



Kumpulan data citra telepon pintar untuk identifikasi varietas cabai merah berbasis daun

Wiwin Suwarningsih¹, Evandri^{2*}, Rinda Kirana³, Purnomo Husnul Khotimah¹, Dianadewi Riswanti¹, Ekasari Nugraheni¹, Andri Fachrur Rozie¹, Andria Arisal¹, Devi Munandar¹, Noor Roufiq Ahmadi⁴

¹Pusat Riset Sains Data dan Informasi BRIN

²Direktorat Repositori, Multimedia dan Penerbitan Ilmiah BRIN

³Pusat Riset Hortikultura dan Perkebunan BRIN

⁴Balai Pengujian Standar Instrumen Tanaman Sayuran KEMENTAN

*Korespondensi: evandri.ahmad@gmail.com

Diajukan: 23-09-2024; **Direview:** 21-12-2024; **Diterima:** tgl-blh-thn; **Direvisi:** 30-12-2024

ABSTRACT

Chili plants play a crucial role in human life as a source of vitamins and minerals for public health. The market demand for red chili continues to rise, prompting seed producers to provide high-quality plant seeds. The selection of chili varieties is tailored to market demands, including taste, color, appearance, size, high productivity, resistance to pests, and suitability for local ecosystem conditions. Based on these considerations, an intelligent approach is needed for the identification of plant varieties to maintain seed purity. To facilitate and streamline the identification of chili varieties based on leaf characteristics, a comprehensive dataset has been compiled. This dataset consists of 3,877 leaf images divided into 12 variety classes, aimed at determining which plants are parent seeds or which have deviated from their variety. Leaf images were collected from the BALITSA garden by observing leaf growth from shoots up to 20 days of plant age. Various stringent measures have been taken to ensure the quality of the dataset and enhance its utility. The leaf images of chilies, taken from various angles and in high resolution, are designed to assist in the development of highly accurate models. By utilizing this curated dataset, it is possible to train models for real-time identification of chili varieties based on leaf characteristics, significantly aiding in the timely identification of these conditions.

ABSTRAK

Tanaman cabai berperan penting dalam kehidupan manusia, yaitu sebagai sumber vitamin dan mineral untuk kesehatan masyarakat. Permintaan pasar akan cabai merah terus meningkat sehingga mendorong produsen benih agar menyediakan benih tanaman yang bermutu. Pemilihan varietas tanaman cabai disesuaikan dengan permintaan pasar yaitu rasa, warna, penampakan, ukuran, produktifitas tinggi, tahan terhadap serangan organisme pengganggu tumbuhan dan cocok ditanam pada kondisi ekosistem setempat. Berdasarkan hal tersebut maka diperlukan suatu cara yang cerdas dalam melakukan identifikasi varietas tanaman untuk mempertahankan kemurnian benih. Untuk memudahkan dan mengefisienkan identifikasi varietas cabai berbasis daun, telah disusun kumpulan data yang komprehensif. Kumpulan data ini, yang terdiri dari 3877 gambar daun yang terbagi menjadi 12 kelas varietas, bertujuan untuk menentukan tanaman mana yang merupakan benih induk atau benih yang mengalami penyimpangan dari varietasnya. Citra daun dikumpulkan dari kebun BALITSA melalui pengamatan pertumbuhan daun dari tunas sampai dengan usia tanaman 20 hari. Berbagai langkah ketat telah diambil untuk memastikan kualitas kumpulan data dan meningkatkan kegunaannya. Citra daun cabai yang diambil dari berbagai sudut dan memiliki resolusi tinggi dirancang untuk membantu pengembangan model yang sangat akurat. Dengan memanfaatkan kumpulan data yang dikurasi ini, dimungkinkan untuk melatih model untuk identifikasi varietas cabai berbasis daun secara *real-time*, yang secara signifikan membantu dalam identifikasi kondisi tersebut secara tepat waktu.

Keywords: chili variety identification; deep learning; balitsa; chili leaves

INFORMASI DATASET

- **Bidang Ilmu:**
Agriculture
- **Bidang Ilmu Spesifik:**
Dataset ini menawarkan sumber daya yang vital untuk pengembangan model *deep learning*, memungkinkan deteksi dini dan akurat terhadap varietas cabai berdasarkan karakteristik daunnya. Model yang digunakan secara otomatis mengekstraksi fitur penting dari citra tanpa memerlukan intervensi manual untuk menentukan tanaman mana yang merupakan benih induk atau benih yang mengalami penyimpangan dari varietasnya.
- **Tipe Data:**
Format dataset ini adalah berbentuk gambar dengan tipe datanya adalah jpeg.
- **Metode Pengumpulan Data:**
Pengumpulan data dilakukan saat usia tanaman cabai berumur 2 minggu sampai dengan usia 5 bulan. Citra daun cabai diambil menggunakan perangkat telepon cerdas (*smartphone*) dengan menggunakan 2 jenis *background* yaitu menggunakan kertas berwarna hitam dan latar belakang alami seperti apa adanya.
- **Lokasi Sumber Data:**
Data diambil dari kebun perakitan varietas cabai Balai Penelitian Tanaman Sayuran (BALITSA) Cikole Lembang Jawa Barat.
- **Aksesibilitas Data:**
Dataset ini diserahkan ke dalam Repositori Ilmiah Nasional yang dikelola oleh Badan Riset dan Inovasi Nasional dan dapat diakses melalui link: <https://data.brin.go.id/dataset.xhtml?persistentId=hdl:20.500.12690/RIN/ND3BJI>
- **Publikasi terkait:**
Suwarningsih, W., Khotimah, P. H., Rozie, A. F., Arisal, A., Riswantini, D., Nugraheni, E., Munandar, D., & Kirana, R. (2022). Ide-cabe: chili varieties identification and classification system based leaf. *Bulletin of Electrical Engineering and Informatics*, 11(1), 445–453. <https://doi.org/10.11591/eei.v11i1.3276>
- **Kebermanfaatan Data:**
 - Dataset ini berharga karena dapat digunakan untuk proses *training* dalam menghasilkan model yang optimal. Setelah model identifikasi ini dikembangkan, maka fokus berikutnya adalah menentukan tanaman mana yang merupakan benih induk atau benih yang mengalami penyimpangan dari varietasnya.
 - Data ini dapat digunakan kembali oleh peneliti lain untuk menghasilkan model yang *robust* sehingga identifikasi varietas memiliki akurasi yang semakin baik.

1. PENDAHULUAN

Smart farming, atau pertanian pintar, adalah pendekatan modern dengan mengintegrasikan teknologi digital dan Internet of Things (IoT) yang bermanfaat untuk meningkatkan efisiensi, produktivitas, dan keberlanjutan dalam pertanian (Doshi et al., 2019). Konsep ini bertujuan untuk memanfaatkan data real-time dan analisis cerdas untuk membuat keputusan yang lebih baik dalam proses pertanian. Menggunakan sensor, kamera, drone, dan sistem otomatis, smart farming dapat memberikan wawasan yang mendalam tentang kondisi tanah, kesehatan tanaman, dan kebutuhan irigasi, serta mengoptimalkan penggunaan sumber daya.

Adanya perkembangan teknologi smart farming, identifikasi varietas tanaman menjadi salah satu tantangan penting (Swami & Shamili, 2022). Salah satu komoditas yang memerlukan perhatian khusus dalam hal identifikasi varietas adalah cabai. Cabai merupakan tanaman hortikultura yang penting, baik dari segi ekonomi maupun konsumsi. Varietas cabai yang berbeda memiliki karakteristik morfologi dan genetik yang bervariasi, mempengaruhi rasa, ukuran, bentuk, dan ketahanan terhadap penyakit. (Hashim et al., 2024). Identifikasi varietas cabai yang akurat sangat penting untuk beberapa alasan:

1. **Pemilihan Varietas yang Tepat:** Petani perlu memilih varietas cabai yang sesuai dengan kondisi iklim dan tanah di wilayah mereka. Identifikasi varietas yang tepat dapat meningkatkan hasil panen dan mengurangi kerugian akibat penyakit atau hama.
2. **Pengendalian Penyakit dan Hama:** Setiap varietas cabai memiliki tingkat ketahanan yang berbeda terhadap penyakit dan hama. Dengan mengidentifikasi varietas secara akurat, petani dapat menerapkan strategi pengendalian yang lebih efektif dan mengurangi penggunaan pestisida.
3. **Kualitas dan Harga:** Varietas cabai tertentu mungkin memiliki kualitas yang lebih tinggi dan harga jual yang lebih baik. Identifikasi varietas yang benar dapat membantu petani memaksimalkan keuntungan dan memenuhi permintaan pasar.
4. **Konservasi dan Pembangunan Varietas:** Identifikasi varietas juga penting dalam upaya konservasi sumber daya genetik dan pengembangan varietas baru yang lebih unggul. Data varietas yang akurat dapat membantu dalam penelitian dan pengembangan varietas cabai yang lebih baik.
5. **Integrasi dengan Teknologi:** Dalam konteks smart farming, teknologi seperti pengolahan citra dan pembelajaran mesin (machine learning) dapat digunakan untuk mengidentifikasi varietas cabai dengan cepat dan akurat. Namun, teknologi ini memerlukan data yang berkualitas tinggi dan sistem yang terintegrasi dengan baik.

Tanaman cabai memiliki banyak varietas, yang setiap varietasnya memiliki ciri-ciri morfologi daun yang unik, seperti bentuk, ukuran, warna, dan tekstur (Paredes et al., 2020). Variabilitas ini penting untuk identifikasi varietas, karena morfologi daun mencerminkan karakteristik genetik dan kondisi lingkungan. Teori ini didukung oleh penelitian di bidang botani dan agronomi, yang menunjukkan bahwa variasi morfologi daun dapat digunakan sebagai penanda untuk membedakan varietas tanaman.

Dalam menghasilkan data citra, variasi morfologi daun cabai diabadikan melalui gambar berkualitas tinggi untuk memastikan bahwa fitur-fitur penting ini dapat diekstraksi dan dianalisis. Setiap gambar diambil dengan memperhatikan variasi sudut, pencahayaan, dan tahap pertumbuhan untuk menangkap seluruh spektrum karakteristik morfologi.

Data citra yang dihasilkan dirancang untuk melatih model deep learning yang dapat membedakan varietas cabai dengan akurasi tinggi. Setiap gambar disiapkan dengan anotasi yang tepat, memastikan bahwa model dapat belajar dari data yang terstruktur dengan baik. Penerapan augmentasi data digunakan untuk memperkaya dataset dan meningkatkan kemampuan generalisasi model.

2. METODE

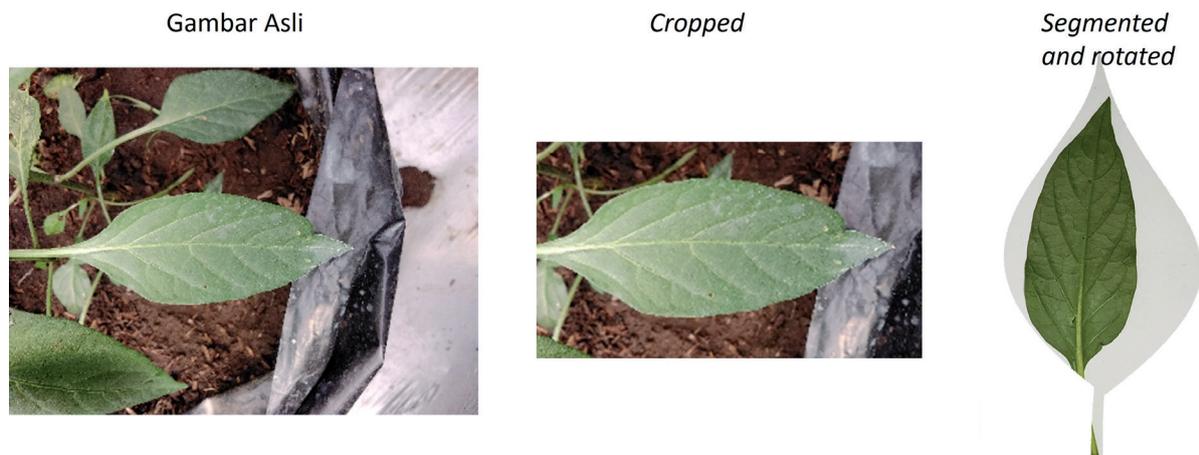
Kumpulan data citra daun cabai diperoleh melalui serangkaian pemotretan terhadap beragam varietas cabai yang ada di kebun perakitan varietas tanamann cabai Balai Pusat Penelitian Tanaman dan Sayuran (BALITSA), Lembang, Jawa Barat. Pemotretan dilakukan menggunakan kamera smartphone resolusi tinggi dengan mengatur pencahayaan yang konsisten untuk mengurangi bayangan dan highlight berlebih. Semua gambar diambil dalam kondisi yang serupa untuk menjaga kualitas data. Pengambilan gambar dilakukan dari berbagai sudut dan jarak untuk mendapatkan variasi yang cukup, sehingga model dapat mengenali daun cabai dari berbagai perspektif. Tanaman yang dipotret memiliki rentang usia dari 2 minggu sampai 5 bulan.

Pengolahan data citra daun cabai melibatkan serangkaian tahapan yang bertujuan untuk mengekstraksi informasi penting dari gambar untuk keperluan identifikasi dan klasifikasi. Tahapan utama dalam pengolahan data citra daun cabai meliputi: preprocessing citra, segmentasi citra, ekstraksi fitur, augmentasi data, klasifikasi dan identifikasi, serta evaluasi dan validasi.

Analisis dataset citra daun varietas cabai terdiri dari beberapa tahapan yang bertujuan untuk memahami, mengekstraksi informasi penting, dan mempersiapkan data untuk model *deep learning*. Tahapan yang dilakukan untuk menganalisis dataset citra daun varietas cabai meliputi: (i) memeriksa jumlah total gambar, jumlah kelas (varietas cabai), dan distribusi gambar dalam setiap kelas. (ii) visualisasi citra daun untuk memahami variasi visual antar varietas dan (iii) Analisis statistik deskriptif seperti ukuran rata-rata gambar, resolusi, serta distribusi warna pada gambar.

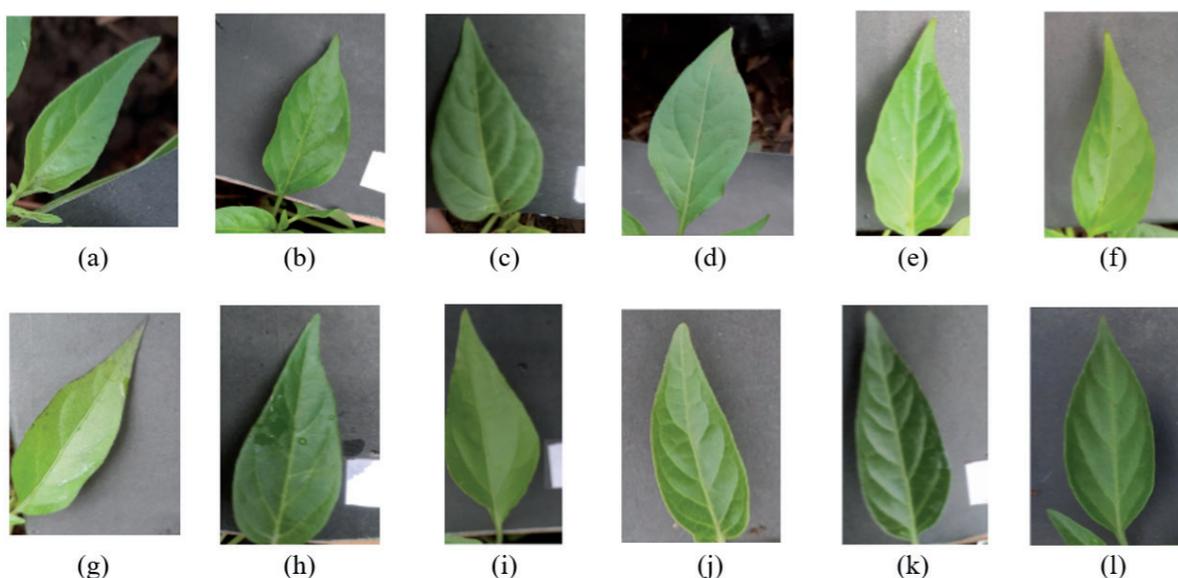
3. DESKRIPSI DATA

Dataset citra daun cabai yang tersedia di Repositori Ilmiah Nasional merupakan dataset hasil dari *pre-processing* yang siap dipakai untuk training deep learning identifikasi varietas. *Pre-processing* dilakukan sesuai gambar 1.



Gambar 1: Tahapan *pre-processing* data

Kumpulan citra berupa gambar asli pada Gambar 1, dikumpulkan sejak usia tanaman cabai berumur 2 minggu sampai dengan usia 5 bulan. Kami memperoleh 1.114 gambar dan terbagi menjadi 12 kelas varietas yaitu Tanjung-1, Tanjung-2, Branang, Ciko, Lingga, Inata Agrihorti, Carvi, Hot Beauty, Carla Agrihorti, Mia, Pilar and Hot Vision (contoh bentuk daun dapat dilihat pada gambar 2). Untuk meningkatkan jumlah dan keragaman gambar, kami melakukan augmentasi dengan cara memotong bagian kecil pinggir daun, memutar dan menggeser.



Gambar 2. Contoh daun berdasarkan kelas varietasnya (a) Tanjung-1, (b) Tanjung-2, (c) Branang, (d) Ciko, (e) Lingga, (f) Inata Agrihorti, (g) Carvi, (h) Hot Beauty, (i) Carla Agrihorti, (j) Mia, (k) Pilar and (l) Hot Vision

Dari proses augmentasi, diperoleh sebanyak 3877 gambar daun dalam format JPEG yang dibagi menjadi 12 folder sesuai jumlah varietas. Jumlah gambar hasil augmentasi per varietas proporsional terhadap jumlah gambar asli. Tabel 1 adalah data jumlah gambar daun dan folder masing-masing varietas pada Repositori Ilmiah Nasional.

Tabel 1. Jumlah gambar daun dan Folder masing-masing varietas di Repositori

No	Varietas	Nama folder di RIN (folder akan muncul jika <i>Tree view</i> diaktifkan)	Jumlah gambar
1	Tanjung-1	1-Tanjung-1	560
2	Tanjung-2	2-Tanjung-2	573
3	Branang	3-Branang	144
4	Ciko	4-Ciko	422
5	Lingga	5-Lingga	412
6	Inata Agrihorti	6-Inata Agrihorti	226
7	Carvi	7-Carvi	677
8	Hot Beauty	8-Hot Beauty	213
9	Carla Agrihorti	9-Carla Agrihorti	142
10	Mia	11-Mia	149
11	Pilar	12-Pilar	208
12	Hot Vision	13-Hot Vision	151

Gambar 3 sampai gambar 14 menampilkan 3 contoh gambar hasil augmentasi dari masing-masing varietas (lihat link aksesibilitas data pada INFORMASI DATASET).



Gambar 3. Daun cabai varietas Tanjung-1

Gambar 3. adalah 3 contoh daun cabai varietas Tanjung-1 yang telah diolah untuk menghilangkan backgroundnya. Karakteristiknya memiliki bentuk daun lanset dengan panjang daun berkisar antara 14,9 hingga 15,9 cm dan lebar daun 5,2 hingga 6,2 cm.



Gambar 4. Daun cabai varietas Tanjung-2

Gambar 4. contoh daun cabai varietas tanjung-2. Varietas ini memiliki karakteristik daun berbentuk lanset dengan panjang daun 15,8 hingga 16,8 cm dan lebar daun 5,1 hingga 6,1 cm.



Gambar 5. Daun cabai varietas Branang

Gambar 5 adalah 3 contoh daun cabai varietas branang. Varietas ini memiliki karakteristik daun berbentuk lanset dengan panjang 7,0 hingga 9,0 cm dan lebar 3,0 hingga 4,0 cm.



Gambar 6 Daun cabai varietas ciko

Gambar 6 adalah contoh daun cabai varietas ciko. Varietas ini memiliki karakteristik bentuk daun lanset dengan panjang 14,8 hingga 15,8 cm dan lebar 5,0 hingga 6,1 cm.



Gambar 7. Daun cabai varietas lingga

Gambar 7 adalah contoh daun cabai varietas lingga. Varietas ini memiliki karakteristik daun berbentuk lanset dengan panjang 12,3 hingga 13,8 cm dan lebar 4,8 hingga 6,1 cm.



Gambar 8. Daun cabai varietas inata agrihorti

Gambar 8 adalah contoh daun cabai varietas Inata . Varietas ini memiliki karakteristik daun lanset dengan panjang 8,09 hingga 9,10 cm dan lebar 3,55 hingga 3,82 cm.



Gambar 9. Daun cabai varietas carvi

Gambar 9 adalah contoh daun cabai varietas carvi. Varietas ini memiliki karakteristik bentuk daun elips (*elliptical*) dengan panjang 6,6 hingga 9,11 cm dan lebar 2,70 hingga 4,03 cm.



Gambar 10. Daun cabai varietas hot beauty

Gambar 10 adalah contoh daun cabai varietas hot beauty. Varietas ini memiliki karakteristik daun lanset dengan panjang daun 8,0 hingga 13,0 cm dan lebar 2,0 hingga 3,5 cm.



Gambar 11. Daun cabai varietas Carla agrihorti

Gambar 11 adalah contoh daun cabai varietas carla agrihorti. Varietas ini memiliki karakteristik bentuk daun lanset dengan panjang daun 6,1 hingga 7,1 cm dan lebar 2,6 hingga 2,9 cm.



Gambar 12. Daun cabai varietas Mia

Gambar 12 adalah contoh daun cabai varietas Mia. Varietas ini memiliki karakteristik daun berbentuk bulat telur dengan panjang daun 8,0 hingga 9,0 cm dan lebar daun 4,7 hingga 5,5 cm.



Gambar 13. Daun cabai varietas Pilar

Gambar 13 adalah contoh daun cabai varietas Mia. Varietas ini memiliki karakteristik bentuk daun jorong dengan panjang daun 8,0 hingga 10,0 cm dan lebar 4,3 hingga 5,6 cm.



Gambar 14. Daun cabai varietas Hot Vision

Gambar 14 adalah contoh daun cabai varietas Hot Vision. Varietas ini memiliki karakteristik bentuk daun memanjang dengan panjang 11,3 hingga 13,0 cm dan lebar daun 4,7 hingga 5,0 cm.

4. LIMITASI

Saat mengumpulkan dataset citra daun cabai, beberapa limitasi yang dihadapi meliputi:

- Variasi Jumlah Gambar: Jumlah gambar untuk setiap varietas atau usia daun mungkin tidak konsisten, yang dapat mempengaruhi hasil analisis.
- Kualitas Gambar: Kualitas gambar bisa bervariasi tergantung pada kondisi pencahayaan, fokus kamera, dan metode pengambilan gambar.
- Kondisi Lingkungan: Faktor lingkungan seperti cuaca, waktu pengambilan gambar, dan kondisi tanaman dapat memengaruhi penampilan daun dalam gambar.

5. UCAPAN TERIMAKASIH

Kami mengucapkan terima kasih kepada kelompok riset Sains Data dan Informasi yang telah memberikan akses untuk menggunakan data hasil pengamatan penelitian berupa citra daun cabai, yang merupakan hasil pengumpulan data di kebun perakitan varietas cabai di BALITSA Lembang.

6. CREDIT (CONTRIBUTOR ROLES TAXONOMY)

Wiwin Suwarningsih: Methodology, Writing, and Reviewing.

Evandri: Curating, writing and Editing

Rinda Kirana: ahli penyelia, konsultasi pakar

Purnomo Husnul Khotimah, Dianadewi Riswantini, Ekasari Nugraheni, Andri Fachrur Rozie, Andria Arisal, Devi Munandar: collecting data

Noor Roufiq Ahmadi: menyediakan lahan tanaman, konsultan metoda

DAFTAR PUSTAKA

- Doshi, J., Patel, T., & Bharti, S. K. (2019). Smart farming using IoT, a solution for optimally monitoring farming conditions. *Procedia Computer Science*, *160*, 746-751. <https://doi.org/10.1016/j.procs.2019.11.016>
- Swami Durai, S. K., & Shamili, M. D. (2022). Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, *3*, 100041. <https://doi.org/10.1016/j.dajour.2022.100041>
- Hashim, A., Rafii, M. Y., Yusuff, O., Harun, A. R., Juraimi, S., Misran, A., Chukwu, S. C., Arolu, F., & Kadar, A. I. (2024). Genetic consequences of chronic gamma irradiation on agro morphological traits in chili under hydrogel enhance media. *Heliyon*, *10(4)*, e25111. <https://doi.org/10.1016/j.heliyon.2024.e25111>
- Paredes Andrade, N. J., Monteros-Altamirano, A., Tapia Bastidas, C. G., & Sørensen, M. (2020). Morphological, sensorial and chemical characterization of chilli peppers (*Capsicum* spp.) from the CATIE genebank. *Agronomy*, *10(11)*, 1732. <https://doi.org/10.3390/agronomy10111732>