JNATIA Volume 3, Nomor 4, Agustus 2025 p-ISSN: 2986-3929
Jurnal Nasional Teknologi Informasi dan Aplikasinya e-ISSN: 3032-1948

Klasifikasi Kematangan Tomat pada Citra Digital Menggunakan DeiT (Data-efficient Image Transformer)

I Gede Made Widi Anditya[a1], Gst. Ayu Vida Mastrika [a2]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]anditya.2308561111@student.unud.ac.id
[2] vida@unud.ac.id

Abstract

This study addresses the critical need for accurate and efficient tomato ripeness classification in agriculture and agribusiness, aiming to overcome the limitations of subjective manual methods. Leveraging advances in Computer Vision, this study implements a Data-efficient Image Transformer (DeiT) model for automatic classification of digital tomato images. DeiT, a Transformer-based architecture developed by Facebook AI Research, was chosen for its superior performance on small to medium-sized datasets, leveraging knowledge distillation. The model was trained using the Kaggle dataset, instrumented to enhance visual diversity, to classify tomatoes into "ripe" and "unripe" categories. Evaluation was performed using standard classification metrics including accuracy, F1-Score, and confusion matrix. The model demonstrated high performance, achieving an overall accuracy of 0.96 on the test dataset.

Keywords: *Classification, Deep Learning, Tomato Ripeness, DeiT*

1. Pendahuluan

Klasifikasi tingkat kematangan buah tomat adalah hal yang sangat penting dalam sektor pertanian serta bisnis agribisnis saat ini. Menentukan tingkat kematangan yang akurat sangat berguna bagi petani dan pemasok dalam tahap panen, pengemasan, hingga pendistribusian, sehingga mutu dan harga jual produk dapat tetap terjaga. Meskipun demikian, metode klasifikasi yang dilakukan secara manual masih sering dipakai di lapangan, yang cenderung subjektif, tidak selalu konsisten, dan kurang efisien pada skala yang lebih besar[1]. Seiring dengan kemajuan teknologi kecerdasan buatan *(Artificial Intelligence),* terutama dalam bidang *Vision Computer*, metode *Deep Learning* sering diterapkan untuk mengotomatisasi klasifikasi gambar buah dan sayuran. Salah satu metode baru yang menunjukkan keunggulan dalam tugas klasifikasi yaitu *Vision Transformer (ViT)* [2]. *ViT* menangani citra seperti halnya teks dalam NLP dengan memecahnya menjadi bagian kecil dan menggunakan self-attention untuk mengenali pola secara keseluruhan. Akan tetapi, model *ViT* yang biasa membutuhkan dataset yang besar agar dapat berfungsi dengan baik. Adapun solusi dari metode *VIT* yang membutuhkan dataset yang besar ini, dari pihak Facebook AI mengembangkan model Data-efficient Image Transformer (DeiT) yang mampu belajar secara efisien dari dataset berukuran kecil hingga menengah melalui teknik distilasi pengetahuan (knowledge distillation) [3]. Pengembangan selanjutnya seperti DeiT III juga menunjukkan kemajuan dalam kinerja dan efisiensi [4][5]. Beberapa studi juga telah mengaplikasikan metode Transformer dalam klasifikasi kematangan buah, termasuk pisang [6], tebu [7], dan tomat [8][9]. Pada Penelitian ini, penulis memiliki tujuan untuk menerapkan model DeiT sebagai cara untuk mengklasifikasikan kematangan tomat pada citra digital menjadi dua kategori label: "matang" dan "tidak matang". Model tersebut dilatih dengan dataset yang didapatkan pada website Kaggle dengan nama "Tomato Maturity Detection and Quality Grading" yang dimana telah dilakukan proses augmentasi, kemudian dievaluasi menggunakan metrik klasifikasi standar.

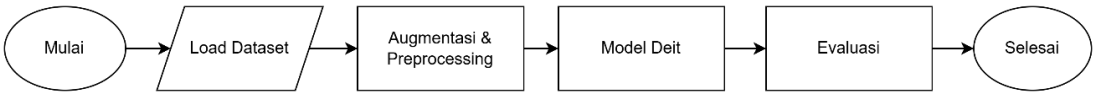
2. Metode Penelitian

2.1. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle dengan judul *Tomato Maturity Detection and Quality Grading* yang dapat diakses melalui tautan https://www.kaggle.com/datasets/sujaykapadnis/tomato-maturity-detection-and-quality-grading. Total dari dataset yang digunakan yaitu berjumlah total 500 gambar per kelas nya. Data yang digunakan dikategorikan ke dalam dua kelas, yaitu "matang" dan "tidak_matang". Gambar-gambar tersebut kemudian dipisahkan ke dalam tiga subset utama dengan rasio data latih (train) sejumlah 350 gambar, validasi (validation) dengan 75 gambar, dan pengujian (test) dengan 75 gambar pada setiap kelasnya.

2.2. Alur Sistem

Dalam mengklasifikasi kematangan pada buah tomat ini, terdapat beberapa tahapan yang perlu dilakukan seperti tahap augmentasi dataset, *preprocessing,* pelatihan model Deit, dan evaluasi menggunakan *confusion matrix*. Alur sistem dapat dilihat pada gambar 1.



Gambar 1. Alur Sistem

Gambar 1 menggambarkan alur sistem untuk mengklasifikasikan kematangan tomat dengan menggunakan model DeiT. Setelah melakukan preprocessing dan augmentasi data, gambar tersebut masuk ke proses pelatihan model *(fine-tuning).* Di fase ini, hanya *layer* klasifikasi model yang dilatih akan kembali, sementara *backbone transformer* tetap tidak berubah. Pelatihan dilakukan dalam beberapa epoch secara berulang (loop), dengan evaluasi dilakukan terhadap data validasi di setiap epoch. Apabila akurasi validasi tidak meningkat selama tiga epoch berturut-turut, pelatihan akan dihentikan secara otomatis dengan menggunakan pendekatan *early stopping*. Model yang memiliki akurasi validasi tertinggi disimpan untuk digunakan dalam evaluasi dan prediksi akhir.

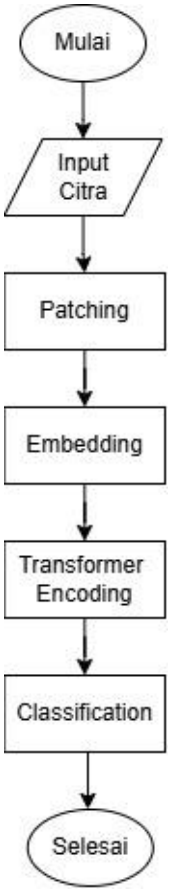
2.3. Augmentasi dan Preprocessing

Tahap augmentasi dan *preprocessing* bertujuan untuk memperbanyak variasi visual data pelatihan dan menyesuaikan ukuran serta format gambar agar sesuai dengan arsitektur input model DeiT. Augmentasi dilakukan secara khusus pada data pelatihan dengan menggunakan berbagai metode transformasi yang bersifat acak, seperti rotasi acak hingga 90 derajat, *Flipping horizontal* dan *vertical*, *Color jitter* (kecerahan, kontras, saturasi), Transformasi perspektif, *Gaussian blur*, dan *Random erasing.* Yang dimana pada data pelatihan yang awalnya hanya berjumlah 500 gambar per kelasnya, setelah di augmentasi jumlah data pelatihan menjadi berjumlah 700 gambar per kelasnya. Yang Dimana ini bermaksud agar dataset yang dipakai lebih beragam dan mencegah *overvitting.* Setelah proses augmentasi, semua gambar akan diubah ukurannya menjadi 224×224 piksel dan dikonversi ke dalam format tensor agar dapat diproses oleh model DeiT. Sedangkan data validasi dan uji hanya melalui proses *resizing* dan konversi tensor tanpa augmentasi tambahan agar pada tahap evaluasi menjadi lebih konsisten.

2.4. Klasifikasi DeiT

Dalam penelitian ini, Klasifikasi tingkat kematangan tomat dilakukan dengan memanfaatkan model *DeiT (Data-efficient Image Transformer)* yang dirancang untuk efisiensi data dan kinerja yang tinggi meskipun dengan data pelatihan yang sedikit. Model ini menerima *input* gambar berukuran 224×224 piksel, kemudian membagi gambar tersebut menjadi *patch* 16×16, mengonversi setiap patch menjadi *vektor embedding*, dan memprosesnya melalui lapisan

encoder transformer. Pada akhir proses, representasi dari *classification token* digunakan untuk memprediksi dua kategori yaitu matang dan tidak_matang. Tahapan dari klasifikasi dengan model DeiT bisa dilihat pada gambar 2.



Gambar 2. Tahapan Model DeiT

Model dilatih melalui metode *fine-tuning*, di mana hanya lapisan klasifikasi yang disempurnakan menggunakan *CrossEntropyLoss* dan *optimizer AdamW*. Evaluasi dilakukan untuk setiap *epoch* dengan memanfaatkan data validasi, dan pelatihan akan berhenti secara otomatis jika tidak ada peningkatan dalam akurasi *(early stopping)*. Hasil klasifikasi dihitung dengan fungsi softmax sebagai berikut :

$$P(y = c|x) = \frac{e^{z_c}}{\sum_{k=1}^{K} e^{z_k}} \quad (1)$$

Dimana $P(y = c|x)$ adalah probabilitas *input* x diklasifikasikan ke kelas c, dan z_k adalah *logit output* untuk kelas ke- k

2.5. Evaluasi

Pada tahap ini dilakukan evaluasi untuk mengukur kinerja dari model DeiT yang dibuat dalam melakukan klasifikasi citra digital. Pada penelitian ini akan digunakan confusion matrix dalam pengukuran peforma model tersebut sehingga dapat diperoleh nilai *accuracy, precision, recall* dan *F1-Score*.

Tabel 1. *Confusion Matrix*

	Positif	Negatif
Positif	TP	FN
Negatif	FN	TP

Keterangan:

TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

Berikut merupakan rumus untuk mencari nilai *accuracy, precision, recall dan F1-Score:*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = 2.\frac{Precision\ .\ Recall}{Precision+Recall} \quad (5)$$

3. Hasil dan Pembahasan

3.1. Hasil Augmentasi

Dataset awal yang terdiri dari sekitar 700 foto tomat yang terbagi rata menjadi dua kategori yaitu matang dan tidak matang. Setiap foto dalam subset data latih mengalami augmentasi sebanyak dua kali dengan menggabungkan transformasi seperti rotasi, *flipping*, perubahan warna, perspektif, buram, dan penghapusan. Setelah dilakukannya augmentasi, jumlah data pelatihan meningkat hingga tiga kali lipat, menjadi sekitar 700 gambar tomat per 2 kelasnya yang dimana total gambar pada data pelatihan berjumlah 1.400 gambar tomat. Semua gambar hasil augmentasi disimpan dalam folder baru dengan struktur direktori yang mengikuti pembagian kategori. Dengan penambahan tersebut, model memperoleh representasi visual yang lebih beragam dan dapat belajar dengan lebih baik terhadap perbedaan tampilan tomat pada berbagai kondisi pencahayaan dan orientasi.

3.2. Hasil Pelatihan Model DeiT

Proses pelatihan model Data-efficient Image Transformer (DeiT) dilakukan dalam 10 *epoch*, dengan pengawasan yang ketat terhadap tingkat akurasi dan *loss* pada data pelatihan *(Train Acc dan Train Loss)* serta data validasi *(Val Acc dan Val Loss).* Pada epoch pertama, model menunjukkan akurasi pelatihan sebesar 0,6421 dan akurasi validasi sebesar 0,7333. Sementara itu, *loss* pelatihan tercatat sebesar 27,7777 dan kerugian validasi sebesar 2,8944. Seiring dengan berjalannya epoch, terdapat kemajuan yang signifikan dalam akurasi dan penurunan dalam *loss*. Akurasi validasi model terus mengalami peningkatan dan mencapai 0,9867 pada *epoch* keenam, kemudian menunjukkan kestabilan pada 0,9933 mulai dari *epoch* ke-7 hingga *epoch* ke-10. Untuk detail dari hasil pelatihan model bisa dilihat pada gambar 3.

```
Epoch 1/10 | Train Acc: 0.6421, Val Acc: 0.7333, Train Loss: 27.7777, Val Loss: 2.8944
Epoch 2/10 | Train Acc: 0.8379, Val Acc: 0.8600, Train Loss: 22.1855, Val Loss: 2.3942
Epoch 3/10 | Train Acc: 0.9236, Val Acc: 0.9533, Train Loss: 18.2440, Val Loss: 2.0145
Epoch 4/10 | Train Acc: 0.9500, Val Acc: 0.9733, Train Loss: 15.3807, Val Loss: 1.7303
Epoch 5/10 | Train Acc: 0.9679, Val Acc: 0.9733, Train Loss: 13.2071, Val Loss: 1.5063
Epoch 6/10 | Train Acc: 0.9764, Val Acc: 0.9867, Train Loss: 11.5915, Val Loss: 1.3311
Epoch 7/10 | Train Acc: 0.9814, Val Acc: 0.9933, Train Loss: 10.3005, Val Loss: 1.1913
Epoch 8/10 | Train Acc: 0.9821, Val Acc: 0.9933, Train Loss: 9.2707, Val Loss: 1.0762
Epoch 9/10 | Train Acc: 0.9829, Val Acc: 0.9933, Train Loss: 8.4323, Val Loss: 0.9807
Epoch 10/10 | Train Acc: 0.9857, Val Acc: 0.9933, Train Loss: 7.7264, Val Loss: 0.9011
Early stopping!
```

Gambar 3. Pelatihan Model

Loss juga tampak stabil sepanjang proses pelatihan. Kerugian dari pelatihan berkurang dari 27. 7777 menjadi 7. 7264, dan kerugian validasi dari 2. 8944 menjadi 0. 9011 pada *epoch* terakhir. Walaupun pelatihan dari model ini direncanakan berlangsung selama 10 epoch, namun proses pelatihannya dihentikan lebih awal *(early stopping)* setelah epoch ke-10 dikarenakan akurasi validasi tidak menunjukkan peningkatan yang terlalu signifikan secara berurutan. Hal ini menunjukkan bahwa model telah mencapai kinerja terbaiknya dan mencegah terjadinya overfitting.

3.3. Hasil Evaluasi

Evaluasi pada model dilaksanakan menggunakan data uji, yang terdiri dari 150 sampel (masing-masing 75 untuk kategori "matang" dan "tidak_matang"), guna mengukur seberapa baik model dapat menggeneralisasi terhadap data baru yang belum pernah dilihat selama proses pelatihan. Hasil dari evaluasi menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi, yaitu 0. 96, yang berarti model ini dapat memprediksi kematangan tomat dengan tingkat keberhasilan sebesar 96% pada data pengujian. Untuk detail dari classification report bisa dilihat pada gambar 4.

	precision	recall	f1-score	support
matang	0.99	0.93	0.96	75
tidak_matang	0.94	0.99	0.96	75
accuracy			0.96	150
macro avg	0.96	0.96	0.96	150
weighted avg	0.96	0.96	0.96	150

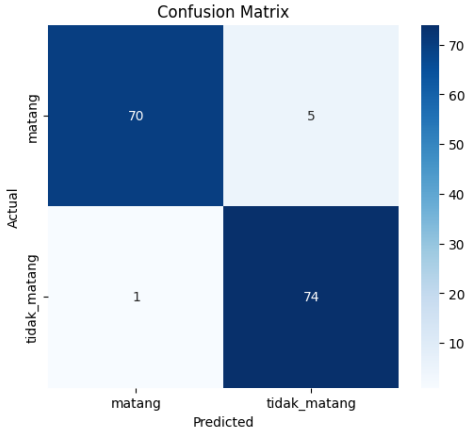
Gambar 4. Classification Report

Untuk kategori "matang", model memperoleh tingkat presisi 0. 99, recall 0. 93, dan F1-Score 0. 96. Hal ini menunjukkan bahwa hampir semua prediksi tentang tomat "matang" yang dilakukan oleh model adalah akurat (memiliki presisi tinggi), meskipun terdapat beberapa kasus tomat "matang" yang tidak teridentifikasi (recall sedikit lebih rendah). Sementara itu, untuk kelas "tidak_matang", model menunjukkan akurasi 0,94, sensitivitas 0,99, dan skor F1 0,96. Tingkat recall yang sangat tinggi dalam kategori ini menunjukkan bahwa model tersebut sangat efisien dalam mengenali sebagian besar tomat yang "tidak matang". Nilai *macro avg* dan *weighted avg* yang sama sama mencapai 0,96 untuk *precision, recall, dan F1-Score* menunjukkan bahwa model memiliki kinerja yang konsisten dan seimbang antara kedua kelas atau labelnya.

3.4. Visualisasi Confusion Matrix

Pada Confusion Matriks ini memberikan gambaran visual dan angka mengenai kinerja model klasifikasi yang menunjukkan jumlah prediksi yang benar *(True Positives, True Negatives)* dan salah *(False Positives, False Negatives)* untuk setiap kategori. Untuk *Confusion Matriks* bisa dilihat pada gambar 5.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 5. Confusion Matrix

Berdasarkan *Confusion Matriks* pada gambar 5 diatas, model berhasil mengklasifikasikan dengan benar 70 dari 75 tomat yang sebenarnya berkategori "matang" sebagai "matang" (*True Positives*), dan 74 dari 75 tomat yang sebenarnya berkategori "tidak matang" sebagai "tidak matang" (*True Negatives* untuk kategori "matang" atau *True Positives* untuk kategori "tidak matang"). Terlihat bahwa model memiliki kecenderungan yang sangat rendah untuk melakukan kesalahan False Positive pada kelas "matang", dengan hanya satu kasus tomat "tidak_matang" yang salah dikategorikan sebagai "matang". Namun, terdapat lima kejadian False Negative pada kategori "matang", di mana tomat yang sebenarnya "matang" keliru diprediksi sebagai "tidak matang". Meskipun demikian, performa dari keseluruhan model sudah sangat kuat, dengan kemampuan yang terpresisi dalam membedakan kedua kategori kematangan tomat, terbukti dari akurasi yang tinggi dan nilai *F1-Score* yang seimbang untuk kedua kelas.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, penelitian ini berhasil menggunakan model Data-efficient Image Transformer (DeiT) untuk mengklasifikasikan tingkat kematangan buah tomat dari citra digital. Model DeiT menunjukkan kinerja yang sangat baik dengan tingkat akurasi keseluruhan sebesar 0,96 pada data pengujian, serta nilai *Precision, Recall* dan *F1-Score* yang tinggi dan seimbang untuk kedua kategori ("matang" dan "tidak_matang") mengindikasikan kemampuan prediksi yang akurat dan konsisten. Pada *Confusion Matriks* menunjukkan jumlah kesalahan klasifikasi yang dilakukan oleh model sangatlah rendah, khususnya pada kelas "tidak matang". Dengan adanya penelitian ini diharapkan dapat membantu pada sektor pertanian atau yang lainnya. Adapun saran dari pengembangan dari system kedepannya, penelitian dapat difokuskan pada penggunaan dataset yang lebih beragam, identifikasi tingkat kematangan yang lebih rinci, serta integrasi sistem untuk aplikasi secara *real-time*.

Daftar Pustaka

[1] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, “Training data-efficient image transformers & distillation through attention,” Jan. 2021, [Online]. Available: http://arxiv.org/abs/2012.12877
[2] H. Touvron, M. Cord, and H. Jégou, “DeiT III: Revenge of the ViT.”
[3] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, and L. Wang, “Vision Transformers for Image Classification: A Comparative Survey,” Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/technologies13010032.
[4] A. Khan *et al.*, “Tomato maturity recognition with convolutional transformers,” *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-50129-w.
[5] L. Papa, P. Russo, I. Amerini, and L. Zhou, “A survey on efficient vision transformers: algorithms, techniques, and performance benchmarking,” Mar. 2024, doi: 10.1109/TPAMI.2024.3392941.

[6] A. Pangestu, B. Purnama, and R. Risnandar, “Vision Transformer untuk Klasifikasi Kematangan Pisang,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 75–84, Feb. 2024, doi: 10.25126/jtiik.20241117389.
[7] İ. Paçal and İ. Kunduracıoğlu, “Data-Efficient Vision Transformer Models for Robust Classification of Sugarcane,” *Journal of Soft Computing and Decision Analytics*, vol. 2, no. 1, pp. 258–271, Jun. 2024, doi: 10.31181/jscda21202446.
[8] P. Li, J. Zheng, P. Li, H. Long, M. Li, and L. Gao, “Tomato Maturity Detection and Counting Model Based on MHSA-YOLOv8,” *Sensors*, vol. 23, no. 15, Aug. 2023, doi: 10.3390/s23156701.
[9] M. Nahiduzzaman *et al.*, “Deep learning-based real-time detection and classification of tomato ripeness stages using YOLOv8 on raspberry Pi,” *Engineering Research Express*, vol. 7, no. 1, Mar. 2025, doi: 10.1088/2631-8695/ada720.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Halaman ini sengaja dibiarkan kosong