

# Analisis Ekstraksi Fitur Klasifikasi Tanaman Rusak Menggunakan Pembelajaran Mesin

Achmad Hadzami Setiawan<sup>1</sup>, Andyan Yogawardhana<sup>2</sup>, Farrel Amroe Azhari<sup>3</sup>, Muhammad Mahdi<sup>4</sup>, Nadia Hasna Azzahra<sup>5</sup>

*Program Studi Ilmu Komputer  
Departemen Ilmu Komputer dan Elektronika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Gadjah Mada*

<sup>1</sup>[achmad.had2003@mail.uqm.ac.id](mailto:achmad.had2003@mail.uqm.ac.id), <sup>2</sup>[andyan.yogawardhana@mail.uqm.ac.id](mailto:andyan.yogawardhana@mail.uqm.ac.id),  
<sup>3</sup>[farrel.amr2003@mail.uqm.ac.id](mailto:farrel.amr2003@mail.uqm.ac.id), <sup>4</sup>[muhammad.mahdi@mail.uqm.ac.id](mailto:muhammad.mahdi@mail.uqm.ac.id),  
<sup>5</sup>[Nadia.Hasna.Azzahra@mail.uqm.ac.id](mailto:Nadia.Hasna.Azzahra@mail.uqm.ac.id)

**Abstrak**—Kerusakan pada tanaman merupakan salah satu masalah yang sering dihadapi pada bidang pertanian dan agrikultur. Dengan adanya kerusakan pada tanaman merupakan masalah serius yang dapat mengancam keberlanjutan pertanian dan kehidupan manusia karena Pertanian adalah sektor penting dalam kehidupan manusia yang secara signifikan berkontribusi terhadap keberlanjutan pangan. Oleh karena itu, deteksi dini dan klasifikasi tanaman rusak menjadi aspek kunci dalam pengendalian penyakit tanaman. Dalam penelitian ini, kami mengusulkan sebuah pendekatan untuk menganalisis dan mengklasifikasikan tanaman rusak berdasarkan fitur yang diambil dari gambar tanaman yang terkena kerusakan. Dengan adanya penelitian ini kami berharap dapat membuat sebuah program yang dapat mengatasi permasalahan dalam sektor pertanian dan agrikultur. Dalam menyelesaikan masalah ini kami akan menggunakan feature extraction, lebih jelasnya kami akan menggunakan fitur warna, fitur tekstur, dan fitur bentuk yang kemudian akan kami latih kepada sebuah model.

**Kata Kunci**—tanaman, kerusakan, agrikultur, pertanian, ekstraksi fitur, model, LBP, GLCM, hu moment, centroid, RGB, HSV

## I. LATAR BELAKANG

Tanaman sangat penting bagi kehidupan manusia yang menyediakan makanan, tempat tinggal, dan oksigen. Namun, tanaman juga rentan terhadap kerusakan akibat berbagai faktor. Tanaman yang rusak dapat berdampak signifikan terhadap hasil panen dan ketahanan pangan. Masalah kerusakan pada tanaman merupakan tantangan serius yang dihadapi oleh sektor pertanian di seluruh dunia. Penyakit, hama, dan faktor lingkungan seperti kekeringan atau polusi udara dapat menyebabkan kerusakan pada tanaman, mengurangi hasil panen, dan mengancam ketahanan pangan. Oleh karena itu, penting untuk mengembangkan metode yang efektif untuk mendeteksi dan mengklasifikasikan jenis kerusakan pada tanaman dengan akurasi yang tinggi. Secara tradisional, penilaian kerusakan tanaman dilakukan secara visual oleh ahli pertanian. Namun, metode ini dapat memakan

waktu, memerlukan keterampilan khusus, dan rentan terhadap subjektivitas. Dalam beberapa tahun terakhir, kemajuan dalam bidang pengolahan citra dan pembelajaran mesin telah membuka peluang baru dalam analisis dan pengenalan pola pada suatu gambar. Dalam konteks ini, ekstraksi fitur dan klasifikasi menggunakan teknik pembelajaran mesin telah muncul sebagai pendekatan yang menjanjikan untuk mengatasi tantangan identifikasi dan klasifikasi tanaman yang mengalami kerusakan.

Berdasar pada The Food and Agriculture Organization of the United Nations (FAO), pada tahun 2017, 2018, dan 2019, Indonesia mengalami kerugian pada sektor pertanian sebesar kurang lebih 4 miliar USD dikarenakan berbagai penyebab. Dengan program Analisis Ekstraksi Fitur Klasifikasi Tanaman Rusak Menggunakan Pembelajaran Mesin yang kami kembangkan diharapkan dapat membantu dalam mengurangi permasalahan kerusakan tanaman supaya dapat memberikan peran dalam kemajuan dunia.

Dalam pembuatan program ini, kami menggunakan berbagai algoritma dalam melakukan ekstraksi fitur. Oleh karena itu, pertama-tama kami harus mengenali apa itu ekstraksi fitur. Ekstraksi fitur adalah proses mengambil informasi yang relevan dan penting dari data mentah untuk digunakan dalam analisis lebih lanjut atau tugas pemrosesan data. Dalam konteks pengolahan citra dan pengenalan pola, ekstraksi fitur melibatkan identifikasi dan pengekstrakan atribut atau karakteristik khusus yang ada dalam citra. Tujuannya adalah untuk mewakili citra dengan fitur-fitur yang mampu membedakan antara objek atau pola yang berbeda. Pada ekstraksi fitur sendiri terdiri dari beberapa macam metode seperti:

- Ekstraksi berbasis intensitas: Fitur-fitur diekstraksi berdasarkan tingkat intensitas piksel dalam citra. Contohnya termasuk metode histogram, filter spasial, dan transformasi Fourier.

- Ekstraksi berbasis tekstur: Fitur-fitur diekstraksi berdasarkan pola dan struktur tekstur citra. Teknik seperti analisis tekstur menggunakan matriks co-occurrence, filter Gabor, dan transformasi wavelet digunakan untuk mengidentifikasi fitur-fitur tekstur.
- Ekstraksi berbasis bentuk: Fitur-fitur diekstraksi berdasarkan bentuk atau kontur objek dalam citra. Teknik seperti transformasi Hough, ekstraksi fitur bentuk invariant, dan ekstraksi fitur morfologis digunakan untuk mengenali bentuk-bentuk objek.
- Ekstraksi berbasis deteksi objek: Fitur-fitur diekstraksi dengan menggunakan algoritma deteksi objek untuk mengidentifikasi objek dalam citra. Algoritma seperti Viola-Jones dan R-CNN (Region-based Convolutional Neural Network) digunakan untuk ekstraksi fitur dan deteksi objek.
- Ekstraksi berbasis warna: Ekstraksi fitur yang fokus pada informasi warna dalam citra. Metode ini melibatkan identifikasi dan pengekstrakan fitur-fitur berdasarkan distribusi dan karakteristik warna dalam citra. Metode berbasis warna yang umum digunakan dalam ekstraksi fitur seperti histogram, model, dan deskriptor warna.

## II. PENELITIAN TERKAIT

Terdapat beberapa penelitian terdahulu yang telah mengaplikasikan ekstraksi fitur pada pendeteksian kesehatan sebuah tanaman dalam sebuah citra.

Pada *research paper* yang berjudul 'Feature Extraction and Naïve Bayes Algorithm for Defect Classification of Manalagi Apples' yang ditulis oleh Moch. Lutfi et al. (2022), mereka menghasilkan sebuah program dengan akurasi terbaik menggunakan metode yang diusulkan menggunakan *cross fold validation* 10 parameter dengan 78,98% dan pengujian lainnya dengan *cross fold validation* 5 parameter hasilnya kurang maksimal yaitu 78.61%. Pada penelitian tersebut dilakukan 2 ekstraksi fitur, yaitu fitur warna dan juga fitur tekstur. Pada ekstraksi warna digunakan *NTSC-based color extraction*. Dalam data citra NTSC terdiri dari tiga komponen, yaitu *luma* (Y), *in-phase* (I), dan *quadrature* (Q). Untuk ekstraksi fitur, mereka menggunakan teknik statistik yang biasa digunakan untuk ekstraksi fitur, yaitu *gray level co-occurrence matrix* (GLCM).

Lalu pada penelitian lain yang dituliskan pada *research paper* berjudul 'An effective feature extraction method for rice leaf disease classification' yang ditulis oleh Muhammad Anwarul Azim et al., mereka menghasilkan sebuah model dengan dengan akurasi berdasarkan kelas yang konsisten. Akurasi 86,58% pada dataset uji dicapai dengan jumlah fitur yang sedikit (26 fitur). Pemilihan fitur yang tepat

menghasilkan peningkatan akurasi yang signifikan dari penelitian sebelumnya. Pada *research paper* tersebut dilakukan evaluasi kinerja berbagai metode ekstraksi fitur pada set pelatihan. Metode-metode ekstraksi fitur tersebut meliputi:

- Fitur warna: Fitur warna diekstraksi dari gambar menggunakan model warna RGB.
- Fitur tekstur: Fitur tekstur diekstraksi dari gambar menggunakan filter Gabor.
- Fitur bentuk: Fitur bentuk diekstraksi dari gambar menggunakan momen inersia.

Penulis kemudian melatih algoritma pembelajaran mesin pada fitur-fitur yang diekstraksi. Algoritma pembelajaran mesin yang digunakan adalah *support vector machine* (SVM). SVM dilatih untuk mengklasifikasikan gambar sebagai gambar yang rusak atau sehat. Lalu dilakukan evaluasi kinerja pada SVM kepada set pengujian. SVM mencapai akurasi 86,58% dalam mengklasifikasikan tanaman yang rusak.

Dari kedua *research paper* diatas, terdapat beberapa kesamaan yang cukup menarik. Secara keseluruhan, kedua makalah tersebut sangat mirip dalam pendekatan mereka terhadap masalah klasifikasi tanaman yang rusak. Kedua makalah tersebut menggunakan dataset gambar yang serupa, serangkaian metode ekstraksi fitur yang serupa, dan algoritma pembelajaran mesin yang serupa.

## III. METODOLOGI

Bagian ini diawali dengan penjelasan tentang dataset yang digunakan dan tahapan-tahapan pra pemrosesan data yang dilakukan sebelum melakukan pelatihan pada model yang dirancang. Pada bagian ini juga akan dijabarkan rancangan arsitektur yang diusulkan untuk melakukan analisis ekstraksi fitur pada citra.

### A. Dataset

Untuk *dataset* yang digunakan pada eksperimen ini bersumber dari data yang disediakan oleh Plant Pathology 2020-FGVC7 Competition di Kaggle. Dataset Plant Pathology 2020-FGVC7 adalah koleksi besar gambar RGB berkualitas tinggi dari penyakit daun apel. Dataset ini berisi sekitar 4000 gambar, termasuk dataset penyakit berlabel yang besar. Dataset ini merupakan sumber daya yang berharga bagi para peneliti dan pengembang yang bekerja pada sistem deteksi penyakit tanaman berbasis penglihatan komputer. Dataset ini mencakup gambar pohon apel dengan berbagai penyakit yang berbeda, termasuk keropeng apel, karat apel, bercak hitam apel, dan embun tepung apel. Gambar-gambar tersebut diambil pada tahap kematangan yang berbeda dan pada waktu yang berbeda dengan pengaturan kamera fokus yang berbeda. Hal ini memungkinkan dataset tersebut digunakan untuk melatih dan mengevaluasi model yang mampu mendeteksi penyakit secara akurat dalam berbagai kondisi yang berbeda. Dataset ini juga bermanfaat bagi para peneliti yang tertarik untuk mempelajari efek dari berbagai faktor lingkungan

terhadap perkembangan penyakit tanaman. Dataset ini tersedia di situs web Kaggle.

### B. Pre Processing

Pada tahapan pra-pemrosesan citra diolah dengan berbagai algoritma. Pertama, kami menerapkan histogram *equalization* pada citra. Histogram *equalization* adalah teknik yang digunakan untuk meningkatkan kontras gambar dengan meregangkan histogram nilai piksel gambar sehingga lebih merata. Hal ini dapat berguna pada proses ekstraksi fitur dan klasifikasi, karena dapat membantu membuat fitur gambar menjadi lebih jelas dan lebih mudah diidentifikasi. Jadi, penerapan histogram *equalization* digunakan untuk meningkatkan kontras gambar tanaman yang rusak, sehingga lebih mudah untuk mengidentifikasi area kerusakan. Hal ini dapat memudahkan proses klasifikasi.



Gambar 1. Perbandingan citra yang dilakukan histogram *equalization* dan citra original.

Lalu, setelah dilakukan histogram *equalization*, akan dilakukan segmentasi menggunakan *grabcut mask*. *Grabcut mask* adalah algoritma yang digunakan dalam segmentasi gambar untuk mengekstrak objek dari sebuah gambar. Alat ini bekerja dengan terlebih dahulu membuat *mask* yang merupakan gambar biner di mana piksel objek berwarna putih dan piksel latar belakang berwarna hitam. Algoritma *grabcut* kemudian menggunakan *mask* secara iteratif untuk menyempurnakan segmentasi gambar.



Gambar 2. Perbandingan citra yang dilakukan pra-pemrosesan berupa histogram *equalization* dan segmentasi dengan citra original.

### C. Ekstraksi Fitur

Pada tahapan ini, data akan dikonversi menjadi *grayscale*, lalu kami menggunakan GLCM (*Gray-Level Co-occurrence Matrix*) yang bertujuan untuk mengukur hubungan spasial antara piksel-piksel dalam citra dan mengekstrak informasi tentang tekstur seperti keteraturan, arah, dan kontras. GLCM bekerja dengan menghitung matriks kemunculan bersama dari pasangan piksel dengan intensitas *gray-level* tertentu dalam suatu jarak dan arah tertentu. Matriks ini menyimpan informasi tentang seberapa sering pasangan piksel dengan kombinasi intensitas tertentu muncul bersamaan dalam citra. Dari matriks ini, berbagai fitur statistik seperti energi, kontras, homogenitas, dan entropi dapat dihitung untuk menggambarkan tekstur citra.

Lalu langkah selanjutnya ada LBP (*Local Binary Pattern*). Metode ini bertujuan untuk menggambarkan dan mengukur tekstur pada citra digital. Tujuannya adalah untuk mengekstrak fitur tekstur yang kuat dan diskriminatif dari citra. LBP bekerja dengan membandingkan nilai intensitas piksel dengan tetangganya dalam suatu area tertentu pada citra. Proses ini melibatkan konversi nilai intensitas piksel menjadi representasi biner berdasarkan perbandingan intensitas piksel dengan tetangganya. Setelah konversi, pola biner yang dihasilkan disusun menjadi bilangan bulat untuk mendapatkan deskriptor LBP. Lalu,

Langkah terakhir adalah melakukan ekstraksi fitur warna dengan mengambil nilai mean dari RGB dan HSV, selain itu kami juga melakukan ekstraksi pada fitur bentuk dengan mengambil nilai momen hu dan centroid.

### D. Seleksi Model

Pada program ini, kami memilih beberapa metode, yang pertama adalah SVM (*Support Vector Machine*). SVM bertujuan untuk memisahkan dua kelas data dengan mencari *hyperplane* terbaik yang memiliki margin maksimal di antara kelas-kelas tersebut. Prinsip dasar SVM adalah dengan mencari *hyperplane* optimal yang dapat memaksimalkan margin antara dua kelas data. Margin merupakan jarak antara *hyperplane* dan titik-titik terdekat dari kedua kelas.

Model kedua pada proyek ini adalah Random Forest. Model ini bertujuan untuk membuat prediksi yang akurat dengan memanfaatkan konsep *ensemble learning* dan penggabungan banyak *decision tree* yang bekerja secara independen. Random Forest bekerja dengan cara menggabungkan prediksi dari beberapa *decision tree* yang dibangun secara acak. Setiap *decision tree* dalam Random Forest dibangun menggunakan *subsampling* acak dari data pelatihan dan subset acak dari fitur yang tersedia. Ketika dilakukan prediksi, setiap *tree* memberikan suara untuk kelas target, dan hasil yang paling banyak muncul menjadi prediksi akhir dari Random Forest.

Model ketiga adalah KNN (*K-Nearest Neighbour*). Model ini bertujuan untuk melakukan klasifikasi atau regresi

berdasarkan data tetangga terdekat. KNN dapat digunakan untuk klasifikasi data dengan beberapa kelas (*multiclass*) maupun dua kelas (*binary*). Selain itu, KNN juga dapat digunakan untuk masalah regresi dengan menghitung rata-rata atau median dari nilai target dari tetangga terdekat.

Model keempat adalah Logistic Regression, model ini bertujuan untuk mempelajari hubungan antara variabel fitur (input) dengan variabel target (output) yang berupa kelas biner. Logistic Regression dapat digunakan untuk mengklasifikasikan data ke dalam dua kelas yang berbeda berdasarkan pola dan hubungan dari fitur-fitur yang ada. Cara kerja model ini adalah dengan menghitung nilai *log-odds* yang diperoleh dari transformasi linear pada nilai-nilai fitur dengan menggunakan bobot (*weights*) yang dipelajari selama proses training. Lalu, *log-odds* yang diperoleh diubah menjadi probabilitas menggunakan fungsi logistik (*sigmoid function*). Fungsi logistik mengubah nilai *log-odds* menjadi nilai probabilitas yang berada dalam rentang 0 hingga 1. Terakhir, menentukan *threshold* atau ambang batas untuk memutuskan klasifikasi. Misalnya, jika probabilitas melebihi *threshold* tertentu (biasanya 0.5), data akan diklasifikasikan ke dalam kelas positif, dan sebaliknya jika probabilitas di bawah *threshold*, data akan diklasifikasikan ke dalam kelas negatif.

Tahap terakhir dari pemilihan model adalah *hyperparameter tuning*. Tahap ini bertujuan untuk mencari kombinasi parameter yang menghasilkan performa terbaik untuk model yang dihasilkan. Performa model dapat diukur dengan menggunakan metrik evaluasi yang relevan, seperti akurasi, presisi, *recall*, atau *F1-score*. Dengan melakukan hyperparameter tuning, kita dapat meningkatkan performa model, mengurangi overfitting atau underfitting, dan meningkatkan generalisasi dari model terhadap data yang belum pernah dilihat sebelumnya. Pada model ini, kami menggunakan *grid search* yang melibatkan pencarian sistematis pada parameter-parameter yang telah ditentukan sebelumnya. Semua kombinasi parameter yang mungkin akan dievaluasi, dan model akan dilatih dan dievaluasi untuk setiap kombinasi tersebut. Meskipun metode ini cukup sederhana, namun bisa memakan waktu yang lama untuk mencari parameter terbaik terutama jika ruang parameter yang diuji cukup besar.

#### IV. KESIMPULAN

Berdasarkan dari berbagai pengujian, fitur yang paling berperan besar adalah *local binary pattern*, *centroid*, dan *red means*. Selain itu untuk model yang terbaik yang telah diuji adalah hyperparameter tuning dengan akurasi 0.80 atau sekitar 80%. Namun tidak menang jauh dari model lain yang akurasinya berkisar antara 70% dari 1820 data yang diuji pada proyek ini. Performa klasifikasi pada kelas tanaman sehat jauh lebih rendah dibandingkan kelas tanaman sakit. Hal ini disebabkan karena distribusi datanya yang belum merata. Untuk kedepannya diperlukan augmentasi untuk meningkatkan performa model. Selain itu bisa juga mencoba mengekstraksi fitur-fitur potensial seperti standar deviasi dan fitur lainnya.

#### V. LAMPIRAN

Untuk *link source code* dari program Analisis Ekstraksi Fitur Klasifikasi Tanaman Rusak Menggunakan Pembelajaran Mesin yang telah kami buat adalah sebagai berikut: [https://colab.research.google.com/drive/1Rie6SWLwfP0S\\_j6XvGeCLAGC4cXsWrzU?usp=sharing](https://colab.research.google.com/drive/1Rie6SWLwfP0S_j6XvGeCLAGC4cXsWrzU?usp=sharing)

#### REFERENSI

- [1] S. Sachar and A. Kumar, "Survey of feature extraction and classification techniques to identify plant through leaves," *Expert Systems with Applications*, vol. 167. Elsevier BV, p. 114181, Apr. 2021 [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2020.114181>
- [2] P. D. Kumar and A. Suhasini, "Feature extraction and classification based crop disease detection using deep neural network and ensemble convolutional neural network architecture," *International journal of health sciences*. Universidad Tecnica de Manabi, pp. 2257–2273, Apr. 13, 2022 [Online]. Available: <http://dx.doi.org/10.53730/ijhs.v6nS3.5997>
- [3] M. A. Azim, M. K. Islam, Md. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2. Universitas Ahmad Dahlan, p. 463, Apr. 01, 2021 [Online]. Available: <http://dx.doi.org/10.12928/TELKOMNIKA.v19i2.16488>
- [4] Moch. Lutfi, H. S. Rizal, M. Hasyim, M. F. Amrulloh, and Z. N. Saadah, "Feature Extraction and Naïve Bayes Algorithm for Defect Classification of Manalagi Apples," *Journal of Physics: Conference Series*, vol. 2394, no. 1. IOP Publishing, p. 012014, Dec. 01, 2022 [Online]. Available: <http://dx.doi.org/10.1088/1742-6596/2394/1/012014>
- [5] N. Kaur and V. Devendran, "Plant Leaf Disease Diagnostic System Built on Feature Extraction and Ensemble Classification," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-3, doi: 10.1109/ICRITO51393.2021.9596070.
- [6] U. Barman, C. Pathak, and N. K. Mazumder, "Comparative assessment of Pest damage identification of coconut plant using damage texture and color analysis," *Multimedia Tools and Applications*. Springer Science and Business Media LLC, Jan. 25, 2023 [Online]. Available: <http://dx.doi.org/10.1007/s11042-023-14369-2>
- [7] E. Sánchez-DelaCruz, J. P. Salazar López, D. Lara Alabazares, E. Tello Leal, and M. Fuentes-Ramos, "Deep learning framework for leaf damage identification," *Concurrent Engineering*, vol. 29, no. 1. SAGE Publications, pp. 25–34, Mar. 2021 [Online]. Available: <http://dx.doi.org/10.1177/1063293X21994953>