Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 09 May 2026
Revised 24 June 2026
Accepted 04 July 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 1490–1506

KEYWORDS:
Biometrics, GAIT recognition, contact less, marker free GAIT, CNN, transfer learning

GAIT recognition using silhouette using CNN, Transfer Learning

Amogha B[1], Rohini Deshpande[2], Prameela Kumari N[2]

[1]Reva University, India
[2]Reva University, India

ABSTRACT:

Background: GAIT means walking pattern. This walking pattern can be used to authenticate person in public areas. This biometric recognition system can identify persons even in low resolution images captured from surveillance camera.

Methods: Paper outlines feasibility in identification of person using GAIT as a biometric security system. Also briefs importance of contactless biometric based security systems are formulated considering videos of person (marker-free GAIT). Paper addresses designing of custom Convolutional Neural Network (CNN) and compare performance transfer learning models. Person identification is performed using silhouette captured indoor & outdoor.

Results: Models employed are Deep Learning method CNN, three transfer learning (VGG16, Xception and MobileNetV2). Result analysis shows that CNN achieved 93% for indoor and 91% for outdoor. Proposed model is tested on CASIA-B standard benchmark's silhouette and achieved accuracy of 87%. Hence propose model is efficient in detecting persons for real time applications. It is implemented using python scripting language.

1. INTRODUCTION

Person identification is possible by inspecting physical and behavioural pattern by analysis walking pattern is titled as GAIT recognition [1]. GAIT can be identified from distance, when video is in low resolution, without target consent. But in other biometrics system such as face, finger, voice and iris need target consent. Also, person need to place their biometrics on equipment to get access. GAIT has various applications in field of forensic identification, security at social places and preventing crimes [2] [3].

To address these problem various researchers proposes their methodology to recognize human using GAIT. It can be broadly classified into event based, color based and skeleton or silhouette based. Event camera is employed in event based methods [4]. This type of cameras is going to capture event frames specific to every pixel. It gives high temporal with low energy consumption [5]. Generative models or GANs are employed to extract texture feature and colour information from color images in RGB model. Skelton based or silhouette based methods are another type of GAIT detection where features are extracted from image and significant results achieved when publicly available dataset for research [6-9]. Skelton based methods extracts body regions but shape features are lost. In silhouette only outline of body is preserved but few body parts are lost due to overlapping of abdomen and legs [10-11]. Fig.1 shows silhouette and skeleton view of human generated from image sequences.

Natural Resources for Human Health
eISSN: 2583-1194

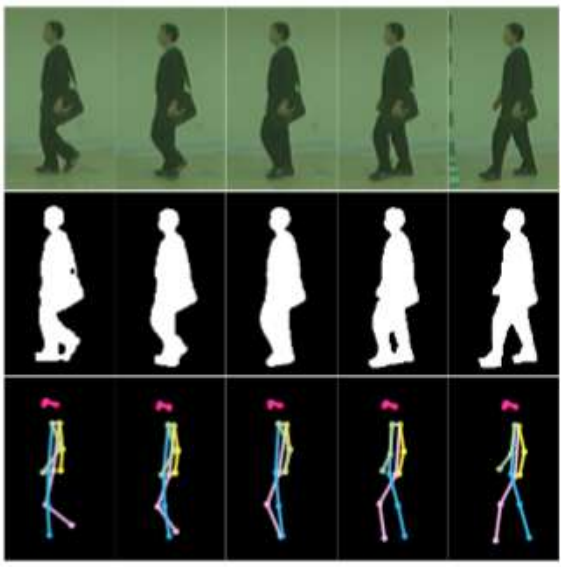


Figure 1 Generation of various forms of GAIT representation on public dataset [12]

Major contributions of this paper are

1. Silhouettes of Indian populations are employed for training and testing Machine Learning, Deep learning and Transfer Learning
2. Comparative analysis is performed between transfer learning, machine learning and deep learning models

1.1 Organization of paper

Section 2 provides research review of appropriate literature on GAIT recognition using machine, transfer and deep learning. Section 3 provides details on proposed methodology to recognition humans using silhouettes. Section 4 is result and discussion section, which gives screenshots of results obtained during implementation. Section 5 provides conclusion and future scope. Lastly reference is provided.

2. LITERATURE SURVEY

Identification of humans based on GAIT can also be classified by using two methods appearance and model based [13]. Individual persons can be identified based on the extraction of unique walking pattern. Methodology employed for model based approaches is through mathematical calculations. It makes use of sequence of GAIT and movement of body parts are used to calculate angles and store it. Other Methodology includes examining kinematics of every person's joint in order to extract characteristics of GAIT. It includes not only hip, knee, ankle movement and also movements of hands. Many methods are employed to extract outline, physical and anatomical patterns extracting model from GAIT frames. To extract these features from the GAIT sequence requires high resources and computation time is needed. These models also need high resolution images [13]. Whereas, appearance based methods extract features directly from GAIT images without need to separate models. It requires lesser computation time and less complex models [13].

Various methods developed using appearance based mainly concentrating on capturing silhouette pattern. In [14] extracted width of the silhouette and given it as a feature to Hidde Markov model for predication. Leg area of silhouette is extracted to recognize humans [15]. In [16] author extracted shape features and also extracted GAIT patterns from edge images. As an alternate methods authors converted a every GAIT cycle into average silhouette [17]. Averaging frames of silhouette is called as GAIT Energy Image (GEI) [18]. Similar works on GEI are GAIT flow image [19], masked GEI [20] and GAIT entropy [21].

Further other feature extraction is performed by using temporal information. Depth images are employed instead of only 2D images. Microsoft Kinect is the input devices employed to capture depth images. Instead of GEI researchers worked on accessing GAIT Energy Volume (GEV) from averaging silhouette frame [1]. Few novel works proposed by using Histogram of Oriented Gradients (HOG)[17].

In recent years deep learning methods had gained its importance in computer vision applications. Because of it generates high scale features from low level patterns [17]. Various fields are using deep learning models to obtain accurate output. A single layer back propagation based Artificial Neural Network (ANN) is employed to categorize GAIT [1]. Later GEI is fed to Convolutional Neural Networks (CNN) and named as GEI to identify specific humans. Deep CNN were employed to increase accuracy compared to CNN in human identification[15].

Later on, research on transfer learning based GAIT detection by using GEI images were performed to authenticate humans [23]. Accuracy of detection is more compared to CNN and Deep CNN.

3. PROPOSED METHODOLOGY

Block diagram of proposed model is as shown in Fig. 2. It also provides information on models employed such as deep learning model CNN and three transfer learning model such as VGG-16, xception and mobilenetV2. Other details are provided in this sub-section.

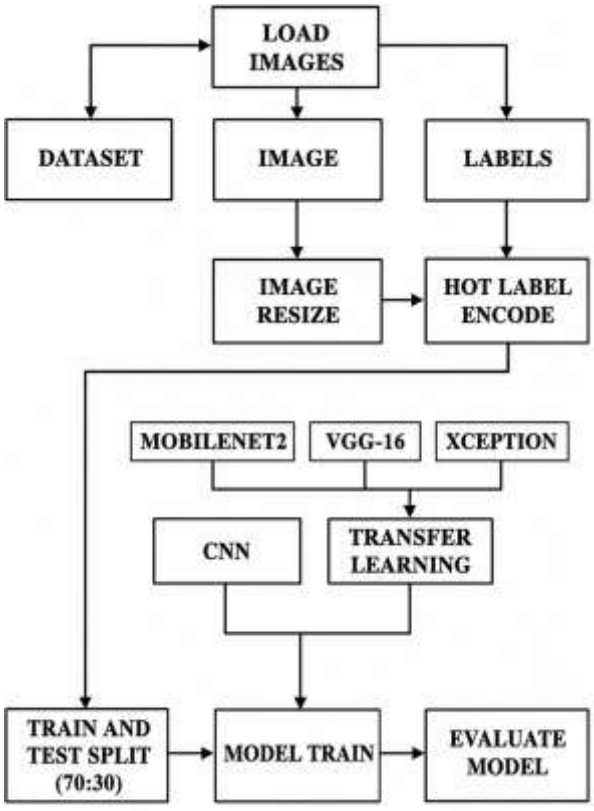


Figure 2 Proposed model

3.1 Collection of Dataset

An own prepared dataset is used for implementation of proposed model. During preparation of dataset 240 subjects were considered. 120 subjects are allowed to walk indoor and outdoor equally. Every subject's pattern is captured by considered four view angles (45, 90, 135 & 180 degrees). Every person walked in three cases such as Normal walking, walking with bag and wearing coat. Individual folders were made and all subjects data is stored. Camera setup to capture walking pattern of subject is as showed in Fig. 3. It has four cameras placed at four different angles. Person walks from position 'A' to 'B' covers angles (45, 90 , 135 and 180 degrees) of person walking. In the same fashion when person walks from position 'B' to 'A' capturing

degree will be 225, 270, 315 and 0 degree. Hence with cropping we can generate eight view angles from four devices. A dataset is prepared in two cases namely indoor and outdoor setup. There was no controlled environment, videos are captured readily in complex environment and different climatic conditions. Each person is allowed to walk without any instructions for three times. First walk is their normal walking, second will be walking with bag and last one will be holding or wearing coat. Each person's four view angles with three different walking styles are captured. Examples of indoor and outdoor scenario is given in Fig 4 (a) shows video captured for indoor environment, it an ITI college having several other complex objects along with subject. Fig 4 (b) shows video capturing of outdoor setup which has bikes, trees, tractors, overhead tanks. These complex objects will be removed in pre-process method. Along with video captured dataset has soft biometrics of every person and their statics is as shown in Fig.5.

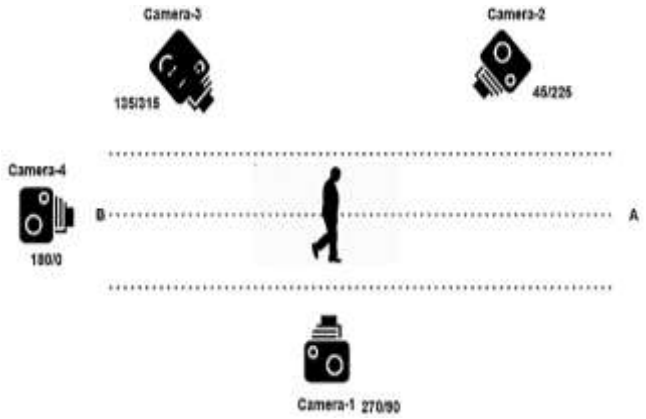
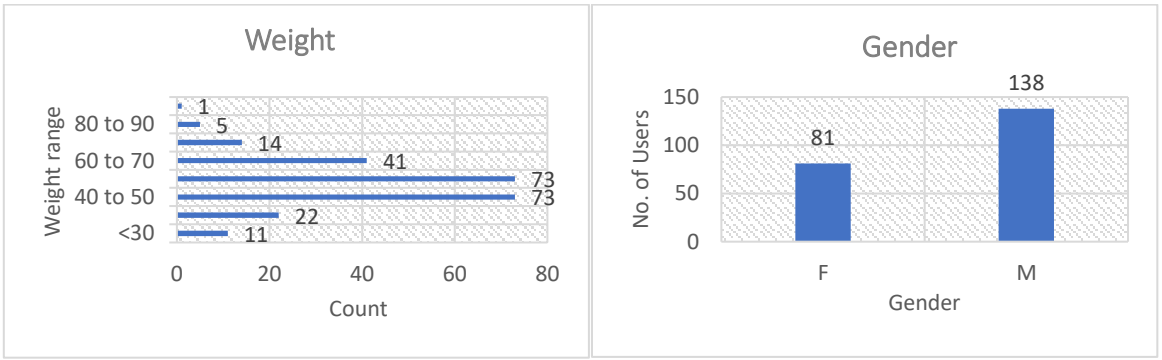


Figure 3 Arrangement for setting up camera

a) **b)**

Figure 4 Scenes of a)indoor region and b) outdoor region



Natural Resources for Human Health
eISSN: 2583-1194

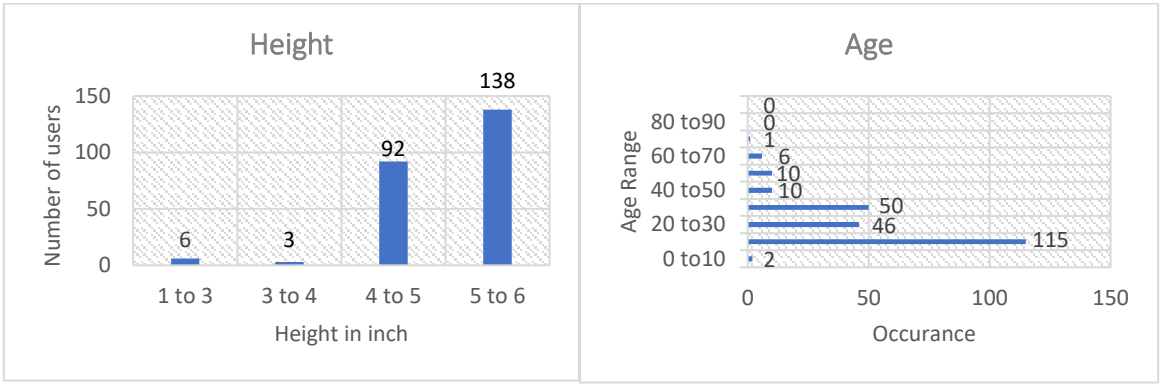


Figure 5 Soft biometrics of subject captured

Methodology employed for extracting silhouette is, it has three steps, foreground pixel extraction, selection of background and remove background. Foreground pixel extraction, this step plays a crucial role in extracting humans and leave rest behind. Logic behind this method is to subtract foreground and background. After subtraction threshold method is applied to convert from black and white image. Threshold is selected automatically for every image by using cv2.ostu method. Threshold is dynamic and chosen as per image. Next step is to select background, first initial few frames of video is considered as background frames using this background next frames are subtracted from background which gives silhouette images. Logic is as showed in Fig. 6.

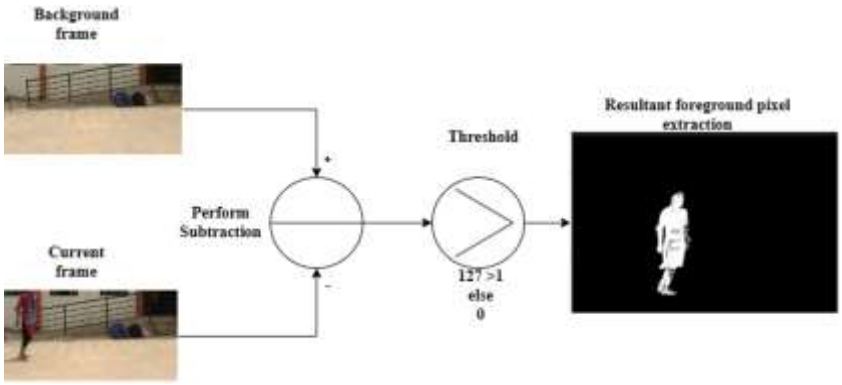


Figure 6 Extraction of foreground pixels

Silhouette of all persons are provided in dataset and extraction process is as showed in Fig.6. A two hundred fifty images from every person are selected for training and testing model. Fig.7. shows capturing of subject from different view angle and three different walking style. For training and testing a total of 29,870 images for indoor and for outdoor scenario 30000 images are employed. For indoor and outdoor scenario 240 persons are considered. Number of images for indoor and outdoor selected is as shown in Fig.8 and Fig. 9 respectively.

Natural Resources for Human Health
eISSN: 2583-1194

Figure 7 Different view angles and walking styles of subject during capturing of video.

```
Loaded 29870 images (max 250 per person) across 120 classes: ['Person121', 'Person122', 'Perso
n123', 'Person124', 'Person125', 'Person126', 'Person127', 'Person128', 'Person129', 'Person130',
'Person131', 'Person132', 'Person133', 'Person134', 'Person135', 'Person136', 'Person137', 'Perso
n138', 'Person139', 'Person140', 'Person141', 'Person142', 'Person143', 'Person144', 'Person145',
'Person146', 'Person147', 'Person148', 'Person149', 'Person150', 'Person151', 'Person152', 'Perso
n153', 'Person154', 'Person155', 'Person156', 'Person157', 'Person158', 'Person159', 'Person160',
'Person161', 'Person162', 'Person163', 'Person164', 'Person165', 'Person166', 'Person167', 'Perso
n168', 'Person169', 'Person170', 'Person171', 'Person172', 'Person173', 'Person174', 'Person175',
'Person176', 'Person177', 'Person178', 'Person179', 'Person180', 'Person181', 'Person182', 'Perso
n183', 'Person184', 'Person185', 'Person186', 'Person187', 'Person188', 'Person189', 'Person190',
'Person191', 'Person192', 'Person193', 'Person194', 'Person195', 'Person196', 'Person197', 'Perso
n198', 'Person199', 'Person200', 'Person201', 'Person202', 'Person203', 'Person204', 'Person205',
'Person206', 'Person207', 'Person208', 'Person209', 'Person210', 'Person211', 'Person212', 'Perso
n213', 'Person214', 'Person215', 'Person216', 'Person217', 'Person218', 'Person219', 'Person220',
'Person221', 'Person222', 'Person223', 'Person224', 'Person225', 'Person226', 'Person227', 'Perso
n228', 'Person229', 'Person230', 'Person231', 'Person232', 'Person233', 'Person234', 'Person235',
'Person236', 'Person237', 'Person238', 'Person239', 'Person240']
```

Figure 8. Total number of indoor images used for training and testing proposed model

```
Loaded 30000 images (max 250 per person) across 120 classes: ['Person1', 'Person10', 'Person10
0', 'Person101', 'Person102', 'Person103', 'Person104', 'Person105', 'Person106', 'Person107', 'Pe
rson108', 'Person109', 'Person11', 'Person110', 'Person111', 'Person112', 'Person113', 'Person11
4', 'Person115', 'Person116', 'Person117', 'Person118', 'Person119', 'Person12', 'Person120', 'Per
son13', 'Person14', 'Person15', 'Person16', 'Person17', 'Person18', 'Person19', 'Person2', 'Person
20', 'Person21', 'Person22', 'Person23', 'Person24', 'Person25', 'Person26', 'Person27', 'Person2
8', 'Person29', 'Person3', 'Person30', 'Person31', 'Person32', 'Person33', 'Person34', 'Person35',
'Person36', 'Person37', 'Person38', 'Person39', 'Person4', 'Person40', 'Person41', 'Person42', 'Pe
rson43', 'Person44', 'Person45', 'Person46', 'Person47', 'Person48', 'Person49', 'Person5', 'Perso
n50', 'Person51', 'Person52', 'Person53', 'Person54', 'Person55', 'Person56', 'Person57', 'Person5
8', 'Person59', 'Person6', 'Person60', 'Person61', 'Person62', 'Person63', 'Person64', 'Person65',
'Person66', 'Person67', 'Person68', 'Person69', 'Person7', 'Person70', 'Person71', 'Person72', 'Pe
rson73', 'Person74', 'Person75', 'Person76', 'Person77', 'Person78', 'Person79', 'Person8', 'Perso
n80', 'Person81', 'Person82', 'Person83', 'Person84', 'Person85', 'Person86', 'Person87', 'Person8
8', 'Person89', 'Person9', 'Person90', 'Person91', 'Person92', 'Person93', 'Person94', 'Person95',
'Person96', 'Person97', 'Person98', 'Person99']
```

Figure 9. Total number of outdoor images used for training and testing proposed model

To Test the proposed CNN model, had employed a CASIA-B dataset. CASIA-B has silhouette of 124 persons. These subjects are walked in controlled conditions all walking is confined to indoor. Persons are allowed to for ten sequences by holding bag, wearing coat and normal walking. Dataset is publicly available and downloaded from official website [25]. For 124 persons 51,624 silhouette images are generated and stored by CASIA-B.

Due to comparison of proposed work with existing dataset instead of all 51,624 images only 12,400 images. Below Fig.10 shows 12400 images are extracted from all folders present in dataset.

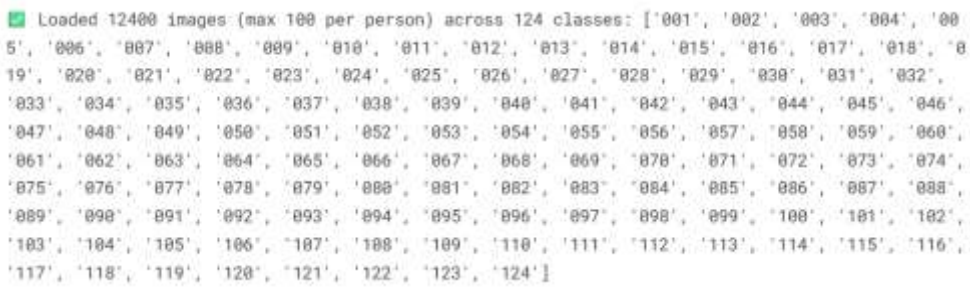


Figure 10. Total images of CASIA-B employed to test performance of proposed model

3.2 Statistics of dataset

Dataset has two scenario subjects allowed to walk indoor and outdoor. For indoor total number classes are 120. Total images for indoor is 29870. For outdoor number of images per classes is 250. Total number of outdoor images are 30000 as showed in Tab.1. To test robustness of model, applied same model to CASIA-B dataset and number of images of CASIA-B considered is 12400.

Table 1 Training and testing images

	Total Number of images	Training	Testing	CASIA – B
Indoor	29870	20909	8961	12400
Outdoor	30000	21000	9000	
Total number of images	59870	41909	17961	

3.3 Load Image

It is the sub-function which loads images and its corresponding labels. Name of labels will be same as name of folders where images are stored. Algorithm steps for loading images are provided in Tab.2. Loaded labels are converted to categorical index by using keras library. It is called hot encoding labels.

Table 2 Algorithm for load image

Step 1: Define data and labels.
Step 2:
Define for loop
List folder names and store in variable class
Define inner loop
Load image by using path
Convert image to array or matric
Normalise image
Data and labels are appended to class
End of inner loop
Return data and images
End of for loop

3.4 Model design and training

The model takes binary images as input with fixed height and width defined as img_height and img_width, correspondingly. It forms the network in a hierarchical fashion where the initial layers extract features through the convolutional and pooling layers while the last layers are connected and clasify the inputs.

The first layer in this architecture is a 2D convolutional layer (Conv2D) which has 32 filters of size 3×3. It scans the input image to derive low-level features such as edges and corners before passing it through the ReLU activation function in order to implement its non-linear capabilities. The input shape of the grayscale images is stated as (img_height, img_width, 1). The next layer is a MaxPooling2D layer which performs down sampling on the feature maps and brings down the spatial dimensions to 2x2 from 3x3 as tabulated in Tab 3. It is useful in: reducing the number of dimensions, enhancing computational efficiency, and preventing data overfitting.

A similar pattern is adopted by the next set of layers, but with more complexity. The second convolutional layer uses 64 filters of 3x3 to extract high level features from the input with max-pooling applied to down-sample the feature maps. These convolutional and pooling layers constitute the feature extraction of the CNN.

Following the feature extraction, the Flatten layer reshapes the 2D feature maps to a 1D vector, making it suitable for the subsequent dense layers. This ensures that the first dense layer has 128 neurons with ReLU activation, where it learns high-level representations of the input. To mitigate overfitting, the Dropout layer is applied, where half of the neurons are turned off randomly during training. This process eliminates overfitting of the data so that the model performs better on unseen data.

Finally, the output layer is a densely connected layer with the number of neurons equal to the number of classes in the dataset (len(class_names)). It applies softmax activation function on the final layer to give probabilities for each class which are normalized to be between 0 and 1 and sum to 1. This assists the network to predict the class of the input image with high confidence.

Table 3 Sequential model layers

Model layers	Descriptions
Convolutional layer 2d	This layer is going to reduce dimensions can extract specific features.
MaxPooling 2D	Pooling is a layer where it is going to down sample features. Maxpooling is going to obtain maximum intensity of sliding window element
2nd Convolutional layer 2d	This layer is going to reduce dimensions can extract specific features.
Maxpooling 2D	Pooling layer
Flatten	This layer is going to convert feature maps to 1D vector
Relu activation function	Non-linear activation function which allows only positive maximum values
Dropout	It is going to drop 50% of neuron information which causes overfitting
Softmax	Last activation function which output same as number of classes. In this case 120 for indoor and 120 persons for outdoor

Training parameters considered for models are provided in Tab 4. It contains batch size, optimizer, epochs and learning rate. Batch size says that how many images are processed before updating weights. Learning rate is the step size used to update weights during optimization. Epochs is number of times model need to be passed through dataset.

Table 4 Training parameters

Training parameter	Details
Batch size	32
Optimizer	adam
Epochs	10
Learning rate	0.001
Criterion	Categorical

Xception model

Xception, also known as Extreme Inception, is a convolutional neural network architecture that improves over the Inception model by incorporating depthwise separable convolutions as shown in Fig.11. This operation breaks down standard convolution into two steps: depthwise convolution and pointwise convolution. This reduces the number of parameters, making the network more efficient. Xception is organized into three main flows: entry flow, middle flow, and exit flow. It employs residual connections in entry and exit paths to solve the vanishing gradient issue and convert dense layers into global average pooling layers, reducing the number of parameters and risk of overfitting. Compared to the Inception model, Xception is simpler and more efficient.

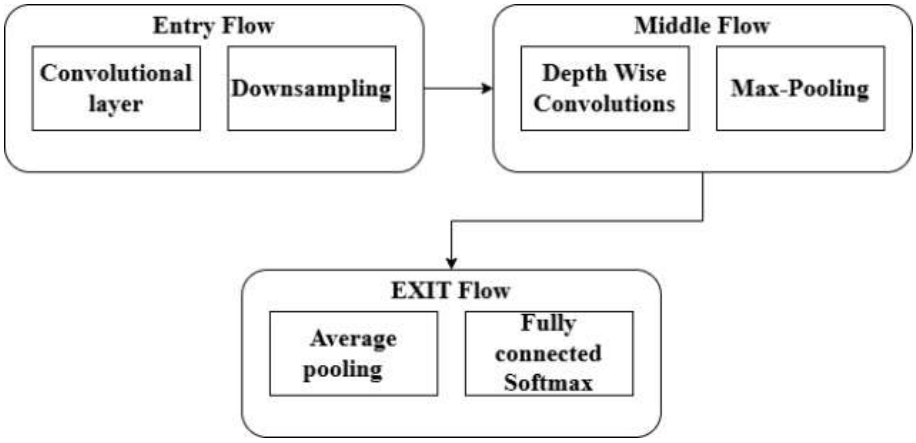


Figure 11. Xception model

VGG-16

VGG16 is a deep convolutional neural network developed by the Visual Geometry Group at the University of Oxford for large-scale image classification. It has 16 layers with learnable parameters and uses 3x3 filters for detail without consuming much computational power as shown in Fig.12. The network gradually moves from simple to complex structures, with the last three layers being fully connected dense layers. Despite its computational intensity, VGG16 remains popular in transfer learning and fine-tuning applications due to its systematic design and pre-trained weights on ImageNet.

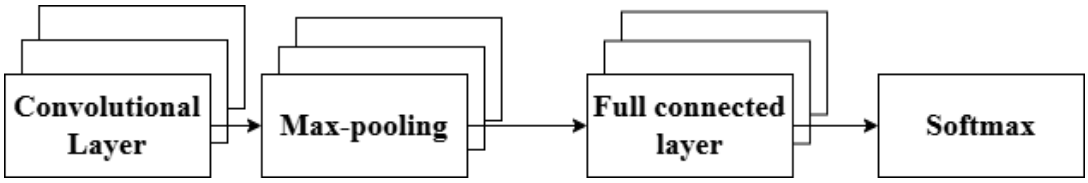


Figure 12 VGG-16 Model

MobileNetV2

MobileNetV2 is a lightweight deep neural network architecture for mobile and embedded vision applications, offering high accuracy and low computational cost. Its main feature is the inverted residual block, which preserves data and removes extra details. The architecture uses depthwise separable convolutions, reducing the number of parameters and calculations. Multiple inverted residual blocks increase feature complexity, and classification is performed using a global average pooling layer, a fully connected layer, and a softmax layer as shown in Fig.13.

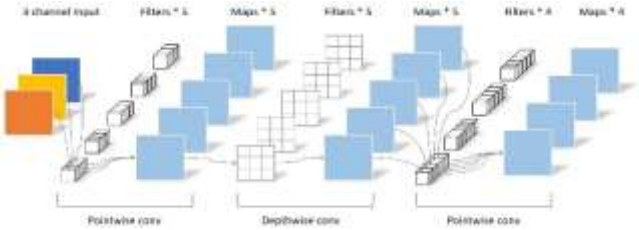


Figure 13 MobileNetV2 architecture [24]

4. RESULT AND DISCUSSION

This section provides detailed performance analysis of proposed model to identify person for indoor and outdoor cases. Number of images employed for training, testing model is as shown in Fig.14. It also shows performance of proposed model implemented on existing benchmark dataset. Employed CASIA-B dataset for testing robustness of trained model on standard dataset. To test model instead of using full images had employed 12400 images for testing model. Details of employed dataset is, prepared dataset has 240 persons walked in both indoor and outdoor equally. Model takes binary image as input for training and testing. To identify person using silhouette models employed are customized CNN, three transfer learning models such as VGG-16, xception MobileNetV2 as shown Tab. 5 and Tab. 6. Performance of proposed model for every epoch is as showed in Fig.15. For training and testing model, split ratio is in 70:30., i.e., 70% of dataset values are reserved for training and 30% are reserved for testing.

Table 5 shows performance of model during training and validation is provided. Epochs are 10 with the batch size 32. Custom CNN gives better results with eight hidden layers. Had tested by increasing hidden layers to twelve due to overfitting model accuracy became less. Other transfer learning has many hidden layers hence accuracy is low. For specific dataset custom CNN had provided a greater performance due to its lesser number of hidden layers. Custom CNN achieved a testing accuracy of 93% for indoor and 91% for outdoor dataset.

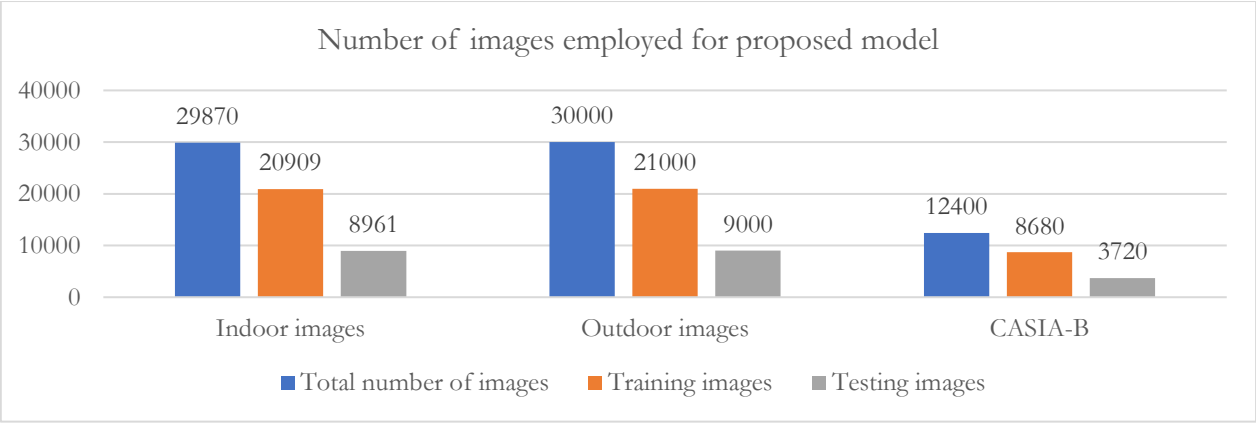


Figure 14 Images employed for training and testing

Performance metrics for outdoor and indoor scenario is provided in Tab.6 and Tab.5 respectively, which gives accuracy, f1 score, precision, recall and f1 score. Compared to other models' performance of CNN model is superior and found efficient.

Table 5 Performance metrics considering outdoor scenario

Name of model	Accuracy (in percentage)	Precision (in percentage)	Recall (in percentage)	F1 score (in percentage)
CNN	91	92	91	91
VGG-16	10	12	13	9
Xception	23	23	24	23
MobileNetV2	20	18	19	16

Table 6 Performance metrics considering indoor scenario

Name of model	Accuracy (in percentage)	Precision (in percentage)	Recall (in percentage)	F1 score (in percentage)
CNN	93	92	93	92
VGG-16	20	16	19	13
Xception	28	21	23	19
MobileNetV2	28	25	25	21

Performance measure of all models for indoor, outdoor and CASIA-B is as shown in Fig.15. Proposed CNN model on CASIA-B dataset had provided an 87% which is realistic accuracy to apply for real world scenario.

Table 7 Performance measure for models with CASIA-B

Name of model	Indoor Accuracy	Outdoor Accuracy	CASIA-B Accuracy
Custom CNN	93	91	87
VGG-16	20.31	13.27	10
Xception	28.39	24	23
MobileNetV2	28	18	20

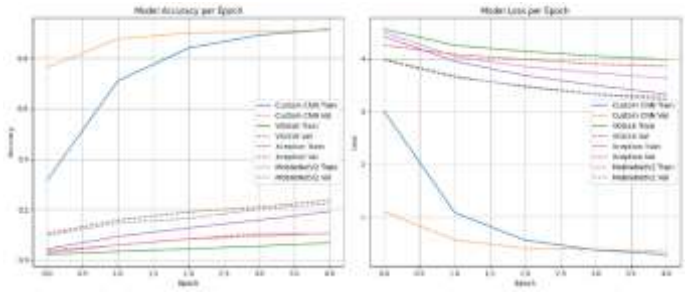


Figure 15 Accuracy comparison of model for indoor, outdoor and CASIA-B dataset

Confusion matrix is another metrics which is used to predict robustness of model in identifying person by using walking style. Confusion matrix for CNN, VGG-16, xception and MobileNetV2 is showed in Fig. 16 to Fig. 19 respectively for indoor dataset.

For outdoor dataset confusion matrix for CNN, VGG-16, xception and MobileNetV2 is showed in Fig. 20 to Fig.23 respectively.

The confusion matrices of the Custom CNN, VGG16, Xception, and MobileNetV2 models for each individual using the indoor dataset were observed to have a strong predicted values along the principal diagonal. The principal diagonal elements represent the number of gait samples of individuals that were correctly classified, while the off-diagonal elements represent the number of gait samples that were misclassified as belonging to other individuals. The Custom CNN model has a particularly high concentration along the principal diagonal, which essentially means a high number of correct predictions and low inter-class confusion. The VGG16, Xception, and MobileNetV2 models also have a high concentration along the principal diagonal, as shown. Thus, the extracted silhouette features contain sufficient discriminative information to identify individuals in the indoor dataset. Having the off-diagonal responses fairly low tells us that only very few gait samples were confused with other identity classes. On comparing the indoor confusion matrices with other types of confusion matrices, we observed their high true-positive classification and low false classification, thus indicating that the deep learning models are very effective under controlled indoor conditions.

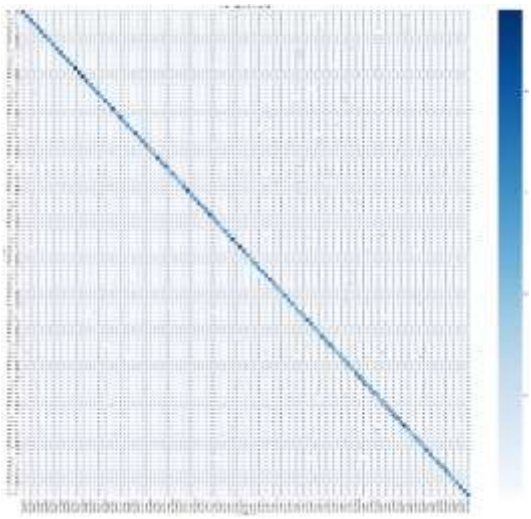


Figure 16 Confusion matrix for indoor dataset employing Custom CNN

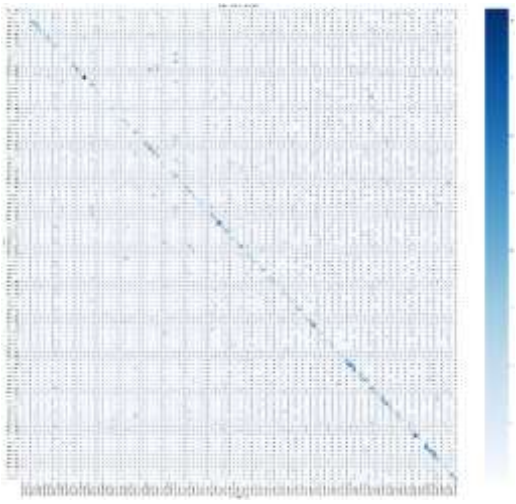


Figure 17 Confusion matrix for indoor dataset employing VGG-16

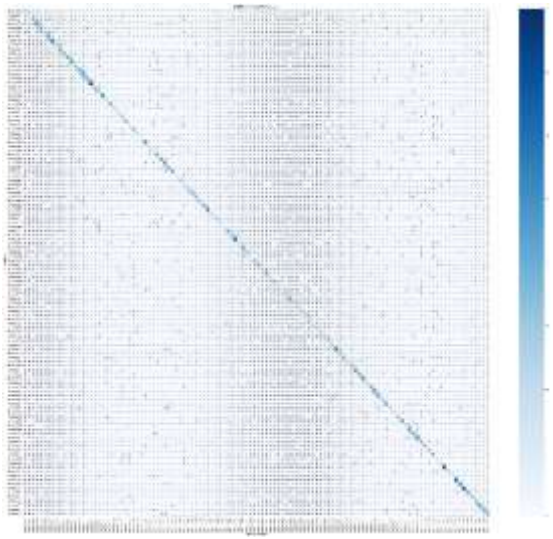


Figure 18 Confusion matrix for indoor dataset employing Xception

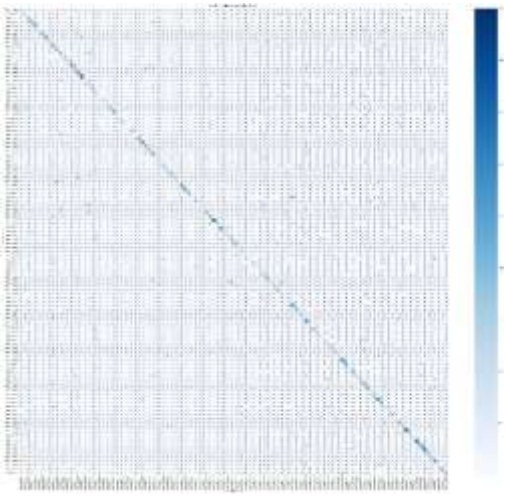


Figure 19 Confusion matrix for indoor dataset employing MobileNetV2

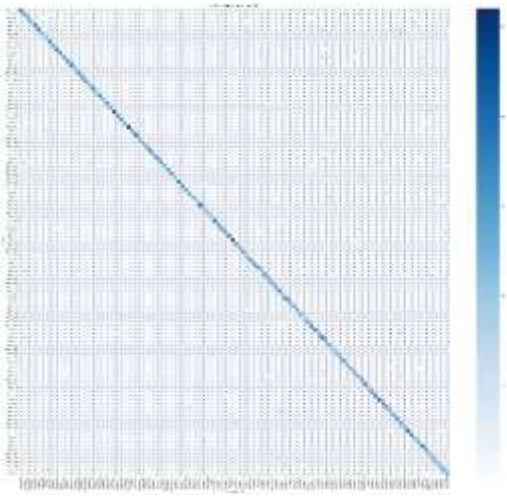


Figure 20 Confusion matrix for outdoor dataset employing Custom CNN

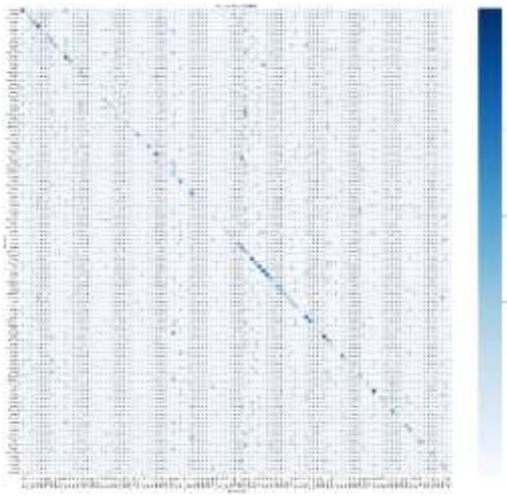


Figure 21 Confusion matrix for outdoor dataset employing VGG16

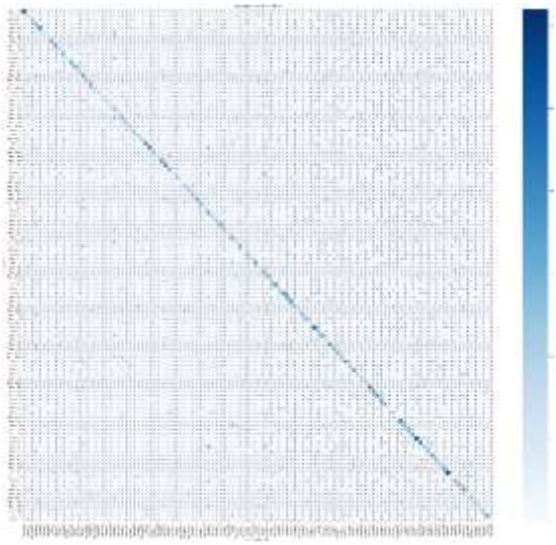


Figure 22 Confusion matrix for outdoor dataset employing Xception

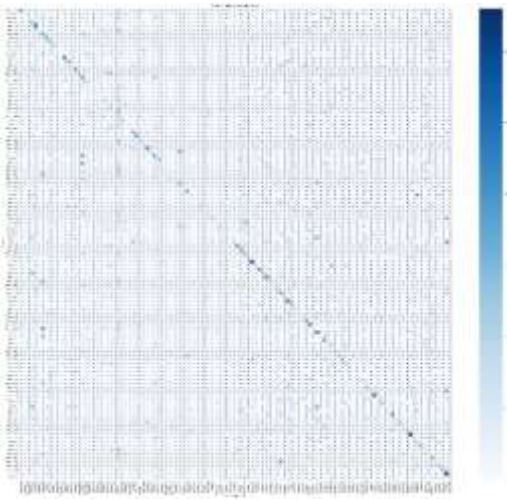


Figure 23 Confusion matrix for outdoor dataset employing MobileNetV2

The proposed silhouette-based deep learning approach produced highly diagonal confusion matrices for both indoor and outdoor datasets (Fig. 16 to Fig. 23). The high number of values on the main diagonal indicated a high individual identification rate in the test, while the small number of values away from the main diagonal indicated low confusion between different persons. This trend, observed under both indoor and outdoor conditions, demonstrated that the proposed model learned the silhouette representation features robustly across different situations. The Custom CNN model produced the most prominent diagonal patterns among all the other models.

4.1 DISCUSSION

This section discusses about the two aspects, firstly applying proposed model on standard benchmark CASIA dataset to find efficiency of model. Secondly comparing existing state of art methods with proposed model. Tab 7 shows accuracy achieved when model is fed with proposed dataset and CASIA dataset. It shows proposed model provided 91% and 93% for outdoor and indoor respectively and achieved 87% when CASIA-B dataset is used, which shows proposed model is able to identify persons with same hidden layers. On benchmark dataset proposed model is identifying persons with good accuracy. This shows that proposed model can work robust in real time environment.. Secondly on comparing existing accuracy with state of arts methods. Author [27] provided that their model achieved 87.2% and 70.4% when CASIA_B dataset is used when person carrying bag and wearing coat [27]. Another scenario author achieved 90% accuracy for CASIA A and C is employed with wearing bag case, person wearing coats achieved 58% accuracy. But they achieved good accuracy of 94% and 96% when person is walking in fast and slow pace [26].

5. CONCLUSION

This paper proposes a custom CNN model to identify person by using walking style. Own dataset is used for current research. Dataset contains 240 people's video who had walked indoor and outdoor environment. Every person's three different walking styles are captured from four input devices placed at four different angles. Walking styles include normal walk, walking by holding bag and walking by wearing or holding coat. Dataset samples undergoes a pre-processing to remove complex objects and generate silhouette by using background subtraction method for every person. Model is trained by silhouette of person. CNN model has eight hidden layers excluding input and output layer. It takes binary image as input. Model is tested for both indoor, outdoor and performance of proposed CNN model is compared with CASIA-B (12,400 binary silhouette images) and obtained 87% accuracy which shows that model is robust in detection when standard benchmark database is employed. Later had compared CNN's accuracy with pre-trained transfer learning models (Vgg-16, MobileNetV2, Xception), found custom CNN performance is better. Accuracy of CNN for Outdoor and indoor is 91 and 93 percentage respectively. Future scope can be usage of infrared or Motion sensor camera to capture outline of body. It helps in capturing outlines of body and makes system efficient in identifying person.

AUTHOR CONTRIBUTIONS

Amogha B: Data curation, analysis, first paper draft, proposing methodology

Rohini Deshpande: Supervision. formal analysis, review

Prameela Kumari N: Supervision, review of draft paper, methodology analysis

CONFLICT OF INTEREST

There is no conflict of interest from authors

ETHICS STATEMENT

Ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

Sample data is available in https://data.mendeley.com/preview/txfb6783j9?a=1f67313e-1796-4d92-8e0c-6eb04ffa1d71

REFERENCES

[1] Zhang, Ziyuan, Luan Tran, Xi Yin, Yousef Atoum, Xiaoming Liu, Jian Wan, Nanxin Wang "GAIT recognition via disentangled representation learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
[2] Larsen, Peter K., Erik B. Simonsen, and Niels Lynnerup. "GAIT analysis in forensic medicine." Journal of forensic sciences 53.5 (2008): 1149-1153.
[3] Bouchrika, Imed, et al. "On using GAIT in forensic biometrics." Journal of forensic sciences 56.4 (2011): 882-889.
[4] Wang, Y., Zhang, X., Shen, Y., Du, B., Zhao, G., Cui, L., & Wen, H. (2022). Event-Stream Representation for Human GAITs Identification Using Deep Neural Networks. IEEE transactions on pattern analysis and machine intelligence, 44(7), 3436–3449.
[5] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., ... & Scaramuzza, D. (2020). Event-based vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(1), 154-180.
[6] Chao, H., He, Y., Zhang, J., & Feng, J. (2019, July). GAITset: Regarding GAIT as a set for cross-view GAIT recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 8126-8133).
[7] Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., ... & He, Z. (2020). GAITpart: Temporal part-based model for GAIT recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14225-14233).
[8] Lin, Beibei, Shunli Zhang, and Xin Yu. "GAIT recognition via effective global-local feature representation and local temporal aggregation." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
[9] Hou, S., Cao, C., Liu, X., & Huang, Y. (2020, August). GAIT lateral network: Learning discriminative and compact representations for GAIT recognition. In European conference on computer vision (pp. 382-398). Cham: Springer International Publishing.
[10] Hou, S., Cao, C., Liu, X., & Huang, Y. (2020, August). GAIT lateral network: Learning discriminative and compact representations for GAIT recognition. In European conference on computer vision (pp. 382-398). Cham: Springer International Publishing.
[11] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5693-5703).
[12] Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S., & Rigoll, G. (2021, September). GAITgraph: Graph convolutional network for skeleton-based GAIT recognition. In 2021 IEEE international conference on image processing (ICIP) (pp. 2314-2318). IEEE.
[13] Singh, J. P., Jain, S., Arora, S., & Singh, U. P. (2021). A survey of behavioral biometric gait recognition: Current success and future perspectives. Archives of Computational Methods in Engineering, 28, 107-148.
[14] Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. IEEE transactions on pattern analysis and machine intelligence, 25(12), 1505-1518.
[15] Sokolova, A., & Konushin, A. (2019). Pose-based deep gait recognition. IET Biometrics, 8(2), 134-143.
[16] Jia, M., Yang, H., Huang, D., & Wang, Y. (2019, October). Attacking gait recognition systems via silhouette guided GANs. In Proceedings of the 27th ACM international conference on multimedia (pp. 638-646).
[17] Xu, C., Makihara, Y., Li, X., Yagi, Y., & Lu, J. (2020). Cross-view gait recognition using pairwise spatial transformer networks. IEEE Transactions on Circuits and Systems for Video Technology, 31(1), 260-274.
[18] Sokolova, A., & Konushin, A. (2017). Gait recognition based on convolutional neural networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 207-212.
[19] Zou, Q., Wang, Y., Wang, Q., Zhao, Y., & Li, Q. (2020). Deep learning-based gait recognition using smartphones in the wild. IEEE Transactions on Information Forensics and Security, 15, 3197-3212.
[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, N. A., Kaiser, L. & Polosukhin, I. "Attention Is All You Need". NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017; 1 (1): 6000 – 6010.
[21] Tran, L., Hoang, T., Nguyen, T., Kim, H., & Choi, D. (2021). Multi-model long short-term memory network for gait recognition using window-based data segment. IEEE Access, 9, 23826-23839.
[22] [Luo, J., Zhang, J., Zi, C., Niu, Y., Tian, H., & Xiu, C. (2015). Gait recognition using GEI and AFDEI. International Journal of Optics, 2015(1), 763908.
[23] Verlekar, T. T., Correia, P. L., & Soares, L. D. (2018, December). Using transfer learning for classification of gait pathologies. In 2018 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 2376-2381). IEEE.

[24] https://yinguobing.com/bottlenecks-block-in-mobilenetv2/
[25] https://service.tib.eu/ldmservice/dataset/casia-b-gait-dataset
[26] Babar Ali, Maryam Bukhari, Muazzam Maqsood, Jihoon Moon, Eenjun Hwang, Seungmin Rho, An end-to-end gait recognition system for covariate conditions using custom kernel CNN, Heliyon, Volume 10, Issue 12, 2024, e32934, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2024.e32934.
[27] Chao, Hanqing, et al. "Gaitset: Regarding gait as a set for cross-view gait recognition." Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019..