JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 1, Januari 2026, pp. 574 - 583
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/169

Implementation of the Vision Transformer (ViT) Method in Soil Type Classification

Abdilah Hasbi*[1], Ardollynata[2], Benelekser Tumanggor[3]
[1] Fakultas lmu Komputer, Universitas Pamulang
Jl. Jalan Raya Puspitek No. 46, 15316, Tangerang Selatan, Banten, Indonesia
[2,3] Fakultas Ilmu Komputer, Universitas Ichsan Satya
Jl. Jombang Raya No.41, 15414, Tangerang Selatan, Banten, Indonesia
e-mail*[1]: dosen03288@unpam.ac.id

ARTICLE INFO

Article History:

Received: 15-11-2025
Revised: 11-12-2025
Accepted: 02-01-2026
Published: 21-01-2026

ABSTRACT

This study applies the Vision Transformer (ViT) method to soil-type classification and evaluates its accuracy using digital images. The Vision Transformer (ViT) is a Deep Learning architecture that uses self-attention to extract global features from images, enabling it to recognize texture and color patterns more comprehensively than other convolutional methods. The dataset used consists of eight soil types, each containing 77 image data in ".jpg" format. Each image was processed and augmented to increase the dataset to 700 images per soil type. This was done to prevent overfitting. The ViT model was trained using a 90%/10 % split of the training and validation datasets. The results show that the Vision Transformer (ViT) architecture achieves an accuracy of 71.4%, with a small difference between training and validation accuracy, indicating that the ViT method generalizes well. This method successfully distinguishes soil types such as alluvial, andosol, laterite, and limestone based on subtle visual differences. These results demonstrate that the Vision Transformer (ViT) is effective for soil type classification and has the potential to serve as a basis for developing a digital image-based soil type classification system to support research and precision agriculture.

Keywords: *Vision Transformer, Soil Classification, Deep Learning.*

Abstrak

Penelitian ini mengimplementasikan metode *Vision Transformer (ViT)* dalam klasifikasi jenis tanah untuk melihat akurasi yang di hasilkan berdasarkan citra digital. Metode *Vision Transformer (ViT)* merupakan asritektur *Deep learning* yang memanfaatkan mekanisme *self-attention* untuk mengetraksi fitur global dari data citra, sehingga mampu mengenali pola tektur dan warna secara lebih menyelurh dibandingkan konvlousional lainnya. Dataset yang digunakan terdiri dari 8 jenis tanah, masing-masing berisi dari 77 data citra dengan format ".jpg". Setiap citra dikembangkan melalui tahap processing dan augmentasi untuk meningkatkan jumlah dataset menjadi 700 data setiap jenis tanah, hal ini dilakukan untuk mencegah overvitting. Model ViT dilatih dengan komposisi 90% dataset training, dan 10% dataset validasi. Hasil penelitiain menunjukan bahwa arsitektur *Vision Transformer (ViT)* mampu mencapai tingkat akurasi 71,4%, dengan selisih kecil antara akurasi training dan validasi menandakan metode ViT cukup mampu

DOI: https://doi.org/10.51903/zqs2q531

p-ISSN : 2803-1507 e-ISSN : 2803-1531

menggeneralisikan data dengan baik. Metode ini berhasil membedakan jenias tanah seperti alluvial, andosol, laterit, dan kapur berdasarkan perbedaan visual yang halus, hasil ini membuktikan bahwa *Vision Transformer* (ViT) efektif dalam klasifikasi jenis tanah dan berpotensi menjadi dasar pengembangan system system klasifikasi jenis tanah berbasi citra digital untuk mendukung riset dan pertanian perisis.

Kata Kunci: *Vision Transformer*, *Klasifikasi Tanah*, *Deep Learning*.

1. PENDAHULUAN

Indonesia dikenal sebagai negara agraris yang memiliki potensi besar dalam sektor pertanian berkat kondisi geografis dan kesuburan tanah yang tinggi. Menurut data Kementerian Pertanian, sebagian besar penduduk Indonesia bergantung pada hasil pertanian sebagai sumber utama ekonomi, sehingga produktivitas lahan menjadi faktor yang sangat penting dalam menunjang ketahanan pangan nasional. Jenis tanah merupakan salah satu faktor kunci dalam menentukan kesesuaian lahan terhadap jenis tanaman tertentu, karena setiap jenis tanah memiliki perbedaan karakteristik fisik, kimia, dan biologis seperti tekstur, struktur, warna, serta kandungan mineral dan unsur hara [1]. Dengan demikian, identifikasi jenis tanah secara akurat menjadi hal yang esensial dalam upaya optimalisasi hasil pertanian dan pengelolaan sumber daya alam secara berkelanjutan [2].

Secara tradisional, proses klasifikasi tanah dilakukan melalui observasi lapangan dan uji laboratorium menggunakan parameter seperti pH, kadar air, kandungan organik, dan struktur partikel. Meskipun metode ini dianggap akurat, prosesnya memerlukan biaya tinggi, waktu lama, serta keterlibatan tenaga ahli di bidang tanah [3]. Di sisi lain, metode manual ini kurang efisien ketika diterapkan pada area pertanian yang luas dan memiliki variasi geografis tinggi. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi yang mampu melakukan identifikasi tanah secara cepat, efisien, dan tetap akurat [4]. Salah satu alternatif yang berkembang adalah penggunaan pengolahan citra digital untuk menganalisis karakteristik visual tanah, seperti warna dan tekstur, guna menentukan klasifikasinya [5].

Dalam beberapa tahun terakhir, kemajuan *deep learning* telah membawa perubahan besar dalam bidang pengolahan citra digital. Salah satu arsitektur yang paling banyak digunakan adalah *Convolutional Neural Network* (CNN), yang terbukti efektif dalam berbagai aplikasi seperti pengenalan objek, deteksi penyakit tanaman, serta klasifikasi hasil pertanian [6], [7]. Penerapan machine learning dalam sektor pertanian berperan penting dalam mendukung sistem pertanian presisi melalui analisis data berbasis citra dan sensor [8]. Namun, CNN memiliki keterbatasan dalam menangkap hubungan spasial global karena sifatnya yang hanya memproses informasi secara lokal [9], [10]. Akibatnya, pada citra dengan pola kompleks seperti tanah yang memiliki tekstur dan warna beragam, CNN sering kali kehilangan konteks global yang penting untuk pengenalan yang akurat [11],[12].

Meskipun metode *Convolutional Neural Network* (CNN) telah banyak digunakan dalam klasifikasi citra tanah, sebagian besar penelitian sebelumnya masih berfokus pada ekstraksi fitur lokal dan menunjukkan keterbatasan dalam menangkap hubungan spasial global pada citra tanah yang memiliki kompleksitas tekstur dan warna tinggi [13], [14], [15]. Di sisi lain, penerapan *Vision Transformer (ViT)* yang mampu memodelkan konteks global melalui mekanisme *self-attention* masih sangat terbatas dalam konteks klasifikasi jenis tanah berbasis citra digital. Selain itu, belum banyak penelitian yang mengevaluasi kinerja ViT secara khusus pada dataset tanah multikelas dengan karakteristik visual yang saling tumpang tindih. Oleh karena itu, penelitian ini secara spesifik bertujuan mengisi celah penelitian tersebut dengan mengkaji implementasi ViT dalam klasifikasi jenis tanah serta menganalisis performanya pada setiap kelas tanah secara lebih mendalam.

Sebagai solusi atas keterbatasan tersebut, *Vision Transformer* (ViT) diperkenalkan sebagai arsitektur baru dalam *deep learning* yang diadaptasi dari konsep Transformer pada bidang Natural Language Processing (NLP) [16]. ViT memecah citra menjadi potongan-potongan kecil (*patches*), kemudian menggunakan mekanisme *self-attention* untuk mempelajari hubungan antar-patch secara global [17], [18], Dengan pendekatan ini, ViT mampu menangkap informasi visual secara menyeluruh dan memiliki performa lebih baik dalam mengenali pola yang kompleks [19], [20]. Penelitian sebelumnya membuktikan bahwa *Vision Transformer* mampu mencapai akurasi tinggi pada berbagai tugas klasifikasi, seperti pengenalan sel darah putih [21], dan deteksi covid dari citra X-ray [22].

Meskipun berbagai penelitian telah mengkaji penerapan *deep learning*, khususnya CNN, dalam klasifikasi citra pertanian dan analisis karakteristik tanah, sebagian besar studi tersebut masih berfokus pada pendekatan berbasis ekstraksi fitur lokal dan belum mampu menangkap hubungan spasial global secara optimal.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Beberapa penelitian terkini mulai mengeksplorasi penggunaan ViT dalam domain pengolahan citra, namun penerapannya masih didominasi oleh bidang medis dan pengenalan objek umum, sementara studi yang secara khusus membahas klasifikasi jenis tanah berbasis citra digital menggunakan arsitektur *Vision Transformer* masih sangat terbatas, terutama dalam konteks pertanian di Indonesia.

Selain itu, penelitian sebelumnya umumnya belum memberikan analisis mendalam mengenai kemampuan *Vision Transformer* dalam mengenali kompleksitas visual tanah yang memiliki variasi warna, tekstur, dan struktur yang tinggi. Perbandingan kinerja *Vision Transformer* dengan pendekatan *deep learning* konvensional pada klasifikasi jenis tanah juga masih jarang dilakukan secara sistematis. Oleh karena itu, terdapat celah penelitian (research gap) berupa kurangnya kajian empiris yang mengimplementasikan dan mengevaluasi efektivitas *Vision Transformer* dalam klasifikasi jenis tanah berbasis citra digital sebagai alternatif metode konvensional dan CNN.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada implementasi metode ViT dalam klasifikasi jenis tanah berbasis citra digital [23], [24]. Beberapa penelitian terbaru menunjukkan bahwa arsitektur *Vision Transformer* telah mulai diterapkan pada bidang pertanian untuk analisis citra dan pengambilan keputusan berbasis data, namun penerapannya secara spesifik pada klasifikasi jenis tanah masih sangat terbatas [25]. Pendekatan ini diharapkan dapat menghasilkan model klasifikasi yang lebih akurat dan efisien dibandingkan metode konvensional. Selain itu, hasil penelitian ini diharapkan dapat mendukung penerapan konsep *precision agriculture* di Indonesia, yaitu pertanian berbasis data dan teknologi, yang memungkinkan petani melakukan pengambilan keputusan yang tepat terkait jenis tanaman dan teknik budidaya sesuai dengan karakteristik tanah yang dimiliki.

2. TINJAUAN PUSTAKA

2.1. *Klasifikasi Jenis Tanah*

Penelitian terkait klasifikasi jenis tanah telah dilakukan dengan berbagai pendekatan, baik menggunakan sensor fisik maupun pengolahan citra digital. [1] mengembangkan sistem identifikasi warna tanah berbasis standar Munsell menggunakan sensor warna dan kelembaban, yang mampu memberikan hasil pengukuran objektif, namun pendekatan ini masih bergantung pada perangkat keras tambahan dan belum memanfaatkan karakteristik visual citra secara menyeluruh lahan. Sementara itu, [3], [5] menerapkan metode pengolahan citra dan *Convolutional Neural Network* (CNN) untuk klasifikasi jenis tanah berbasis mobile, yang menunjukkan peningkatan efisiensi dibanding metode konvensional, namun masih memiliki keterbatasan dalam menangkap variasi pola visual tanah yang kompleks.

2.2. *Pengolahan Citra Digital dalam Klasifikasi Jenis Tanah*

Pengolahan citra digital merupakan teknik yang digunakan untuk mengekstraksi informasi dari gambar digital guna mendapatkan karakteristik tertentu, seperti warna, tekstur, dan pola visual. Dalam konteks klasifikasi jenis tanah, citra digital dapat digunakan untuk merepresentasikan kondisi permukaan tanah secara visual, sehingga memungkinkan analisis otomatis berbasis komputer [5]. Beberapa penelitian telah menunjukkan bahwa karakteristik visual tanah, seperti perbedaan warna dan tekstur, dapat digunakan sebagai fitur penting dalam proses klasifikasi. Dengan dukungan teknologi komputer, pengolahan citra digital mampu mempercepat proses identifikasi tanah dan mengurangi ketergantungan terhadap analisis manual yang bersifat subjektif.

2.3. *Deep Learning dan Convolution Network*

Beberapa penelitian lain menunjukkan bahwa CNN, termasuk arsitektur ringan seperti MobileNetV3 dan pengembangannya dengan Squeeze-and-Excitation, mampu memberikan performa yang baik pada berbagai tugas klasifikasi citra [6], [7]. Namun demikian, CNN secara fundamental bekerja dengan ekstraksi fitur lokal sehingga kurang optimal dalam memahami hubungan spasial global pada citra dengan tekstur dan warna yang heterogen, seperti citra tanah [9], [12]. Keterbatasan ini menjadi tantangan utama ketika diterapkan pada objek dengan pola tidak teratur dan memiliki variasi visual tinggi.

2.4. *Vision Transformer (ViT)*

Sebagai alternatif, *Vision Transformer* (ViT) dan turunannya seperti Swin Transformer dan Pyramid *Vision Transformer* telah diperkenalkan untuk mengatasi keterbatasan CNN melalui mekanisme *self-attention* yang mampu mempelajari hubungan global antar-patch citra [16] [20]. Berbagai studi membuktikan bahwa *Vision*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Transformer memberikan performa unggul pada tugas klasifikasi citra kompleks, seperti klasifikasi sel darah putih [18] dan deteksi COVID-19 dari citra X-ray [22]. Meskipun demikian, sebagian besar penelitian *Vision Transformer* masih berfokus pada domain medis dan pengenalan objek umum, sementara penerapannya pada bidang pertanian, khususnya klasifikasi jenis tanah berbasis citra digital, masih sangat terbatas.

2.5. *Penelitian Terkait Klasifikasi Tanah Berbasi Citra*

Penelitian terkait klasifikasi jenis tanah berbasis citra telah banyak dilakukan dengan menggunakan metode machine learning dan CNN. [5] mengimplementasikan CNN untuk klasifikasi jenis tanah berbasis Android dan menunjukkan bahwa pendekatan berbasis citra mampu memberikan hasil yang cukup baik. Namun, penelitian yang secara khusus menerapkan *Vision Transformer* dalam klasifikasi jenis tanah masih relatif terbatas. Beberapa studi terbaru mulai mengkaji potensi Transformer dalam bidang pertanian dan pengolahan citra lingkungan [25]. Hal ini membuka peluang untuk mengeksplorasi penggunaan ViT sebagai alternatif yang lebih unggul dalam mengatasi keterbatasan CNN, khususnya dalam menangkap fitur global pada citra tanah.

Berdasarkan tinjauan tersebut, dapat disimpulkan bahwa penelitian sebelumnya belum secara spesifik mengkaji efektivitas *Vision Transformer* dalam mengenali kompleksitas visual tanah yang memiliki variasi warna, tekstur, dan struktur yang tinggi, serta belum membandingkannya secara implisit sebagai solusi atas keterbatasan CNN dalam konteks klasifikasi jenis tanah. Oleh karena itu, penelitian ini memposisikan diri sebagai upaya untuk mengisi celah penelitian tersebut dengan mengimplementasikan metode *Vision Transformer* (ViT) pada klasifikasi jenis tanah berbasis citra digital, sehingga diharapkan mampu memberikan kontribusi ilmiah dalam pengembangan metode klasifikasi tanah yang lebih akurat dan mendukung penerapan pertanian presisi berbasis teknologi.

3. METODOLOGI PENELITIAN

Model ADDIE pada penelitian ini digunakan sebagai kerangka konseptual untuk mengorganisasi tahapan pengembangan sistem klasifikasi, bukan sebagai metode eksperimen utama. Sementara itu, alur eksperimen *deep learning* tetap mengikuti praktik standar yang meliputi pengumpulan data, preprocessing, pelatihan model, validasi, dan evaluasi kinerja. Dengan pendekatan ini, ADDIE berfungsi untuk memastikan sistem dikembangkan secara sistematis, sedangkan evaluasi kinerja model dilakukan berdasarkan metrik machine learning yang terukur [11].

3.1. *Tahap Penelitian*

Pada bagian ini juga dapat dijelaskan tahapan-tahapan penelitian disertai dengan hasil yang didapatkan pada setiap tahapan penelitian tersebut.

3.1.1. *Tahap Analysis (Analisis)*

Untuk menemukan masalah utama, kebutuhan sistem, dan tujuan penelitian ini, tahap analisis dilakukan. Problem yang ditimbulkan adalah kesulitan untuk mengidentifikasi jenis tanah secara cepat dan akurat di lapangan, yang biasanya dilakukan secara manual melalui pengamatan fisik dan analisis laboratorium. Jenis tanah, jumlah dataset, format gambar, dan parameter model *Vision Transformer* adalah beberapa variabel yang digunakan. Variabel-variabel ini dinilai melalui analisis kebutuhan. Dataset yang digunakan terdiri dari delapan kelas tanah: andosol, entisol, humus, inceptisol, laterit, kapur, dan pasir, dengan 77 gambar berformat.jpg untuk masing-masing.

3.1.2. *Tahap Design (Perancangan)*

Pada tahap ini, perhatian terutama diberikan pada perancangan alur kerja sistem klasifikasi tanah yang menggunakan *Vision Transformer*. Setiap gambar tanah melalui tahapan preprocessing seperti pengurangan, normalisasi, dan peningkatan data (rotasi, flipping, dan pencahayaan) untuk meningkatkan variasi data dan mengurangi risiko overfitting. Perancangan termasuk pembuatan diagram alir proses (flowchart). Model ViT dibuat untuk menerima gambar berukuran 224×224 piksel dan kemudian mengubahnya menjadi patch 16×16 piksel yang diproses oleh mekanisme *self-attention* untuk menghasilkan fitur visual luas. Seluruh eksperimen dilakukan menggunakan parameter yang konsisten untuk memastikan reprodusibilitas penelitian.

3.1.3. *Tahap Development (Pengembangan)*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Pada titik ini, dibangun model klasifikasi menggunakan arsitektur *Vision Transformer*. Model *Vision Transformer* dilatih menggunakan optimizer Adam dengan nilai learning rate sebesar 0,001 untuk mempercepat konvergensi serta menjaga stabilitas proses pembelajaran, ukuran batch 128, dan 100 epoch digunakan untuk mengembangkan model [26]. Dalam proses pelatihan, 90% data digunakan untuk pelatihan dan 10% untuk validasi. Tahap ini juga mencakup proses penyesuaian khusus pada model ViT yang telah dilatih sebelumnya untuk menyesuaikannya dengan karakteristik dataset tanah lokal.

3.1.4. *Tahap Implementation (Implementasi)*

Implementasi dilakukan dengan menggunakan model *Vision Transformer* yang telah dilatih untuk mengklasifikasikan citra tanah baru. Model diuji menggunakan matrix confusion untuk mengukur kinerjanya dalam mengenali setiap kelas tanah berdasarkan parameter kecepatan, ketepatan, recall, dan skor F1. Tahap implementasi bertujuan untuk menunjukkan bahwa ViT dapat mengenali pola tekstur dan warna tanah secara konsisten di seluruh kelas, dengan kinerja terbaiknya pada tanah humus dan andosol.

3.1.5. *Tahap Evaluation (Evaluasi)*

Tujuan dari tahap evaluasi adalah untuk mengevaluasi keberhasilan sistem secara keseluruhan dan menemukan area yang dapat diperbaiki. Proses evaluasi dilakukan dengan menguji kinerja model pada data uji baru dan membandingkan hasil klasifikasi dengan data nyata. Selain itu, setiap tahap dalam model ADDIE dievaluasi untuk memastikan bahwa setiap langkah penelitian telah berjalan sesuai dengan tujuan. Peningkatan jumlah dataset dan penggunaan model hybrid ViT-CNN untuk meningkatkan efisiensi komputasi dapat menjadi fokus penelitian berikutnya.

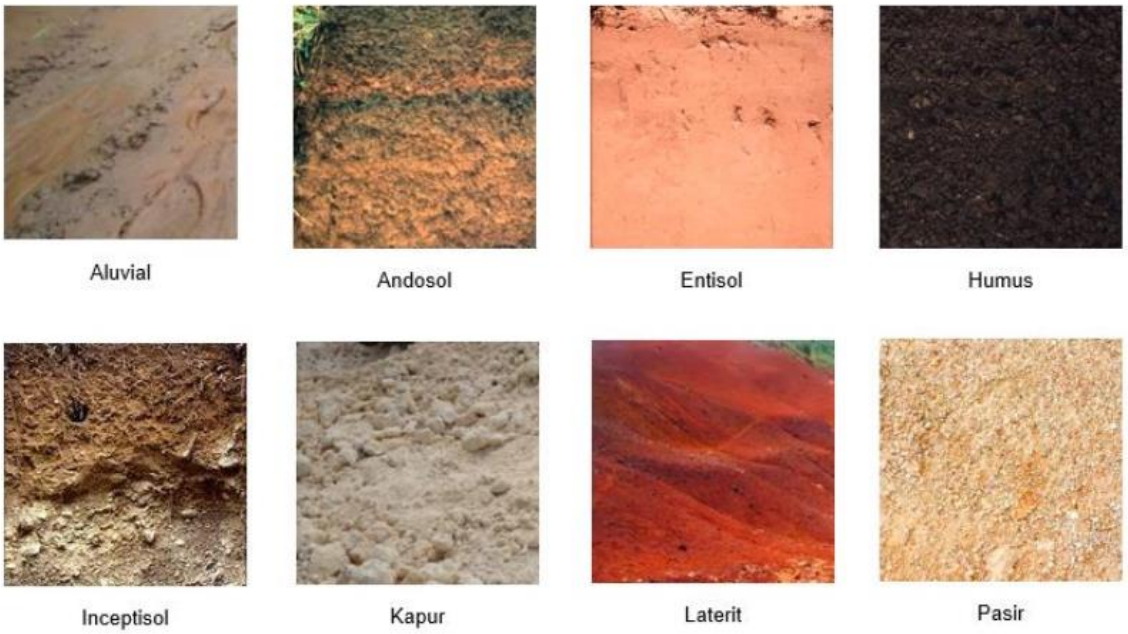
4. HASIL DAN PEMBAHASAN

Data penelitian ini di peroleh dari Kaggle, sebuah platform online yang menyediakan beberpa dataset demi menunjang kebutuhan riset dan machine learning, data yang digunakan berupa .jpg dengan 8 jenis tanah. Model *Vision Transformer* (ViT) digunakan untuk melakukan klasifikasi citra jenis tanah berdasarkan dataset berisi delapan kelas, yaitu Aluvial, Andosol, Entisol, Humus, Inceptisol, Kapur, Laterit, dan Pasir yang dapat di lihat pada Gambar 1. Dataset dibagi menjadi 90% data pelatihan dan 10% data validasi. Sebelum proses pelatihan, dilakukan augmentasi citra untuk memperbanyak variasi data dan mengurangi risiko overfitting yang terlihat pada Tabel 1. Proses augmentasi meliputi rotasi, translasi, zoom, dan flipping horizontal, sehingga jumlah citra bertambah signifikan dan model dapat belajar dari variasi tekstur dan warna tanah yang beragam [27].

Tabel 1. Data Augmentasi

No	Nama	Data Citra	Data citra Augmentasi	Data Training	Data Validasi
1	Aluvial	77	729	656	73
2	Andosol	77	699	629	70
3	Entisol	77	769	692	77
4	Humus	77	740	666	74
5	Inceptisol	77	770	693	77
6	Kapur	77	740	666	74
7	Laterit	77	769	692	77
8	Pasir	77	775	697	78

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 1. Contoh Citra Jenis Tanah

Selama proses pelatihan, model ViT menggunakan ukuran citra 224×224 piksel, batch size 128, learning rate 0,001, dan weight decay 0,0001 yang terlihat pada Tabel 2. Model dilatih hingga 100 epoch dengan optimizer Adam dan loss function categorical crossentropy. Parameter-parameter tersebut menghasilkan konvergensi yang stabil, di mana akurasi validasi meningkat secara bertahap hingga mencapai titik optimal pada epoch ke-100.

Tabel 2. Hyperparameter Pelatihan

Variabel	Nilai
image_size	244
batch_size	128
patch_size	16
num_patch	(image_size / patch_size) * 2
epoch	100
learning_rate	0.001
weigh_decay	0.0001

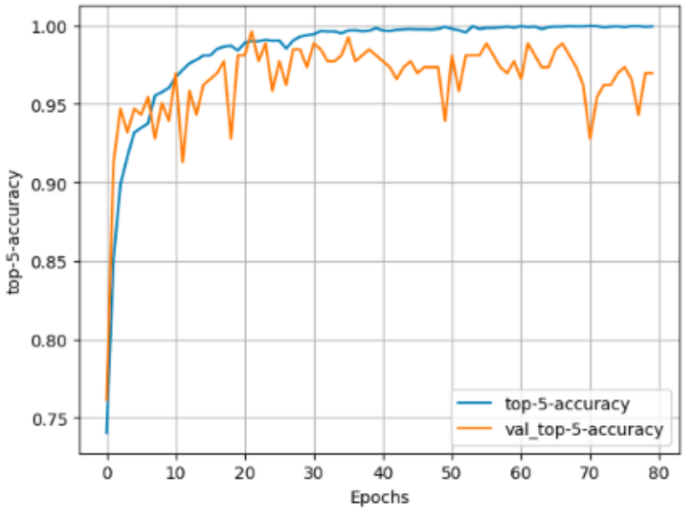
Evaluasi dilakukan untuk menguji keakuratan model yang dibagun. Pada proses pelatihan model, evaluasi kinerja dilakukan menggunakan metrik precision, recall, dan F1-score untuk memastikan keseimbangan performa pada setiap kelas, terutama pada kondisi distribusi data yang tidak merata untuk setiap kelas tanah [28]. Hasil evaluasi dapat dilihat pada Tabel 3 yang menunjukkan bahwa nilai precision rata-rata mencapai 73%, recall sebesar 71%, dan F1-score sebesar 72%. Nilai tersebut menandakan bahwa model ViT cukup efektif dalam membedakan antarjenis tanah, meskipun terdapat beberapa kelas seperti Aluvial dan Inceptisol yang masih memiliki nilai di bawah rata-rata.

Tabel 3. Hasil Training Vision Transformer (ViT)

Kelas	Precision	Recall	F1-Score
Aluvial	0.47	0.52	0.49
Andosol	0.88	0.67	0.76
Entisol	0.86	0.76	0.81
Humus	0.89	0.94	0.91
Inceptisol	0.42	0.45	0.43
Kapur	0.93	0.75	0.83
Laterit	0.89	0.97	0.93
Pasir	0.53	0.64	0.58
Rata-rata	0.73	0.71	0.72

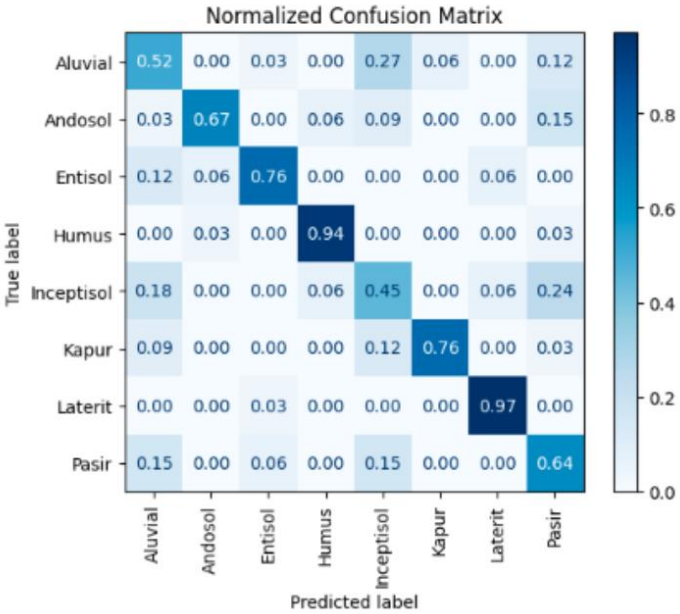
p-ISSN : 2803-1507 e-ISSN : 2803-1531

Selain itu, hasil grafik akurasi pada Gambar 2 memperlihatkan kemampuan model klasifikasi menunjukan keakuratan antara tes dan validasi untuk setiap interval, dimana validasi keakuratan diambil pada proses. Gambar 2 menunjukan model ini dilatih sangat baik dan mencapai akurasi yang sangat tinggi, meskipun terdapat sedikit indikasi overfitting ringan, Meskipun model *Vision Transformer* menunjukkan kemampuan klasifikasi yang cukup baik secara keseluruhan, hasil pengujian juga mengindikasikan bahwa performa model masih bersifat moderat, khususnya pada kelas aluvial dan inceptisol yang memiliki kemiripan visual tinggi. Hal ini menunjukkan bahwa model masih memiliki keterbatasan dalam membedakan kelas tanah tertentu dan memerlukan pengembangan lebih lanjut.



Gambar 2. Grafik Akurasi

Berdasarkan informasi yang tlah diperoleh dengan klasifikasi jenis tanah yang dilakukan oleh model Vision Transfomer (ViT) dengan hasil klasifikasi, maka peneliti menggunakan confusion matrix untuk mengukur kinerja model seperti pada Gambar 3.

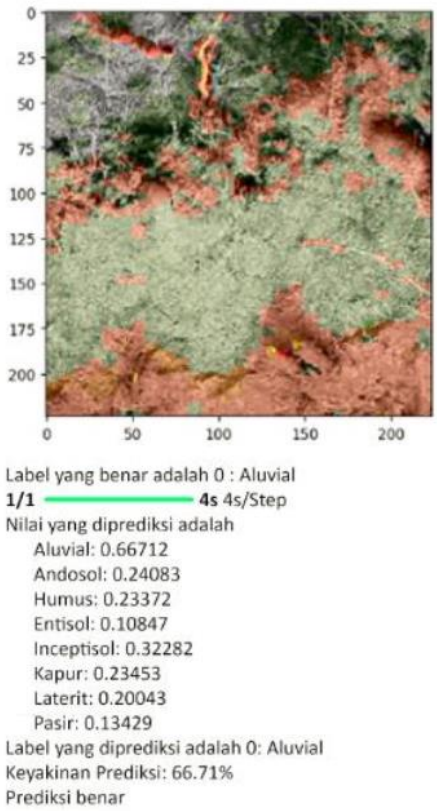


Gambar 3. Confuison Matrix

Berdasarkan hasil akurasi yang yang ditampilkan dalam bentuk confusion matrix pada Gambar 3, dapat dikatehui bahwa akuarasi kelas aluvial 52%, kelas andosol 67%, kelas entisol 76%, kelas humus 94%, kelas

p-ISSN : 2803-1507 e-ISSN : 2803-1531

inceptisol 45%, kelas kapur 76%, kelas laterit 97%, dan pasir 64%, hal ini menunjukan bahwa model sudah bisa membedakan setipap kelas secara baik.



Gambar 4. Hasil Prediksi Model ViT

Gambar 4 menampilkan hasil jenis tanah aluvial dengan keyakinan prediksi 66.71% yang di olah model *Vision Transformer* (ViT), secara sefesifik gambar ini menujukan area dengan berbagai tesktur dan warna yang mepersenatsikan karakterisitik tanah dan vegatsi.

5. DISKUSI

Hasil klasifikasi menggunakan Vision Transformer (ViT) dengan akurasi rata-rata sebesar 71,4% menunjukkan bahwa arsitektur berbasis *self-attention* mampu mempelajari karakteristik visual tanah yang kompleks, khususnya variasi warna dan tekstur permukaan. Temuan ini berkaitan dengan tujuan penelitian dalam mengevaluasi kemampuan ViT untuk menangkap konteks global citra tanah yang sebelumnya sulit dimodelkan oleh pendekatan berbasis *local feature extraction*. Mekanisme pembagian citra menjadi *patch* dan pemodelan hubungan antar-*patch* sebagaimana dijelaskan oleh [17] memungkinkan model memahami pola visual tanah secara menyeluruh. Dengan pendekatan tersebut, ViT dapat menghubungkan area-area citra yang berjauhan namun memiliki korelasi visual yang relevan dalam proses klasifikasi.

Jika dibandingkan dengan penelitian terdahulu yang menggunakan Convolutional Neural Network (CNN), hasil penelitian ini menunjukkan perbedaan pendekatan dalam memahami informasi visual. Penelitian [5], [3] melaporkan bahwa CNN cukup efektif dalam klasifikasi jenis tanah, namun performanya sangat bergantung pada ekstraksi fitur lokal. Keterbatasan CNN dalam menangkap hubungan spasial global sebagaimana dibahas oleh [9], [19] menjadi salah satu faktor yang membedakan pendekatan ViT. Dalam konteks ini, hasil penelitian memperkuat temuan bahwa arsitektur Transformer menawarkan sudut pandang pemodelan yang berbeda terhadap citra dengan pola visual tidak teratur seperti tanah.

Variasi performa antar kelas tanah, khususnya pada kelas aluvial dan inceptisol yang menunjukkan nilai *precision* dan *recall* lebih rendah, mengindikasikan adanya tantangan pada citra dengan kemiripan visual tinggi. Fenomena ini sejalan dengan penjelasan [9] bahwa objek dengan karakteristik visual heterogen dan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

saling tumpang tindih cenderung sulit dibedakan oleh model berbasis citra. Meskipun teknik augmentasi data telah diterapkan untuk meningkatkan variasi citra, hasil pengujian menunjukkan bahwa augmentasi visual belum sepenuhnya mengatasi kompleksitas perbedaan antar kelas tanah tertentu. Kondisi ini menunjukkan adanya kebutuhan eksplorasi pendekatan tambahan, seperti penggabungan fitur non-visual atau arsitektur hibrida, sebagaimana mulai dibahas dalam studi Vision Transformer terkini [18], [20].

Dari sisi implikasi, penerapan Vision Transformer dalam klasifikasi jenis tanah berbasis citra digital memberikan kontribusi teoretis terhadap pengembangan metode *deep learning* di bidang pertanian presisi. Pendekatan ini memperluas domain penerapan ViT yang sebelumnya lebih banyak digunakan pada bidang medis dan pengenalan objek umum [21], [22]. Secara praktis, sistem klasifikasi berbasis citra berpotensi mendukung proses identifikasi tanah secara lebih efisien tanpa ketergantungan penuh pada analisis laboratorium, sebagaimana dibutuhkan dalam pengelolaan lahan pertanian. Namun, keterbatasan penelitian ini terletak pada penggunaan satu arsitektur model dan satu sumber dataset, sehingga variasi kondisi lingkungan dan karakteristik tanah di luar dataset penelitian belum sepenuhnya terwakili.

6. KESIMPULAN DAN SARAN

6.1. *Kesimpulan*

Penelitian ini membahas implementasi metode *Vision Transformer* (ViT) dalam klasifikasi jenis tanah berbasis citra digital sebagai upaya pemanfaatan kecerdasan buatan untuk mendukung sektor pertanian dan geologi digital. Hasil pengujian menunjukkan bahwa arsitektur ViT mampu mengenali karakteristik visual tanah secara efektif melalui mekanisme *self-attention* yang mengekstraksi hubungan global antarbagian citra. Dengan menggunakan dataset delapan kelas tanah yang telah diperkaya melalui teknik augmentasi data, model berhasil mencapai akurasi rata-rata sebesar 71,4%, dengan nilai precision 73%, recall 71%, dan F1-score 72%. Penerapan augmentasi data terbukti meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting. Dengan demikian, hasil penelitian ini menunjukkan bahwa *Vision Transformer* memiliki potensi sebagai pendekatan alternatif dalam klasifikasi jenis tanah berbasis citra digital. Namun, temuan ini masih terbatas pada satu arsitektur model dan satu dataset, sehingga generalisasi hasil perlu dilakukan secara hati-hati. Penelitian lanjutan dengan pembandingan model dan dataset yang lebih beragam sangat diperlukan untuk memperkuat temuan ini.

6.2. *Saran*

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya. Peningkatan jumlah dan keragaman dataset citra tanah sangat disarankan agar model *Vision Transformer* (ViT) dapat mempelajari karakteristik visual tanah secara lebih representatif, khususnya pada kelas tanah yang memiliki tingkat kemiripan tinggi seperti aluvial dan inceptisol. Selain itu, eksplorasi arsitektur hybrid yang mengombinasikan *Vision Transformer* dengan Convolutional Neural Network (CNN) perlu dilakukan guna menggabungkan keunggulan ekstraksi fitur lokal dan pemodelan konteks global. Optimalisasi hyperparameter, seperti ukuran patch, jumlah layer transformer, dan learning rate, juga berpotensi meningkatkan stabilitas pelatihan serta performa model. Penelitian lanjutan disarankan untuk mengintegrasikan data non-visual, seperti parameter kimia tanah dan data lingkungan, guna menghasilkan sistem klasifikasi yang lebih komprehensif. Selain itu, pengembangan sistem ke dalam bentuk aplikasi berbasis web atau mobile diharapkan dapat meningkatkan pemanfaatan hasil penelitian secara praktis dalam mendukung pertanian presisi dan pengelolaan lahan berkelanjutan.

DAFTAR PUSTAKA

[1] M. Lamsani, R. A. Pangestika, M. Cahyanti, and E. R. Swedia, “Sistem Identifikasi Warna Tanah Munsell Menggunakan Sensor Warna TCS3200 Dan Kelembaban YL-69,” *Sebatik*, vol. 27, no. 1, pp. 379–389, Jun. 2023, doi: 10.46984/sebatik.v27i1.2249.

[2] L. Lauer *et al.*, “Investigating the Usability of a Head-Mounted Display Augmented Reality Device in Elementary School Children,” *Sensors*, vol. 21, no. 19, Oct. 2021, doi: 10.3390/s21196623.

[3] M. Firdaus and R. Afwani, “Pengembangan Restful API untuk Aplikasi Klasifikasi Jenis Tanah Berbasis Mobile pada Google Cloud (Restful API Development for Mobile-Based Soil Type Classification Applications on Google Cloud),” *Jurnal Teknologi Informasi, Komputer dan Aplikasinya*, vol. 6, no. 1, pp. 275–287, 2024. [Online]. Available: http://jtika.if.unram.ac.id/index.php/jtika/275

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[4] V. Birukov and V. Mostovoy, "Optimisation of Acrylic Fiber Temperature Regime Thermostabilisation with Computer Modelling," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 578–581. doi: 10.1016/j.procs.2020.02.207.

[5] Y. P. Astuti, I. Wardatunizza, E. R. Subhiyakto, and E. Kartikadarma, "Implementasi Algoritma Convolutional Neural Network (CNN) untuk Klasifikasi Jenis Tanah Berbasis Android," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, vol. 8, no. 3, pp. 220–225, sep. 2023, doi: 10.30591/jpit.v8i3.5026.

[6] R. Shalehuddin Albawani, F. Tri Anggraeny, and M. Muharrom Al Haromainy, "Implementasi SEBlock pada Klasifikasi Citra Penyakit Mata Manusia dengan Arsitektur MobileNetV3-Small," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 1123–1128, 2024, doi: 10.36040/jati.v8i1.8916.

[7] G. Ashari Rakhmat and M. Fikri Haekal, "MIND (Multimedia Artificial Intelligent Networking Database Peningkatan Performa MobilenetV3 dengan Squeeze-and-Excitation (Studi Kasus Klasifikasi Kesegaran Ikan Berdasarkan Mata Ikan)," *Journal MIND Journal | ISSN*, vol. 8, no. 1, pp. 27–41, 2023, doi: 10.26760/mindjournal.v8i1.27-41.

[8] Y. Kim, J. Park, J. Yoon, and J. Kim, "Improved Q Network Auto-Scaling in Microservice Architecture," *Applied Sciences (Switzerland)*, vol. 12, no. 3, p. 1206, Feb. 2022, doi: 10.3390/app12031206.

[9] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Jun. 2021, doi: 10.1109/tnnls.2021.3084827.

[10] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," *arXiv preprint*, May 2019. doi:10.48550/arxiv.1905.02244.

[11] H. Syuhada, S. Hidayat, S. Mulyati, and A. Giri Persada, "Pengembangan Gamifikasi pada Pelajaran Matematika SD dengan Metode Addie untuk Meningkatkan Minat Belajar Siswa," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 1, pp. 1–14, Dec. 2023, doi: 10.36341/rabit.v9i1.466.

[12] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint.* doi:10.48550/arxiv.1905.11946.

[13] K. Naseer and H. N. Ahmed, "Effectiveness and Reliability of Artificial Intelligence in Fraud Detection: A Mixed-Method Study on Financial Audit," *Journal of Management and Informatics*, vol. 4, no. 1, pp. 706–722, Apr. 2025, doi: 10.51903/jmi.v4i1.168.

[14] A. Syafiq, M. Zuhri, N. Nafiiyah, and A. Setia Budi, "Identification of Road Damage Using the Convolutional Neural Network (CNN) Method," *Jurnal Ilmiah Sistem Informasi*, vol. 4, no. 3, pp. 802–817, Nov. 2025, doi: 10.51903/jgn2sd35.

[15] W. L. Tan and A. Menon, "3D Medical Image Reconstruction through Transformer-Based Neural Networks: A Comparative Study," *Journal of Technology Informatics and Engineering*, vol. 4, no. 3, pp. 448–462, Dec. 2025, doi: 10.51903/jtie.v4i3.453.

[16] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 10012–10022, oct. 2021, doi: 10.1109/iccv48922.2021.00986.

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint*, Oct. 2020. doi:10.48550/arxiv.2010.11929.

[18] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 568–578, doi: 10.1109/iccv48922.2021.00061.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[19] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–41, jan. 2022, doi: 10.1145/3505244.

[20] N. Park and S. Kim, "How Do Vision Transformers Work?," *arXiv preprint*, Feb. 2022. doi: 10.48550/arxiv.2202.06709.

[21] J. Daud Halim and Risal, "Klasifikasi Sel Darah Putih Menggunakan Vision Transformer (ViT)," *Jurnal Strategi: Jurnal Maranatha*, vol. 6, no. 2, pp. 291–307, Nov. 2024. [Online]. Available: https://strategi.itmaranatha.org/index.php/strategi/article/view/521

[22] J. A. Figo, N. Yudistira, and A. W. Widodo, "Deteksi Covid-19 dari Citra X-Ray Menggunakan Vision Transformer (ViT)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1116–1125, mar. 2023. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12388

[23] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT Pre-Training of Image Transformers," *arXiv preprint*, Jun. 2021. doi: 10.48550/arxiv.2106.08254.

[24] F. Sigloch, S. Sangiao, P. Orús, and J. M. de Teresa, "Large Output Voltage to Magnetic Flux Change in NanoSQUIDs Based on Direct-Write Focused Ion Beam Induced Deposition Technique," *Nanoscale Advances*, vol. 4, pp. 4628–4634, sep. 2022, doi: 10.1039/d2na00602b.

[25] L. O. Villegas, C. Moreno, M. A. Pajkos, M. Zanolin, and J. M. Antelis, "Parameter Estimation From the Core-Bounce Phase of Rotating Core Collapse Supernovae in Real Interferometer Noise," *arXiv preprint*, Apr. 2023. doi: 10.48550/arxiv.2304.01267.

[26] E. Parisotto, H. F. Song, J. W. Rae, R. Pascanu, C. Gulcehre, S. M. Jayakumar, M. Jaderberg, R. Lopez Kaufman, A. Clark, S. Noury, M. M. Botvinick, N. Heess, and R. Hadsell, "Stabilizing Transformers for Reinforcement Learning," *arXiv preprint*, oct. 2019. doi: 10.48550/arxiv.1910.06764.

[27] J. W. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," in *Proc. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP)*, Nov. 2019, pp. 6382–6388, doi: 10.18653/v1/d19-1670.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2018, doi: 10.1109/tpami.2018.2858826.