

MODEL KLASIFIKASI LIGHTWEIGHT UNTUK DETEKSI HAMA PERTANIAN DENGAN EFFICIENT NET, SPINAL NET FC, DAN SHARPNESS-AWARE MINIMIZATION

Naufal Ihsan Pratama^{1*}, Fityan Azizi², Adila Alfa Krisnadhi³

^{1*,2,3} Universitas Indonesia, Indonesia

Email: ^{1*}naufal.ihsan11@ui.ac.id, ²fityan.azizi@ui.ac.id, ³adila@cs.ui.ac.id

Abstrak

Petani Indonesia menghadapi tantangan besar yang disebabkan oleh hama, yang menyebabkan gagal panen, busuk batang, kerusakan daun, dan busuk buah. Mengembangkan model pendeteksian hama yang lightweight menjadi penting untuk membantu petani dalam peningkatan program pengendalian hama. Tujuan utama dari model ini adalah untuk mengklasifikasikan hama secara akurat dengan memanfaatkan kumpulan data berskala besar. Kumpulan data ini mencakup berbagai spesies dengan beragam skala, bentuk, latar belakang kompleks, dan tingkat kesamaan visual yang tinggi di antara spesies serangga. Penelitian ini menggunakan model klasifikasi lightweight berbasis Efficient Net. Model ini menggabungkan Spinal Net FC sebagai classifier dan mengadopsi Sharpness-Aware Minimization sebagai optimizer untuk meningkatkan kinerjanya. Model yang dikembangkan mencapai akurasi sebesar 68,2%. Selain itu, dengan mengimplementasikan metode yang diusulkan, performa model mengalami peningkatan yang signifikan, menghasilkan peningkatan akurasi tambahan sebesar 4%.

Kata kunci: Efficient Net; Spinal Net; klasifikasi gambar fine-grained

Abstract

Indonesian farmers face major challenges caused by pests, which cause crop failure, stem rot, leaf damage, and pod rot. Developing a lightweight pest detection model is important to assist farmers in improving pest control programs. The main goal of this model is to classify pests accurately by leveraging large-scale data sets. This data set includes a wide range of species with varying scales, shapes, complex backgrounds, and high levels of visual similarity among insect species. This study uses a lightweight classification model based on Efficient Net. This model combines SpinalNet FC as a classifier and adopts Sharpness-Aware Minimization as an optimizer to improve its performance. The developed model achieves an accuracy

How to cite: Naufal Ihsan Pratama, Fityan Azizi, Adila Alfa Krisnadhi (2022) Model Klasifikasi Lightweight Untuk Deteksi Hama Pertanian Dengan Efficient NET, Spinal Net FC, dan Sharpness-Aware Minimization, (7) 09, DOI 10.36418/syntax-literate.v7i9.15119

E-ISSN: 2548-1398

Published by: [Ridwan Institute](#)

of 68.2%. In addition, by implementing the proposed method, model performance was significantly improved, resulting in an additional 4% increase in accuracy.

Keywords: *Efficient Net; Spinal Net; fine-grained image classification*

Pendahuluan

Indonesia dikenal sebagai negara agraris, di mana mayoritas penduduknya bekerja di bidang pertanian dan memiliki lahan pertanian yang luas serta sumber daya alam yang beragam dan melimpah. Namun, salah satu masalah utama yang dihadapi petani adalah hama. Hama merupakan penyebab utama gagal panen, menyebabkan busuk batang, kerusakan daun, dan busuk buah sebelum masa panen. Oleh karena itu, diperlukan sistem deteksi dini hama tanaman yang dapat membantu petani dalam program pengendalian hama guna mengurangi masalah yang disebabkan oleh hama. Hal ini menjadikan penting untuk membangun model pendeteksian hama yang lightweight, efisien, dan memiliki kinerja yang baik agar petani dapat menggunakannya dengan mudah.

Klasifikasi hama serangga sulit dilakukan karena adanya keragaman spesies, sisik, bentuk, serta latar belakang yang kompleks di lapangan, sehingga seringkali terjadi kesamaan penampilan tinggi di antara spesies serangga. Untuk mengatasi tantangan ini, beberapa penelitian telah dilakukan menggunakan deep learning, seperti Convolutional Neural Network (CNN), MobileNetV2 (Sandler et al., 2018) dan Transformer (Vaswani et al., 2017).

Ung et al. (2022) telah melakukan penelitian terhadap beberapa teknik berbasis CNN untuk mengidentifikasi hama serangga, antara lain Residual Attention Network (RAN), Feature Pyramid Network (FPN), Multi-Branch and Multi-Scale Attention Network (MMAL-Net), dan Ensemble Technique (ET). Dengan menggabungkan model-model tersebut, teknik ensemble mampu mencapai akurasi sebesar 74,13% pada dataset IP102.

Peng et al. (2022) mengusulkan arsitektur Transformer dengan CNN sebagai backbone network dan attention-based classification head untuk pengenalan hama serangga. Selain itu, juga menjelajahi strategi pelatihan yang lebih efektif, model pra-pelatihan, dan metode augmentasi data untuk mencapai akurasi klasifikasi tertinggi. Kinerja klasifikasi model dievaluasi pada dataset IP102 dan mencapai akurasi sebesar 74,89%.

Setiawan et al. (2022) menunjukkan bahwa model yang lebih lightweight seperti Mobile Net mampu menghasilkan akurasi yang menjanjikan dengan menggabungkan beberapa metode selama pelatihan. Dengan memanfaatkan augmentasi CutMix, sparse regularization, dan learning rate dinamis, Mobile Net V2 unggul dibandingkan dengan model lain dengan parameter lebih besar. Ketika ketiga metode tersebut digabungkan selama pelatihan, akurasinya mencapai 71,32%, yang efisien karena menggunakan parameter paling sedikit dibandingkan dengan model lain yang lebih besar.

Dalam penelitian ini, tujuan penulis adalah mengembangkan model deep-learning yang ringan untuk mengklasifikasikan hama serangga dengan detail lebih baik dan

terintegrasi dengan program pengendalian hama terpadu. Kontribusi utama mencakup modifikasi pada model Efficient Net, yang merupakan model berbasis CNN paling efisien, dengan menggunakan SpinalNet FC sebagai classifier untuk mencapai akurasi yang lebih tinggi. Selain itu, penelitian ini menerapkan metode Sharpness-Aware Minimization (SAM) sebagai optimizer dalam model yang dikembangkan untuk meningkatkan generalisasi dengan efisien.

Metode Penelitian

A. Dataset

Dataset serangga hama IP102 memiliki total 75.222 gambar yang mewakili 102 spesies hama serangga tanaman yang umum dan berbeda (Wu et al., 2019). Contoh dari 12 kelas spesies ini dapat dilihat pada Gambar 1 dalam dataset IP102. Penelitian ini difokuskan pada pelatihan dan pengujian model menggunakan dataset IP102. Tugas klasifikasi dilakukan dengan memanfaatkan 45.095 gambar untuk data latih, 7508 gambar untuk data validasi, dan 22.619 gambar untuk data uji, yang semuanya diambil dari pemisahan data set IP102. Dalam kumpulan data IP102, terdapat beberapa tantangan praktis yang harus diatasi:

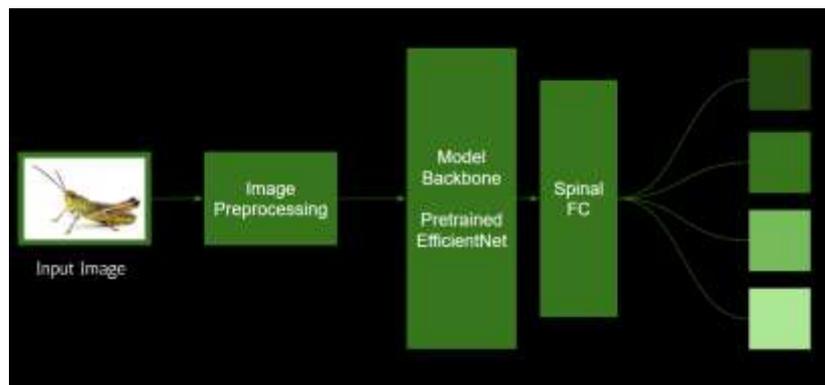
1. Ketidakseimbangan jumlah sampel pada kelas-kelas tertentu, di mana beberapa kelas memiliki jumlah sampel yang jauh lebih banyak daripada yang lain. Hal ini dapat mempengaruhi kinerja model dan akurasi klasifikasi pada kelas-kelas yang kurang representatif.
2. Perbedaan kecil antara gambar spesies serangga yang mirip dan adanya banyak gambar berkualitas rendah, seperti gambar dengan tanda air atau gambar yang buram. Hal ini bisa menyulitkan model dalam membedakan spesies yang serupa dan mempengaruhi akurasi keseluruhan.
3. Kumpulan data mencakup gambar-gambar dari berbagai tahap siklus hidup hama, termasuk telur, larva, pupa, dan hama dewasa. Hal ini meningkatkan kesulitan dalam klasifikasi, terutama ketika mengidentifikasi hama pada tahap larva yang sering memiliki perbedaan bentuk yang signifikan dengan tahap dewasa.



Gambar 1. Contoh Gambar dalam Dataset IP102

B. Arsitektur Model

Untuk membangun model klasifikasi fine-grained yang mudah digunakan dan efisien, penelitian ini mengadopsi pendekatan yang menggabungkan model Efficient Net (Tan et al., 2019) sebagai model backbone dengan pemanfaatan SpinalNet FC sebagai classifier (Kabir et al., 2022) untuk meningkatkan kinerja model. Selain itu, pelatihan model juga menggunakan optimizer SAM (Bahri et al., 2022) untuk mencapai hasil generalisasi yang lebih baik. Secara keseluruhan, ilustrasi arsitektur model yang diusulkan dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur Model

1. Image Preprocessing

Dalam penelitian ini, metode pra-pemrosesan gambar digunakan untuk koreksi data dengan mengurangi distorsi yang tidak diinginkan dan meningkatkan fitur-fitur penting pada gambar sebelum dilakukan proses lanjutan. Untuk tahap pra-pemrosesan ini, digunakan pustaka TorchVision (Marcel et al., 2010) yang menyediakan berbagai teknik, seperti pencerminan horizontal, rotasi, penyesuaian kecerahan, dan penyesuaian ukuran gambar agar berukuran 224x224 sesuai dengan ukuran yang disarankan.

2. Efficient Net

Efficient Net adalah arsitektur CNN dan metode penskalaan yang mengubah ukuran semua dimensi kedalaman, lebar, atau resolusi secara seragam menggunakan koefisien majemuk. Pendekatan penskalaan ini didasarkan pada asumsi bahwa ketika gambar input lebih besar, jaringan memerlukan lebih banyak lapisan untuk memperluas bidang reseptif dan lebih banyak saluran untuk menangkap pola fine-grained pada gambar yang lebih besar.

Dalam studi ini, digunakan EfficientNetB0 sebagai backbone model, yang terdiri dari 9 layer seperti yang terlihat pada Tabel 1 dan memiliki 5.3M parameter. Model EfficientNetB0 telah dilatih terlebih dahulu pada dataset ImageNet sebagai langkah pra-pelatihan, dan selanjutnya dilakukan transfer knowledge.

Tabel 1
Layer EfficientB0

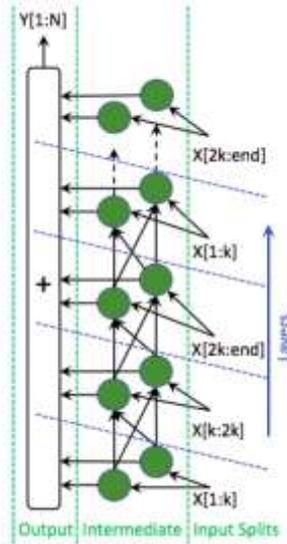
Layer	Resolusi	# Channels
Conv3x3	224x224	32
MBCConv1, k3x3	112x112	16
MBCConv6, k3x3	112x112	24
MBCConv6, k5x5	56x56	40
MBCConv6, k3x3	28x28	80
MBCConv6, k5x5	14x14	112
MBCConv6, k5x5	14x14	192
MBCConv6, k3x3	7x7	320
Conv1x1 & Pooling & FC	7x7	1280

3. Spinal Net FC

Spinal Net didesain dengan mensimulasikan sistem saraf chordate, yang secara unik menghubungkan banyak titik data penginderaan dan membuat keputusan lokal. Struktur SpinalNet dibagi menjadi tiga bagian, yaitu input split, intermediate split, dan output split, sebagaimana ditunjukkan dalam Gambar 3. Penelitian yang dilakukan oleh Kabir et al. (2022) mengusulkan Spinal Net sebagai lapisan Fully Connected (FC) pada VGG-5, dan berhasil mencapai kinerja state-of-the-art (SOTA) pada empat kumpulan data MNIST. Mereka menggunakan inisialisasi yang ditransfer dengan lapisan SpinalNet yang terhubung sepenuhnya untuk mencapai kinerja SOTA pada kumpulan data STL-10, Fruits 360, dan Caltech-101. Penelitian ini akan menerapkan pendekatan serupa dengan menggunakan Spinal Net FC untuk jaringan Efficient Net.

4. Sharpness Aware Minimization

Penggunaan model dengan jumlah parameter yang sangat besar memberikan hanya sedikit jaminan terhadap kemampuan generalisasi model. Untuk mengatasi hal ini, telah dikembangkan metode optimasi baru yang disebut Sharpness-Aware Minimization (SAM) yang mencari parameter di lingkungan dengan loss rendah yang seragam, formulasi ini dapat mengatasi masalah optimisasi min-maks di mana penurunan gradien dapat dilakukan secara efisien. Penggunaan SAM telah berhasil meningkatkan kemampuan generalisasi model pada berbagai dataset benchmark dan model, termasuk CIFAR-10, CIFAR-100, dan ImageNet (Bahri et al., 2022). Dalam penelitian ini, SAM digunakan sebagai optimizer untuk model.



Gambar 3. Arsitektur Spinal Net FC

Sumber: (Kabir et al., 2015)

5. Metrik Evaluasi

Model yang diusulkan dievaluasi menggunakan beberapa metrik yang cocok untuk klasifikasi multi-kelas dengan data IP102 yang tidak seimbang. Metrik-metrik tersebut meliputi macro-average precision (MPre), macro-average recall (MRec), macro-average F1 (MF1), dan accuracy (Acc) (Grandini et al., 2020).

Untuk memperlakukan kelas secara setara, dilakukan perhitungan recall untuk setiap kategori, kemudian hasilnya dirata-ratakan untuk mendapatkan nilai MRec, sebagaimana ditunjukkan di bawah ini:

$$Rec_c = \frac{TP_c}{TP_c + FN_c}$$

$$MRec = \frac{\sum_{c=1}^C MRec_c}{C}$$

C mengacu pada jumlah kelas, dalam penelitian ini yang dalam hal ini pembelajaran berjumlah 102 kelas. Huruf TP_c dan FN_c mewakili true positive dan false negative dari kelas c^{th} , masing-masing secara aktif. Demikian pula, Prec dan MPre dihitung sebagai berikut:

$$Pre_c = \frac{TP_c}{TP_c + FP_c}$$

$$MPre = \frac{\sum_{c=1}^C MPre_c}{C}$$

di mana FP_c adalah false positive dari kelas c^{th} . MF1 adalah rata-rata harmonik dari MRec dan MPre sebagai berikut:

$$MF1 = 2 \frac{MRec \cdot MPre}{MRec + MPre}$$

Acc dihitung dengan nilai positif sebenarnya di antara semua kelas sebagai berikut:

$$Acc = \frac{TP}{N}$$

di mana N adalah jumlah sampel.

C. Konfigurasi dan Skenario Pelatihan

Proses training pada penelitian ini menggunakan 100 epoch, batch size 16, learning rate 1e-3, dan fungsi loss cross-entropy. Proses pelatihan menggunakan GPU Nvidia 1650 dengan kapasitas memori 4GB. Terdapat tiga skenario pelatihan yang dilakukan. Pertama, dilakukan pelatihan model EfficientNetB0 tanpa menambahkan SpinalNet FC dan tanpa menggunakan optimizer SAM. Kedua, dilakukan pelatihan model EfficientNetB0 menggunakan optimizer SAM, tetapi tanpa menambahkan SpinalNet FC. Terakhir, dilakukan pelatihan model EfficientNetB0 dengan menambahkan SpinalNet FC dan menggunakan optimizer SAM. Optimizer dalam skenario pertama menggunakan optimizer Adam. Ketiga skenario menggunakan pre-trained model dari ImageNet.

Hasil dan Pembahasan

Performa model diukur menggunakan metrik akurasi, macro-average F1, macro-average precision, dan macro-average recall. Hasil kinerja model pada ketiga skenario yang telah ditentukan ditampilkan dalam Tabel 2. Model EfficientNetB0 mencapai akurasi sebesar 64.74%, dengan macro-average F1 sebesar 56.86%, macro-average precision sebesar 59.27%, dan macro-average recall sebesar 55.56%. Selanjutnya, model EfficientNetB0 yang menggunakan optimizer SAM mencapai akurasi sebesar 67.7%, dengan macro-average F1 sebesar 60.1%, macro-average precision sebesar 63.01%, dan macro-average recall sebesar 58.46%. Terakhir, model EfficientNetB0 yang ditambahkan SpinalNet FC sebagai classifier dan menggunakan optimizer SAM memperoleh hasil tertinggi dengan akurasi sebesar 68.22%, macro-average F1 sebesar 60.52%, macro-average precision sebesar 63.28%, dan macro-average recall sebesar 59.19%. Berdasarkan percobaan yang telah dilakukan, penggunaan SpinalNet FC sebagai classifier dan penggunaan optimizer SAM terbukti mampu meningkatkan performa model EfficientNetB0 secara signifikan dengan meningkatkan akurasi sekitar 4%.

Selanjutnya, Tabel 3 menunjukkan kinerja model dibandingkan dengan penelitian sebelumnya. Dalam perbandingan ini, kinerja model hanya dapat dibandingkan berdasarkan nilai akurasi, karena tidak semua penelitian sebelumnya memberikan hasil kinerja selain akurasi. Hasil yang tercantum dalam Tabel 3 menunjukkan bahwa model yang diusulkan pada penelitian ini tidak menghasilkan kinerja terbaik jika dibandingkan dengan hasil dari ketiga penelitian sebelumnya.

Tabel 2
Perbandingan Ketiga Skenario

Model	Acc	MF1	MRec	MPre
EfficientNetB0	64.74%	56.86%	55.56%	59.27%

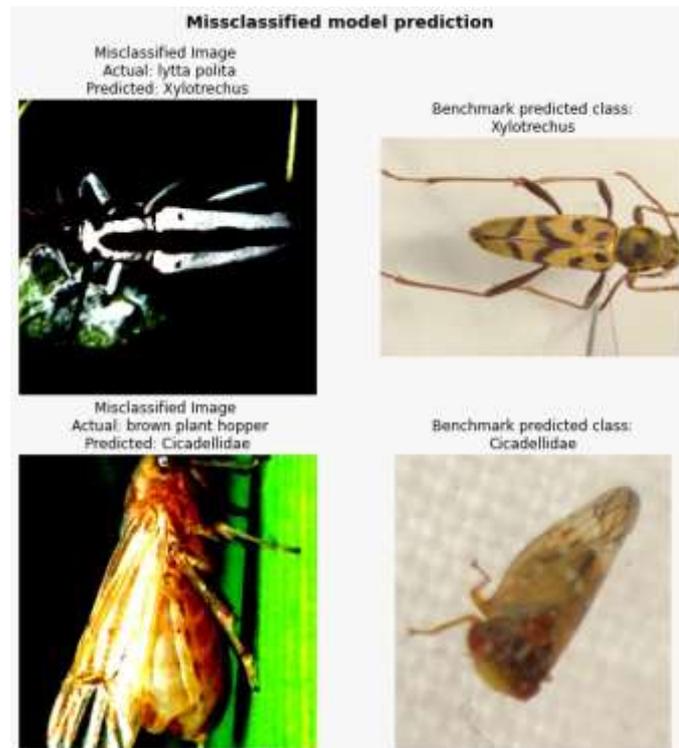
EfficientNetB0 + SAM	67.7%	60.1%	58.46%	63.01%
EfficientNetB0 + Spinal Net FC + SAM	68.22%	60.52%	59.19%	63.28%

Hasil terdekat dari model yang diusulkan adalah hasil penelitian dari Setiawan et al. (2022), yang memodifikasi model MobileNetV2 dan mencapai nilai akurasi sebesar 71.32%. Sedangkan, model MobileNetV2 tanpa modifikasi mencapai nilai akurasi sebesar 67.04%. Hasil ini menunjukkan perbedaan nilai yang signifikan dibandingkan dengan model EfficientNetB0 pada Tabel 3. Oleh karena itu, model MobileNetV2 dapat dipertimbangkan untuk penelitian selanjutnya karena memiliki jumlah parameter yang sama kecilnya dengan EfficientNetB0.

Tabel 3
Perbandingan Model Usulan dengan Penelitian Sebelumnya

Model	Acc	Params	Referensi
MobileNetV2	67.02%	4.2M	(Setiawan et al., 2022)
MobileNetV2 + Regularization	71.32%		
Ensemble CNN	74.13%	25M	(Ung et al., 2022)
Transformer + CNN	74.89%	28M	(Peng et al., 2022)
EfficientNetB0 + Spinal Net FC + SAM	68.22%	5.3M	

Gambar 4 menunjukkan hasil misklasifikasi oleh model yang diusulkan, di mana sisi kiri adalah gambar input dari data uji yang salah diklasifikasikan, dan sisi kanan adalah gambar aktual dari hasil klasifikasi, atau kelas prediksi benchmark. Dua contoh pada Gambar 4 menunjukkan kesamaan objek antara kelas aktual dan kelas prediksi, membuat model sulit untuk mengklasifikasikannya dengan benar. Contoh pada gambar kiri atas merupakan gambar masukan yang seharusnya tergolong ke dalam kelas *Lytta Polita*. Namun, model memprediksi gambar tersebut sebagai kelas *Xylotrechus*. Contoh pada gambar di kanan atas adalah salah satu gambar yang seharusnya dari kelas *Xylotrechus*. Dapat dilihat bahwa terdapat kesamaan bentuk objek antara kedua kelas tersebut. Selanjutnya, contoh pada gambar kiri bawah merupakan gambar masukan yang seharusnya tergolong ke dalam kelas *Brown Plant Hopper*. Model memprediksi gambar tersebut sebagai kelas *Cicadellidae*. Contoh pada gambar kanan bawah adalah gambar dari kelas *Cicadellidae*. Seperti yang dijelaskan sebelumnya, kelas *Brown Plant Hopper* dan *Cicadellidae* juga memiliki kesamaan.



Gambar 4. Contoh Gambar yang Salah Klasifikasi

Kesimpulan

Dalam penelitian ini, model EfficientNetB0 dikembangkan dengan menambahkan SpinalNet FC sebagai classifier dan menggunakan optimizer SAM untuk membangun model klasifikasi fine-grained yang lightweight. Model tersebut kemudian dilatih dan diuji menggunakan dataset IP102, menghasilkan akurasi sebesar 68.22%. Penambahan Spinal Net FC dan penggunaan optimizer SAM terbukti meningkatkan kinerja model secara signifikan dengan peningkatan sekitar 4% dalam akurasi. Disarankan agar penelitian di masa mendatang mengadopsi model lightweight lainnya, seperti MobileNetV2, untuk mencapai hasil yang lebih baik.

BIBLIOGRAFI

- Bahri, D., Mobahi, H., & Tay, Y. (2022). Sharpness-aware minimization improves language model generalization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers, pp. 7360-7371). Association for Computational Linguistics.
- Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview.
- Kabir, H. M. D., Abdar, M., Khosravi, A., Jalali, S. M. J., Atiya, A. F., Nahavandi, S., & Srinivasan, D. (2022). SpinalNet: Deep neural network with gradual input. IEEE Transactions on Artificial Intelligence, 1-13.
- Marcel, S., & Rodriguez, Y. (2010). Torchvision the machine-vision package of torch. In Proceedings of the 18th ACM International Conference on Multimedia (MM '10, pp. 1485-1488). ACM.
- Peng, Y., & Wang, Y. (2022). CNN and transformer framework for insect pest classification. Ecological Informatics, 72, 101846.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. Retrieved from <https://arxiv.org/abs/1801.04381>
- Setiawan, A., Yudistira, N., & Wihandika, R. C. (2022). Large scale pest classification using efficient convolutional neural network with augmentation and regularizers. Computers and Electronics in Agriculture, 200, 107204.
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In K. Chaudhuri & R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning (Vol. 97, pp. 6105-6114). PMLR.
- Ung, H. T., Ung, H. Q., Nguyen, T. T., & Nguyen, B. T. (2022). An efficient insect pest classification using multiple convolutional neural network-based models. Frontiers in Artificial Intelligence and Applications.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), Advances in Neural Information Processing Systems (Vol. 30, pp. 5998-6008). Curran Associates, Inc.

Naufal Ihsan Pratama, Fityan Azizi, Adila Alfa Krisnadhi

Wu, X., Zhan, C., Lai, Y.-K., Cheng, M.-M., & Yang, J. (2019). IP102: A large-scale benchmark dataset for insect pest recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Copyright holder:

Naufal Ihsan Pratama, Fityan Azizi, Adila Alfa Krisnadhi (2022)

First publication right:

Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

