

Application of Transfer Learning Method on Convolutional Neural Network (CNN) to Identify Genuine and Fake Diplomas

Rifa Awaludin^{1*)}, Anggun Fergina²⁾, Gina Purnama Insany³⁾

¹⁾²⁾³⁾ Teknik Informatika, Fakultas Komputer dan Desain, Universitas Nusa Putra

*)Correspondence author: Rifa Awaludin, rifa.awaludin_ti19@nusaputra.ac.id, Sukabumi, Indonesia

DOI : <https://doi.org/10.37012/jtik.v12i1.3295>

Abstract

The authenticity of diplomas plays a crucial role in maintaining the integrity of the education system and ensuring that recognized academic competencies align with an individual's actual achievements. Diplomas are not merely administrative documents, but strategic instruments in job recruitment and professional qualification assessment. However, with increasing educational mobility, document misuse through diploma forgery is becoming increasingly prevalent, potentially undermining public trust in educational institutions. Currently, the verification process is still largely carried out manually through visual inspection of document elements such as layout and stamps. The reliance on the examiner's experience makes this method vulnerable to inconsistencies and human error, especially when dealing with fake diplomas with visual qualities that increasingly resemble genuine documents. Diploma forgery is a problem that impacts the credibility of educational institutions and the validity of academic data. Manual inspection is often inconsistent and time-consuming. This study develops a model for classifying genuine and fake diplomas using a Convolutional Neural Network (CNN) with a transfer learning scheme. The performance of the ResNet50, VGG16, and MobileNetV2 architectures is comparatively analyzed. Data preprocessing included resizing, normalization, and augmentation. Test results showed the ResNet50 architecture achieved optimal performance with 92.63% accuracy, 92.16% precision, 94.00% recall, and 93.07% F1-score. The system was implemented in a Streamlit-based web application to facilitate the verification process.

Keywords: Diploma, Image Classification, Convolutional Neural Network, Transfer Learning, Document Forgery

Abstrak

Keaslian ijazah memiliki peran krusial dalam menjaga integritas sistem pendidikan dan memastikan kompetensi akademik yang diakui sesuai dengan capaian riil individu. Ijazah bukan sekadar dokumen administratif, melainkan instrumen strategis dalam rekrutmen tenaga kerja dan penilaian kualifikasi profesional. Namun, seiring meningkatnya mobilitas pendidikan, penyalahgunaan dokumen melalui pemalsuan ijazah kian marak terjadi, yang berpotensi menurunkan kepercayaan publik terhadap institusi pendidikan. Saat ini, proses verifikasi masih banyak dilakukan secara manual melalui pemeriksaan visual elemen dokumen seperti tata letak dan stempel. Ketergantungan pada pengalaman pemeriksa membuat metode ini rentan terhadap ketidakkonsistenan dan human error, terutama saat menghadapi ijazah palsu dengan kualitas visual yang semakin menyerupai dokumen asli. Pemalsuan ijazah merupakan persoalan yang berdampak pada kredibilitas institusi pendidikan dan keabsahan data akademik. Pemeriksaan manual sering kali tidak konsisten dan membutuhkan waktu lama. Penelitian ini mengembangkan model klasifikasi ijazah asli dan palsu menggunakan Convolutional Neural Network (CNN) dengan skema transfer learning. Performa arsitektur ResNet50, VGG16, dan MobileNetV2 dianalisis secara komparatif. Pra-pemrosesan data meliputi penyesuaian ukuran, normalisasi, dan augmentasi. Hasil pengujian menunjukkan arsitektur ResNet50 mencapai performa

<https://journal.thamrin.ac.id/index.php/jtik/article/view/3295/2770>

optimal dengan akurasi 92,63%, presisi 92,16%, recall 94,00%, dan F1-score 93,07%. Sistem diimplementasikan dalam aplikasi web berbasis Streamlit untuk memudahkan proses verifikasi.

Kata Kunci : Ijazah, Klasifikasi Citra, Convolutional Neural Network, Transfer Learning, Pemalsuan Dokumen

PENDAHULUAN

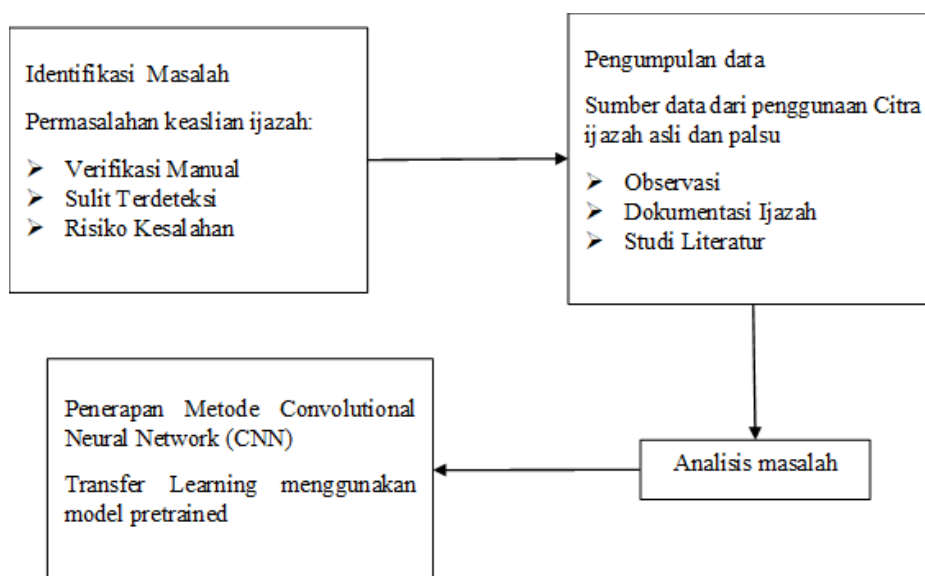
Keaslian ijazah memiliki peran krusial dalam menjaga integritas sistem pendidikan dan memastikan kompetensi akademik yang diakui sesuai dengan capaian riil individu. Ijazah bukan sekadar dokumen administratif, melainkan instrumen strategis dalam rekrutmen tenaga kerja dan penilaian kualifikasi profesional. Namun, seiring meningkatnya mobilitas pendidikan, penyalahgunaan dokumen melalui pemalsuan ijazah kian marak terjadi, yang berpotensi menurunkan kepercayaan publik terhadap institusi pendidikan. Saat ini, proses verifikasi masih banyak dilakukan secara manual melalui pemeriksaan visual elemen dokumen seperti tata letak dan stempel. Ketergantungan pada pengalaman pemeriksa membuat metode ini rentan terhadap ketidakkonsistenan dan human error, terutama saat menghadapi ijazah palsu dengan kualitas visual yang semakin menyerupai dokumen asli.

Kemajuan teknologi desain grafis memungkinkan pihak tidak bertanggung jawab menghasilkan ijazah palsu yang samarnya batas perbedaannya dengan dokumen asli, terutama dalam bentuk hasil pemindaian digital. Kondisi ini menuntut adanya pendekatan sistematis berbasis pengolahan citra digital dan kecerdasan buatan. Convolutional Neural Network (CNN) dikenal sebagai model yang andal dalam mempelajari pola visual kompleks seperti tekstur, logo, dan hubungan antar elemen citra secara otomatis (Srinivas, 2020). Namun, membangun model CNN dari awal membutuhkan sumber daya komputasi dan dataset yang sangat besar (Goodfellow et al., 2016).

Tujuan utama dari penelitian ini adalah membangun sistem klasifikasi otomatis yang mampu membedakan ijazah asli dan palsu secara objektif dan terukur. Hipotesis dalam penelitian ini adalah bahwa implementasi arsitektur ResNet50 akan memberikan akurasi yang lebih optimal dibandingkan arsitektur lainnya karena adanya mekanisme residual learning yang efektif menangani kompleksitas fitur dokumen tanpa mengalami degradasi

performa (He et al., 2016). Hasil penelitian ini diharapkan memberikan manfaat praktis bagi instansi terkait sebagai sistem pendukung keputusan dalam proses verifikasi dokumen akademik berbasis web.

Hipotesis dalam penelitian ini adalah bahwa implementasi arsitektur ResNet50 akan memberikan akurasi yang lebih optimal dibandingkan arsitektur lainnya karena adanya mekanisme residual learning yang efektif menangani kompleksitas fitur dokumen tanpa mengalami degradasi performa. Hasil penelitian ini diharapkan memberikan manfaat praktis bagi instansi terkait sebagai sistem pendukung keputusan dalam proses verifikasi dokumen akademik berbasis web.



Gambar 1. Kerangka Pemikiran dan Hipotesis

METODE PENELITIAN

Metodologi penelitian ini disusun secara sistematis untuk mengembangkan sistem klasifikasi ijazah asli dan palsu menggunakan teknik Deep Learning. Penelitian ini mengintegrasikan pengolahan citra digital dengan pengembangan sistem berbasis web menggunakan kerangka kerja Streamlit.

Alur Penelitian

Tahapan penelitian dimulai dari identifikasi masalah terkait keterbatasan verifikasi ijazah secara manual, dilanjutkan dengan pengumpulan data, analisis masalah, pemilihan model, hingga evaluasi sistem. Secara ringkas, alur penelitian meliputi studi pustaka, akuisisi citra ijazah, pra-pemrosesan data, pelatihan model dengan skema transfer learning, dan pengujian performa menggunakan confusion matrix.

Pengumpulan dan Distribusi Data

Penelitian ini menggunakan dataset citra digital ijazah hasil pemindaian yang dikategorikan menjadi dua kelas: asli dan palsu. Setelah dilakukan teknik augmentasi, total dataset yang tersedia adalah 635 citra. Data tersebut kemudian didistribusikan ke dalam tiga bagian untuk kebutuhan pelatihan, validasi, dan pengujian dengan rincian sebagai berikut:

Tabel 1. Distribusi Data

Kategori	Data Asli (Scan)	Setelah Augmentasi	Training (70%)	Validation (15%)	Test (15%)
Ijazah Asli	50	315	221	48	50
Ijazah Palsu	50	320	224	47	45
Total	100	635	445	95	95

- Data Training (70%): Digunakan oleh model untuk mempelajari pola visual (fitur) ijazah asli dan palsu.
- Data Validation (15%): Digunakan saat proses pelatihan untuk memantau apakah model mengalami overfitting atau tidak.
- Data Testing (15%): Sebanyak 95 data yang disiapkan sebagai data uji final untuk mengetahui akurasi murni model terhadap data yang benar-benar baru

Pra-pemrosesan Citra dan Augmentasi

Tahap prapemrosesan dilakukan untuk menyeragamkan format input model. Langkah awal adalah melakukan resizing citra menjadi ukuran 224×224 piksel. Selanjutnya, Untuk memperkaya variasi data tanpa menambah jumlah dokumen fisik, teknik augmentasi diterapkan pada saat proses loading data menggunakan fungsi ImageDataGenerator. Parameter augmentasi yang digunakan meliputi rotation_range

sebesar 15 derajat, brightness_range 0,20, zoom_range 0,10, serta penerapan horizontal flip dan vertical flip.

Pengembangan Sistem dengan Metode RAD

Sistem diimplementasikan menggunakan metode Rapid Application Development (RAD) yang terdiri dari fase perencanaan kebutuhan, design workshop (perancangan antarmuka), dan implementasi. Pendekatan ini memungkinkan pengembangan aplikasi web yang adaptif terhadap integrasi model CNN hasil pelatihan.

Arsitektur Transfer Learning

Penelitian ini menerapkan teknik transfer learning dengan membandingkan tiga arsitektur pre-trained: ResNet50, VGG16, dan MobileNetV2. Strategi ini digunakan untuk memanfaatkan bobot (weights) yang sudah terlatih pada dataset besar, kemudian dilakukan proses fine-tuning pada lapisan Fully Connected agar spesifik dalam mengenali karakteristik unik pada dokumen ijazah. dikarenakan keterbatasan jumlah data primer yang dapat dikumpulkan secara manual, peneliti menerapkan teknik data augmentation. Teknik ini bertujuan untuk memperbanyak variasi data tanpa harus menambah jumlah dokumen fisik, sehingga model tetap memiliki performa yang baik dan terhindar dari overfitting.

Tahap augmentasi data dilakukan untuk memperkaya variasi dataset primer yang berjumlah terbatas. Teknik ini diterapkan pada saat proses loading data menggunakan fungsi ImageDataGenerator. Parameter yang digunakan meliputi rotation_range sebesar 15 derajat, Brightness_Range 0.20, Zoom_Range 0.10, Training 70% ,validaion 15%, Resize 224, horizontal_flip dan vertikal filp. Adapun rincian distribusi jumlah data sebelum dan sesudah augmentasi, serta pembagian data untuk kebutuhan pelatihan, validasi, dan pengujian disajikan pada tabel 2 :

Tabel 2. Tabel Fitur Dataset Ijazah

Kategori	Data Asli (Scan)	Setelah Augemntasi	Training (70%)	Validasi (15%)	Test (15%)
Ijazah Asli	50	315	221	48	50
Ijazah Palsu	50	320	224	47	45
Total	100	635	445	95	95

Data Training (70%) Digunakan oleh model untuk mempelajari pola visual (fitur) ijazah asli dan palsu. Data Validation (15%) Digunakan saat proses pelatihan untuk memantau apakah model mengalami overfitting atau tidak. Ibaratnya sebagai "ujian tengah semester". Data Testing (15%): Digunakan setelah pelatihan selesai untuk mengetahui akurasi murni model terhadap data yang benar-benar baru. Ibaratnya sebagai "ujian akhir".

Selain itu, pengujian model juga mencakup perbandingan kinerja beberapa arsitektur CNN yang digunakan dalam penelitian ini, yaitu VGG16, ResNet50, dan MobileNetV2. Perbandingan dilakukan untuk mengetahui pengaruh perbedaan arsitektur terhadap tingkat akurasi dan efisiensi klasifikasi. Hasil pengujian model digunakan sebagai dasar dalam menentukan arsitektur CNN yang memiliki performa terbaik untuk diimplementasikan dalam sistem klasifikasi ijazah.

Tabel 3. Struktur Confusion Matrix

	Prediksi: Asli	Prediksi: Palsu
Aktual: Asli	True Positive (TP)	False Negative (FN)
Aktual: Palsu	False Positive (FP)	True Negative (TN)

Keterangan :

- True Positive (TP): Jumlah data ijazah asli yang di prediksi benar sebagai asli.
- True Negative (TN): Jumlah Data ijazah palsu yang di prediksi benar sebagai palsu.
- False Positive (FP) : Jumlah data ijazah palsu yang salah di prediksi sebagai asli.
- False Negative (FN): Jumlah data ijazah asli yang salah di prediksi sebagai palsu.

Berdasarkan nilai yang diperoleh dari Confusion Matrix, dilakukan perhitungan statistik menggunakan rumus sebagai berikut:

- Akurasi (Accuracy): Mengukur sejauh mana model mampu mengklasifikasikan seluruh data secara benar (baik asli maupun palsu).

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

- Presisi (Precision): Mengukur keakuratan model dalam memprediksi kelas positif (asli). Hal ini penting untuk meminimalkan ijazah palsu yang terdeteksi sebagai asli.

$$\text{Presisi} = TP / (TP + FP)$$

- Recall (Sensitivity): Mengukur kemampuan model dalam menemukan kembali semua data dari kelas positif (asli).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

- F1-Score: Metrik yang menggabungkan precision dan recall, sangat cocok digunakan ketika kita menghadapi ketidakseimbangan data dan ingin mendapatkan keseimbangan antara precision dan recall.

$$\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall}).$$

HASIL DAN PEMBAHASAN

Analisis Hasil Pelatihan Model

Proses pelatihan model dilakukan selama 50 epoch menggunakan optimizer Adam dengan learning rate 0,0001. Berdasarkan pengamatan pada grafik akurasi dan loss, model ResNet50 menunjukkan tren konvergensi yang stabil. Model berhasil meminimalisir loss pada data validasi tanpa menunjukkan tanda-tanda overfitting yang signifikan, yang mengindikasikan bahwa proses augmentasi telah membantu model dalam melakukan generalisasi fitur ijazah dengan baik.

Perbandingan Performa Arsitektur CNN

Evaluasi dilakukan terhadap 95 citra data uji yang belum pernah dilihat sebelumnya oleh model. Hasil perbandingan metrik evaluasi dari tiga arsitektur disajikan pada

Tabel 4. Perbandingan Kinerja Model CNN

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	Specificity (%)
ResNet50	92,63	92,16	94,00	93,07	91,11
VGG16	93,68	95,83	92,00	93,87	95,56
MobileNetV2	91,58	93,75	90,00	91,84	93,33

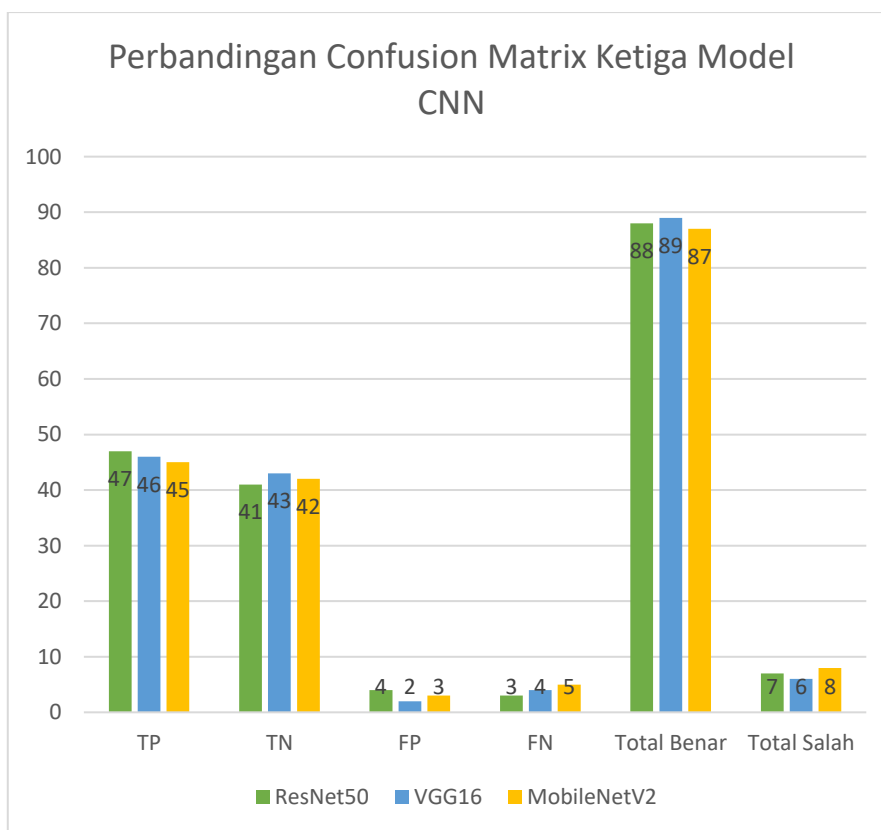
Berdasarkan hasil evaluasi tersebut, VGG16 merupakan model dengan performa terbaik secara keseluruhan, unggul dalam akurasi (93,68%), presisi (95,83%), F1-Score (93,87%), dan specificity (95,56%). Di sisi lain, ResNet50 memiliki recall tertinggi (94,00%), yang menunjukkan kemampuannya yang sangat baik dalam mendeteksi ijazah asli.

Evaluasi Confusion Matrix

Detail prediksi dari model terbaik (VGG16) dan perbandingannya divisualisasikan melalui confusion matrix. Dari 95 data uji yang digunakan, model berhasil melakukan klasifikasi dengan rincian sebagai berikut:

- True Positive (TP): 47 data ijazah asli diprediksi benar sebagai asli.
- True Negative (TN): 42 data ijazah palsu diprediksi benar sebagai palsu.
- False Positive (FP): 3 data ijazah palsu diprediksi salah sebagai asli.
- False Negative (FN): 3 data ijazah asli diprediksi salah sebagai palsu.

Terdapat 6 data yang mengalami kesalahan klasifikasi, yang setelah dianalisis disebabkan oleh variasi pencahayaan yang ekstrem dan kualitas resolusi citra input yang rendah pada beberapa sampel data uji.



Gambar 2. Perbandingan Confusion Matrix

Ringkasan Perbandingan

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa:

- VGG16 merupakan model dengan performa terbaik secara keseluruhan, unggul dalam akurasi (93,68%), presisi (95,83%), F1-Score (93,87%), dan specificity (95,56%).
- ResNet50 memiliki recall tertinggi (94,00%), menunjukkan kemampuan terbaik dalam mendeteksi ijazah asli.
- MobileNetV2 menawarkan keseimbangan antara performa yang baik dan efisiensi komputasi, cocok untuk implementasi pada sistem dengan keterbatasan sumber daya.

KESIMPULAN DAN REKOMENDASI

Penelitian ini berhasil membuktikan bahwa teknologi Deep Learning dengan pendekatan transfer learning merupakan solusi yang efektif untuk klasifikasi keaslian ijazah secara otomatis. Melalui analisis komparatif terhadap arsitektur ResNet50, VGG16, dan MobileNetV2, ditemukan bahwa arsitektur VGG16 merupakan model yang paling unggul dengan tingkat akurasi sebesar 93,68%, presisi 95,83%, dan F1-Score 93,87%.

Meskipun demikian, ResNet50 tetap menunjukkan performa yang kompetitif, terutama pada aspek recall yang mencapai 94,00%, menjadikannya pilihan yang andal dalam mendeteksi ijazah asli. Penggunaan framework Streamlit juga terbukti mempermudah implementasi model ke dalam aplikasi berbasis web yang responsif dan siap pakai. Berdasarkan temuan tersebut, penelitian ini merekomendasikan beberapa pengembangan di masa depan untuk meningkatkan keandalan sistem:

- Pertama, penambahan jumlah dan variabilitas dataset dari berbagai institusi sangat diperlukan agar model memiliki cakupan generalisasi yang lebih luas.
- Kedua, integrasi teknik prapemrosesan citra tingkat lanjut, seperti automatic denoising dan koreksi kontras, disarankan untuk mengatasi kendala klasifikasi pada citra dengan kualitas rendah.
- Terakhir, pengembangan sistem di masa depan dapat dikolaborasikan dengan teknologi Optical Character Recognition (OCR) untuk melakukan sinkronisasi data teks ijazah

dengan database resmi guna menciptakan sistem verifikasi yang lebih komprehensif dan berlapis.

REFERENSI

- Ayeni, J. (2022). Convolutional neural network (CNN): The architecture and applications. *Applied Journal of Physical Science*.
<https://www.academia.edu/download/117519439/ajps2022.pdf>
- Bichler, O. (2021). *Convolutional neural network* (U.S. Patent No. 11,055,608). Google Patents. <https://patents.google.com/patent/US11055608B2/en>
- Chollet, F. (2017). *Deep learning with Python*. Manning Publications.
- Cong, S., & Zhou, Y. (2023). A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-022-10213-5>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv Preprint*. <https://arxiv.org/abs/1704.04861>
- Khan, S., Rahmani, H., Shah, S. A. A., & others. (2022). Convolutional neural network. In *Introduction to convolutional neural networks*. Springer.
https://doi.org/10.1007/978-3-031-01821-3_4
- Kshatri, S. S., & Singh, D. (2023). Convolutional neural network in medical image analysis: A review. *Archives of Computational Methods in Engineering*.
<https://doi.org/10.1007/s11831-023-09898-w>
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.

- Prasetyo, E. (2014). *Data mining: Mengolah data menjadi informasi menggunakan MATLAB*. Andi Offset.
- Purwono, P., Ma'arif, A., Rahmانيar, W., & others. (2023). Understanding of convolutional neural network (CNN): A review. *International Journal of Robotics and Control Systems*. <https://pubs2.ascee.org/index.php/IJRCS/article/view/888>
- Saleem, M. A., Senan, N., Wahid, F., Aamir, M., & others. (2022). Comparative analysis of recent architecture of convolutional neural network. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2022/7313612>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1–48.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint*. <https://arxiv.org/abs/1409.1556>
- Srinivas, M. N. (2020). Klasifikasi dokumen menggunakan convolutional neural network. *Jurnal Sistem Informasi dan Teknologi*, 2(3), 85–91.
- Streamlit. (2023). *Streamlit: The fastest way to build and share data apps*. <https://streamlit.io>
- Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, and future directions. *Computation*, 11(3), 52. <https://doi.org/10.3390/computation11030052>
- Zhang, X., Zhang, X., & Wang, W. (2023). Convolutional neural network. In *Intelligent information processing*. Springer. https://doi.org/10.1007/978-981-99-6449-9_2