

IMPLEMENTASI VISION TRANSFORMER (ViT) UNTUK KLASIFIKASI CITRA WAJAH ASLI DAN HASIL MANIPULASI ARTIFICIAL INTELLIGENCE

Winesti Widasari¹⁾, Alun Sujjada²⁾, Ivana Lucia Kharisma³⁾

^{1,2,3} Program Studi Teknik Informatika, Universitas Nusa Putra.

Corresponding Author: ¹ winesti.widasari_ti22@nusaputra.ac.id

Article Info

Article history:

Received: Aug 02, 2026

Revised: Aug 20, 2026

Accepted: Oct 11, 2026

Published: Oct 15, 2026

Keywords:

Vision Transformer,
Deepfake Detection,
Image Classification,
Artificial Intelligence,
Deep Learning.

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has enabled the creation of highly realistic manipulated facial images, making them increasingly difficult to distinguish from authentic images. This study aims to implement the Vision Transformer (ViT) model for classifying real and AI-generated facial images. An experimental research approach was employed using a publicly available dataset consisting of real and fake facial images. The preprocessing stage included image resizing, normalization, and data augmentation to improve model generalization. The dataset was divided into training, validation, and testing subsets prior to model training. The Vision Transformer Base Patch16-224 model was fine-tuned to learn global visual representations of facial images. Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. The experimental results demonstrate that the proposed model achieved an accuracy of 99.94%, with precision, recall, and F1-score of 99.94%, indicating excellent capability in distinguishing authentic facial images from AI-generated manipulated images. Furthermore, the ROC analysis produced an Area Under the Curve (AUC) of 1.0000, confirming the model's outstanding classification performance. These findings indicate that the Vision Transformer model is highly effective for facial image authenticity verification and has strong potential for supporting digital forensic applications and enhancing digital security systems against AI-generated image manipulation.



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. PENDAHULUAN

Perkembangan Artificial Intelligence (AI) telah mendorong kemajuan yang signifikan dalam berbagai bidang, khususnya pada pengolahan citra digital (*computer vision*). Salah satu teknologi yang berkembang pesat adalah deep learning, yaitu pendekatan pembelajaran mesin yang mampu mempelajari representasi data secara otomatis melalui jaringan saraf berlapis. Berkat kemampuannya dalam mengekstraksi karakteristik visual secara efektif, *deep learning* telah banyak diterapkan pada berbagai tugas, seperti klasifikasi citra, deteksi objek, segmentasi citra, dan pengenalan wajah [1], [2], [3]

Kemajuan deep learning juga melahirkan teknologi deepfake yang mampu menghasilkan citra wajah dengan tingkat realisme yang semakin tinggi. Teknologi ini memberikan manfaat pada berbagai bidang, seperti industri kreatif, hiburan, dan pendidikan. Namun, penyalahgunaan teknologi

deepfake dapat menimbulkan berbagai dampak negatif, antara lain penyebaran informasi palsu (*misinformation*), pemalsuan identitas, pelanggaran privasi, hingga ancaman terhadap keamanan digital. [4], [5] Seiring meningkatnya kualitas citra hasil manipulasi AI, proses identifikasi secara visual menjadi semakin sulit sehingga diperlukan metode deteksi otomatis yang memiliki tingkat akurasi tinggi.

Berbagai penelitian telah mengembangkan metode deteksi citra manipulasi AI menggunakan pendekatan berbasis *deep learning*. Dang *et al.* mengkaji berbagai teknik deteksi manipulasi wajah digital dan menunjukkan bahwa model berbasis jaringan saraf mampu mengenali karakteristik manipulasi dengan performa yang baik [6]. Penelitian Hao Lin *et al.* kemudian mengusulkan pendekatan yang menggabungkan Multi-scale Convolution dengan Vision Transformer untuk meningkatkan kemampuan deteksi *deepfake* melalui kombinasi ekstraksi fitur lokal dan global [7]. Selain itu, Das *et*

al. mengembangkan Edge-Enhanced Vision Transformer yang memanfaatkan informasi tepi citra untuk meningkatkan akurasi deteksi gambar hasil generasi AI [8]. Hasil dari penelitian-penelitian tersebut menunjukkan bahwa pendekatan berbasis Transformer memiliki potensi yang besar dalam mendeteksi citra manipulasi AI.

Salah satu arsitektur Transformer yang banyak digunakan pada tugas klasifikasi citra adalah Vision Transformer (ViT) yang diperkenalkan oleh Dosovitskiy *et al.* [9]. Berbeda dengan arsitektur *Convolutional Neural Network* (CNN), Vision Transformer membagi citra menjadi beberapa bagian (*patch*) kemudian memanfaatkan mekanisme *self-attention* untuk mempelajari hubungan antar-*patch* secara global. Pendekatan tersebut memungkinkan model memperoleh representasi visual yang lebih menyeluruh sehingga mampu meningkatkan performa pada berbagai tugas klasifikasi citra. Penelitian Ghita *et al.* juga menunjukkan bahwa Vision Transformer mampu memberikan performa yang kompetitif dalam mendeteksi citra wajah hasil manipulasi AI [10].

Meskipun berbagai penelitian telah berhasil menerapkan CNN maupun Vision Transformer untuk mendeteksi citra manipulasi AI, sebagian besar penelitian masih berfokus pada pengembangan arsitektur model atau membandingkan beberapa metode klasifikasi. Penelitian oleh Dharma *et al.* menggunakan EfficientNet untuk mendeteksi citra *deepfake*, sedangkan Aria *et al.* membandingkan performa Vision Transformer dan EfficientNet-B0 pada klasifikasi citra asli dan citra hasil generasi AI [11], [12]. Selain itu, Rumapea mengevaluasi kinerja CNN dan Vision Transformer pada klasifikasi citra resolusi tinggi [13]. Namun, implementasi Vision Transformer menggunakan dataset 140K Real and Fake Faces dengan evaluasi komprehensif menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *ROC Curve* masih perlu dikaji lebih lanjut.

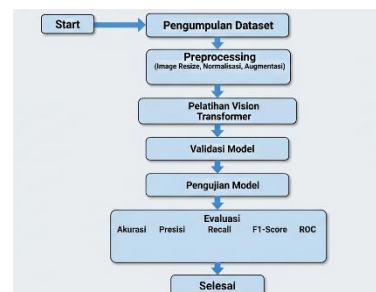
Vision Transformer (ViT) merupakan arsitektur deep learning yang mengadaptasi mekanisme Transformer dari bidang pemrosesan bahasa alami (Natural Language Processing/NLP) ke dalam tugas pengolahan citra. Berbeda dengan *Convolutional Neural Network* (CNN) yang mengekstraksi fitur menggunakan operasi konvolusi, ViT membagi citra menjadi sejumlah *patch* berukuran tetap, kemudian setiap *patch* direpresentasikan sebagai vektor dan diproses menggunakan mekanisme *self-attention* untuk mempelajari hubungan global antarbagian citra. Pendekatan ini memungkinkan model menangkap informasi spasial secara lebih menyeluruh sehingga efektif untuk berbagai tugas klasifikasi citra [9].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengimplementasikan Vision Transformer (ViT) untuk mengklasifikasikan citra wajah asli (*real*) dan citra wajah hasil manipulasi AI (*fake*). Penelitian menggunakan dataset publik yang terdiri atas 161.972

citra wajah, meliputi 81.000 citra Real dan 80.972 citra Fake. Model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *Receiver Operating Characteristic (ROC) Curve*. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan metode deteksi citra manipulasi AI yang akurat serta mendukung penerapan sistem verifikasi keaslian citra pada berbagai aplikasi keamanan digital.

2. MATERIALS AND METHODS

Penelitian ini menggunakan metode eksperimen (*experimental research*) untuk mengimplementasikan model Vision Transformer (ViT) dalam mengklasifikasikan citra wajah asli (*Real*) dan citra wajah hasil manipulasi AI (*Fake*). Metode eksperimen dipilih karena memungkinkan evaluasi performa model secara objektif berdasarkan data pelatihan, validasi, dan pengujian yang telah disiapkan. Tahapan penelitian meliputi pengumpulan dataset, *preprocessing*, pelatihan model, pengujian, serta evaluasi performa menggunakan beberapa metrik pengujian [6], [7]. Diagram alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alur penelitian

2.1 Dataset

Dataset yang digunakan merupakan dataset publik yang terdiri atas citra wajah asli (*Real*) dan citra wajah hasil manipulasi AI (*Fake*). Dataset tersebut telah banyak dimanfaatkan pada penelitian klasifikasi citra wajah berbasis *deep learning* dan deteksi citra hasil manipulasi AI [2], [14]

Berdasarkan proses identifikasi dan penyiapan data, diperoleh 161.972 citra wajah, yang terdiri atas 81.000 citra Real dan 80.972 citra Fake. Dataset kemudian dibagi menjadi data pelatihan (*training*), data validasi (*validation*), dan data pengujian (*testing*) dengan rincian seperti pada Tabel 1.

Tabel 1. Distribusi Dataset

Subset	Real	Fake	Total
Training	64.800	64.776	129.576
Validation	8.100	8.098	16.198
Testing	8.100	8.098	16.198
Total	81.000	80.972	161.972

Distribusi data yang relatif seimbang pada kedua kelas diharapkan mampu membantu model mempelajari karakteristik citra secara optimal dan mengurangi potensi bias selama proses pelatihan.

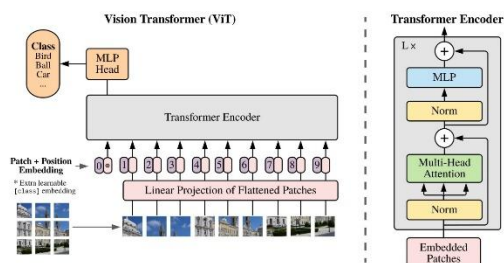
2.2 Preprocessing Data

Tahap *preprocessing* dilakukan untuk menghasilkan data masukan yang sesuai dengan kebutuhan model Vision Transformer. Seluruh citra diubah menjadi ukuran 224×224 piksel, kemudian dilakukan normalisasi menggunakan ViT Image Processor. Selain itu, diterapkan teknik *data augmentation* berupa Random Horizontal Flip, Random Rotation, dan Color Jitter untuk meningkatkan variasi data pelatihan dan mengurangi risiko *overfitting*. Teknik *data augmentation* tersebut bertujuan meningkatkan kemampuan generalisasi model dan mengurangi risiko *overfitting* selama proses pelatihan [15]

2.3 Vision Transformer

Model yang digunakan adalah Vision Transformer Base Patch16-224 (ViT-B/16) yang diperkenalkan oleh Dosovitskiy et al.[9]. Berbeda dengan Convolutional Neural Network (CNN), Vision Transformer membagi citra menjadi beberapa bagian (*patch*) berukuran 16×16 piksel. Setiap *patch* diubah menjadi representasi vektor (*embedding*) dan diproses menggunakan mekanisme self-attention pada Transformer Encoder, sehingga model mampu mempelajari hubungan global antarbagian citra.

Pendekatan tersebut memungkinkan model menghasilkan representasi fitur yang lebih komprehensif dibandingkan metode konvolusi yang berfokus pada ekstraksi fitur lokal. Kemampuan tersebut menjadikan Vision Transformer efektif untuk berbagai tugas klasifikasi citra, termasuk deteksi citra hasil manipulasi AI [9], [10] Arsitektur Vision Transformer yang digunakan pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Arsitektur Vision Transformer

2.4 Pelatihan Model

Pelatihan model dilakukan menggunakan teknik *fine-tuning* terhadap model Vision Transformer Base Patch16-224 yang telah melalui proses *pre-trained* [9]. Proses optimasi menggunakan algoritma AdamW dengan fungsi kerugian (*loss function*) CrossEntropyLoss. Model dilatih menggunakan *batch size* sebesar 32, *learning rate* sebesar 2×10^{-5} , dan

weight decay sebesar 0,01 selama 3 epoch. Penyesuaian laju pembelajaran dilakukan menggunakan CosineAnnealingLR dengan nilai *eta_min* sebesar 1×10^{-6} untuk membantu proses konvergensi model selama pelatihan. Konfigurasi parameter pelatihan yang digunakan pada penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Parameter Pelatihan

Parameter	Nilai
Epoch	3
Batch Size	32
Learning Rate	2×10^{-5}
Optimizer	AdamW
Loss Function	CrossEntropyLoss
Learning Rate Scheduler	CosineAnnealingLR
Weight Decay	0.01

2.5 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik Accuracy, Precision, Recall, dan F1-Score untuk mengukur performa klasifikasi. Selain itu, digunakan Confusion Matrix untuk menganalisis hasil prediksi pada setiap kelas serta Receiver Operating Characteristic (ROC) Curve dan Area Under Curve (AUC) untuk mengevaluasi kemampuan model dalam membedakan citra wajah asli (*Real*) dan citra wajah hasil manipulasi AI (*Fake*). Penggunaan metrik tersebut telah banyak diterapkan pada penelitian klasifikasi citra dan deteksi *deepfake* karena mampu memberikan evaluasi yang komprehensif terhadap performa model [6], [7], [10], [11]

3. RESULTS AND DISCUSSION

3.1. Hasil Pelatihan Model

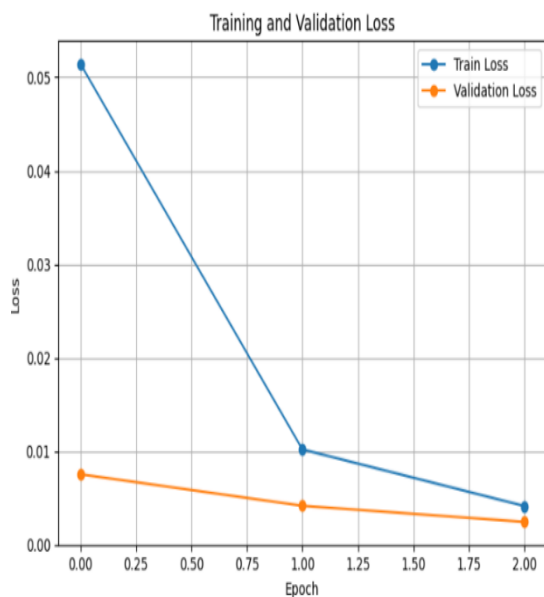
Vision Transformer dilatih menggunakan 129.576 data pelatihan dan 16.198 data validasi selama 3 epoch. Proses pelatihan dilakukan menggunakan *optimizer* AdamW dengan *learning rate* sebesar 2×10^{-5} , *batch size* 32, serta fungsi kerugian (*loss function*) CrossEntropyLoss. Selama proses pelatihan dilakukan pengamatan terhadap nilai *training loss*, *training accuracy*, *validation loss*, dan *validation accuracy* pada setiap epoch untuk mengevaluasi proses pembelajaran model.

Tabel 3. Hasil Pelatihan Model

Epoch	Train Loss	Train Acc(%)	Val Loss	Val Acc(%)
1	0.0493	98.12%	0.0052	99.85%
2	0.0117	99.61%	0.0063	99.78%
3	0.0039	99.86%	0.0027	99.92%

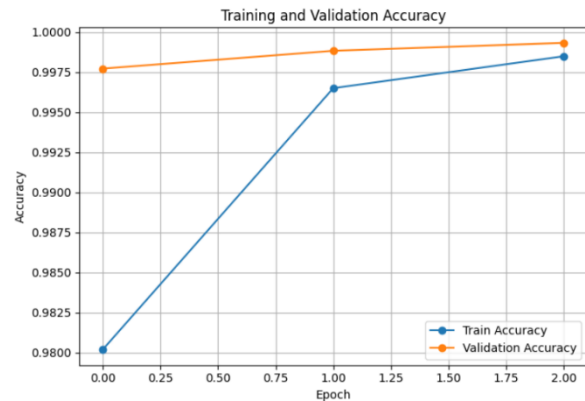
Berdasarkan Tabel 3, nilai *training loss* mengalami penurunan dari 0.0493 pada epoch pertama menjadi 0.0039 pada epoch ketiga. Penurunan tersebut menunjukkan bahwa model semakin mampu meminimalkan kesalahan prediksi selama proses pelatihan. Di sisi lain, *training accuracy* meningkat dari 98.12% menjadi 99.86%, yang menunjukkan bahwa model mampu mempelajari karakteristik citra wajah secara efektif.

Pada data validasi, model memperoleh *validation accuracy* sebesar 99.85%, 99.78%, dan 99.92% pada masing-masing epoch. Meskipun terjadi sedikit penurunan pada epoch kedua, nilai *validation accuracy* kembali meningkat pada epoch ketiga dengan *validation loss* yang semakin kecil. Hasil tersebut menunjukkan bahwa model memiliki kemampuan generalisasi yang baik tanpa mengalami *overfitting* yang signifikan.



Gambar 3. Grafik Training Loss dan Validation Loss

Gambar 3. menunjukkan perubahan nilai *training loss* dan *validation loss* selama proses pelatihan. Nilai *training loss* mengalami penurunan dari 0.0493 pada epoch pertama menjadi 0.0039 pada epoch ketiga. Sementara itu, *validation loss* tetap berada pada nilai yang rendah dengan sedikit fluktuasi pada epoch kedua sebelum kembali menurun pada epoch ketiga. Hasil tersebut menunjukkan bahwa model mampu meminimalkan kesalahan prediksi selama proses pelatihan dan memiliki kemampuan generalisasi yang baik.



Gambar 4. Grafik Training Accuracy dan Validation Accuracy

Gambar 4. menunjukkan perkembangan *training accuracy* dan *validation accuracy* selama proses pelatihan. Nilai *training accuracy* meningkat dari 98.12% pada epoch pertama menjadi 99.86% pada epoch ketiga. Sementara itu, *validation accuracy* tetap berada di atas 99% pada setiap epoch dan mencapai 99.92% pada akhir pelatihan. Hasil tersebut menunjukkan bahwa model mampu mempelajari karakteristik citra dengan baik tanpa mengalami *overfitting* yang signifikan.

3.2 Evaluasi Model

Evaluasi dilakukan menggunakan 16.198 data pengujian yang tidak digunakan selama proses pelatihan maupun validasi. Kinerja model diukur menggunakan metrik Accuracy, Precision, Recall, F1-Score, Confusion Matrix, dan ROC-AUC.

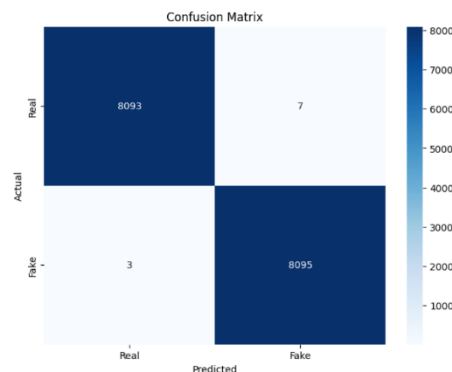
Model memperoleh test accuracy sebesar 99.94% dengan test loss sebesar 0.0016. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan citra wajah asli (*Real*) dan citra hasil manipulasi AI (*Fake*) dengan tingkat ketepatan yang sangat tinggi. Konsistensi antara hasil validasi (99.92%) dan hasil pengujian (99.94%) juga menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah digunakan sebelumnya.

Tabel 4. Hasil Evaluasi Model

Metrik	Nilai
Test Loss	0.0016
Accuracy	99.94%
Precision	99.94%
Recall	99.94%
F1-Score	99.94%
AUC-ROC	1.0000

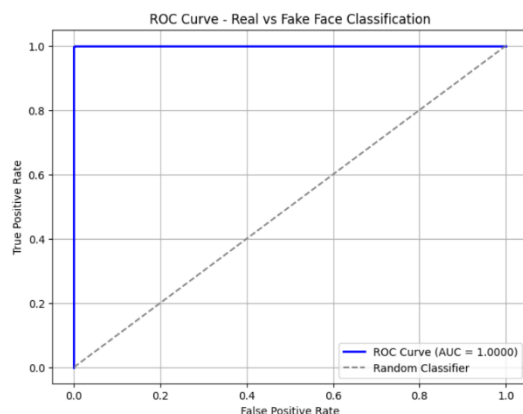
Berdasarkan *classification report*, model memperoleh nilai precision, recall, dan F1-score sebesar 1.00 untuk kedua kelas, yaitu Real dan Fake. Hasil tersebut menunjukkan bahwa model mampu mengidentifikasi kedua kelas secara sangat baik dengan tingkat kesalahan yang sangat rendah.

Selain itu, kurva ROC menghasilkan nilai AUC sebesar 1.0000, yang menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat tinggi dalam membedakan citra wajah asli dan citra hasil manipulasi AI. Nilai AUC yang mendekati 1.0 mengindikasikan bahwa model mampu mempertahankan tingkat *True Positive Rate* yang tinggi dengan *False Positive Rate* yang sangat rendah. Hasil ini konsisten dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh pada tahap evaluasi.



Gambar 5. Confusion Matrix Vision Transformer

Selain itu, hasil evaluasi menggunakan ROC Curve menghasilkan nilai Area Under Curve (AUC) sebesar 1.0000. Nilai tersebut menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat tinggi dalam membedakan citra wajah asli dan citra hasil manipulasi AI. Nilai AUC yang mendekati 1 mengindikasikan bahwa model mampu mempertahankan *True Positive Rate* yang tinggi dengan *False Positive Rate* yang sangat rendah.



Gambar 6. Kurva ROC Vision Transformer

Secara keseluruhan, hasil penelitian menunjukkan bahwa model Vision Transformer mampu mengklasifikasikan citra wajah asli (*Real*) dan citra

hasil manipulasi AI (*Fake*) dengan tingkat akurasi yang sangat tinggi. Nilai *accuracy* sebesar 99.94%, *precision*, *recall*, dan *F1-score* sebesar 1.00, serta nilai AUC sebesar 1.0000 menunjukkan bahwa model memiliki kemampuan klasifikasi dan generalisasi yang sangat baik. Hasil ini menunjukkan bahwa mekanisme *self-attention* pada Vision Transformer efektif dalam menangkap representasi global citra sehingga berpotensi diterapkan sebagai metode deteksi citra hasil manipulasi AI pada berbagai aplikasi keamanan digital.

4. CONCLUSION

Penelitian ini berhasil mengimplementasikan Vision Transformer (ViT) untuk mengklasifikasikan citra wajah asli (*Real*) dan citra wajah hasil manipulasi AI (*Fake*). Model dilatih menggunakan 161.972 citra wajah yang terdiri atas 81.000 citra Real dan 80.972 citra Fake, dengan pembagian data menjadi 129.576 data pelatihan, 16.198 data validasi, dan 16.198 data pengujian. Hasil pengujian menunjukkan bahwa model memperoleh *accuracy* sebesar 99.94%, *precision* sebesar 1.00, *recall* sebesar 1.00, *F1-score* sebesar 1.00, serta AUC-ROC sebesar 1.0000. Hasil tersebut menunjukkan bahwa Vision Transformer memiliki kemampuan yang sangat baik dalam membedakan citra wajah asli dan citra hasil manipulasi AI.

Berdasarkan hasil penelitian, Vision Transformer terbukti mampu memanfaatkan mekanisme *self-attention* untuk menangkap hubungan global antarbagian citra sehingga menghasilkan performa klasifikasi yang tinggi. Penelitian selanjutnya dapat mengembangkan model dengan menggunakan dataset yang lebih beragam, mengevaluasi berbagai varian Vision Transformer, serta membandingkan performanya dengan arsitektur *deep learning* lainnya untuk memperoleh model deteksi citra manipulasi AI yang lebih andal dan mampu melakukan generalisasi dengan lebih baik.

ACKNOWLEDGEMENTS

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika Universitas Nusa Putra atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini. Penulis juga mengucapkan terima kasih kepada seluruh pihak yang telah memberikan bantuan serta kepada penyedia dataset publik yang digunakan dalam penelitian ini.

REFERENCES

- [1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.

- [2] H. Sharma, "Detection of AI-Generated Images Using Deep Learning."
- [3] F. Badri, M. Taqijuddin Alawiy, and E. M. Yuniarno, "DEEP LEARNING ARCHITECTURE BASED ON CONVOLUTIONAL NEURAL NETWORK (CNN) ON ANIMAL IMAGE CLASSIFICATION," vol. 12, no. 2, 2023.
- [4] F. Folorunsho and B. F. Boamah, "DEEPFAKE TECHNOLOGY AND ITS IMPACT: ETHICAL CONSIDERATIONS, SOCIETAL DISRUPTIONS, AND SECURITY THREATS IN AI-GENERATED MEDIA," *INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY AND MANAGEMENT INFORMATION SYSTEMS*, vol. 16, no. 1, pp. 1060–1080, Feb. 2025, doi: 10.34218/ijitmis_16_01_076.
- [5] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. Law Rev.*, vol. 107, no. 6, pp. 1753–1820, 2019, doi: 10.15779/Z38RV0D15J.
- [6] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, "On the Detection of Digital Face Manipulation."
- [7] Hao Lin Wenmin Huang Weiqi Luo Wei Lu, "DeepFake detection with multi-scale convolution and vision transformer," *Digit. Signal Process.*, vol. 134, 2023.
- [8] D. Das, M. Yahan, T. Zaman, and M. R. Bayesh, "Edge-Enhanced Vision Transformer Framework for Accurate AI-Generated Image Detection A R T I C L E I N F O." [Online]. Available: <https://ssrn.com/abstract=5556795>
- [9] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2010.11929>
- [10] B. Ghita, I. Kuzminykh, A. Usama, T. Bakhshi, and J. Marchang, "Deepfake Image Detection using Vision Transformer Models," IEEE. [Online]. Available: <http://shura.shu.ac.uk/information.html>
- [11] E. M. Dharma, N. M. S. Iswari, and I. P. R. A. Arimbawa, "DeepFake Image Detection Using Convolutional Neural Network with EfficientNet Architecture," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 4, pp. 1829–1838, Oct. 2025, doi: 10.70609/g-tech.v9i4.7727.
- [12] M. S. A. Aria, C. Slamet, and M. D. Firdaus, "Klasifikasi Fake dan Real Menggunakan Vision Transformer dan EfficientNet-B0 pada Gambar Asli dan Generatif AI," *SMATIKA JURNAL*, vol. 15, no. 01, pp. 179–192, Jun. 2025, doi: 10.32664/smatika.v15i01.1531.
- [13] H. Rumaepa, "Evaluasi Kinerja CNN dan Vision Transformer pada Klasifikasi Citra Resolusi Tinggi Berbasis Deep Learning," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 9, no. 2, pp. 372–379, Oct. 2025, doi: 10.46880/jmika.Vol9No2.pp372-379.
- [14] L. El Mouna, M. Cheikh Tourad, M. Farouk Nanne, H. Silkan, and Y. Hanyf, "AI-Generated Fake Image Detection using Pre-trained CNN Models."
- [15] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0197-0.