

PENGENALAN AKTIVITAS MANUSIA DENGAN METODE RESNET50 DAN VGG16

Alvian Wijaya

Teknologi Informasi, Fakultas Teknik Informatika, Universitas Tarumanagara

Jl. Letnan Jendral. S. Parman No. 1, Jakarta 11440

*e-mail: alvian.535200026@stu.untar.ac.id

ABSTRAK

Perkembangan dalam dunia AI atau kecerdasan buatan berkembang pesat di zaman modern sekarang ini. Mulai dari pengenalan aktivitas manusia, pengenalan produk melalui citra, pendeteksian suara, dan lainnya. Pengenalan aktivitas manusia juga merupakan sebuah tren yang sedang banyak mendapatkan perhatian terutama dalam bidang *Intelligence System*. Pengenalan aktivitas manusia memiliki tujuan yang bisa diimplementasikan kedalam berbagai macam tujuan mulai dari pengawasan anak, perhitungan kepadatan manusia dalam suatu ruangan, dan sebagai kamera pengawas. Dalam penelitian ini pengenalan aktivitas manusia akan dideteksi dengan *Convolutional Neural Network* (CNN) dengan algoritma ResNet50 dan VGG16 sebagai perbandingan. Aktivitas yang diteliti terbagi dalam 15 kelas seperti: duduk, menggunakan laptop, tidur, menari, berlari, makan, dan lain-lain. Data dibagi menjadi dua yaitu data latih dan data validasi lalu kemudian dengan algoritma ResNet50 dan VGG16 akan didapatkan akurasi dari model yang dilatih. Pada hasil penelitian didapatkan hasil bahwa model yang dilatih dengan algoritma ResNet50 memiliki hasil yang sedikit lebih baik dibandingkan VGG16 dengan akurasi ResNet50 95% dan VGG16 sebesar 94%.

Kata kunci: Pengenalan Aktivitas Manusia, CNN, ResNet50, VGG16.

ABSTRACT

The field of Artificial Intelligence (AI) is rapidly advancing in the modern era. Various applications such as human activity recognition, product recognition through images, voice detection, and more have gained significant attention. Human activity recognition, in particular, is a trending area of focus, especially in the domain of Intelligence Systems. The goal of human activity recognition can be implemented for diverse purposes, including child monitoring, crowd density estimation in a room, and surveillance cameras. This research focuses on human activity recognition using Convolutional Neural Networks (CNN), employing the ResNet50 algorithm and comparing it with VGG16. The studied activities are categorized into 15 classes, such as sitting, using a laptop, sleeping, dancing, running, eating, and others. The dataset is divided into training and validation data. Subsequently, the ResNet50 and VGG16 algorithms are employed to train models, and the accuracy of the trained models is evaluated. The research results indicate that the model trained with the ResNet50 algorithm performs slightly better than VGG16, achieving an accuracy of 95% compared to VGG16's 94%.

Keywords: Human Activity Recognition, CNN, ResNet50, VGG16.

1 PENDAHULUAN

Dalam kesehariannya aktivitas manusia merupakan hal yang penting dalam interaksi dan komunikasi antar manusia [1]. Ditambah dengan kemajuan teknologi dalam bidang ilmu kecerdasan juga membuat penelitian dalam bidang interaksi antara manusia dengan komputer atau robot dengan interaksi manusia dalam mengenali aktivitas yang dilakukan [2]. Teknologi kecerdasan buatan yang dikembangkan memiliki kecerdasan layaknya seorang manusia dalam memahami perintah, mengenali citra, hingga mengawasi sebuah aktivitas [3]. Salah satu pengimplementasian kecerdasan buatan adalah dengan pengenalan aktivitas manusia. Pengenalan aktivitas manusia merupakan proses dalam mengidentifikasi bentuk tubuh dari objek manusia [4]. Dalam melakukan pengenalan aktivitas manusia ada beberapa cara untuk melakukan tahapan sensornya seperti mengambil *key-point*, *heatmap*, sensor berdasarkan lingkungan, dan lainnya [5] [6]. Penelitian ini mengklasifikasikan total

sebanyak 15 kelas aktivitas manusia yang dilakukan seperti berdiri, bermain laptop, minum, menari, dan lainnya. Dalam proses perancangannya pengenalan aktivitas manusia memiliki dua basis yaitu visual dan sensor [7].

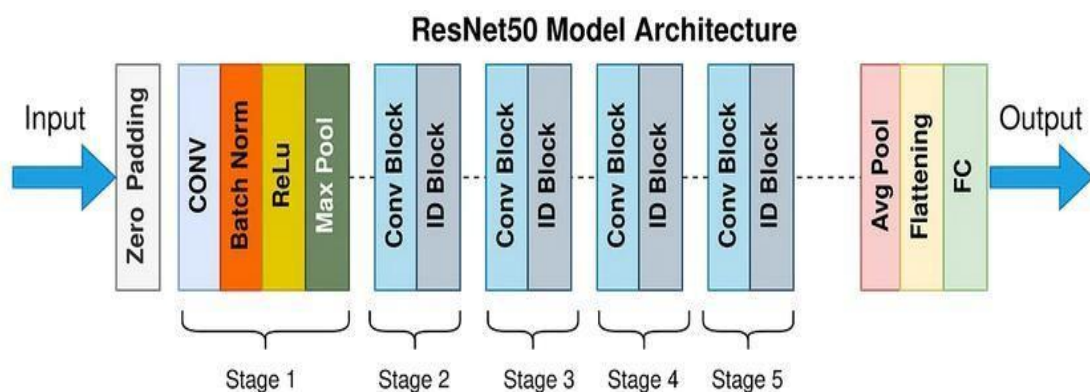
Pengenalan aktivitas manusia dapat bermanfaat atau berguna dalam era yang serba digital sekarang ini. Misalnya untuk mengenali aktivitas manusia di supermarket yang bisa digunakan sebagai kamera pengawas atau diimplementasikan seperti *amazon go* [8]. Pengenalan aktivitas di daerah pertambakan udang sebagai CCTV pengawas [9]. Pengenalan aktivitas olahraga kebugaran [10]. Pengenalan aktivitas manusia khususnya anak-anak sebagai bentuk kamera pengawasan aktivitas anak-anak [11].

Secara garis besar dalam proses pengenalan aktivitas manusia terbagi menjadi empat tahapan yaitu menangkap aktivitas dari sumber, *pre-processing*, model training, dan label aktivitas yang sudah di training [12]. Model Pelatihan training yang dikenal secara luas umumnya seperti ResNet, VGG, *Inception*, *EfficientNet*, dan *MobileNet* [13]. Penelitian ini menggunakan dua model pelatihan *supervised learning* yaitu VGG16 dan ResNet50. Perbedaan antara *supervised learning* dan *unsupervised learning* terdapat pada tujuan untuk melakukan pemisahan dan mengelompokan data yang akan di prediksi sesuai dengan penelitian [14]. Dalam pembelajaran *unsupervised learning* data yang dimiliki hanya berupa input namun hasil keluarannya tidak diidentifikasi dengan contoh kelas [15]. Pada penulisan paper ini jumlah kelas berjumlah total 15 kelas dengan masing-masing data sebanyak 840 foto yang terdiri dari: duduk, menggunakan laptop, berpelukan, tidur, minum, tepuk-tangan, menari, bersepeda, menelepon, ketawa, makan, bertengkar, mendengarkan musik, berlari, dan mengetik.

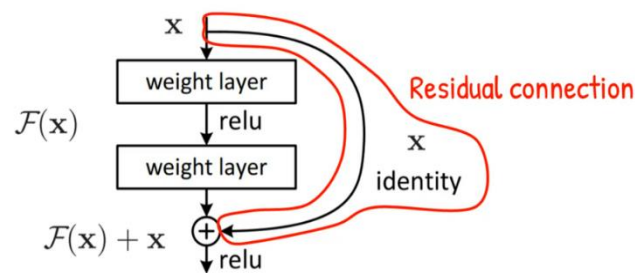
2 TINJAUAN LITERATUR

Pengenalan aktivitas olahraga dengan LSTM dan mediapipe oleh Tanugraha dkk [16] yang menghasilkan akurasi sebesar 91% dengan loss sebesar 29% untuk melakukan pengenalan Gerakan *T-Pose*, *Warrior II Pose*, dan *Tree Pose* dalam yoga. Selanjutnya adalah analisis performa algoritma untuk mengenali kegiatan manusia oleh Azhar dkk [17] dengan membandingkan algoritma SVM, CNN, dan LSTM dengan hasil algoritma CNN mencapai hasil tertinggi dengan akurasi 97.58%. Pengenalan aktivitas manusia pada tambak udang dengan CNN oleh Arhan dkk [18] dengan memanfaatkan algoritma SQRT untuk melacak dan mendeteksi manusia untuk mencegah kasus pencurian yang ada dalam Kawasan tambak mencapai akurasi pengujian sebesar 81,03%.

ResNet50 merupakan satu jenis arsitektur *Convolutional Neural Network* (CNN) yang memiliki keunggulan dalam mengoptimalkan *neural network* dan dapat memperoleh akurasi yang lebih tinggi dibandingkan arsitektur lainnya dengan memiliki 50 lapisan dalam jaringannya [19].



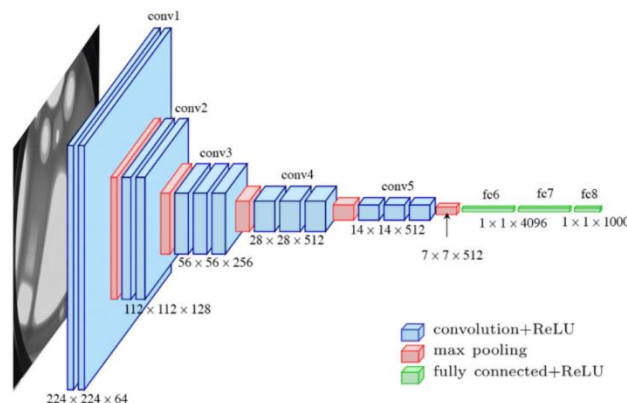
Gambar.1 Arsitektur ResNet50



Gambar.2 Residul Connection dalam ResNet-50

ResNet memiliki keunggulan dalam mengoptimalkan *neural network* dikarenakan adanya permasalahan gradien yang hilang saat melatih *neural network* yang sangat mendalam sehingga penerapan ResNet ini menjadi solusi dalam melatih jaringan yang dalam agar *neuron* dapat mengingat dan menyimpan nilai dengan optimal. Agar menjadi lebih optimal, dalam ResNet terdapat fungsi yang dinamakan sebagai *skip network*. Dalam ResNet juga terdapat yang Namanya Residual connection yang berguna dalam mengurangi masalah *vanishing gradient* dikarenakan *gradient* dapat mengalir dari dua cabang [19].

VGG16 adalah versi dari jaringan saraf konvolusional populer yang disebut VGGNet. VGG16 terdiri dari beberapa lapisan, termasuk 13 lapisan konvolusional dan 3 lapisan terhubung penuh. Oleh karena itu, ia harus mempelajari bobot dari 16 lapisan. Ini menerima gambar berwarna dengan ukuran 224x224 piksel sebagai input dan mengklasifikasikannya ke salah satu dari 1000 kelas. Oleh karena itu, hasilnya adalah vektor dengan ukuran 1000, yang berisi probabilitas untuk setiap kelas [20].



Gambar 3 Arsitektur VGG16

Convolutional Layer: Merupakan lapisan pertama dalam CNN dalam melakukan ekstraksi fitur. Gambar input berupa matrik akan dihitung dengan persamaan seperti berikut:

$$x(i, j) = \sum_m \sum_n w_{m,n}^l * o_{i+m, j+n}^{l-1} + b \quad (1)$$

Keterangan:

$x(i, j)$ = hasil perhitungan konvolusi pada titik (x, y)

l = layer

$o_{i,j}$ = input citra

$w_{m,n}$ = filter yang dipakai

b = bias

i = baris *pixel* citra

j = kolom *pixel* citra

Selanjutnya *feature map* untuk menghitung matrik hasil *convolution* dengan rumus:

$$Output = \frac{w-n}{s} + 1 \quad (2)$$

Keterangan:

w = Panjang atau tinggi *input*

n = Panjang atau tinggi *filter*

p = *Padding*

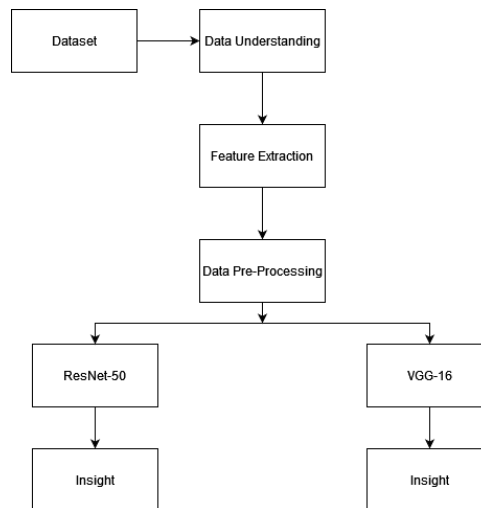
s = *Stride*

Pooling Layer digunakan untuk mengurangi dimensi gambar dengan memilih nilai yang penting dari satu wilayah. Teknik-teknik umum yang terdapat dalam *Pooling Layer* antara lain: *Max Pooling*, *Average Pooling*, *Min Pooling*, dan *Global Average Pooling*.

Flattening adalah tindakan mengubah matrik dua dimensi menjadi sebuah vector.

3 METODE PENELITIAN

Gambar untuk melakukan proses training dapat diakses secara online pada Kaggle dataset. Kemudian gambar akan dilakukan tahapan *Data Understanding*, *Feature Extraction*, *Data Pre-processing*, *Modelling* dengan 2 algoritma, dan *Insight*.



Gambar 4 Diagram Alur Penelitian

3.1 Dataset Penelitian

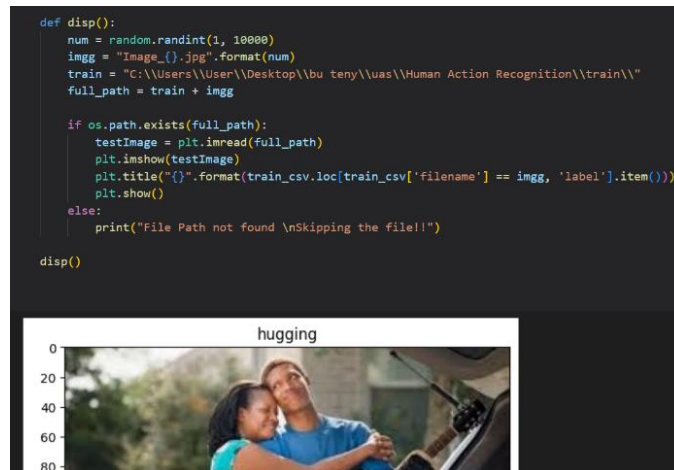
Dalam tahap ini akan dilakukan pengumpulan data yang akan digunakan sebagai data *training* dan data validasi. Dalam penelitian ini jumlah data *training*nya ada 12.600 data dan 5.400 data validasi yang terbagi dalam 15 kategori.

3.2 Feature Extraction

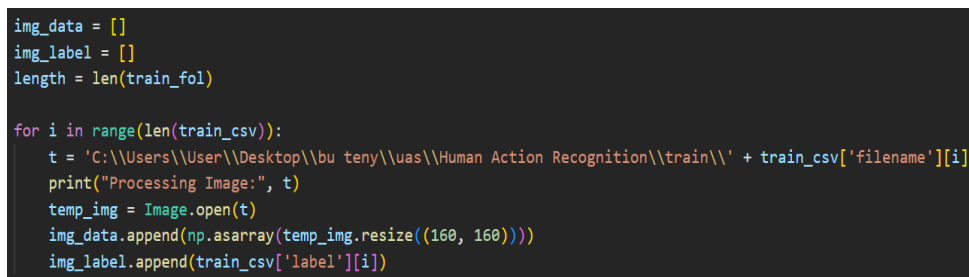
Dalam *feature extraction* akan digunakan total 15 kategori dataset yang dipakai dengan tujuan mengenali aktivitas. Terdapat tiga proses dalam *feature extraction* yakni *selection*, *extraction/engineering*, dan *transformation* [21]. *Selection* bertujuan untuk membuang fitur yang tidak relevan, *extraction/engineering* membuat fitur baru dari fitur yang ada, dan terakhir adalah *transformation* untuk mengubah fitur ke bentuk yang lebih muda dipakai model [22].

3.3 Pre-Processing

Tahapan memahami data yang akan dilatih selanjutnya adalah *pre-processing* yang merupakan tahapan untuk transformasi data agar dapat dijalankan kedalam algoritma yang dipakai nantinya.



Gambar 5. Pre-Processing



Gambar 6. Gambaran Pre-Processing data

3.4 Modelling

Model: "sequential_17"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 512)	14714688
flatten_17 (Flatten)	(None, 512)	0
dense_34 (Dense)	(None, 512)	262656
dense_35 (Dense)	(None, 15)	7695

=====
Total params: 14985039 (57.16 MB)
Trainable params: 270351 (1.03 MB)
Non-trainable params: 14714688 (56.13 MB)

Gambar 7. Model VGG16

Model: "sequential_18"

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2048)	23587712
flatten_18 (Flatten)	(None, 2048)	0
dense_36 (Dense)	(None, 512)	1049088
dense_37 (Dense)	(None, 15)	7695

=====
Total params: 24644495 (94.01 MB)
Trainable params: 1056783 (4.03 MB)
Non-trainable params: 23587712 (89.98 MB)

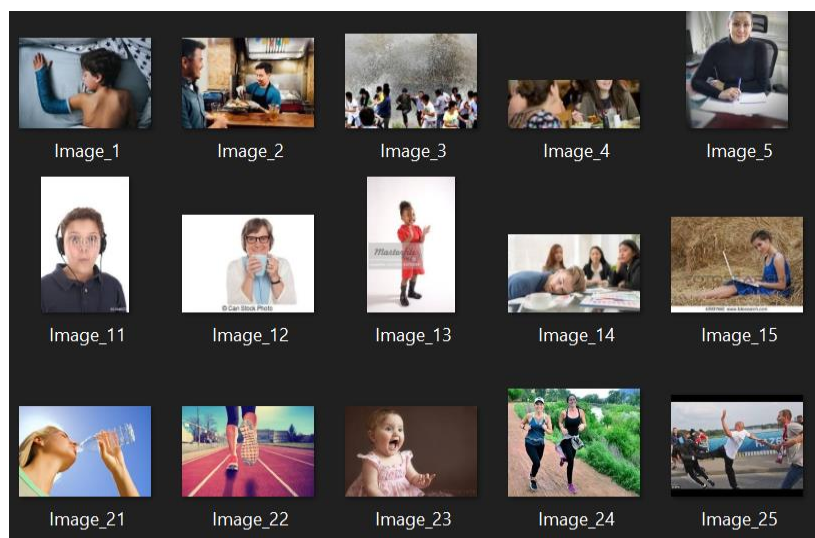
Gambar 8. Model ResNet50

4 HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan hasil dari model VGG16 dan ResNet-50 yang dipakai dalam penelitian yang dirancang. Dalam penelitian ini jumlah total citra yang dipakai sebanyak 18.000 citra yang terdiri dari 12.600 citra latih dan 5.400 citra validasi yang dibagi menjadi 15 kelas.

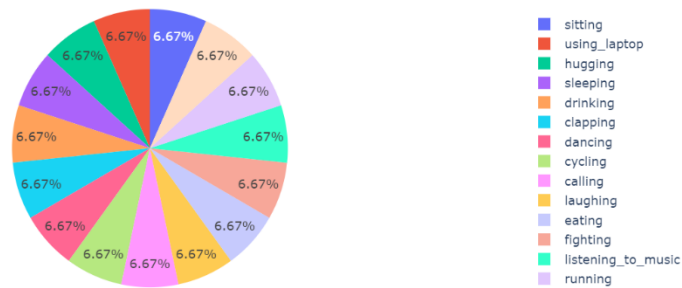


Gambar 9. Citra Training



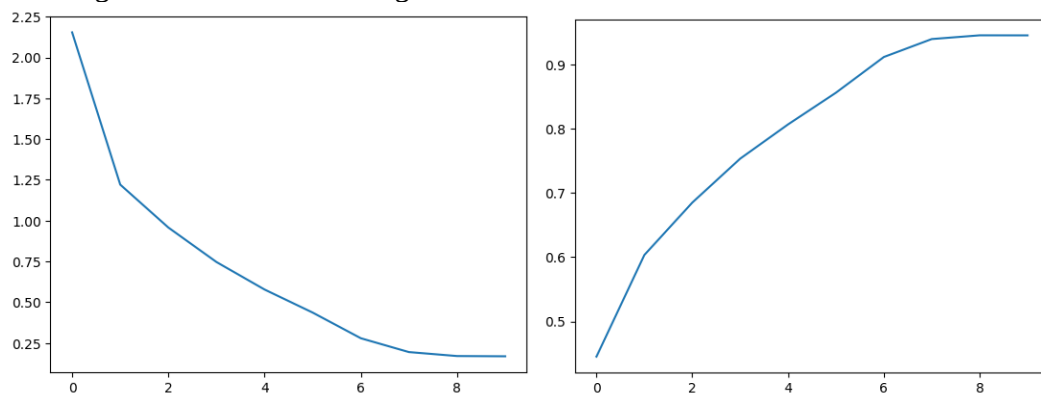
Gambar 10. Citra Validasi

Distribution of Human Activity



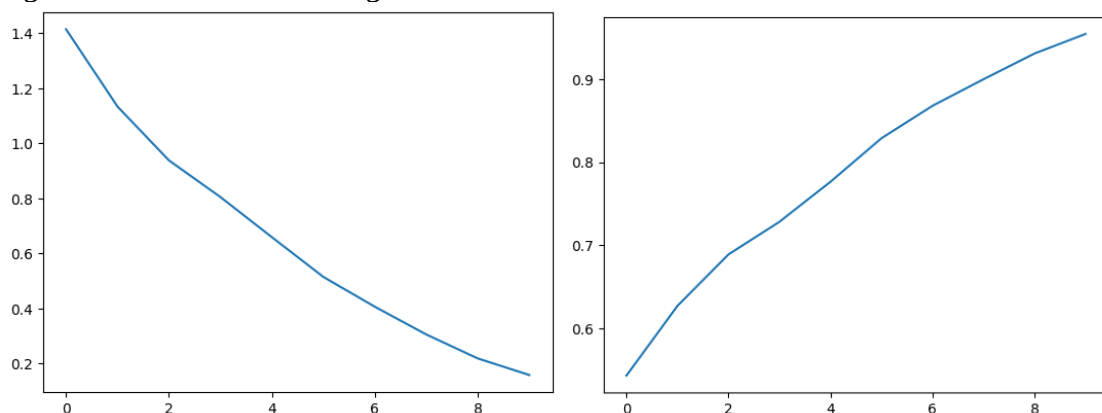
Gambar 11. Chart Pie Kelas aktivitas

Setelah dilakukan tahapan training selanjutnya didapatkan hasil dari tiap model. Model VGG16 menghasilkan akurasi 0.9451 dengan loss 0.1690.



Gambar 12. Loss dari VGG16 dan Akurasi dari VGG16

Selanjutnya setelah model pertama selesai dilanjutkan untuk Model ResNet50 yang menghasilkan akurasi 0.9541 dengan loss 0.1584.



Gambar 13. Loss dari ResNet50 dan Akurasi dari ResNet50

Kemudian hasil dari model yang sudah dilatih kemudian akan dimasukkan kedalam data validasi untuk diprediksi hasilnya.

	image_name	prediction
0	Image_1.jpg	texting
1	Image_10.jpg	drinking
2	Image_100.jpg	running
3	Image_1000.jpg	eating
4	Image_1001.jpg	drinking
...	...	...
5405	Image_995.jpg	sitting
5406	Image_996.jpg	using laptop
5407	Image_997.jpg	clapping
5408	Image_998.jpg	dancing
5409	Image_999.jpg	running

5410 rows × 2 columns

Gambar 14. VGG16

	image_name	prediction
0	Image_1.jpg	sleeping
1	Image_10.jpg	drinking
2	Image_100.jpg	running
3	Image_1000.jpg	eating
4	Image_1001.jpg	listening_to_music
...	...	...
5405	Image_995.jpg	dancing
5406	Image_996.jpg	sitting
5407	Image_997.jpg	clapping
5408	Image_998.jpg	dancing
5409	Image_999.jpg	texting

Gambar 15. ResNet-50

5 KESIMPULAN

Dari penelitian ini, pengenalan aktivitas manusia sebanyak 15 kelas

DAFTAR PUSTAKA

- [1] R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 302-311, 2020.
- [2] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, dan H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.
- [3] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, dan H. Ghayvat, "CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021.
- [4] J. K. Deepak, A. Mahanti, P. Shamsolmoali, dan R. Manikandan, "Deep neural learning techniques with long short-term memory for gesture recognition," *Neural Computing and Applications*, vol. 32, pp. 16073-16089, 2020.

- [5] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, dan Y. Liu, "Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities," *ACM Computing Surveys*, vol. 54, no. 4, p. 77, 2021.
- [6] C. Jobanputra, J. Bavishi, dan N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, p. 698–703, 2019.
- [7] M. H. Arshad, M. Bilal, dan A. Gani, "Human activity recognition: Review, taxonomy and open challenges," *Sensors*, vol. 22, no. 17, p. 6463, 2022.
- [8] K. D. Runtu and Lina, "Pengenalan Aktivitas Manusia di Supermarket dengan Metode Long Short Term Memory," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 2, 2022.
- [9] M. A. Zulfikar, M. Somantri, dan Sudjadi, "Penerapan Metode Convolutional Neural Network (CNN) dan Long Short Term Memory (LSTM) Untuk Pengenalan Aktivitas Manusia pada CCTV di Area Tambak Udang," *Transient: Jurnal Ilmiah Teknik Elektro*, vol. 10, no. 1, pp. 98-105, 2021.
- [10] Irfan, M. A. Muthalib, Kartika, and S. Meliala, "Pengiraan Pose Model Manusia Pada Repetisi Kebugaran Ai Pemograman Python Berbasis Komputerisa," *Infotech*, vol. 9, no. 1, pp. 11-19, 2023.
- [11] A. Lay and L. Lina, "Pendeteksian Aktivitas Manusia dengan Human Pose Estimation dan Convolutional Neural Network," *Computatio*, vol. 6, no. 1, pp. 51-60, 2021.
- [12] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, "Human activity recognition in artificial intelligence framework: A narrative review," *Artificial intelligence review*, vol. 55, no. 6, pp. 4755-4808, 2022.
- [13] A. Krizhevsky, I. Sutskever, dan G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [14] J. Sanjaya, E. Renata, V. E. Budiman, F. Andreson, dan M. Ayub, "Prediksi Kelalaian Pinjaman Bank Menggunakan Random Forest dan Adaptive Boosting," *JuTISI (Jurnal Teknik Informatika dan Sistem Informasi)*, pp. 50-60, 2020.
- [15] Z. Lin, D. Qiu, D. Ergu, Y. Cai, dan K. Liu, "A Study on Predicting Loan Default Based on the Random Forest Algorithm," *Procedia Computer Science*, pp. 503-513, 2019.
- [16] F. D. Tanugraha, H. Pratikno, Musyannah, dan W. I. Kusumawati, "Pengenalan Gerakan Olahraga Berbasis (Long Short-Term Memory) Menggunakan Mediapipe," *Journal of Advances in Information and Industrial Technology*, vol. 4, no. 1, pp. 37-46, 2022.
- [17] A. Y. Ramadhan, F. A. D. Fitrah, F. Y. Suratman, dan Istiqomah, "Analisis Performansi Algoritma Svm, Cnn, dan Lstm Untuk Pengenalan Kegiatan Manusia Dengan Urad Fmcw Radar," *Teknika - Jurnal Penelitian Dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, Dan Elektronika*, Vol. 8, no. 1, pp. 27-34, 2023.
- [18] M. Arfan, A. Nurjalal, M. Somantri, dan Sudjadi, "Pengenalan Aktivitas Manusia pada Area Tambak Udang dengan Convolutional Neural Network," *Resti Rekayasa Sistem dan Teknologi Informasi*, vol. 5, no. 1, pp. 174-179, 2021.
- [19] S. Mukherjee, "The Annotated ResNet-50," *Toward Data Science*, 18 August 2022. [Online]. Available: <https://towardsdatascience.com/the-annotated-ResNet-50-a6c536034758>. [Diakses 11 September 2023].
- [20] M. Aatila, M. Lachgar, H. Himech, dan A. Kartit, "Diabetic Retinopathy Classification Using ResNet50 and VGG-16 Pretrained Networks," *International Journal of Computer Engineering and Data Science (IJCEDS)*, vol. 1, no. 1, pp. 1-7, 2021.
- [21] K. B. Simarmata dan K. D. Hartomo, "Analisa Rekomendasi Fitur Persetujuan Pinjaman Perusahaan Financial Technology Menggunakan Metode Random Forest," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, pp. 2055-2070, 2022.
- [22] Y. Religa, A. Rusdi, I. Romli, dan A. Mazid, "Religia, Y., Rusdi, A., Feature Extraction Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine Dan K-Nearest Neighbor," *Pelita Teknologi*, pp. 85-92, 2019.