menjadi dua segmen, yaitu data latih dan data uji. Proporsi yang diterapkan adalah 80% untuk data latih dan 20% untuk data uji. Tujuan dari pemisahan ini adalah agar model dilatih dengan sebagian besar data, kemudian diuji dengan data yang belum pernah dikenali sebelumnya. Dengan cara ini, kinerja model yang dievaluasi akan lebih merefleksikan kemampuannya untuk menggeneralisasi pada data yang baru.

3.8. Klasifikasi dan Pelatihan Model (KNN dan SVM)

Pada tahap ini, dataset yang telah melewati ekstraksi fitur dipisahkan menjadi dua segmen, yaitu data latih dan data uji. Proporsi yang diterapkan adalah 80% untuk data latih dan 20% untuk data uji. Tujuan dari pemisahan ini adalah agar model dilatih dengan sebagian besar data, kemudian diuji dengan data yang belum pernah dikenali sebelumnya. Dengan cara ini, Penelitian ini menggunakan dua motode klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), untuk membedakan antara sampah organik dan anorganik. Pada metode KNN, setiap data yang diuji diklasifikasikan dengan melihat data latih yang lebih dekat berdasarkan ukuran jarak tertentu seperti jarak Euclidean. Dalam studi ini, beberapa jumlah tetangga (k) ditetapkan untuk menentukan kategori data yang diuji.

Di sisi lain, SVM bertujuan untuk menemukan pemisah (*hyperplane)* yang optimal, sehingga jarak antar kelas (*margin*) dapat diperbesar. Parameter C diatur untuk menyeimbangkan antara leabar *margin* dan tingkat kesalahan dalam klasifikasi. Dalam penelitian ini, digunakan kernel RBF dengan nilai C = 1 dan gamma = "*scale*" agar data dapat dipindahkan ke ruang yang memiliki dimensi lebih tinggi, sehingga pemisahan kelas jadi lebih efisien. Dengan kedua metode ini, proses pelatihan model dilakukan dengan data latih yang telah melalui tahap ekstraksi fitur, sedangkan data uji digunakan untuk mengukur kinerja akhir dari model yang telah dibuat.

3.9. Evaluasi Model

Evaluasi model dilakukan dengan mengukur metrik seperti akurasi, precision, recall, dan F1-score. *Precision* mengukur persentase prediksi positif yang benar, sedangkan *recall* mengukur kemampuan model mendeteksi sampel positif. F1-*score* merupakan rata-rata harmonis dari *precision* dan *recall.* Metrik-metrik tersebut digunakan untuk membandingkan kinerja model KNN dan SVM dalam mengklasifikasikan sampah organik dan anorganik.

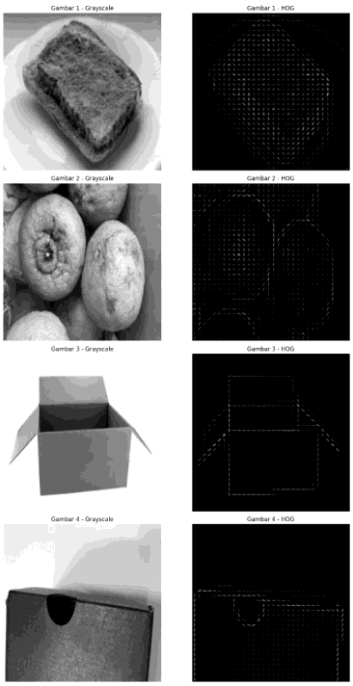
4. HASIL DAN PEMBAHASAN

4.1. *Preprocessing* Data

Gambar 4. Tahapan Preprocessing pada Citra Sampah

Tahap *preprocessing* yang digunakan pada gambar menunjukkan kondisi warna yang sebenarnya, gambar asli ditampilkan dalam format RGB. Kemudian, gambar diubah menjadi resolusi 256 x 256 piksel agar setiap gambar memiliki ukuran yang sama, sehingga proses ekstraksi fitur dan pelatihan model dapat dilakukan secara konsisten. Untuk menyederhanakan informasi warna menjadi intensitas piksel, gambar diubah ke format *grayscale*. Oleh karena itu, *preprocessing* ini membantu menekankan bentuk dan tekstur objek dalam gambar sekaligus mengurangi kompleksitas data, sehingga model dapat lebih dapat fokus pada karakteristik visual yang relevan untuk proses klasifikasi.

4.2. Ekstraksi Fitur HOG



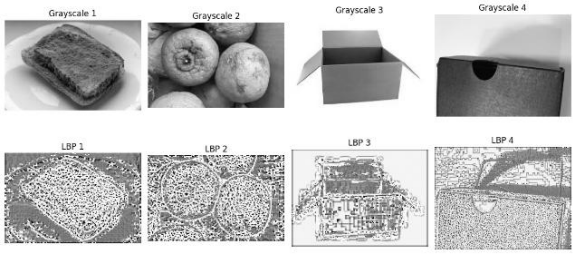
Gambar 5. Contoh Hasil Ekstraksi Fitur HOG

Setelah tahap *preprocessing* selesai, dimana setiap citra diubah ke format *grayscale* dan fiturnya diekstraksi menggunakan metode Histogram of

Oriented Gradients (HOG). HOG tidak hanya menghasilkan vektor fitur, tetapi juga dapat divisualisasikan untuk menunjukkan pola gradien di setiap piksel dalam bentuk garis putih yang menunjukkan arah dan intensitas perubahan intensitas piksel.

Pada Gambar 5 di atas menunjukkan 4 gambar (di kolom kiri) beserta representasi HOG (di kolom kanan). Meskipun perbedaan antar kelas tidak selalu tampak jelas secara visual, data gradien yang diperoleh dari HOG secara efisien menangkap pola tepi dan kontur objek. Data ini selanjutnya dapat digunakan oleh algoritma klasifikasi seperti KNN dan SVM untuk membedakan antara sampah organik dan anorganik.

4.3. Ekstraksi Fitur LBP



Gambar 6. Contoh Hasil Ekstraksi Fitur LBP

Pada Gambar 6 empat gambar di baris pertama dari atas adalah gambar grayscale dari empat objek yang berbeda: pertama adalah sampah organik seperti sisa roti, kedua adalah buah, ketiga adalah boks atau kardus kosong, dan keempat adalah bagian kardus yang berbeda. Intensitas abu-abu dalam setiap gambar grayscale menunjukkan pola tekstur dan bentuk secara keseluruhan.

Di baris kedua, semua gambar telah diubah dengan Local Binary Patterns (LBP). LBP membuat pola biner dengan membandingkan intensitas piksel pusat dengan piksel tetangga dalam radius tertentu, yang kemudian divisualisasikan. Pola biner tersebut seperti "titik-titik" atau area kontras di mana intensitas piksel di sekitar piksel pusat lebih tinggi atau lebih rendah daripada piksel pusat. Oleh karena itu, gambar LBP menekankan pola tekstur lokal dari setiap objek. Ini termasuk struktur permukaan roti, buah, dan kardus. Selanjutnya, hasil transformasi ini dapat digunakan sebagai fitur dalam proses klasifikasi atau analisis tekstur.

4.4. Hasil Normalisasi

Pada langkah ini, kedua jenis fitur yang telah diekstraksi, HOG dan LBP, dinormalisasi dengan menggunakan StandardScaler dari scikit-learn. Tujuan normalisasi ini adalah untuk menyeragamkan skala data dengan memaksa nilai rata-rata, atau mean, dari masing-masing fitur menjadi 0 dan standar deviasi, atau standard deviasi, menjadi 1.

Tabel 1. Hasil Normalisasi

Fitur	Rata – Rata Setelah Normalisasi	Standar Deviasi Setelah Normalisasi
HOG	1.27×10^{-16}	0.9999999999992
LBP	-3.41×10^{-16}	1.0

Pada Tabel 1 di atas menunjukkan bahwa normalisasi telah berhasil, karena rata-rata kedua fitur mendekati 0 dan standar deviasi mendekati 1. Oleh karena itu, data sudah dalam skala yang seimbang dan siap untuk diproses lebih lanjut. Ini termasuk menerapkan Principal Component Analysis (PCA) pada karakteristik HOG dan proses klasifikasi.

4.5. Reduksi Dimensi HOG dan Penggabungan Fitur

Setelah tahap normalisasi selesai, Principal Component Analysis (PCA) digunakan untuk melakukan reduksi dimensi pada fitur HOG. Metode ini mencari kombinasi variabel linear, juga dikenal sebagai fitur HOG, untuk memaksimalkan variansi data. Studi ini menetapkan jumlah komponen (n_components) menjadi 30. Dengan demikian, fitur HOG yang sebelumnya memiliki 34.596 dimensi per gambar dapat dikurangi menjadi 30 komponen utama. Karena data berukuran terlalu besar seringkali menyulitkan proses pembelajaran mesin, reduksi dimensi ini sangat penting untuk mengurangi beban komputasi dan risiko overfitting.

Tabel 2. Dimensi Fitur Ekstraksi

Jenis Fitur	Dimensi Sebelum PCA	Dimensi Setelah PCA	Dimensi Setelah Penggabungan
HOG	(1800, 34596)	(1800, 30)	-
LPB	(1800, 26)	-	-
Gabungan (HOG + LBP)	-	-	(1800, 56)

Dengan menggabungkan hasil PCA (30 dimensi) dan LBP (26 dimensi) fitur HOG, vektor fitur akhir berukuran 56 per citra diperoleh. Ini secara signifikan mengurangi beban komputasi dan risiko *overfitting* sekaligus menjaga informasi penting untuk proses klasifikasi.

4.6. Pembagian Dataset

Dataset dibagi menjadi dua kelompok yaitu data latih dan data uji, dengan rasio 80:20. Tujuan dari proses ini adalah untuk membedakan data yang akan digunakan untuk pengajaran (training) dan data yang akan digunakan untuk menguji kinerja model. Tabel berikut menunjukkan pembagian dataset ini.

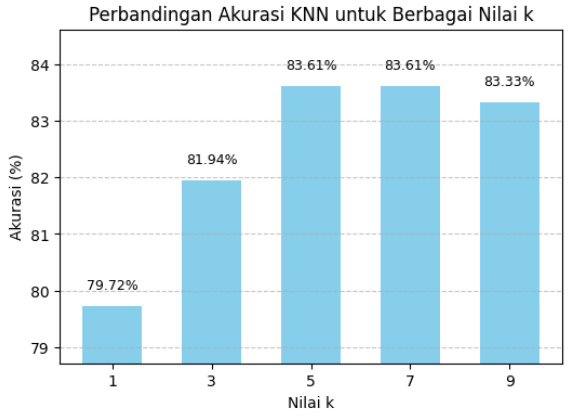
Tabel 3. Jumlah Data Latih dan Uji

Data Latih	1440
Data Uji	360

Pembagian ini memungkinkan model untuk dilatih pada sebagian besar data yang tersedia (1440 sampel) dan dievaluasi menggunakan data uji yang terpisah (360 sampel), sehingga pengukuran performa model dapat dilakukan secara lebih objektif.

4.7. Hasil dan Evaluasi Model Klasifikasi

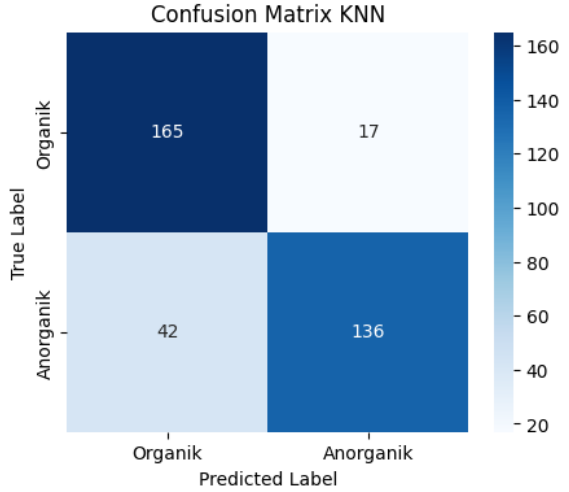
Pada penelitian ini terlebih dahulu melakukan percobaan pada beberapa nilai k (1, 3, 5, 7, dan 9). Ini mengevaluasi kinerja K-Nearest Neighbor (KNN). Perbandingan akurasi yang diperoleh untuk masing-masing "k" ditunjukkan pada Gambar 7.



Gambar 7. Perbandingan Akurasi KNN untuk Berbagai Nilai k

Berdasarkan Gambar 7, tingkat akurasi paling tinggi terlihat pada k = 5 dan k = 7, yaitu 83,61%. Angka "k" yang lebih rendah, seperti "k=1", cenderung membuat model terlalu sensitive terhadap noise, sedangkan angka "k" yang terlalu tinggi bisa mengurangi sensitivitas model terhadap pola minoritas, sehingga performanya tidak maksimal.

Selanjutnya, model KNN dengan "k=5" diuji untuk mendapatkan hasil yang lebih mendalam. *Confusion matrix* yang ada pada Gambar 7.



Gambar 8. Confusion Matrix KNN (k=5)

Menurut Confusion Matrix Gambar 8, model KNN mencapai akurasi keseluruhan sebesar 83,61%. Nilai recall yang lebih tinggi untuk kelas organik menunjukkan bahwa model ini lebih baik dalam mengidentifikasi sampah organik daripada sampah anorganik.

```
=== Evaluasi KNN ===
              precision    recall  f1-score   support

     Organik       0.80      0.91      0.85       182
   Anorganik       0.89      0.76      0.82       178

    accuracy                           0.84       360
   macro avg       0.84      0.84      0.84       360
weighted avg       0.84      0.84      0.84       360
```

Gambar 9. Classification Report KNN (k=5) pada Data Uji

Hasil menunjukkan bahwa KNN dengan k=5 memiliki akurasi mendekati 84%, dengan recall kelas organik lebih tinggi (0,91) daripada recall kelas anorganik (0.76), menunjukkan bahwa model cenderung lebih tepat mengenali sampah organik, sementara recall kelas anorganik sedikit cenderung salah dalam mengklasifikasikannya.

Setelah melakukan pengujian KNN, penelitian berlanjut dengan menerapkan *Support Vector Machine* (SVM). Dalam studi ini, digunakan kernel RBF dengan parameter C dan gamma yang ditentukan melalui percobaan awal seperti *grid search* atau metode lainnya.

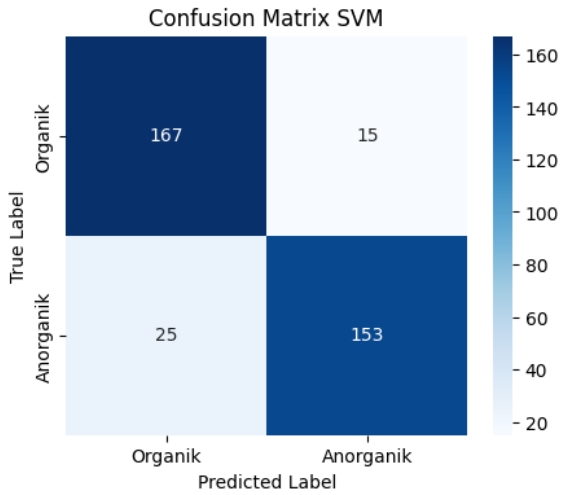
```
=== Evaluasi SVM ===
              precision    recall  f1-score   support

     Organik       0.87      0.92      0.89       182
   Anorganik       0.91      0.86      0.88       178

    accuracy                           0.89       360
   macro avg       0.89      0.89      0.89       360
weighted avg       0.89      0.89      0.89       360
```

Gambar 10. Classification Report SVM (kernel RBF) pada Data Uji.

Hasil penilaian menunjukkan bahwa akurasi SVM 88,89% atau mendekati 89%, seperti yang tertera pada Gambar 10, bahwa *precision*, *recall*, dan F1-*score* untuk setiap kelas juga cukup tinggi dan seimbang.

Pada gambar 11, Sebagian besar sampel organik dan anorganik terklasifikasi dengan tepat. Kesalahan dalam klasifikasi terlihat lebih sedikit dibandingkan dengan KNN, menunjukkan bahwa SVM dapat melakukan pemisahan yang lebih efektif.

Secara keseluruhan, dalam penelitian ini, SVM lebih baik daripada KNN. Ini mungkin karena kemampuan SVM untuk memaksimalkan margin saat memisahkan kelas organik dan anorganik dengan baik, terutama ketika fitur HOG dapat mengesktraksi pola-pola yang signifikan. Akurasi juga meningkat dengan penemuan parameter yang tepat, seperti gamma dan C.



Gambar 11. Confusion Matrix SVM

Berdasarkan hasil klasifikasi dari kedua model tersebut, SVM menunjukkan tingkat akurasi yang terbaik yaitu 88,89% . Di sisi lain, KNN dengan "k" yang bernilai 5 menghasilkan akurasi sebesar 83,61%. Perbedaan dalam kinerja ini menunjukkan bahwa metode SVM lebih cocok untuk pola-pola nya dihasilkan dari fitur HOG dalam dataset sampah, baik organik maupun anorganik. Meskipun begitu, hasil dari KNN masih bisa diperbaiki lebih lanjut dengan mengoptimalkan parameter, menambah data, atau menggunakan teknik prapemrosesan yang lebih baik.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis dan membandingkan efektivitas K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam mengklasifikasikan sampah organik dan anorganik dengan menggunakan fitur HOG dan LBP. Temuan menunjukkan bahwa SVM mendapatkan akurasi tertinggi yaitu 88,89%, sedangkan KNN mencatat akurasi 83,61%. SVM lebih mampu dalam memisahkan kedua kategori sampah berkat pemilihan parameter C dan gamma yang optimal pada kernel RBF. Sebaliknya, KNN menghadapi tantangan terutama pada klasifikasi sampah anorganik, yang ditandai dengan nilai recall yang lebih rendah dibandingkan dengan limbah organik. Walaupun demikian, kinerja KNN dapat ditingkatkan dengan melakukan optimasi parameter, menambah jumlah data, serta menerapkan metode prapemrosesan yang lebih efisien.

DAFTAR PUSTAKA

[1] K. Ariadi, F. T. Anggraeny, and A. N. Sihananto, "PERBANDINGAN PERFORMA METODE SVM DAN KNN DALAM MENGKLASIFIKASIKAN CITRA INFEKSI TELINGA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11342–11347, 2024, doi: https://doi.org/10.36040/jati.v8i6.11350.

[2] A. Arfian, J. Siregar, and D. Wijayanti, "INTEGRASI MODEL NEURAL NETWORK DENGAN SVM DAN KNN UNTUK DETEKSI DINI PENYAKITA JANTUNG PADA PENDERITA HYPERTENSI," *Technologia : Jurnal Ilmiah*, vol. 16, no. 1, pp. 104–109, Jan. 2025, doi: 10.31602/tji.v16i1.17046.

[3] S. A. B. Nasution, D. Lestari, D. P. Azzahra, and D. Kiswanto, "Deteksi Jenis Tanaman Berdasarkan Bentuk Daun Menggunakan KNN," *Jurnal Sistem Informasi dan Sistem Komputer*, vol. 10, no. 1, pp. 31–38, Jan. 2025, doi: 10.51717/simkom.v10i1.646.

[4] A. Marlina, A. N. Sari, N. A. Syahira, P. Syafarina, and R. S. Bintang, "Edukasi Mengenai Pentingnya Pemilahan Serta Pengolahan Sampah Untuk Mengurangi Dampak Negatif Terhadap Lingkungan," *Darmabakti: Jurnal Inovasi Pengabdian dalam Penerbangan*, vol. 4, no. 1, pp. 11–17, 2023, doi: https://doi.org/10.52989/darmabakti.v4i1.108.

[5] Y. Ritonga and Usiono, "SAMPAH DAN PENYAKIT : SYSTEMATIC LITERATURE REVIEW," *JURNAL KESEHATAN TAMBUSAI*, vol. 4, no. 4, pp. 5148–5157, 2023, doi: https://doi.org/10.31004/jkt.v4i4.19608.

[6] A. Z. S. Azizah, F. N. T. Yalis, I. S. Narayana, K. A. Syalom, and P. Rosyani, "Pengolahan Citra Digital Dengan Penerapan Teknik Ambang Batas: Studi Kasus Menggunakan Opencv," *Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan*, vol. 1, no. 4, pp. 283–287, 2024, [Online]. Available: https://jurnalmahasiswa.com/index.php/aidansp k

[7] J. Jumadi, Yupianti, and D. Sartika, "PENGOLAHAN CITRA DIGITAL UNTUK IDENTIFIKASI OBJEK MENGGUNAKAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING," *Jurnal Sains dan Teknologi*, vol. 10, no. 2, pp. 148–156, 2021, doi: https://doi.org/10.23887/jstundiksha.v10i2.3363 6.

[8] T. A. M and Q. N. Azizah, "Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine," *Jurnal Infortech*, vol. 4, no. 1, pp. 45–50, 2022, doi: https://doi.org/10.31294/infortech.v4i1.12813.

[9] F. A.-I. A. Putra, A. G. Sulaksono, L. T. Utomo, and A. R. Khamdani, "Klasifikasi Buah dan Sayur menggunakan fitur ekstraksi HOG dan Metode KNN," *JIP (Jurnal Informatika Polinema)*, vol. 10, no. 1, pp. 45–52, 2023, doi: https://doi.org/10.33795/jip.v10i1.1433.

[10] M. T. Kanugroho, M. A. Rahman, and R. C. Wihandika, "Klasifikasi Batik dengan Ekstraksi Fitur Tekstur Local Binary Pattern dan Metode K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 10, pp. 4788–4794, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[11] R. Kosasih and C. Daomara, "Pengenalan Wajah dengan Menggunakan Metode Local Binary Patterns Histograms (LBPH)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, pp. 1258–1264, Oct. 2021, doi: 10.30865/mib.v5i4.3171.

[12] D. N. B. Ginting, R. Faristyawan, S. N. M. Safitri, and G. Winarso, "Identifikasi Spesies Mangrove dengan menggunakan Metode Principal Component Analysis (PCA) pada Citra Landsat-8 di Taman Nasional Sembilang, Sumatera Selatan, Indonesia," *Jurnal Kelautan Tropis*, vol. 27, no. 2, pp. 345–356, Jun. 2024, doi: 10.14710/jkt.v27i2.22830.

[13] M. Rais, R. Goejantoro, and S. Prangga, "Optimalisasi K-Means Cluster dengan Principal Component Analysis pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Tingkat Pengangguran Terbuka," *Jurnal EKSPONENSIAL*, vol. 12, no. 2, pp. 129–136, 2021, doi: https://doi.org/10.30872/eksponensial.v12i2.805

[14] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021, doi: https://doi.org/10.31294/ijcit.v6i2.10438.

[15] M. I. Mubarok, Purwantoro, and Carudin, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DALAM KLASIFIKASI PENILAIAN JAWABAN UJISAN ESAI," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3446–3452, 2023, doi: https://doi.org/10.36040/jati.v7i5.7676.

[16] Rosdiana, M. Ula, and H. A. K. Aidilof, "IMPLEMENTASI PEMODELAN CITRA MODEL SVM (SUPPORT VECTOR MACHINE) DALAM PENENTUAN PENGKLASIFIKASIAN JENIS SUARA KONTES BURUNG," *Jurnal Informatika Kaputama (JIK)*, vol. 5, no. 2, pp. 317–324, 2021, doi: https://doi.org/10.59697/jik.v5i2.264.

[17] F. Abdusyukur, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER," *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 73–82, 2023, doi: https://doi.org/10.34010/komputa.v12i1.9418.

[18] R. Damasela, B. P. Tomasouw, and Z. A. Leleury, "PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MENDETEKSI PENYALAHGUNAAN NARKOBA," *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, vol. 1, no. 2, pp. 111–122, Oct. 2022, doi: 10.30598/parameterv1i2pp111-122.

[19] R. Nurhidayat and K. E. Dewi, "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN FITUR EKSTRAKSI N-GRAM DALAM ANALISIS SENTIMEN BERBASIS ASPEK," *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 91–100, 2023, doi: https://doi.org/10.34010/komputa.v12i1.9458.

[20] R. Maulana, R. Narasati, R. Herdiana, R. Hamonangan, and S. Anwar, "KOMPARASI ALGORITMA DECISION TREE DAN NAIVE BAYES DALAM KLASIFIKASI PENYAKIT DIABETES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3865–3870, 2023, doi: https://doi.org/10.36040/jati.v7i6.8265