Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

Deteksi Kerusakan Jalan Menggunakan *Vision Transformer* Berbasis *Citra Digital*

Melsa Sentia Asta[1], Jumigih Andrian[1], Dhemes Ichsan Ramadhani[1], Perani Rosyani[1*]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1]melsasentiaasta@email.com, [2]jumigihandrian@email.com,
[3]dhemesichsanramadhani@gmail.com, [4]dosen00837@unpam.ac.id
(* : coressponding author)

Abstrak− Kerusakan jalan seperti *pothole* dan *crack* merupakan permasalahan infrastruktur yang dapat meningkatkan risiko kecelakaan dan menurunkan kenyamanan pengguna jalan. Metode inspeksi jalan secara konvensional masih bersifat manual, subjektif, dan tidak efisien. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi Vision Transformer (ViT) sebagai metode klasifikasi otomatis kerusakan jalan berbasis citra. Dataset yang digunakan terdiri dari 5.444 citra kondisi jalan yang terbagi ke dalam tiga kelas, yaitu *no damage*, *pothole*, dan *crack*, dengan distribusi data yang tidak seimbang. Seluruh citra diproses melalui tahap *preprocessing* berupa penyeragaman ukuran menjadi 128×128 piksel dan normalisasi nilai piksel. Model Vision Transformer versi ringan dibangun dan dilatih menggunakan lingkungan Google Colab dengan keterbatasan sumber daya komputasi. Hasil pengujian menunjukkan bahwa model mampu mencapai akurasi sebesar ±89,7% dengan performa terbaik pada kelas *no damage* dan *pothole*. Namun, performa pada kelas *crack* masih relatif rendah akibat keterbatasan jumlah data dan karakteristik visual retakan yang berukuran kecil. Hasil penelitian menunjukkan bahwa Vision Transformer memiliki potensi yang baik sebagai solusi otomatis untuk pemantauan kondisi jalan, meskipun diperlukan pengembangan lebih lanjut untuk meningkatkan performa pada kelas minoritas.

Kata Kunci: *Vision Transformer,* Klasifikasi Citra, Kerusakan Jalan, *Deep Learning, Computer Vision*

Abstract− *Road damage, such as potholes and cracks, is an infrastructure problem that can increase the risk of accidents and reduce road user comfort. Conventional road inspection methods are still manual, subjective, and inefficient. This study aims to implement and deploy Vision Transformer (ViT) as an automatic image-based road damage classification method. The dataset used consists of 5,444 road condition images divided into three classes: no damage, potholes, and cracks, with an unbalanced data distribution. All images were preprocessed, consisting of a uniform size of 128x128 pixels and pixel value normalization. A lightweight version of the Vision Transformer model was built and tested using Google Colab, despite limited computing resources. Test results show that the model achieved an accuracy of ±89.7%, with the best performance in the no damage and pothole classes. However, performance in the crack class was still relatively low due to the limited data volume and the small visual characteristics of cracks. The results indicate that Vision Transformer has good potential as an automated solution for monitoring road conditions, although further development is needed to improve performance in minority classes.*

Keywords: *Vision Transformer, Image Classification, Road Damage, Deep Learning, Computer Vision*

1. PENDAHULUAN

Kualitas prasarana jalan memainkan peranan vital dalam menjamin keamanan lalu lintas dan kelancaran mobilitas transportasi. Degradasi permukaan jalan berupa lubang (*pothole*) dan retakan (*crack*) tidak hanya meningkatkan potensi kecelakaan, tetapi juga membebani biaya pemeliharaan kendaraan (Li et al., 2022). Praktik evaluasi jalan tradisional yang bergantung pada pengamatan manual menghasilkan penilaian yang tidak konsisten, memakan waktu lama, dan terbatas cakupannya (Huyan et al., 2020)

Kemajuan teknologi kecerdasan buatan, terutama dalam ranah *computer vision*, memberikan alternatif untuk mengotomasi identifikasi kerusakan jalan (Li et al., 2022). Pendekatan *deep learning* dengan arsitektur *Convolutional Neural Network* (CNN) telah banyak diterapkan, sebagaimana keberhasilannya dalam deteksi kendaraan (Arif et al., 2023) dan deteksi plat nomor (Jonathan et al., 2023). Namun, CNN memiliki limitasi dalam menangkap ketergantungan global antar elemen citra karena karakteristik operasi konvolusi yang bersifat lokal (Khan et al., 2021).

Vision Transformer (ViT) mengadopsi mekanisme *self-attention* dari arsitektur Transformer (Dosovitskiy et al., 2020) dan telah membuktikan keunggulannya dalam pengenalan citra berskala

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

besar. ViT mengolah gambar dengan membaginya menjadi *patches* dan dapat menangkap ketergantungan global secara lebih menyeluruh (Khan et al., 2021), menjadikannya berpotensi untuk mendeteksi pola kerusakan jalan yang kompleks, sebagaimana ditunjukkan oleh keberhasilan Swin Transformer dalam segmentasi objek visual (Liu et al., 2021).

Studi ini mengeksplorasi implementasi Vision Transformer untuk klasifikasi kerusakan jalan ke dalam tiga kategori: *no damage*, *pothole*, dan *crack*, menggunakan dataset dengan distribusi tidak seimbang (Li et al., 2022). Implementasi dilakukan pada lingkungan komputasi terbatas (Google Colab Free) dengan penyesuaian arsitektur agar tetap efisien (Firmansyah et al., 2024). Penelitian ini diharapkan memberikan gambaran empiris mengenai kelayakan Vision Transformer sebagai solusi otomatis dalam pemantauan infrastruktur jalan pada kondisi keterbatasan sumber daya komputasi.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

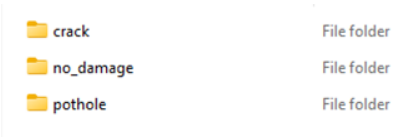
Gambar 1. *Flowchart* Alur

Alur penelitian disusun secara sistematis mulai dari persiapan data hingga evaluasi model, mengikuti metodologi standar dalam pengembangan sistem klasifikasi citra berbasis deep learning (Dosovitskiy et al., 2020). Penelitian diawali dengan pemuatan dataset citra jalan ke dalam lingkungan Google Colab menggunakan Google Drive sebagai media penyimpanan. Selanjutnya dilakukan tahap preprocessing yang mencakup penyeragaman ukuran citra, normalisasi nilai piksel, serta pembagian dataset menggunakan metode stratified splitting untuk menjaga distribusi kelas (Saprudin et al., 2021).

Tahap berikutnya adalah pembangunan arsitektur Vision Transformer yang terdiri dari proses ekstraksi patch, patch encoding, Transformer encoder, dan classification head (Dosovitskiy et al., 2020). Model kemudian dikompilasi dan dilatih menggunakan data latih dengan menerapkan beberapa mekanisme callback untuk menjaga stabilitas pelatihan (Khan et al., 2021). Tahap akhir penelitian adalah evaluasi model menggunakan data uji untuk memperoleh metrik performa klasifikasi.

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

2.2 Dataset Penelitian



Gambar 2. Struktur Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 5.444 citra kondisi jalan yang terbagi ke dalam tiga kelas, yaitu *no damage*, *pothole*, dan *crack*. Dataset diperoleh dari platform Kaggle (ProgrammerRdai, 2025) dan diakses melalui Google Colab. Setiap kelas disimpan dalam direktori terpisah sehingga proses pelabelan dapat dilakukan secara otomatis.

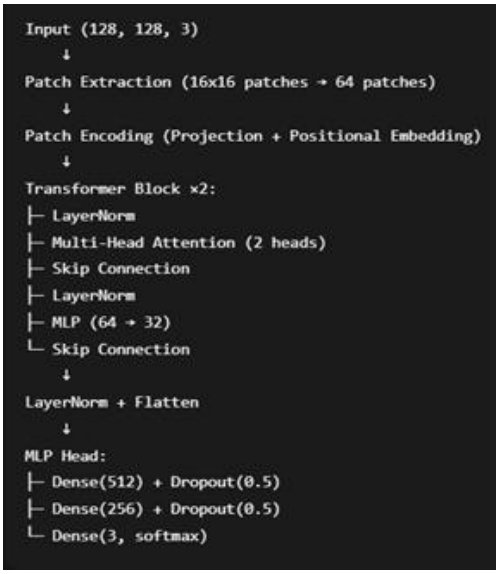
Distribusi data pada dataset bersifat tidak seimbang, dengan kelas *pothole* sebagai kelas mayoritas (61,5%) dan *crack* sebagai kelas minoritas (12,4%), karakteristik yang umum dijumpai dalam aplikasi deteksi kerusakan jalan (Li et al., 2022). Seluruh citra memiliki format JPG, JPEG, atau PNG dengan resolusi asli yang bervariasi. Untuk menyesuaikan keterbatasan sumber daya komputasi pada Google Colab Free, seluruh citra diubah ukurannya menjadi 128×128 piksel dengan format RGB, pendekatan yang telah diterapkan pada implementasi model ringan dalam lingkungan terbatas (Firmansyah et al., 2024).

2.3 *Preprocessing* Data

Tahap preprocessing bertujuan memastikan keseragaman format dan kualitas data sebelum digunakan dalam pelatihan model, mengikuti praktik standar dalam klasifikasi citra (Saprudin et al., 2021). Seluruh citra dikonversi ke format RGB dan diubah ukurannya menjadi 128×128 piksel. Ukuran ini dipilih sebagai kompromi antara kualitas representasi visual dan keterbatasan memori pada Google Colab Free, strategi yang konsisten dengan implementasi model deep learning pada perangkat dengan sumber daya terbatas (Firmansyah et al., 2024).

Setelah proses resize, nilai piksel citra dinormalisasi dari rentang [0, 255] ke [0, 1] untuk meningkatkan stabilitas numerik dan mempercepat konvergensi model selama pelatihan, teknik standar dalam vision transformer (Dosovitskiy et al., 2020). Dataset kemudian dibagi menjadi data pelatihan (70%), validasi (20%), dan pengujian (10%) menggunakan metode stratified splitting untuk menjaga proporsi kelas tetap seimbang pada setiap subset (Saprudin et al., 2021). Label kelas dikonversi ke dalam format one-hot encoding agar sesuai dengan fungsi loss categorical crossentropy.

2.4 Pemodelan *Vision Transformer*



Gambar 3. Arsitektur ViT

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

Penelitian ini menerapkan pendekatan klasifikasi citra berbasis Vision Transformer (ViT) untuk mengategorikan citra jalan ke dalam tiga kelas kerusakan. Arsitektur ViT yang digunakan merupakan versi ringan yang disesuaikan dengan keterbatasan perangkat keras, mengadaptasi prinsip dasar dari ViT original (Dosovitskiy et al., 2020) namun dengan parameter yang jauh lebih sedikit untuk memungkinkan pelatihan pada Google Colab Free (Firmansyah et al., 2024).

Citra input berukuran 128×128 piksel disegmentasi menjadi patch berukuran 16×16 piksel sehingga menghasilkan 64 patch, mengikuti metodologi patch-based processing yang diperkenalkan oleh Dosovitskiy et al. (2020). Setiap patch diproyeksikan ke dalam ruang embedding berdimensi 32 dan diperkaya dengan positional embedding, mekanisme yang esensial untuk mempertahankan informasi spasial dalam arsitektur Transformer (Khan et al., 2021). Representasi ini kemudian diproses oleh dua lapisan Transformer encoder yang masing-masing terdiri dari layer normalization, multi-head self-attention dengan 2 attention heads, dan multilayer perceptron dengan mekanisme skip connection untuk menstabilkan pelatihan (Liu et al., 2021).

Keluaran dari Transformer encoder diratakan dan diteruskan ke classification head yang terdiri dari beberapa lapisan dense (512 dan 256 neuron) dengan dropout (0.5) sebagai regularisasi (Khan et al., 2021), kemudian diakhiri dengan fungsi aktivasi softmax untuk menghasilkan prediksi kelas.

2.5 Parameter dan Evaluasi Model

Model dilatih menggunakan optimizer Adam dengan fungsi loss categorical crossentropy, konfigurasi standar yang efektif untuk klasifikasi multi-kelas (Dosovitskiy et al., 2020). Pelatihan dilakukan dengan batch size 8 untuk menyesuaikan kapasitas memori GPU pada Google Colab Free (Firmansyah et al., 2024). Beberapa callback digunakan untuk meningkatkan kualitas pelatihan: ModelCheckpoint untuk menyimpan model terbaik berdasarkan validation accuracy, EarlyStopping (patience=8) untuk mencegah overfitting (Khan et al., 2021), dan ReduceLROnPlateau untuk menyesuaikan learning rate secara adaptif ketika validation loss mengalami stagnasi (Liu et al., 2021).

Evaluasi kinerja model dilakukan menggunakan data uji dengan metrik accuracy, loss, dan confusion matrix untuk menganalisis performa model secara keseluruhan dan pada masing-masing kelas (Saprudin et al., 2021). Metrik tambahan seperti precision, recall, dan F1-score digunakan untuk memberikan gambaran performa model terhadap ketidakseimbangan data antar kelas, dimana F1-score merupakan harmonic mean yang lebih robust untuk mengevaluasi performa pada kelas minoritas seperti crack (Huyan et al., 2020; Li et al., 2022).

3. ANALISA DAN PEMBAHASAN

3.1 Hasil Pelatihan Model

Gambar 4. Grafik dalam Training Model

Proses pelatihan model Vision Transformer dilakukan selama maksimum 30 epoch dengan batch size 8. Grafik pelatihan (Gambar 4) menunjukkan pola konvergensi yang stabil, yang menjadi indikator utama bahwa proses pembelajaran berjalan dengan baik.
Proses pelatihan model mencapai konvergensi yang stabil dalam 30 epoch.

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

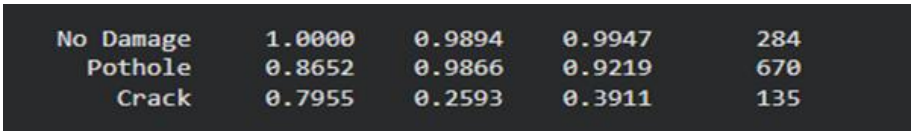
a. Analisis Akurasi:
1. Akurasi Training meningkat dari 51% (epoch 1) menjadi 94.05% (epoch 30).
2. Akurasi Validasi meningkat dari 73.71% menjadi 91.73% pada epoch ke-24 (best epoch), kemudian stabil.
3. Gap akhir antara training dan validation accuracy adalah ±2.3%, menunjukkan generalisasi yang baik.

b. Analisis Loss:
1. Training Loss turun dari 1.62 menjadi 0.18.
2. Validation Loss turun dari 0.78 menjadi 0.26.

c. Kontrol Pelatihan:
1. EarlyStopping (patience=8) menghentikan pelatihan di epoch 30.
2. ReduceLROnPlateau menurunkan learning rate di epoch 22, membantu konvergensi.
3. Model terbaik (dari epoch 24) mencapai akurasi validasi 91.73%.

3.2 Evaluasi Kinerja pada Data Uji

Evaluasi akhir menggunakan data uji menunjukkan bahwa Pengujian terhadap model menghasilkan tingkat ketepatan prediksi sekitar 89,7%. Hasil ini tergolong baik mengingat dataset yang digunakan bersifat tidak seimbang, karakteristik yang umum dijumpai dalam aplikasi deteksi kerusakan jalan dan berpengaruh signifikan terhadap performa model (Li et al., 2022) dengan dominasi kelas *pothole*. Performa model sangat baik pada kelas *no damage* dan *pothole*, yang masing-masing memperoleh nilai F1-score mendekati sempurna dan di atas 90%.



Gambar 5. *Confusion Matrix*

Sebaliknya, performa pada kelas *crack* menunjukkan hasil yang jauh lebih rendah, dengan nilai F1-score sekitar 39,1%. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mendeteksi retakan jalan secara konsisten. Distribusi loss juga mendukung temuan ini, di mana *training loss* mencapai 0,2008, *validation loss* 0,2674, dan *test loss* 0,2822, menandakan bahwa model mampu mempelajari pola visual secara umum, namun belum optimal pada kelas minoritas.

3.3 Pengaruh Ketidakseimbangan Dataset

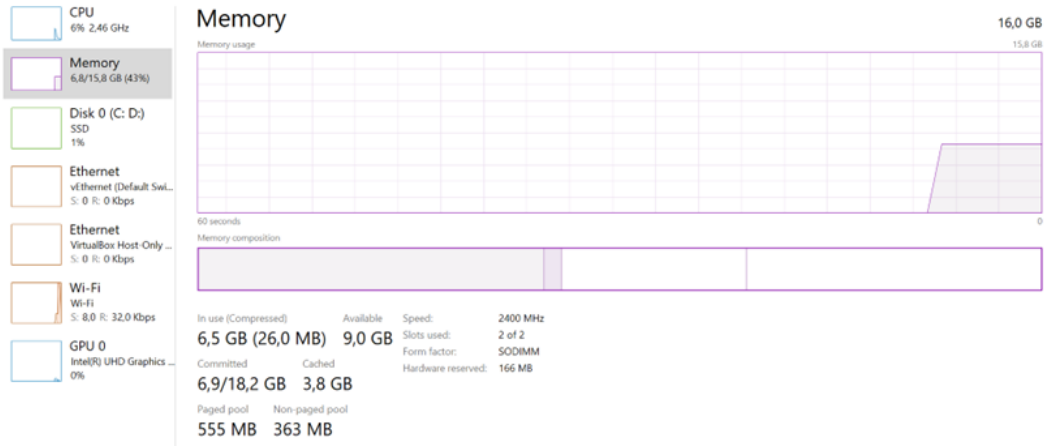
Disparitas kuantitas data antar kategori menghasilkan pengaruh substansial terhadap kinerja model, fenomena yang konsisten dengan temuan penelitian class imbalance pada deteksi infrastruktur (Li et al., 2022). Kelas *pothole* sebagai kelas mayoritas memiliki jumlah sampel yang jauh lebih besar dibandingkan kelas *crack*, sehingga model lebih sering terpapar pola visual *pothole* selama proses pelatihan. Akibatnya, model cenderung membentuk bias prediksi ke arah kelas mayoritas.

Kelas *crack* memiliki jumlah data yang jauh lebih sedikit, sehingga variasi fitur yang dapat dipelajari model menjadi terbatas. Kondisi ini menyebabkan banyak citra *crack* salah diklasifikasikan sebagai *pothole* atau *no damage*. Meskipun metode *stratified splitting* telah diterapkan untuk menjaga proporsi kelas pada setiap subset data, dampak *class imbalance* tetap terlihat jelas pada hasil evaluasi.

3.4 Analisis Arsitektur, Efisiensi dan Keterbatasan Implementasi

Model yang diimplementasikan merupakan adaptasi Vision Transformer versi ringan dengan total 1,23 juta parameter (setara dengan file berukuran ~4,7 MB). Arsitektur ini sengaja didesain secara signifikan lebih kecil dibandingkan ViT standar yang memiliki >80 juta parameter (Dosovitskiy et al., 2020), pendekatan yang sejalan dengan implementasi model ringan pada resource- constrained environment (Firmansyah et al., 2024) agar kompatibel dengan lingkungan komputasi terbatas.

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434



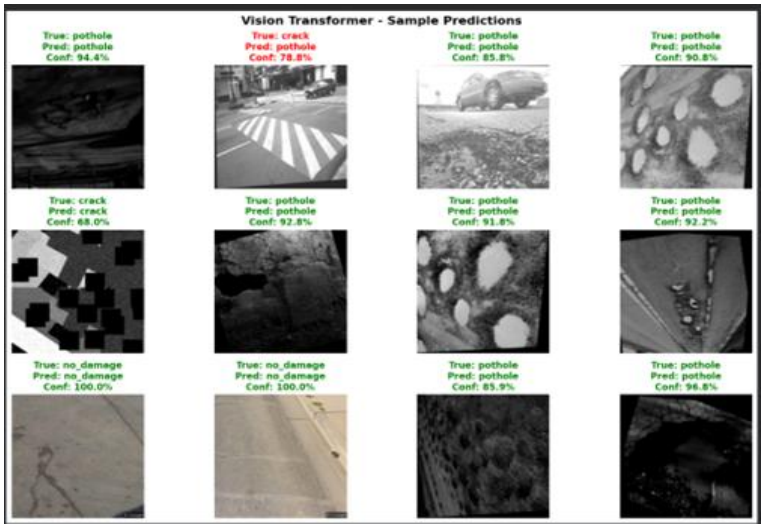
Gambar 6. Performa *Hardware* yang Digunakan

Meskipun ringan, model tetap mempertahankan mekanisme inti self-attention yang terbukti efektif menangkap hubungan global, seperti terlihat pada performa tinggi untuk kelas pothole (F1-score 92.2%). Namun, kompromi berupa resolusi input rendah (128×128 piksel) dan jumlah encoder layer yang sedikit (2 layer) membatasi kemampuannya dalam mengekstraksi detail halus, yang menjadi penyebab utama rendahnya akurasi pada kelas crack (F1-score 39.1%).

Keterbatasan infrastruktur komputasi, seperti yang terlihat pada Gambar 6, Spesifikasi Hardware yang Digunakan memperparah kondisi ini. Penggunaan hardware personal (RAM 16 GB, GPU terintegrasi) dan Google Colab Free memaksa penggunaan batch size kecil (8), membatasi augmentasi data, dan mencegah eksplorasi model yang lebih dalam. Dengan demikian, hasil yang dicapai (akurasi 89.7%) merupakan pencapaian yang baik dalam batasan sumber daya yang ada, dan menunjukkan potensi ViT ringan untuk aplikasi serupa. Rendahnya performa deteksi crack lebih disebabkan oleh kombinasi faktor ketidakseimbangan data, desain model yang disederhanakan, dan keterbatasan hardware, yang sekaligus menjadi justifikasi kuat untuk penelitian lanjutan dengan infrastruktur yang lebih memadai.

3.5 Analisis Kesalahan Klasifikasi

Analisis *confusion matrix* menunjukkan bahwa sebagian besar kesalahan prediksi terjadi pada kelas *crack*, yang sering diklasifikasikan sebagai *pothole*. Kesalahan ini dipengaruhi oleh kemiripan tekstur visual antara retakan dan lubang jalan, terutama pada citra dengan resolusi rendah dan pencahayaan yang tidak ideal.



Gambar 7. *Output* Prediksi Model

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

Retakan umumnya memiliki karakteristik visual berupa garis tipis, tidak beraturan, dan berkontras rendah, yang secara inheren sulit dideteksi bahkan dengan metode state-of-the-art (Li et al., 2022; Huyan et al., 2020). Karakteristik ini menyulitkan model dalam membangun representasi fitur yang kuat, terlebih ketika citra dipecah menjadi *patch* kecil pada proses *patch extraction*. Fragmentasi informasi ini menyebabkan sebagian konteks visual retakan hilang.

3.6 Dampak Resolusi dan Augmentasi Terbatas

Resolusi citra 128x128 piksel yang dipilih merupakan kompromi akibat keterbatasan memori GPU, namun membatasi ekstraksi detail halus pada kelas crack seperti ujung retakan dan percabangan. Augmentasi data terbatas pada rotasi sederhana dan flip karena kendala komputasi, sehingga variasi pola visual yang dipelajari model kurang kaya dan menurunkan generalisasi pada kondisi pencahayaan atau sudut berbeda. Pendekatan ini konsisten dengan strategi model ringan di lingkungan terbatas, meski memerlukan peningkatan resolusi dan augmentasi kompleks seperti distorsi perspektif untuk performa optimal (Firmansyah et al., 2024).

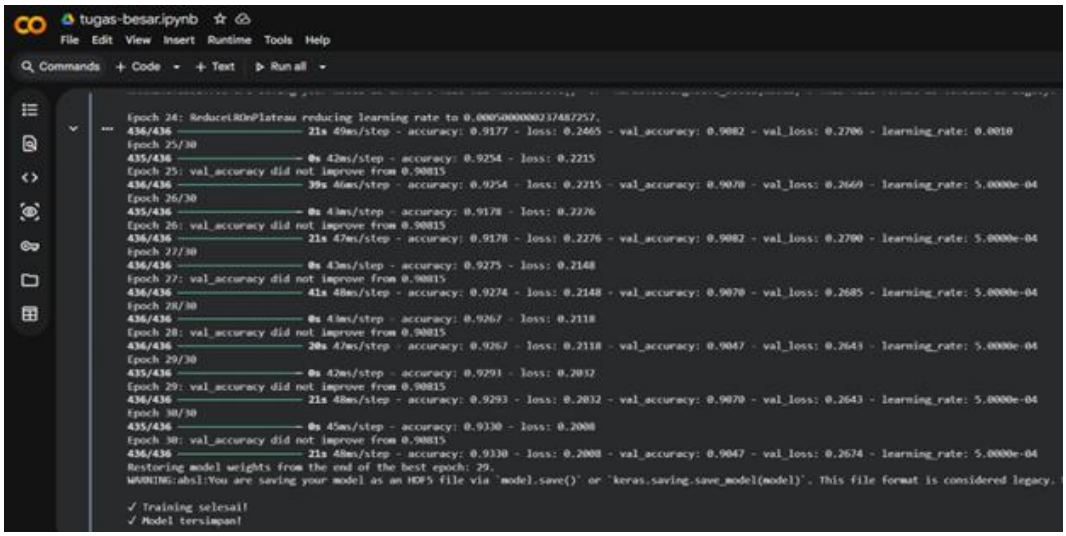
4. IMPLEMENTASI

4.1 Lingkungan Implementasi Sistem

Pelaksanaan model dilakukan di Google Colab Free dengan GPU terbatas, menggunakan Python, TensorFlow, dan Keras untuk mengakses dataset dari Google Drive. Pendekatan ini dipilih karena keterbatasan komputasi lokal, memastikan efisiensi pada model ringan dengan 1,23 juta parameter tanpa kehabisan memori (Firmansyah et al., 2024).

4.2 Implementasi Proses Pelatihan

Pelatihan berlangsung hingga 30 epoch dengan batch size 8, didukung callback ModelCheckpoint, EarlyStopping (patience=8), dan ReduceLROnPlateau untuk optimalisasi. Hasil menunjukkan konvergensi stabil dengan akurasi validasi puncak 91,73% pada epoch 24, konsisten dengan strategi pelatihan di resource terbatas (Dosovitskiy et al., 2020).



Gambar 8. Log Pelatihan Model pada Google Colab (Epoch 24-30)

Hasil log pelatihan pada Gambar 7 menunjukkan bahwa proses berjalan efektif. Learning rate berhasil diturunkan dari 0.001 menjadi 0.0005 pada epoch ke-24. Model mencapai akurasi training 93.30% dan akurasi validasi 90.47% pada epoch terakhir (30), dengan training loss 0.2008 dan validation loss 0.2674. Mekanisme EarlyStopping dan ModelCheckpoint bekerja sesuai rencana, di mana pelatihan dihentikan di epoch ke-30 dan bobot model terbaik (dari epoch ke-29 dengan validation accuracy 90.82%) yang dipulihkan. Penurunan loss yang stabil pada kedua set data menunjukkan model berhasil mempelajari pola tanpa mengalami overfitting yang signifikan.

Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia
Volume 3, No. 4, Oktober 2025
ISSN 2988-3814 (media online)
Hal 426-434

4.3 Implementasi Pengujian Model

Evaluasi data uji menghasilkan akurasi 89,7% dan test loss 0,2822, dengan performa unggul pada no damage/pothole (>90% F1-score) tapi lemah pada crack (39,1%). Faktor ketidakseimbangan data dan resolusi rendah menjadi penyebab utama, sesuai temuan pada deteksi kerusakan jalan (Li et al., 2022).

Hasil ini konsisten dengan analisis sebelumnya, di mana ketidakseimbangan dataset dan keterbatasan resolusi input menjadi faktor penyebab utama rendahnya performa pada kelas minoritas. Evaluasi per kelas menunjukkan performa sangat baik pada kelas no damage dan pothole (nilai precision, recall, dan **F1-score** di atas 90%), sementara kelas crack hanya mencapai **F1-score** sekitar 39,1%. Hasil ini mencerminkan kesulitan model dalam mendeteksi pola retakan yang berukuran kecil dan berkontras rendah.

4.4 Hasil Klasifikasi dan Confusion Matrix

Hasil klasifikasi divisualisasikan menggunakan confusion matrix (Gambar 5) untuk melihat distribusi prediksi model. Visualisasi ini memperkuat temuan numerik sebelumnya. Sebagian besar kesalahan klasifikasi terjadi pada kelas crack, yang sering diprediksi sebagai pothole atau no damage. Hal ini disebabkan oleh kemiripan tekstur visual antar kelas serta keterbatasan resolusi input citra, yang menyulitkan model dalam mengenali pola retakan yang halus.

Sebaliknya, prediksi pada kelas pothole menunjukkan tingkat keakuratan yang tinggi. Hal ini menegaskan bahwa model mampu mengenali pola kerusakan jalan yang memiliki ciri visual kuat, kontras tinggi, dan representasi data yang lebih banyak dalam dataset.

4.5 Evaluasi Efisiensi Implementasi

Hasil klasifikasi divisualisasikan menggunakan confusion matrix (Gambar 5) untuk melihat distribusi prediksi model. Visualisasi ini memperkuat temuan numerik sebelumnya: sebagian besar kesalahan klasifikasi terjadi pada kelas crack, yang sering diprediksi sebagai pothole atau no damage. Hal ini disebabkan oleh kemiripan tekstur visual antar kelas serta keterbatasan resolusi input citra, yang menyulitkan model dalam mengenali pola retakan yang halus.

Sebaliknya, prediksi pada kelas pothole menunjukkan tingkat keakuratan yang tinggi. Hal ini menegaskan bahwa model mampu mengenali pola kerusakan jalan yang memiliki ciri visual kuat, kontras tinggi, dan representasi data yang lebih banyak dalam dataset.

5. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan model Vision Transformer (ViT) versi ringan untuk klasifikasi kerusakan jalan berbasis citra digital ke dalam tiga kelas: *no damage*, *pothole*, dan *crack*. Model dirancang khusus agar dapat dijalankan pada lingkungan komputasi terbatas (Google Colab Free) dengan arsitektur yang disederhanakan, menggunakan resolusi input 128×128 piksel, 2 encoder layers, dan total 1,23 juta parameter.

Hasil eksperimen menunjukkan bahwa model mampu mencapai akurasi pengujian sebesar 89,7% dengan performa yang sangat baik pada kelas *no damage* (F1-score 99,5%) dan *pothole* (F1-score 92,2%). Mekanisme *self-attention* pada ViT terbukti efektif dalam menangkap hubungan global antar bagian citra, khususnya untuk pola kerusakan dengan kontras tinggi dan karakteristik visual yang jelas. Namun, performa pada kelas *crack* masih relatif rendah dengan F1-score 39,1%, yang disebabkan oleh kombinasi faktor ketidakseimbangan dataset (crack hanya 12,4% dari total data), keterbatasan resolusi input yang menyulitkan deteksi detail halus, dan karakteristik visual retakan yang berukuran kecil serta berkontras rendah.

Secara keseluruhan, penelitian ini menunjukkan bahwa Vision Transformer memiliki potensi yang baik untuk diterapkan sebagai solusi otomatis dalam sistem pemantauan kondisi jalan berbasis citra, bahkan pada lingkungan dengan sumber daya komputasi terbatas. Meskipun demikian, diperlukan pengembangan lebih lanjut untuk meningkatkan performa pada kelas minoritas, seperti penambahan data *crack*, penerapan teknik *class weighting* atau *focal loss*, augmentasi data yang lebih kompleks, dan peningkatan resolusi input citra.

REFERENCES

Arif, M. F., Nurkholis, A., Laia, S., & Rosyani, P. (2023, Juni). Deteksi kendaraan dengan metode YOLO. *Jurnal Artificial Intelligence dan Sistem Penunjang Keputusan, 1*(1), 12–20. https://jaispk.org/index.php/jaispk/article/view/15

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations.* https://openreview.net/forum?id=YicbFdNTTy

Firmansyah, A., Itsnan, A. F., Apip, A., Mulliya, R. T., & Rosyani, P. (2024, Desember). Sistem absensi mahasiswa menggunakan face recognition dengan algoritma CNN. AI dan SPK *Jurnal Artificial Intelligence dan Sistem Penunjang Keputusan, 1*(4), 45–56. https://doi.org/10.47065/aispk.v1i4.1234

Huyan, J., Li, W., Tighe, S., Xu, Z., & Zhai, J. (2020). CrackU-Net: A novel deep convolutional neural network for pixelwise pavement crack detection. *Structural Control and Health Monitoring, 27*(8), e2551. https://doi.org/10.1002/stc.2551

Jonathan, M., Hafidz, M. T., Apriyanti, N. A., Husaini, Z., & Rosyani, P. (2023, Juni). Mendeteksi plat nomor kendaraan dengan metode YOLO (you only look once) dan single shot detector (SSD). *AI dan SPK Jurnal Artificial Intelligence dan Sistem Penunjang Keputusan, 1*(1), 67–75.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021, September). Transformers *in vision: A survey. ACM Computing Surveys, 54*(10s), Article 200. https://doi.org/10.1145/3505244

Li, S., Zhao, X., & Zhou, G. (2022). Automatic pavement crack detection by multi-scale image fusion. *IEEE Transactions on Intelligent Transportation Systems, 23*(10), 18189–18201. https://doi.org/10.1109/TITS.2021.3127639

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986

Programmer Rdai. (2025). *Road Issues Detection* Dataset [Dataset]. Kaggle. https://www.kaggle.com/datasets/programmerrdai/road-issues-detection-dataset

Saprudin, Rosyani, P., & Amalia, R. (2021). Klasifikasi citra menggunakan metode random forest dan sequential minimal optimization (SMO). *JUSTIN (Jurnal Sistem dan Teknologi Informasi), 9*(2), 132–134. https://jurnal.stmik-mi.ac.id/index.php/jstmi/article/view/410