**GUB Journal of Science and Engineering (GUBJSE)**
**p-ISSN: 2409-0476**
**vol. 10, no. 01, 2023**

# Evaluating CNN Models for Gait Recognition: A Study on the CASIA-B Dataset

Md. Mehedi Hasan[1], Mohammad Asif Ul Haq[1], Md. Hasan Maruf[1], Nakib Aman[2]

1 Department of CSE, Varendra University, Rajshahi, Bangladesh
2 Department of CSE, Pabna University of Science and Technology, Pabna, Bangladesh

*Corresponding author's email: nakibaman@gmail.com

Received: 12/02/2023
Accepted: 01/12/2023
Published: 15/07/2024

Data Availability: The data are available on request from the corresponding author.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DOI: 10.3329/gubjse.v10i1.74941



## Abstract

Gait recognition, a form of biometric identification, has garnered considerable interest for its ability to identify individuals from a distance. Yet, its effectiveness depends significantly on the precision of the underlying categorization models. This study thoroughly evaluates many Convolutional Neural Network (CNN) models to determine their effectiveness in properly identifying and categorizing gait patterns. We train and test several CNN architectures using advanced deep learning techniques on the widely known CASIA-B dataset, which is a benchmark dataset for gait recognition research. The analyses involve models like as VGG16, VGG19, NASNetLarge, NASNetMobile, EfficientNetB0, and Xception, each well-known for their effectiveness in picture categorization tasks. We evaluate the performance of these models by conducting thorough experiments and detailed analysis, focusing on accuracy, validation loss, and other pertinent metrics. The Xception model demonstrated the highest accuracy of 97.17% of the models assessed. This model frequently surpasses similar models, demonstrating its strength and effectiveness in precisely recognizing gait patterns. On the other hand, the accuracy of other models varied from 9.78% to 93.85%, demonstrating the variable levels of efficacy among various designs. The findings have significant implications for the creation and implementation of gait recognition systems. The high precision demonstrated by the Xception model highlights the promise of CNN-based methods in progressing gait identification technology. Our work highlights the significance of choosing suitable model architectures for achieving the best performance in gait detection tasks. In the future, research may focus on using larger datasets and exploring other CNN architectures to improve the accuracy and reliability of gait detection systems. Our goal is to enhance gait recognition technology to improve biometric identification for security, surveillance, and healthcare applications.

**Keywords:** Gait recognition, CNN model, CASIA B dataset, Deep learning, Computer Vision.

## Highlights

- CNN models achieve 97.17% accuracy in gait recognition.
- Xception model outperforms others with highest accuracy.
- Importance of selecting suitable model architectures.
- Potential for advancing gait identification technology.

# 1 Introduction

GAIT recognition is a biometric identification method that uses an individual's unique walking pattern to identify them. This method provides a unique benefit compared to other biometric methods like face, fingerprint, and iris identification because it does not require physical touch with the individual. Gait recognition may be used in distant and uncontrolled contexts to categories unfamiliar objects and recognize familiar persons.

Gait recognition has many uses in several areas such as security, surveillance, and medical diagnosis, contributing to its increasing popularity in recent times. It may be used for immediate suspect identification in surveillance and security, identifying patient gait anomalies in medical diagnostics, and analyzing athletes' performance in sports.

Gait recognition involves extracting information from gait sequences to identify and categories individuals. The process consists of three essential steps: data gathering, feature extraction, and classification.

Data gathering is done via wearable sensors, depth sensors, and video cameras. The gathered data is preprocessed to remove any noise and outliers. Feature extraction is the process of extracting pertinent data from the gait sequence. Possible extracted characteristics include spatiotemporal, frequency domain, and shape-based features. These characteristics are subsequently utilized to recognize and categories individuals.

Deep learning approaches have greatly improved the efficacy and efficiency of gait recognition systems. Convolutional Neural Networks (CNNs), a deep learning model, have demonstrated encouraging outcomes in gait identification. These algorithms, which can automatically extract pertinent information from sequences of walking patterns and have been trained on extensive datasets, have enhanced the precision and efficiency of gait detection.

This research introduces a method designed to categorize human gaits, helping in recognizing people by their walking patterns. We initially isolate silhouette pictures from the CASIA B collection, which includes recordings of humans walking at different speeds and angles. The images are utilized for training and recognizing a deep learning model, especially a Convolutional Neural Network (CNN).

We found that the CNN model outperformed other models in accuracy and speed throughout our test. In this work, we utilized the Xception model, known for its effectiveness in many images' recognition tasks. The Xception model helps extract important information from gait sequences, improving the accuracy of gait identification.

Ultimately, gait recognition shows promise as a biometric identification method that might be applied in many industries. Deep learning techniques, such as Convolutional Neural Networks (CNNs), have greatly enhanced the precision and efficiency of gait detection systems. The suggested approach for categorizing human walking.

# 2 Related Work

There are two main techniques to gait recognition: model-free and model-based. A model-free technique is simpler and may be identified in low-resolution photos. GEI is one of the techniques in this area. It is the most often utilized method. A gait cycle is required to calculate a GEI. Several aspects are affected by the precision of gait recognition. Gait recognition would not be successful in this scenario. J. Man and B. Bhanu [1] utilized a novel gait representation known as Gait Energy Image (GEI) and compared it with previous gait representation methods using the USF HumanID Dataset [2]. The results indicate that GEI is a successful and efficient method for representing gait, and it performs competitively compared to other gait representation approaches. Another research introduced a novel approach for human gait detection called Gait Gaussian Image (GGI), which is based on spatiotemporal information. GGI is a gait analysis approach that focuses on extracting features of gait images within a single gait cycle. The suggested characteristics are utilised to calculate Euclidean distances among the closest neighbouring categories. Gaussian gait signatures provide a notable enhancement in the identification accuracy of CASIA-B and Soton datasets under typical walking situations when compared to alternative techniques. Y. Guan et al. [3] proposes a strategy for addressing corruption involving unknown locations using an RSM-based classification ensemble as a partial feature. Corruption's location may vary depending in the direction of the questions, causing important elements to become irrelevant when the walking position changes. To address the issue, a classifier ensemble technique was suggested utilising the random subspace method (RSM) and majority voting (MV). Two techniques, Local Enhancement (LE) and Hybrid Decision Level Fusion (HDF), were employed to reduce the ratio of false votes to real votes in the suggested approach before Majority Voting (MV). The USF dataset and the OU-ISIR-B dataset assess approaches against difficult factors including clothes, walking surface, and time elapsed, which are more effective than other advanced algorithms.

Model-based gait classification approaches generate gait representations from video without explicitly analysing the underlying structure of the human body. L. Yao et al. [4] introduces a sophisticated model called Skeleton Gait Energy Image (SGEI), which is based on robust skeleton points generated from a two-branched multi-stage CNN network. The suggested SGEI model has demonstrated improved strength and precision. The proposed CNN-based structures with SGEI are effective in improving the robustness of gait detection in various contexts, such as different views and clothing variations.

# 3 Data Collection and Preprocessing

## 3.1 Dataset

The entire experiment was performed on the CASIA B dataset [5]. This dataset contains 124 subsets, where every subject contains three variations, i.e., view angle, clothing, and carrying condition and each variation contains 11 views. This proposed work used a total of 54150 images in 10 subjects. These images from the CASIA B dataset were resized into 72 X 72. As a result, a total of 50000 images were used for training, and 4150 images for testing.

**Table 1.** Data Distribution Table

| Classes | Training Image per Class | Testing Image per Class | Total |
|---|---|---|---|
| 10 distinct subjects | 50000 | 4150 | 54150 |

## 3.2 Environment Setup

For this experiment, the hardware specification is the i5 processor, RAM is 8 GB and has a 2 GB Nvidia graphics card. Keras neural network library was used for the image classification. The selected images were sorted in particular directory order.

# 4 Methodology

The CASIA dataset B [5] used in our experiments provided segmented silhouette images. So, in this paper, we focus on training and classification. For person identification the silhouette images were selected as the parameter because the frames give a summary of effective parts of the subjects. In the flowchart in Figure 1, the workflow of segmentation, training, and classification is mentioned. The whole process is divided into the steps of gait image segmentation and then training the calculated data for the classification of effective parts of each subject. We will first prepare the dataset and separate the images. We'll first split the contents of the folder into the train and validation directories. Then, in each directory, we'll create a separate directory for each individual subject that will contain only images for that subject. After cropping the unwanted portion of images, the images were resized to 72×72. The training and validation data have a ratio of 80:20.

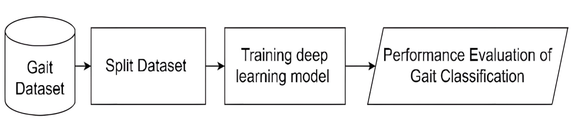



**Fig 1.** Methodology flowchart.

We will perform gait classification on the CASIA B datasets with 54150 images inside 10 subjects, comparing which model best classifies gait subjects with various CNN models.

Deep learning has completely changed artificial intelligence, especially when it comes to speech and image recognition. Deep learning uses a lot of data to train sophisticated models that can automatically identify patterns and characteristics in the data.

Deep learning models known as convolutional neural networks (CNNs) are particularly effective for picture recognition tasks like gait recognition. Convolutional, pooling, and fully linked layers are among the many layers that make up CNNs.

The fundamental units of CNNs are convolutional layers, which are a collection of filters used to extract local characteristics from the input image. The following layer receives the feature maps that were produced as a result of the filters' sliding window scan of the input image.

In order to down sample the feature maps and decrease the number of parameters in the model and avoid overfitting, pooling layers are employed to minimize the spatial dimensions of the feature maps. The output from preceding layers is finally provided to the completely connected layer. All of the neurons in this layer are interconnected, creating a high-dimensional feature vector that is then applied to classification.

The silhouette images from the CASIA-B dataset are classified using CNNs in the suggested method for gait identification. The fully connected layer of the CNN architecture creates a high-dimensional feature vector that is utilized for classification while also automatically extracting features and patterns from the images.

In general, CNNs have shown to be quite efficient at tasks requiring image recognition, and their application to gait recognition holds considerable potential for enhancing the reliability and accuracy of biometric identification systems.

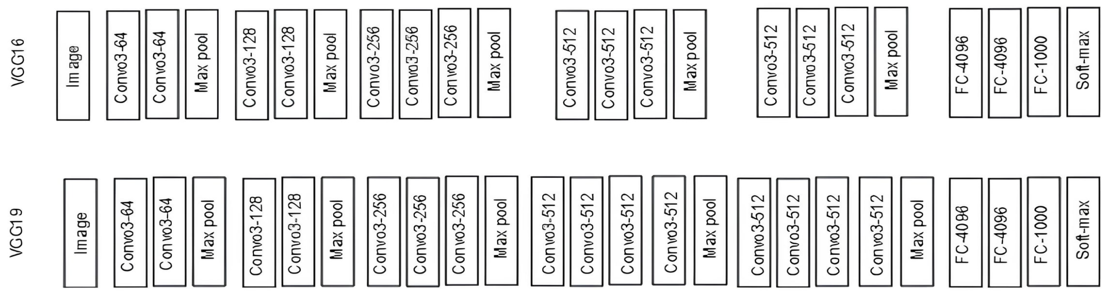



**Fig 2.** VGG 16 and VGG 19 Architecture

## 4.1 VGG (Visual Geometry Group)

In computer vision, the VGG (Visual Geometry Group) network architecture is commonly used in deep learning models. In 2014, the VGG network was initially suggested by the Visual Geometry Group at the University of Oxford. The design consists of many convolutional layers with small 3x3 filters, followed by layers utilizing max-pooling, and concluding with fully linked layers. To enhance network depth without compromising receptive field size, more layers can be stacked using compact filters.
The VGG network has been utilized for many tasks such as segmentation, object identification, and image classification. It has shown remarkable performance on benchmark datasets such as ImageNet [6]. The pre-trained model is fine-tuned for a specific task via transfer learning on a smaller dataset, based on the VGG architecture [7].
The VGG network has shown superior performance in image classification tasks compared to other CNN models [8]. The VGG16 and VGG19 models, with 16 and 19 weight layers respectively, are commonly used in computer vision applications.

## 4.2 NASNet (Neural Architecture Search Network)

The Google Brain team developed the neural network structure called NASNet, short for Neural Architecture Search Network. Barret Zoph and Quoc V. Le initially addressed this topic in their 2018 publication "Learning Transferable Architectures for Scalable Image Recognition" [9]. The NASNet architecture was developed using an advanced neural architecture search technique that outperformed previous state-of-the-art models on the ImageNet dataset.
There are two fundamental cell kinds in the NASNet architecture: normal cells and reduction cells. Feature extraction is performed with normal cells, and the feature maps are then down sampled using reduction cells. This design offers superior precision and reduced computational complexity among its other advantages.
There are two iterations of NASNet: NASNetLarge and NASNetMobile. NASNetLarge is more suitable for deployment on high-performance computing equipment such as desktop PCs and servers due to its higher parameter count compared to NASNetMobile. NASNetMobile is designed for low-power mobile devices such as smartphones and tablets. The NASNet architecture has shown significant improvements in accuracy for large-scale image identification tasks compared to previous state-of-the-art models. As a result, it is commonly utilized in gait recognition and other computer vision applications.

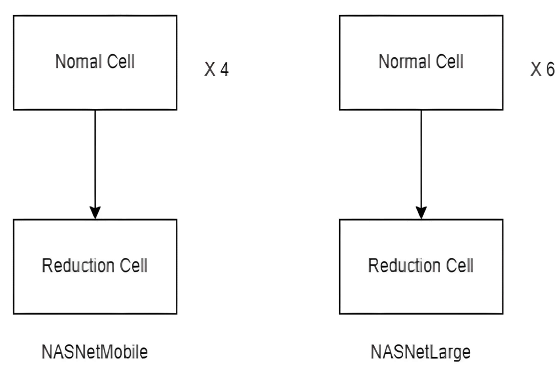



**Fig 3.** CNASNet Cell.

## 4.3 EfficientNets

Mingxing Tan and Quoc V. Le's EfficientNets model scaling approach has gained significant attention for its outstanding performance in computer vision challenges. EfficientNets use predefined scaling coefficients to equally scale each dimension, unlike conventional methods that randomly scale network dimensions. This method allows

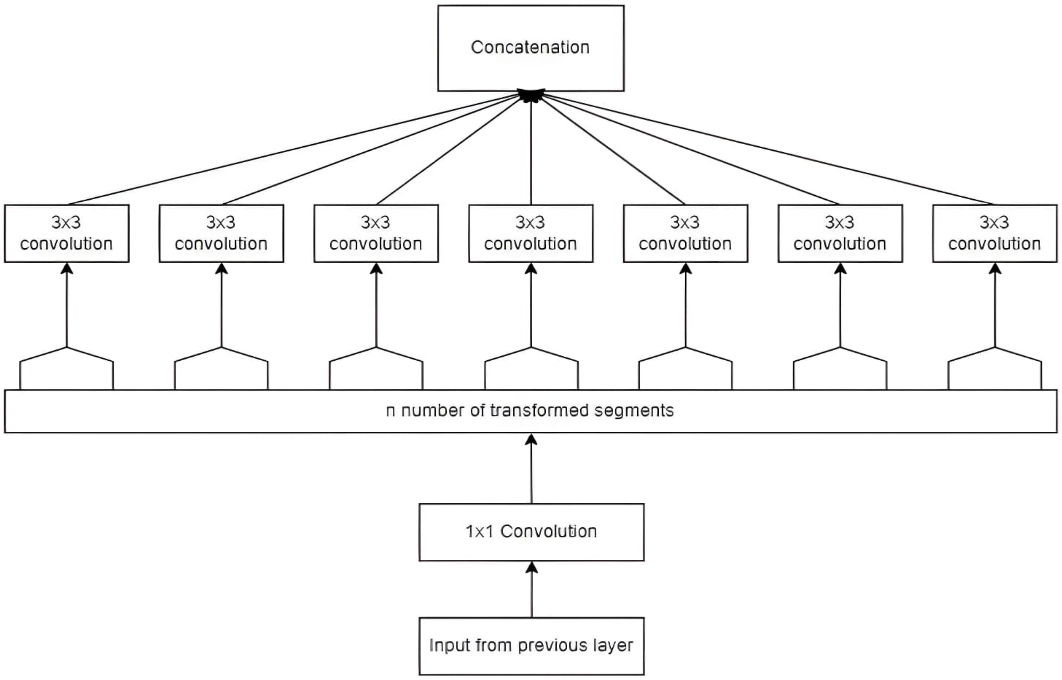
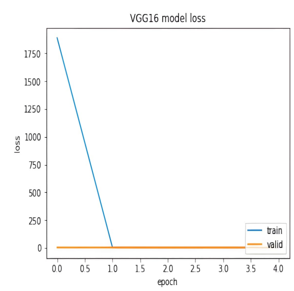


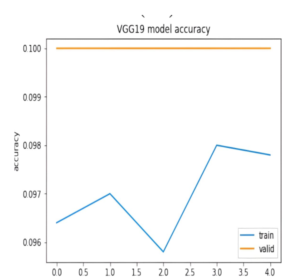

**Fig 4.** Xception block's architecture.

for more structured and efficient scaling of CNNs. The authors developed a new baseline network optimized for accuracy and efficiency by utilizing the AutoML MNAS framework.
EfficientNets have shown high performance on several computer vision benchmarks, particularly ImageNet [10]. EfficientNetB0 shares a comparable structure with NASNetMobile but use the same bottlenecks as MobileNetV2. The EfficientNet models have demonstrated high accuracy and efficiency in a range of image classification applications.

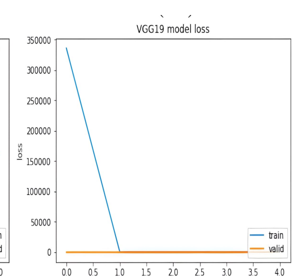

(a1)




(a2)



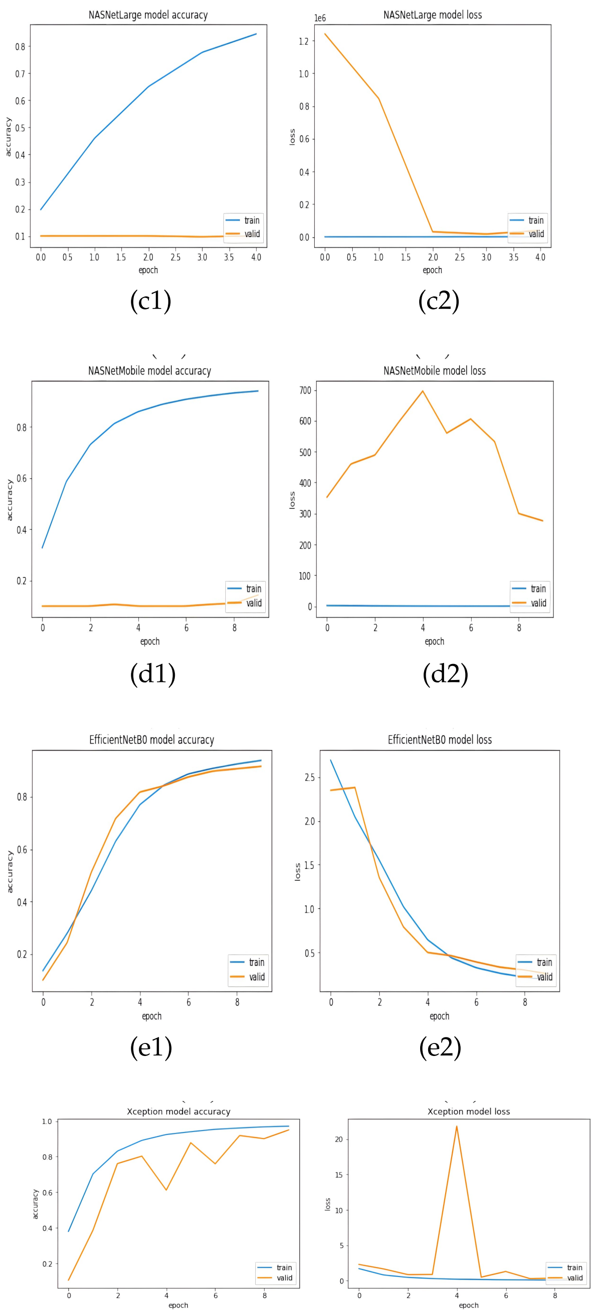

(b1)




(b2)

## 4.4 Xception

Chollet et al. introduced the Xception convolutional neural network architecture in 2016 [11]. The Inception architecture [12] has been expanded by including depth wise separable convolutions instead of the conventional Inception modules. The Xception architecture simplifies computation by using pointwise convolution followed by separate convolution of each feature-map along spatial axes. This method enhances performance and boosts the network's computational efficiency. Xception utilizes a series of depth-wise separable convolutions in its transformation process. Operating on the whole input feature map in a traditional convolutional layer might be compu-




(c1) (c2)

(d1) (d2)

(e1) (e2)

(f1) (f2)

**Fig 5.** Overall caption for Figure 5 covering (a1), (a2), (b1), (b2), etc.

tationally costly. The depth wise separable convolutional layer performs individual spatial convolutions on each channel and then combines the results using a pointwise convolution. This approach yields superior performance and necessitates less computational effort.

The Xception architecture has shown strong performance across several computer vision tasks such as semantic segmentation, object recognition, and picture classification. It has achieved state-of-the-art performance on many benchmark datasets like as ImageNet and CIFAR-10.

**Table 2.** A holistic comparison of different CNN models

| Model | Loss | Accuracy | Val_loss | Val_ accuracy |
|---|---|---|---|---|
| VGG16 | 2.159 | 0.1845 | 2.0334 | 0.2333 |
| VGG19 | 2.303 | 0.0978 | 2.3026 | 0.1000 |
| NASNet-Large | 0.448 | 0.8441 | 39880.16 | 0.1000 |
| NASNet Mobile | 0.178 | 0.9385 | 276.1620 | 0.1431 |
| Efficient-NetB0 | 0.183 | 0.9380 | 0.2498 | 0.9154 |
| Xception | 0.086 | 0.9717 | 0.1718 | 0.9501 |

## 5 Result and Performance Analysis

Classifying specific persons based on their walking patterns in computer vision requires the use of advanced algorithms to achieve high levels of accuracy. The authors proposed a deep learning approach for classifying gait silhouettes using basic convolutional neural networks (CNNs). The CNN models utilized in this work include VGG16, VGG19, NASNetLarge, NASNetMobile, EfficientNetB0, and Xception.

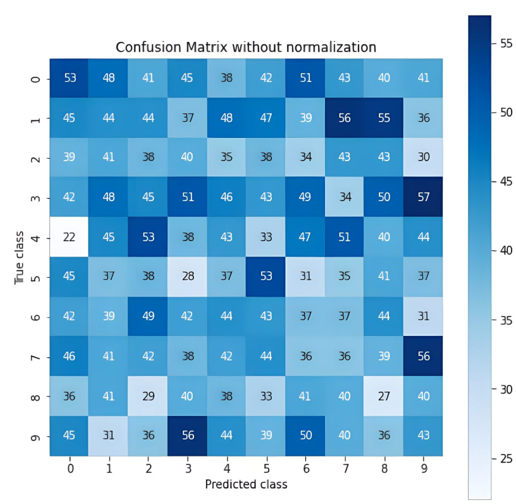



**Fig 6.** Confusion Matrix of Xception Model without Normalization

This study's main contribution is the use of pixel values from gait silhouettes, which proves to be a more effective approach compared to depending on overall features. Various measures including as accuracy, loss, validation accuracy, and validation loss are employed for training and evaluating CNN models. The results showed that the Xception model outperformed other models in terms of accuracy and had the lowest loss values. Moreover, all models exhibited improved performance following the implementation of data augmentation approaches.

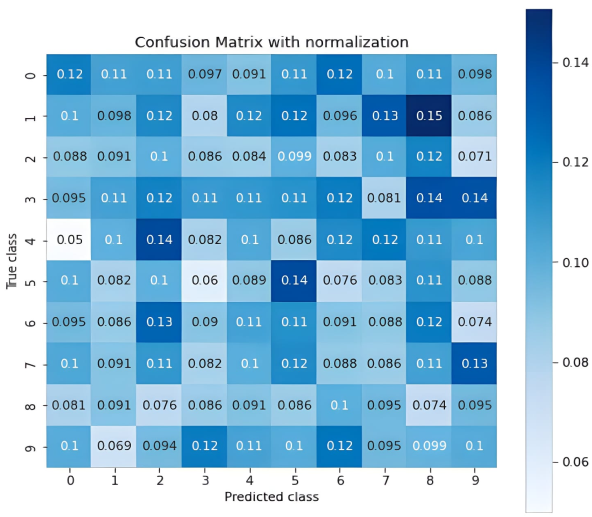



**Fig 7.** Confusion Matrix of Xception Model with Normalization

The study illustrates the efficiency of deep learning-based strategies for gait recognition and demonstrates how vital it is to pick the correct CNN model in order to obtain high accuracy. This study's findings might aid in creating systems capable of automatically recognizing individuals by their unique walking patterns, with potential uses in security, surveillance, and healthcare.

The Xception model outperformed other models in terms of accuracy and loss values, demonstrating the effectiveness of using CNN models for gait identification. Data augmentation approaches improved the efficacy of each model. The results of this study might be quite informative for scholars and experts engaged in the research of gait recognition.

The findings be seen in Table 2. The algorithms are compared based on the utilization of data augmentation, training accuracy, training loss, validation accuracy, and validation loss.

The analysis findings showed that the Xception model had the highest accuracy compared to all other models examined. This model generated the highest and most precise result, with an output accuracy of around 97.17%. Six distinct CNN models were trained on the dataset to classify individuals based on their gait silhouettes. The

**Table 3.** Output of the xception models

| Epoch no. | Loss | Accuracy | Val_loss |
|---|---|---|---|
| 1 | 1.7148 | 0.3803 | 2.3009 |
| 2 | 0.8184 | 0.7031 | 1.6763 |
| 3 | 0.4714 | 0.8314 | 0.8683 |
| 4 | 0.3072 | 0.8919 | 0.8830 |
| 5 | 0.2176 | 0.9249 | 21.8194 |
| 6 | 0.1700 | 0.9405 | 0.5073 |
| 7 | 0.1352 | 0.9541 | 1.3042 |
| 8 | 0.1123 | 0.9615 | 0.3024 |
| 9 | 0.0989 | 0.9681 | 0.3709 |

models used were VGG16, VGG19, NASNetLarge, NASNetMobile, EfficientNetB0, and Xception.

The study analyzed several models based on the use of data augmentation and the values of training accuracy, training loss, validation accuracy, and validation loss. The results indicated that Xception outperformed the other models in terms of accuracy. The Xception model has superior performance in recognizing gaits from silhouette data, as indicated by this study.

Figures 5 (a) - (f) display the training histories of the models, showcasing model accuracy and model loss. The plots illustrate the evolution of accuracy and loss for each model during training. The Xception model consistently achieved the best accuracy and lowest loss during the training procedure. The model successfully learned from the training data and demonstrated excellent accuracy when applied to the validation dataset, indicating its ability to generalist.

The study highlights the effectiveness of deep learning techniques using simple convolution to identify gaits from silhouette data. Furthermore, it suggests that the Xception model is very accurate and effective for this task. These findings may impact the advancement of more accurate and efficient gait recognition systems used in surveillance, biometric identification, and healthcare monitoring

Table 2I's comparison of the six models' performances shows that the Xception model had the highest accuracy and lowest validation loss, making it the top performer. Figures 5 (f) and 6 (f) display the training and validation history curves for the Xception model, illustrating that the model underwent training for 10 epochs using a batch size of 200. The Xception model achieved a training accuracy of 97.17% and the lowest training loss (categorical_crossentropy), indicating successful learning from the dataset. The Xception model had the lowest validation loss of 0.1718, indicating its superior accuracy in predicting the dataset compared to other models.

Table 3 displays the training and validation accuracy and loss data of the Xception model. The Xception model had the highest accuracy and the lowest loss values in the findings, establishing it as the optimal model for predicting the gait silhouette dataset. The Xception model, utilizing separable convolutions to optimize computation without compromising accuracy, outperformed the other models. This model is suitable for applications such as gait identification in picture classification due to its ability to process high-dimensional input data.

This study's results demonstrate the effectiveness of deep learning models in gait detection tasks and underscore the importance of selecting the appropriate model architecture for the specific challenge. Table 3 displays the accuracy and loss figures for training and validation of the Xception model.

A confusion matrix is a tabular representation that compares the observed values of the target variable with the predicted values to evaluate the performance of a classification model. To calculate performance metrics such as precision, recall, and accuracy, the system presents the counts of true positives, false positives, true negatives, and false negatives.

Recall is the ratio of true positive predictions to all real positive values, whereas accuracy is the ratio of true positive predictions to all positive predictions. The metrics derived from the confusion matrix can evaluate the efficiency of a classification system and identify areas for improvement. If all positive classes are correctly predicted, the algorithm is doing well with high accuracy and recall for that specific class. The effectiveness of search results may be assessed by the number of good outcomes.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})} \tag{1}$$

where TP and FP define the true positive and false positive values in (1). Here FN is called False Negative result in (2).

$$\mathrm{Recall} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \tag{2}$$

Upon completion of model evaluation, it generates a N $\times$ N confusion matrix for multiclass classification. If the dataset has N classes, the matrix will have dimensions of N $\times$ N. Figure 6 displays a confusion matrix for the Xception model with dimensions of $10 * 10$, representing 10 classes. The confusion matrix consists of four categories: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). When the actual value matches the expected value, it is considered really positive. The sum of the numbers in the corresponding rows, less the actual positive values, is the false negative value for a class. The false-positive value for a class is calculated by summing the values in the appropriate columns, excluding the actual positive values. The sum of all column and row values, minus the values of the class being analyzed, will provide the true negative value for that class.

A confusion matrix is a table that displays the effectiveness of a classification model by comparing predicted labels against actual labels. The confusion matrix is shown in Figure 6 without any normalizing. The instances in each row of the matrix correspond to a predicted class, whereas the examples in each column correspond to an actual class. The matrix's diagonal displays the model's accurate predictions, while the off-diagonal components highlight the occurrences that were incorrectly categorized.

A confusion matrix without normalization, however, cannot provide the relative frequencies of examples in each class that are incorrectly identified. A normalized confusion matrix can be used to remedy this, as seen in Figure 7. We may compare the mistakes made by the model in other classes thanks to the normalization of the confusion matrix, which gives a more realistic depiction of the model's performance. The model is marginally more accurate at predicting the first class (walking forward) than the second class (walking with a bag), according to the normalized confusion matrix. For the first class, the model accurately predicted 8% of cases, but only 81% for the second class. Additionally, the model frequently misclassified instances from the second class as belonging to the first class due to a higher rate of false negatives for the second class. The normalized confusion matrix offers a more thorough insight of the model's performance overall and may be used to guide decisions about how to enhance the model.

**Table 4.** Classification report of xception model

| No. | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 0.13 | 0.12 | 0.12 | 442 |
| 1 | 0.11 | 0.10 | 0.10 | 451 |
| 2 | 0.09 | 0.10 | 0.10 | 381 |
| 3 | 0.12 | 0.11 | 0.12 | 465 |
| 4 | 0.10 | 0.10 | 0.10 | 416 |
| 5 | 0.13 | 0.14 | 0.13 | 382 |
| 6 | 0.09 | 0.09 | 0.09 | 408 |
| 7 | 0.09 | 0.09 | 0.09 | 420 |
| 8 | 0.07 | 0.07 | 0.07 | 365 |
| 9 | 0.10 | 0.10 | 0.10 | 420 |

Figure 7 confusion matrix provides a clearer insight into the model's performance compared to the confusion matrix in Figure 7 without normalization. The efficiency of different classes may be compared by scaling the values in the confusion matrix using the normalizing approach. The true positive rate (TPR) in Figure 7 represents the proportion of positive instances correctly identified by the model. TPR is displayed for each class. Class 5 has the highest TPR while class 8 has the lowest TPR, with TPR values in

each class ranging from 0.074 to 0.14.
Table 4 contains supplementary performance metrics for each category, including F1-score, recall, and accuracy. The F1-score is calculated as the harmonic mean of accuracy and recall. accuracy represents the ratio of correctly predicted positive instances to all predicted positive instances, while recall indicates the proportion of actual positive cases that are correctly detected. In addition to accuracy, these metrics provide a comprehensive evaluation of the model's performance. The model exhibits varying levels of performance across different classes, indicating room for improvement in some areas.

# 6 Conclusion

The research presented a convolutional neural network (CNN) gait classification system that achieved an accuracy level of 97.17% on the CASIA B dataset. The study used pixel values from the gait silhouette as input for the CNN models instead of global characteristics. Various CNN models such as VGG16, VGG19, NASNetLarge, NASNetMobile, EfficientNetB0, and Xception were evaluated to identify the top-performing model.
Nevertheless, the study also identified several constraints. The model exhibits misclassifications, seen from the confusion matrix without normalization, indicating the necessity for further enhancements. To overcome this limitation, it is recommended to use a confusion matrix with normalization.
To enhance the accuracy of the models, it is advisable to compare several CNN models and utilize a larger gait dataset in future studies. They recommended examining the influence of additional parameters, such as batch size and number of epochs, on the accuracy of the models.
The Xception model demonstrated superior performance compared to all other models, achieving the greatest accuracy. This signifies a significant progress compared to previous studies that manually designed features and achieved lower accuracy. The results also showed that utilizing data augmentation improved the accuracy of the models.
The suggested CNN-based gait classification technique has significant potential for application in biometric identification and surveillance systems. The high accuracy of the Xception model suggests its potential use in real-world scenarios. The study establishes a foundation for future research to improve the accuracy and efficiency of gait classification systems.

# References


[1] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316–322, 2005.

[2] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 2, pp. 162–177, 2005.

[3] Y. Guan, C.-T. Li, and F. Roli, "On reducing the effect of covariate factors in gait recognition: A classifier ensemble method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 7, pp. 1521–1528, 2014.

[4] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, and Z. Tang, "Robust cnn-based gait verification and identification using skeleton gait energy image," in *2018 digital image computing: techniques and applications (DICTA)*, IEEE, 2018, pp. 1–7.

[5] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th international conference on pattern recognition (ICPR'06)*, IEEE, vol. 4, 2006, pp. 441–444.

[6] J. Deng, "A large-scale hierarchical image database," *Proc. of IEEE Computer Vision and Pattern Recognition, 2009*, 2009.

[7] N. Bar, D. Blas, K. Blum, and H. Kim, "Assessing the fornax globular cluster timing problem in different models of dark matter," *Physical Review D*, vol. 104, no. 4, p. 043 021, 2021.

[8] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[9] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[10] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

**Md. Mehedi Hasan** is a Software Engineer at Developer eXperience Hub in Bangladesh, starting in August 2023. His research interests focus on Machine Learning, Image Processing, and Computer Vision, often working with Python and libraries such as Keras, scikit-learn, TensorFlow, and OpenCV. He is currently engaged in Gait Recognition research. Md. Mehedi Hasan holds a Bachelor of Science degree in Computer Science and Engineering from Varendra University, where he studied from January 2019 to February 2023. His expertise includes classification within Artificial Intelligence, emphasizing practical applications in his field.

**Mohammad Asif Ul Haq** holds a Bachelor of Science degree in Computer Science and Engineering from Varendra University. His expertise includes Biometrics, Artificial Intelligence, and Robotics.

**Md. Hasan Maruf** holds a Bachelor of Science degree in Computer Science and Engineering from Varendra University. His expertise includes Computer Vision, Machine Learning, and Cloud Computing.

**Nakib Aman Turzo** is currently a Lecturer at Pabna University of Science and Technology, having started in March 2024. Prior to this, he held multiple roles at the National Academy for Computer Training and Research (NACTAR), including Focal Point Officer for Innovation and Research Coordinator. Nakib also served as an Information Technology Trainer and Lecturer at Varendra University for nearly five years. His technical expertise spans Software development, Data analysis, and the Internet of Things. Nakib holds a Master's degree in Computer Science from Rajshahi University of Engineering & Technology and a Bachelor's degree in Computer Engineering from Pabna University of Science and Technology. He has contributed to numerous research publications, focusing on IoT, Machine Learning, and Data privacy.