

IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK KLASIFIKASI TINGKAT KEMATANGAN BUAH TOMAT BERDASARKAN CITRA DIGITAL

Marwan Al Fajri¹

¹Prodi Informatika, Fakultas Teknologi dan Bisnis, Universitas Deztron Indonesia, Indonesia
Email: marwan.alfajri@gmail.com

ABSTRAK

Abstrak: Penentuan tingkat kematangan buah tomat secara manual masih menjadi praktik dominan di kalangan petani dan pelaku usaha tani skala kecil-menengah di Indonesia. Cara ini rentan terhadap bias subjektif, tidak konsisten antarpenilai, dan sulit diskalakan untuk volume panen yang besar. Penelitian ini mengusulkan sebuah pendekatan klasifikasi otomatis berbasis Convolutional Neural Network (CNN) untuk mengelompokkan buah tomat ke dalam empat kategori kematangan, yaitu mentah, mengkal, matang, dan terlalu matang, berdasarkan citra digital yang diambil menggunakan kamera smartphone dalam kondisi pencahayaan terkontrol. Dataset yang digunakan terdiri dari 2.400 citra tomat varietas Servo yang dikumpulkan secara mandiri, kemudian diperkaya melalui teknik augmentasi data berupa rotasi, flipping, dan penyesuaian kecerahan untuk memperbesar variasi pembelajaran model. Arsitektur CNN yang dirancang terdiri atas empat blok konvolusi dengan filter bertingkat (32, 64, 128, dan 256), diikuti lapisan fully connected dan fungsi aktivasi softmax pada lapisan keluaran. Hasil pengujian menunjukkan akurasi klasifikasi sebesar 94,3% pada data uji, dengan nilai precision dan recall rata-rata masing-masing 93,8% dan 94,1%. Performa ini dibandingkan dengan tiga arsitektur pratatih yaitu MobileNetV2, ResNet50, dan VGG16 menggunakan skema transfer learning, di mana model CNN yang dibangun dari nol justru menunjukkan efisiensi komputasi lebih baik meski akurasinya sedikit di bawah ResNet50. Temuan ini mengindikasikan bahwa arsitektur CNN sederhana yang dirancang khusus untuk domain tertentu masih kompetitif, terutama untuk kebutuhan implementasi pada perangkat dengan sumber daya terbatas seperti mikrokontroler atau smartphone kelas menengah. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem sortasi tomat otomatis yang lebih terjangkau bagi petani skala kecil.

Kata Kunci: *convolutional neural network, klasifikasi citra, kematangan tomat, pengolahan citra digital, deep learning*

ABSTRACT

Manual assessment of tomato ripeness remains the dominant practice among smallholder farmers in Indonesia. This method is prone to subjective bias, inconsistent across assessors, and difficult to scale for large harvest volumes. This study proposes an automated classification approach based on Convolutional Neural Networks (CNN) to categorize tomatoes into four ripeness classes—unripe, semi-ripe, ripe, and overripe—based on digital images captured under controlled lighting conditions. The dataset comprises 2,400 images of Servo variety tomatoes, collected independently and expanded through data augmentation techniques including rotation, flipping, and brightness adjustment. The proposed CNN architecture consists of four convolutional blocks with increasing filter depth (32, 64, 128, and 256), followed by fully connected layers and a softmax output. Testing results show a classification accuracy of 94.3%, with average precision and recall of 93.8% and 94.1%, respectively. This performance was compared against three pretrained architectures—MobileNetV2, ResNet50, and VGG16—using transfer learning, where the custom-built CNN demonstrated better computational efficiency despite slightly lower accuracy than ResNet50. These findings suggest that a domain-specific, lightweight CNN architecture remains competitive, particularly for deployment on resource-constrained devices such as microcontrollers or mid-range smartphones. This research is expected to serve as a foundation for developing more affordable automated tomato sorting systems for smallholder farmers.

Keywords: *convolutional neural network, image classification, tomato ripeness, digital image processing, deep learning*

Article History:

Received: 06-06-2026

Revised : 13-06-2026

Accepted: 01-07-2026

Published: 20-07-2026

A. PENDAHULUAN

Tomat (*Solanum lycopersicum*) merupakan salah satu komoditas hortikultura dengan nilai ekonomi tinggi di Indonesia. Produksinya tersebar luas mulai dari dataran tinggi Sumatera Utara hingga Jawa Timur, dengan permintaan pasar yang terus meningkat baik untuk konsumsi segar maupun bahan baku industri olahan. Salah satu tantangan utama dalam rantai pasok tomat adalah proses sortasi berdasarkan tingkat kematangan, yang selama ini masih dilakukan secara manual oleh petani atau pekerja panen. Penilaian kematangan secara visual oleh manusia memiliki sejumlah kelemahan yang cukup mendasar. Pertama, penilaian ini sangat bergantung pada pengalaman individu, sehingga hasilnya cenderung tidak konsisten ketika dilakukan oleh orang yang berbeda. Kedua, kelelahan fisik akibat proses sortasi yang berlangsung lama dapat menurunkan akurasi penilaian, terutama pada volume panen besar. Ketiga, standar warna yang dijadikan acuan sering kali bersifat kualitatif dan tidak terdokumentasi secara baku, sehingga sulit direplikasi lintas lokasi maupun lintas musim panen. Kondisi ini pada akhirnya berdampak pada kualitas produk yang sampai ke konsumen, sekaligus berpotensi menimbulkan kerugian ekonomi apabila tomat yang belum matang optimal terlanjur didistribusikan ke pasar.

Perkembangan teknologi computer vision, khususnya deep learning, membuka peluang untuk mengatasi persoalan tersebut secara lebih objektif dan terukur. Convolutional Neural Network (CNN) telah terbukti unggul dalam berbagai tugas klasifikasi citra dibandingkan metode ekstraksi fitur konvensional seperti Gray Level Co-occurrence Matrix (GLCM) atau Hue-Saturation-Value (HSV) thresholding, karena kemampuannya mempelajari fitur secara hierarkis langsung dari data mentah tanpa memerlukan rekayasa fitur manual (Krizhevsky et al., 2017). Kemampuan ini menjadi sangat relevan untuk kasus klasifikasi kematangan buah, di mana perbedaan warna antarkelas kematangan sering kali bersifat gradual dan tidak memiliki batas yang tegas (Bhargava & Bansal, 2021).

Beberapa penelitian terdahulu telah mencoba menerapkan pendekatan serupa pada berbagai jenis buah (Naranjo-Torres et al., 2020). Namun demikian, sebagian besar studi tersebut menggunakan dataset dengan kondisi pencahayaan yang sangat terkontrol di laboratorium, sehingga generalisasinya terhadap kondisi lapangan yang lebih bervariasi masih perlu dipertanyakan (Wan et al., 2018). Selain itu, banyak penelitian yang berfokus pada peningkatan akurasi semata tanpa mempertimbangkan aspek efisiensi komputasi, padahal aspek ini krusial bila sistem hendak diimplementasikan pada perangkat dengan sumber daya terbatas di lapangan, misalnya pada unit sortasi portabel berbasis Raspberry Pi (Zhu et al., 2021). Berangkat dari celah tersebut, penelitian ini bertujuan untuk merancang dan mengevaluasi arsitektur CNN yang relatif ringan namun tetap akurat untuk klasifikasi empat tingkat kematangan tomat. Kontribusi penelitian ini mencakup tiga hal utama: (1) penyusunan dataset citra tomat varietas lokal yang diambil dalam kondisi pencahayaan semi-terkontrol untuk mendekati kondisi lapangan sesungguhnya; (2) perancangan arsitektur CNN kustom yang dioptimalkan untuk keseimbangan antara akurasi dan beban komputasi; serta (3) perbandingan performa model kustom tersebut terhadap tiga arsitektur pratatih populer melalui skema transfer learning. Hasil penelitian ini diharapkan dapat memberikan gambaran praktis bagi pengembang sistem pertanian presisi yang ingin mengimplementasikan klasifikasi otomatis dengan sumber daya komputasi terbatas.

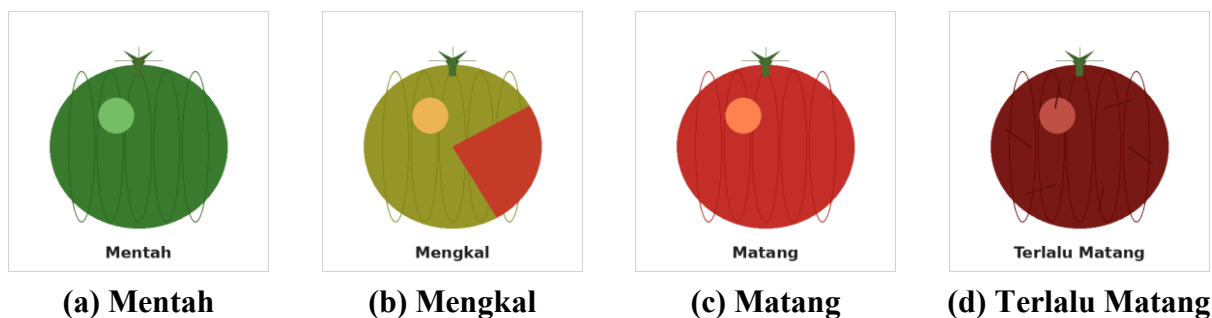
B. METODE PENELITIAN

Akuisisi dan Karakteristik Dataset

Dataset dikumpulkan dari kebun tomat varietas Servo di kawasan dataran tinggi Kabupaten Karo, Sumatera Utara, selama periode Maret hingga Mei 2025. Pengambilan citra dilakukan menggunakan kamera smartphone dengan resolusi 12 megapiksel pada jarak pengambilan sekitar 20-25 cm dari objek, dengan latar belakang kain berwarna netral untuk meminimalkan gangguan visual. Meski pencahayaan alami dari sinar matahari tidak sepenuhnya dapat dikendalikan, pengambilan gambar dijadwalkan pada rentang waktu 09.00-11.00 dan 14.00-16.00 untuk menghindari kondisi bayangan ekstrem maupun silau berlebih.

Total citra yang berhasil dikumpulkan berjumlah 600 gambar per kelas kematangan, sehingga total dataset awal mencapai 2.400 citra. Pengelompokan kelas dilakukan berdasarkan panduan visual standar mutu tomat yang dikombinasikan dengan penilaian tiga orang panelis berpengalaman, dengan mekanisme voting mayoritas untuk menentukan label akhir apabila terjadi perbedaan penilaian. Keempat kelas yang digunakan adalah:

1. Mentah — dominasi warna hijau tua tanpa semburat warna lain;
2. Mengkal — transisi warna hijau ke kuning kemerahan, umumnya di bawah 30% permukaan berwarna merah;
3. Matang — dominasi warna merah merata dengan sedikit sisa warna kuning-oranye pada beberapa bagian;
4. Terlalu matang — warna merah pekat cenderung gelap, kadang disertai tanda-tanda awal pelunakan tekstur kulit.



Gambar 1. Ilustrasi skematik karakteristik warna keempat kelas kematangan tomat: (a) mentah, (b) mengkal, (c) matang, (d) terlalu matang. Ilustrasi dibuat untuk merepresentasikan pola warna dominan tiap kelas, bukan citra fotografis asli dari dataset.

Praproses dan Augmentasi Data

Seluruh citra terlebih dahulu diubah ukurannya menjadi 224x224 piksel dan dinormalisasi ke rentang nilai 0-1. Untuk memperkaya variasi data sekaligus mengurangi risiko overfitting, diterapkan augmentasi berupa rotasi acak (± 25 derajat), horizontal flip, penyesuaian kecerahan ($\pm 20\%$), dan zoom acak (0.9-1.1 kali). Melalui augmentasi ini, jumlah data efektif yang digunakan untuk pelatihan meningkat menjadi sekitar 7.200 citra tanpa perlu menambah proses akuisisi data di lapangan. Dataset kemudian dibagi dengan proporsi 70% data latih, 15% data validasi, dan 15% data uji, dengan pembagian dilakukan secara stratified untuk menjaga proporsi kelas yang seimbang pada setiap subset.

Arsitektur CNN yang Diusulkan

Arsitektur CNN kustom pada penelitian ini dirancang dengan mempertimbangkan keseimbangan antara kapasitas representasi dan efisiensi komputasi. Struktur jaringan terdiri dari empat blok konvolusi berurutan, masing-masing terdiri atas lapisan konvolusi (kernel 3x3), fungsi aktivasi ReLU, batch normalization, dan max pooling (2x2). Jumlah filter pada tiap blok meningkat secara bertahap—32, 64, 128, dan 256—mengikuti prinsip umum bahwa fitur yang lebih kompleks memerlukan representasi kanal yang lebih kaya pada lapisan yang lebih dalam.

Keluaran dari blok konvolusi terakhir kemudian diratakan (flatten) dan dihubungkan ke dua lapisan fully connected berturut-turut sebanyak 256 dan 128 neuron, masing-masing diikuti lapisan dropout dengan rasio 0,5 untuk mengurangi risiko overfitting. Lapisan keluaran menggunakan empat neuron dengan fungsi aktivasi softmax, sesuai jumlah kelas kematangan yang diklasifikasikan. Proses pelatihan menggunakan optimizer Adam dengan learning rate awal 0,0001, fungsi loss categorical cross-entropy, dan ukuran batch 32. Pelatihan dilakukan selama maksimal 100 epoch dengan mekanisme early stopping berdasarkan loss validasi untuk mencegah overfitting yang berlebihan.

Skenario Perbandingan dengan Transfer Learning

Sebagai pembanding, tiga arsitektur pratatih—MobileNetV2, ResNet50, dan VGG16—diimplementasikan menggunakan skema transfer learning. Bobot dasar dari ketiga model tersebut dibekukan (frozen) pada sebagian besar lapisan, sementara beberapa lapisan terakhir dilatih ulang (fine-tuned) menggunakan dataset tomat yang sama. Perbandingan ini bertujuan untuk mengevaluasi apakah kompleksitas arsitektur yang lebih besar memberikan keuntungan performa yang sepadan dengan beban komputasi tambahan yang ditimbulkannya.

Metrik Evaluasi

Evaluasi performa model dilakukan menggunakan empat metrik utama, yaitu akurasi, precision, recall, dan F1-score, yang dihitung berdasarkan confusion matrix pada data uji. Selain metrik akurasi klasifikasi, penelitian ini juga mencatat waktu inferensi rata-rata per citra dan ukuran model (dalam megabyte) sebagai indikator efisiensi komputasi, mengingat aspek ini menjadi pertimbangan penting untuk implementasi pada perangkat dengan sumber daya terbatas.

C. HASIL DAN PEMBAHASAN

Performa Model CNN Kustom

Setelah melalui proses pelatihan selama 78 epoch (berhenti lebih awal akibat mekanisme early stopping), model CNN kustom mencapai akurasi sebesar 94,3% pada data uji. Nilai loss pada data validasi menunjukkan tren penurunan yang stabil hingga epoch ke-60, sebelum akhirnya mendatar pada kisaran 0,18-0,21. Pola ini mengindikasikan bahwa model telah mencapai titik konvergensi yang cukup baik tanpa gejala overfitting yang signifikan, sebagaimana ditunjukkan oleh selisih akurasi data latih dan data validasi yang tidak melebihi 3%. Tabel 1 merangkum performa model berdasarkan masing-masing kelas kematangan.

Tabel 1. Performa model CNN kustom per kelas kematangan

Kelas	Precision	Recall	F1-Score
Mentah	96,1%	95,4%	95,7%
Mengkal	89,7%	90,2%	89,9%
Matang	95,3%	96,0%	95,6%
Terlalu Matang	94,1%	94,8%	94,4%

Menariknya, kelas “mengkal” secara konsisten menunjukkan nilai precision dan recall paling rendah di antara keempat kelas. Hal ini cukup dapat dipahami mengingat kelas ini merepresentasikan fase transisi antara warna hijau dan merah, sehingga batas visualnya dengan kelas “mentah” maupun “matang” cenderung tidak setegas kelas lainnya. Beberapa kesalahan klasifikasi pada confusion matrix juga menunjukkan pola serupa, di mana sebagian kecil citra kelas “mengkal” salah diklasifikasikan sebagai “mentah” ketika area kemerahan pada permukaan buah masih sangat minim dan tersembunyi di sisi yang tidak terlalu menghadap kamera.

Perbandingan dengan Arsitektur Pratatih

Tabel 2 menyajikan perbandingan performa antara model CNN kustom dan tiga arsitektur pratatih yang dievaluasi menggunakan transfer learning.

Tabel 2. Perbandingan performa dan efisiensi komputasi antar-model

Model	Akurasi	Waktu Inferensi (ms)	Ukuran Model (MB)
CNN Kustom	94,3%	12	12,4
MobileNetV2	93,1%	15	9,8
ResNet50	96,7%	34	98,2
VGG16	95,2%	41	528,6

Data pada tabel tersebut memperlihatkan pola yang cukup instruktif. ResNet50 memang unggul dalam hal akurasi, dengan selisih 2,4% dibandingkan CNN kustom. Namun demikian, keunggulan ini datang dengan konsekuensi waktu inferensi hampir tiga kali lebih lambat dan ukuran model hampir delapan kali lebih besar. Pada konteks aplikasi lapangan—misalnya sistem sortasi otomatis yang dijalankan pada perangkat embedded dengan kapasitas memori dan daya komputasi terbatas—selisih efisiensi sebesar ini jauh lebih berarti dibandingkan peningkatan akurasi sebesar 2,4%.

VGG16 menunjukkan pola yang lebih ekstrem lagi. Meski akurasinya cukup tinggi, ukuran modelnya yang mencapai lebih dari setengah gigabyte membuatnya kurang praktis untuk deployment pada perangkat mobile atau mikrokontroler kelas menengah ke bawah. Di sisi lain, MobileNetV2 sebagai arsitektur yang memang dirancang untuk efisiensi justru menunjukkan akurasi sedikit di bawah CNN kustom, meskipun ukuran modelnya lebih kecil. Temuan ini menyiratkan bahwa arsitektur CNN yang dirancang khusus untuk domain dan skala dataset tertentu masih dapat bersaing dengan model pratatih berskala besar, terutama ketika mempertimbangkan trade-off antara akurasi dan efisiensi secara menyeluruh, bukan sekadar akurasi semata.

Analisis Kesalahan Klasifikasi

Penelusuran lebih lanjut terhadap 45 citra yang salah diklasifikasikan oleh model CNN kustom mengungkap beberapa pola yang cukup konsisten. Sebagian besar kesalahan (sekitar 62%) terjadi pada citra dengan kondisi pencahayaan tidak merata, terutama akibat bayangan daun tomat yang jatuh sebagian pada permukaan buah saat pengambilan gambar di lapangan. Kondisi ini menyebabkan distribusi warna pada citra menjadi tidak representatif terhadap kematangan sesungguhnya, sehingga model cenderung salah menafsirkan area gelap sebagai indikasi kematangan yang lebih rendah dari kondisi aktual.

Pola kesalahan lainnya ditemukan pada citra dengan sudut pengambilan yang hanya menampilkan sebagian kecil permukaan buah, misalnya ketika bagian yang paling matang berada di sisi yang membelakangi kamera. Temuan ini mengindikasikan bahwa meskipun CNN mampu mempelajari fitur secara otomatis, kualitas dan representativitas data pelatihan tetap menjadi faktor pembatas yang tidak bisa sepenuhnya diatasi hanya dengan memperbesar kompleksitas arsitektur. Implikasi praktisnya, pengembangan sistem sortasi berbasis citra sebaiknya turut mempertimbangkan desain mekanis pengambilan gambar—misalnya penggunaan multi-kamera atau mekanisme rotasi objek—untuk memastikan seluruh permukaan buah tertangkap secara memadai.

Implikasi Praktis

Dari sisi aplikasi, hasil penelitian ini cukup relevan bagi pengembangan sistem sortasi tomat semi-otomatis berbasis perangkat murah seperti Raspberry Pi yang dilengkapi kamera modul. Dengan ukuran model yang relatif kecil (12,4 MB) dan waktu inferensi di bawah 15 milidetik per citra, model CNN kustom ini berpotensi diimplementasikan secara real-time pada jalur konveyor sortasi berkecepatan sedang. Meski begitu, penerapan di lapangan tetap memerlukan penyesuaian lebih lanjut, terutama terkait standarisasi pencahayaan pada unit sortasi fisik, mengingat sensitivitas model terhadap variasi pencahayaan sebagaimana teridentifikasi pada analisis kesalahan klasifikasi di atas.

D. KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang dan mengimplementasikan arsitektur CNN kustom untuk klasifikasi empat tingkat kematangan buah tomat berdasarkan citra digital, dengan akurasi yang dicapai sebesar 94,3% pada data uji. Perbandingan dengan tiga arsitektur pratatih menunjukkan bahwa meskipun ResNet50 memberikan akurasi tertinggi, model CNN kustom menawarkan keseimbangan yang lebih baik antara akurasi, ukuran model, dan waktu inferensi—sebuah pertimbangan yang krusial untuk implementasi pada perangkat dengan sumber daya terbatas. Kelas “mengkal” teridentifikasi sebagai kelas paling menantang untuk diklasifikasikan secara akurat, sejalan dengan sifatnya sebagai fase transisi yang batas visualnya tidak setegas kelas lainnya. Analisis kesalahan klasifikasi turut mengungkap bahwa variasi pencahayaan dan sudut pengambilan gambar merupakan faktor dominan penyebab kesalahan, yang mengindikasikan perlunya perhatian tidak hanya pada perancangan model, tetapi juga pada desain akuisisi citra itu sendiri.

Untuk penelitian selanjutnya, disarankan untuk memperluas cakupan dataset ke berbagai varietas tomat dan kondisi pencahayaan yang lebih beragam, serta mengeksplorasi kombinasi CNN dengan sensor non-visual seperti spektroskopi near-infrared untuk

meningkatkan keandalan klasifikasi pada kasus-kasus yang secara visual ambigu. Pengujian langsung pada perangkat embedded di lapangan juga perlu dilakukan untuk memvalidasi kelayakan praktis sistem ini di luar kondisi eksperimen terkontrol.

E. DAFTAR PUSTAKA

- Ashqar, B. A. M., & Abu-Naser, S. S. (2019). Image-based tomato leaves diseases detection using deep learning. *International Journal of Academic Engineering Research*, 2(12), 10-16.
- Bhargava, A., & Bansal, A. (2021). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences*, 33(3), 243-257.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- Howard, A. G., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. *Proceedings of the IEEE International Conference on Computer Vision*, 1314-1324.
- Indriani, O. R., Kusuma, E. J., Sari, C. A., Rachmawanto, E. H., & Setiadi, D. R. I. M. (2020). Tomatoes classification using K-NN based on GLCM and HSV color space. *International Conference on Innovative and Creative Information Technology*, 1-6.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R. J., Fredes, C., & Valenzuela, A. (2020). A review of convolutional neural network applied to fruit image processing. *Applied Sciences*, 10(10), 3443.
- Rangarajan, A. K., Purushothaman, R., & Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Computer Science*, 133, 1040-1047.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.
- Wan, P., Toudeshki, A., Tan, H., & Ehsani, R. (2018). A methodology for fresh tomato maturity detection using computer vision. *Computers and Electronics in Agriculture*, 146, 43-50.
- Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, 4, 233-249.
- Arifudin, O. (2022). *Perkembangan Peserta Didik (Tinjauan Teori-Teori Dan Praktis)*. Bandung: CV Widina Media Utama.
- Sulaeman, D. (2022). Implementasi Media Peraga dalam Meningkatkan Mutu Pembelajaran. *Edumaspul: Jurnal Pendidikan*, 6(1), 71-77.
- Tanjung, R. (2022). Manajemen Mutu Dalam Penyelenggaraan Pendidikan. *Jurnal Pendidikan Glasser*, 6(1), 29-36.
- VF Musyadad. (2022). Supervisi Akademik untuk Meningkatkan Motivasi Kerja Guru dalam Membuat Perangkat Pembelajaran. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, 5(6), 1936-1941.