

# Deteksi Emosi pada Ekspresi Wajah Manusia Menggunakan *Convolutional Neural Network (CNN)* Berbasis *Android*

1<sup>st</sup> Kristoforus Ferel Viandito

Program Studi Teknologi Informasi  
Universitas Telkom, Kampus Surabaya

Surabaya, Jawa Timur, Indonesia

[viandito@student.telkomuniversity.ac.id](mailto:viandito@student.telkomuniversity.ac.id)

d

2<sup>nd</sup> Farah Zakiyah Rahmantti

Program Studi Teknologi Informasi  
Universitas Telkom, Kampus Surabaya

Surabaya, Jawa Timur, Indonesia

[farahzakiyah@telkomuniversity.ac.id](mailto:farahzakiyah@telkomuniversity.ac.id)

3<sup>rd</sup> Yohanes Setiawan

Program Studi Teknologi Informasi  
Universitas Telkom, Kampus Surabaya

Surabaya, Jawa Timur, Indonesia

[yohanessetiawan@telkomuniversity.ac.id](mailto:yohanessetiawan@telkomuniversity.ac.id)

**Abstrak** — Pada era teknologi digital saat ini, teknologi pengolahan citra menjadi salah satu yang berkembang pesat. Hal ini bersamaan juga dengan perkembangan pesat pada teknologi bidang *Artificial Intelligence (AI)* yang relevan pada bidang teknologi pengenalan wajah dan deteksi emosi. Penelitian ini mengembangkan sebuah sistem deteksi emosi pada wajah dengan menggunakan *Convolutional Neural Network (CNN)* yang didesain untuk mengolah data dua dimensi yang dijalankan pada sistem *Android*. Dengan mengetahui ekspresi emosi seseorang, maka seseorang dapat dinilai kondisi psikologinya dengan bantuan mesin. Sistem ini dapat digunakan pada beberapa bidang seperti bidang kesehatan. Dengan sistem deteksi emosi maka tenaga medis dapat memantau dan menentukan pemberian dukungan apa yang dibutuhkan pasien. Model pada sistem penelitian ini dijalankan dengan model *CNN* dengan konfigurasi hyperparameter batch size 256, jumlah *epoch* 76, dan *dropout rate* sebesar 50%, serta menggunakan *optimizer Adam*. Hasil dari pengujian model ini menunjukkan angka akurasi sebesar 84% yang cukup baik. Melalui penelitian ini, diharapkan memberi kontribusi dalam pengembangan teknologi citra digital untuk deteksi emosi manusia menggunakan *CNN* yang diimplementasikan pada aplikasi *Android*. Diharapkan juga dengan penelitian ini dapat mendorong motivasi penelitian dengan teknologi yang serupa di masa depan.

**Kata kunci**— *Android, CNN, AI, hyperparameter*.

## I. PENDAHULUAN

Pada era digital yang semakin berkembang ini, salah satu bidang yang paling berkembang pesat adalah teknologi citra digital. Salah satu bidang dari teknologi ini adalah pengenalan wajah dan deteksi emosi. Teknologi ini menjadi dibutuhkan karena dengan adanya teknologi ini dapat melakukan deteksi ekspresi emosi seseorang. Dengan mengetahui ekspresi seseorang, maka dapat dinilai kondisi psikologi dari orang tersebut dengan mesin[1]. Maka dari itu, teknologi pendekripsi emosi menjadi relevan untuk dimanfaatkan pada bidang kesehatan. Gejala gangguan kesehatan mental seperti depresi, kecemasan, dan *post-traumatic stress disorder (PTSD)* dapat tampak dari ekspresi wajah dan perilaku. Teknologi deteksi emosi akan menjadi alat yang dapat memantau perubahan emosi pasien dengan baik yang berguna bagi tenaga medis dalam memantau

kondisi pasien pada sesi konsultasi. Hal ini dapat membantu tenaga medis melakukan pertimbangan keputusan dalam melakukan dukungan apa yang diperlukan untuk pasien dengan lebih baik.

Oleh sebab itu, peneliti membuat sebuah sistem deteksi emosi pada ekspresi wajah manusia dengan menggunakan pendekatan *Convolutional Neural Network (CNN)*. *CNN* adalah sebuah pengembangan dari *Multilayer Perceptron (MLP)* yang didesain untuk mengolah data dua dimensi [2]. *CNN* pertama kali diciptakan pada tahun 1960-an[3] dan telah terbukti menghasilkan kinerja yang menjanjikan dalam sistem *Computer Vision*[4]. *CNN* telah menjadi arsitektur yang paling representatif dalam sebuah sistem *Deep Learning*[5]. *CNN* dapat digunakan dalam tugas visual yang rumit dengan komputasi yang tinggi, terutama dalam melakukan klasifikasi citra[6], segmentasi, deteksi objek, pengolahan video, pengolahan bahasa alami, dan pengenalan suara[7]. Maka dari itu, dengan menggunakan *CNN* memungkinkan untuk mengimplementasikan sebuah sistem yang dapat mendekripsi emosi seseorang berdasarkan ekspresi wajah manusia. Untuk memudahkan dalam penggunaannya, aplikasi ini dibuat dalam basis *Android*.

Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana membangun model deteksi emosi berbasis *CNN* yang dapat diintegrasikan dengan perangkat *Android*, serta seberapa tinggi akurasi yang dapat dicapai oleh sistem tersebut dalam mengenali ekspresi emosi dari citra wajah. Tujuan dari penelitian ini adalah mengembangkan model *CNN* untuk deteksi ekspresi emosi dengan akurasi di atas 80%, serta menyajikan fondasi awal sistem yang dapat digunakan dalam berbagai skenario nyata.

Penelitian ini bertujuan untuk memberikan kontribusi dalam pengembangan ilmu teknologi citra digital pengenalan emosi pada ekspresi wajah yang dapat diimplementasikan dalam kehidupan sehari-hari jika menggunakan sistem ini. Melalui deteksi emosi pada ekspresi wajah manusia dengan menggunakan *CNN* berbasis *Android*, sistem yang dikembangkan mampu mengenali emosi dari ekspresi wajah dalam berbagai kondisi tertentu seperti kondisi pencahayaan yang berbeda. Diharapkan juga hasil dari penelitian ini dapat membantu dalam bidang lainnya seperti dapat digunakan dalam praktik bidang kesehatan.

## II. KAJIAN TEORI

### A. Deep Learning

*Deep Learning* bidang dalam *Machine Learning* yang berkembang pesat, terutama dengan perkembangan teknologi *GPU acceleration*. *Deep Learning* memiliki keunggulan dalam menjalankan *computer vision*[2].

### B. Artificial Intelligence

*Artificial Intelligence (AI)* atau kecerdasan buatan adalah sistem komputer yang dibuat untuk mampu melakukan tugas yang biasanya memerlukan kecerdasan manusia[8]. Dalam perkembangannya, *AI* memiliki efek signifikan di berbagai industri seperti otomotif. Dimana *AI* dikembangkan untuk sistem mobil *self-driving* yang memungkinkan mobil bisa berjalan dengan otomatis melalui kombinasi sensor, *machine learning*, dan *computer vision*.

### C. Convolutional Neural Network

*Convolutional Neural Network (CNN)* adalah algoritma pada *deep learning* yang biasa digunakan pada data dua dimensi. *CNN* juga memiliki arsitektur dimana terdapat fitur untuk ekstraksi dan klasifikasi untuk melakukan data training.

### D. Confusion Matrix

*Confusion matrix* adalah perhitungan dari nilai akurasi machine learning. Dengan *confusion matrix*, informasi hasil klasifikasi yang telah dilakukan oleh sistem dapat ditemukan dalam model. Ada beberapa matriks performansi yang sering digunakan, diantaranya adalah *accuracy*, *precision*, dan *recall*[9]. Berikut adalah rumus ketiganya:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

#### Keterangan:

TP adalah nilai pada hasil model yang benar dalam melakukan prediksi kelas positif.

TN adalah nilai pada hasil model yang benar dalam melakukan prediksi kelas negatif.

FP adalah nilai hasil pada model yang salah dalam memprediksi kelas positif.

FN adalah nilai hasil pada model yang salah dalam memprediksi kelas negatif.

## III. METODE

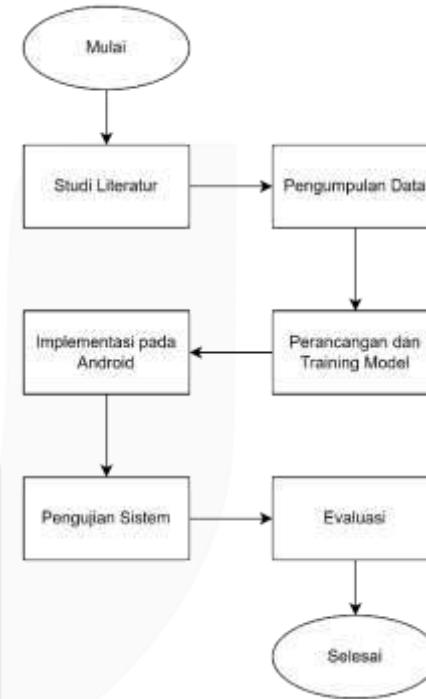
Penelitian dimulai dengan studi literatur untuk meningkatkan aspek-aspek teoritis dan aspek-aspek manfaat praktis dari penelitian[10]. Selain itu, studi literatur digunakan juga untuk identifikasi masalah yang akan diselesaikan pada penelitian ini.

Penelitian dilanjutkan dengan mengumpulkan data berupa dataset publik, yang menggunakan dataset utama *Face Expression Recognition Dataset* oleh Jonathan Oheix dan dataset kedua *FER-2013* oleh Manas Sambare yang keduanya memiliki total 57.528 *data training*, 7.066 *data validation*, dan 7.178 *data test*. Keduanya didapat melalui situs *Kaggle* dan berisikan citra wajah beresolusi 48x48

piksel yang memuat tujuh kategori ekspresi wajah, yaitu marah, jijik, takut, bahagia, netral, sedih, dan terkejut.

Setelah itu, penelitian dilanjutkan dengan perancangan dan *training model* dengan *ResNet50*, model diuji dalam beberapa kali percobaan yang akan menentukan *hyperparameter* seperti ukuran citra, *batch size*, *epoch*, *learning rate*, dan *dropout rate*, yang kemudian akan dipilih satu model yang memiliki performa terbaik berdasarkan *confusion matrix* dan *classification report*. Setelah model selesai ditrain, kemudian model diimplementasikan dengan disambungkan pada *Android* melalui *API* yang sistemnya dibangun dengan *Flutter*.

Penelitian dilanjutkan dengan pengujian sistem yang dilakukan di dunia nyata untuk menguji seberapa baik model mengidentifikasi ketujuh ekspresi. Pengujian dilakukan di lingkungan sekolah TK SD Katolik Santa Maria II Sidoarjo dengan mengambil citra murid dalam proses belajar mengajar. Tahap terakhir dilakukan dengan evaluasi berdasarkan hasil pengujian untuk menilai seberapa baik sistem dalam mengenali emosi pada ekspresi wajah murid. Pada gambar 1 ditampilkan *flowchart* yang menjadi prosedur pada penelitian ini.



GAMBAR 1  
(FLOWCHART PENELITIAN)

## IV. HASIL DAN PEMBAHASAN

### A. Preprocessing

Data yang digunakan dalam penelitian ini diambil dari dataset publik yang didapat dari situs web *Kaggle*, yang meliputi *Face Expression Recognition Dataset* oleh Jonathan Oheix dan dataset kedua *FER-2013* oleh Manas Sambare. Kedua dataset tersebut dipilih karena sudah memiliki berbagai variasi wajah seperti umur, jenis kelamin, dan bentuk wajah, sehingga kedua dataset ini cocok digunakan dalam model untuk pengenalan berbagai emosi pada ekspresi

wajah, tidak hanya untuk lingkungan sekolah, tapi juga bisa digunakan dalam berbagai penerapan dunia nyata. Dalam *preprocessing*, dilakukan augmentasi data dengan *rotate*, *zoom*, dan *flip* secara horizontal, hal ini dilakukan untuk menambah variasi supaya model bisa mengenali lebih banyak kondisi dalam penerapannya. Selain itu, citra dalam *dataset* juga diperbesar ke 128x128 piksel untuk memperjelas detail guna meningkatkan akurasi hasil.

### B. Model Training

Pada tahap *model training*, *dataset* yang telah melalui proses *preprocessing* digunakan untuk dilatih dalam arsitektur *CNN*, yaitu *ResNet50*. *ResNet50* dipilih karena memiliki hasil akurasi paling tinggi saat dibandingkan dengan model *CNN* lain seperti *ResNet18* dan *ResNet34*.

Selama proses *model training*, beberapa *hyperparameter* ditetapkan untuk memastikan bahwa model dapat berfungsi dengan optimal dan proses *training* berjalan secara efisien. Pemilihan *hyperparameter* yang tepat sangat penting karena hal ini merupakan salah satu yang paling mempengaruhi hasil akhir dari model. *Hyperparameter* yang dicoba dalam berbagai konfigurasi dalam penelitian ini adalah nilai ukuran citra, *batch size*, dan *epoch*. Sedangkan *learning rate* disamakan sebesar 0.001, yang berlaku juga pada *dropout rate* sebesar 50% dengan *optimizer Adam*. Beberapa konfigurasi *hyperparameter* yang dicoba dalam proses *training* dapat dilihat pada tabel 1. Percobaan ketiga dipilih menjadi model dalam penelitian ini karena memiliki angka akurasi tertinggi, yaitu 84%.

TABEL 1  
(KONFIGURASI TRAINING)

Percobaan ke-	Batch Size	Epoch	Ukuran Citra	Akurasi
1	128	13	48x48	67%
2	256	56	128x128	80%
3	256	78	128x128	84%

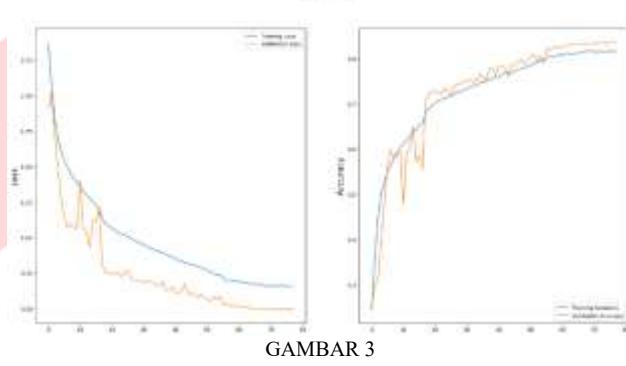
TABEL 2  
(CLASS WEIGHT)

No	Kelas Ekspresi	Class Weight
1	Marah	150%
2	Jijik	150%
3	Takut	200%
4	Bahagia	80%
5	Netral	100%
6	Sedih	150%
7	Terkejut	100%

Dalam *training model* juga dilakukan penambahan *class weight* yang komposisinya dapat dilihat pada tabel 2. *Class weight* digunakan untuk mengatur porsi *training* setiap class pada model. Dalam konteks model yang dibuat, *weight* pada ekspresi bahagia diturunkan karena jumlah data pada ekspresi bahagia yang dominan, sehingga fokus pada proses class ini dapat diturunkan untuk mengalihkan fokus pada jenis ekspresi yang ingin ditingkatkan, yaitu marah, jijik, sedih, dan takut. Hal ini menjadi solusi terhadap ketidakseimbangan jumlah data citra pada tiap *class* yang ada dalam *dataset*.

	precision	recall	f1-score	support
angry	0.82	0.78	0.80	1918
disgust	0.97	0.89	0.93	222
fear	0.88	0.79	0.79	2842
happy	0.93	0.93	0.93	3599
neutral	0.78	0.79	0.79	2449
sad	0.75	0.80	0.77	2386
surprise	0.91	0.87	0.89	1628
accuracy			0.84	14244
macro avg	0.85	0.84	0.84	14244
weighted avg	0.84	0.84	0.84	14244

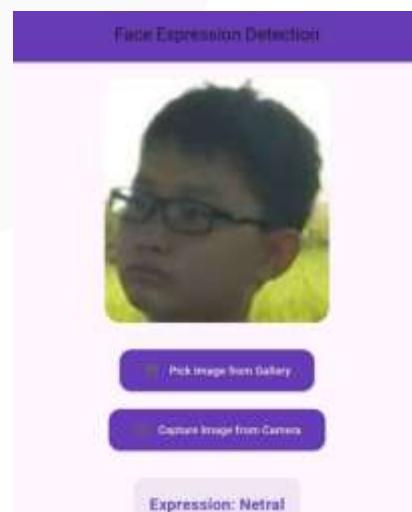
GAMBAR 2  
(CLASSIFICATION REPORT)



GAMBAR 3  
(GRAFIK LOSS DAN ACCURACY)

Dalam hasil *classification report* pada gambar 2 menunjukkan bahwa performa model sudah baik dengan keseluruhan angka berada di 75% keatas. Model berhasil menunjukkan performa yang baik dengan angka akurasi keseluruhan 84%, yang berarti model ini dapat digunakan dalam implementasi dunia nyata. Adapun grafik *loss* dan *accuracy* pada gambar 3 yang menunjukkan pergerakan garis yang menggambarkan bahwa *model training* berjalan dengan baik dan minim *overfitting*.

### C. Pengujian Dunia Nyata



GAMBAR 4  
(CONTOH TAMPILAN APLIKASI ANDROID DALAM PERCOBAAN DUNIA NYATA)

Pengujian sistem secara nyata dilakukan di lingkungan TK SD Katolik Santa Maria II Sidoarjo sebagai uji coba penggunaan sistem dalam konteks kesehatan dalam pendidikan. Tujuan utamanya adalah untuk melihat seberapa baik sistem berfungsi dalam mendeteksi ekspresi wajah siswa di kondisi nyata. Pengujian dilakukan dengan mengambil sampel sebanyak 50 citra murid saat proses kegiatan belajar mengajar yang kemudian diproses pada program. Setelah 50 citra diproses menggunakan sistem ini seperti pada gambar 4.

TABEL 3  
(HASIL PENGUJIAN DUNIA NYATA)

No	Emosi	Jumlah	Percentase
1	Bahagia	19	38%
2	Netral	16	32%
3	Sedih	6	12%
4	Takut	2	4%
5	Marah	2	4%
6	Terkejut	4	8%
7	Jijik	1	2%

Dapat dilihat pada tabel 3 bahwa ekspresi bahagia adalah yang paling dominan, diikuti oleh ekspresi netral yang mencerminkan proses belajar mengajar yang positif. Ekspresi sedih muncul dalam jumlah yang signifikan, namun masih dalam batas wajar. Sedangkan terdapat 4 ekspresi takut dan marah, yang dapat menjadi perhatian pihak guru untuk diamati secara berkala.

#### D. Analisis Hasil

Berdasarkan *classification report*, model menunjukkan performa yang baik dan seimbang pada semua metrik evaluasi dengan seluruh skor *precision* dan *recall* berada di atas 75%, yang menandakan bahwa model tidak hanya mampu mengenali ekspresi dengan baik, tetapi juga mampu menghindari kesalahan klasifikasi terhadap ekspresi lain. *F1-score* juga menunjukkan skor yang baik dengan angka yang paling rendah adalah 77%.

Pada *precision*, performa tertinggi model ada pada kelas jijik dan bahagia sebesar 97% dan 93%, yang berarti model sangat jarang salah dalam memprediksi ekspresi tersebut. Kelas dengan *precision* terendah adalah sedih yang sebesar 75%, yang menandakan adanya beberapa kesalahan minor prediksi ke kelas ini dari ekspresi lainnya.

Pada *recall*, performa tertinggi model ada pada kelas bahagia dan terkejut sebesar 93% dan 87%, menandakan bahwa sebagian besar data berhasil dikenali oleh model dengan baik. *Recall* terendah berada pada kelas marah sebesar 78%, menunjukkan bahwa sebagian kecil ekspresi marah gagal dikenali dan teridentifikasi sebagai kelas lain.

Pada *f1-score*, performa tertinggi model ada pada kelas bahagia dan jijik dengan keduanya sebesar 93%. Sedangkan nilai terendah ada pada kelas sedih sebesar 77%, yang menunjukkan pada kelas itu masih belum sepenuhnya optimal, namun sudah cukup baik.

Jika dilihat dari sisi sensitivitas terhadap data baru, nilai *precision* yang cukup tinggi menunjukkan bahwa model tidak mudah salah mengklasifikasikan ekspresi lain sebagai satu

kelas tertentu. Ini menjadi indikasi bahwa model cenderung stabil. Selain itu, nilai *macro average* dan *weighted average* keduanya sebesar 84%, yang menandakan bahwa model tidak bias walaupun terdapat beberapa kelas dengan jumlah data yang lebih besar.

Dalam pengujian dunia nyata di lingkungan TK SD Katolik Santa Maria II Sidoarjo, ditemukan beberapa tantangan, yang utamanya dikarenakan faktor yang dapat mempengaruhi kelengkapan fitur wajah. Beberapa citra memiliki pencahayaan yang kurang optimal, sehingga fitur wajah siswa/siswi tidak jelas dan menyebabkan prediksi sistem menjadi kurang akurat. Selain cahaya, adapun faktor lain seperti gerakan saat dipotret, dan wajah yang tertutup oleh tangan atau rambut. Hal-hal tersebut membuat model tidak selalu mengenali ekspresi emosi dengan tepat. Ini menunjukkan bahwa meskipun performa model tinggi secara angka, lingkungan nyata tetap memberikan pengaruh besar terhadap akurasi sistem.

Maka dari itu, skenario penggunaan terbaik dari sistem ini ada pada kondisi pencahayaan yang baik serta ekspresi yang jelas dan tidak tertutup oleh elemen lain seperti masker atau tangan. Dengan kondisi seperti ini, sistem mampu memberikan prediksi yang akurat, sehingga cocok untuk diimplementasikan dalam aplikasi deteksi ekspresi wajah berbasis *Android* yang menggunakan *API* di sisi *backend*.

#### V. KESIMPULAN

Penelitian ini berhasil membuat sistem untuk deteksi ekspresi wajah dengan arsitektur *CNN* melalui aplikasi *Android*. Sistem ini dirancang dengan metode *client-server*, dimana pemrosesan analisis ekspresi dilakukan di *PC server*, sedangkan smartphone *Android* berperan sebagai antarmuka untuk pengguna mengirim citra dan menampilkan hasil ekspresi. Metode ini dipilih untuk mengatasi keterbatasan kemampuan smartphone *Android* dalam menjalankan model. Dengan demikian, sistem dapat dijalankan pada smartphone *Android* berspesifikasi rendah dengan baik tanpa mempengaruhi performa prediksi wajah.

Hasil model menunjukkan angka performa yang baik, yaitu akurasi sebesar 84% sehingga model dapat mengenali ekspresi wajah dengan baik, terutama dalam kondisi ideal. Namun, masih terdapat banyak ruang untuk improvisasi berdasarkan tantangan dari penelitian ini, seperti menambahkan berbagai variasi kondisi dalam *dataset* sehingga dapat mengatasi tantangan kondisi yang tidak ideal seperti dalam pencahayaan yang sangat gelap dan wajah yang tidak jelas seperti tertutup objek.

Sistem juga dapat dikembangkan lebih lanjut agar model dapat dijalankan langsung di perangkat *Android* dengan *TensorFlow Lite (TFLite)*. Hal ini akan membuat sistem dapat dijalankan tanpa bergantung pada ketersediaan *PC server*. Sistem ini juga dapat diperluas fungsinya untuk aplikasi di bidang pendidikan atau lainnya, di mana deteksi emosi secara *real-time* dan dapat tidak hanya citra statis, namun juga dapat mendeteksi video yang dapat mendukung lebih dalam interaksi manusia dan teknologi.

## REFERENSI

- [1] A. L. Sigit Guntoro, E. Julianto, and D. Budiyanto, “Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network,” 2022.
- [2] I. W. E. P. W. A. Y. S. R. Suartika, “Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101,” 2016.
- [3] M. Wu *et al.*, “Prediction of the Remaining Time and Time Interval of Pebbles in Pebble Bed Htrgs Aided by CNN via DEM Datasets,” *Nuclear Engineering and Technology*, vol. 55, no. 1, pp. 339–352, Jan. 2023, doi: 10.1016/j.net.2022.09.019.
- [4] M. M and S. P, “COVID-19 Infection Prediction from CT Scan Images of Lungs Using Iterative Convolution Neural Network Model,” *Advances in Engineering Software*, vol. 173, p. 103214, 2022, doi: <https://doi.org/10.1016/j.advengsoft.2022.103214>.
- [5] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” *IEEE Trans Neural Netw Learn Syst*, vol. 33, no. 12, pp. 6999–7019, 2022, doi: 10.1109/TNNLS.2021.3084827.
- [6] M. K. Bohmrah and H. Kaur, “Classification of COVID-19 Patients Using Efficient Fine-Tuned Deep Learning Densenet Model,” *Global Transitions Proceedings*, vol. 2, no. 2, pp. 476–483, 2021, doi: <https://doi.org/10.1016/j.gltip.2021.08.003>.
- [7] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, “A Survey of The Recent Architectures of Deep Convolutional Neural Networks,” *Artif Intell Rev*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.
- [8] M. Sobron *et al.*, Implementasi Artificial Intelligence Pada System Manufaktur Terpadu. 2021.
- [9] R. J. Gunawan, B. Irawan, and C. Setianingsih, “Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network Dengan Model Arsitektur VGG16,” 2021.
- [10] A. F. Subianto, “Penerapan Gerbang Spektrum Untuk Mengurangi Noise Pada Sistem Penghitung Pembicara,” 2024.