

Gender classification using GAIT

A Project Report

submitted by

**VIMALRAJ S, E S SARVESWARA RAO (CS23M1007,
CS23M1008)**

*in partial fulfilment of requirements
for the award of the degree of*

MASTER OF TECHNOLOGY



**Department of Computer Science and Engineering with
specialization in Data Science and AI
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING KANCHEEPURAM**

MAY 2024

DECLARATION OF ORIGINALITY

I, **Vimalraj S, E S Sarveswara Rao**, with Roll No: **CS23M1007, CS23M1008** hereby declare that the material presented in the Project Report titled **Gender classification using GAIT** represents original work carried out by me in the **Department of Computer Science and Engineering with specialization in Data Science and AI** at the Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram.

With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Vimalraj S, E S Sarveswara Rao

Place: Chennai

Date: 15.05.2024

CERTIFICATE

This is to certify that the report titled **Gender classification using GAIT**, submitted by **Vimalraj S, E S Sarveswara Rao (CS23M1007, CS23M1008)**, to the Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, for the award of the degree of **MASTER OF TECHNOLOGY** is a bona fide record of the work done by him/her under my supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Rahul Raman

Project Guide

Assistant Professor

Department of Computer Science and Engineering

IIITDM Kancheepuram, 600 127

Place: Chennai

Date: 15.05.2024

ABSTRACT

Gender classification is in high demand because of its versatility. It can be used as a surveillance system for security. It can also be used to classify customers in retail establishments. In this work, we propose a new gender classification technique from a gait silhouette using observation angle-based GEIs. The system consists of two main parts: the observation angle classification model and the gender classification model. The observation angle classification generates 10 observation angle-based GEIs. The gender classification model uses angle-based GEIs to predict gender. The proposed methods perform well with freestyle walks which contain a viewpoint issue.

KEYWORDS: Gender classification, Gait, Gender recognition, GEI, CNN, Freestyle Walk

TABLE OF CONTENTS

ABSTRACT	i
LIST OF TABLES	iv
LIST OF FIGURES	v
ABBREVIATIONS	vi
NOTATION	vii
1 INTRODUCTION	1
1.1 Gender Classification	1
1.1.1 Existing approaches	1
1.1.2 Using Gait	1
1.2 Gait Analysis	1
2 DATASET AND PREPROCESSING	3
2.1 Dataset	3
2.1.1 CASIA-B	3
2.2 Preprocessing	3
2.2.1 Gait Energy Image (GEI)	3
2.2.2 Multi channel GEI	4
3 TRAINING	5
3.1 Model	5
3.1.1 Convolution Layer	5
3.1.2 Relu Activation	6
3.1.3 Max Pooling Layer	7
3.1.4 Fully connected layers	7
3.1.5 Sigmoid Activation	7

3.1.6	Loss Function	8
3.2	Training	8
3.2.1	Training History	8
4	EVALUATION AND COMPARISON	12
4.1	Evaluation Metrics	12
4.1.1	Confusion Matrix	12
4.1.2	True Positives	12
4.1.3	True Negatives	13
4.1.4	False Positives	13
4.1.5	False Negatives	13
4.1.6	Accuracy	13
4.1.7	Precision	13
4.1.8	Recall	13
4.1.9	F1 Score	14
4.2	Comparison between models trained on different angles	14
5	CONCLUSION AND FUTURE SCOPE	16
5.1	Conclusion	16
5.2	Future Scope	16

LIST OF TABLES

4.1	Comparison of metrics while using different angles	14
4.2	Comparison of metrics while using different angles, using 3-channel GEI	15

LIST OF FIGURES

2.1	Gait Energy Image	4
2.2	3-channel GEI	4
3.1	Training history for 90 degrees	9
3.2	Training history for 72 degrees	9
3.3	Training history for 54 degrees	9
3.4	Training history for 36 degrees	9
3.5	Training history for 18 degrees	10
3.6	Training history for 90 degrees, 3-channel GEI	10
3.7	Training history for 72 degrees, 3-channel GEI	10
3.8	Training history for 54 degrees, 3-channel GEI	10
3.9	Training history for 36 degrees, 3-channel GEI	11
3.10	Training history for 18 degrees, 3-channel GEI	11

ABBREVIATIONS

GEI	Gait Energy Image
CNN	Convolutional Neural Network

NOTATION

λ wavelength

CHAPTER 1

INTRODUCTION

1.1 Gender Classification

1.1.1 Existing approaches

Gender is an important fundamental attribute of humans. Nowadays, most gender recognition methods are based on facial features. This does not perform well when the subject is far away from cameras. And when the resolution of the cameras is low.

1.1.2 Using Gait

So in order to overcome the above discussed issues, we are using GAIT analyses in order to find the gender of a human.

1.2 Gait Analysis

Gait analysis, the study of human walking patterns, has become a useful bio metric technique for identification and recognition tasks. One important gait representation called Gait Energy Images (GEIs) has gained popularity. GEIs are averaged silhouette images over a full gait cycle that capture a person's gait pattern in a compact image.

However, GEIs are typically captured from a lateral side view, which may not always be possible in real-world scenarios. This has led to research into generating GEIs from arbitrary viewing angles to make the approach more practical.

Gender classification from gait is an important preliminary task that can improve the performance of subsequent identification systems. This project proposes a method

to accurately classify an individual's gender based solely on their gait patterns captured from any viewing angle using GEIs.

The proposed approach combines the strengths of GEIs in representing gait with advanced machine learning techniques to achieve accurate gender classification from certain observation angles.

CHAPTER 2

DATASET AND PREPROCESSING

2.1 Dataset

Many gait datasets are available like CASIA, OU-ISIR. For our project we have used CASIA-B Gait silhouette images as input.

2.1.1 CASIA-B

Dataset B[1] is a large multi view gait database. It is Created in January 2005. There are 124 subjects (93 males, 31 females). Gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered. Human silhouettes extracted from video files are also available.[1]

2.2 Preprocessing

2.2.1 Gait Energy Image (GEI)

After categorizing silhouette images of a walk into 11 groups, the silhouette images in the same group (except the rejected group) are combined to form 10 GEIs. For each GEI, its value at position (x, y) is defined by;

$$GEI(x, y) = \sum_{t=1}^n I_t(x, y)$$

where N is the number of silhouette images in a group, $I_t(x, y)$ is the value at position (x, y) of the tth image in the group.

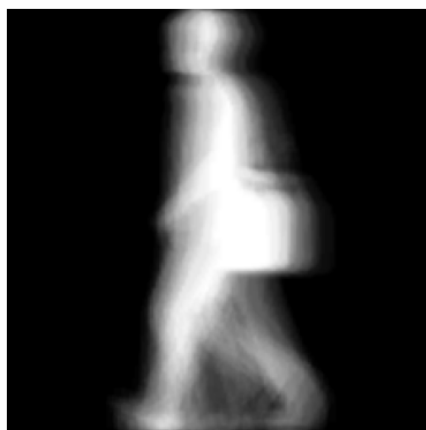


Figure 2.1: Gait Energy Image

2.2.2 Multi channel GEI

Multi-channel GEI involves extracting features from the multi-dimensional gait data. This often involves calculating the covariance matrices of the data from each sensor/channel and then finding the eigenvectors and eigenvalues of these matrices. These eigenvectors represent the principal components of variation in the gait data, while the eigenvalues indicate the significance of each component.[2]

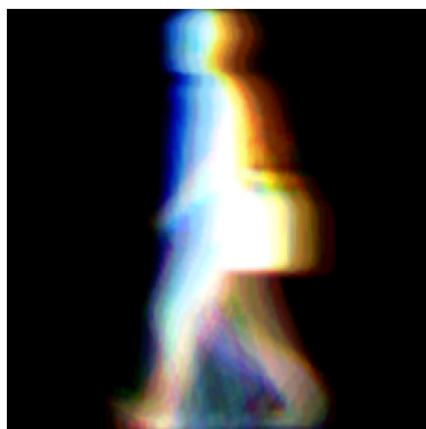


Figure 2.2: 3-channel GEI

CHAPTER 3

TRAINING

3.1 Model

A simple Convolutional Neural Network (CNN) with the following layers are used

- Input layer of size 224 x 224
- 4 blocks of Convolution, ReLU, MaxPooling
- 2 Dense layers using ReLU activation
- Output layer using Sigmoid activation

3.1.1 Convolution Layer

The convolutional layer is the core building block of a CNN, and it is where the majority of computation occurs. It requires a few components, which are input data, a filter, and a feature map. Lets assume that the input will be a color image, which is made up of a matrix of pixels in 3D. This means that the input will have three dimensions a height, width, and depth which correspond to RGB in an image. We also have a feature detector, also known as a kernel or a filter, which will move across the receptive fields of the image, checking if the feature is present. This process is known as a convolution.

The feature detector is a two-dimensional (2-D) array of weights, which represents part of the image. While they can vary in size, the filter size is typically a 3x3 matrix; this also determines the size of the receptive field. The filter is then applied to an area of the image, and a dot product is calculated between the input pixels and the filter. This dot product is then fed into an output array. Afterwards, the filter shifts by a stride, repeating the process until the kernel has swept across the entire image. The final output from the series of dot products from the input and the filter is known as a feature map, activation map, or a convolved feature.

Note that the weights in the feature detector remain fixed as it moves across the image, which is also known as parameter sharing. Some parameters, like the weight values, adjust during training through the process of back propagation and gradient descent. However, there are three hyper parameters which affect the volume size of the output that need to be set before the training of the neural network begins. These include:

1. The number of filters affects the depth of the output. For example, three distinct filters would yield three different feature maps, creating a depth of three.

2. Stride is the distance, or number of pixels, that the kernel moves over the input matrix. While stride values of two or greater is rare, a larger stride yields a smaller output.

3. Zero-padding is usually used when the filters do not fit the input image. This sets all elements that fall outside of the input matrix to zero, producing a larger or equally sized output. There are three types of padding: Valid padding: This is also known as no padding. In this case, the last convolution is dropped if dimensions do not align.

Same padding: This padding ensures that the output layer has the same size as the input layer.

Full padding: This type of padding increases the size of the output by adding zeros to the border of the input.

After each convolution operation, a CNN applies a Rectified Linear Unit (ReLU) transformation to the feature map, introducing nonlinearity to the model.

3.1.2 Relu Activation

The rectified linear unit (ReLU) or rectifier activation function introduces the property of non linearity to a deep learning model and solves the vanishing gradients issue. It interprets the positive part of its argument. It is one of the most popular activation functions in deep learning.

The ReLU Function is given as,

$$f(x) = \max(0, x)$$

3.1.3 Max Pooling Layer

Max pooling is a downsampling technique commonly used in convolutional neural networks (CNNs) to reduce the spatial dimensions of an input volume. It is a form of non-linear down-sampling that serves to make the representation smaller and more manageable, and to reduce the number of parameters and computation in the network. Max pooling operates independently on each depth slice of the input and resizes it spatially.

The primary objective of max pooling is to reduce the amount of information in an image while maintaining the essential features necessary for accurate image recognition. This process helps to make the detection of features in input data invariant to scale and orientation changes and also aids in preventing overfitting.

3.1.4 Fully connected layers

A fully connected layer refers to a neural network in which each neuron applies a linear transformation to the input vector through a weights matrix. As a result, all possible connections layer-to-layer are present, meaning every input of the input vector influences every output of the output vector.

3.1.5 Sigmoid Activation

Sigmoid function is as the activation for the output layer of binary classification models. It squashes the output to a probability value between 0 and 1, which can be interpreted as the probability of the input belonging to a particular class.

The Sigmoid Function is given as,

$$f(x) = \frac{1}{1 + e^{-x}}$$

3.1.6 Loss Function

Binary Cross Entropy is a loss function used in machine learning and deep learning to measure the difference between predicted binary outcomes and actual binary labels. It quantifies the dissimilarity between probability distributions, aiding model training by penalizing inaccurate predictions. Its widely used in tasks like binary classification, where the goal is to categorize data into two classes.

$$Loss = \frac{-1}{N} * \sum_{i=1}^N y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)$$

Binary Cross Entropy, also known as Binary Log Loss or Binary Cross-Entropy Loss, is a commonly used loss function in machine learning, particularly in binary classification problems. It is designed to measure the dissimilarity between the predicted probability distribution and the true binary labels of a dataset.

3.2 Training

The CNN model is trained separately for 5 different angles: 90 degrees, 72 degrees, 54 degrees, 36 degrees, and 18 degrees.

The models are trained using two types of input data:

- GEI (Gait Energy Image)
- 3-channel GEI

The only difference between the two inputs is the number of channels - GEI is a single-channel image, while 3-channel GEI has 3 color channels. So the model also differs only at the Input Layer.

3.2.1 Training History

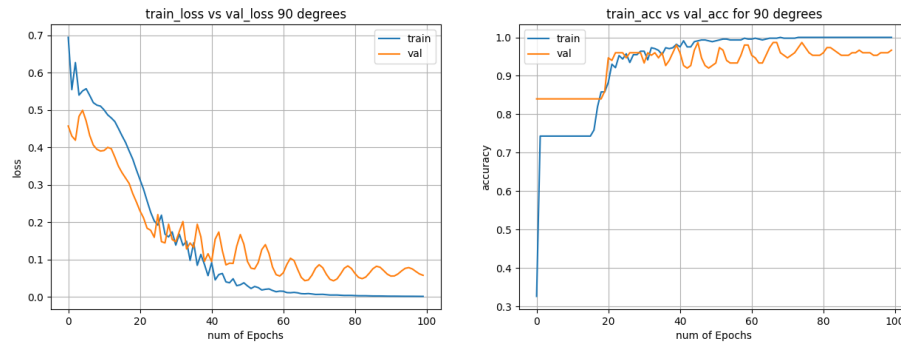


Figure 3.1: Training history for 90 degrees

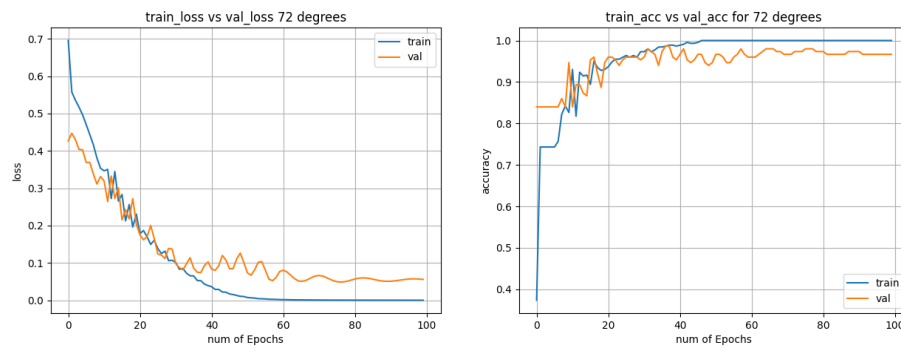


Figure 3.2: Training history for 72 degrees

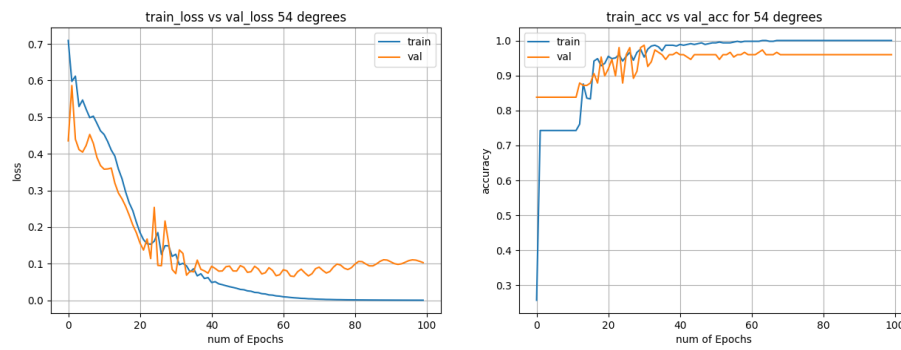


Figure 3.3: Training history for 54 degrees

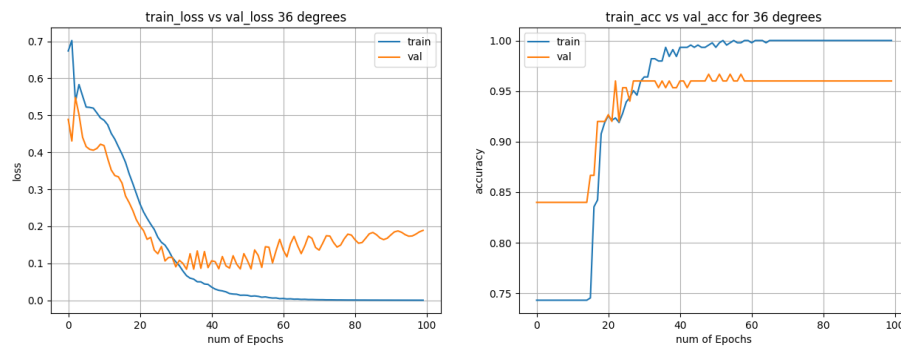


Figure 3.4: Training history for 36 degrees

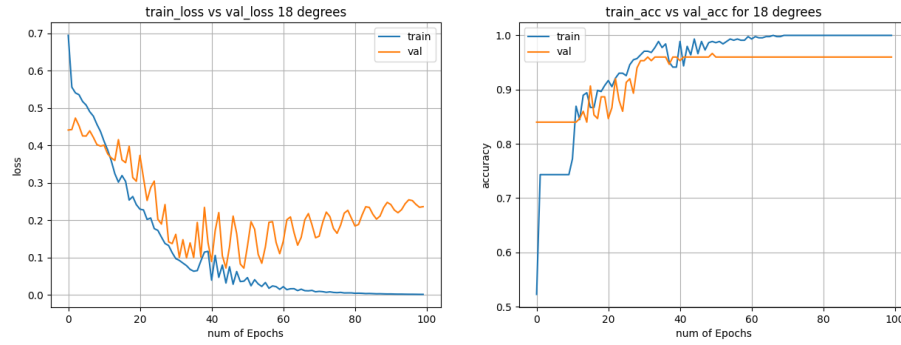


Figure 3.5: Training history for 18 degrees

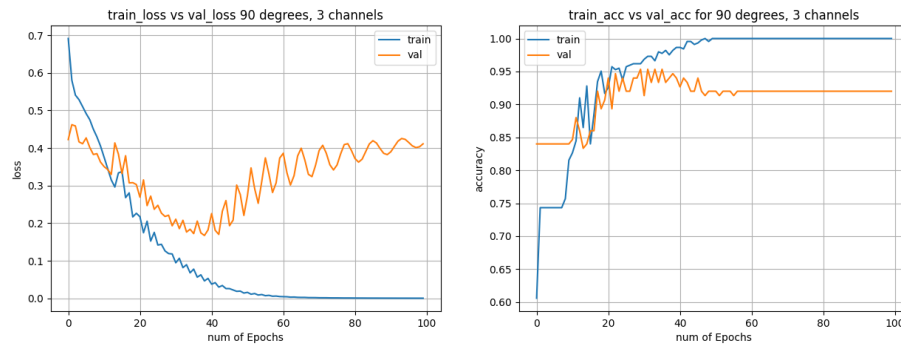


Figure 3.6: Training history for 90 degrees, 3-channel GEI

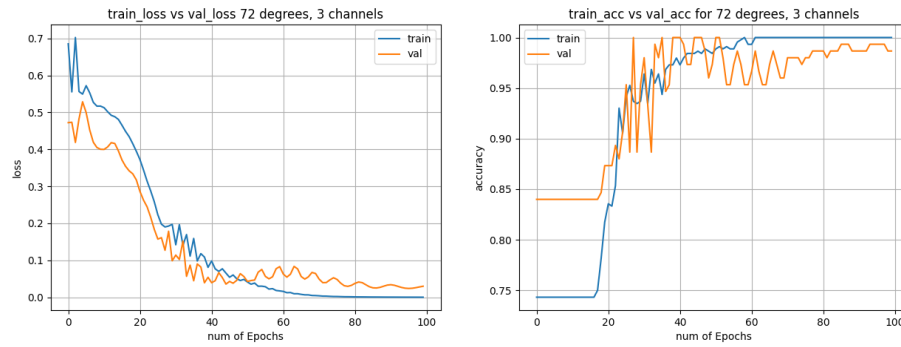


Figure 3.7: Training history for 72 degrees, 3-channel GEI

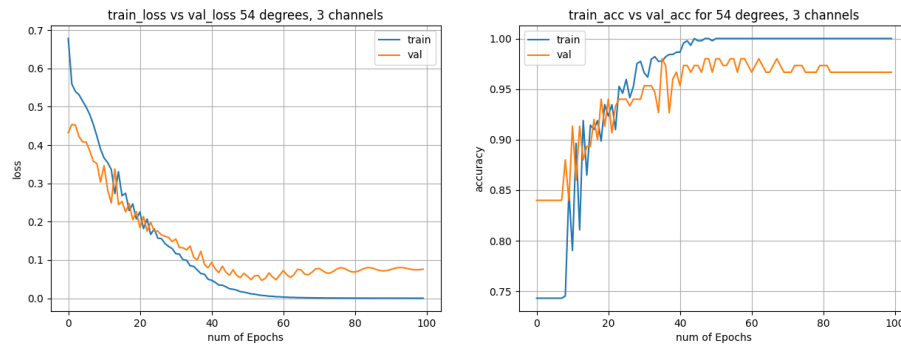


Figure 3.8: Training history for 54 degrees, 3-channel GEI

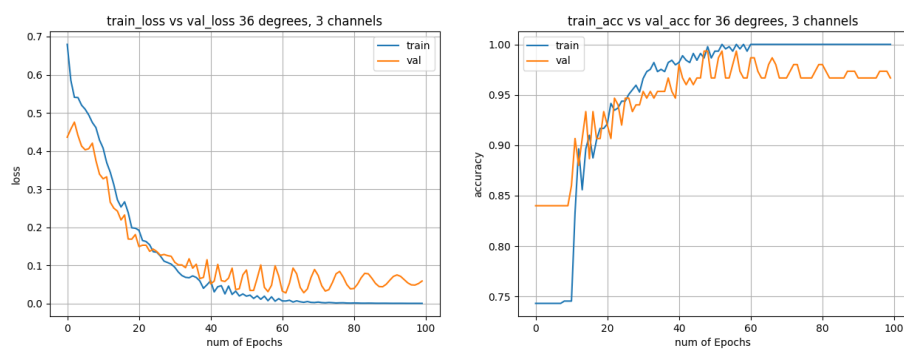


Figure 3.9: Training history for 36 degrees, 3-channel GEI

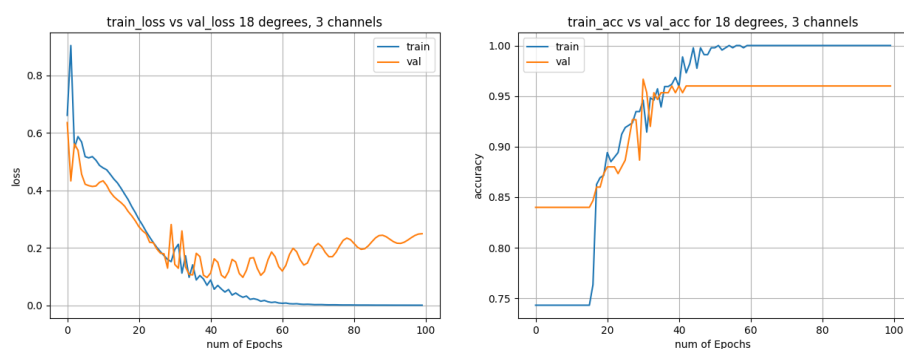


Figure 3.10: Training history for 18 degrees, 3-channel GEI

CHAPTER 4

EVALUATION AND COMPARISON

4.1 Evaluation Metrics

4.1.1 Confusion Matrix

For the gender classification from gait using observation angle-based GEIs project, the confusion matrix will be a valuable metric to evaluate the performance of the proposed method. A confusion matrix provides a comprehensive breakdown of the classification results, allowing for a detailed analysis of the model's accuracy and potential errors.

In the context of this binary classification task (male or female), the confusion matrix will be a 2x2 table that summarizes the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The rows of the matrix represent the actual classes (male or female), while the columns represent the predicted classes.

The confusion matrix will provide the following information:

- True Positives
- True Negatives
- False Positives
- False Negatives
- Accuracy
- Precision
- Recall
- F1 Score

4.1.2 True Positives

The number of instances where the model correctly predicted the gender as male for actual male samples.

4.1.3 True Negatives

The number of instances where the model correctly predicted the gender as female for actual female samples.

4.1.4 False Positives

The number of instances where the model incorrectly predicted the gender as male for actual female samples.

4.1.5 False Negatives

The number of instances where the model incorrectly predicted the gender as female for actual male samples.

From the confusion matrix, several evaluation metrics can be derived to assess the performance of the proposed method:

4.1.6 Accuracy

The overall proportion of correct predictions, calculated as $(TP + TN) / (TP + TN + FP + FN)$.

4.1.7 Precision

The proportion of true positives among all positive predictions, calculated as $TP / (TP + FP)$ for each class.

4.1.8 Recall

The proportion of actual positives that are correctly identified, calculated as $TP / (TP + FN)$ for each class.

angle	accuracy	precision	recall	f1_score
90	0.927	1.000	0.771	0.871
72	0.900	0.780	0.958	0.860
54	0.833	0.658	1.000	0.793
36	0.460	0.372	1.000	0.542
18	0.320	0.320	1.000	0.485

Table 4.1: Comparison of metrics while using different angles

4.1.9 F1 Score

The harmonic mean of precision and recall, providing a balanced measure of overall performance.

By analyzing the confusion matrix and these derived metrics, the researchers can gain insights into the strengths and weaknesses of the proposed method. They can identify specific cases where the model struggles, such as mis classifying males as females or vice versa, and potentially investigate the underlying reasons behind these errors.

Furthermore, the confusion matrix can be used to compare the performance of the proposed method with existing techniques or baseline models, allowing for a comprehensive evaluation and bench marking of the approach.

Overall, the confusion matrix will serve as a valuable tool for assessing the gender classification performance of the proposed observation angle-based GEI method, enabling the researchers to quantify and analyze the results, identify areas for improvement, and demonstrate the effectiveness of their approach.

4.2 Comparison between models trained on different angles

As shown in 4.1 and 4.2 the accuracy increases while using 3-channel GEI only at 90 degrees, stays the same at 72 degrees and decreases for all other angles.

angle	accuracy	precision	recall	f1_score
90	0.967	1.000	0.896	0.945
72	0.907	0.783	0.979	0.870
54	0.507	0.388	0.938	0.549
36	0.320	0.318	0.979	0.480
18	0.313	0.315	0.979	0.477

Table 4.2: Comparison of metrics while using different angles, using 3-channel GEI

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The model trained on 90° GEIs perform better than other models. Using 3-channel GEIs improve accuracy only at 90 degrees. 90° is the ideal angle to capture gait, but even for 72° gait sequence the model gives accuracy around 90

5.2 Future Scope

This approach can be extended to perform different types of classification. This simple model achieves good accuracy for the binary classification task. Other complex models may be required to extract features in a multi class classification problem.

REFERENCES

- [1] 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)*, vol. 4, 2006, pp. 441–444.
- [2] K. Kitchat, N. Khamsemanan, and C. Nattee, “Gender classification from gait silhouette using observation angle-based geis,” in *2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, 2019, pp. 485–490.