Scientific Journal of Informatics
Vol. 13, No. 1, February 2026
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Robust Human Gait Recognition with Convolutional Neural Network based on Gait Energy Image

Fandy Indra Pratama[1*], Akhmad Pandhu Wijaya[2], Gilar Pandu Annanto[3], Avira Budianita[4], Hairudin Farid Sunanda[5]
[1, 2, 5]Department of Informatics Engineering, Universitas Wahid Hasyim, Indonesia
[3] Department of Mechanical Engineering, Universitas Wahid Hasyim, Indonesia
[4] Department of Information System, Universitas Muhammadiyah Kudus, Indonesia

Abstract.
Purpose: Human gait recognition is one of the developments in artificial intelligence technology. Gait recognition is a biometric recognition technique that uses no direct interaction with an object, allowing for identification of individuals based on their gait. However, this recognition faces challenges, including varying camera angles (0^0 - 180^0), so this requires a more in-depth introduction.
Methods: Therefore, based on the references, this study proposes using the Gait Energy Image (GEI) and Convolutional Neural Network (CNN) features for in-depth extraction and recognition of each image in the Casia B Dataset, which is then compared with the results of previous studies.
Result: The results of this study, with the division of the Casia B Dataset 80% as training data and 20% as testing data and 11 camera angles between 0^0 - 180^0 produced an accuracy rate of 99.48%.
Novelty: So the accuracy achieved with this deep learning technique exceeds that of previous research using conventional methods and this gait pattern recognition technique can be used to be implemented in a biometric recognition system based on human gait patterns.

Keywords: Authentication, Biometrics, CNN, GEI, Gait recognition
Received November 2025 / **Revised** December 2025 / **Accepted** January 2026

INTRODUCTION

Implementations of artificial intelligence technology include systems that can recognize a person's biometrics based on their gait, fingerprint, and iris patterns. Fingerprint and iris recognition require direct interaction with a sensor or system, while human gait recognition is a biometric recognition process that requires no direct interaction [1], [2]. Therefore, human gait recognition can be used for authentication or identification of individuals on criminal wanted lists. This recognition remains a significant challenge due to the impact of viewing angle on the recognition rate. Related research addresses this challenge using a model-free approach, as this approach captures the characteristics or features of human movement without modelling individual parts. [3]–[6].

In the model-free approach, the preprocessing stage of the study uses data from real images of people walking, converted into Gait Energy Image (GEI) feature, which is then used as input for each person [7]. The results from the GEI feature are then processed using several methods, including the PCA (Principal Component Analysis) algorithm, which can extract features and reduce them to low-dimensional images. [6], [8], [9]. At the introduction stage, research [3], [10]–[12] using the K-NN method because its simple calculations are capable of performing optimal recognition.

Apart from using this method in technological developments, there are deep learning techniques such as the use of Convolutional Neural Network (CNN) models which are often used for object recognition [13]–[16]. In the research [17] analyzed and compared the results of the accuracy of human gait recognition using four deep learning models CNN, MLP, SOM, and EfficientNet, from this research the CNN model produced the highest accuracy level of 97.12%. In using the CNN model, epoch is a major consideration because it affects the level of object recognition accuracy [18], [19]

*Corresponding author.
Email addresses: fandy@unwahas.ac.id (Pratama)
DOI: 10.15294/sji.v13i1.37383

Therefore, research references on human gait recognition have gaps in dataset selection and implementation techniques. In this study, biometric detection of human gait patterns is carried out based on GEI feature, which is then recognized using a CNN model. This research aims to improve the level of accuracy in gait recognition and the results will be compared with previous research models that used datasets from CASIA-B (viewpoints vary). Therefore, the results of the research on the development of human gait recognition detection are better recommended for long-range biometric recognition techniques.

METHODS

The method used in solving this research is shown in Figure 1 below. The public dataset used is the Casia-B dataset consisting of 0^0 - 180^0 and each dataset has 6 gait models as an example of the dataset is shown in Figure 2. Prior to training with CNN, the dataset was segmented using the GEI algorithm, which extracts human gait models using averaging techniques. This allowed for the evaluation of gait recognition accuracy, with 80% testing data and 20% training data.

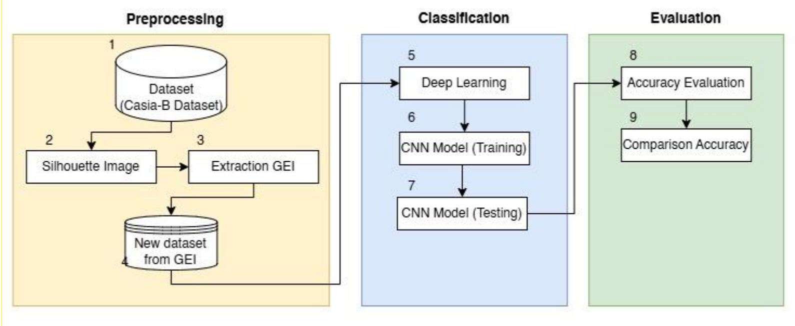

Figure 1. Proposed method flow

Gait energy image (GEI)

GEI is a widely used algorithm to obtain human walking frequencies based on sequences of images of humans walking. The following is the formula for obtaining GEI.

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \tag{1}$$

Where N is the total number of frames in the gait cycle image. T is the frame number in the gait cycle image. (x,y) are the coordinate points in the 2D image. The following is an example of 2 sequence images of a person walking with sequence number 2 from the right to the leftmost image being a frame in the sequence image, and the leftmost image is the image resulting from the GEI algorithm.

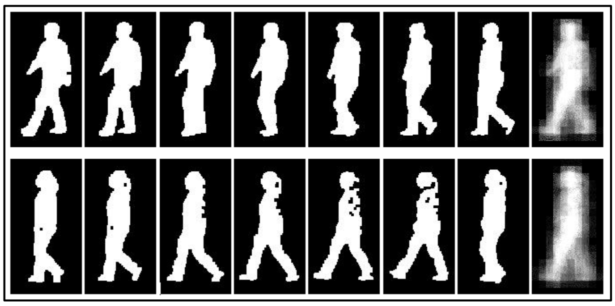

Figure 2. Gait cycle and GEI feature results (right)

GEI reflects the character of the human walking cycle or can be called Gait Energy Image because:
1. The silhouette image shows a walking human that has been converted to black and white.
2. The GEI is the accumulation of each walking human image over a complete cycle.
3. Pixels with low intensity values represent dynamic parts of the human.

GEI targets a specific representation of human gait and we use it as the basis for human gait recognition. [4].

Convolutional neural network (CNN)

This research uses a Convolutional Neural Network (CNN) architecture designed for individual recognition of human gait patterns, consisting of a convolutional layer section and a dense layer section for classification [17].

Convolutional layer

Multi-layer Conv2D with filter dimensions between 32, 64, and 128 is used to capture the spatial characteristics of the main image. ReLU activation is added to help capture complex patterns in the data (adding nonlinearity to the model). MaxPooling layers are used to reduce the resolution without losing any characteristic information. The first step is to use a selected set of frames as input to a convolutional layer. The convolutional layer is connected with a set of M filters into a set of T channels., and $A \times B$ compared to a mini-image set X with a channel T of size $Height \times Width$. The filter element is written as $E_{w,x,y,z}$, symbols and the image elements are symbolized by $F_{g,h,i,j}$. So the convolutional computation formula is defined as follows:

$$Y_{g,u,i,j} = \sum_{c=1}^{T}\sum_{u=1}^{R}\sum_{v=1}^{S} F_{g,h,i,j+u,j+z} - E_{w,x,y,z} \tag{2}$$

The result of the overall output filter image combination can be defined as follows:

$$Y_{g,w} = \sum_{c=1}^{T} F_{g,h} * E_{w,x} \tag{3}$$

where $*$ is a two-dimensional correlation. The figure 3 below shows the arrangement of convolutional layers in the proposed convolutional neural network.

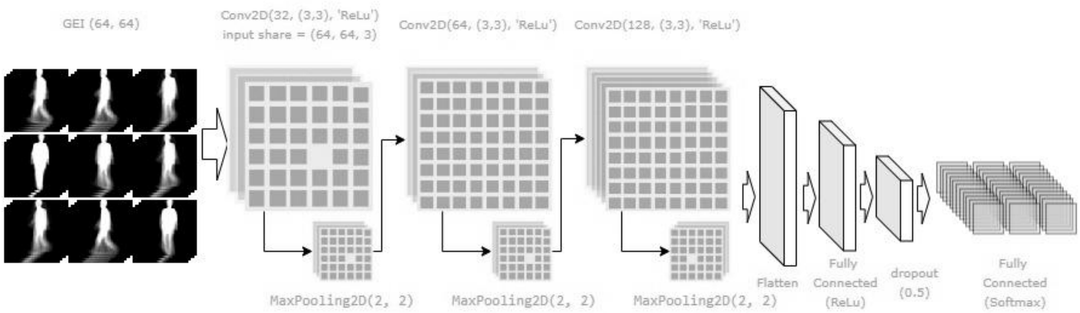

Figure 3. Convolutional neural network for human gait recognition

This CNN model starts with three convolutional layers (32, 64, and 128 filters) to capture more detailed and in-depth filters. These convolutional layers are combined with a 3x3 kernel to learn and capture important details, allowing for accurate distinction between various features.

From this convolutional layer, a 2x2 MaxPooling process is then performed to reduce the spatial dimensionality of the features. MaxPooling helps reduce sampling, making computation easier, and makes the model more robust to spatial variations in the image. This 2x2 pooling method preserves important features without compromising the character of each feature.

Dense layers

After the convolutional layer, dense layers with a total of 256 units are used to learn the complex data representation of the extracted features. The ReLU activation function is used to add nonlinearity, and dropout layers with a ratio of 0.5 are added to prevent overfitting.

Fully connected layer (dense)

The feature extraction results are then assembled into a fully connected layer. $f_{full} = f_{CNN}$. The results of the subsequent recognition process are generated through a SoftMax operation, the formula for which is as follows:

$$p = Softmax(f_{full} * w_p + b_p) \quad (4)$$

where p is the sample result, the output layer weight matrix is denoted by w_p, and the output bias is symbolized by b_p. Fully connected layer (FC) works on flattened inputs, where all neurons are connected to each input.

Output layer

The dense layer has the appropriate number of classes (human) for the dataset used. The SoftMax activation function is used to classify multi-class probabilities, enabling this model to predict all classes in the dataset. The model utilizes three convolutional layers that capture the characteristics of each image from the base to the deepest layers. Therefore, the use of three convolutional layers is able to maintain a balance between model complexity and computational efficiency, while reducing the risk of overfitting during the training process.

This study used a learning rate of 0.001 for the training process. This value reflects the complexity of parameter selection. A learning rate of 0.001 is considered neutral, aiming to find a balance between learning outcomes quickly. An epoch is the process of training large amounts of data by dividing it into smaller units [20]. If there is only one epoch, the training process is performed only once on all data. However, if there are more than one epoch, the training process is divided according to the number of epochs. This study compared epochs from 1 to 20 to determine which epoch achieves maximum accuracy [21]–[25]. This model uses a batch sample size of 16 samples, which are processed before the model is updated. This size maintains computational efficiency and achieves faster convergence by maintaining stochastic values and gradient reduction [26].

RESULTS AND DISCUSSIONS

This study employed deep learning techniques with a CNN model. Prior to extraction and recognition with the CNN, the original images from the Casia B dataset were segmented using the Gait Energy Image feature, which captures the characteristics of each gait. The GEI feature provides strong color for static body parts like the torso and head, and weak color for dynamic parts like the hands and feet. This can be used as a reference in the recognition process.

The GEI feature recognition used a CNN model with three filter parameters. The first filter used 32 filters, followed by a second filter with 64 filters, and the final filter with 128 filters. This allows for detailed and in-depth object recognition, resulting in extraction that further reveals characteristics.

The extracted filter results were mapped with varying fully connected parameters (128, 256, and 512) to determine recognition complexity. In the test, the epochs were divided between 1 and 20. The number of epochs indicates that the greater the number of epochs, the more accurate the recognition. The tests with these parameters showed varying accuracy and recognition speed, as shown in the figure 4.

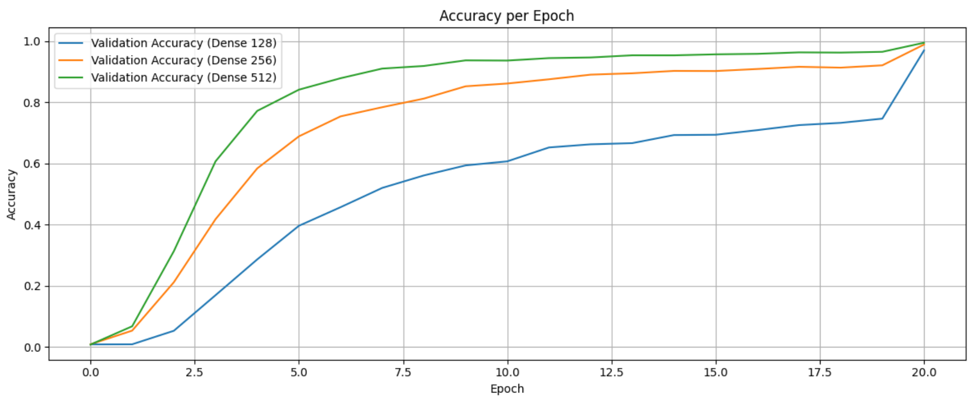

Figure 3. CNN test results

The figure 4 shows that the more epochs, the better the recognition. However, the more epochs, the longer the processing time. For comparison, the dense 512 and epochs 20 and 19 have a difference of 120 ms/step

(198 ms-78 ms). However, this study considers this time difference to be tolerable when compared to the accuracy difference between epochs 20 and 19 (99.48% and 96.51%). Therefore, this increase in accuracy is due to the large dataset size, which requires a larger number of epoch configurations to divide the tasks and will increase the recognition process time. The results of this study, when compared with previous research, can be formulated in the table 1 below.

Table 1. Comparison results of accuracy from previous research

No	Reasearch	Method	Accuracy
1	M. Karg [8]	Casia B Dataset - PCA - SVM	95%
2	C.-H. Huang [6]	Casia B Dataset - GEI - Gabor Wavelet - PCA	88,5%
3	F. I. Pratama [12]	Casia B Dataset - GEI - Gabor Wavelet - PCA - KNN	98.50%
5	N. Aman [17]	Casia B Dataset - CNN	97.12%
6	**F. I. Pratama (Ours)**	**Casia B Dataset - GEI - CNN**	**99.48%**

The table 1 above shows an evaluation of the increasingly accurate performance of human gait recognition using multiple viewpoints, with an accuracy of 99.48%. This deep learning technique outperforms conventional techniques used in previous studies due to the CNN model's ability to read detailed characteristics of objects, as well as the appropriate epoch selection in human gait recognition.

CONCLUSION

The conclusion of this study shows that deep learning technology is able to demonstrate better performance than conventional models. Deep learning technology using the CNN method is able to perform good extraction and recognition supported by the GEI algorithm to produce a complete and continuous gait cycle characteristic. Deep learning technology using the CNN method is capable of extracting and recognizing a maximum amount of data supported by the GEI algorithm as a feature extraction that will be classified using CNN to produce complete and continuous step cycle characteristics. The ideal epoch parameter for the dataset used is 20, and the ideal fully connected selection is 512 units. Therefore, the combination of GEI and CNN from this study is ideal for use in gait pattern recognition with various viewing angles from 0^0 - 180^0. Based on the analysis, the selection of CNN method parameters affects the level of accuracy and time. Therefore, this CNN has advantages in image extraction and its deep recognition is able to distinguish between images that produce an accuracy level of 99.48%. Therefore, the combination of algorithms and parameter selection can be used for recognition of every human gait pattern.

The development suggestion for this research is to add human walking characteristics to the recognition of human gait patterns, such as the condition of humans carrying various items and clothes to the training data, so this requires the development of recognition techniques that can be used to recognize various human patterns.

ACKNOWLEDGEMENTS

The author would like to thank for Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) of Wahid Hasyim University for providing support in funding this research activity through DIPA at Competitive Research Scheme in 2025.

REFERENCES

[1] N. M. Bouchrika I, Goffredo M, Carter J, "On using gait in forensic biometrics," *J. Forensic Sci.*, vol. 56, no. 4, pp. 882–889, 2011.

[2] Y. Makihara, Y.; Nixon, M.S.; Yagi, "Gait Recognition: Databases, Representations, and Applications.," in *Springer: Cham Switzerland*, 2021, pp. 1–15.

[3] I. Rida, X. Jiang, and G. L. Marcialis, "Human body part selection by group lasso of motion for model-free gait recognition," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 154–158, 2016, doi: 10.1109/LSP.2015.2507200.

[4] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, 2006, doi: 10.1109/TPAMI.2006.38.

[5] Z. Lai, Y. Xu, Z. Jin, and D. Zhang, "Human gait recognition via sparse discriminant projection learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1651–1662, 2014, doi: 10.1109/TCSVT.2014.2305495.

[6] C.-H. Huang, Deng-Yuan; Lin, Ta-Wei; Hu, Wu-Chih; Cheng, "Gait Recognition Based On Gabor Wavelets And Modified Gait Energy Image For Human Identification," *J. Electron. Imaging*, vol. 22, 2013.

[7] R. H. dan T. T. S. Zheng, J. Zhang, K. Huang, “Robust View Transformation Model for Gait Recognition,” 2011.
[8] M. Karg, K. Kühnlenz, and M. Buss, “Recognition of affect based on gait patterns,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 40, no. 4, pp. 1050–1061, 2010, doi: 10.1109/TSMCB.2010.2044040.
[9] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette Analysis-Based Gait Recognition for Human Identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, 2003, doi: 10.1109/TPAMI.2003.1251144.
[10] A. Budianita and F. I. Pratama, “Penerapan Algoritma Klasifikasi Dengan Fitur Seleksi Weight By Information Gain Pada Pemodelan Prediksi Kelulusan Mahasiswa,” *Infotekmesin*, vol. 11, no. 2, pp. 80–86, Jun. 2020, doi: 10.35970/INFOTEKMESIN.V11I2.255.
[11] R. R. I. dan a. A. A. Z. A. L. Kurniawan, “Perancangan Sistem Pengenalan Wajah Menggunakan Metode Ekstraksi Ciri Susunan Tapis Wavelet Gabor 2D Dengan Euclidean Distance,” *TRANSIENT*, vol. 4, no. 1, p. 40, 2015.
[12] F. I. Pratama and A. Budianita, “Optimization of K-Nn classification in human gait recognition,” 2020.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[14] A. Mehmood *et al.*, “Prosperous Human Gait Recognition: an end-to-end system based on pre-trained CNN features selection,” *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 14979–14999, 2024, doi: 10.1007/s11042-020-08928-0.
[15] W. Kong, M. H. Saad, M. A. Hannan, and A. Hussain, “Human gait state classification using artificial neural network,” *2014 IEEE Symp. Comput. Intell. Multimedia, Signal Vis. Process.*, pp. 1–5, 2014, doi: 10.1109/CIMSIVP.2014.7013287.
[16] C. C. Wu, C. W. Tsai, F. E. Wu, C. H. Chiang, and J. C. Chiou, “Plantar Pressure-Based Gait Recognition with and Without Carried Object by Convolutional Neural Network-Autoencoder Architecture,” *Biomimetics*, vol. 10, no. 2, 2025, doi: 10.3390/biomimetics10020079.
[17] N. Aman, M. R. Islam, M. F. Ahamed, and M. Ahsan, “Performance Evaluation of Various Deep Learning Models in Gait Recognition Using the CASIA-B Dataset,” *Technologies*, vol. 12, no. 12, 2024, doi: 10.3390/technologies12120264.
[18] Y. Sari, Y. F. Arifin, Novitasari, and M. R. Faisal, “The Effect of Batch Size and Epoch on Performance of ShuffleNet-CNN Architecture for Vegetation Density Classification,” *ACM Int. Conf. Proceeding Ser.*, pp. 39–46, Nov. 2022, doi: 10.1145/3568231.3568239;ISSUE:ISSUE:DOI.
[19] M. K. Qetheth, A. A. D. Al-Zuky, and B. H. Hamza, “Exploring the Impact of Epoch Value Variation on Human Detection System Training across Diverse Distances using the Faster R-CNN Technique,” *AIP Conf. Proc.*, vol. 3211, no. 1, May 2025, doi: 10.1063/5.0257225/3346007.
[20] A. Johny and K. N. Madhusoodanan, “Dynamic Learning Rate in Deep CNN Model for Metastasis Detection and Classification of Histopathology Images,” *Comput. Math. Methods Med.*, vol. 2021, no. 1, p. 5557168, Jan. 2021, doi: 10.1155/2021/5557168.
[21] R. Gaetano, D. Ienco, K. Ose, and R. Cresson, “A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery,” *Remote Sens. 2018, Vol. 10, Page 1746*, vol. 10, no. 11, p. 1746, Nov. 2018, doi: 10.3390/RS10111746.
[22] T. Karim *et al.*, “CNN Applied In Public Transport For The Protection Against The Covid-19 Spread,” *Int. J. Eng. Trends Technol. - IJETT*, vol. 69, no. 10, pp. 33–37, Oct. 2021, doi: 10.14445/22315381/IJETT-V69I10P205.
[23] A. Y. Saleh and E. Liansitim, “Palm oil classification using deep learning,” *Sci. Inf. Technol. Lett.*, vol. 1, no. 1, pp. 1–8, 2020, doi: https://doi.org/10.31763/sitech.v1i1.1.
[24] M. H. Ghaeminia, A. Badiezadeh, and S. B. Shokouhi, “An Efficient Energy Model for Human Gait Recognition,” pp. 1–6, 2016.
[25] N. L. Husni *et al.*, “Real-Time Littering Activity Monitoring Based on Image Classification Method,” *Smart Cities 2021, Vol. 4, Pages 1496-1518*, vol. 4, no. 4, pp. 1496–1518, Dec. 2021, doi: 10.3390/SMARTCITIES4040079.
[26] R. Lin, “Analysis on the Selection of the Appropriate Batch Size in CNN Neural Network,” *Proc. - 2022 Int. Conf. Mach. Learn. Knowl. Eng. MLKE 2022*, pp. 106–109, 2022, doi: 10.1109/MLKE55170.2022.00026.