Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Hakikat Realitas Visual Penyakit Daun Kentang dalam Perspektif Filsafat Ilmu melalui Analisis Attention Map Model Vision Transformer

The Nature of Visual Reality in Potato Leaf Disease from a Philoshopy of Science Perspective through Vision Transformers Attention Map Analysis

Puji Utami Rakhmawati[1,2,*], Syaad Patmanthara[1]

[1]Teknik Elektro dan Informatika, Universitas Negeri Malang, Malang, Indonesia
Jl. Cakrawala No.5, Sumbersari, Kec. Lowokwaru, Kota Malang, Jawa Timur, Indonesia
[2]Ilmu Komputer, Universitas PGRI Wiranegara, Pasuruan, Indonesia
Jl. Ki Hajar Dewantara No.27-29, Tembokrejo, Kec. Purworejo, Kota Pasuruan, Jawa Timur, Indonesia
Email: puji.utami.2605349@students.um.ac.id
Email Penulis Korespondensi: puji.utami.2605349@students.um.ac.id
Received: **16/09/2026**, Revised: **30/10/2026**, Accepted: **30/10/2026**, Published**: 31/10/2026**

Abstrak-Hasil prediksi kecerdasan buatan merupakan pengetahuan komputasional yang perlu diuji melalui data, metrik evaluasi, dan interpretasi visual. Dalam pengenalan penyakit daun kentang, Attention Map dapat digunakan untuk mengamati bagaimana Vision Transformer membentuk representasi visual yang mendukung keputusan klasifikasi. Pada aksiologis, penelitian ini menekankan nilai kebermanfaatan, transparansi, dan tanggung jawab penggunaan kecerdasan buatan melalui implementasi sistem PotatoScan berbasis web sebagai sarana pendukung identifikasi awal penyakit daun kentang. Penelitian ini bertujuan menganalisis kemampuan ViT-B/16 dalam mengenali tiga kondisi daun kentang, yaitu Early, Healthy, dan Late serta menginterpretasikan pola perhatian internal model. Dataset terdiri atas 65 citra yang terbagi secara seimbang ke dalam tiga kelas. Citra diubah menjadi 224 × 224 piksel, dinormalisasi, dan digunakan pada proses fine-tuning. Model terbaik diperoleh pada epoch ke-8 dengan validation loss 0,3547 dan akurasi validasi 77,78%. Pengujian terhadap 12 citra menghasilkan akurasi 75%, macro precision 78%, macro recall 75%, dan macro F1-score 74%. Kelas bercak kering menyebar memperoleh recall 100%, sedangkan kelas bercak tepi daun memperoleh recall 50%. Attention Map digunakan untuk mengamati distribusi perhatian model terhadap bagian citra yang mendukung prediksi dan tidak diposisikan sebagai bukti penyebab biologis penyakit. Model diimplementasikan pada aplikasi web PotatoScan menggunakan PyTorch dan timm. Pada satu contoh pengujian, sistem menghasilkan confidence 90,5% dengan waktu inferensi 1,352 detik. Kontribusi penelitian ini terletak pada penerapan Vision Transformer (ViT-B/16) yang dipadukan dengan analisis attention map untuk mengevaluasi hasil klasifikasi penyakit daun kentang. Penelitian ini tidak hanya mengukur akurasi model, tetapi juga mengkaji keterbatasannya dalam membedakan pola visual Early Blight dan Late Blight. Melalui perspektif filsafat ilmu, penelitian ini memberikan pemahaman mengenai hubungan antara representasi visual, hasil prediksi, dan batas interpretasi model kecerdasan buatan.

Kata Kunci: Vision Transformer; Attention Map; Penyakit Daun Kentang; Klasifikasi Citra; Explainable Artificial Intelligence

Abstract-Artificial intelligence predictions represent computational knowledge that requires validation through data, evaluation metrics, and visual interpretation. In potato leaf disease classification, attention maps can be used to examine how Vision Transformers develop visual representations that support classification decisions. From an axiological perspective, this study emphasizes the practical value, transparency, and responsible use of artificial intelligence through the implementation of PotatoScan, a web-based system designed to support the preliminary identification of potato leaf diseases. This study aims to evaluate the performance of ViT-B/16 in classifying three potato leaf conditions, namely Early Blight, Healthy, and Late Blight, while interpreting the model's internal attention patterns. The dataset consists of 65 images distributed across the three classes. The images were resized to 224 × 224 pixels, normalized, and used for model fine-tuning. The best model was obtained at epoch 8, achieving a validation loss of 0.3547 and a validation accuracy of 77.78%. Testing on 12 images resulted in an accuracy of 75%, macro precision of 78%, macro recall of 75%, and macro F1-score of 74%. The Early Blight class achieved a recall of 100%, whereas the Late Blight class achieved a recall of 50%. Attention maps were used to examine the distribution of the model's attention across image regions associated with its predictions, rather than as evidence of the biological causes of disease. The model was implemented in the PotatoScan web application using PyTorch and timm. In one test case, the system achieved a prediction confidence of 90.5% with an inference time of 1.352 seconds. The main contribution of this study lies in integrating ViT-B/16 classification with attention map analysis to evaluate both predictive performance and limitations in distinguishing visually similar disease classes. Beyond classification accuracy, the study provides a philosophical perspective on the relationship between visual representations, computational predictions, and the limitations of artificial intelligence interpretation. From an axiological standpoint, PotatoScan demonstrates the potential of web-based artificial intelligence as an accessible tool for preliminary disease identification while emphasizing the importance of transparent interpretation and responsible decision-making.

Keywords: Vision Transformer; Attention Map; Potato Leaf Disease; Image Classification; Explainable Artificial Intelligence

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 236
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

1. PENDAHULUAN

Kajian terhadap Attention Map menimbulkan pertanyaan yang lebih mendasar mengenai pengetahuan yang dihasilkan oleh model. Pertanyaan filsafat yang diajukan dalam penelitian ini adalah: (1) secara ontologis, apa yang sebenarnya direpresentasikan oleh ViT-B/16 dan Attention Map ketika mengenali penyakit daun kentang? (2) secara epistemologis, bagaimana hasil klasifikasi dan Attention Map dapat diperoleh, diuji, dan dipercaya sebagai pengetahuan berbasis data? dan (3) secara aksiologis, bagaimana hasil klasifikasi dan Attention Map dapat dimanfaatkan secara bertanggung jawab sebagai alat bantu identifikasi penyakit daun kentang? Ketiga pertanyaan tersebut berkaitan dengan ontologi yang membahas hakikat realitas, epistemologi yang membahas cara memperoleh dan memvalidasi pengetahuan, serta aksiologi yang membahas nilai dan manfaat ilmu [1], [2] .

Perkembangan kecerdasan buatan memberikan peluang besar dalam pengolahan citra untuk membantu klasifikasi penyakit tanaman. Pada daun kentang, kondisi Healthy, Early Blight, dan Late Blight memiliki karakteristik visual yang dapat dikenali melalui citra, tetapi pemeriksaan secara langsung masih bergantung pada pengalaman pengamat dan berpotensi bersifat subjektif. Zeynalov dan Çakmak [3] menunjukkan bahwa model berbasis Vision Transformer mampu memberikan performa tinggi pada klasifikasi penyakit kentang. Hal ini menunjukkan bahwa pendekatan berbasis citra dan kecerdasan buatan dapat digunakan sebagai alat bantu klasifikasi yang lebih konsisten.

Vision Transformer (ViT) memproses citra dengan membaginya menjadi sejumlah patch yang kemudian diubah menjadi token dan diproses menggunakan mekanisme self-attention [4]. Mekanisme ini memungkinkan model mempelajari hubungan antarbagian citra secara global. Sejumlah penelitian menunjukkan potensi ViT pada klasifikasi penyakit tanaman. Model hybrid CNN-ViT pada daun cabai memperoleh akurasi 93% [5], sedangkan Attention Score Based Multi Vision Transformer mencapai akurasi lebih dari 99% pada beberapa dataset tanaman [6]. Pada penyakit daun kentang, EfficientRMT-Net yang menggabungkan ResNet-50 dan ViT mencapai akurasi 99,12% pada dataset khusus kentang [7]. Penelitian lain juga menekankan pentingnya evaluasi performa, generalisasi, augmentasi, dan kondisi dataset dalam penerapan ViT pada citra penyakit tanaman [8], [9], [10]. Pendekatan attention berbasis fitur tekstur juga dilaporkan dapat meningkatkan ketahanan representasi pada klasifikasi penyakit daun [11]. Pendekatan Explainable Artificial Intelligence (XAI) menjadi bagian penting dalam pengembangan model klasifikasi berbasis deep learning karena memungkinkan hasil prediksi dianalisis melalui interpretasi visual. Li et at. [12] . Thakur et al. mengkaji integrasi Convolutional Neural Network (CNN) dan Vision Transformer (ViT) serta memanfaatkan metode interpretabilitas seperti Gradient-weighted Class Activation Mapping (Grad-CAM) dan Local Interpretable Model-agnostic Explanations (LIME) untuk membantu memahami hasil klasifikasi. Pendekatan tersebut menunjukkan pentingnya visualisasi dalam mengevaluasi informasi visual yang digunakan model [13] .

Meskipun performa klasifikasi penting, angka akurasi saja belum menjelaskan bagaimana model memproses citra. Karena itu, penelitian ini menggunakan Attention Map untuk mengamati distribusi perhatian internal ViT terhadap patch citra. Pada ViT-B/16 dengan masukan 224 × 224 piksel, citra dibagi menjadi 196 patch berukuran 16 × 16 piksel. Bobot self-attention kemudian digunakan untuk melihat bagian citra yang memperoleh perhatian relatif lebih tinggi. Visualisasi heatmap pada klasifikasi penyakit daun telah digunakan untuk memperlihatkan area citra yang berkontribusi terhadap keputusan model [14], sedangkan pengawasan menggunakan explainable attention map dapat mendorong pola perhatian ViT yang lebih konsisten secara semantik [15] . Namun, Attention Map dalam penelitian ini tidak diposisikan sebagai bukti penyebab biologis penyakit, melainkan sebagai visualisasi pola perhatian internal model ketika menghasilkan klasifikasi.

Dalam penelitian ini, penyakit daun kentang dipandang sebagai realitas fisik, citra daun sebagai representasi digital dari realitas tersebut, sedangkan Attention Map merupakan representasi komputasional mengenai bagian citra yang memperoleh perhatian relatif lebih tinggi dari model. Dengan demikian, penelitian tidak hanya menilai apakah model mampu menghasilkan prediksi yang benar, tetapi juga mempertanyakan apa yang direpresentasikan oleh model, bagaimana hasil prediksi tersebut dapat dipercaya, dan bagaimana hasilnya digunakan secara bertanggung jawab.

Kesenjangan dalam penelitian ini berkaitan dengan kemampuan model dalam menghasilkan prediksi yang akurat dan bagaimana hasil prediksi tersebut dapat dipahami secara ilmiah. Tingkat akurasi menunjukkan kemampuan model dalam mengenali penyakit daun kentang berdasarkan data pengujian, tetapi belum dapat menjelaskan apakah model benar-benar memperhatikan karakteristik visual yang berkaitan dengan penyakit. Melalui Attention Map, pola perhatian model terhadap bagian tertentu pada citra dapat diamati, meskipun visualisasi tersebut belum sepenuhnya menjelaskan alasan di balik keputusan yang dihasilkan. Oleh karena itu, pendekatan filsafat ilmu diperlukan untuk memahami hubungan antara kondisi penyakit yang sebenarnya, informasi visual yang diproses oleh model, dan pengetahuan yang dihasilkan melalui proses komputasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengevaluasi kemampuan model Vision Transformer (ViT-B/16) dalam mengklasifikasikan tiga kondisi daun kentang, yaitu Healthy, Early Blight, dan Late Blight, serta menganalisis hasil Attention Map melalui perspektif ontologi, epistemologi, dan aksiologi. Penelitian ini tidak hanya berfokus pada ketepatan hasil klasifikasi, tetapi juga berupaya memahami bagaimana model memproses informasi visual dan membentuk pola perhatian terhadap bagian tertentu pada citra daun kentang. Selain itu, penelitian ini bertujuan menerapkan model ViT-B/16 pada sistem PotatoScan berbasis web

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 237
This work is licensed under a Creative Commons Attribution 4.0 International License.

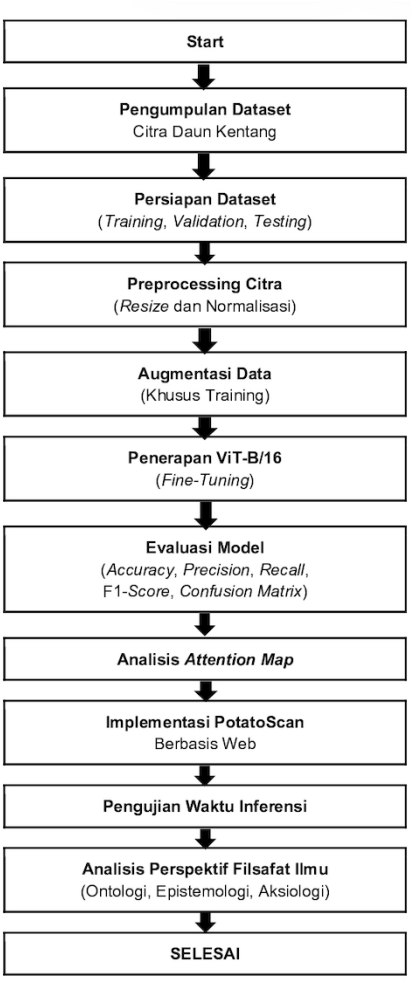
Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

agar proses identifikasi awal penyakit daun kentang dapat dilakukan secara cepat, praktis, dan mudah diakses oleh pengguna. Kontribusi penelitian ini mencakup analisis hubungan antara citra digital, representasi visual yang dipelajari model, dan hasil klasifikasi, serta evaluasi keandalan prediksi berdasarkan pengujian empiris. Melalui PotatoScan, pengguna dapat memperoleh hasil klasifikasi, tingkat keyakinan prediksi, dan visualisasi Attention Map dalam satu sistem. Dengan demikian, penelitian ini diharapkan dapat mendukung pemanfaatan kecerdasan buatan yang tidak hanya memperhatikan akurasi dan kecepatan pemrosesan, tetapi juga transparansi, kemudahan penggunaan, serta tanggung jawab dalam menginterpretasikan hasil prediksi.

2. METODOLOGI PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif dengan analisis interpretabilitas model. Tahapan penelitian terdiri atas persiapan dataset citra daun kentang, preprocessing dan augmentasi, pelatihan serta evaluasi ViT-B/16, ekstraksi Attention Map, analisis pola perhatian, dan implementasi hasil klasifikasi pada aplikasi berbasis web. Tahapan ini ditunjukkan pada gambar 1. ViT-B/16 dipilih karena memproses citra sebagai sekumpulan patch dan memanfaatkan mekanisme self-attention untuk mempelajari hubungan global antar bagian citra [2]. Fokus penelitian tidak hanya pada performa klasifikasi, tetapi juga pada pola perhatian internal model ketika membedakan kelas Healthy, Early Blight, dan Late Blight.



Gambar 1. Diagram Alur Penelitian

2.2 Dataset

Dataset yang digunakan terdiri atas 65 citra daun kentang yang berasal dari penelitian sebelumnya. Dataset dibagi secara seimbang menjadi tiga kelas, yaitu Healthy sebanyak 20 citra, Early Blight sebanyak 20 citra, dan Late Blight sebanyak 25 citra. Keseimbangan jumlah data digunakan untuk mengurangi dominasi salah satu kelas pada proses pembelajaran. Dataset selanjutnya dipisahkan menjadi data pelatihan, validasi, dan pengujian secara terstratifikasi agar proporsi ketiga kelas tetap terjaga. Persentase pembagian data mengikuti konfigurasi eksperimen yang digunakan pada implementasi akhir dan perlu dilaporkan sesuai kode pelatihan yang benar-benar

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 238
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

dijalankan. Langkah ini penting karena karakteristik dan pembagian dataset dapat memengaruhi performa serta kemampuan generalisasi model [9], [10].

2.3 Preprocessing dan Augmentasi Citra

Seluruh citra diubah ukurannya menjadi 224 × 224 piksel agar sesuai dengan ukuran masukan ViT-B/16. Citra kemudian dikonversi menjadi tensor dan dinormalisasi mengikuti konfigurasi model pralatih yang digunakan. Augmentasi hanya diterapkan pada data pelatihan untuk menambah variasi citra dan membantu mengurangi risiko overfitting. Transformasi yang digunakan mencakup rotasi, horizontal flip, dan transformasi geometris atau zoom sesuai konfigurasi pada kode eksperimen. Nilai parameter augmentasi tidak ditetapkan secara terpisah dalam naskah ini sebelum diverifikasi dari implementasi akhir agar metode yang dilaporkan tetap sama dengan proses pelatihan yang benar-benar dilakukan. Penggunaan augmentasi dan fine-tuning juga telah digunakan pada penelitian ViT untuk meningkatkan ketahanan model terhadap variasi citra [9].

2.4 Penerapan Vision Transformer

Model utama yang digunakan adalah Vision Transformer ViT-B/16. Arsitektur ViT mengubah citra menjadi sekumpulan patch yang kemudian diperlakukan sebagai urutan token dan diproses menggunakan Transformer [4] . Pada citra dengan tinggi H, lebar W, dan ukuran patch P × P, jumlah patch dapat dinyatakan sebagai:

$$N = \frac{(H \times W)}{P^2} \tag{1}$$

Pada penelitian ini, citra berukuran 224 × 224 piksel dan ukuran patch ViT-B/16 adalah 16 × 16 piksel, sehingga:

$$N = \frac{(224 \times 224)}{16^2} = 196 \tag{2}$$

Dengan demikian, satu citra menghasilkan 196 patch. Persamaan (2) merupakan penerapan konfigurasi ViT-B/16 pada penelitian ini, sedangkan konsep pembagian citra menjadi patch mengacu pada Dosovitskiy et al. [4]. Setiap patch kemudian diratakan dan diproyeksikan secara linear menjadi patch embedding. Sebuah class token ditambahkan pada awal urutan dan positional embedding ditambahkan agar informasi posisi setiap patch tetap dipertahankan. Formulasi masukan ViT mengacu pada Dosovitskiy et al. [4] :

$$z_0 = \left[x_{class}; x_p^1 E; x_p^2 E; \ldots; x_p^N E\right] + E_{pos} \tag{3}$$

dengan x_p^i merupakan patch citra ke-i, E adalah matriks proyeksi patch, E_pos adalah positional embedding, dan x_class merupakan class token.

Selanjutnya, token diproses menggunakan mekanisme self-attention. Setiap token diproyeksikan menjadi Query (Q), Key (K), dan Value (V). Scaled dot-product attention dirumuskan sebagai [6]:

$$\text{Attention}(Q, K, V) = \left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

dengan d_k merupakan dimensi vektor key. Baek [6] menggunakan formulasi self-attention tersebut dalam konteks Vision Transformer untuk klasifikasi penyakit tanaman. ViT menggunakan beberapa attention head secara paralel. Mekanisme Multi-Head Attention diimplementasikan oleh Baek [6] sebagai:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{5}$$

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{6}$$

Mekanisme ini memungkinkan attention head mempelajari hubungan antartoken dari representasi yang berbeda. Dalam ViT, hubungan tersebut digunakan untuk menangkap informasi dari berbagai bagian citra [4]. Pada setiap blok Transformer, hasil multi-head self-attention diproses menggunakan Layer Normalization, koneksi residual, dan Multilayer Perceptron (MLP). Formulasi blok encoder ViT mengacu pada Dosovitskiy et al. [4] :

$$z'_l = \{MSA\}(\{LN\}(z_\{l-1\})) + z_\{l-1\} \tag{7}$$

$$z_l = \{MLP\}(\{LN\}(z'_l)) + z'_l \tag{8}$$

dengan l menunjukkan blok Transformer ke-l. Setelah seluruh blok Transformer dilalui, representasi akhir class token diunakan untuk proses klasifikasi [4] . Pada penelitian ini, classification head disesuaikan menjadi tiga kelas, yaitu Healthy, Early Blight, dan Late Blight.

2.5 Ekstraksi dan Analisis Attention Map

Bobot perhatian pada mekanisme self-attention berasal dari hasil normalisasi hubungan antara Query dan Key. Berdasarkan Persamaan (4), matriks bobot perhatian dapat dinyatakan sebagai [6]:

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 239
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

$$A = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (9)$$

Matriks A menunjukkan bobot relatif hubungan antartoken. Baek [6] menjelaskan bahwa bobot attention menentukan seberapa besar informasi dari suatu patch berkontribusi terhadap patch lainnya dalam proses self-attention. Mekanisme visual attention juga dapat dipadukan dengan fitur tekstur untuk memperkuat representasi area penting pada citra penyakit tanaman [11] . Dalam penelitian ini, bobot perhatian tersebut digunakan sebagai dasar untuk membentuk Attention Map sehingga distribusi perhatian internal ViT terhadap patch citra dapat diamati.

Attention Map diposisikan sebagai visualisasi pola perhatian internal ViT, bukan sebagai bukti penyebab biologis penyakit. Analisis dilakukan dengan mengamati distribusi perhatian pada citra untuk masing-masing kelas Healthy, Early Blight, dan Late Blight. Area dengan nilai perhatian relatif lebih tinggi diamati untuk mengetahui apakah perhatian model cenderung berada pada bagian daun yang memiliki pola visual tertentu. Agar hasil dapat direproduksi, implementasi akhir perlu menyebutkan secara jelas layer dan head yang digunakan, apakah attention dirata-ratakan antar-head, serta cara token kelas dihubungkan dengan patch citra. Informasi tersebut harus mengikuti kode Attention Map yang benar-benar digunakan dalam eksperimen.

2.6 Evaluasi Model

Kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score. Accuracy menunjukkan proporsi keseluruhan prediksi yang benar. Precision menunjukkan ketepatan prediksi pada suatu kelas, recall menunjukkan kemampuan model menemukan sampel yang benar pada kelas tersebut, sedangkan F1-score merupakan rata-rata harmonis antara precision dan recall. Confusion matrix digunakan sebagai pendukung untuk melihat pola kesalahan klasifikasi antara Healthy, Early Blight, dan Late Blight. Waktu inferensi juga dicatat untuk menilai efisiensi pemrosesan sistem, sebagaimana evaluasi aplikasi ViT pada skenario waktu nyata [16] . Evaluasi tidak hanya diarahkan pada nilai akurasi tertinggi karena penelitian penyakit tanaman menunjukkan pentingnya generalisasi, kondisi dataset, interpretabilitas, dan kesiapan implementasi model [8], [9], [10], [17], [18]. Penggabungan ViT ringan dan modul attention pada objek pertanian lain juga menunjukkan kemampuan mengenali fitur visual yang halus dengan waktu inferensi yang tetap kompetitif [19] .

2.7 Implementasi Sistem Berbasis Web dan Pengujian Waktu

Model yang telah dilatih diintegrasikan ke dalam aplikasi berbasis web menggunakan Python. Pengguna mengunggah citra daun kentang melalui halaman web, kemudian sistem melakukan preprocessing dengan ukuran dan normalisasi yang sama seperti pada tahap pengujian. Citra diproses oleh ViT-B/16 untuk menghasilkan kelas prediksi. Pada tahap berikutnya, sistem membentuk Attention Map dari model dan menampilkannya bersama hasil klasifikasi. Integrasi model diagnosis penyakit tanaman ke dalam aplikasi web dapat menjembatani performa model dengan pemanfaatan praktis oleh pengguna [17]. Alur sistem secara umum adalah: unggah citra, preprocessing, inferensi ViT-B/16, prediksi kelas, ekstraksi Attention Map, visualisasi hasil pada web. Alur tersebut ditunjukkan pada gambar 1. Waktu pemrosesan dicatat sebagai informasi tambahan mengenai respon sistem, tetapi pengakuan kecepatan hanya diberikan berdasarkan hasil pengukuran aktual pada perangkat yang digunakan.

2.8 Analisis dalam Perspektif Filsafat Ilmu

Analisis filsafat ilmu dalam penelitian ini diarahkan untuk menjawab tiga pertanyaan yang telah dirumuskan pada bagian pendahuluan. Pertama, pada aspek ontologis, analisis dilakukan untuk menjawab apa yang sebenarnya direpresentasikan oleh ViT-B/16 dan Attention Map ketika mengenali penyakit daun kentang. Penyakit daun kentang diposisikan sebagai realitas fisik, citra sebagai representasi digital dari kondisi tersebut, sedangkan Attention Map dipahami sebagai representasi komputasional dari distribusi perhatian internal model terhadap patch citra. Kedua, pada aspek epistemologis, analisis dilakukan untuk menjawab bagaimana hasil klasifikasi dan Attention Map diperoleh, diuji, dan dapat dipercaya sebagai pengetahuan berbasis data. Hal ini dilakukan melalui dataset citra, preprocessing, pelatihan ViT-B/16, serta evaluasi menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan pemeriksaan Attention Map. Dengan demikian, kepercayaan terhadap hasil model tidak hanya didasarkan pada nilai akurasi, tetapi juga pada proses pengujian dan interpretasi pola perhatian model. Ketiga, pada aspek aksiologis, analisis dilakukan untuk menjawab bagaimana hasil klasifikasi dan Attention Map dapat dimanfaatkan secara bertanggung jawab. Hasil prediksi dan visualisasi Attention Map diposisikan sebagai alat bantu dalam mengenali kondisi penyakit daun kentang dan bukan sebagai pengganti mutlak pengetahuan ahli. Attention Map juga tidak diperlakukan sebagai bukti penyebab biologis penyakit, tetapi sebagai visualisasi pola perhatian komputasional model. Pendekatan ini mengikuti landasan ontologi, epistemologi, dan aksiologi dalam filsafat ilmu [1], [2].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan dan Pengujian ViT-B/16

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 240
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Proses fine-tuning ViT-B/16 yang ditunjukkan pada tabel 1 dilakukan dengan menyimpan bobot berdasarkan validation loss terendah dan menerapkan early stopping. Pada tabel 1 ditunjukkan model mencapai validation loss terbaik sebesar 0,3547 pada epoch ke-8 dengan akurasi validasi 77,78%. Bobot pada epoch tersebut dipilih sebagai model terbaik untuk tahap pengujian. Setelah epoch ke-8, akurasi pelatihan terus meningkat hingga 96,08% pada epoch ke-15, tetapi akurasi validasi tetap 77,78% dan validation loss meningkat menjadi 0,5417. Pola ini menunjukkan bahwa model mulai mengalami overfitting.

Tabel 1. Ringkasan hasil fine-tuning ViT-B/16

Parameter	Hasil
Epoch terbaik	8
Akurasi training pada epoch terbaik	80,39%
Akurasi validasi	77,78%
Validation loss terbaik	0,3547
Akurasi training tertinggi	96,08%
Epoch penghentian	15

Model terbaik kemudian diuji menggunakan 12 citra yang tidak digunakan pada proses pelatihan. Pada tabel 2 ditunjukkan data uji terdiri atas kelas early, healthy, dan late dengan masing-masing empat citra hasil tersebut ditunjukkan pada tabel 2. Model mengklasifikasikan 9 dari 12 citra secara benar sehingga memperoleh akurasi pengujian 75%. Nilai macro precision mencapai 78%, macro recall 75%, dan macro F1-score 74%. Nilai macro average dan weighted average sama karena jumlah data pada setiap kelas seimbang.

Tabel 2. Hasil classification report ViT-B/16

Kelas	Precision	Recall	F1-score	Support
Early	0,67	1	0,8	4
Healthy	1	0,75	0,86	4
Late	0,67	0,5	0,57	4
Macro average	0,78	0,75	0,74	12
Weighted average	0,78	0,75	0,74	12

Hasil per kelas menunjukkan bahwa early memperoleh recall 100% karena seluruh empat citra berhasil dikenali. Kelas healthy memperoleh precision 100%, recall 75%, dan F1-score tertinggi sebesar 86%. Pada kelas late, hanya dua dari empat citra yang dikenali dengan benar, sehingga didapatkan recall 50% dan F1-score 57%. Dengan demikian, kelas late merupakan kelas yang paling sulit dikenali oleh model.

Tabel 3. Tabel Hasil Confusion Matrix

Aktual / Prediksi	Early Blight	Healthy	Late Blight
Early Blight	4	0	0
Healthy	0	3	1
Late Blight	2	0	2

Berdasarkan hasil confusion matrix yang ditunjukkan pada tabel 2, model ViT-B/16 berhasil mengklasifikasikan 9 dari 12 citra pengujian dengan benar, sehingga memperoleh akurasi sebesar 75%. Pada kelas Early Blight, seluruh citra berhasil dikenali dengan baik, sedangkan pada kelas Healthy, terdapat satu citra yang salah diklasifikasikan sebagai Late Blight. Kesalahan terbanyak terjadi pada kelas Late Blight, yaitu dua dari empat citra dikenali sebagai Early Blight, sehingga nilai recall hanya mencapai 50%. Hasil ini menunjukkan bahwa model masih mengalami kesulitan membedakan kedua jenis penyakit tersebut. Kemiripan pola visual antara Early Blight dan Late Blight diduga menjadi salah satu penyebabnya, sehingga diperlukan analisis lebih lanjut melalui attention map.

Confusion matrix memperlihatkan bahwa empat citra early diklasifikasikan secara benar tanpa kesalahan ditunjukkan pada tabel 3. Pada kelas healthy, tiga citra diprediksi dengan benar dan satu citra diprediksi sebagai late. Pada kelas late, dua citra diprediksi dengan benar, sedangkan dua citra lainnya diprediksi sebagai early. Kesalahan yang dominan antara late dan early mengindikasikan bahwa kedua kelas memiliki karakteristik visual yang berdekatan, seperti perubahan warna, pola bercak kecokelatan, dan distribusi kerusakan pada permukaan daun.

3.2 Analisis Attention Map

Analisis Attention Map dilakukan untuk mengamati distribusi perhatian internal model ViT-B/16 ketika mengklasifikasikan citra daun kentang. Visualisasi disajikan untuk tiga kelas, yaitu early, healthy, dan late. Setiap hasil visualisasi terdiri atas citra asli, Attention Map, hasil overlay, kelas prediksi, dan nilai confidence. Warna merah dan kuning menunjukkan bagian citra yang memperoleh perhatian relatif tinggi, sedangkan warna hijau dan biru menunjukkan bagian yang memperoleh perhatian lebih rendah.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 241
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Pada kelas early, perhatian model diharapkan terkonsentrasi pada bagian daun yang memiliki bercak kering berwarna kecokelatan dan pola kerusakan yang menyebar. Apabila wilayah dengan perhatian tinggi berimpit dengan bagian bercak, hal tersebut menunjukkan bahwa model menggunakan karakteristik visual yang relevan untuk menentukan kelas early. Namun, perhatian yang dominan pada latar belakang dapat menunjukkan bahwa prediksi dipengaruhi oleh fitur yang tidak berhubungan langsung dengan penyakit.

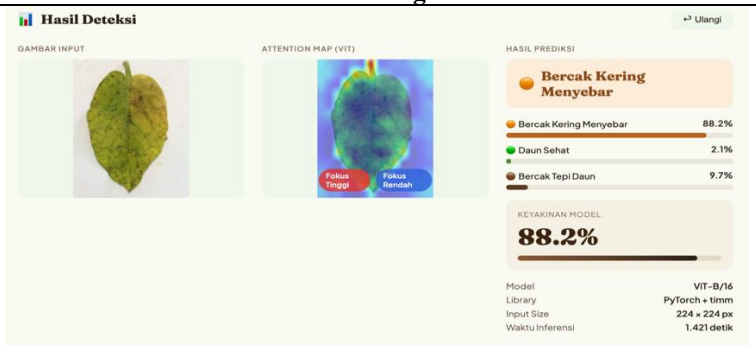
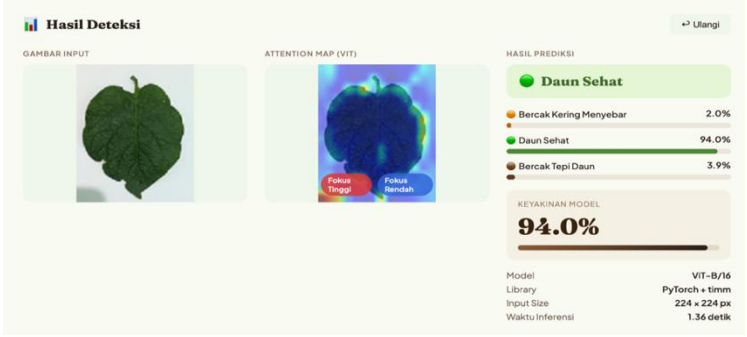
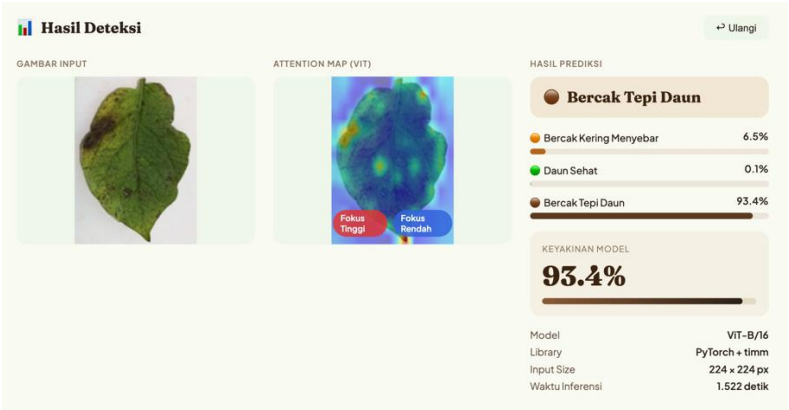
Pada kelas healthy, distribusi perhatian dapat terlihat lebih menyebar pada permukaan daun karena tidak terdapat lesi penyakit yang dominan. Model dapat memberikan perhatian pada bentuk daun, warna hijau, struktur tulang daun, dan keseragaman tekstur permukaan. Attention Map pada kelas ini perlu diperiksa untuk memastikan bahwa model mengenali kondisi sehat berdasarkan karakteristik daun dan bukan berdasarkan latar belakang citra.

Pada kelas late, perhatian model diharapkan berada pada perubahan warna atau bercak yang muncul di bagian tepi daun. Berdasarkan confusion matrix, kelas late memiliki recall terendah sebesar 50% dan dua citranya salah diprediksi sebagai early. Oleh karena itu, Attention Map pada kelas late penting untuk mengetahui apakah model telah memberikan perhatian pada tepi daun atau justru pada bercak yang memiliki kemiripan dengan kelas early.

Secara keseluruhan, Attention Map membantu memberikan interpretasi visual terhadap proses klasifikasi ViT-B/16. Meskipun demikian, area dengan perhatian tinggi tidak secara langsung membuktikan bahwa wilayah tersebut menjadi penyebab biologis penyakit. Visualisasi hanya menunjukkan distribusi perhatian komputasional model terhadap patch citra yang digunakan dalam menghasilkan prediksi.

Untuk mengetahu penerapan model ViT-B/16 pada aplikasi PotatoScan, dilakukan pengujian prediksi citra daun kentang melalui website. Hasil pengujian pada tig akelas, yaitu Early, Healthy dan Late, disajikan pada tabel 4.

Tabel 4. Tabel Pengujian dengan Website

Kelas	Hasil Prediksi dengan Website
Early	
Healthy	
Late	

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

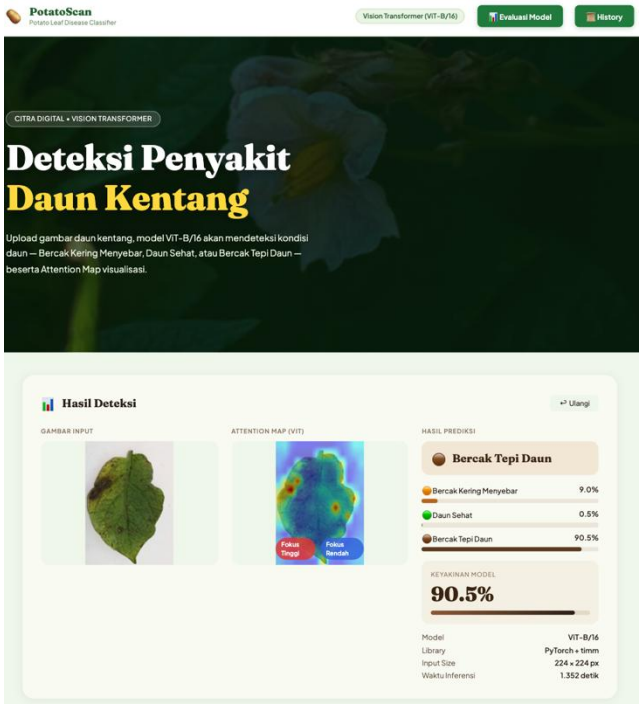
Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Berdasarkan Tabel 4, aplikasi PotatoScan dapat menampilkan hasil prediksi untuk ketiga kelas penyakit daun kentang. Pengujian tersebut menunjukkan bahwa model ViT-B/16 telah berhasil diintegrasikan ke dalam website sehingga pengguna dapat mengunggah citra daun kentang dan memperoleh hasil klasifikasi secara langsung.

3.3 Implementasi dan Waktu Inferensi Sistem

Model ViT-B/16 diimplementasikan ke dalam aplikasi berbasis web bernama PotatoScan menggunakan PyTorch dan library timm. Sistem dirancang agar pengguna dapat melakukan klasifikasi penyakit daun kentang dengan mengunggah citra melalui halaman web. Citra yang diterima sistem menjalani tahap preprocessing berupa perubahan ukuran menjadi 224 × 224 piksel, konversi citra menjadi tensor, dan normalisasi sesuai konfigurasi model. Citra hasil preprocessing selanjutnya diproses oleh ViT-B/16 untuk menghasilkan probabilitas pada tiga kelas, yaitu Bercak Kering Menyebar (Early), Daun Sehat (Healthy), dan Bercak Tepi Daun (Late).

Gambar 2 menunjukkan salah satu hasil implementasi sistem pada citra daun kentang. Model memprediksi citra tersebut sebagai kelas Bercak Tepi Daun dengan tingkat keyakinan sebesar 90,5%. Probabilitas untuk kelas Bercak Kering Menyebar adalah 9,0%, sedangkan probabilitas kelas Daun Sehat adalah 0,5%. Hasil tersebut menunjukkan bahwa model memberikan probabilitas tertinggi pada kelas Bercak Tepi Daun dengan selisih yang cukup besar dibandingkan dua kelas lainnya. Selain menampilkan kelas prediksi dan nilai keyakinan, sistem juga menghasilkan visualisasi Attention Map. Visualisasi ini digunakan untuk menunjukkan distribusi perhatian internal model ketika melakukan klasifikasi. Area berwarna merah dan kuning menunjukkan bagian citra yang memperoleh perhatian relatif lebih tinggi, sedangkan area berwarna biru menunjukkan bagian yang memperoleh perhatian lebih rendah. Pada contoh pengujian, perhatian model tersebar pada beberapa bagian permukaan daun, terutama pada area yang memiliki perubahan warna dan pola bercak. Attention Map membantu pengguna mengamati apakah keputusan model berkaitan dengan bagian daun yang relevan. Namun, visualisasi tersebut hanya menggambarkan pola perhatian komputasional model dan tidak menunjukkan penyebab biologis penyakit.



Gambar 2. Hasil Deteksi pada Halaman Web PotatoScan

Waktu inferensi yang tercatat pada contoh pengujian gambar 2 adalah 1,352 detik. Nilai tersebut menunjukkan bahwa sistem dapat menghasilkan prediksi kelas, nilai probabilitas, dan Attention Map dalam waktu relatif singkat setelah citra selesai diproses. Waktu tersebut merupakan hasil dari satu contoh pengujian sehingga belum dapat dianggap sebagai rata-rata kinerja keseluruhan sistem. Pengukuran lebih lanjut perlu dilakukan menggunakan sejumlah citra dan perangkat yang sama agar diperoleh nilai rata-rata, minimum, maksimum, serta simpangan baku waktu inferensi. Secara fungsional, implementasi berbasis web telah berhasil mengintegrasikan tahap unggah citra, preprocessing, inferensi ViT-B/16, prediksi kelas, ekstraksi Attention Map, dan visualisasi hasil dalam satu antarmuka.

3.4 Pembahasan

Akurasi validasi sebesar 77,78% dan akurasi pengujian sebesar 75% memiliki selisih yang relatif kecil, sehingga model terbaik masih menunjukkan kemampuan generalisasi pada data yang tidak digunakan dalam pelatihan.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 243
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Namun, jumlah data uji yang hanya 12 citra membuat hasil ini perlu ditafsirkan secara hati-hati karena satu kesalahan prediksi dapat mengubah nilai akurasi secara cukup besar. Kecenderungan overfitting juga terlihat dari akurasi training yang meningkat hingga 96,08%, sedangkan akurasi validasi tidak lagi meningkat setelah epoch ke-8.

Hasil tersebut sejalan dengan karakteristik Vision Transformer yang membutuhkan jumlah dan keragaman data memadai agar mampu membentuk representasi visual yang stabil [4] . Pada dataset terbatas, ViT lebih mudah menyesuaikan parameter terhadap data training. Oleh karena itu, penyimpanan bobot berdasarkan validation loss terendah dan penggunaan early stopping menjadi penting agar model pada epoch terakhir yang telah mengalami overfitting tidak digunakan sebagai model akhir.

Kinerja setiap kelas menunjukkan bahwa kemampuan model tidak merata. Recall 100% pada kelas early membuktikan bahwa seluruh citra bercak kering menyebar dapat ditemukan, tetapi precision 67% menunjukkan adanya citra late yang ikut diprediksi sebagai early. Sebaliknya, kelas healthy menghasilkan precision 100% dan F1-score 86%, sehingga menjadi kelas dengan keseimbangan precision dan recall terbaik. Kelas late memiliki recall 50% dan F1-score 57%, yang menunjukkan perlunya peningkatan data dan representasi ciri untuk pola penyakit pada tepi daun.

Baek [6] menunjukkan bahwa mekanisme attention pada Vision Transformer dapat membantu model mempelajari hubungan antarbagian citra dalam klasifikasi penyakit tanaman. Sementara itu, EfficientRMT-Net menggabungkan fitur konvolusional dan Transformer untuk memperoleh representasi lokal dan global secara bersamaan [7]. Dibandingkan pendekatan tersebut, penelitian ini menggunakan ViT-B/16 dan focus pada evaluasi pola perhatian internal model. Akurasi 75% yang diperoleh belum dimaksudkan untuk menandingi penelitian dengan dataset lebih besar, tetapi menjadi dasar untuk menguji kesesuaian prediksi dan fokus perhatian model pada tiga kondisi daun kentang. Meng et al. [20] menggunakan kombinasi CNN dan ViT untuk mengenali penyakit tanaman dengan memanfaatkan fitur lokal dan keseluruhan citra. Namun, hasil penelitian tersebut tidak dapat dibandingkan secara langsung dengan penelitian ini karena terdapat perbedaan dataset dan metode pengujian.

Prashanthi et al. [8] melaporkan bahwa model berbasis CNN dapat menghasilkan akurasi tinggi, sedangkan model berbasis Vision Transformer memberikan keunggulan pada interpretabilitas dan kemampuan menangkap konteks global. Sudheer et al. [11] juga menunjukkan manfaat penggabungan fitur tekstur dengan mekanisme attention dalam mendeteksi penyakit daun. Temuan tersebut mendukung dugaan bahwa perbedaan early dan late tidak cukup diselesaikan hanya melalui keputusan kelas, tetapi perlu ditinjau dari pola tekstur serta wilayah daun yang menerima perhatian model. Sinamenye et al. [21] mengembangkan model hybrid yang menggabungkan EfficientNetV2B3 dan Vision Transformer untuk mengenali penyakit daun kentang. Model tersebut menghasilkan akurasi sebesar 85,06%, lebih tinggi dibandingkan EfficientNetV2B3 sebesar 79,55% dan ViT sebesar 77,92%. Hasil ini menunjukkan bahwa penggabungan kedua arsitektur berpotensi meningkatkan kemampuan model dalam mengenali pola visual daun kentang. Sementara itu, pada penelitian ini, model ViT-B/16 masih mengalami kesulitan membedakan Late Blight dan Early Blight, terlihat dari dua citra Late Blight yang salah diklasifikasikan sebagai Early Blight. Meskipun hasil penelitian Sinamenye et al. lebih tinggi, perbedaan jumlah dan karakteristik dataset perlu diperhatikan sehingga kedua hasil tersebut tidak dapat dibandingkan secara langsung.

Mondal et al. [22] juga mengembangkan model hybrid bernama PLDNet yang menggabungkan DenseNet dan Transformer untuk klasifikasi penyakit daun kentang. Model tersebut memperoleh akurasi 99,54% pada dataset PlantVillage dan 87,50% pada dataset Mendeley. Perbedaan hasil tersebut menunjukkan bahwa karakteristik dataset dapat memengaruhi kemampuan model dalam mengenali penyakit. Pada penelitian ini, ViT-B/16 memperoleh akurasi sebesar 75% dari 12 citra pengujian. Jumlah data yang terbatas membuat kemampuan model pada kondisi citra yang lebih beragam belum dapat dipastikan. Oleh karena itu, diperlukan penambahan dataset, pengujian yang lebih luas, serta analisis attention map untuk memahami kesalahan klasifikasi, terutama pada kelas Late Blight.

Dalam penelitian ini, attention map diposisikan sebagai visualisasi distribusi perhatian internal ViT dan bukan sebagai bukti penyebab biologis penyakit. Pendekatan ini sesuai dengan penelitian explainable AI pada penyakit tanaman yang menggunakan visualisasi untuk memeriksa apakah model berfokus pada lesi atau bagian daun yang relevan [14] . Analisis attention map selanjutnya diperlukan untuk menilai apakah kesalahan prediksi late sebagai early terjadi karena perhatian model berada pada area bercak yang serupa atau justru dipengaruhi oleh latar belakang citra.

Kinerja model dapat ditingkatkan melalui penambahan citra, variasi pencahayaan dan sudut pengambilan, serta representasi tingkat keparahan penyakit, terutama pada kelas late. Pengujian berikutnya juga perlu menggunakan stratified cross-validation dan optimasi learning rate, batch size, weight decay, dropout, serta jumlah blok yang di fine-tuning. Dengan data yang lebih besar dan evaluasi yang lebih stabil, hubungan antara hasil klasifikasi dan pola attention map dapat dianalisis secara lebih kuat. Dalam perspektif aksiologi, Grant et al. [23] menjelaskan bahwa penggunaan sistem kecerdasan buatan tidak cukup hanya dinilai dari transparansi dan kemampuan memberikan penjelasan, tetapi juga perlu mempertimbangkan tanggung jawab terhadap pihak yang terdampak oleh keputusan sistem. Dalam penelitian ini, PotatoScan diharapkan dapat membantu identifikasi awal penyakit daun kentang, dengan tetap memperhatikan keterbatasan hasil prediksi agar pengguna tidak sepenuhnya bergantung pada keputusan model.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 244
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

3.5 Pembahasan dalam Perspektif Filsafat Ilmu

Berdasarkan hasil penelitian, tiga pertanyaan filsafat yang diajukan pada pendahuluan dapat dijelaskan melalui temuan eksperimen. Secara ontologis, ViT-B/16 tidak mengamati penyakit daun kentang secara langsung, melainkan memproses representasi digital berupa citra yang dibagi menjadi patch. Attention Map kemudian menunjukkan distribusi perhatian internal model terhadap patch tersebut. Oleh karena itu, area dengan perhatian tinggi merupakan representasi komputasional dari fokus model dan tidak dapat disamakan dengan penyebab biologis penyakit. Secara epistemologis, pengetahuan yang dihasilkan model diperoleh melalui proses pelatihan dan pengujian terhadap dataset. Akurasi pengujian sebesar 75%, macro precision 78%, macro recall 75%, dan macro F1-score 74% menunjukkan bahwa hasil prediksi memiliki dasar empiris, tetapi belum dapat dianggap sebagai pengetahuan yang sepenuhnya pasti. Hal ini terlihat pada dua citra kelas late yang salah diprediksi sebagai early serta jumlah data uji yang hanya 12 citra. Attention Map melengkapi evaluasi numerik dengan memberikan informasi mengenai bagian citra yang memperoleh perhatian model, sehingga hasil klasifikasi dapat diperiksa tidak hanya dari benar atau salahnya prediksi, tetapi juga dari pola perhatian yang terbentuk. Secara aksiologis, implementasi model pada aplikasi web PotatoScan menunjukkan manfaat praktis hasil penelitian sebagai alat bantu pengenalan kondisi daun kentang. Namun, nilai confidence dan Attention Map perlu dipahami sebagai keluaran komputasional yang mendukung analisis, bukan sebagai pengganti penilaian ahli. Dengan demikian, penggunaan ViT-B/16 dan Attention Map perlu mempertimbangkan keterbatasan dataset, kemungkinan kesalahan klasifikasi, serta interpretasi yang hati-hati agar hasil kecerdasan buatan digunakan secara bertanggung jawab.

4. KESIMPULAN

Penelitian ini Penelitian ini menunjukkan bahwa model Vision Transformer (ViT-B/16) dapat digunakan untuk mengklasifikasikan tiga kondisi daun kentang, yaitu Early Blight, Healthy, dan Late Blight, serta memberikan gambaran mengenai proses klasifikasi melalui Attention Map. Proses fine-tuning dihentikan pada epoch ke-15 menggunakan early stopping, sedangkan model terbaik diperoleh pada epoch ke-8 dengan validation loss sebesar 0,3547 dan akurasi validasi 77,78%. Pemilihan model tersebut dilakukan karena peningkatan akurasi pelatihan hingga 96,08% tidak diikuti perbaikan akurasi validasi, yang menunjukkan adanya kecenderungan overfitting. Pengujian terhadap 12 citra menghasilkan akurasi 75%, dengan macro precision 78%, macro recall 75%, dan macro F1-score 74%. Kelas Early Blight memperoleh recall tertinggi sebesar 100%, sedangkan Healthy menghasilkan precision 100% dan F1-score 86%. Performa terendah ditemukan pada Late Blight dengan recall 50% dan F1-score 57%. Berdasarkan confusion matrix, dua citra Late Blight salah diklasifikasikan sebagai Early Blight, yang mengindikasikan bahwa model masih mengalami kesulitan membedakan pola visual kedua penyakit tersebut. Analisis Attention Map digunakan untuk mengamati distribusi perhatian model terhadap bagian citra, bukan sebagai bukti penyebab penyakit. Temuan ini menegaskan pentingnya menggabungkan evaluasi performa dengan interpretasi visual agar hasil klasifikasi dapat dipahami secara lebih menyeluruh. Namun, keterbatasan jumlah data pengujian membuat kemampuan generalisasi model belum dapat dipastikan. Penelitian selanjutnya disarankan memperluas dataset, menerapkan stratified cross-validation, mengoptimalkan hyperparameter, serta membandingkan Attention Map pada prediksi benar dan salah. Langkah tersebut diharapkan dapat meningkatkan ketepatan klasifikasi dan keandalan interpretasi model dalam mendukung identifikasi awal penyakit daun kentang.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: Puji Utami Rakhmawati;
Metodologi: Puji Utami Rakhmawati;
Perangkat Lunak: Puji Utami Rakhmawati;
Validasi: Puji Utami Rakhmawati, Syaad Patmanthara;
Analisis Formal: Puji Utami Rakhmawati, Syaad Patmanthara;
Investigasi: Puji Utami Rakhmawati, Syaad Patmanthara;
Sumber Daya: Puji Utami Rakhmawati, Syaad Patmanthara;
Kurasi Data: Puji Utami Rakhmawati;
Penulisan Draf Awal: Puji Utami Rakhmawati;
Penulisan, Peninjauan, dan Penyuntingan: Puji Utami Rakhmawati, Syaad Patmanthara;
Visualisasi: Puji Utami Rakhmawati, Syaad Patmanthara;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 245
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (Not applicable).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (Not applicable).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCE

[1] A. S. Mardhatillah and R. Ahida, “Kajian Ontologi, Epistemologi, dan Aksiologi sebagai Fondasi Filsafat Ilmu dalam Pengembangan Ilmu Pengetahuan Kontemporer,” JECCO, vol. 5, no. 3, pp. 670–676, 2025.
[2] A. Salsabila, “Landasan Filsafat Ilmu Modern: Eksplorasi Ontologi, Epistemologi, dan Aksiologi,” MIJ, vol. 4, no. 6, pp. 2412–2420, 2026.
[3] J. Zeynalov and Y. Çakmak, “Efficient and Accurate Potato Disease Classification Using Lightweight Vision Transformers: A Comparative Benchmark Against a Deep CNN Architecture,” Journal of Computer Science and Digital Technologies, vol. 2, no. 1, pp. 5–15, 2026, doi: 10.30546/UNECCSDT.2026.001.210. [4] A. Dosovitskiy et al., “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale,” arXiv preprint, arXiv:2010.11929, 2020, doi: 10.48550/arXiv.2010.11929.
[4] A. Dosovitskiy et al., “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale,” arXiv preprint, arXiv:2010.11929, 2020, doi: 10.48550/arXiv.2010.11929.
[5] I. Fernanda and T. A. Saputri, “Klasifikasi Penyakit Daun pada Tanaman Cabai Menggunakan Model Hybrid CNN-Vision Transformer,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 10, no. 4, 2026, doi: 10.36040/jati.v10i4.18624.
[6] E. T. Baek, “Attention Score-Based Multi-Vision Transformer Technique for Plant Disease Classification,” Sensors, vol. 25, no. 1, Art. no. 270, 2025, doi: 10.3390/s25010270.
[7] K. Shaheed et al., “EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases,” Sensors, vol. 23, no. 23, Art. no. 9516, 2023, doi: 10.3390/s23239516.
[8] B. Prashanthi, A. V. Praveen Krishna, and C. Mallikarjuna Rao, “A Comparative Study of Fine-Tuning Deep Learning Models for Leaf Disease Identification and Classification,” Engineering, Technology and Applied Science Research, vol. 15, no. 1, pp. 19661–19669, 2025, doi: 10.48084/etasr.9017.
[9] S. Ulukaya and S. Deari, “A Robust Vision Transformer-Based Approach for Classification of Labeled Rices in the Wild,” Computers and Electronics in Agriculture, vol. 231, Art. no. 109950, 2025, doi: 10.1016/j.compag.2025.109950.
[10] N. Chihaoui, A. Waleed, A. S. S. Mufleh, S. M. Alshraah, L. H. R. Alajmi, and S. Altuwayjiri, “A Comprehensive Review on AI-Based Crop Disease Detection Using Leaf Image Classification and Explainable AI,” Discover Applied Sciences, 2026, doi: 10.1007/s42452-026-08684-0.
[11] C. L. Sudheer, T. Sreelatha, and P. V. V. Kishore, “Texture Feature Guided Attention Based Fusion Representations for Crop Leaf Disease Detection,” Computers and Electronics in Agriculture, vol. 239, Art. no. 111144, 2025, doi: 10.1016/j.compag.2025.111144.
[12] Y. Li et al., “How Does Attention Work in Vision Transformers? A Visual Analytics Attempt,” IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 6, pp. 2888–2900, 2023, doi: 10.1109/TVCG.2023.3261935.
[13] P. S. Thakur, S. Chaturvedi, P. Khanna, T. Sheorey, and A. Ojha, “Vision Transformer Meets Convolutional Neural Network for Plant Disease Classification,” Ecological Informatics, vol. 77, Art. no. 102245, 2023, doi: 10.1016/j.ecoinf.2023.102245.
[14] G. Loganathan and M. Palanivelan, “An Explainable Lightweight Convolutional Neural Network with Global–Local Context-Enhanced Channel Attention for Grape Leaf Disease Detection,” Computers and Electrical Engineering, vol. 128, Art. no. 110745, 2025, doi: 10.1016/j.compeleceng.2025.110745.
[15] H. Ye et al., “Beyond Efficient Fine-Tuning: Efficient Hybrid Fine-Tuning of CLIP Models Guided by Explainable ViT Attention,” Information Processing & Management, vol. 63, no. 4, Art. no. 104628, 2026, doi: 10.1016/j.ipm.2026.104628.
[16] R. Suresha, N. Manohar, M. Priyanka, and B. R. Pushpa, “Hybrid YOLO-ViT Model with Digital Maps for Efficient Small Traffic Sign Detection in Real Time Road Scenarios,” MethodsX, vol. 15, Art. no. 103694, 2025, doi: 10.1016/j.mex.2025.103694.
[17] A. M. S. Kheir, A. Koubaa, V. Kolluru, S. Mungara, and T. Feike, “Smart Plant Disease Diagnosis Using Multiple Deep Learning and Web Application Integration,” Journal of Agriculture and Food Research, vol. 21, Art. no. 101948, 2025, doi: 10.1016/j.jafr.2025.101948.
[18] M. U. I. Tamim, S. A. Hamim, S. Malik, M. F. Mridha, and S. Mahmood, “InsightNet: A Deep Learning Framework for Enhanced Plant Disease Detection and Explainable Insights,” Plant Direct, vol. 9, no. 5, Art. no. e70076, 2025, doi: 10.1002/pld3.70076.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 246
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Information System Research (JOSH)
ISSN 2686-228X (media online)
Volume 8, No. 1, Oktober 2026, pp 236−247
DOI 10.47065/josh.v8i1.11194
https://ejurnal.seminar-id.com/index.php/josh/

[19] X. Li, L. Chen, S. Zhou, Y. Zhang, and J. Li, "RepVitASF12: A Network of ViT and Attention Scale Sequence Modules to Identify the Big and Small Ends of Eggs," Information Processing in Agriculture, 2026, doi: 10.1016/j.inpa.2026.03.017.
[20] Q. Meng, J. Guo, H. Zhang, Y. Zhou, and X. Zhang, "A Dual-Branch Model Combining Convolution and Vision Transformer for Crop Disease Classification," PLOS ONE, vol. 20, no. 4, Art. no. e0321753, 2025, doi: 10.1371/journal.pone.0321753.
[21] J. H. Sinamenye, A. Chatterjee, and R. Shrestha, "Potato Plant Disease Detection: Leveraging Hybrid Deep Learning Models," BMC Plant Biology, vol. 25, no. 1, 2025, doi: 10.1186/s12870-025-06679-4.
[22] A. Mondal, A. Chatterjee, and N. Avazov, "A Hybrid CNN-Transformer Model with Adaptive Activation Function for Potato Leaf Disease Classification," Scientific Reports, vol. 16, no. 1, 2026, doi: 10.1038/s41598-025-34406-4.
[23] D. G. Grant, J. Behrends, and J. Basl, "What We Owe to Decision-Subjects: Beyond Transparency and Explanation in Automated Decision-Making," Philosophical Studies, vol. 182, no. 1, pp. 55–85, 2025, doi: 10.1007/s11098-023-02013-6.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 247
This work is licensed under a Creative Commons Attribution 4.0 International License.