

Klasifikasi Kualitas Citra Cabai dengan Menggunakan Algoritma Gradien Boosting

Umi Mahdiyah¹, Lilia Sinta Wahyuniar², Siti Rochana³

^{1,2,3}Program Studi Teknik Informatika,
Fakultas Teknik, Universitas Nusantara PGRI
Kediri,
e-mail : umimahdiyah@gmail.com

Penulis Korespondensi. Umi Mahdiyah, Program
Studi Teknik Informatika, Fakultas Teknik,
Universitas Nusantara PGRI Kediri,
e-mail : umimahdiyah@gmail.com

A B S T R A K

Objektif. Kementerian Pertanian Indonesia memiliki tujuan untuk meningkatkan penggunaan teknologi oleh petani dengan target 65-80% pada tahun 2020 dan 80-95% pada tahun 2024. Hal ini diharapkan dapat meningkatkan efisiensi dan ekonomi dalam sektor pertanian yang merupakan sektor penting dalam perekonomian Indonesia. Salah satu upaya untuk mencapai hal tersebut adalah dengan melakukan klasifikasi kualitas cabai berbasis citra.

Material and Metode. Dalam penelitian ini algoritma Gradien Boosting digunakan untuk melakukan klasifikasi citra cabai. Untuk simulasinya menggunakan aplikasi python.

Hasil. Rata-rata performa dari Gradient Boosting cukup baik untuk klasifikasi kualitas cabai berbasis citra. Akan tetapi pada kasus kelas *dried* selalu ada perbedaan dibandingkan kelas lain, karena banyak datanya jauh lebih besar dibandingkan yang lain, serta variasi data citranya lebih banyak.

Kesimpulan. Performa dari Gradien boosting cukup baik untuk klasifikasi kualitas cabai dengan nilai precision, recall, F-Score dan akurasinya adalah sekitar 69,7%, 69,1% 69,7%, dan $\pm 76\%$.

Kata kunci :

klasifikasi, kualitas cabai, gradien boosting

A B S T R A C T

Objective The Indonesian Ministry of Agriculture aims to increase the use of technology by farmers with a target of 65-80% in 2020 and 80-95% in 2024. This is expected to improve efficiency and the economy in the agricultural sector, which is an important sector in the Indonesian economy. One effort to achieve this is by conducting image-based chili quality classification..

Materials and Methods. In this research, the Gradient Boosting algorithm is used to classify chili images. Python application is used for simulation.

Results. The average performance of Gradient Boosting was good enough for image-based chili pepper quality classification. However, in the case of the "dried", there were always differences compared to other classes, as there were significantly more data points, and the image data had more variations.

Conclusion. The performance of Gradient Boosting was good enough for chili pepper quality classification with a precision, recall, F-score, and accuracy values of approximately 69.7%, 69.1%, 69.7%, and $\pm 76\%$.

Keywords :

Classification, chilli quality, gradien boosting

1. PENDAHULUAN

Pertanian merupakan sektor yang sangat penting dalam perekonomian Indonesia. Salah satu tujuan Kementan yang tercantum dalam Renstra Kementan tahun 2020-2024 adalah meningkatkan penggunaan teknologi oleh petani dengan target 65-80% pada tahun 2020 dan 80-95% pada tahun 2024 (Keputusan Menteri Pertanian Republik Indonesia No. 259/Kpts/RC.020/-M/05/2020 Tentang Rencana Strategis

Kementerian Pertanian Tahun 2020–2024, 2020). Teknologi untuk petani ini diharapkan dapat meningkatkan nilai efisiensi dan ekonomi dalam dunia pertanian.

Langkah awal dalam mengembangkan teknologi pertanian, salah satunya adalah mengenali objek pertanian yang ada. Salah satu objek pertanian di Indonesia yang sangat melimpah adalah tanaman cabai, karena cabai termasuk tanaman hortikultura yang banyak dibudidayakan oleh masyarakat petani di Indonesia. Selain sebagai penambah cita rasa dalam masakan atau sebagai sayuran yang banyak dibudidayakan oleh masyarakat petani di Indonesia. Selain sebagai penambah cita rasa dalam masakan terutama masakan khas Indonesia.

Untuk pengenalan cabai ini ada banyak metode, diantaranya dapat memanfaatkan machine learning. Selama ini machine learning banyak sekali dimanfaatkan dalam berbagai bidang dalam pertanian terutama dengan objek cabai, diantaranya adalah pengenalan jenis cabai yang telah dilakukan Ilyas Perlindungan dkk yaitu klasifikasi jenis cabai (Perlindungan, 2020) selain itu juga ada penelitian yang telah dilakukan oleh Tuti Purwaningsih dkk tentang klasifikasi cabe dengan CNN (Purwaningsih, Anjani, dan Utami 2018), pengenalan penyakit pada daun cabai (Naik et al., 2022), dll.

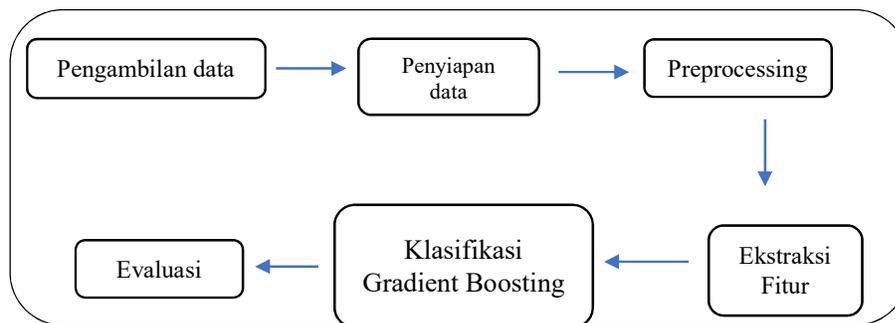
Pada penelitian ini dilakukan klasifikasi kualitas dari cabai menggunakan machine learning. Alasan dari pengambilan kasus klasifikasi kualitas cabai diantaranya adalah karena Kualitas cabai yang baik cenderung memiliki harga yang lebih tinggi dibandingkan dengan cabai yang kurang berkualitas. Dengan mengetahui kualitas cabai, petani dapat memperbaiki teknik budidaya dan teknik pemanenan yang sesuai sehingga dapat meningkatkan efisiensi produksi dan mengurangi kerugian akibat produksi cabai yang kurang berkualitas (Aziz dkk. 2021). Kualitas cabai yang baik dapat meningkatkan daya saing petani atau perusahaan di pasar global (Saptana, Muslim, dan Susilowati 2018). Cabai yang berkualitas tinggi biasanya memiliki permintaan yang lebih tinggi dan dapat memperkuat posisi pasar dari produsen atau eksportir. Selain itu juga dalam bidang pangan konsumen mengandalkan kualitas cabai sebagai salah satu faktor penting dalam menentukan keamanan pangan. Kualitas cabai yang buruk dapat mengandung bahan kimia berbahaya atau bakteri yang dapat membahayakan kesehatan manusia.

Gradient boosting merupakan algoritma klasifikasi *machine learning* yang menggunakan *ensemble* dari *decision tree* untuk memprediksi nilai. *Gradient boosting* termasuk *supervised learning* berbasis *decision tree* yang dapat digunakan untuk klasifikasi. *Gradient boosting* dimulai dengan menghasilkan pohon klasifikasi awal dan terus menyesuaikan pohon baru melalui minimalisasi fungsi kerugian (Kumar dan Reddy 2021).

Gradient Boosting dipilih dalam penelitian ini, karena dalam artikel sebelumnya algoritma ini memiliki performa yang cukup baik, diantaranya adalah untuk Prediksi Risiko Diabetes Menggunakan OCT Image Attributes menghasilkan akurasi 98% (Vidhyasree dan Parameswari 2021), serta penggunaan Gradient Boosting untuk deteksi penyakit jantung memiliki akurasi 97,1% (Kumar dan Reddy 2021), sehingga dengan menggunakan metode ini diharapkan mendapat performa yang baik pula. Pada penelitian ini objek cabai tidak hanya pengenalan citra cabai tunggal saja, akan tetapi ada dataset kumpulan cabai.

2. MATERIAL DAN METODE

Alur dari penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengambilan Data

Pada proses ini adalah proses pengambilan data. Data pada penelitian kali ini merupakan *public data* yang diambil dari public dataset yang dibuat oleh Yogesh Suryawanshi dkk pada tahun 2022 dan dipublikasikan dalam artikel data yang berjudul VegNet: Dataset of vegetable quality images for machine learning applications (Suryawanshi, Patil, dan Chumchu 2022). Data tersebut dapat diambil pada link <https://data.mendeley.com/datasets/73n5hrn8hh>, dengan rincian data dapat dilihat pada Tabel 1.

Tabel 1. Detail dataset yang digunakan

No	Jenis Kualitas Cabai	Jumlah Citra	Citra Cabai Tunggal	Citra Kumpulan Cabai
1	<i>Unripe</i> (Mentah)	189		
2	<i>Ripe</i> (matang)	183		
3	Old (Tua)	200		
4	Dried (Kering)	593		

No	Jenis Kualitas Cabai	Jumlah Citra	Citra Cabai Tunggal	Citra Kumpulan Cabai
5	Damaged(rusak)	121		
Total		1286		

2.2 Penyiapan Data

Pada tahap ini data dipilah dan dianalisis serta dipertimbangkan untuk mengambil berapa persen data training dan testing. Sehingga karena data yang ada jumlahnya tidak terlalu banyak sehingga untuk proses training testing, data digunakan 90% data training, 10% data testing. Selain itu juga pada tahap ini penyiapan dan pemilihan data kualitas cabai.

2.3 Preprocessing

Pada pemrosesan awal data citra cabai yang digunakan dalam penelitian ini akan dikonversi dari format data berwarna RGB (*Red, Green, Blue*) menjadi keabuan(*grayscale*).

2.4 Ekstraksi fitur

Ekstraksi ciri merupakan tahap yang harus dilakukan sebelum melakukan klasifikasi. Proses ini berkaitan dengan kuantitasi karakteristik citra kedalam sekelompok nilai ciri yang sesuai. Ciri citra diekstraksi dalam bentuk vektor ciri.

- Sebelum melakukan ekstraksi ciri citra cabai, tahap awal yang dilakukan adalah melakukan normalisasi citra cabai terlebih dahulu. Hal ini dilakukan untuk mengubah format, ukuran, dan dimensi citra cabai agar lebih mudah dalam proses pengambilan ciri.
- Menghitung gradien citra: Gradien citra dihitung dengan mengaplikasikan operator Sobel pada citra pada arah x dan y
- Menghitung magnitudo dan orientasi gradien: Magnitudo gradien dihitung sebagai akar kuadrat dari jumlah kuadrat gradien x dan y. Orientasi gradien dihitung sebagai arctan dari gradien y dibagi dengan gradien x
- Membagi citra menjadi sel: Citra kemudian dibagi menjadi sel-sel kecil dengan ukuran tertentu. Setiap sel menghitung histogram orientasi gradien berdasarkan magnitudo gradien pada setiap piksel di dalam sel. Ukuran sel dan jumlah bin histogram dapat disesuaikan tergantung pada jenis citra dan tujuan ekstraksi ciri.
- Selanjutnya, beberapa sel digabungkan menjadi blok. Untuk setiap blok, nilai setiap sel dinormalisasi dengan nilai blok untuk mengurangi pengaruh perbedaan pencahayaan dan kontras. Normalisasi dapat dilakukan dengan metode L1-norm atau L2-norm
- Vektor ciri untuk setiap blok digabungkan menjadi satu vektor fitur yang mewakili citra.

2.5 Klasifikasi Gradient Boosting

Pada tahap ini dilakukan proses training testing data. Proposisi data yang digunakan untuk testing pada penelitian ini adalah 10%, jadi 90% data untuk training, 10% untuk testing. Selain itu dalam pemilihan data training dan testing menggunakan *5-fold cross validation*. Setelah dilakukan pemilihan data training dan testing selanjutnya pemodelan dilakukan menggunakan Gradient Boosting.

Proses iteratif dalam algoritma Gradient Boosting terdiri dari beberapa tahap, yaitu:

1. **Inisialisasi**: Pada tahap ini, model awal dibuat sebagai model konstan yang merupakan rata-rata atau median dari target variabel.
2. **Membuat *weak model***: Pada tahap ini, *weak model* dibuat sebagai model yang dapat memprediksi kesalahan dari model sebelumnya. *Weak model* berupa pohon keputusan yang dangkal dengan satu atau dua cabang.
3. **Menghitung kesalahan residual**: Setelah *weak model* dibuat, kesalahan residual dihitung sebagai selisih antara nilai prediksi dari model sebelumnya dan nilai sebenarnya dari target variabel.
4. **Mengatur ulang data pelatihan**: Pada tahap ini, data pelatihan diatur ulang dengan menggunakan kesalahan residual sebagai target variabel.
5. **Membuat model baru**: Pada tahap ini, model baru dibuat dengan memprediksi kesalahan residual yang dihasilkan dari model sebelumnya.
6. **Menggabungkan model**: Model baru yang dibuat pada tahap sebelumnya digabungkan dengan model sebelumnya untuk membentuk model yang lebih baik.
7. **Iterasi**: Tahapan di atas diulang sampai memenuhi kondisi berhenti, seperti jumlah iterasi yang telah ditentukan atau ketika model tidak lagi mengalami peningkatan yang signifikan.

2.6 Evaluasi

Dalam proses evaluasi untuk mengukur performa algoritma, digunakan metode validasi *cross-validation* dengan menggunakan skema *5-fold*. Metode *cross-validation* ini akan menghasilkan nilai persentase keberhasilan klasifikasi dan *confusion matrix*. Dalam uji *cross-validation*, algoritma akan dievaluasi pada lima set data yang berbeda, dan hasil evaluasi akan dihitung dan diambil rata-rata untuk menghasilkan performa yang lebih akurat. *Confusion matrix* akan menunjukkan distribusi hasil klasifikasi pada setiap kelas, sehingga dapat digunakan untuk menganalisis kesalahan klasifikasi yang terjadi.

3. HASIL DAN PEMBAHASAN

Dari beberapa proses yang dilakukan yaitu pengklasifikasian kualitas cabai dengan menggunakan Gradient Boosting, hasilnya dapat dilihat pada tabel 2. Dalam penelitian ini ada 2 skenario berbeda yaitu pada bagian resize, pada penelitian eksperimen dilakukan dengan 100x100 piksel serta 150x150 piksel.

Tabel 2. Hasil Klasifikasi Kualitas Cabai dengan Algoritma Gradient Boosting

No	Kualitas Cabai	Precision	Recall	F1-score	Akurasi	Waktu Training(detik)
Resize 100x100						
1	Damaged(rusak)	0.50	0.36	0.42	75.97%	570,3045539855957
2	Dried(Kering)	0.86	0.87	0.86		
3	Old(Tua)	0.86	0.92	0.89		
4	Ripe(matang)	0.47	0.56	0.51		
5	Unripe (Mentah)	0.82	0.78	0.80		
Resize 150x150						
1	Damaged(rusak)	0.50	0.64	0.56	76.74%	1456.5981187820435
2	Dried(Kering)	0.84	0.85	0.85		
3	Old(Tua)	0.73	0.85	0.79		
4	Ripe(matang)	0.56	0.31	0.40		
5	Unripe (Mentah)	0.83	0.83	0.83		
	Rata-rata	0,697	0,691	0,697	76,35%	

Dengan *confusion matrix* untuk resize 100x100 sebagai berikut:

$$\begin{bmatrix} 5 & 4 & 0 & 3 & 2 \\ 4 & 59 & 2 & 3 & 0 \\ 0 & 0 & 12 & 1 & 0 \\ 1 & 5 & 0 & 9 & 1 \\ 0 & 1 & 0 & 3 & 14 \end{bmatrix}$$

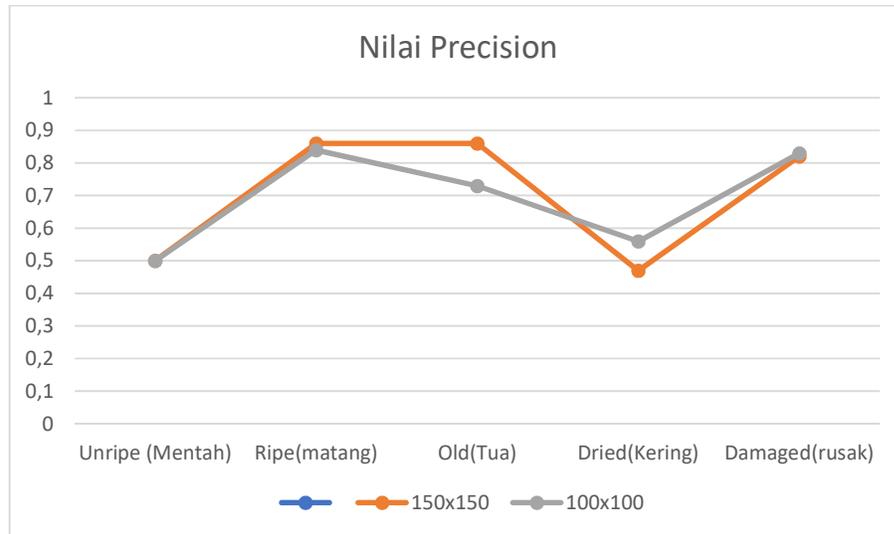
Sedangkan resize 150x150 sebagai berikut:

$$\begin{bmatrix} 9 & 1 & 0 & 2 & 2 \\ 6 & 58 & 4 & 0 & 0 \\ 0 & 1 & 11 & 1 & 0 \\ 2 & 8 & 0 & 5 & 1 \\ 1 & 1 & 0 & 1 & 15 \end{bmatrix}$$

Dari Tabel 2 dan *confusion matrix* dapat dilihat bahwa pada data dengan kelas *Damaged* banyak tidak dikenali karena data kelas *damaged* ada kemiripan ciri dengan kelas-kelas yang lain ada data cabai yang rusak saat masih muda, kering, atau sudah matang.

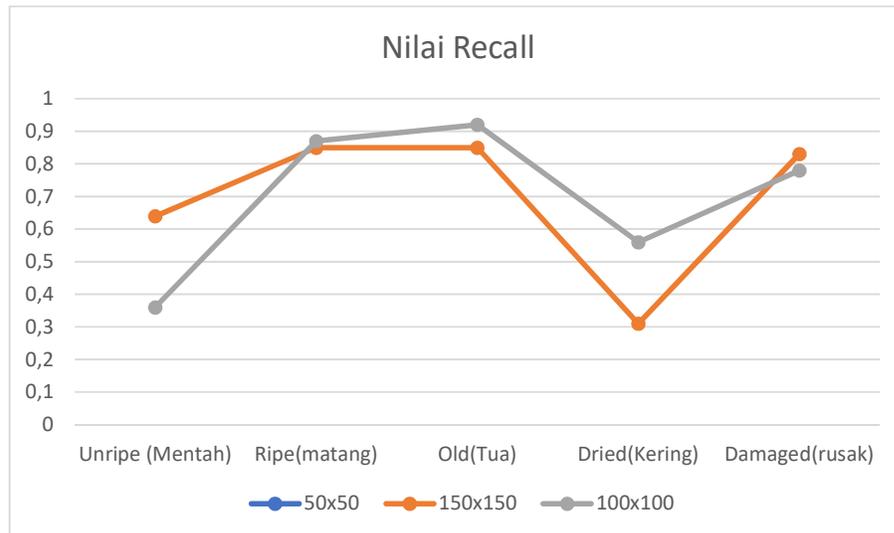
Tabel 2. dapat digambarkan dalam bentuk diagram yang dapat dilihat pada Gambar 2. Dari gambar 2 dapat diketahui nilai *precision* untuk citra yang dilakukan resize 150x150 hampir selalu lebih tinggi nilai *precisionnya* dibandingkan citra yang di resize 100x100. Artinya, dari semua data yang terklasifikasi kelas tersebut nilai kebenarannya rata-rata adalah 69%. Akan tetapi, pada kelas *Dried* ada nilai *precisionnya* lebih rendah, bisa jadi karena memang data pada kelas ini

sangat banyak dibandingkan data kelas lain, serta variasi data juga mempengaruhi,

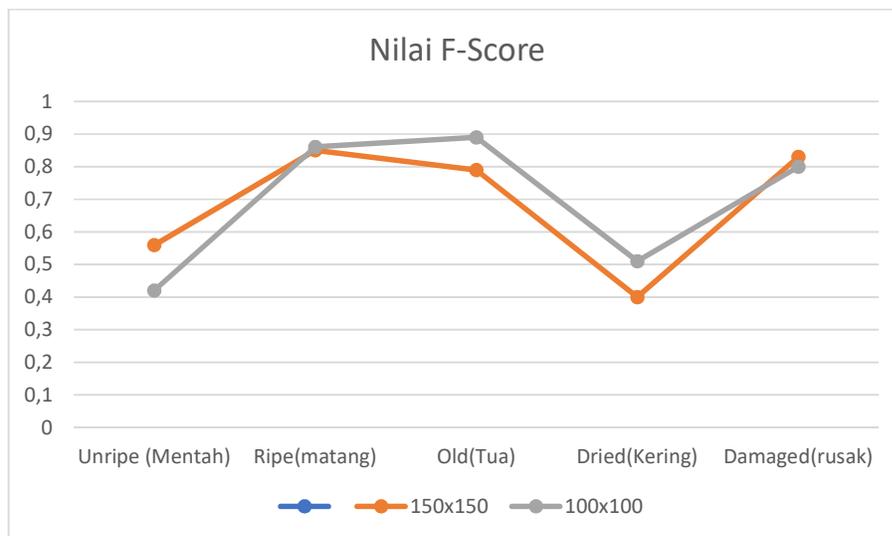


Gambar 2. Grafik Nilai Precision

Pada Gambar 3. Dapat dilihat bahwa tidak ada nilai recall yang selalu tinggi untuk masing-masing kelas, hanya saja pada kelas ripe nilai recall hampir sama,. Selain itu pada data dengan resize 150x150 nilai recall paling tinggi pada titik data unripe , ripe , old dan damaged .



Gambar 3. Grafik Nilai Recall



Gambar 4. Grafik Nilai F-Score

Nilai F-Score hampir sejalan dengan recall, pada Gambar 4 terlihat bahwa pada hampir semua atribut citra dengan resize 150x150 memiliki nilai F-Score tinggi dibandingkan yang lain, kecuali pada kelas dried. Penurunan tersebut terjadi karena banyaknya data yang ada pada kelas dried serta dipengaruhi jumlah data kelompok dan tunggal.

4. KESIMPULAN

Berdasarkan uraian sebelumnya, maka dapat disimpulkan bahwa, performa Gradien Boosting cukup bagus untuk klasifikasi kualitas cabai dengan banyak kelas ada 5 dengan rata-rata nilai precision, recall, F-Score dan akurasinya adalah sekitar 69,7%, 69,1%, 69,7%, dan $\pm 7\%$.

UCAPAN TERIMAKASIH

Terimakasih kepada semua pihak yang membantu terselesaikannya penelitian sekaligus penulisan artikel ini. Terimakasih juga kami sampaikan nantinya untuk pembaca artikel ini, semoga apa yang dibaca memberikan manfaat yang baik.

DAFTAR PUSTAKA

- Aziz, Maslina Abdul, Wan Muhammad Arif Mohamad Nazir, Azliza Mohd Ali, dan Jemal Abawajy. 2021. "Chili Ripeness Grading Simulation Using Machine Learning Approach." Dalam *2021 IEEE International Conference on Computing (ICOCO)*, , 253–58.
- Kementerian Pertanian. 2020. *Keputusan Menteri Pertanian Republik Indonesia No. 259/Kpts/RC.020/-M/05/2020 tentang Rencana Strategis Kementerian Pertanian tahun 2020–2024*. Indonesia: [https://rb.pertanian.go.id/upload/file/RENSTRA%20KEMANTAN%202020-2024%20REVISI%20%20\(26%20Agt%202021\).pdf](https://rb.pertanian.go.id/upload/file/RENSTRA%20KEMANTAN%202020-2024%20REVISI%20%20(26%20Agt%202021).pdf).

- Kumar, Kamarthi Lava, dan B Eswara Reddy. 2021. "Heart Disease Detection System Using Gradient Boosting Technique." Dalam *2021 International Conference on Computing Sciences (ICCS)*, , 228–33.
- Naik, B Nageswararao, R Malmathanraj, dan P Palanisamy. 2022. "Detection and classification of chilli leaf disease using a squeeze-and-excitation-based CNN model." *Ecological Informatics* 69: 101663. <https://www.sciencedirect.com/science/article/pii/S1574954122001121>.
- Perlindungan, Ilyas. 2020. *PENGENALAN TANAMAN CABAI DENGAN TEKNIK KLASIFIKASI MENGGUNAKAN METODE CNN*.
- Purwaningsih, Tuti, Imania Ayu Anjani, dan Pertiwi Bakti Utami. 2018. "Convolutional Neural Networks Implementation for Chili Classification." Dalam *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, , 190–94.
- Saptana, NFN, Chaerul Muslim, dan Sri Hery Susilowati. 2018. "Manajemen Rantai Pasok Komoditas Cabai pada Agroekosistem Lahan Kering di Jawa Timur." *Analisis Kebijakan Pertanian* 16(1): 19.
- Suryawanshi, Yogesh, Kailas Patil, dan Prawit Chumchu. 2022. "VegNet: Dataset of vegetable quality images for machine learning applications." *Data in Brief* 45.
- Vidhyasree, M, dan R Parameswari. 2021. "Meta Learning Gradient Boosted Neural Network Model Based Diabetes Risk Prediction with Bias Reduction Using OCT Image Attributes." Dalam *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, , 1188–95.