

Implementasi penggunaan seleksi fitur pada klasifikasi daun sirih.

Mulyanto¹, Asrina Astagani², Aam Shodiqul Munir ^{*3} dan April Line Viona⁴

1-4 Program Studi D3 Teknik Informatika
Politeknik Negeri Samarinda
Jalan Cipto Mangunkusumo Kampus Gn. Lipan Samarinda
mulyanto@polnes.ac.id; astagani.asrina@gmail.com;
aamshodiqulmunir@gmail.com ; aprillineviona07@gmail.com;

Abstrak

Daun sirih (*Piper Betle*) memiliki beberapa varian yang berbeda seperti sirih merah, hijau, dan hitam, yang masing-masing memiliki kegunaan berbeda dalam keperluan medis hingga tradisional. Penelitian ini bertujuan untuk mengembangkan metode klasifikasi jenis daun sirih berdasarkan fitur warna dan tekstur, dengan memanfaatkan metode seleksi fitur Algoritma Genetika dan *Correlation Based Feature Selection* (CFS) untuk meningkatkan performa klasifikasi. Data citra daun sirih dikumpulkan melalui proses akuisisi citra digital, diikuti dengan segmentasi menggunakan metode thresholding untuk memisahkan daun dari latar belakang. Selanjutnya, dilakukan preprocessing dan ekstraksi fitur. Ekstraksi fitur warna dilakukan dalam ruang warna RGB, sedangkan ekstraksi fitur tekstur didapatkan dari *Gray Level Co-occurrence Matrix* (GLCM). Penelitian ini mengimplementasikan dua teknik seleksi fitur yaitu Algoritma Genetika yang bekerja berdasarkan prinsip evolusi biologis, dan *Correlation Based Feature Selection* (CFS) yang menggunakan pendekatan berbasis korelasi untuk mengurangi dimensi fitur dan mempertahankan fitur yang paling relevan. Metode klasifikasi *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan kelas daun sirih. Hasil penelitian menunjukkan bahwa seleksi fitur dengan Algoritma Genetika meningkatkan akurasi klasifikasi secara signifikan, dengan KNN mencapai akurasi tertinggi 98,33% dari sebelumnya 93,3%, sedangkan CFS memberikan peningkatan moderat dengan akurasi 93,3% pada KNN dan 85% pada Naive Bayes dan SVM. Hal ini membuktikan bahwa implementasi seleksi fitur menggunakan kedua metode tersebut efektif meningkatkan performa akurasi pada klasifikasi daun sirih, dengan Algoritma Genetika menunjukkan hasil yang lebih unggul dalam penelitian ini.

Kata Kunci ketidakstabilan, lereng, RMR, kinematika, RQD

Digital Object Identifier 10.36802/jnanaloka.2025.v6-no1-27-37

1 Pendahuluan

Perkembangan teknologi informasi telah memberikan dampak signifikan terhadap berbagai bidang kehidupan, termasuk dalam bidang identifikasi dan klasifikasi tanaman. Salah satu tanaman yang memiliki nilai penting dalam khasanah pengobatan tradisional Indonesia adalah daun sirih (*Piper Betle*) [1]. Daun sirih merupakan daun tunggal yang memiliki bentuk menyerupai hati, dengan permukaan yang mengkilap dan ujung yang runcing. Daun ini tumbuh secara berselang-seling, memiliki tangkai, dan mengeluarkan aroma khas ketika

* Corresponding author.



diremas. Ukuran daun sirih berkisar antara 6 hingga 17,5 cm untuk panjangnya dan 3,5 hingga 10 cm untuk lebar [2].

Tanaman ini telah dimanfaatkan secara turun-temurun sebagai bahan pengobatan dan memiliki potensi yang besar untuk dikembangkan dalam industri farmasi modern. Keberadaan varietas daun sirih yang beragam, seperti sirih hijau, sirih hitam dan sirih merah, memerlukan metode identifikasi yang akurat untuk membedakan karakteristik morfologinya. Seiring dengan majunya perkembangan zaman, daun sirih dapat dibedakan / diklasifikasikan dengan memanfaatkan berbagai teknik teknik pengolahan citra digital dan machine learning, seperti dengan menggunakan metode ekstraksi fitur warna [3] dan tekstur [4], menggunakan metode klasifikasi seperti KNN [3], Naive bayes, Support Vector Machine dan ANN [4].

Tantangan utama dalam klasifikasi berbasis pengolahan citra digital terletak pada dimensionalitas data yang tinggi. Hasil ekstraksi fitur dari citra daun sirih seringkali menghasilkan vektor fitur berdimensi besar yang dapat mempengaruhi performa algoritma klasifikasi. Permasalahan dimensi tinggi ini berdampak pada penurunan akurasi, peningkatan kompleksitas komputasi, dan kemungkinan terjadinya fenomena *curse of dimensionality*. Oleh karena itu, implementasi teknik reduksi dimensi menjadi langkah krusial dalam proses pengembangan sistem klasifikasi yang efektif [5]. Seleksi fitur merupakan pendekatan strategis untuk mengatasi permasalahan dimensionalitas data. Proses ini bertujuan untuk mengidentifikasi dan memilih subset fitur yang paling relevan dan informatif, sehingga dapat meningkatkan performa klasifikasi sekaligus mengurangi kompleksitas komputasi [6].

Berbagai metode seleksi fitur dapat digunakan untuk mereduksi dimensi dari suatu data, seperti menggunakan metode Feature Selection Correlation Based [5], menggunakan Information Gain [7] dan lain lain. Penelitian oleh [8] dan [9] menunjukkan bagaimana penggunaan metode seleksi fitur, seperti *Correlation-based Feature Selection* (CFS) dan ReliefF, dapat meningkatkan akurasi model Support Vector Machine (SVM). Penelitian tersebut mengeksplorasi berbagai teknik seleksi fitur untuk memahami fitur mana yang paling berkontribusi terhadap hasil klasifikasi, menegaskan bahwa pemilihan fitur yang cermat adalah fundamental dalam mencapai hasil terbaik dalam analisis data.

Penelitian yang dilakukan oleh [10] menunjukkan penerapan CFS dalam memprediksi performa akademik mahasiswa menggunakan decision tree. Hasil penelitian mengungkapkan bahwa CFS yang dikombinasikan dengan algoritma C4.5 berhasil mencapai akurasi sebesar 76,92%, membuktikan keefektifan metode seleksi fitur ini dalam meningkatkan ketepatan prediksi. kemudian [11] menekankan pentingnya metode seleksi fitur yang tepat, seperti CFS, untuk mengoptimalkan proses klasifikasi dengan menghilangkan fitur-fitur yang tidak relevan dan redundan, yang pada gilirannya akan meningkatkan performa model. Pada penelitian [5] melakukan analisis komparatif antara CFS dengan Information Gain, yang semakin menegaskan efektivitas metode berbasis korelasi dalam mengembangkan teknik seleksi fitur.

Algoritma genetika adalah suatu algoritma yang dapat digunakan untuk mereduksi data / seleksi fitur sehingga dapat meningkatkan akurasi dari klasifikasi dan mereduksi waktu komputasi [12]. Algoritma Genetika merupakan salah satu algoritma optimasi, yang diciptakan dengan meniru beberapa proses yang terjadi pada evolusi alam. Algoritma Genetika juga merupakan algoritma stochastic yang bekerja berdasarkan prinsip-prinsip seleksi alam yang sering diterapkan dalam masalah machine learning dan optimasi [13].

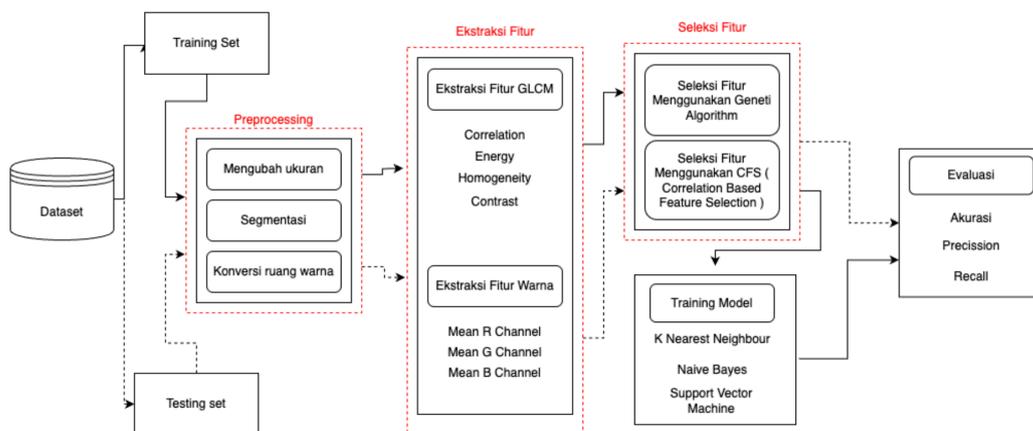
Algoritma genetika telah menjadi salah satu metode yang populer dalam fitur seleksi dan optimasi parameter dalam berbagai domain penelitian. Prinsip dasar algoritma ini terinspirasi oleh seleksi alam dan mekanisme evolusi biologis, yang memungkinkan pencarian

solusi optimal melalui representasi genetik dan proses seleksi [14, 15]. Proses ini mencakup tahapan seperti inialisasi populasi, evaluasi fitness individu, seleksi, crossover, dan mutasi untuk menghasilkan generasi baru yang lebih baik [16].

Dalam konteks fitur seleksi, algoritma genetika telah terbukti sangat efektif. Misalnya, dalam penelitian [17], Algoritma Genetika digunakan untuk mengoptimalkan klasifikasi kerusakan pada tulang belakang dengan meningkatkan akurasi model k-nearest neighbours (k-NN) mencapai 93%. Penelitian ini juga menunjukkan bahwa seleksi fitur yang dilakukan melalui Algoritma Genetika tidak hanya meningkatkan akurasi tetapi juga efisiensi model. Pendekatan ini sejalan dengan hasil dari Penelitian [18], yang membandingkan Algoritma genetika dengan algoritma greedy dalam konteks prediksi harga minyak kelapa sawit (CPO) dan menemukan bahwa algoritma genetika memberikan hasil yang lebih baik dalam pemilihan fitur-fitur yang relevan. Berdasarkan beberapa uraian tersebut, penelitian ini ingin mengkombinasikan dua buah metode seleksi fitur yaitu Algoritma genetika dan *Correlation based feature selection* kedalam klasifikasi daun sirih untuk mereduksi dimensi dari data hasil ekstraksi menggunakan ekstraksi fitur tekstur dan warna.

2 Metodologi

Metodologi penelitian yang digunakan dalam penelitian ini dibagi dalam dua bagian besar yaitu fase training dan fase testing. Dua fase tersebut memiliki beberapa langkah langkah yang sama yaitu akuisisi data, preprocessing, ekstraksi fitur, training model, testing model kemudian evaluasi [3, 19]. Pada penelitian ini, langkah langkah tersebut akan ditambahkan suatu proses yaitu seleksi ciri sehingga proses tersebut menjadi akuisisi data, preprocessing, ekstraksi fitur training model, testing model dan evaluasi (Gambar 1).



■ **Gambar 1** Langkah langkah penelitian.

Pada langkah akuisisi citra, penelitian ini menggunakan data citra daun sirih sebagai masukan yang dibagi menjadi dua bagian data yaitu data training dan data testing. Data daun sirih yang digunakan terdiri dari 3 kelas / kategori yaitu daun sirih hijau, daun sirih merah dan daun sirih hitam. Setelah langkah akuisisi data selesai, dilakukan langkah *preprocessing*. langkah preprocessing dilakukan untuk mengeliminasi background (latar belakang) dari *foreground* (objek daun sirih) sehingga proses ekstraksi fitur yang akan dilakukan menjadi lebih akurat. Langkah yang selanjutnya adalah ekstraksi fitur / ciri, metode yang digunakan pada langkah ekstraksi adalah ekstraksi ciri tekstur dan ekstraksi ciri war-

na. Kemudian langkah seleksi fitur digunakan untuk mendapatkan fitur-fitur yang relevan pada hasil ekstraksi fitur tersebut. Langkah selanjutnya adalah training klasifikasi sekaligus testing klasifikasi untuk membedakan tipe-tipe daun sirih berdasarkan fitur-fitur yang didapatkan dan diseleksi sebelumnya. Langkah terakhir adalah evaluasi untuk mengukur performa dari langkah klasifikasi.

Akuisisi data

Pada langkah pengumpulan data, data yang digunakan pada penelitian ini berjumlah 300 buah data citra daun sirih. Data tersebut dibagi menjadi 3 kategori yaitu 100 buah data citra sirih merah, 100 buah data citra sirih hijau dan 100 buah data citra sirih hitam. Ilustrasi dari contoh masing-masing kelas daun sirih dapat dilihat pada gambar 2.



■ **Gambar 2** Contoh sampel data citra daun sirih (a) Daun sirih Merah (b) Daun sirih Hitam (c) Daun sirih Hijau.

Preprocessing

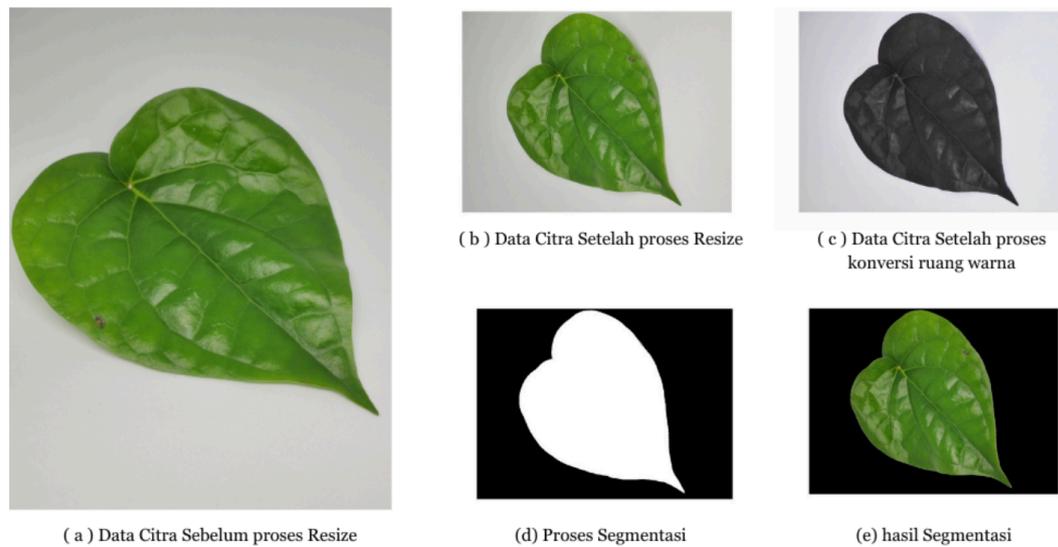
Pada langkah preprocessing dilakukan beberapa proses, yaitu resizing (mengubah ukuran), segmentasi dan juga konversi ruang warna. Proses resizing bertujuan untuk memperkecil ukuran gambar sehingga dapat mempercepat proses ekstraksi dan training model. Proses perubahan ukuran tersebut mengubah data citra sirih yang memiliki ukuran 3000 x 4000 piksel menjadi 480 x 640 piksel [3]. Setelah diubah ukurannya, data citra sirih kemudian mengalami konversi ruang warnanya menjadi grayscale. Tujuan konversi ini adalah untuk menjalankan proses segmentasi dan ekstraksi fitur.

Setelah dilakukan perubahan ukuran dan konversi ruang warna grayscale, dilakukan tahapan segmentasi. Tahapan segmentasi menjadi sangat penting untuk memisahkan object yang akan diekstraksi dari background [20]. Proses segmentasi juga dapat dianggap sebagai sebuah upaya untuk mengelompokkan object-object yang sama dalam suatu citra. Pada penelitian ini digunakan metode segmentasi yaitu thresholding.

Thresholding adalah salah satu metode segmentasi citra yang digunakan untuk memisahkan objek dari latar belakang dalam sebuah citra, berdasarkan perbedaan tingkat cerah atau gelapnya objek tersebut [21]. Gambar 3 memperlihatkan hasil dari proses preprocessing yang dilakukan pada penelitian ini.

Ekstraksi fitur

Ekstraksi fitur dilakukan untuk memperoleh karakteristik unik atau ciri khas dari suatu objek. Agar fitur tersebut dapat dianggap baik, informasi yang dihasilkan harus mampu membedakan satu objek dari objek lainnya. Dalam proses pengambilan fitur, penting untuk



■ **Gambar 3** Ilustrasi dari proses preprocessing yang dilakukan. (a) Data citra awal (b) Data citra setelah di Resize (c) Data citra setelah konversi ruang warna (d) Proses segmentasi (e) Hasil segmentasi.

mempertimbangkan kompleksitas komputasinya, sifat invariannya terhadap transformasi agar tidak terikat, serta jumlah fitur yang dihasilkan harus sedikit. Pada penelitian ini digunakan ekstraksi fitur warna menggunakan nilai mean dari setiap ruang warna RGB. Selain menggunakan nilai mean dari ruang warna RGB, digunakan juga ekstraksi ciri tekstur menggunakan 4 matrix dari metode *Gray Level Co-occurrence matrix*, yakni [22]:

1. **Contrast**, menunjukkan ukuran penyebaran (momen inersia) dari elemen-elemen matriks citra. Jika elemen-elemen tersebut terletak jauh dari diagonal utama, maka nilai kontras menjadi besar. Secara visual, kontras dapat dipahami sebagai ukuran variasi antara derajat keabuan dalam suatu area citra.
2. **Correlation**, menunjukkan ukuran suatu ketergantungan linier dari derajat keabuan citra, yang dapat memberikan indikasi / informasi adanya struktur linier dalam citra tersebut.
3. **Homogeneity**, menggambarkan keseragaman variasi intensitas dalam citra. Nilai homogenitas akan tinggi jika semua piksel dalam citra memiliki nilai intensitas yang serupa. Ketika homogenitas mendekati satu, hal ini menunjukkan bahwa distribusi elemen sangat dekat dengan diagonalnya, menandakan bahwa intensitas piksel dalam citra tersebut relatif konsisten dan tidak bervariasi secara signifikan [23].
4. **Energi**, menunjukkan penyebaran intensitas piksel terhadap jangkauan grayscale. Energi bernilai tinggi bila nilai pikselnya mirip dengan piksel lainnya, sebaliknya bernilai rendah yang berarti nilai pikselnya bersifat heterogen [23].

Total dari keseluruhan fitur adalah 19 buah fitur dengan 16 fitur didapatkan dari ekstraksi fitur tekstur dan 3 fitur didapatkan dari fitur warna. Tabel 1 menunjukkan sampel contoh dari proses ekstraksi fitur.

■ **Tabel 1** Contoh sampel hasil dari langkah ekstraksi fitur.

Jenis Daun	Fitur			
Daun sirih hijau	Contrast 0	0,330681622442	Correlation 90	0,9713581732
	Contrast 45	0,500911923141	Correlation 135	0,4195953935
	Contrast 90	0,162906470204	Homogeneity 0	0,41609814479
	Contrast 135	0,11311422845691	Homogeneity 45	0,42391121982
	Energi 0	0,1388428158923	Homogeneity 90	0,4178408426
	Energi 45	0,080304609218	Homogeneity 135	0,9673247828
	Energi 90	0,11688306472	Mean R	0,9626249158
	Energi 135	0,972258905528	Mean G	0,9755459585
	Correlation 0	0,965977005413	Mean B	0,96517102340
	Correlation 45	0,9803054153202		
Daun sirih hitam	Contrast 0	0,270118	Correlation 90	0,952296
	Contrast 45	0,345258	Correlation 135	0,493962
	Contrast 90	0,181759	Homogeneity 0	0,491634
	Contrast 135	0,076898	Homogeneity 45	0,494839
	Energi 0	0,083365	Homogeneity 90	0,490024
	Energi 45	0,068481	Homogeneity 135	0,975135
	Energi 90	0,091453	Mean R	0,972307
	Energi 135	0,959850	Mean G	0,976602
	Correlation 0	0,956515	Mean B	0,970320
	Correlation 45	0,964245		
Daun sirih merah	Contrast 0	0,445674	Correlation 90	0,523217
	Contrast 45	0,286582	Correlation 135	0,520434
	Contrast 90	0,139615	Homogeneity 0	0,523091
	Contrast 135	0,175887	Homogeneity 45	0,519739
	Energi 0	0,140778	Homogeneity 90	0,953786
	Energi 45	0,186409	Homogeneity 135	0,945005
	Energi 90	0,961406	Mean R	0,952527
	Energi 135	0,951442	Mean G	0,943742
	Correlation 0	0,961084	Mean B	
	Correlation 45	0,948537		

Seleksi Fitur

Setelah didapatkan data hasil ekstraksi fitur, maka dilakukan proses seleksi fitur. Pada penelitian ini mengimplementasikan dua teknik seleksi fitur yaitu menggunakan metode algoritma genetika dan menggunakan *Correlation Based feature Selection*.

Algoritma genetik adalah metode yang terinspirasi oleh teori evolusi Darwin, dan merupakan salah satu paradigma perangkat lunak yang relatif baru dan populer. Metode ini dapat diterapkan dalam sistem data mining untuk mengklasifikasikan data, sehingga menghasilkan informasi yang berguna.

Algoritma genetika berfungsi dengan cara meniru proses seleksi alam, di mana individu terbaik dipilih untuk menghasilkan generasi baru yang lebih baik, memungkinkan penyelesaian masalah optimasi yang kompleks dengan efisien [12,13]. Gambaran Umum proses seleksi fitur menggunakan algoritma genetika dapat dilihat pada gambar 4.

Correlation based feature Selection merupakan salah satu metode seleksi fitur yang mengimplementasikan pendekatan pencarian berbasis peringkat [5]. Metode ini secara khusus



■ **Gambar 4** Gambaran umum proses seleksi fitur.

memperhatikan kelas target dalam proses evaluasi. Korelasi antara setiap atribut dan kelas target diukur menggunakan Metode Korelasi Pearson. Setiap nilai yang dihasilkan berfungsi sebagai indikator relevansi atribut, dengan mempertimbangkan atribut nominal dalam konteks basis nilai [24].

Klasifikasi (*Training & Testing*)

Proses klasifikasi menggunakan 3 buah metode klasifikasi yaitu naive bayes, KNN dan juga support Vector Machine. Proses klasifikasi dilakukan dua kali, klasifikasi pertama menggunakan input data tanpa dilakukan seleksi fitur dan klasifikasi kedua menggunakan input data yang sudah menjalani proses seleksi fitur.

Evaluasi

Langkah evaluasi adalah langkah terakhir dari penelitian ini. Langkah evaluasi menggunakan confusion matrix. *Confusion matrix* adalah suatu metode yang dapat mengukur performa dari suatu klasifikasi. Confusion matrix berbentuk matrix (MXM) Dimana, M adalah jumlah kelas klasifikasi, baris dan kolom mewakili nilai data dari setiap kelas. Penelitian ini menggunakan 3 buah parameter evaluasi dari *Confusion matrix*. Parameter tersebut antara lain [3]: *Precision* (presisi), *Recall* (recall) dan *Accuracy* (Akurasi).

True Positive (TP) dan *True Negative* (TN) merupakan jumlah data uji yang berhasil diklasifikasikan dengan tepat, di mana kelas target sesuai dengan hasil prediksi. *False Positive* (FP) terjadi ketika data yang sebenarnya memiliki kelas negatif tetapi diprediksi sebagai positif oleh metode yang digunakan. Sebaliknya, *False Negative* (FN) terjadi ketika data yang seharusnya bernilai positif justru diprediksi sebagai negatif oleh model. Nilai TP, TN, FP, dan FN diperoleh melalui analisis *confusion matrix*.

3 Hasil dan pembahasan

Penelitian ini menghasilkan sistem klasifikasi daun sirih dengan menggunakan beberapa metode seleksi fitur dan klasifikasi. Hasil penelitian ini memperlihatkan performa dari masing-masing metode, baik tanpa seleksi fitur maupun dengan penerapan seleksi fitur menggunakan Algoritma Genetika dan CFS.

Proses seleksi fitur menggunakan 2 buah metode yaitu menggunakan Algoritma genetik dan menggunakan Correlation based feature selection. Seleksi fitur digunakan untuk mengeliminasi fitur-fitur yang tidak relevan dan tidak berkorelasi. Hasil dari seleksi fitur dapat dilihat pada Tabel 2.

■ **Tabel 2** Hasil Seleksi Fitur menggunakan Algoritma genetika dan CFS.

Metode Seleksi Fitur	Nama Fitur
Algoritma Genetika	Mean R , Mean B, Contrast 0 derajat, Correlation 45 derajat, Correlation 135, derajat Energy 135 Derajat
<i>Correlation Based Feature Selection</i>	Energi 0 derajat, Energi 135 derajat, correlation 0 derajat, correlation 45 derajat, correlation 90 derajat

■ **Tabel 3** Hasil Evaluasi klasifikasi tanpa menggunakan seleksi fitur

Classifier	Sirih hijau		Sirih hitam		Sirih merah		Akurasi
	Presisi	Recall	Presisi	Recall	Presisi	Recall	
KNN	90%	90%	90%	90%	100%	100%	93%
Naive Bayes	80%	69,57%	65%	76,47%	95%	95%	80%
SVM	90%	60%	65%	86,67%	75%	100%	76,67%

Berdasarkan Tabel 2, Algoritma Genetika memilih enam fitur yang terdiri dari Mean R, Mean B, Contrast 0 derajat, Correlation 45 derajat, Correlation 135 derajat, dan Energy 135 derajat. Sementara itu, metode CFS memilih lima fitur yang meliputi Energi 0 derajat, Energi 135 derajat, Correlation 0 derajat, Correlation 45 derajat, dan Correlation 90 derajat. Meskipun kedua metode menghasilkan fitur-fitur yang berbeda, terdapat irisan berupa fitur Correlation 45 derajat dan beberapa fitur energi dan korelasi pada sudut yang berbeda. Algoritma Genetika cenderung memilih fitur-fitur yang berkaitan dengan nilai rata-rata komponen warna (Mean R, Mean B) dan fitur tekstur pada berbagai sudut. Sementara CFS lebih berfokus pada fitur-fitur energi dan korelasi pada berbagai sudut. Perbedaan ini diakibatkan karena kedua metode tersebut memiliki pendekatan yang berbeda dalam mengevaluasi relevansi fitur. Algoritma Genetika menggunakan pendekatan evolusioner untuk menemukan kombinasi fitur optimal, sedangkan CFS lebih berfokus pada korelasi antara setiap fitur dengan fitur yang lain dan korelasi kelas dan fitur.

Setiap kategori / kelas pada klasifikasi daun sirih terdiri dari 100 buah data citra daun sirih hijau, 100 buah data daun sirih hitam dan 100 buah data daun sirih merah. Seluruh data tersebut kemudian di bagi menjadi dua buah bagian yaitu dataset training dan dataset testing. Data set training terdiri dari 80 % (240 Buah Data) keseluruhan data dan dataset testing terdiri dari 20 % keseluruhan data (60 Buah data). Klasifikasi juga menggunakan 3 buah metode klasifikasi yaitu KNN, Naive Bayes dan *Support Vector Machine* (SVM). Hasil evaluasi klasifikasi tanpa menggunakan seleksi fitur disajikan pada Tabel 3. Berdasarkan tabel tersebut, metode KNN mencapai akurasi tertinggi sebesar 93%, diikuti oleh Naive Bayes dengan akurasi 80%, dan SVM dengan akurasi 76,67%.

Tabel 4 menunjukkan hasil evaluasi klasifikasi menggunakan seleksi fitur CFS. Metode KNN tetap mencapai akurasi tertinggi sebesar 93,3%, sementara Naive Bayes dan SVM mengalami peningkatan akurasi menjadi 85%.

Hasil evaluasi klasifikasi menggunakan seleksi fitur Algoritma Genetika ditampilkan pada Tabel 5. Metode KNN menunjukkan peningkatan signifikan dengan akurasi mencapai 98,33%. Naive Bayes mencapai akurasi 89%, dan SVM mencapai akurasi 91,67%.

Berdasarkan Tabel 5, 4, dan 5, Hasil evaluasi klasifikasi tanpa seleksi fitur, metode KNN menunjukkan performa terbaik dengan akurasi 93%. Hal ini mengindikasikan bahwa metode KNN mampu memanfaatkan keseluruhan fitur dengan baik untuk mengklasifikasikan daun sirih. Namun, Naive Bayes dan SVM menunjukkan performa yang lebih rendah, yang mengindikasikan bahwa keberadaan fitur-fitur yang tidak relevan mungkin mempengaruhi

■ **Tabel 4** Hasil evaluasi klasifikasi menggunakan hasil seleksi fitur Algoritma Correlation Based Feature Selection

Classifier	Sirih hijau		Sirih hitam		Sirih merah		Akurasi
	Presisi	Recall	Presisi	Recall	Presisi	Recall	
KNN	90%	90%	90%	90%	100%	100%	93,3%
Naive Bayes	85%	73,91%	75%	83%	95%	100%	85%
SVM	80%	76,19%	95%	95%	80%	84,21%	85%

■ **Tabel 5** Hasil evaluasi klasifikasi menggunakan hasil seleksi fitur Algoritma genetika

Classifier	Sirih hijau		Sirih hitam		Sirih merah		Akurasi
	Presisi	Recall	Presisi	Recall	Presisi	Recall	
KNN	95%	100%	100%	95,24%	100%	95,24%	98,33%
Naive Bayes	85%	85%	85%	90%	95%	100%	89%
SVM	90%	85%	90%	85%	95%	100%	91,67%

performa kedua metode tersebut. Penerapan seleksi fitur terbukti meningkatkan akurasi klasifikasi pada semua metode. CFS memberikan peningkatan pada Naive Bayes dan SVM, sementara KNN hanya mengalami sedikit peningkatan. Di sisi lain, Algoritma Genetika memberikan peningkatan yang signifikan pada semua metode, terutama pada KNN yang mencapai akurasi tertinggi sebesar 98,33%. Peningkatan performa ini mengindikasikan bahwa seleksi fitur mampu mengeliminasi fitur-fitur yang tidak relevan dan berpotensi menghasilkan noise, sehingga model klasifikasi dapat berfokus pada fitur-fitur yang memiliki daya pembeda yang tinggi.

4 Kesimpulan dan saran

Berdasarkan Implementasi dan percobaan yang telah dilakukan, Penelitian ini membuktikan bahwa penggunaan seleksi fitur dapat meningkatkan akurasi dari proses klasifikasi daun sirih. Peningkatan terjadi di semua metode klasifikasi, dengan KNN menggunakan seleksi fitur Algoritma Genetika menunjukkan peningkatan paling signifikan yaitu mencapai akurasi 98,33% dari sebelumnya 93,3% tanpa seleksi fitur. Penelitian ini membuktikan bahwa memilih fitur yang tepat dalam meningkatkan performa klasifikasi daun sirih. Kombinasi metode seleksi fitur Algoritma Genetika dan klasifikasi KNN terbukti menjadi pendekatan yang paling efektif untuk klasifikasi daun sirih berdasarkan fitur warna dan tekstur.

Pustaka

- 1 F. J. Rachmawaty, D. A. C. Mahardina, B. Nirwani, T. Nurmasitoh, dan E. T. Bowo, "Manfaat sirih merah (*piper crocatum*) sebagai agen anti bakterial terhadap bakteri gram positif dan bakteri gram negatif," *JKKI: Jurnal Kedokteran Dan Kesehatan Indonesia*, pp. 12–20, 2009.
- 2 T. M. Sarjani, M. Mawardi, E. S. Pandia, dan D. Wulandari, "Identifikasi morfologi dan anatomi tipe stomata famili piperaceae di kota langsa," *JUPI (Jurnal IPA dan Pembelajaran IPA)*, vol. 1, no. 2, pp. 182–191, 2017.
- 3 H. Hamdani, A. Septiarini, N. Puspitasari, A. Tejawati, dan F. Alameka, "The color features and k-nearest neighbor algorithm for classifying betel leaf image," *IAES International Journal of Robotics and Automation*, vol. 13, no. 3, pp. 330–337, 2024.

- 4 N. Puspitasari, A. Septiarini, U. Hairah, A. Tejawati, dan H. Sulastri, "Betel leaf classification using color-texture features and machine learning approach," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 5, pp. 2939–2947, 2023.
- 5 A. N. Puteri, A. Arizal, dan A. D. Achmad, "Feature selection correlation-based pada prediksi nasabah bank telemarketing untuk deposito," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, pp. 335–342, 2021.
- 6 J. Jupriyadi, "Implementasi seleksi fitur menggunakan algoritma fvbrm untuk klasifikasi serangan pada intrusion detection system (IDS)," *Prosiding Semnastek*, 2018.
- 7 I. M. B. Adnyana, "Penerapan feature selection untuk prediksi lama studi mahasiswa," *Jurnal Sistem dan Informatika (JSI)*, vol. 13, no. 2, pp. 72–76, 2019.
- 8 N. Afifah, S. Deris, dan A. Bardadi., "Perbandingan kinerja kernel RBF dan linear pada algoritma support vector machine (SVM) untuk prediksi serangan ransomware locker," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 3118–3123, 2023.
- 9 I. Fitriani, S. Basuki, dan A. E. Minarno, "Seleksi fitur relief pada klasifikasi malware android menggunakan support vector machine (SVM)," *Jurnal Repositor*, vol. 2, no. 11, 2020.
- 10 A. S. B. Asmoro, W. S. G. Irianto, U. Pujiyanto, J. S. No, dan J. Timur, "Perbandingan kinerja hasil seleksi fitur pada prediksi kinerja akademis siswa berbasis pohon keputusan," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, pp. 84–89, 2018.
- 11 D. Pramadhana, "Klasifikasi penyakit diabetes menggunakan metode CFS dan ROS dengan algoritma J48 berbasis Adaboost," *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 1, pp. 89–98, 2021.
- 12 N. Amini, T. H. Saragih, M. R. Faisal, A. Farmadi, dan F. Abadi, "Implementasi algoritma genetika untuk seleksi fitur pada klasifikasi genre musik menggunakan metode random forest," *Jurnal Informatika Polinema*, vol. 9, no. 1, pp. 75–82, 2022.
- 13 P. Guo, X. Wang, dan Y. Han, "The enhanced genetic algorithms for the optimization design," in *2010 3rd international conference on biomedical engineering and informatics*, vol. 7. IEEE, 2010, pp. 2990–2994.
- 14 L. Sari, A. Romadloni, dan R. Listyaningrum, "Penerapan data mining dalam analisis prediksi kanker paru menggunakan algoritma random forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023.
- 15 A. Nazarius, R. D. Pratama, R. A. Soebagijo, R. Priskila, dan V. H. Pranatawijaya, "Peningkatan implementasi algoritma genetika dalam sistem penjadwalan praktikum," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3237–3243, 2024.
- 16 T. Taslim, D. Toresa, D. Jollyta, D. Suryani, dan E. Sabna, "Optimasi k-means dengan algoritma genetika untuk target pemanfaat air bersih provinsi riau," *Indonesian Journal of Computer Science*, 2021.
- 17 R. T. Prasetyo, "Seleksi fitur dan optimasi parameter k-nn berbasis algoritma genetika pada dataset medis," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 2, no. 2, pp. 213–221, 2020.
- 18 D. I. Puspitasari, "Penerapan data mining menggunakan perbandingan algoritma greedy dengan algoritma genetika pada prediksi rentet waktu harga crude palm oil," *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 2, no. 1, pp. 21–26, 2017.
- 19 N. Puspitasari, A. Septiarini, dan A. R. Aliudin, "Metode k-nearest neighbor dan fitur warna untuk klasifikasi daun sirih berdasarkan citra digital," *Jurnal PROSISKO*, vol. 10, no. 2, pp. 165–172, 2023.
- 20 E. Maria, Y. Yulianto, Y. P. Arinda, J. Jumiaty, dan P. Nobel, "Segmentasi citra digital bentuk daun pada tanaman di poltani samarinda menggunakan metode thresholding," *Jurnal Rekayasa Teknologi Informasi (JURTI)*, vol. 2, no. 1, pp. 37–46, 2018.

- 21 D.-H. Pham dan S. Meignen, "A novel thresholding technique for the denoising of multicomponent signals," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4004–4008.
- 22 M. Widyarningsih, "Identifikasi kematangan buah apel dengan gray level co-occurrence matrix (GLCM)," *Jurnal Saintekom: Sains, Teknologi, Komputer dan Manajemen*, vol. 6, no. 1, pp. 71–88, 2016.
- 23 Y. F. Achmad, A. Yulfitri, P. Maharani *et al.*, "Penerapan algoritma GLCM dan KNN dalam pengenalan jenis jerawat," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 6, no. 2, pp. 74–82, 2022.
- 24 G. Surya Aditya *et al.*, "Perbandingan performansi dengan metode correlation based feature selection pada lvq 2," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 8, no. 1, pp. 170–179, 2023.