

## ANALISIS KOMPARATIF KLASIFIKASI BUAH BERBASIS CITRA: K-MEANS + RANDOM FOREST VS CONVOLUTIONAL NEURAL NETWORK (CNN)

Octaviana Sholikhah<sup>1</sup>, Khoirul Anam<sup>2</sup>, Edi Faisal<sup>3</sup>

[octaviana0056@gmail.com](mailto:octaviana0056@gmail.com)<sup>1</sup>, [akunnyakhoirulanam@gmail.com](mailto:akunnyakhoirulanam@gmail.com)<sup>2</sup>, [faisal@dsn.dinus.ac.id](mailto:faisal@dsn.dinus.ac.id)<sup>3</sup>

Universitas Dian Nuswantoro

### ABSTRAK

Klasifikasi buah berbasis citra merupakan salah satu penerapan teknologi pengolahan citra digital yang banyak digunakan dalam bidang pertanian, industri makanan, dan sistem otomatisasi. Penelitian ini membandingkan performa dua pendekatan klasifikasi citra, yaitu kombinasi segmentasi K-Means dan Random Forest, serta metode Convolutional Neural Network (CNN). Pada pendekatan pertama, K-Means digunakan untuk segmentasi citra dan ekstraksi fitur seperti bentuk dan warna, yang kemudian diklasifikasikan menggunakan Random Forest. Sementara itu, CNN digunakan sebagai metode end-to-end berbasis deep learning yang secara otomatis mempelajari fitur dari citra tanpa proses ekstraksi manual. Eksperimen dilakukan menggunakan dataset yang dikurasi dari platform Kaggle, terdiri dari empat kelas buah utama: Anggur (988 citra), Apel (991 citra), Jeruk (656 citra), dan Tomat (992 citra). Hasil pengujian menunjukkan bahwa model K-Means + Random Forest mencapai akurasi sebesar 99,72% pada data uji, sementara model CNN memperoleh akurasi sebesar 84,62%. Performa CNN yang lebih rendah terutama diduga disebabkan oleh overfitting yang signifikan dan kebutuhan tuning parameter yang lebih kompleks. Dengan demikian, penelitian ini menunjukkan bahwa meskipun CNN memiliki potensi besar dalam menangani variasi citra yang kompleks, metode tradisional seperti K-Means + Random Forest masih sangat efektif dan bahkan dapat memberikan performa superior dalam kondisi pencitraan yang terkontrol. Hasil ini dapat menjadi acuan dalam pemilihan pendekatan klasifikasi yang tepat berdasarkan kebutuhan aplikasi dan sumber daya yang tersedia.

**Kata Kunci:** Klasifikasi Citra, K-Means, Random Forest, Convolutional Neural Network, Pengolahan Citra Digital, Segmentasi Citra, Deep Learning.

### PENDAHULUAN

Kemajuan teknologi pengolahan citra digital dan kecerdasan buatan telah memberikan kontribusi besar dalam otomatisasi berbagai bidang, khususnya di sektor pertanian dan industri pengolahan makanan. Salah satu aplikasi penting dari teknologi ini adalah sistem klasifikasi buah berbasis citra, yang bertujuan untuk membantu proses sortasi, pengawasan mutu, dan pengenalan jenis buah secara otomatis[4]. Penerapan teknologi ini dapat meningkatkan efisiensi dan akurasi dibandingkan dengan metode manual yang memerlukan waktu dan tenaga lebih banyak.

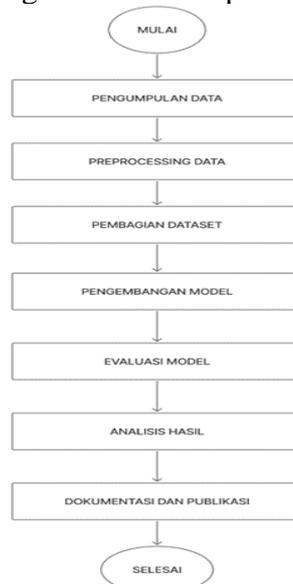
Dalam pengembangan sistem klasifikasi citra buah, terdapat dua pendekatan utama yang banyak digunakan. Pendekatan pertama adalah metode tradisional yang mengkombinasikan algoritma segmentasi citra seperti K-Means dengan algoritma klasifikasi seperti Random Forest[9],[11]. Metode ini memanfaatkan ekstraksi fitur manual berdasarkan karakteristik citra, seperti warna, bentuk, dan tekstur. Model yang dihasilkan cenderung lebih mudah dipahami dan diinterpretasikan[2]. Namun demikian, pendekatan ini memiliki keterbatasan dalam menangani variasi kondisi pencahayaan dan latar belakang yang sering ditemui dalam pengambilan citra buah secara nyata. Pendekatan kedua adalah metode berbasis deep learning, khususnya Convolutional Neural Network (CNN), yang dapat melakukan ekstraksi fitur secara otomatis dari citra input

tanpa perlu intervensi manual[10]. CNN telah terbukti unggul dalam berbagai tugas pengenalan citra karena kemampuannya mengenali pola visual yang kompleks serta adaptif terhadap variasi visual[1]. Meski demikian, metode ini memerlukan dataset yang besar dan sumber daya komputasi yang lebih tinggi. Permasalahan yang sering dihadapi adalah kurangnya analisis komprehensif yang membandingkan efektivitas kedua pendekatan tersebut, terutama pada dataset dengan ukuran dan variasi terbatas sebagaimana umum ditemukan dalam aplikasi pertanian skala kecil hingga menengah[4]. Selain itu, tidak semua aplikasi menuntut penggunaan metode deep learning yang kompleks dan mahal.

Oleh karena itu, penting untuk mengevaluasi secara langsung performa dan kelebihan masing-masing pendekatan agar dapat memberikan panduan yang tepat bagi pengembang sistem klasifikasi citra buah. Penelitian ini bertujuan untuk melakukan analisis komparatif antara metode segmentasi K-Means yang digabungkan dengan algoritma Random Forest dan Convolutional Neural Network (CNN) dalam mengklasifikasikan berbagai jenis buah berdasarkan citra digital. Evaluasi dilakukan berdasarkan akurasi klasifikasi, waktu proses, serta ketahanan terhadap variasi kondisi citra. Selain itu, penelitian ini juga menyediakan dataset dan kode implementasi yang dapat dimanfaatkan untuk pengembangan dan penelitian lebih lanjut. Dengan hasil penelitian ini, diharapkan dapat memberikan kontribusi dalam pemilihan metode klasifikasi citra yang tepat sesuai kebutuhan dan sumber daya yang tersedia di bidang pertanian dan industri makanan.

## METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan utama yang ditunjukkan dalam diagram alir penelitian (Gambar 1). Setiap tahapan dijelaskan secara sistematis, mulai dari pengumpulan data, pra-pemrosesan citra, pembagian dataset, pelatihan model, hingga evaluasi dan analisis hasil. Diagram alir ini digunakan untuk memperjelas alur proses klasifikasi dan pendekatan yang digunakan dalam penelitian.



Gambar 1. Tahapan Penelitian

Penelitian ini terdiri dari beberapa tahapan utama yang disusun secara sistematis untuk mencapai tujuan penelitian. Penjelasan terkait tahapan penelitian pada gambar 1, dapat dijelaskan sebagai berikut :

## 1. Pengumpulan Data

Tahap awal dalam penelitian ini dimulai dengan proses pengumpulan dataset citra buah. Dataset ini dikurasi dari platform Kaggle (berdasarkan dataset Fruits-360), terdiri dari empat kelas utama: Anggur (988 citra), Apel (991 citra), Jeruk (656 citra), dan Tomat (992 citra)[5]. Dataset ini mencakup kumpulan gambar yang telah dikategorikan dan diberi label. Citra-citra dalam dataset ini memiliki anotasi yang jelas untuk keperluan pelatihan dan evaluasi model klasifikasi, memastikan akurasi dan kualitas data yang digunakan dalam penelitian.

## 2. Pra-pemrosesan Data (Preprocessing Data)

Pada tahap pra-pemrosesan data, dilakukan beberapa langkah penting untuk memastikan kualitas dan konsistensi input. Pertama, seluruh gambar diubah ukurannya (resize) menjadi 100x100 piksel untuk memastikan dimensi yang seragam. Selanjutnya, untuk metode Convolutional Neural Network (CNN), dilakukan augmentasi data seperti rotasi (hingga 20 derajat), horizontal flipping, dan zoom (hingga 20%) guna meningkatkan variasi gambar pelatihan dan mengurangi risiko overfitting. Selain itu, nilai piksel dinormalisasi ke rentang [0,1] dengan membagi setiap nilai piksel dengan 255. Sementara itu, pada pendekatan K-Means + Random Forest, dilakukan ekstraksi fitur citra yang meliputi fitur bentuk (seperti metric dan eccentricity) serta histogram warna pada kanal hijau dan biru. Tahapan ini bertujuan untuk menghasilkan data fitur yang representatif dan siap digunakan pada proses pelatihan model.

## 3. Pembagian Dataset

Setelah dataset berhasil dikumpulkan dan diproses, langkah selanjutnya adalah membagi dataset tersebut ke dalam subset utama. Untuk metode CNN, dataset dibagi menjadi data latih (80%), data validasi (10%), dan data uji (10%). Sementara itu, untuk metode Random Forest, dataset dibagi menjadi data latih (80%) dan data uji (20%). Pembagian ini dilakukan menggunakan teknik stratified sampling guna memastikan bahwa distribusi kelas tetap seimbang di setiap subset, sehingga model dapat belajar dan dievaluasi secara adil pada seluruh kategori buah.

## 4. Pengembangan Model

Untuk pendekatan pertama, model K-Means + Random Forest diawali dengan melakukan segmentasi citra menggunakan algoritma K-Means guna memisahkan objek buah dari latar belakang dan mengekstraksi fitur awal seperti warna dan bentuk dari gambar buah[6]. Fitur-fitur tersebut kemudian digunakan untuk melatih model Random Forest. Selama proses pelatihan, dilakukan tuning parameter seperti kedalaman pohon dan jumlah fitur yang dipertimbangkan agar performa model dapat dioptimalkan. Sementara itu, pendekatan kedua menggunakan Convolutional Neural Network (CNN). Model CNN dibangun dari awal (atau, jika Anda menggunakan transfer learning, jelaskan: "menggunakan arsitektur MobileNetV2 yang telah dilatih sebelumnya (pretrained) pada dataset ImageNet, dengan lapisan output disesuaikan untuk empat kelas buah[7]. Lapisan-lapisan awal MobileNetV2 dibekukan untuk feature extraction, sedangkan lapisan dense di akhir dilatih ulang"). Model ini terdiri dari beberapa lapisan konvolusi, pooling, dan fully connected. Proses pelatihan dilakukan selama 10 epoch menggunakan algoritma optimasi Adam dengan learning rate awal 0.001 dan fungsi loss categorical crossentropy, yang sesuai untuk klasifikasi multi-kelas. Selama pelatihan, data validasi digunakan untuk memantau performa dan mencegah overfitting, serta diterapkan teknik early stopping (jika digunakan) untuk menghentikan pelatihan jika loss validasi tidak membaik setelah beberapa epoch.

## 5. Evaluasi Model

Tahap evaluasi dilakukan dengan mengukur performa masing-masing model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score pada data uji untuk menilai kemampuan klasifikasi secara menyeluruh. Khusus untuk model Random Forest, diterapkan teknik 5-fold cross-validation guna memperoleh rata-rata akurasi dan mengevaluasi stabilitas model terhadap variasi data latih[8]. Selanjutnya, hasil dari kedua pendekatan, yaitu K-Means + Random Forest dan CNN, dibandingkan berdasarkan aspek akurasi, waktu komputasi, serta performa dalam mengklasifikasikan berbagai jenis buah.

## 6. Analisis Hasil

Tahap analisis dilakukan dengan mengkaji kelebihan dan kekurangan masing-masing metode berdasarkan hasil evaluasi yang telah diperoleh. Analisis ini mencakup interpretasi terhadap hasil klasifikasi, serta identifikasi faktor-faktor yang mempengaruhi performa model, seperti kompleksitas arsitektur, jumlah data, kualitas fitur, dan sensitivitas terhadap kondisi pencahayaan atau latar belakang citra. Pembahasan juga akan mencakup analisis overfitting yang teramati pada model CNN dan implikasinya. Dengan demikian, dapat diperoleh pemahaman yang lebih mendalam mengenai efektivitas dan efisiensi kedua pendekatan dalam konteks klasifikasi buah berbasis citra.

## 7. Dokumentasi dan Publikasi

Setelah proses pelatihan selesai, model yang telah terlatih disimpan dalam format yang dapat digunakan kembali, seperti file .h5 untuk CNN atau file model .pkl untuk Random Forest. Penyimpanan ini memungkinkan penggunaan model pada tahap implementasi tanpa perlu melatih ulang dari awal. Selain itu, untuk mendukung transparansi dan pengembangan penelitian lanjutan, dataset, kode program, dan dokumentasi hasil penelitian turut dipublikasikan. Langkah ini bertujuan agar penelitian dapat direplikasi, divalidasi, serta dikembangkan lebih lanjut oleh peneliti lain di bidang yang sama.

## HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan dua model klasifikasi buah berbasis citra, yaitu model tradisional menggunakan kombinasi segmentasi K-Means dan klasifikasi Random Forest, serta model deep learning menggunakan Convolutional Neural Network (CNN). Evaluasi dilakukan berdasarkan metrik akurasi, precision, recall, F1-score, serta analisis terhadap performa masing-masing model pada Dataset Standar (Dataset-STD) yang sama.

### 1. Kinerja Model pada Dataset Standar (Dataset-STD)

Dataset-STD merujuk pada citra buah yang dikurasi dari platform Kaggle, yang memiliki kondisi pengambilan gambar relatif terkontrol dan latar belakang seragam.



Gambar 2. Daftar Sebagian Dataset Dataset-STD

### 2. Hasil dan Pembahasan K-Means + Random Forest pada Dataset-STD

Pendekatan klasifikasi menggunakan kombinasi segmentasi K-Means dan klasifikasi Random Forest diterapkan pada Dataset-STD untuk empat kelas buah (Anggur, Apel, Jeruk, Tomat). Setelah proses segmentasi K-Means untuk memisahkan objek buah dan ekstraksi fitur bentuk serta warna, dataset fitur dibagi menjadi data latih (80%) dan data uji (20%). Total dataset terdiri dari 3.627 citra, dengan 726 citra digunakan sebagai data uji.

a. Validasi Model dan Performa Umum

Model Random Forest dilatih dan dievaluasi stabilitasnya menggunakan 5-fold cross-validation pada data latih. Hasil validasi silang menunjukkan akurasi rata-rata yang sangat tinggi, yaitu 99.72%. Ini mengindikasikan bahwa model memiliki konsistensi yang baik dalam mengenali pola pada berbagai subset data latih dan tidak cenderung overfit. Selanjutnya, model dievaluasi menggunakan data uji yang terdiri dari 726 citra (20% dari total data). Performa model pada data uji disajikan sebagai berikut:

- Akurasi pada Data Uji: 99.72%
- Jumlah Kesalahan Klasifikasi: 2 dari 726 citra uji.

Akurasi pada data uji yang mencapai 99.72% konsisten dengan hasil validasi silang, menunjukkan kemampuan generalisasi model yang sangat baik pada data yang belum pernah dilihat sebelumnya. Dua kesalahan klasifikasi yang terjadi adalah:

- Satu citra buah Apel diprediksi sebagai Tomat.
- Satu citra buah Apel lainnya diprediksi sebagai Tomat.

Kesalahan ini kemungkinan disebabkan oleh adanya kemiripan fitur warna atau bentuk antara beberapa sampel Apel dan Tomat dalam dataset, mungkin karena variasi kematangan, sudut pengambilan gambar, atau kondisi pencahayaan tertentu yang membuat kedua buah tersebut memiliki kemiripan fitur yang terekstraksi.

b. Analisis Performa Per-Kelas

Laporan klasifikasi detail pada data uji memberikan rincian performa model untuk setiap kelas buah, seperti yang disajikan dalam Tabel 1.

Tabel 1. Laporan Klasifikasi K-Means + Random Forest pada Dataset-STD

Kelas	Presisi	Recall	F1-score	Support
Anggur	1.00	1.00	1.00	198
Apel	1.00	0.99	0.99	198
Jeruk	1.00	1.00	1.00	131
Tomat	0.99	1.00	0.99	199
Accuracy			1.00	726
Macro Avg	1.00	1.00	1.00	726
Weighted Avg	1.00	1.00	1.00	726

Laporan klasifikasi menunjukkan bahwa model K-Means + Random Forest mencapai nilai precision, recall, dan F1-score yang sangat tinggi, mendekati sempurna untuk semua kelas. Untuk kelas Anggur dan Jeruk, semua metrik mencapai 1.00. Kelas Apel memiliki recall 0.99 karena dua kesalahan prediksi, dan kelas Tomat memiliki precision 0.99 karena dua prediksi salah tersebut masuk ke kelas Tomat. Secara keseluruhan, baik rata-rata makro maupun rata-rata tertimbang menunjukkan performa yang mendekati 1.00, menandakan kemampuan klasifikasi yang seimbang dan sangat baik di semua kelas.

c. Pembahasan Performa K-Means + Random Forest pada Dataset-STD

Hasil eksperimen pada Dataset-STD menunjukkan bahwa pendekatan K-Means untuk segmentasi, diikuti dengan ekstraksi fitur bentuk dan warna, dan diklasifikasikan menggunakan Random Forest, sangat efektif untuk tugas klasifikasi empat jenis buah ini. Akurasi yang sangat tinggi (99.72%) dan kesalahan klasifikasi yang minimal (2 dari 726) pada kondisi citra yang relatif terkontrol menunjukkan bahwa fitur-fitur yang diekstraksi mampu merepresentasikan perbedaan antar kelas buah dengan sangat baik. Keberhasilan ini didukung oleh stabilitas model yang ditunjukkan oleh akurasi cross-validation yang tinggi. Ini menyiratkan bahwa dengan dataset yang bersih dan fitur yang relevan, metode

pembelajaran mesin tradisional seperti Random Forest dapat mencapai performa yang sangat kompetitif, bahkan mendekati sempurna.

### 3. Hasil dan Pembahasan CNN pada Dataset-STD

Model CNN dibangun dari awal dengan arsitektur yang terdiri dari beberapa lapisan konvolusi, max pooling, dan fully connected. Proses pelatihan dilakukan selama 10 epoch dengan algoritma Adam (learning rate 0.001) dan fungsi loss categorical crossentropy. Dataset dibagi menjadi 80% data latih, 10% data validasi, dan 10% data uji.

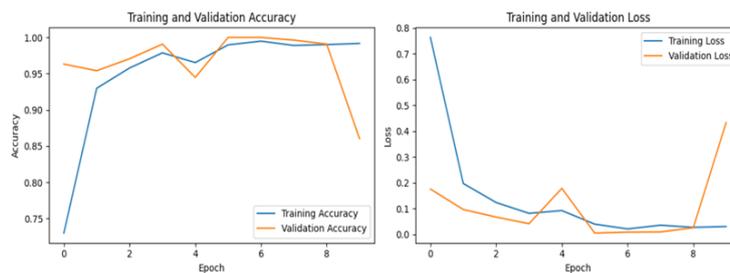
#### a. Analisis Kemajuan Pelatihan (Kurva Loss dan Akurasi)

Proses pelatihan model CNN menunjukkan peningkatan performa pada data latih, namun teramati gejala overfitting yang signifikan pada epoch akhir, yang ditunjukkan oleh divergensi antara performa pada data latih dan data validasi. Tabel 2 merangkum kinerja model CNN per epoch:

Tabel 2. Ringkasan Kinerja Model CNN per-Epoch

Epoch	Akurasi Latih	Loss Latih	Akurasi Validasi	Loss Validasi
1	57.64%	1.3290	96.32%	1.763
2	91.77%	2.288	95.40%	971
3	96.62%	1.063	97.06%	679
4	97.98%	794	99.08%	421
5	97.14%	746	94.49%	1.791
6	97.96%	620	100.00%	58
7	99.35%	257	100.00%	94
8	99.35%	268	99.63%	102
9	98.85%	264	99.08%	265
10	99.55%	210	86.03%	4.341

Dari Tabel 2, terlihat bahwa akurasi pelatihan terus meningkat hingga mencapai 99.55% pada epoch ke-10, dengan loss pelatihan yang menurun secara signifikan. Namun, akurasi validasi menunjukkan pola yang berbeda; mencapai puncaknya pada epoch 6 dan 7 dengan 100%, lalu mengalami penurunan drastis pada epoch ke-10 menjadi 86.03%. Peningkatan loss validasi yang tajam pada epoch akhir (dari 0.0058 menjadi 0.4341) semakin mengindikasikan terjadinya overfitting. Hal ini berarti model terlalu menghafal data latih dan kehilangan kemampuan untuk menggeneralisasi pola pada data yang tidak dikenalnya.



Gambar 3. Grafik Pelatihan CNN (Akurasi dan Loss)

#### b. Performa Akhir pada Data Uji

Setelah 10 epoch pelatihan, model dievaluasi pada data uji (10% dari total dataset). Hasil evaluasi akhir adalah sebagai berikut:

Akurasi Uji Akhir: 84.62%

Loss Uji Akhir: 0.5122

Akurasi uji yang sebesar 84.62% ini jauh lebih rendah dibandingkan akurasi pelatihan (99.55%) dan akurasi validasi pada epoch terbaik (100%), yang secara jelas

mengkonfirmasi adanya masalah overfitting. Meskipun CNN memiliki potensi besar, dalam implementasi ini model gagal menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya karena terlalu kompleks atau kurangnya strategi regularization yang efektif.

#### 4. Analisis Komparatif Menyeluruh dan Dampak Variasi Dataset

Berdasarkan hasil yang diperoleh, perbandingan antara pendekatan K-Means + Random Forest dan CNN menunjukkan perbedaan kinerja yang signifikan pada Dataset-STD:

Tabel 3. Hasil dari Pendekatan K-Means + Random Forest dan CNN

Metrik	<i>K-Means + Random Forest</i>	CNN
Akurasi Uji	99.72%	84.62%
Kondisi	Dataset Standar (STD)	Dataset Standar (STD)

##### a. Perbandingan Kinerja

Terlihat bahwa model K-Means + Random Forest mencapai akurasi uji yang jauh lebih tinggi (99.72%) dibandingkan model CNN (84.62%). Performa superior dari K-Means + Random Forest pada dataset ini dapat diatribusikan pada beberapa faktor:

- Kualitas Fitur Manual: Dataset-STD, dengan kondisi pencitraan yang relatif terkontrol, memungkinkan ekstraksi fitur bentuk dan warna secara manual (melalui segmentasi K-Means) yang sangat efektif dalam membedakan antar kelas buah. Fitur-fitur ini mungkin secara langsung merepresentasikan karakteristik kunci buah tanpa banyak "noise" dari latar belakang atau variasi visual lainnya.
- Kesederhanaan Model: Random Forest adalah model ensemble yang kuat namun relatif sederhana, dan dalam kondisi data yang bersih dengan fitur yang jelas, ia mampu membangun batas keputusan yang sangat akurat tanpa cenderung overfit.
- Keterbatasan Implementasi CNN: Akurasi CNN yang lebih rendah dan gejala overfitting yang jelas menunjukkan bahwa model CNN yang diimplementasikan belum dioptimalkan secara memadai untuk dataset ini. Meskipun CNN memiliki kemampuan otomatisasi ekstraksi fitur, model memerlukan tuning parameter yang tepat, strategi regularization (seperti dropout atau early stopping) yang efektif, dan mungkin volume data yang lebih besar atau diversifikasi melalui data augmentation yang lebih agresif untuk dapat menggeneralisasi dengan baik.

##### b. Implikasi Hasil

Hasil ini menunjukkan bahwa meskipun deep learning (CNN) sering dianggap sebagai solusi state-of-the-art untuk klasifikasi citra, metode tradisional seperti K-Means + Random Forest masih sangat kompetitif dan bahkan dapat mengungguli CNN, terutama pada dataset dengan karakteristik tertentu:

- Dataset Bersih dan Terkontrol: Untuk citra dengan latar belakang seragam dan fitur yang jelas, dimana ekstraksi fitur manual efektif, metode tradisional dapat memberikan hasil yang sangat baik dengan sumber daya komputasi yang lebih ringan dan interpretabilitas yang lebih tinggi.
- Pentingnya Optimasi CNN: Performa CNN yang suboptimal menggarisbawahi pentingnya hyperparameter tuning dan strategi regularization yang cermat. Model CNN yang tidak dioptimalkan dengan baik dapat dengan mudah overfit dan gagal dalam tugas generalisasi.

Ini juga mengindikasikan bahwa pemilihan metode klasifikasi harus disesuaikan dengan karakteristik dataset dan ketersediaan sumber daya. Untuk aplikasi yang tidak memerlukan robustness ekstrim terhadap variasi lingkungan dan memiliki dataset yang

terkontrol, metode tradisional bisa menjadi pilihan yang lebih efisien dan efektif.

## KESIMPULAN

Penelitian ini telah berhasil melakukan analisis komparatif antara dua pendekatan klasifikasi citra buah pada dataset standar (Dataset-STD) yang terdiri dari empat kelas buah : Anggur, Apel, Jeruk, dan Tomat. Kedua pendekatan yang dibandingkan adalah kombinasi segmentasi K-Means dengan klasifikasi Random Forest, serta Convolutional Neural Network (CNN).

Berdasarkan hasil eksperimen, dapat disimpulkan beberapa poin utama:

1. Kinerja K-Means + Random Forest: Model K-Means + Random Forest menunjukkan performa yang sangat superior pada Dataset-STD, mencapai akurasi uji sebesar 99.72%. Metode ini efektif dalam mengekstraksi fitur bentuk dan warna dari citra buah yang relatif bersih dan terkontrol, sehingga mampu mengklasifikasikan buah dengan akurasi yang mendekati sempurna. Stabilitas model juga ditunjukkan oleh hasil cross-validation yang tinggi.
2. Kinerja Convolutional Neural Network (CNN): Model CNN yang diimplementasikan pada Dataset-STD menunjukkan akurasi uji sebesar 84.62%. Meskipun akurasi pelatihan mencapai 99.55%, analisis epoch-wise pada data validasi dan uji mengindikasikan terjadinya overfitting yang signifikan. Model CNN terlalu menghafal pola pada data latih dan kehilangan kemampuan generalisasi pada data baru, yang menyebabkan akurasi yang lebih rendah dibandingkan metode tradisional dalam kasus ini.
3. Analisis Komparatif: Meskipun CNN secara umum dikenal unggul dalam tugas klasifikasi citra, hasil penelitian ini menunjukkan bahwa untuk dataset dengan kondisi citra yang terkontrol dan fitur pembeda yang jelas, metode tradisional seperti K-Means + Random Forest dapat memberikan performa yang jauh lebih baik dengan komputasi yang lebih ringan dan interpretabilitas yang lebih tinggi. Kasus ini juga menyoroti pentingnya optimasi dan regularisasi yang tepat pada model deep learning untuk mencegah overfitting dan memastikan kemampuan generalisasi.

## DAFTAR PUSTAKA

- Y. Jumaryadi, A. M. Ihsan, dan B. Priambodo, "Klasifikasi Jenis Buah-Buahan Menggunakan Citra Digital Dengan Metode Convolutional Neural Networks," *Kajian Ilmiah Informatika dan Komputer (KLIK)*, vol. 4, no. 3, pp. 1737–1746, Des. 2023, doi: 10.30865/klik.v4i3.1421.
- A. R. Maulana, H. Suhartono, dan R. R. Oktaviani, "Penggunaan Algoritma Random Forest dalam Klasifikasi Buah Segar dan Busuk," *Jurnal Teknik Informatika dan Sistem Informasi (JATISI)*, vol. 9, no. 2, pp. 621–628, 2023.
- Y. F. Wijaya and D. Hindarto, "Advancing Fruit Image Classification with State-of-the-Art Deep Learning Techniques," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, 2023. doi: 10.33395/sinkron.v8i2.13604.
- R. Riyanshu, S. S. Nagaraj, S. Ritesh, T. A. Thushar, and V. M. Aparanji, "Fruit Classification Comparison Based on CNN and YOLO," in *IOP Conference Series: Materials Science and Engineering*, vol. 1187, no. 1, 2021, Art. no. 012031. doi: 10.1088/1757-899X/1187/1/012031.
- H. Mureşan and M. Oltean, "Fruits-360: A dataset of images containing fruits and vegetables," *Kaggle*, 2018.
- Z. Zhang, "Exploring the Application of K-means Machine Learning Algorithm in Fruit Classification," *The Computer Science and Information Systems Review*, vol. 2, no. 3, pp. 45–50, 2024.

- Y. Zhang et al., "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, p. 1906, 2023. doi: 10.3390/su15031906
- S. R. Das, "A Gentle Introduction to k-fold Cross-Validation," *Machine Learning Mastery*, 2023.
- R. Premana and H. Wijaya, "Klasifikasi jenis buah mangga menggunakan metode K-Means clustering dan jaringan syaraf tiruan," *Jurnal Teknik Komputer*, vol. 8, no. 2, pp. 45–52, 2022.
- A. M. Ihsan, "Klasifikasi jenis buah-buahan menggunakan citra digital dengan metode Convolutional Neural Network (CNN)," Skripsi, Universitas Mercu Buana, Jakarta, 2023.
- A. Saputro and S. Sumantri, "Klasifikasi buah anggur menggunakan metode K-Nearest Neighbor (KNN) dan augmentasi data," *Jurnal Ilmiah Teknologi dan Komputer (JITEK)*, vol. 6, no. 1, pp. 12–17, 2023.