

MODEL HIBRIDA CNN BERBOBOT DAN MODEL ADABOOST DECISION TREE UNTUK KLASIFIKASI PENYAKIT KUBIS PADA DATASET TIDAK SEIMBANG

Atiek Iriany¹, Nabila Ayunda Sovia^{*2}, Ni Wayan Surya Wardhani³, Ani Sumarminingsih⁴

^{1,2,3,4,5}Universitas Brawijaya, Malang

Email: ¹atiekirianny@ub.ac.id, ² nabilaayundasovia@gmail.com, ³wswardhani@ub.ac.id, ⁴eni_stat@ub.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 20 Maret 2025, diterima untuk diterbitkan: 30 Oktober 2025)

Abstrak

Metode klasifikasi berbasis gambar banyak digunakan dalam bidang pertanian untuk mendeteksi penyakit pada berbagai tanaman, termasuk jenis yang rentan terhadap infeksi seperti kubis. Namun, performa optimal dalam klasifikasi gambar sangat dipengaruhi oleh jumlah dan keseimbangan data. Ketidakseimbangan data dalam klasifikasi penyakit tanaman kubis dapat menyebabkan model lebih memfokuskan perhatian pada kelas mayoritas, sehingga mengabaikan kelas minoritas, terutama pada klasifikasi multi-kelas. CNN sering digunakan dalam klasifikasi gambar, tetapi memiliki kelemahan dalam menangani data tidak seimbang karena cenderung lebih fokus pada kelas mayoritas. Untuk mengatasi permasalahan ini, dikembangkan model *Hybrid* yang mengombinasikan metode pembobotan pada CNN untuk ekstraksi fitur, model estimasi menggunakan *Decision Tree*, serta teknik *ensemble Adaboost* pada tahap klasifikasi. Pendekatan ini dirancang untuk meningkatkan kemampuan generalisasi terhadap kelas minoritas serta menghasilkan distribusi prediksi yang lebih seimbang. Hasil penelitian menunjukkan bahwa model *Hybrid* yang diusulkan mampu meningkatkan performa klasifikasi sebagaimana terlihat dari peningkatan *Weighted Average F1-score* menjadi 97%, dibandingkan model CNN tunggal dengan pembobotan yang hanya mencapai 63%. Peningkatan ini menunjukkan bahwa model *Hybrid* tidak hanya lebih efektif dalam menangani ketidakseimbangan data, tetapi juga mampu melakukan generalisasi yang lebih baik.

Kata kunci: Adaboost, CNN, Data Tidak Seimbang, Decision Tree, Kubis

HYBRID WEIGHTED CNN AND ADABOOST DECISION TREE MODEL FOR CABBAGE CLASSIFICATION ON IMBALANCED DATASETS

Abstract

Image-based classification methods are widely used in agriculture to detect diseases in various plants, including those susceptible to infections, such as cabbage. However, achieving optimal performance in image classification is highly influenced by the quantity and balance of the data. Data imbalance in cabbage disease classification can cause the model to focus more on the majority class while neglecting the minority class, especially in multi-class classification. CNN is commonly used for image classification but struggles with imbalanced data, as it tends to prioritize the majority class. To address this issue, a hybrid model has been developed by combining weighting techniques in CNN for feature extraction, a Decision Tree for estimation, and the Adaboost ensemble technique for classification. This approach is designed to enhance generalization for minority classes and produce a more balanced prediction distribution. The results of the study indicate that the proposed Hybrid model is capable of improving classification performance, as evidenced by an increase in the Weighted Average F1-score to 97%, compared to the weighted CNN model, which achieved only 63%. This improvement demonstrates that the Hybrid model is not only more effective in addressing data imbalance but also better at generalizing the data.

Keywords: Adaboost, Cabbage, CNN, Decision Tree, Imbalance Data

1. PENDAHULUAN

Dalam klasifikasi gambar, untuk mendapatkan hasil yang maksimal, diperlukan metode yang optimal serta penyesuaian algoritma model terhadap

berbagai jenis data. Salah satu metode yang telah banyak digunakan dalam beberapa tahun terakhir adalah *Convolutional Neural Network* (CNN) (El Sakka et al., 2024). CNN terbukti unggul dalam mengekstraksi fitur-fitur mendalam dari gambar,

memungkinkan identifikasi pola yang sulit terdeteksi oleh mata manusia (Reya et al., 2022). Keunggulan ini menjadikan CNN penting dan diterapkan di berbagai bidang, termasuk di sektor pertanian, sebagai deteksi dini hama dan penyakit pada tanaman kubis untuk membantu petani menanggulangi dengan tindakan yang tepat (Myna et al., 2023).

Namun, meskipun CNN sangat efektif dalam analisis gambar, model ini menghadapi keterbatasan ketika berhadapan dengan data yang tidak seimbang. CNN cenderung lebih memfokuskan perhatian pada kelas mayoritas, sehingga akurasi klasifikasi kelas minoritas seringkali menurun (Dablain et al., 2023). Penelitian oleh Dablain et al (2023) menunjukkan bahwa hal ini disebabkan oleh ketidakmampuan CNN untuk melakukan generalisasi pada data tidak seimbang dengan baik. Untuk mengoptimalkan generalisasi data, Dablain menunjukkan pembobotan yang disesuaikan dapat membuat model memberi perhatian lebih pada kelas minoritas tetapi diperlukan strategi lanjutan untuk memperkuat model.

Dalam klasifikasi pada data tidak seimbang, jenis data multi-kelas lebih meningkatkan kompleksitas model, sehingga model kesulitan mempelajari karakteristik data minoritas dan menyebabkan salah mengklasifikasi kelas (Dablain et al., 2023). Oleh karena itu diperlukan penanganan lebih, seperti penerapan *Hybrid* untuk meningkatkan kerja CNN. Misalnya, studi yang dilakukan oleh Jijo and Abdulazeez (2021), menunjukkan bahwa penggabungan kekuatan CNN dalam ekstraksi fitur dengan *Decision Tree* yang dapat membagi data secara efektif, menjadi pilihan yang efektif untuk klasifikasi multi-kelas (Sharma and Brar, 2024). Namun, model ini juga memiliki kelemahan, yaitu sensitivitas terhadap distribusi kelas yang tidak seimbang (Chabalala et al., 2023).

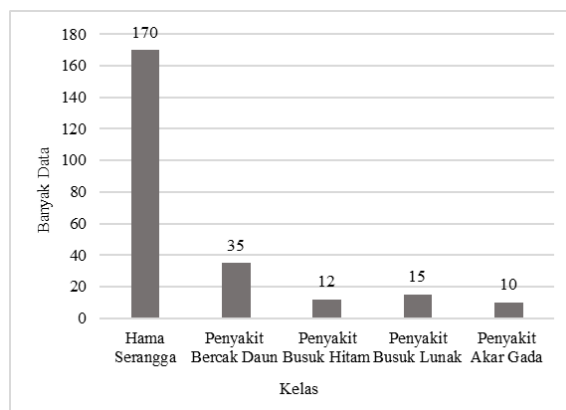
Untuk mengatasi kelemahan ini, teknik ensemble seperti *Adaboost* digunakan. Studi oleh Wu et al. (2020), menggabungkan beberapa *Decision Tree* sebagai estimator dasar dan menyatakan bahwa *Adaboost* mampu meningkatkan performa model melalui mekanisme pembobotan dan pelatihan berulang, yang secara khusus membantu model menjadi lebih adaptif terhadap kelas minoritas (Taherkhani et al., 2020). Dengan demikian, studi ini bertujuan untuk meningkatkan akurasi dalam klasifikasi data tidak seimbang, terutama dalam klasifikasi multi-kelas, dengan melalui kombinasi antara pembobotan pada CNN, kekuatan *Decision Tree* dalam klasifikasi, serta teknik ensemble *Adaboost* untuk menyesuaikan distribusi kelas.

2. DATA

Data yang digunakan pada penelitian ini adalah data primer yang didapatkan di Poncokusumo, Malang pada bulan Juli tahun 2023. Jenis data

adalah data citra tanaman kubis terserang beberapa jenis hama dan penyakit. Data didapatkan dengan menggunakan kamera *Canon Mirrorless M100* pada lahan seluas 1000 meter persegi. Data citra didapatkan pada masa mendekati panen tanaman kubis yang berumur 70 sampai 80 hari. Jumlah data adalah 242 data, terbagi menjadi 5 kelas yang dibedakan dari ciri-ciri hama dan penyakit yang nampak dari luar tanaman kubis, yaitu hama serangga, penyakit bercak daun, penyakit busuk hitam, penyakit busuk lunak, dan penyakit akar gada. Dengan jumlah data dapat dilihat pada Gambar 1.

Informasi jumlah data pada Gambar 1 menunjukkan ketidakseimbangan pada setiap distribusi antar kelas. Setelah mengetahui keadaan data tersebut maka diperlukan penanganan lebih lanjut untuk mengatasi data tidak seimbang. Selanjutnya dijelaskan metode yang digunakan pada penelitian ini.



Gambar 1. Informasi Data Kubis yang Terserang Hama dan Penyakit

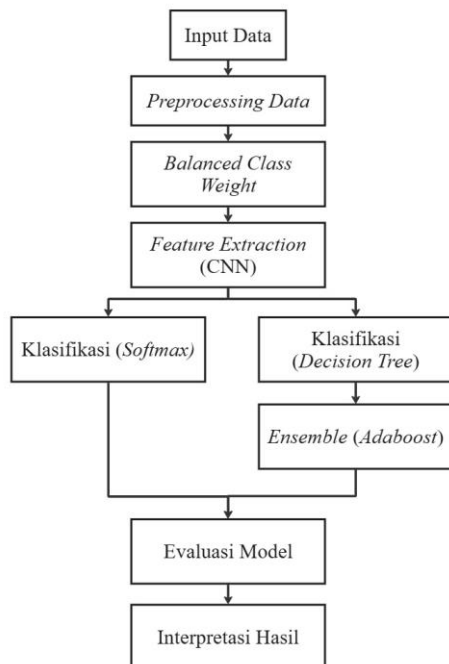
3. METODE PENELITIAN

Penelitian ini menggunakan beberapa metode untuk mengatasi data tidak seimbang, sebagaimana ditunjukkan pada Gambar 2.

Gambar 2 menunjukkan langkah-langkah penelitian dalam menangani klasifikasi data tidak seimbang menggunakan pengoptimalan pembobotan melalui ekstraksi fitur gambar oleh CNN, serta kombinasi metode klasifikasi *Decision Tree* dan *Ensemble Adaboost*. Proses dimulai dengan input data atau akuisisi data yang kemudian diproses melalui tahap *preprocessing* untuk meningkatkan kualitas data. Selanjutnya, dilakukan penyesuaian *Balanced Class Weight* untuk menangani ketidakseimbangan kelas. Fitur diekstraksi menggunakan CNN, yang kemudian digunakan untuk dua pendekatan klasifikasi, sebagai berikut.

1. Klasifikasi langsung menggunakan *Softmax* pada CNN.

2. Klasifikasi menggunakan *Decision Tree*, yang selanjutnya diperkuat dengan *Ensemble Adaboost* untuk meningkatkan kinerja model.



Gambar 2. Ilustrasi Langkah-Langkah Penelitian

Selanjutnya dijelaskan lebih detail langkah-langkah dalam penelitian ini sebagai berikut.

3.1. Preprocessing Data

Untuk mendapatkan hasil klasifikasi yang maksimal, maka data yang digunakan perlu diolah dan dibersihkan terlebih dahulu. Langkah *preprocessing* dibagi menjadi tiga, yaitu pengolahan gambar mentah, pelabelan, dan segmentasi gambar. Pengolahan gambar mentah dilakukan secara manual dengan menggunakan *photoshop* untuk menghilangkan *background* dari tanaman kubis. Pelabelan didasarkan pada penyakit dan hama yang menyerang tanaman kubis dibedakan dari ciri-ciri yang nampak dari luar, dengan label 0 (hama serangga), 1 (penyakit bercak daun), 2 (penyakit busuk hitam), 3 (penyakit busuk lunak), 4 (penyakit akar gada).

Segmentasi gambar dilakukan menggunakan teknik *Morphology* dari *library OpenCV*. Meskipun untuk kelas biner, tetapi penyesuaian teknik ini dengan *threshold* pada kelas yang didefinisikan dapat membantu model beradaptasi pada data multi-kelas (Lotufo et al., 2023). Teknik ini unggul dalam memastikan keutuhan fitur dari setiap kelas dan tidak menghilangkan informasi yang mendasar dari data citra (Bhutada et al., 2022). Kemudian sesuai dengan prinsip Pareto, maka data yang telah diolah dibagi menjadi data latih dan data uji yang dilakukan secara *random sampling* dengan proporsi 80:20 (Sharp et al., 2019).

3.2. Model Pembobotan

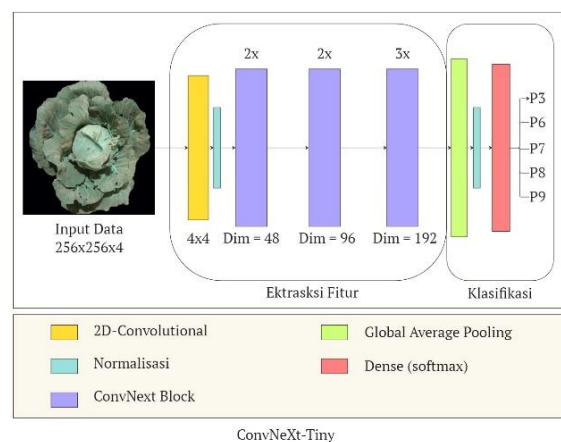
Untuk meningkatkan performa model, dilakukan pembobotan agar model memahami fitur laten sehingga akurasi pada kelas minoritas lebih diperhatikan (Dablain et al., 2023). Pembobotan ini dihitung menggunakan konsep *balanced class weight* yang bobotnya disesuaikan pada tingkat ketidakseimbangan pada data, ditunjukkan pada persamaan (1).

$$W_k = \frac{N}{J \times n_k} \quad (1)$$

Dengan N adalah jumlah total sampel pada dataset. k menunjukkan kelas ke- k dan J adalah jumlah kelas keseluruhan (Rokhana et al., 2022). Besarnya perhitungan bobot kelas minoritas kemungkinan menyebabkan terjadinya *overfitting*, karena bobot rendah pada kelas mayoritas (Alzammam et al., 2020). Untuk mengatasi ini maka dibutuhkan strategi lanjutan seperti teknik *Hybrid* untuk menangani generalisasi data.

3.3. Ekstraksi Fitur dengan Conv-NeXt Tiny

Meskipun CNN tidak menghasilkan akurasi yang baik pada data tidak seimbang, tetapi metode ini memiliki kekuatan teknik ekstraksi fitur sehingga metode ini dapat digabungkan dengan metode lain untuk memperkuat akurasi (Almarshdi et al., 2023). Terdapat berbagai algoritma arsitektur CNN, salah satunya adalah *Conv-Next-Tiny* dengan kelebihan parameter yang sedikit untuk mereduksi kompleksitas model dan meningkatkan efisiensi (Jiang et al., 2023). Arsitektur ini digunakan karena data yang digunakan pada penelitian ini memiliki jumlah data lebih sedikit. Ilustrasi dari ekstraksi fitur CNN dengan *Conv-NeXt Tiny* dapat dilihat pada Gambar 3.

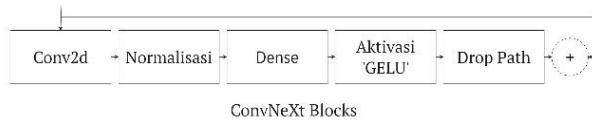


Gambar 3. Ilustrasi Model Dasar CNN dengan *Conv-Next Tiny*

Tahapan ekstraksi fitur pada Gambar 3 dimulai dengan memasukkan data citra berukuran 256x256 piksel dan kedalaman 4 piksel pada lapisan 2D

convolutional yang mereduksi dimensi spasial untuk mendapatkan fitur level rendah pada gambar seperti tekstur, tepian dan pola dari gambar kubis dengan kernel berukuran 4x4 (Todi et al., 2023). Selanjutnya, model diproses pada *ConvNext Block* yang terdiri dari lapisan *Conv2D* untuk menerapkan filter berkelompok yang jumlahnya tergantung kedalaman lapisan piksel. Kemudian diterapkan normalisasi untuk menstabilkan proses pelatihan

Proses ini dilanjutkan dengan lapisan *Dense* untuk lebih mereduksi dan mengekspansi dimensi, diterapkan ketika akan dilakukan penyesuaian data tidak linier oleh fungsi aktivasi 'GELU'. Terakhir diterapkan *drop path* untuk menghindari terjadinya *overfitting* (Liu et al., 2022). Berikut ilustrasi langkah-langkah *ConvNeXt Blocks* ditunjukkan pada Gambar 4.



Gambar 4. Ilustrasi *Conv-NeXt Blocks*

3.4. Model Klasifikasi

Penelitian ini mengusulkan model *Hybrid* dengan menggabungkan kekuatan ekstraksi fitur CNN dengan model kuat lainnya. *Hybrid* CNN dengan *Decision Tree*, dalam melakukan klasifikasi multi-kelas menggunakan cabang untuk memfokuskan pada aspek visual dengan kriteria pemisahan *Gini*, sehingga meningkatkan generalisasi model (Sharma and Brar, 2024). Berikut ditunjukkan perhitungan *Gini* pada model *Decision Tree* CART (*Classification and Regression Tree*) pada persamaan (2).

$$I_G(p) = \sum_{k=1}^K p_i (1 - p_k) \quad (2)$$

Dengan p_i adalah proporsi kelas ke- i dalam node. Kriteria klasifikasi *Gini* mengukur elemen yang salah diklasifikasikan jika berlabel acak, dimana jika nilai *Gini* rendah maka kelas dikatakan homogen dan sebaliknya (Sathiyapriya and Shanthi, 2022).

Untuk memaksimalkan klasifikasi, *Decision Tree* sebagai model estimasi dasar dioptimalkan dengan ensemble *Adaboost* untuk mengatasi distribusi kelas yang tidak seimbang (Taherkhani et al., 2020). *Adaboost* meningkatkan kinerja *Decision Tree* dengan pembobotan secara berurutan untuk memperbaiki kesalahan dari model prediksi dasar sebelumnya, sehingga mengurangi *overfitting* dan meningkatkan generalisasi model (Wu et al., 2020). Perhitungan *Adaboost* ditunjukkan pada persamaan (3).

$$H(x) = \text{sign} \left(\sum_{k=1}^J \alpha_k h_k(x) \right) \quad (3)$$

Dengan $h_t(x)$ adalah estimasi dasar oleh *Decision Tree* dan α_t adalah bobot dalam klasifikasi yang dihitung dengan persamaan (4).

$$\alpha_k = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_k}{\varepsilon_k} \right) \quad (4)$$

Dengan ε_k adalah pembobotan yang eror dalam klasifikasi (Mienye and Sun, 2022).

3.6. Hyperparameter Model

Untuk melakukan klasifikasi diperlukan *hyperparameter* untuk mengoptimalkan klasifikasi, selain bobot yang telah diatur terdapat beberapa *hyperparameter* digunakan untuk model CNN dan *Decision Tree*, ditunjukkan pada Tabel 1.

Tabel 1. *Hyperparameter Model*

Model	Hyperparameter	Pengaturan
CNN	Model Optimasi	Adam
	Epoch	100
	Ukuran Batch	32
<i>Decision Tree</i>	Kedalaman Pohon	7

Untuk mengoptimalkan CNN, digunakan optimizer Adam dengan batch size sebesar 32 dan jumlah epoch sebanyak 100, yang artinya proses pelatihan berjalan penuh selama 100 epoch, dengan *Adam optimizer* bertugas memperbarui bobot berdasarkan gradien yang dihitung. Pada *Decision Tree*, kedalaman maksimum pohon ditetapkan sebesar 7 untuk mengontrol kompleksitas.

3.5. Evaluasi Model

Untuk mengevaluasi model klasifikasi multi-kelas digunakan *Generalized Confusion Matrix* ditunjukkan pada Tabel 2. Dengan elemen diagonal *True Positive* (TP) digunakan untuk menghitung prediksi benar untuk setiap kelas. Elemen luar diagonal menunjukkan jumlah prediksi salah yang dibagi dua yaitu diagonal bawah untuk *False Positive* (FP) dan diagonal atas untuk *False Negative* (FN) (Grandini et al., 2020).

Tabel 2. *Generalized Confusion Matrix*

Aktual	Prediksi			
	Kelas 1	Kelas 2	...	Kelas-k
Kelas 1	X_{11}	X_{12}	...	X_{1k}
Kelas 2	X_{21}	X_{22}	...	X_{2k}
...	...	...	...	...
Kelas-k	X_{k1}	X_{k2}	...	X_{kk}

Pada data tidak seimbang, evaluasi model biasa seperti akurasi tidaklah cukup. *Balanced Accuracy* melakukan evaluasi yang adil pada setiap kelas karena perhitungan setara antara *true positive* dan

true negative ditunjukkan pada persamaan (5) (Byeon, 2021).

$$\text{Balanced Accuracy} = \frac{1}{J} \sum_{k=1}^J \text{Recall}_k \quad (5)$$

Dengan *Recall* adalah perhitungan prediksi akurasi untuk kelas positif yang ditunjukkan pada persamaan (5)

$$\text{Recall}_k = \frac{TP_k}{TP_k + FN_k} \quad (6)$$

Untuk evaluasi lebih lanjut, maka digunakan *Macro Average F1-score* yang lebih memfokuskan prediksi pada kelas minoritas tanpa mempertimbangkan proporsi label, ditunjukkan pada persamaan (7) (Grandini et al., 2020).

$$\text{Macro Average F1-score} = \frac{1}{J} \sum_{k=1}^J F1\text{-score}_k \quad (7)$$

Untuk memastikan bahwa model juga bekerja dengan baik pada kelas mayoritas maka digunakan *Weighted Average F1-score*, ditunjukkan pada persamaan (8) (Andriyanto, 2024).

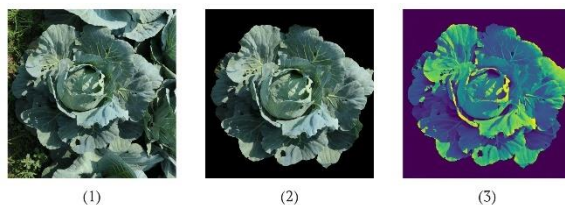
$$\text{Weighted Average F1-score} = \sum_{k=1}^J F1\text{-score} \quad (8)$$

4. HASIL

Pada penelitian ini digunakan CNN yang diterapkan dengan pembobotan *Balance Accuracy* untuk meningkatkan performa model, kemudian dilakukan penggabungan model CNN setelah pembobotan untuk ekstraksi fitur dengan *Decision Tree* sebagai model estimasi dasar dan *Adaboost* sebagai eksekusi akhir. Hasil dibandingkan antar model tunggal dan model gabungan untuk melihat model yang paling optimal. Dimana model dengan akurasi tertinggi dinyatakan sebagai model paling baik dalam mengatasi data tidak seimbang.

4.1. Analisis Preprocessing Gambar

Sebelum dilakukan klasifikasi maka data perlu diproses terlebih dahulu, berikut ditunjukkan hasil dari proses *preprocessing* pada Gambar 5.



Preprocessing Gambar Tanaman Kubis

Gambar 5. Contoh Gambar Setelah *Preprocessing*

Gambar (1) menunjukkan data citra mentah yang masih terlihat latar belakang dari tanaman kubis dengan warna dasar yang mirip dengan kubis, hal ini menyebabkan model salah mengklasifikasikan data citra apabila tidak diolah terlebih dahulu. Selanjutnya pada gambar (2) dilakukan penghilangan latar belakang tanaman kubis dengan *photoshop* menyebabkan fitur-fitur pada tanaman kubis terlihat jelas. Terakhir untuk merubah jenis gambar menjadi RGB maka diterapkan teknik segmentasi pada gambar (3) serta dilakukan peningkatan kualitas gambar dengan mempertajam gambar.

4.2. Hasil Perhitungan Bobot

Sebelum dilakukan klasifikasi maka Hasil pembobotan dengan persamaan (1) ditunjukkan pada Tabel 3.

Tabel 3. Hasil Perhitungan Bobot

Kelas	Bobot
0	0,283
1	1,379
2	4,289
3	3,217
4	4,825

Hasil pembobotan pada Tabel 3, menunjukkan perbedaan ukuran dimana jumlah sampel yang kecil mendapatkan bobot yang besar dan sebaliknya. Kemudian bobot ini digunakan sebagai *hyperparameter* pada proses CNN dengan tujuan menanganai ketidakseimbangan data.

4.3. Analisis Hasil Klasifikasi

Berikut ini ditunjukkan hasil klasifikasi model CNN dengan pembobotan, CNN dengan pembobotan yang dikombinasikan dengan *Decision Tree*, serta CNN dengan pembobotan yang dikombinasikan dengan *Decision Tree* dan *Adaboost*. Hasil evaluasi berdasarkan confusion matrix ditampilkan pada Tabel 4.

Tabel 4. Hasil *Confusion Matrix* Model

Model	Kelas				
	0	1	2	3	4
CNN	108	23	7	12	20
	6	14	3	3	9
	2	0	8	0	2
	6	4	0	5	0
	0	2	0	0	8
CNN+ Decision Tree	164	4	1	1	0
	1	32	2	0	0
	0	0	10	2	0
	0	0	0	15	0
	0	0	0	0	10
CNN+Decision Tree+Adaboost	169	1	0	0	0
	5	30	0	0	0
	0	0	10	0	2
	0	0	0	15	0
	0	0	0	0	10

Hasil confusion matrix menunjukkan bahwa model CNN dengan pembobotan mengalami kesalahan

klasifikasi dengan jumlah besar yang ditunjukkan pada kelas 4 sebesar 20 data dan kelas 1 sebesar 23 data. Tabel confusion matrix menunjukkan bahwa model *Hybrid CNN* dengan model lainnya menghasilkan ketepatan prediksi lebih banyak untuk seluruh kelas, termasuk kelas minoritas seperti kelas 3 dan 4, dimana untuk model *CNN-Decision Tree* terdapat 11 data yang mengalami kesalahan klasifikasi. Untuk gabungan model *CNN-Decision Tree-Adaboost* menunjukkan kesalahan klasifikasi sebanyak 8 data.

Berikutnya ditunjukkan hasil evaluasi model dengan menggunakan *Balanced Accuracy*, *Macro Average F1-score* dan *Weighted Average F1-score*, yang ditunjukkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Model

Model	Balanced Accuracy	Macro Average F1-score	Weighted Average F1-score
CNN	56,71%	45%	63%
CNN+ Decision Tree	94,25%	92%	96%
CNN+ Decision Tree+ Adaboost	93,69%	94%	97%

Hasil menunjukkan bahwa model CNN dengan pembobotan memiliki akurasi paling rendah pada ketiga model evaluasi. Model gabungan CNN berbobot dengan *Decision Tree* memiliki akurasi tinggi diatas 90% pada ketiga model evaluasi, tetapi hasil akurasi *Weighted Average* yang lebih tinggi dari pada *Macro Average* menunjukkan bahwa model masih lebih baik dalam mengklasifikasikan data mayoritas. Model gabungan CNN berbobot dengan *Decision Tree* dan *Adaboost* menunjukkan akurasi yang sedikit lebih tinggi daripada model sebelumnya, dan sama dengan model *CNN Decision Tree*, model ini menunjukkan bahwa akurasi oleh *Weighted Average* yang lebih tinggi daripada akurasi pada *Macro Average*, yang artinya model masih lebih baik dalam mengklasifikasikan data mayoritas daripada data minoritas.

Berdasarkan hasil *Confusion Matrix*, *balanced accuracy*, *Macro Average F1-score* dan *Weighted Average F1-score* menunjukkan bahwa model CNN dengan pembobotan masih belum mampu melakukan klasifikasikan data dengan baik, sesuai dengan penelitian oleh Dablain et al. (2023). Kemudian upaya *Hybrid* dan *ensemble* dengan *Decision Tree* dan *Adaboost* menunjukkan peningkatan yang signifikan dalam klasifikasi, dimana kesalahan dalam klasifikasi semakin sedikit

dan akurasi model meningkat pada data minoritas maupun data mayoritas. Hal ini menunjukkan bahwa penggabungan *Decision Tree* dan *Adaboost* dapat membantu model dalam generalisasi data, distribusi ketidakseimbangan kelas, dan lebih memperhatikan kelas yang tidak terlihat atau minoritas sesuai dengan penelitian oleh Jijo dan Abdulazeez (2021) mengenai penggabungan CNN sebagai ekstraksi fitur dengan *Decision Tree*, serta penelitian oleh Wu et al. (2022) mengenai model *Adaboost* yang dapat mengatasi data tidak seimbang pada model *Decision Tree*.

Meskipun pembobotan pada CNN bertujuan untuk meningkatkan perhatian model terhadap kelas minoritas, hasil menunjukkan bahwa model masih lebih cenderung mengklasifikasikan data mayoritas dengan lebih baik. Hal ini sesuai dengan penelitian oleh Dablain et al. (2023), bahwa CNN pada dasarnya mengoptimalkan bobot berdasarkan distribusi awal data, sehingga model tetap mengalami bias terhadap kelas mayoritas. Penggunaan *Decision Tree* dan *Adaboost* membantu mengurangi ketidakseimbangan ini dengan cara meningkatkan perhatian terhadap kesalahan klasifikasi di kelas minoritas melalui mekanisme *boosting* yang memperkuat pengaruh sampel yang sulit diklasifikasikan.

6. KESIMPULAN

Berdasarkan hasil dan pembahasan, penelitian ini berhasil mengembangkan model *Hybrid CNN* berbobot, *Decision Tree*, dan *Adaboost* untuk data multi-kelas dengan ketidakseimbangan. Model ini menunjukkan akurasi tertinggi sebesar 94,25%, meningkat secara signifikan dibandingkan model tunggal CNN berbobot dengan akurasi 56,71%. Evaluasi metrik seperti *balanced accuracy* dan *Macro F1-score* mengindikasikan bahwa metode *Hybrid* ini mampu meningkatkan sensitivitas terhadap data minoritas, tanpa mengabaikan performa data mayoritas. Keunggulan ini disebabkan oleh kombinasi kemampuan CNN untuk ekstraksi fitur, *Decision Tree* untuk menangani distribusi kelas, serta *Adaboost* untuk memperbaiki kelemahan model dalam iterasi sebelumnya. Untuk penelitian selanjutnya, peneliti menyarankan menggunakan data seimbang dan data tidak seimbang untuk dapat membandingkan performa model di dua kondisi data, serta upaya penanganan data tidak seimbang pada algoritma klasifikasi juga diusulkan pada penelitian selanjutnya.

DAFTAR PUSTAKA

ALMARSHDI, R., NASSEF, L., FADEL, E., ALOWIDI, N., 2023. Hybrid Deep Learning Based Attack Detection for Imbalanced Data Classification. *Intell. Autom. Soft Comput.* 35, pp.297–320.

ALZAMMAM, A., BINSALLEEH, H.,

- ASSADHAN, B., KYRIAKOPOULOS, K.G., LAMBOTHARAN, S., 2020. Comparative analysis on imbalanced multi-class classification for malware samples using CNN. 2019 Int. Conf. Adv. Emerg. Comput. Technol. AECT 2019, pp.1–6.
- ANDRIYANTO, R., 2024. Enhancing Sentiment Analysis Effectiveness with LSTM Variants , and Stratified K-Fold on Imbalanced Dataset 6, pp.981–990.
- BHUTADA, S., YASHWANTH, N., DHEERAJ, P., SHEKAR, K., 2022. Opening and closing in morphological image processing. *World J. Adv. Res. Rev.* 14, pp.687–695.
- BYEON, H., 2021. Comparing the Balanced Accuracy of Deep Neural Network and Machine Learning for Predicting the Depressive Disorder of Multicultural Youth. *Int. J. Adv. Comput. Sci. Appl.* 12, pp.584–588.
- CHABALALA, Y., ADAM, E., ALI, K.A., 2023. Exploring the Effect of Balanced and Imbalanced Multi-Class Distribution Data and Sampling Techniques on Fruit-Tree Crop Classification Using Different Machine Learning Classifiers. *Geomatics* 3, pp.70–92.
- DABLAİN, D., JACOBSON, K.N., BELLINGER, C., ROBERTS, M., CHAWLA, N. V., 2023. Understanding CNN Fragility When Learning With Imbalanced Data. *Mach. Learn.* pp.1–26.
- EL SAKKA, M., IVANOVICI, M., CHAARI, L., MOTHE, J., 2024. A Review of CNN Applications in Smart Agriculture Using Multimodal Data.
- GRANDINI, M., BAGLI, E., VISANI, G., 2020. Metrics for Multi-Class Classification: an Overview. pp.1–17.
- JIANG, L., YUAN, B., MA, W., WANG, Y., 2023. JujubeNet: A high-precision lightweight jujube surface defect classification network with an attention mechanism. *Front. Plant Sci.* 13, pp.1–16.
- JJO, B.T., ABDULAZEEZ, A.M., 2021. Classification Based on Decision Tree Algorithm for Machine Learning. *J. Appl. Sci. Technol. Trends* 2, pp.20–28.
- LIU, Z., MAO, H., CHRISTOPH, C.W., TREVOR, F., SAINING, D., BERKELEY, U.C., WU, C.-Y., FEICHTENHOFER, C., DARRELL, T., XIE, S., 2022. [ConvNeXt_CVPR22] A ConvNet for the 2020s. *Cvpr* 11976–11986.
- LOTUFO, R.A., AUDIGIER, R., SAÚDE, A. V., MACHADO, R.C., 2023. Chapter Six - Morphological Image Processing, in: Merchant, F.A., Castleman, K.R. (Eds.), *Microscope Image Processing (Second Edition)*. Academic Press, pp. 75–117.
- MIENYE, I.D., SUN, Y., 2022. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access* 10, pp.99129–99149.
- MYNA, A.N., K., M., K., M., S., R.H., JAIN, Y.D., 2023. Classification and Detection of Cabbage Leaf Diseases from Images Using Deep Learning Methods. *Autom. Control Intell. Syst.* 11, pp.1–7.
- REYA, S.S., MALEK, M.D.A., DEBNATH, A., 2022. Deep Learning Approaches for Cabbage Disease Classification, in: 2022 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET). pp. 1–5.
- ROKHANA, R., HERULAMBANG, W., INDRASWARI, R., 2022. Classification of Ocular Diseases on Fundus Images Using Weighted MobileNetV2. *IES 2022 - 2022 Int. Electron. Symp. Energy Dev. Clim. Chang. Solut. Clean Energy Transition, Proceeding* pp.570–575.
- SATHIYAPRIYA, G., SHANTHI, S.A., 2022. Image Classification using Convolutional Neural Network. 2022 1st Int. Conf. Electr. Electron. Inf. Commun. Technol. ICEEICT 2022, pp.6867–6884.
- SHARMA, R., BRAR, M., 2024. Agricultural Informatics: Assessing CNN and Decision Tree Effectiveness in Indian French Bean Variety Classification. 2024 4th Int. Conf. Intell. Technol. CONIT 2024, pp.1–5.
- SHARP, B., ROMANIUK, J., GRAHAM, C., 2019. Marketing's 60/20 Pareto Law. *SSRN Electron. J.*
- TAHERKHANI, A., COSMA, G., MCGINNITY, T.M., 2020. AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. *Neurocomputing* 404, pp.351–366.
- TODI, A., NARULA, N., SHARMA, M., GUPTA, U., 2023. ConvNext: A Contemporary Architecture for Convolutional Neural Networks for Image Classification. *Proc. - 2023 3rd Int. Conf. Innov. Sustain. Comput. Technol. CISCT* 2023.
- WU, Y., KE, Y., CHEN, Z., LIANG, S., ZHAO, H., HONG, H., 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena* 187, 104396.

Halaman ini sengaja dikosongkan