

Literatur Review: Penerapan *Random Forest* untuk Klasifikasi Penyakit Tanaman Padi

Anang Muhamad Lutfi¹, Eko Purwadi¹, Kamaluddin¹, Yusuf Ali Hanaan¹, Perani Rosyani^{1*}

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}ananglutfi2003@gmail.com, ²ekopurwadi688@gmail.com, ³kamaludn.69@gmail.com, ⁴yusufalihanaan7@gmail.com, ⁵dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak—Indonesia merupakan negara agraris yang menjadikan sektor pertanian sebagai elemen penting dalam ekonomi. Penyakit pada tanaman padi menjadi ancaman serius bagi petani karena dapat menurunkan kualitas dan hasil panen secara signifikan. *Random Forest*, sebagai salah satu metode pembelajaran mesin, telah diimplementasikan dalam penelitian untuk mengklasifikasi jenis-jenis penyakit pada tanaman padi secara efektif. Studi ini mengkaji berbagai literatur terkait penerapan metode *Random Forest* dan beberapa algoritma lain seperti CNN, Decision Tree, dan SVM, dalam mendeteksi dan mengidentifikasi penyakit tanaman padi. Hasil kajian ini menunjukkan bahwa metode *Random Forest* memiliki performa akurasi tinggi, yang menjadikannya metode yang direkomendasikan untuk deteksi dini penyakit tanaman padi. Studi ini diharapkan dapat menjadi panduan untuk penelitian lebih lanjut dalam meningkatkan akurasi dan efisiensi metode klasifikasi penyakit padi.

Kata Kunci: *Random Forest*, Klasifikasi Penyakit, Tanaman Padi, Pembelajaran Mesin, Indonesia

Abstract—Indonesia is an agrarian country where the agricultural sector plays a vital role in the economy. Diseases in rice plants pose a serious threat to farmers as they can significantly reduce the quality and yield of the harvest. *Random Forest*, one of the machine learning methods, has been implemented in research to effectively classify types of diseases in rice plants. This study reviews various literatures related to the application of the *Random Forest* method and several other algorithms such as CNN, Decision Tree, and SVM in detecting and identifying rice plant diseases. The review shows that the *Random Forest* method has high accuracy performance, making it a recommended method for early detection of rice plant diseases. This study is expected to serve as a guide for further research to improve the accuracy and efficiency of rice disease classification methods.

Keywords: *Random Forest*, Disease Classification, Rice Plants, Machine Learning, Indonesia

1. PENDAHULUAN

Indonesia merupakan negara agraris yang dimana pertanian ini merupakan bidang yang memegang peranan penting dan harus dilindungi kestabilan pemrosesannya. Karena Sebagian besar penduduk Indonesia masih bergantung pada beras sebagai makanan pokok sehari harinya. Pertanian merupakan suatu aktivitas untuk memproses pangan dengan menggunakan sumber daya alam yang tersedia yang dilakukan oleh manusia. Pertanian ini dijadikan sebagai wadah yang dapat memberikan lapangan pekerjaan dan alat pembayaran negara melalui bidang ekspor dan impor.

Penyakit pada tanaman padi merupakan ancaman serius bagi para petani, yang dapat mengurangi kualitas beras yang dihasilkan dan bahkan menimbulkan kerugian besar. Tanaman yang terinfeksi penyakit biasanya menunjukkan gejala berupa bercak dengan pola dan warna yang khas. Gejala tersebut dapat terlihat pada berbagai bagian tanaman padi, seperti telinga, daun, batang, akar

Melalui tinjauan literatur ini, kami berharap untuk memberikan pemahaman yang lebih baik tentang kontribusi metode *random forest* dalam mengklasifikasi penyakit pada tanaman daun padi. Temuan kami diharapkan bisa dapat memberikan panduan untuk pengembangan dan penerapan metode yang lebih baik di bidang tanaman padi. Oleh karena itu, diharapkan penelitian ini dapat menambahkan peran yang bermakna untuk terciptanya metode yang akurat serta efisien.

2. METODOLOGI PENELITIAN

2.1 *Systematic Literatur Review* (SLR)

Sistematisasi Literatur Review (SLR) adalah suatu metode untuk mengumpulkan, dan mengevaluasi literatur yang relevan dengan bidang studi yang di pelajari. Metode digunakan untuk memberikan bukti yang relevan dengan topik penelitian yang telah ditentukan.

2.2 *Random Forest*

Random Forest merupakan algoritma yang andal untuk menangani data dalam tugas-tugas klasifikasi dan regresi. Dengan menggabungkan banyak pohon keputusan melalui proses training pada sampel data, algoritma ini memberikan prediksi yang akurat serta mengurangi risiko overfitting. *Random Forest* melakukan klasifikasi menggunakan metode voting keputusan terbanyak berdasarkan dari pohon-pohon yang telah dibentuk (Rosyani et al., 2021). Meskipun membutuhkan waktu komputasi lebih lama, *Random Forest* tetap menjadi solusi yang fleksibel untuk berbagai keperluan analisis.

2.3 *Convolutional Neural Network (CNN)*

Convolutional Neural Network (CNN) adalah algoritma deep learning yang terinspirasi dari sistem visual manusia, dirancang untuk mendeteksi dan mengenali pola serta fitur penting dalam citra secara otomatis. *CNN* melakukan ekstraksi fitur melalui lapisan konvolusi dan secara simultan melakukan klasifikasi. Dalam konteks penelitian ini, *CNN* diterapkan untuk deteksi penyakit padi, mengintegrasikan proses deteksi dan ekstraksi fitur menjadi satu rangkaian (Auza et al., 2024). Pendekatan ini memungkinkan *CNN* untuk tidak hanya mengenali fitur-fitur yang relevan dari gambar tanaman padi, tetapi juga belajar dari fitur-fitur tersebut dengan akurasi tinggi.

2.4 *Naïve Bayes*

Naive Bayes merupakan metode yang menganalisis bukti-bukti secara terpisah. Metode ini juga dapat disetiap fitur bersifat tidak terikat dan baik untuk klasifikasi teks dan data kategorikal.

2.5 *Decision Tree*

Algoritma *Decision Tree* ini membuat "cabang-cabang" keputusan berdasarkan pertanyaan ya/tidak, serta juga menjelaskan bahwa kelebihan utamanya adalah interpretabilitas yang tinggi serta kemampuan dalam menangani berbagai tipe data.

2.6 *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah metode solusi optimal untuk pemisahan kelas data yang menggunakan prinsip hyperplane untuk menciptakan batas keputusan dengan jarak maksimum antar kelas, lalu dapat ditingkatkan kemampuannya dengan menggunakan teknik kernel agar dapat menangani data non-linear.

2.7 *K-Nearest Neighbors (KNN)*

Metode ini menggunakan konsep klasifikasi berdasarkan kedekatan karakteristik, yang nantinya data baru akan diklasifikasikan berdasarkan mayoritas label dari sejumlah k data terdekat yang ada di sekitarnya, metode ini sudah terbukti menghasilkan yang memuaskan, terutama untuk dataset dengan skala kecil sampai ke skala menengah.

3. ANALISA DAN PEMBAHASAN

Pada bagian ini berisi tujuan penelitian, hasil dari topik penelitian

Tabel 1. *Systematic Literatur Review (SLR)*

No	Nama dan Tahun	Metode yang Dibahas	Tujuan Penelitian	Hasil yang didapat
1	(Rozi & Badri Tamam, 2024)	Convolutional Neural Network (CNN) Dan Random Forest	Jurnal ini bertujuan untuk mengidentifikasi penyakit tanaman padi dengan metode CNN dan random forest untuk menciptakan model yang dapat mengenali penyakit padi dengan	1. Metode CNN Dengan hasil data training dan validation pada metode CNN diperoleh hasil akurasi yang baik, yaitu dengan nilai akurasi pelatihan dan validasi mencapai 0,91 pada epoch 10. Serta akurasi untuk

			akurasi tinggi. Sehingga petani dapat membuat tindakan yang tepat	<p>meneteksi penyakit pada daun padi sebesar 87,33%. Model ini memberikan akurasi yang baik dalam klasifikasi penyakit daun padi</p> <p>2. Metode random forest Pada hasil dari data training dengan metode ini menunjukkan 4 kategori yaitu, sehat, sakit ringan, sakit sedang, dan sakit berat dengan memberikan kinerja bernilai 0.99976 dimana hal ini mendekati 1, sehingga menunjukkan akurasi yang tinggi</p>
2	(Agustiani et al., n.d, 2022)	Random Forest dan Color Histogram	Penelitian pada jurnal ini bertujuan untuk menemukan metode klasifikasi penyakit daun padi, agar dapat mengurangi resiko agar penyakit dapat di deteksi dini sehingga dapat meningkatkan produksi .	Pada hasil yang di dapatkan dari penelitian, klasifikasi terhadap penyakit daun padi ini memberikan kesimpulan pengujian dari beberapa perbandingan. Dengan hasil akurasi tertinggi didapatkan sebesar 99.65% hal ini menunjukkan pengklasifikasian yang akurat .
3	Purnamawati et al., 2020)	Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN	<p>Penelitian ini bertujuan untuk menggunakan pembelajaran mesin untuk membantu petani mendeteksi penyakit padi dengan lebih cepat dan akurat. Penelitian ini mencoba untuk menentukan metode yang paling dapat diandalkan untuk penerapan di bidang ini dengan membandingkan lima algoritma utama: Naive Bayes, Random Forest, Decision Tree, SVM, dan KNN.</p> <p>Selain itu, penelitian ini bertujuan untuk memberikan solusi praktis dan berbasis informasi untuk mengurangi risiko kehilangan hasil panen akibat penyakit, sekaligus</p>	Hasil ini menunjukkan bahwa KKN memiliki konsistensi yang baik pada data pelatihan dan pengujian tanpa masalah overfitting. Sebaliknya, algoritma seperti Decision Tree, Random Forest, dan Naive Bayes cenderung terlalu menyesuaikan data pelatihan (overfitting), sehingga kurang akurat pada data baru. Algoritma SVM menghasilkan performa terendah karena mengalami underfitting

			mempromosikan penggunaan teknologi mutakhir untuk meningkatkan efisiensi dan produktivitas petani.	
4	(Habib Hawari et al., 2022)	Algoritma CNN (Convolutional Neural Network)	Penelitian ini memiliki tujuan untuk mempelajari dan membantu bagaimana cara petani untuk tahu cara mengenali berbagai macam penyakit tanaman padi dan meminimalisasi resiko gagal panen akibat penyakit daun padi. Penelitian ini dicapai menggunakan data training, data testing, dan data validation.	Hasil dari penelitian ini adalah, metode deep learning CNN dapat diimplementasikan untuk meneliti citra daun padi yang terkena penyakit. dari nilai data tertinggi yang didapat pada data training yaitu 85%, lalu untuk data testing nya 86%, dan nilai data validation yaitu 95%. dengan nilai setinggi ini maka dapat di terapkan nya proses citra penyakit daun padi dengan cukup maksimal.
5	(Julianto et al., n.d, 2022)	Algoritma Convolutional Neural Network (CNN)	Penelitian ini bertujuan untuk mengembangkan sistem deteksi penyakit tanaman padi yang efektif kita menggunakan arsitektur MobileNet-V2 yang akan berfokus ke optimasi hyperparameter, tujuannya sendiri untuk meningkatkan akurasi klasifikasi 3 jenis penyakit utamanya yaitu blast, blight dan tungro	Setelah peneliti melakukan eksperimen yang mendalam, dapat saya temukan konfigurasi optimal yang sangat menjanjikan untuk model metode ini. Dengan menggunakan epoch sebanyak 100, ukuran batch 32, learning rate yang hanya mencapai 0.001, dan juga pada optimizer RMSProp, model ini berhasil mencapai hasil yang menurut saya sangat signifikan. Lalu pada akurasi yang tercatat sebesar 97.56%, denan oresisi yang mencapai angka 97.64%, dengan recall di angka 97.57%, dan f1-score juga yang mencapai angka yang mirip yaitu menyentuh 97.57%. Hail ini juga menunjukkan bahwa pengaturan parameter yang sangat tepat sehingga berpengaruh terhadap performa model nya itu sendiri, dan memberikan dasar yang kuat untuk si peneliti agar dapat meneliti lebih jauh lagi dalam bidang ini.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa metode Random Forest memiliki potensi besar dalam meningkatkan akurasi dan efisiensi dalam klasifikasi penyakit tanaman padi dibandingkan dengan beberapa algoritma lain. Dengan penerapan yang tepat, metode ini dapat membantu petani mendeteksi penyakit pada tanaman lebih awal, sehingga mencegah kerugian besar dan mendukung peningkatan produksi pertanian. Studi ini juga menunjukkan bahwa penggunaan metode lain seperti CNN dan Decision Tree memberikan hasil yang baik, namun Random Forest secara konsisten memberikan akurasi yang lebih tinggi dalam sebagian besar kasus uji. Penelitian lanjutan disarankan untuk mengeksplorasi lebih lanjut kombinasi algoritma dan optimasi hyperparameter untuk memperoleh hasil yang lebih optimal dalam deteksi penyakit tanaman padi.

REFERENCES

- Agustiani, S., Tajul Arifin, Y., Junaidi, A., Khotimatul Wildah, S., & Mustopa, A. (n.d.). *Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram 1,** (Vol. 10, Issue 1). <https://www.kaggle.com/vbookshelf/rice-leaf-1>
- Auza, H., Bagus Arisila Putra, M., Azril Saputra, M., Hartono, R., & Rosyani, P. (2024). Implementasi Deep Learning untuk Deteksi Wajah dan Ekspresi menggunakan Algoritma Convolutional Neural Network (CNN) dengan OpenCV. *In Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan (Vol. 1, Issue 4)*. <https://jurnalmahasiswa.com/index.php/aidanspk>
- Habib Hawari, F., Fadillah, F., Rifqi Alviandi, M., & Arifin, T. (2022). KLASIFIKASI PENYAKIT PADI MENGGUNAKAN ALGORITMA CNN (CONVOLUTIONAL NEURAL NETWORK). *JURNAL RESPONSIF*, 4(2), 184–189. <https://ejurnal.ars.ac.id/index.php/jti>
- Julianto, A., Sunyoto, A., Ferry, D., & Wibowo, W. (n.d.). OPTIMASI HYPERPARAMETER CONVOLUTIONAL NEURAL NETWORK UNTUK KLASIFIKASI PENYAKIT TANAMAN PADI (OPTIMIZATION OF CONVOLUTIONAL NEURAL NETWORK HYPERPARAMETERS FOR CLASSIFICATION OF RICE PLANT DISEASES). <https://www.kaggle.com/tedisetiady/leaf-rice-dis>
- Purnamawati, A., Nugroho, W., Putri, D., & Hidayat, W. F. (2020). Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan Attribution-NonCommercial 4.0 International. Some rights reserved* 5(1). <https://doi.org/10.30743/infotekjar.v5i1.2934>
- Rosyani, P., Saprudin, S., & Amalia, R. (2021). Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO). *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 9(2), 132. <https://doi.org/10.26418/justin.v9i2.44120>
- Rozi, K., & Badri Tamam, M. (2024). KLASIFIKASI PENYAKIT TANAMAN PADI MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) DAN RANDOM FOREST (Vol. 7, Issue 2).