University of Essex

School of Mathematics, Statistics
and Actuarial Science

MA981 DISSERTATION

GENDER IDENTIFICATION BASED ON GAIT ANALYSIS

RISHWANTH MITHRA (2311566)

Supervisor: **Dr Saideh Ferdowsi**

September, 2024
Colchester

Abstract

Gait is a well-known bio-metric feature that has been broadly utilized for recognizing the individuals. Over the past few decades, gait biometrics has made an appearance as a promising replacement for traditional identification methods, contributing developments in monitoring, surveillance and analysis techniques. However, identifying a person based on an individual's walking style is a challenging task in computer vision applications. This paper gives a powerful and flexible approach to tackle the issue using gait analysis. The experiment is evaluated using the CASIA-B gait dataset, which consists of de-noised images of males and females. The methodology involves deep neural networks such as Convolutional Neural Network (CNN) and machine learning models to train and evaluate the data. For machine learning algorithms, feature extraction is performed using pre-trained models, followed by the classification of data using a Decision Tree, Random Forest and k-nearest Neighbor (KNN). Convolutional Neural Network (CNN) has attained the best results with a score of 98.8% compared to all other models. Therefore, gait analysis can be measured well using deep learning algorithms with better results and be used in applications such as biometric recognition and image classification. Gait Analysis learns a person's walking or running pattern, helps to recognize their unique movements, diagnoses problems that cause pain in the body, and corrects abnormalities by evaluating the treatments.

Acknowledgement

First and foremost, I would like to convey my sincere thanks to my supervisor, Dr. Saideh Ferdowsi, for all of her invaluable advice, support, and encouragement in assisting me in preparing this dissertation report and making the process easier for me. Her guidance and suggestions helped me learn more profoundly about the concepts in the right direction.

This acknowledgement is a sincere tribute to appreciating her for being in every phase of this dissertation journey. I want to thank the University of Essex for giving me the opportunity to pursue my master's and research on this dissertation. I would also love to thank my friends and my family for giving me complete love and support throughout my master's journey.

Contents

List of Figures

List of Tables

List of Abbreviations

CNN - Convolutional Neural Network

GEI - Gait Energy Images

SVM - Support Vector Machine

IA - Image Augmentation

PCA - Principal Component Analysis

HOG - Histogram of Oriented Gradients

LBP - Local Binary Pattern

CS-LBP - Centre Symmetric Local Binary Pattern

KNN - k-nearest neighbours

DTW - Dynamic Time Warping

VTM - View Transformation Model

SVD - Singular Value Decomposition

LDA - Linear Discriminant Analysis

AGI - Average Gait Images

DCV - Discriminative Common Vectors

ReLU - Rectified Linear Unit

SGD - Stochastic Gradient Unit

CHAPTER 1

Introduction

1.1 History of Gait Analysis

The study of how humans walk has been an area of interest for a long time. Aristotle (384-322 BCE) made some early observations on the topic, and more progress was made during the Renaissance by Giovanni Borelli (1608-1679) [2]. The Weber brothers, Wilhelm (1804-1891) and Eduard (1806-1871), also made significant contributions by conducting basic measurements while working in Leipzig [24].

Originally, a lot of research on how people walk depended on the use of film cameras. However, in the 1970s, more people started using video cameras to study a person's walking style. This made it simple and easy to do more detailed studies on individual patients with conditions like cerebral palsy, Parkinson's disease, and neuromuscular disorders without spending too much money or time [25].

1.2 Methods and Applications of Gait Analysis

The word 'Gait' represents the distinctive pattern of human walking, which is impacted by the complex interaction of the nervous, musculoskeletal, and cardiorespiratory systems. Recognizing the gender of the object for a far-away intelligent monitoring system is the most important application of gait recognition [1]. This capacity of the system can help enhance the understanding of the system to monitor the environment, building it as a powerful tool for security and surveillance.

Gait is used as a biometric measure, determining the gender of an individual or recognizing the individuals who are familiar with it, which is sensible as the walking patterns of humans enclose stylistic and informative interpretations. Gait recognition is a non-intrusive biometric technology that is becoming immensely popular. [2] Unlike traditional biometrics, which are face, fingerprint, vein pattern, and iris scan, which measure physical biometric characteristics, gait recognition is a social biometric method that is used to identify people based on their features of gait and motion.

Classification of Gender is critical for numerous applications such as analysis of the customer and security. Capturing the information of the face from a very far distance can be crucial in the natural world video surveillance systems. In these instances, human gait analysis can work better, which helps detect gender easily and is non-invasive. It is also very helpful in detecting individuals who are lost, especially people who are young at airports, bus stops, social events places and other public locations [4]. Gender classification using human gait is also useful in investigations as well.

In the field of healthcare, it is very important for the providers of healthcare to use instruments that are supported by evidence to help patients. Computerised gait analysis is one such tool that provides a detailed and objective assessment of a person's walking pattern. This analysis helps us understand the causes of any walking problems and guides us in selecting the most effective treatments.[3]. Human gait gender classification technology can also be installed in talking robots in supermarkets. This application helps the ability to track the gender of a person's robot through the cameras and helps customers provide gender-specific assistance such as giving directions to washrooms, men's and women's wear shops, ice cream shops, food stores, and other stores.

In real-life scenarios, when we use systems such as gait recognition, we usually encounter people from different viewing angles. We can use techniques like deep learning and machine learning to determine the problem of identifying a person's gender based on the style of walk. A person who walks can look different, which in turn depends on the particular view that you look from, such as side, front and back. By this, we can create a robust model by capturing images of a person's walking pattern from various different angles. This aids us in recognizing people accurately, which helps gait-based identification systems to be more trustworthy.

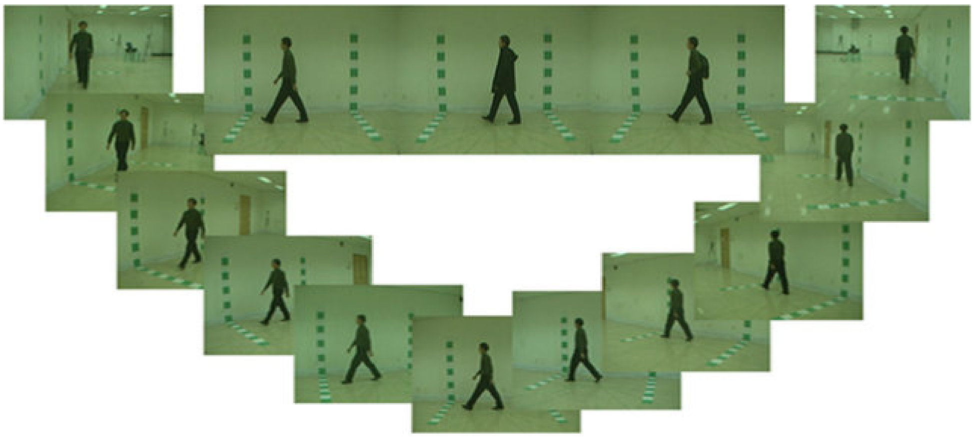

Figure 1.1: Images of multiple viewing angles [5]

Most of the research on gender classification has made use of average gait energy images (GEI), which are viewed from different angles. These average GEI combines silhouettes of a complete walking cycle from a particular angle, as shown in figure 3.2. The brightness of each and every image of different angles in GEI acts as a pattern cycle of the entire motion of the human gait.

The Institute of Automation at the Chinese Academy of Sciences (CASIA) presents the gait database to researchers who are interested in working on human gait recognition and other related fields [5]. CASIA-B is a large multiview gait database that contains the walking images of 124 people from 11 different angles. These are the images of human silhouettes extracted from the video files. Also, the images in the dataset include variations such as normal, clothing, and carrying conditions [6].

Machine Learning can help the analysis of gait images to be trained explicitly. It has helped people in the medical field to diagnose fast with accurate results being exceptional to human experts [9]. Machine learning has shown the capacity to automate tasks to provide better predictions.

In recent years, computer programs such as deep learning have become dominant in the field of computer vision. Specifically, deep convolutional neural networks (CNN), which are highly complex, are being used to resolve various tasks [7]. A few notable achievements working with images involve Krizhevsky [8] and his team training the highly developed image classifier, a network made up of three fully connected layers and five layers that analyze the images.

1.3 Research problems

The study focuses on the following research problems:

- How can deep learning and machine learning algorithms identify the walking pattern of an individual?
- Extracting valuable features to detect the gender of a person for machine learning algorithms.
- Tuning the hyperparameters for machine and deep learning models to obtain a better accuracy score.
- Ensuring that the models perform well in different conditions such as clothing, carrying and normal walking.

1.4 Contributions in this Study

The outline of this paper's contributions is as follows:

1. We propose a technique for the classification of gait that leverages an approach based on a Convolutional Neural Network. This method has the capacity to autonomously train the most distinctive changes in features of gait. Therefore, the network enables us to accurately predict the similarity between the pair of gait features. 2. We also provide machine learning models such as Decision Tree, Random Forest and k-nearest Neighbors to train the data containing images of multi-view or cross-walking conditions. We use pre-trained models and other techniques to extract the features of Gait energy images. 3. The accuracy scores that we have received for all the methods that we have performed are pretty good. The accuracy is about 90 percent for every model, as the dataset we have taken has denoised images. These images help the experiments to be trained correctly by extracting the edges of multi-view images. The experiment that we have performed shows the ability to be used for many practical applications as we have attained high results under multi-view conditions for CASIA-B dataset.

In the remaining part of this paper, we will discuss the related works in the Literature review from Chapter 2, demonstrate the methodology in Chapter 3, explain the results in Chapter 4, provide Discussions in Chapter 5 and finally, give a conclusion with Future Scopes in Chapter 6.

CHAPTER 2

Literature Review

2.1 Overview of the Literature review

The literature review is an in-depth review and assessment of the current state of the study on a specific subject. It assists in recognising trends, gaps in the information, and significant ideas in the field of study and provides perspective on the study's findings. It provides suggestions on how new developments can be produced and establishes the research conducted by the writer throughout the wider academic conversation.

2.2 Studies involved using Machine Learning and Deep Learning Methods

2.2.1 Related works using deep learning techniques

Convolutional Neural Networks have been greatly advanced in areas such as image classification and biometric recognition. Researchers have analysed the use of a support vector machine to replace the softmax layer function in CNN, and it has given promising results. A study [12] has investigated the performance of a CNN along with a linear SVM classifier for gender recognition. The dataset used is CASIA-B, and the best result that they have obtained is 87.94 percent, combining deep features with SVM. The outcomes demonstrated that SVM outperformed the softmax layer in gender recognition.

The Research investigates identifying the gender classification of people using a deep CNN [13]. The investigators used labelled multi-view human walking videos to train the network and performed thorough empirical experiments involving cross-view and cross-walking condition analysis. They also used many different pre-processing techniques and discrete network architectures with local, mid-level, and global features to get better outcomes.

Research focuses on enhancing people's identification by examining their walking patterns. It utilises deep convolutional neural network and image augmentation (IA) methods to improve the training data. This improves the algorithm's handling of image variations and increases its accuracy during identification [30]. The dataset's testing results showed major accuracy improvements with image augmentation achieving up to 96.23% for 200 people. This gives information on how the methods handle issues such as variations in human poses and partially covered objects.

In a paper [15], the researchers have introduced a gait recognition system called HEATGait, which enhances multi-scale graphs using a hop-extraction technique. They have used ResGCN architecture as the baseline model, combined with preprocessing and hop feature extractor. The accuracy they obtained is 93.3 percent using the CASIA-B gait dataset.

A work investigates gait analysis using machine learning and deep learning and compares their approaches under diverse circumstances. It illustrates the pipelines of deep learning utilized for the identification of people. The analysis demonstrates common approaches, difficulties, and unsolved problems using the behavioural features for gait recognition [35]. It also provides researchers with educated choices about data collection approaches while developing gait detection systems. In this particular field, this article is an invaluable resource for learning the distinctions between machine learning and deep learning. Feature extractors such as Histogram of Oriented Gradients (HOG) and Principal Component Analysis (PCA) are used to get relevant information on gait analysis.

This paper discusses the identification of abnormal behaviour by analysing how a person walks using computer vision methods. Gait modelling is developed by evaluating three feature extraction techniques: Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), and Center Symmetric Local Binary Pattern (CS-LBP) [37].

These techniques were chosen as they are effective in computer vision applications. Convolutional Neural Networks were used to classify the extracted features, resulting in 99% accuracy in detecting abnormal behaviour. This highlights the importance of feature extraction in enhancing anomaly detection systems.

In a study, numerous model-based experiments, such as static body measurements, installing ellipses, and encoding the shapes, were conducted to understand the human walking pattern. Principal component analysis was used to reduce the features of the collected data [27]. Deep belief networks were used to group the walking patterns. Compared to models such as k-nearest neighbour (KNN) and dynamic times warping (DTW), the deep belief network trained better using the CASIA-B dataset. Analysis was compared for different walking styles such as normal condition, wearing clothes and carrying bags.

The study of gait analysis has multiple applications, which are used for personal individual verification in forensic science and security protocols, handling gait-based queries in healthcare industries and analysing changes in the walking pattern of a person concerned with ethnic origin and gender [49]. Principal Component Analysis (PCA) is used to perform gait recognition on images gathered from the RGB camera. With the help of PCA, the key characteristics were extracted from the gait patterns captured by the camera. The neural network (NN) model is generated using the data reduced by PCA for the purpose of identification of an individual.

2.2.2 Related Works using Machine Learning and Other Techniques

The method proposed in [16] illustrates feature extraction of GEI using pre-trained models such as DenseNet, VGG16, VGG19 and others. Further to the feature extraction, they have trained the XGBoost classifier and tuned the parameters. The best score they are achieving is 93.88 percent, with DenseNet as the feature extractor with a tuned XGBoost Classifier.

An approach is developed for classifying human gender based on spatial-temporal reasoning using the Casia gait dataset [17]. Two feature extraction methods, model-based and free model-based, are used for the extraction of temporal and spatial data, respectively. The data is classified using SVM, and accuracy was obtained at 97.63 percent for processing spatial and temporal information.

A novel pattern is proposed from a study which is called Gait Principal Component Image, which represents spatiotemporal information of gait appearance [19]. The images are grey-level, which amplifies the dynamic variations of different body parts. The timing of the step is detected using the LLE coefficients, and the KNN classifier is used for the prediction of gender.

In a research paper, a new strategy is implemented for recognizing the gender of a person by using smartphone sensors, particularly the accelerometer. The method used reduces the direct interaction with the individuals, who only need to hold their smartphones and walk normally. The paper [20] utilizes different data mining techniques and train machines such as Decision Trees, k-nearest Neighbor (KNN), and deep learning algorithms.

A paper addresses the analysis of human gait performance using various classification techniques, such as Decision Tree, Random Forest and KNN Models [23]. A tool called RapidMiner Studio is used to compare classification techniques based on various parameters. The results show that the random forest algorithm performs better compared to other models classifying the gaits.

A study points to the effectiveness of Machine Learning (ML) approaches to the researchers in gait analysis [10]. Research by scholars and other articles is being used to explain the use of supervised machine learning algorithms to detect the neurological effects in the field, such as gait disorders, gait asymmetry, events and other activities. The paper also gives more information about how machine-learning techniques are applied to detect different medical conditions, predict the patient's recovery duration, and supervise the instruments of clinical diagnostics.

In recent decades, many computer vision systems have developed to identify the classification of gender using gait analysis. The paper [11] gives information on gender recognition based on gait analysis using the systems of computer vision. The method used here combines the information about a person's walking style from multiple views at the feature level. From this, gait energy images (GEI) were developed from multiple camera views and analyzed using multi-linear component analysis to merge all views, resulting in effective performance for multi-view gender classification using gait.

A new approach to multi-view gait recognition has been presented in a study using View Transformation Model (VTM) based on gait energy image (GEI) and Singular value decomposition (SVD). Linear Discriminant Analysis (LDA) is used to increase the performance of VTM by optimizing the GEI feature vectors [14]. The proposed method achieves up to 90 per cent accuracy for a particular viewing angle.

A study has gathered interest in identifying a person's gender depending on their natural walking patterns [21]. From uncontrolled sequences of gait, human silhouettes are extracted using background subtraction and clustering them into many different groups. Average Gait Image (AGI) is computed as a feature for each group. Then, the distance metric is used to minimize the variations within the gender groups and maximize the variations between the gender groups so that effective information is generated for gender classification.

A work examining the effectiveness of different viewing angles for the classification of gender using gait biometrics is developed. Videos of each person walking from seven different angles are captured and stored in a database [22]. Gait features are extracted using robust techniques, and class separability is reviewed from different viewing angles. Experiments are designed to evaluate the performance of gender classification with changes in angles.

A study explains that applying advanced feature extraction techniques to gait energy images (GEI) increases the identification of a person in gait detection. Feature extraction methods such as LBP give detailed knowledge from the data used [36]. Discriminative Common Vectors (DCV) are effective dimensionality reduction techniques for handling high dimensions of LBP features. This experiment demonstrated that the CASIA-B database has provided high recognition accuracy.

2.3 Challenges and Improvements in Gender Identification using Gait Analysis

A trend has been developed in techniques of biometric identification, which is explained in many papers by numerous authors that illustrate the growing acceptance of recognising individuals based on their walking pattern. Models based on deep learning, feature extraction techniques, and machine learning algorithms are some of the different strategies that are explored and evaluated. Although there are a few challenges associated with the speed of walking, clothing, and carrying conditions, modern methods that involve Principal Component Analysis, Conventional Neural Networks, and Gait Energy Images have been implemented with identical performance. To improve these techniques and explore more regarding the mentioned hurdles, more research needs to be conducted to discover effective approaches to gender identification.

CHAPTER 3

Methodology

3.1 Dataset Description

CASIA, Institute of Automation of the Chinese Academy of Sciences in China, is one of the largest institutes of research which has accomplished field works related to computer vision, and they have also created multiple biometric databases. The main objective of creating the CASIA-B gait database is to develop a large-scale dataset for the research of gait and the evaluation of algorithms related to gait recognition. The CASIA-B gait database was published to perform research in the field of biometrics in the year 2005. At that moment, biometric technologies like facial recognition and fingerprints were widely used, while gait recognition was newly developed.

The dataset that is used in this study is CASIA-B [26], which contains average gait energy images formed by combining the dissimilarities in silhouettes between adjacent frames for 11 different viewing angles from 0 degrees to 180 degrees having 18 degrees differences for each multi-view gait images, as shown in figure 1.1. In total, there are 124 subjects, with 31 female subjects and 93 male subjects, as shown in figure 3.1. This bar graph figure is taken as a screenshot from the output of the code.

These images offer an advantage in preserving both kinetic and static information about a person's walking style. Therefore, average GEI as shown in figure 3.2, serve as feature images as they are better suited to capture silhouette variations in a complete walking cycle.

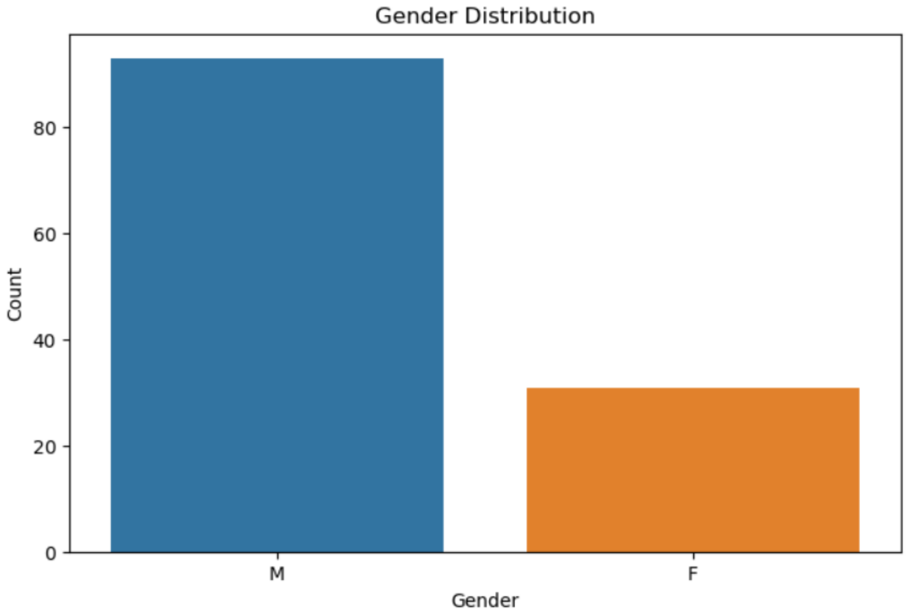

Figure 3.1: Number of male and female subjects

Figure 3.2: Combining Silhouettes to form Average GEI [58]

There are three variations in the images of people such as normal walking, images wearing coats and carrying bags. These images which are used here have been cleaned to remove noise. These denoised images have several benefits that help to clear out unwanted visual distortion while having specific information to train the model. These simplified images assist in maintaining the structures, edges, and important features of an image.

3.2 Deep Learning Model

To study the pattern of how a person walks using the CASIA-B dataset, we should first understand the images present in the dataset [26], organize the images, and we have to load the images in the environment where the analysis is done, for example, Jupyter notebook.

Organizing the data is nothing, but the dataset usually contains folders having unique IDs for each person, and we have to separate them into male and female folders locally [29]. Inside these folders, there are sub-folders for different walking conditions, such as normal walking, wearing a coat and carrying a bag, as shown in figure 3.3. These sub-folders consist of 11 average silhouette images from 0 degrees to 180 degrees of different angles.

Figure 3.3: Walking Conditions: Normal, Coat, Bag [5]

Now, the images of males and females can be loaded into the working environments. After uploading the data, the loaded images of males and females and their labels are combined, and the images are adjusted by resizing and normalizing them. Resizing ensures that all the images are in a uniform condition and the same size, which helps machine learning and neural network models train better and saves memory usage. This makes the models to be designed that require the fixed input. Normalizing the images to a fixed range from -1 to 1 helps the algorithms learn better and faster.

In gait analysis, it is essential to convert the labels to categories, as this helps with tasks such as classification, which differentiates people by identifying their gender. This type of conversion is needed as machine learning and deep learning algorithms normally need the format in binary values that are one-hot encoded. 'Keras' is used to perform one-hot encoding, and further, the data is been split into 80% train where the algorithms will learn to understand the patterns and features of the data and 20% test to evaluate the performance of the model.

Following the train and test split, Data augmentation is applied to the training set. It is very relevant to experiments like gait analysis, which helps to modify the classes of existing data and provides a diverse view of the training set, as explained in [30]. Data augmentation improves the model's ability to hold the data and gives better results. This increases the size of the dataset, enhances the robustness of the model and reduces the overfitting.

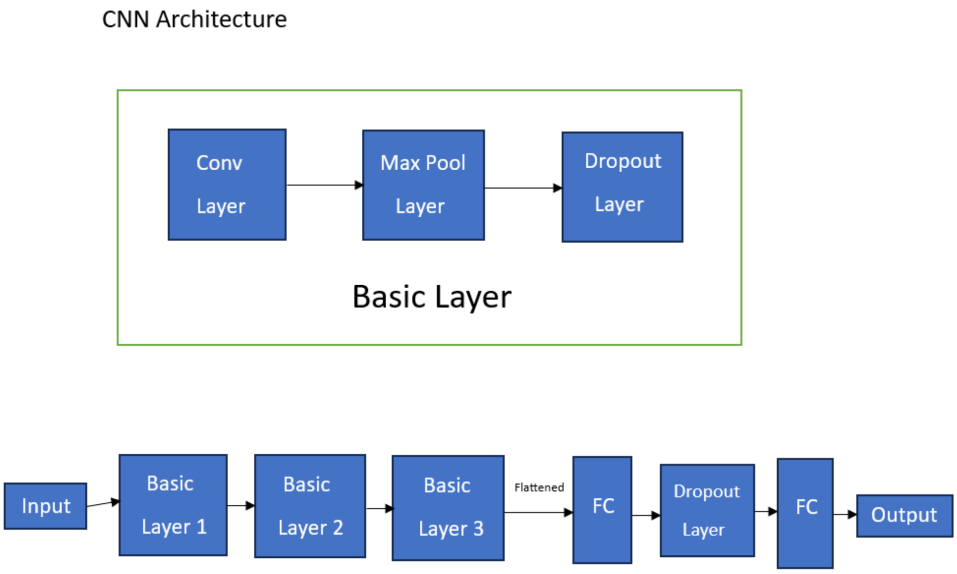

Figure 3.4: CNN Model

3.2.1 Convolutional Neural Network

A Convolutional Neural Network is a deep neural network algorithm that can analyze people's walking patterns. Gait analysis entails extracting significant information from sequences of silhouette images captured of a person's movement. In this study, due to the computational power of the running system, we used 2000 male images and 2000 female images, which also balanced the data. The figure 3.4 is designed to explain the architecture of CNN.

Input

In this experiment, the images used here are normalised by dividing the input data with each pixel by the value of 255. The batch size given here is 32, which is considerably smaller, as it leads to better performance of the model. This study has 12 layers in CNN architecture: three 2d convolutional layers, three max-pooling layers, and four dropout layers with two fully connected layers. The complete CNN architecture of this study is shown in the figure 3.4. The size of the input image is given with the pixel value of (128 X 128).

Convolutional Layer

A part of the neural network layer is a convolutional layer that performs convolutional operations using a filter known as the kernel to the input data. These filters slide over the input data to carry out element-wise multiplication and sum these results to produce a feature map. This mathematical work enables the structure to identify image patterns such as edges, textures, and shapes in the input. The size of the filter in every 2d convolutional layer is (3x3). These filters help in the extraction of useful information about the image and bring reasonable changes to the input pattern.

In this study, there are 3 sequential triplets of the 2D Convolutional layer with Rectified Linear Unit (ReLU) as the activation function, the Max-pooling layer, and the Dropout layer. The first sequence of conv2d has 32 neurons, and the second sequence has 64 neurons, followed by 128 neurons in the final sequence. A similar architecture is shown in [31], with two sequential triplets involving convolutional, max-pooling and normalization layers.

$$f(x) = \text{ReLU}(x) = \max(0, x) \tag{3.1}$$

The activation function used in the referred study is also ReLU, as it helps to activate only the positive input neurons. Negative input neurons are converted to zero and are deactivated. The mathematical equation of ReLU is given in equation (3.1).

Max pooling Layer

The Max pooling layer is part of neural networks that performs operations after every individual convolutional layer. It lowers the dimension of the image from the previous output image by reducing the number of pixels. In gait analysis, using max pooling is an advantage as the shape of the body changes or fluctuates according to the movements. Pooling units can be used with a factor of 2 when subjected to max pooling [32]. This illustrates that the image is down-sampled with the filter size having the dimension (2x2) and the value of stride is 2 so that the pooling windows do not overlap.

In this paper, the size of every max pooling layer is (2x2), and the input passed into this layer takes the maximum pixel value of the four values. After completing all the strides, this gives the complete output from the max pooling layer. The operation for max pooling is given in equation (3.2).

$$y_{i,j} = \max_{m,n} (x_{i+m,j+n}) \tag{3.2}$$

Where x(i,j) represents the values of input, y(i,j) represents the values of output, and m and n are the dimensional values of the pooling window.

Dropout Layer

The word dropout in neural networks refers to ignoring the particular nodes present in the layers randomly during the training of the data as shown in figure 3.5. The nodes are removed in terms of probability. It is used as a regularization technique to avoid overfitting, which means the noise present in the data. Generally, if the hidden layer has 1000 neurons and the dropout operation has been conducted with the probability value of 0.5, 500 neurons will be dropped in every consistent batch of the iteration [33].

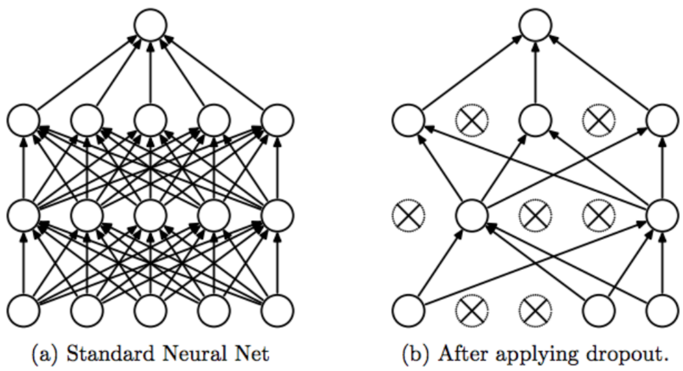

Figure 3.5: Dropping particular Nodes [59]

In this study, the dropout layer has been used with a probability value of 0.25 in the first three sequential convolutional layers of triplets. This will remove 25% of the neurons present in each layer of convolution. Another dropout method has been performed after the flattening of the data between the two dense layers. The probability value is given to be 0.5, which drops the 50% of the neurons coming from the output of first dense layer.

Fully Connected Layer

A fully connected (FC) layer, also called a dense layer, is a part of neural networks that is placed at the end of a Convolutional Neural Network. In the dense layer, each neuron is connected to every other neuron from the previous layer. This illustrates that the output of each neuron from the previous layer acts as the input to every neuron in the present layer, having certain weights and biases for each neuron [34]. This enables the model to train non-linear and complex operations. Two Fully connected layers are used in this study. The first dense network has 128 neurons with ReLU as the activation function, as shown in equation (3.1). The second dense network has only two neurons with softmax as the activation function.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{3.3}$$

Softmax is used when an input needs to be assigned to one of several classes by allocating probabilities to each class. This indicates how likely an input belongs to a particular class. Softmax function is given as shown in equation (3.3).

Output Layer

The final recognition or classification of a person according to their walking pattern analysis is performed in the last layer of CNN, which is the output layer. The probability distribution in gait analysis is used to make a binary judgement of a particular person.

While compiling the CNN model, Adam optimizer is used, which is a systematic algorithm that handles large datasets with several parameters to optimize the gradient descent. The optimizer incorporates momentum and root mean square propagation algorithms to speed up the gradient descent using the exponential moving average.

The accuracy of the CNN model is about 91%, where the classification of Males gives a precision score of 0.86, a recall score of 0.99, and an f1 score of 0.92. The female classification precision score is 0.99, the recall score is 0.80, and the f1 score is 0.89. The loss of the CNN model is 0.2551.

3.2.2 Tuning the hyperparameters of CNN model with K-fold cross validation

Fine-tuning the hyperparameters is required in deep neural networks, especially convolutional neural networks, to maximise the performance of the model. This method gives the opportunity to choose the combination of hyperparameters to achieve the best accuracy and generalize the data.

Hyperparameters in CNN

- In the CNN model, filters, also known as kernels, are used in subsequent layers of convolution, which detects features from the input images, such as edges, textures, and walking patterns, in gait recognition. The extent of the generated feature maps relies on the number of filters used. In this study, three sets of filters are used, which are:

  filters1: It is used in the first convolutional layer having 32 or 64 filters.

  filters2: The second convolutional layer is used with 64 or 128 filters.

  filters3: The last convolutional layer has 128 or 256 filters.

  The capacity of the model is determined from the above sets of filters, and the model is well-trained using the input images.

- The dimensions of the filters are determined by the kernel size. In this study, we used two kernel sizes: (3,3) and (5,5). The small kernel, which is 3x3, will capture finer information from the image, while the large kernel size, 5x5, collects broader global patterns of the images.

- Pooling layers help to reduce the computational cost and the overfitting issues by lowering the spatial dimensions of the feature maps. The pool sizes used in this study are (2,2) and (3,3), which determine the size of the window over which max pooling or other types of pooling are performed.

- The optimizer plays an important role during model training by determining how the model weights are adjusted. Two widely recognized optimizers are used in this study:

  Adam: It is known as adaptive moment estimation and is renowned for its efficiency and convergence speed.

  SGD: Stochastic Gradient Descent is simpler and sometimes more robust, specifically when coupled with learning rate and momentum.

- The batch size determines the amount of training samples used during one iteration of training the model. The batch sizes used are 16 and 32. A smaller batch size (16) frequently updates the weights of the model, resulting in noisier updates with fast convergence. On the other hand, larger batch sizes, such as 32, produce more consistent updates, but additional memory is needed.

- Epochs represent the number of complete iterations through the whole training dataset. This study uses 10 or 20 epochs to determine the right balance between underfitting and overfitting.

This study utilizes k-fold cross-validation with five splits to ensure the stability and usefulness of the model for different datasets. Using this technique, the training data is divided into five different folds, four of which are used for training the data, and one fold is used for testing purposes. This method is iterated five times, with all different folds used for testing in each iteration. By averaging all five iterations, the final performance metric is computed.

After manually hyper-tuning the parameters using the CNN model with the k-fold cross-validation technique, this study achieved the best accuracy of 98.83%.

3.3 Machine Learning Models

To perform each of the machine learning models on gait analysis in this study, 3000 male and 3000 female images were loaded concerning the computational power of the system. After the data is uploaded, the male images are assigned to the value zero and female images are assigned to the value of one. All the pictures of males and females are combined, and labels assigned to the images are merged.

3.3.1 Random Forest

Machine Learning (ML) algorithms are being used in various applications in our daily lives, offering recommendations in various domains. The amazing fact is that ML techniques have been proven to be more effective and productive than human interventions in dealing with multiple sectors, like in healthcare, solving issues such as the diagnosis of skin and breast cancer [38] and also demonstrate promising performance in terms of classification. In gait analysis, machine learning develops the connection between inputs and outputs in a biomechanics simulation [39].

Once the data is uploaded, the pictures of males and females and their labels are merged into different variables. Further, the images are resized to 64x64 to ensure uniformity. With this, machine learning models train more efficiently and use less storage space.

Feature Extraction Methods

The procedures for feature extraction are very important in the development of systems for detecting human walking patterns. The frames of the gait video were utilized to establish an array of features. Spatial and temporal characteristics are equally essential within gait features [40]. For the purpose of obtaining a smaller number of features which contain most of the details included within the original image, the method known as feature extraction is used. This reduces the dimensionality of the data.

The feature extraction methods used before training the random forest model are:

- Histogram of Oriented Gradients (HOG):

  The initial objective of using the HOG feature classifier was to identify the pedestrians. It records the number of instances in the gradient direction that occur within a particular detection window or cell. This cell is very small, where the rectangular shape of the image utilizes the HOG feature computation. To create a feature vector for the classification of the image, the gradient direction histograms must be evaluated first for each cell and then combined [37]. In this study, the number of pixels per cell used is (8,8) and (2,2) cells per block by enabling the feature vector for each and every image. These images are returned in the form of an array. The primary phases in computing the HOG features are:

  Computing Gradients: Gradients in vertical and horizontal directions are evaluated in this stage. The gradient's magnitudes and angles are then calculated using the equations shown in equations (3.4) and (3.5).

$$G_x = I(x+1, y) - I(x-1, y) \tag{3.4}$$

$$G_y = I(x, y+1) - I(x, y-1) \tag{3.5}$$

where the gradient components are denoted as Gx and Gy, which are in horizontal and vertical axes, respectively. It is important to note, as stated in (3.4) and (3.5), that the convolution of images with the Sobel mask improves the calculation of gradient components. The features of the image can then be thoroughly examined by utilizing equations (3.6) and (3.7) to determine the gradient's size and direction.

$$M = |G_x| + |G_y| \tag{3.6}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{3.7}$$

Where θ (x,y) represents the gradient's direction, and M indicates the magnitude.

The implementation of HOG is that the GEI is separated into small interlinked cells. The edge direction is computed for each cell and partitioned into eight angular bins, indicating the distribution of gradient directions. Further, pixel values are normalised to ensure similarity. The vertical and horizontal gradients are determined and utilised to detect a person's walking pattern in both the x and y axes [41]. This approach helps to differentiate between different people and presents accurate information on motion dynamics in a person's walking cycle.

- Local Binary Pattern (LBP):

  LBP, which was initially presented by Ojala et al. [42], is an effective technique used for describing the texture, which depends on statistical analysis and has shown its usefulness in this study. Although numerous LBP improvements have been commonly employed to face analysis because of their better performance in classification, it remains to be seen how well compact they are [37]. From this, feature integration is shown to be a powerful technique. In particular, higher accuracy in face analysis is demonstrated for the LBP type, which increases information sharing.

  The analysis of gray-level intensity between a core pixel and its outer neighbours is essential for the computation of LBP. A binary value is allocated to every nearby pixel according to the degree to which its intensity exceeds that of the core pixel. The binary values are combined in order to create the binary pattern that serves as a distinctive character of the image. The mathematical calculation and interpretation of LBP results can be stated using equations (3.8) and (3.9) [43].

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{n=0}^{7} \delta(g_n - g_c) \cdot 2^n \tag{3.8}$$

where, within a radius R, gn specifies the value of one of eight neighbouring pixels surrounding the central pixel, whereas gc indicates the central pixel value at coordinates (xc, yc). P defines the total amount of pixels in the neighbourhood. In addition, the sign function $\delta(\cdot)$ is expressed as:

$$\delta(x) = \begin{cases} 10 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

The histogram of LBP values having specified range is computed in this study. Using particular number of points and a radius, it is calculated for every pixel in the image. LBP images are flattened to a 1-dimensional array by specifying the bin edges with uniform and non-uniform patterns.

- Principal Component Analysis (PCA):

  PCA is a simple and efficient technique that assists in solving issues and retaining valuable information by lowering the data's dimensionality and maintaining the original information as much as possible [44]. PCA is the most widely utilized feature extraction method used in gait analysis. In order to identify the individuals, the PCA approach usually analyzes a person's walking patterns to a known database [45]. PCA helps to reduce the amount of computation needed and it will help run gait energy images fast with less space and memory.

  Using PCA, the original data is converted to new sets of orthogonal coordinates [46]. The key stages of PCA are:

  Covariance Matrix: To analyze the interactions between numerous features, the covariance matrix Σ of data X is calculated as shown in equation (3.10).

$$\Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)(X_i - \mu)^T \tag{3.10}$$

Where X_i and μ are the i-th data point and the mean of the data, respectively.

Decomposition of the eigenvalue: The covariance matrix is decomposed into eigenvalues and eigenvectors as shown in the equation (3.11).

$$\Sigma\mathbf{v} = \lambda\mathbf{v} \tag{3.11}$$

Where λ represents eigenvalues and v represents eigenvectors.

In this study, the flattened images are extracted using principal component analysis (PCA), with the number of components equal to 50.

All these methods, which are HOG, LBP and PCA are used for extracting the features from the combined images of males and females one by one and are concatenated together as combined features.

The advantage of merging HOG and LBP is that the produced feature set includes both the local texture and the global structure information from LBP and HOG, respectively. Later, using PCA to lower the dimensionality provides an effective approach for gait analysis. PCA uses the characteristics of both HOG and LBP to acquire comprehensive features and to boost the computational performance and adaptation of the model.

These images are later split into training and testing sets containing 80% train data and 20% test data.

Random Forest Model

Random Forest is one of the popular ways of combining classifiers that provides a high degree of performance with a reasonable computational cost. In a random forest, each and every classifier is a decision tree, and every single tree is generated by independently assigning the attributes of the nodes to identify the decision of classifications [47]. Therefore, each decision tree is randomly sampled independently, but they all possess the same vector distributions. The architecture of the Random Forest model, which is designed, is shown in figure 3.6.

Bootstrapping is a method of building a decision tree by initially randomly picking a part of the training set. This implies that a unique set of data is minimally found in each tree. Further, a subset of features is randomly chosen to split the nodes of the trees,

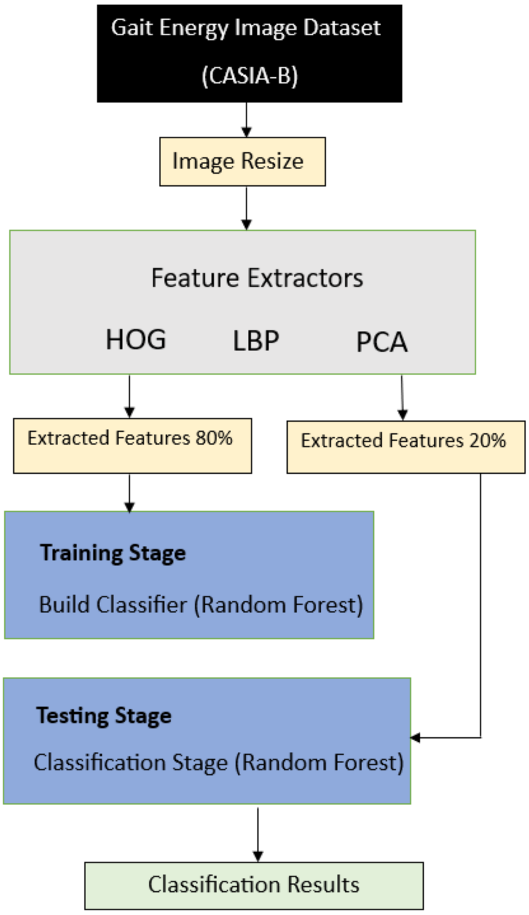

Figure 3.6: Random Forest Model Architecture

which is known as random feature selection. This guarantees that each tree focuses on different features. Using a particular subset of characteristics and data, a decision tree is constructed, which terminates when the tree achieves its maximum depth or a minimum number of samples per leaf [48]. This process is repeated many times to build numerous trees and is used as a hyperparameter by modifying the number of trees. As a result, the classification is predicted by averaging the outputs of all the trees.

The overall accuracy given by random forest after predicting the classification is 96.44%. The precision, recall and f1 score of male prediction are 0.96, 0.98, and 0.97, respectively. The female prediction has a precision of 0.97, a recall value of 0.94 and an f1 score of 0.96.

3.3.2 Tuning the hyperparameters for Random Forest using Grid Search method

For a Random Forest model, hyperparameter tuning requires changing different sets of combinations of parameters with the aim of utilizing the grid search method to identify the best set of values. The grid search technique identifies the most effective configuration for the random forest model by systematically evaluating the performance of the model with numerous combinations of hyperparameters.

Hyperparameters in Random Forest

- The number of estimators determines the number of trees used in the random forest. If more trees are used, the computation time increases, resulting in better performance. The number of trees used here are 100, 200 and 300.
- The maximum depth parameter where the values are 10, 20 and 30 used in this study controls the maximum length of each tree that can develop. The maximum information is captured depending on the depth of each tree, which is also likely to be overfitted.
- The minimum number of sample splits used here are 2, 5, and 10. It controls the sensitivity of the model with respect to the changes in the data. More common trees are generated with greater values.
- The complexity of the model is concerned with the number of samples required to be at a leaf node, which are used as 1,2 and 4. This stops from generating nodes that include very few samples.
- Bootstrap parameter determines the samples of the bootstraps used when creating trees. Each tree is generated as a replacement sample if it is set to the value True. All the trees are created from the original dataset if bootstrap is adjusted to False.

Cross-validation with the value 3 demonstrates that the data will be divided into three folds while tuning the parameters. The model will contain two training folds and one testing fold, which cycles through every single fold for each pair of hyperparameters. After tuning the parameters for the Random Forest model, the overall accuracy obtained is 97.94%.

3.3.3 Decision Tree

A decision tree is an algorithm obtained by a supervised machine-learning technique that can be utilized to solve problems related to classification and regression. The decision tree method is typically employed to solve classification issues. It is a tree-structured predictor in which every single leaf node indicates the results, features of the dataset are represented by the internal nodes, and branches show the decision rules [50]. Decision nodes and leaf nodes are the two distinct types of nodes present in the Decision Tree. Leaf nodes are the outcomes of decisions, and they do not include branches, as shown in figure 3.7, while decision nodes are used to generate decisions and have multiple branches.

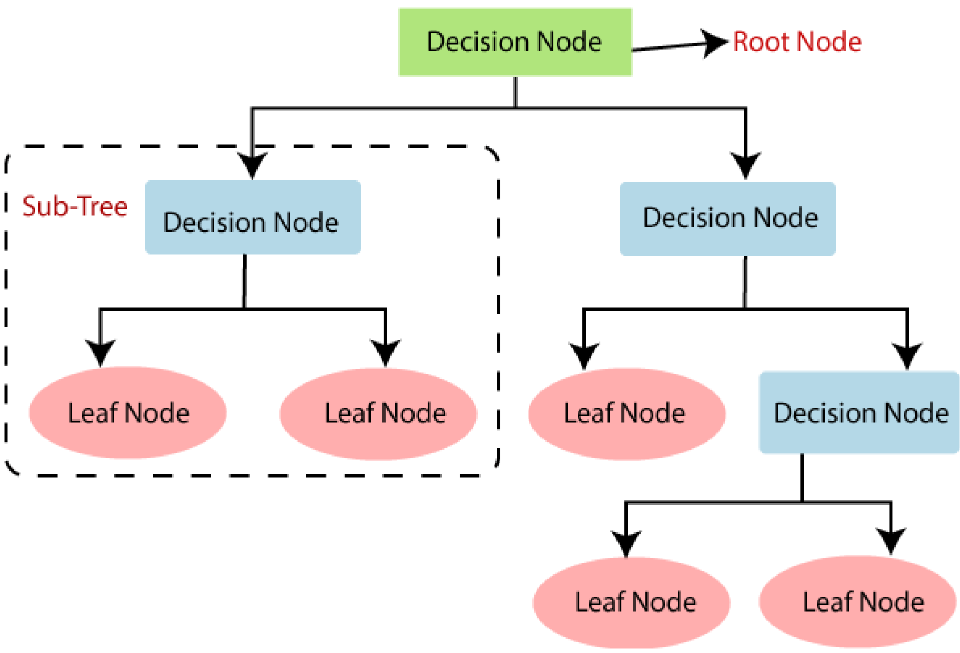

Figure 3.7: Structure of Decision Tree [50]

To perform the Decision Tree algorithm on gait analysis, the data used are resized to the value of 224x224 to make sure consistency is present to get better results. Finally, the images are split into 20% test data and 80% train data before feature extraction.

Feature Extraction using Pre-trained DenseNet-121 model

DensNet, which stands for Dense Convolutional Network, is an architecture of deep learning designed for convolutional neural networks. By proposing an important connectivity structure within CNNs and handling problems such as reusing features,

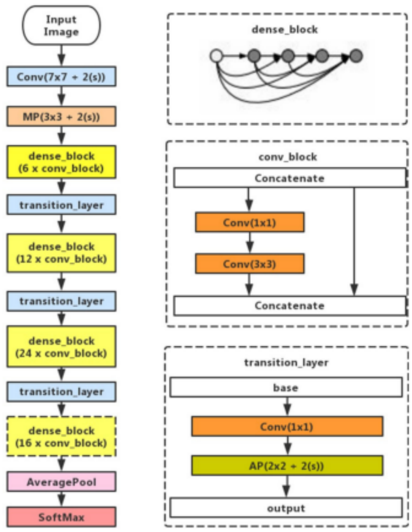

Figure 3.8: Left is DenseNet model Architecture, right is DenseBlock, Convolutional Block and Transition Block [52]

decreasing gradients, and parameter efficiency, DenseNet redesigned the area of computer vision. DenseNet provides direct connections between each and every layer throughout the block, compared to the traditional CNN designs, where every single layer is connected to the following layer. Due to this strong connection, a large quantity of data flows through the network, allowing each layer to absorb mappings of features from all the previous layers [51].

In this study, the features of the images are extracted for training and testing data separately using the DenseNet-121 base model. The DenseNet-121 architecture, as shown in figure 3.8, includes three distinct types of blocks. The initial block is the convolutional block, an essential dense block. The convolutional block has more similarities to the identity block in the ResNet model. The second block is the dense block, in which the convolutional blocks are connected strongly in dense and is one of the important components in the DenseNet model. The last and final block is the transition block, which interconnects two adjacent dense blocks. The transition block lowers the dimensions of feature maps as the dense block's map with feature size remains the same [52]. The bottleneck design method is utilized in all the blocks of the model.

The Configuration information of the parameters used for feature extraction are:

The weights parameter has been utilized, which was previously trained on the ImageNet dataset as reflected in the model. This allows the model to take advantage of the features that are learned from a large dataset consisting of more than 14 million images with 1,000 classes.

The fully connected layers that had been designed for ImageNet classification are removed if the include top parameter is been assigned to False. By this, only the convolutional base is maintained, therefore making it suitable for extracting generic features from the data.

At last, a global average pooling layer is added after the final convolutional layer by modifying the pooling parameter to 'avg'. This layer offers a simplified version of the image by aggregating the spatial dimensions of the 2D feature maps to generate a 1D feature vector. This helps to represent machine learning models such as K-nearest Neighbors (KNN), Decision Tree and other algorithms.

These feature-extracted images are finally fit to the Decision Tree classifier, and the overall accuracy obtained by the model is 89.3%. The precision, recall, and f1 score for predicting the male classification are 0.88, 0.92 and 0.90, respectively. While the values of precision, recall and f1 score for female classification are o.87, 0.91 and 0.89.

3.3.4 Tuning the hyperparameters for Decision Tree using Grid Search

Using the grid search method, the decision tree classifier can be fine-tuned by changing the multiple combinations of parameters and identifying the most effective values. By systematically evaluating the performance of the decision tree classifier, best configuration of the model can be adopted utilising the various combinations of parameters through grid search method.

Hyperparameters of Decision Tree

- The extent to which a tree can grow is measured by its maximum depth. In this study, trees are created using varying depth values, such as 5, 10, 15, and 20 and an infinite depth value which is None.

- The threshold values such as 5, 10 and 15 specify the minimum sample sizes that are required to split an internal node. These threshold values represent distinct sample sizes required to split.

- The minimum number of samples used in a leaf node prevents overfitting, limiting the growth of the tree.

- The number of features to consider while determining the best split is by testing all characteristics with the value 'None', the number of features doubled using the value 'sqrt', and the total number of features logarithms with base two, which is 'log2'.

- The optimum split at each node is evaluated by the degree of quality of a split using either entropy or Gini impurity.

The grid search approach splits the training data into five separate groups with the objective of using a 5-fold cross-validation technique. By this, the model is trained five times, having each subset used as a validation set once and the other four subsets used to train the data. This method guarantees that the overfitting issues can be reduced and provide better performance of the model by training the images rigorously.

The method runs in parallel by making use of all the CPU cores that are available which significantly speeds the process of grid search by splitting the task to all different processors. The users can track the information on the grid search's progress by setting the verbose level to two. The information includes the total number of parameter combinations attempted and the duration taken for each combination. After performing the parameter hypertuning for the decision tree, the best accuracy obtained is 87.54%.

3.3.5 K-nearest Neighbors (KNN)

KNN is one of the favoured and adaptable machine learning algorithms which is mainly used for classification and regression tasks as it is simple and easy to implement. It is an adaptable option for various types of datasets as it is capable of handling categorical as well as numerical information. Its non-parametric approach focuses on predictions based on the similarity of the data points in a particular dataset. It is also less sensitive to outliers when compared to the other machine learning techniques [53].

Utilising the distance metric system, which can be Euclidean distance, the KNN technique identifies the k neighbors which is closest to a specific data point as shown in the figure 3.9. The model compares this newly pointed data to the instances of the existing data, and based on the correlation, the algorithm classifies this new data point according to its target class [54].

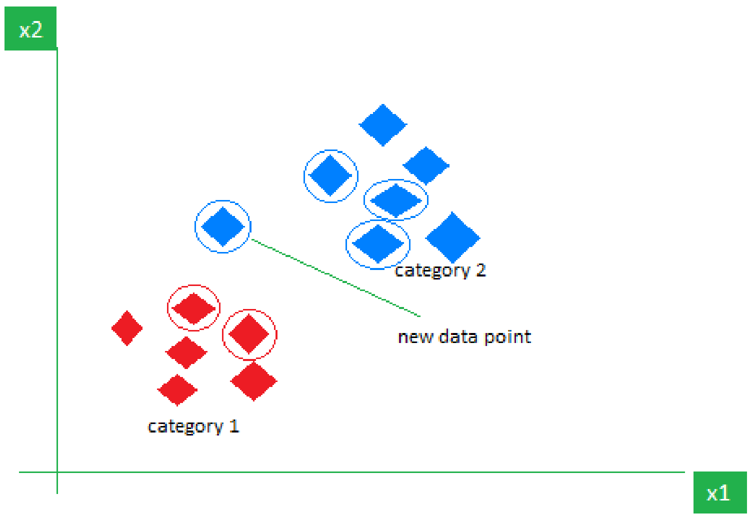

Figure 3.9: KNN working visualization [53]

To work on the KNN algorithm, for better data performance, the images are resized to a value of (224,224), which ensures the consistency of the images while training. At last, the images are split into 80% train data and 20% test data.

Feature Extraction using Pre-trained ResNet model

A CNN architecture, ResNet-50, belongs to the group of Residual Networks, which is combined with a group of models created to overcome the issues while training the deep neural network. This algorithm, developed by researchers at Microsoft Research Asia, is widely recognised for its complicated structure and efficiency in industries such as image classification [55].

The ResNet-50 architecture has four main components, which are the convolutional layers, convolutional blocks, identity blocks and fully connected layers as shown in the figure 3.10. The features of the input data are extracted from the convolutional layers, and the extracted features are modified using the convolutional block and identity block. At last, the images are classified using the fully connected layers [56].

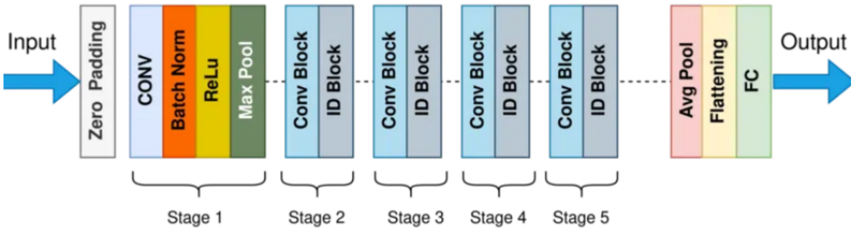

Figure 3.10: ResNet-50 Architecture [56]

The convolutional layers in ResNet50 are constructed using multiple sequential convolutional layers, and these layers are triggered by batch normalisation and ReLU activation function. The edges, textures, and patterns are extracted from the input data by these layers. The spatial dimensions of the feature map are minimised, securing the essential features by the max pooling layers, which appear after the convolutional layers.

The convolutional block and the identity block are the two fundamental parts of the ResNet50. The identity block is a basic unit that puts the initial input data back into the output while analysing the input images across several convolutional layers. This is helpful in the process of developing the network by correctly mapping input images to output images [56]. The identity block and convolutional block are similar to each other, but the convolutional block contains an added 1x1 convolutional layer that minimises the number of filters before using the 3x3 convolutional layer.

The final classification of the images is carried out by the fully connected layers, which is the last section of the ResNet-50 architecture. The ultimate probabilities of classes are generated by adding a softmax activation function to the result of the final fully connected layer.

The extracted images from ResNet-50 are finally fit to the KNN classifier to predict the overall accuracy of the KNN model, and the result obtained is about 95.26%. The precision, recall and f1-score of male classification are 0.94, 0.98, and 0.96, respectively. The female classification has the scores of 0.97, 0.9 and 0.93, which are precision, recall, and f1-score, respectively.

3.3.6 Tuning Hyperparameters for KNN Model using Grid Search

The widely used hyperparameters in the KNN algorithm are the distance metric, the weighting scheme, and the number of neighbours, which are applied for classification and regression problems. Techniques that involve grid search can be used for choosing the best combination of hyperparameters. By exploring multiple hyperparameter results, researchers can recognize reliable and precise models [57].

Hyperparameters in KNN

- The number of nearest neighbours is applied to generate predictions. The values used are 3,5,7, and 9, which indicates multiple options for the value of k. Achieving the perfect balance between overfitting when the value of k is too high and underfitting when the value of k is too low might be obtained by adjusting k.

- The weights parameter has two options, which are 'distance' and 'uniform'. Uniform indicates that every neighbour influences the predictions similarly, while the distance option provides the adjacent neighbors more significance, which is helpful when the points are nearby.

- The metric parameter calculates the neighbor distances by specifying the distance metric. It offers testing in several ways using options such as 'euclidean', 'Manhattan', and 'minkowski' to calculate the distance between the points, which can help improve the performance of the model.

The grid search method involves 5-fold cross-validation, exploring the hyperparameter grid and the known model. This technique produces an accurate evaluation of the performance of the model by analysing multiple training and validation splits, which helps choose the best hyperparameter combination.

3.4 Evaluation metrics

3.4.1 Overview of Evaluation Metrics

An essential tool to assess the efficiency of machine learning algorithms is called an evaluation metric. They help to identify the regions that require enhancement and provide an understanding on how effectively a model is producing the results. Here is an overview of some common evaluation measures used for classification tasks in machine learning.

3.4.2 Classification Metrics

- **Accuracy:** It calculates the number of instances that have been properly recognized across all the scenarios such as true positives and true negatives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.12}$$

  Where,

  - TP refers to True Positive values
  - TN refers to True Negative values
  - FP refers to False Positive values
  - FN refers to False Negative values

  Accuracy is very helpful data when the data is almost balanced between the number of instances in every single class.

- **Precision:** It measures the percentage of true positive forecasts among the total number of positive forecasts. This illustrates the number of actual positive cases among all the predicted positive cases.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.13}$$

  Precision matters when the number of False Positive is essential.

- **Recall:** It determines the number of actual positive instances that are predicted accurately by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.14}$$

  Recall is very effective when the number of False Negative is high.

- **F1-score:** The harmonic average of precision and recall gives the f1-score that achieves an acceptable balance between the two units of measurement.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.15}$$

When the precision and recall needs to be balanced, the f1-score becomes very useful, especially, when the dataset is perfectly imbalanced containing a class which has significantly more number of instances than the other.

CHAPTER 4

Comparison of Results and Discussions

4.1 Table demonstrating the accuracy values of different models

Table 4.1: Overall Accuracy of different Models

Model Name	Overall Accuracy	Hyperparameter Accuracy
CNN	91%	98.83%
Random Forest	96.44%	97.94%
Decision Tree	89.30%	87.54%
KNN	95.26%	97.23%

In this study, we focus on the overall accuracy of four models as shown in table 4.1, which are CNN, Random Forest, Decision Tree, and KNN. The values include the accuracy of the model and the overall accuracy of the model while considering the specific hyperparameters.

4.2 Explaining the results obtained for machine and deep learning models

Convolutional neural network (CNN):

In terms of general response with respect to the accuracy, the CNN model attained 91%. This demonstrates that the algorithm worked very well in attaining such accuracy and is capable of recognising 91% of the instances extracted from the dataset. While tuning the hyperparameters for CNN, we used a k-fold cross-validation technique with five splits to enhance the stability and robustness of the model. The Hyperparameter tuning is done manually using the 'for' loop. After tuning the hyperparameters for the CNN model, the accuracy that is achieved is 98.83%, which is the highest accuracy among all the models in this study. This illustrates the possibility of the CNN model being improvised after optimising the hyperparameters.

Random Forest:

To extract valuable features from the dataset images, we have combined three feature extraction methods, which are HOG, LBP, and PCA. After feature extraction, the Random Forest model gives an accuracy value of about 96.44%, which is the best accuracy among the other fundamental models. This information shows that the random forest algorithm can recognise images very well. The grid search method is used for tuning the parameters with a cross-validation value of three, and the value of the accuracy after tuning the hyperparameters for the model is 97.94%, which is just a percent more than the result obtained by its basic model. This minor difference indicates that the random forest model is effectively tuned, and modifying the hyperparameters does not have much impact on the algorithm's performance.

Decision Tree:

The Decision Tree model gives an accuracy of 89.30%, which is a lower accuracy score compared to all other models. Before fitting the data into the model, we extracted features using the DenseNet-121 pre-trained model. This highlights that the Decision Tree could perform well while classifying the different classes of images, as it does not record the interactions of the data well. Tuning the parameters is done using a grid search technique with a validation value of five. The hyperparameter tuning does not much impact the Decision Tree model, as the accuracy score is 87.54%, which is lesser compared to its basic algorithm. The optimal values used during the tuning phase do not considerably differ from the default values. Compared to algorithms such as Random Forest and CNN, the accuracy of the Decision Tree is significantly less.

K-nearest Neighbors (KNN): In this study, we have used ResNet-50 pre-trained

model to extract the edges and patterns from the input images and then, fit the data to the model. The KNN model gives an accuracy of 95.26%, which shows its ability to recognise the data. The Grid Search method is used to find the best parameters for attaining high accuracy with five splits of cross-validation. After performing the hyperparameter tuning, the accuracy score improves to 97.23%, indicating that fine-tuning can enhance the model's performance. This result demonstrates the effectiveness of changing the hyperparameters for the KNN model, pointing to the ability to get a better outcome with proper parameter optimisation.

4.3 Discussing our results with the research papers referred in Literature review

Studying the analogies between the CNN-based gait identification analysis, in which we are able to obtain 98.83% accuracy, illustrates different techniques and accomplishments in the area of gait recognition. In a study, Taocheng Liu, Xiangbin Ye, and Bei Sun [12] utilised the information from gait to determine gender using a CNN-SVM model. Based on the parameters of the model and the datasets offered, the accuracy could possibly reach 85% to 95%. Considering the fact that gender identification is usually simpler to perform than person identification, the CNN model we used appears to be more sophisticated in recognizing individuals as compared to genders, considering the accuracy of 98%. Basic person identification involving the CNN and image augmentation methods that are normally employed to enhance the range of the training data by improving the robustness of the model is the primary objective of Abeer Mohsin Saleh and Talal Hamoud's study [30]. The accuracy of their paper is very high which is around 95% to 98%. The research of Zifeng Wu et al. [13] emphasises cross-view gait identification, which is more complex than single-view gait identification since there are strong variations in silhouette representation when viewed from different angles.

The Heat Gait model, which utilises GCNs for gait recognition, was developed in 2022 by Md. Bakhtiar Hasan et al [15]. It is an innovative method for dealing with structured data. Although GCN models typically generate the best results around the 90% to 95% range, the CNN algorithm that we used produces higher accuracy, about 98.83%. This demonstrates that the CNN models remain powerful, especially in controlled circumstances. Z. A. A. Alhuseen et al. [37] differentiated among the control group of unusual behaviour recognition using the analysis of gait and person identification, although maintaining similar thoughts regarding applying the CNN models. The accuracy varies between 85% to 90% in the detection of abnormalities, indicating that the CNN model is able to correctly identify individuals based on their walking patterns, although the requirements are complex. Ultimately, 98.83% obtained by the CNN architecture places it in the best position of gait recognition methods compared to other innovative methods.

It demonstrates the manner in which we obtained 96.44% accuracy using random forest model measures compared to other studies regarding feature evaluations for machine learning algorithms used in the classification of gender based on gait. The researchers in the paper, C. Vora, V. Katkar, and M. Lunagaria [16], used pre-trained models to recognise the gender and implemented methods of transfer learning to improve the outcomes with a large number of datasets involving ImageNet. The pre-trained models are highly efficient for classification problems as their average accuracy values are very high between the range of 95% to 98%. The accuracy value of the random forest model obtained from our study, which is 96.44% explains that properly tuned machine learning algorithms do not need to go through transfer learning as imagined earlier. A unique topic is highlighted by Azhin T. Sabir et al. [20], who specialize in the classification of gender through accelerometer data collected by smartphones. This is a more difficult attempt since the information from the sensors appears to be more noisy. This technique commonly produces accuracies of 80% to 90% due to numerous aspects such as speed of walking and location of sensors that may impact the results of the accelerometer. This technique significantly overtook the random forest algorithm that is used in our study due to its visual-based gait parameters, which provide additional data based on gender classification.

Similarly, a study from Kohei Arai and Rosa Andrie Asmara [17] utilised temporal and spatial reasoning to determine the gender attained a moderate level of accuracy, result around 80% to 90%. The way they approached utilised simpler feature extraction techniques compared to modern machine learning and deep learning methods. Improvements in feature engineering and machine learning models have made it easy for algorithms such as Random Forests to deal with the complex information of the gait more effectively, which enhanced the results of this study. The work of Sk Md Alfayeed and Baljit Singh Saini [10] provides an extensive overview of machine learning methods for the analysis of gait, possibly covering a wide variety of approaches from modern deep learning techniques to traditional methods such as k-nearest neighbours and SVM. The selection of the algorithm and number of datasets impact the overall 85% to 95% accuracy of scientific classifications in this area of study. The accuracy obtained, which is 96.44% by the Random Forest model in our research, puts to the highest end, which explains the tuning adjustment and implementation with proper features, and combined techniques such as Random Forest can be particularly effective for this analysis. In summary, the model used demonstrates the ability and operates around an appropriate range of values considering the similar information obtained from the early studies.

CHAPTER 5

Conclusions

Based on the evaluation of the findings, it is concluded that the denoised images used in this study from the CASIA-B dataset are extremely useful for gender recognition and have produced better results through various machine learning and deep learning techniques. The Convolutional Neural Network model with the optimisation of hyperparameters is particularly an efficient model, achieving the best accuracy score of 98.83%, according to the results of our study. This result highlights the improvement in deep learning models, especially CNN, which can be utilised to identify the complex patterns and characteristics in gait analysis. CNN models are very effective in gathering spatial data from the images. In this study, the hyperparameter tuning allowed a suitable network architecture to be achievable, which led to the generation of the best results while comparing with other machine learning models.

On the other hand, traditional machine learning models, such as Random Forest, Decision Tree, and K-nearest neighbours, produced acceptable accuracy values. The Random Forest consists of multiple decision-making strategies which were effective in minimizing the complexity of gait information but were eventually overpowered by the CNN algorithm. This illustrates that CNN methods are more suitable for image-based applications such as gait analysis. The accuracy of the CNN algorithm is better compared to the models, such as the Decision Tree, which utilises the idea of splitting the information into uniform groups of data and KNN, which performs on the concept of feature space relationship.

In general, the study demonstrates that the incorporation of deep learning models such as CNN is highly effective and can exhibit outstanding results when used with preprocessed datasets of better quality, which in turn helps to enhance the performance of recognising the individuals based on their walking pattern. The two most useful approaches, which are denoised data and the modification of Hyperparameter tuning of the CNN, helped the accuracy to increase to 98.83%. This highlights that the CNN model is the most ideal and reliable technique for the identification of individuals based on gait analysis.

CHAPTER 6

Future Steps

The improvements in gait analysis in future will concentrate on various fields that would aid in enhancing the productivity of gait technology, improve the values of precision, and adaptability of this scientific knowledge in real world applications.

6.1 Gait Analysis in real time applications

Innovative gait analysis for the identification of gender has a good future mainly through the developments in subjects pertaining to devices that are wearable and video analysis. While walking, non-invasive sensors in the shoe or bracelets could send measurable information on gait in real time. This data can be consistently transmitted to other systems for the analysis of recognising the gender with no governed surroundings or fixed cameras. In large crowds, the identification of an individual based on their walking pattern can be precisely determined by evaluating live feed from regions that are accessible to the public using a machine incorporated with Artificial Intelligence (AI).

6.2 Advancements in Deep Learning

Recent Developments in deep learning have significantly enhanced the area of gender identification using gait analysis. Early approaches involving techniques and components such as handcrafted features and the use of shallow models needed to be revised to represent the minute details that make a human gait. However, there are multiple possibilities for modelling gait data, which is technologically complicated with the appearance of more advanced and intricate designs such as transformers. By analysing the temporal and spatial patterns and interactions, in addition to the mechanism of attention, the transformers can improve our understanding and the way people walk.

6.3 Applications in Clinical and Health Industries

Gait analysis can make significant benefits to the areas of health care and therapy, significantly in recognising the differences in gender for gait, which suggests clarifying and understanding several types of different human diseases and disorders. For example, changes in the structures of musculoskeletal or walking or gait dynamics could result from factors such as gender. These differences could be examined and compared in a manner in which gender impacts the development of early diseases such as arthritis, hip dislocation and also problems with balance. In addition, it explains that by understanding the treatment setting, medical providers can better adapt physical treatment programs to the distinct gait characteristics of both genders.

CHAPTER 7

Appendix

The code file has been submitted.

Bibliography

[1] A. Ali Saleem, H. U. R. Siddiqui, R. Sehar, and S. Dudley. Gender classification based on gait analysis using ultrawide band radar augmented with artificial intelligence. *Expert Systems with Applications*, 249:123843, 2024. https://www.sciencedirect.com/science/article/pii/S0957417424007097.

[2] P. Connor and A. Ross. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding*, 167:1–27, 2018. https://www.sciencedirect.com/science/article/pii/S1077314218300079.

[3] J. Feng, J. Wick, E. Bompiani, and M. Aiona. Applications of gait analysis in pediatric orthopaedics. *Current Orthopaedic Practice*, 27(4):455–464, 2016. 10.1097/BCO.0000000000000386.

[4] D. Sethi, S. Bharti, and C. Prakash. A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work. *Artificial Intelligence in Medicine*, 129:102314, 2022. https://www.sciencedirect.com/science/article/pii/S0933365722000793.

[5] Institute of Automation, Chinese Academy of Sciences. CASIA Gait Database. http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp. Accessed: 2024-08-10.

[6] V7 Labs. CASIA-B Gait Dataset. https://www.v7labs.com/open-datasets/casia-b. Accessed: 2024-08-10.

[7] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 253–256, 2010. IEEE.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, volume 25, pages 1106–1114, 2012.

[9] M. E. Ozates, A. Yaman, F. Salami, et al. Identification and interpretation of gait analysis features and foot conditions by explainable AI. *Scientific Reports*, 14:5998, 2024. https://doi.org/10.1038/s41598-024-56656-4.

[10] S. M. Alfayeed and B. S. Saini. Human Gait Analysis Using Machine Learning: A Review. In *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pages 550–554, 2021. 10.1109/ICCIKE51210.2021.9410678.

[11] D. Zhang and Y. Wang. Using multiple views for gait-based gender classification. In *The 26th Chinese Control and Decision Conference (2014 CCDC)*, pages 2194–2197, 2014. 10.1109/CCDC.2014.6852532.

[12] T. Liu, X. Ye, and B. Sun. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition. In *2018 Chinese Automation Congress (CAC)*, pages 3477–3481, 2018. 10.1109/CAC.2018.8623118.

[13] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, 2017. 10.1109/TPAMI.2016.2545669.

[14] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using View Transformation Model based on optimized Gait Energy Image. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 1058–1064, 2009. 10.1109/ICCVW.2009.5457587.

[15] M. B. Hasan, T. Ahmed, and M. H. Kabir. HEATGait: Hop-Extracted Adjacency Technique in Graph Convolution based Gait Recognition. In *2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, pages 1–6, 2022. 10.1109/CTISC54888.2022.9849799.

[16] C. Vora, V. Katkar, and M. Lunagaria. GAIT analysis based on GENDER detection using pre-trained models and tune parameters. *Discovery Artificial Intelligence*, 4(19), 2024. Springer Nature.

[17] K. Arai and R. A. Asmara. Human gait gender classification in spatial and temporal reasoning. *International Journal of Advanced Research in Artificial Intelligence*, 1(6):1–6, 2012. Citeseer.

[18] J.-H. Yoo, D. S. Hwang, and M. S. Nixon. Gender Classification in Human Gait Using Support Vector Machine. In *Advanced Concepts for Intelligent Vision Systems (ACIVS 2005)*, volume 3708 of *Lecture Notes in Computer Science*, pages 138–145. Springer, 2005. Berlin, Heidelberg. 10.1007/11558484-14. https://link.springer.com/chapter/10.1007/11558484_14.

[19] M. Hu and Y. Wang. A New Approach for Gender Classification Based on Gait Analysis. In *2009 Fifth International Conference on Image and Graphics*, pages 869–873, 2009. 10.1109/ICIG.2009.80.

[20] A. T. Sabir, H. S. Maghdid, S. M. Asaad, M. H. Ahmed, and A. T. Asaad. Gait-based gender classification using smartphone accelerometer sensor. In *2019 5th International Conference on Frontiers of Signal Processing (ICFSP)*, pages 12–20, 2019. DOI: 10.1109/ICFSP48124.2019.8938033.

[21] J. Lu, G. Wang, and T. S. Huang. Gait-based gender classification in unconstrained environments. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3284–3287, 2012.

[22] D. Zhang and Y. Wang. Investigating the separability of features from different views for gait based gender classification. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, 2008. DOI: 10.1109/ICPR.2008.4761872.

[23] A. Gupta, A. Jadhav, S. Jadhav, and A. Thengade. Human gait analysis based on decision tree, random forest and KNN algorithms. In *Applied Computer Vision and Image Processing: Proceedings of ICCET 2020, Volume 1*, pages 283–289. Springer, 2020.

[24] R. Baker. The history of gait analysis before the advent of modern computers. *Gait & posture*, 26(3):331–342, 2007. Publisher: Elsevier.

[25] O. Fischer and W. Braune. *Der Gang des Menschen: Versuche am unbelasteten und belasteten Menschen*. Volume 1. Hirzel Verlag, 1895. Language: German.

[26] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan. Robust view transformation model for gait recognition. In *International Conference on Image Processing (ICIP)*, Brussels, Belgium, 2011.

[27] X. Xing, K. Wang, T. Yan, and Z. Lv. Complete canonical correlation analysis with application to multi-view gait recognition. *Pattern Recognition*, 50:107–117, 2016. DOI: 10.1016/j.patcog.2015.08.011.

[28] J. Upadhyay and T. Gonsalves. Robust and lightweight system for gait-based gender classification toward viewing angle variations. *AI*, 3(2):538–553, 2022. DOI: 10.3390/ai3020031.

[29] M. Benouis, S. Mohamed, R. Tlemsani, and L. Mostefai. Gait recognition based on model-based methods and deep belief networks. *International Journal of Biometrics*, 8:237, 2016. DOI: 10.1504/IJBM.2016.082598.

[30] A. M. Saleh and T. Hamoud. Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image augmentation. *Journal of Big Data*, 8(1):1, 2021. DOI: 10.1186/s40537-020-00387-6. ISSN: 2196-1115.

[31] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. GEINet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, 2016. DOI: 10.1109/ICB.2016.7550060.

[32] M. Alotaibi and A. Mahmood. Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164:103–110, 2017. DOI: https://doi.org/10.1016/j.cviu.2017.10.004. ISSN: 1077-3142.

[33] A. Goyal. Dropout in neural networks. *Towards Data Science*, 2018. URL: https://towardsdatascience.com/dropout-in-neural-networks-47a162d621d9.

[34] What is fully connected layer in deep learning? 2021. URL: https://www.geeksforgeeks.org/what-is-fully-connected-layer-in-deep-learning/. Accessed: 2024-08-15.

[35] A. Parashar, A. Parashar, W. Ding, M. Shabaz, and I. Rida. Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. *Pattern Recognition Letters*, 172:65–73, 2023. DOI: https://doi.org/10.1016/j.patrec.2023.05.021. ISSN: 0167-8655.

[36] Z. Liu, G. Yang, and G. Feng. Gait recognition based on gabor wavelet and local binary pattern. *Zhongshan Daxue Xuebao/Acta Scientiarum Natralium Universitatis Sunyatseni*, 53:1–7, 2014.

[37] Z. A. A. Alhuseen, F. A. Joda, and M. A. Naser. Abnormal behavior detection in gait analysis using convolutional neural networks. *IngÃ©nierie des SystÃ¨mes d Information*, 28(5):1151–1159, 2023. DOI: 10.18280/isi.280504.

[38] D. Slijepcevic, F. Horst, S. Lapuschkin, B. Horsak, A. Raberger, A. Kranzl, W. Samek, C. Breiteneder, W. I. SchÃ¶llhorn, and M. Zeppelzauer. Explaining machine learning models for clinical gait analysis. *ACM Trans. Comput. Healthcare*, 3(2):14, 2021. DOI: 10.1145/3474121. Publisher: Association for Computing Machinery, New York, NY, USA.

[39] S. Alfayeed and B. Saini. Human gait analysis using machine learning: A review. In *2021 3rd International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 550–554, 2021. DOI: 10.1109/ICCIKE51210.2021.9410678.

[40] K. Sugandhi, F. F. Wahid, and G. Raju. Feature extraction methods for human gait recognition A survey. In *Applied Computer Vision and Image Processing: Proceedings*

of ICCET 2020, Volume 1, pages 377–385. Springer, Singapore, 2016. DOI: 10.1007/978-981-10-5427-3_40.

[41] M. S. Joul and G. M. Sheeba. Gait based authentication with HOG feature extraction. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pages 1478–1483, 2018. DOI: 10.1109/ICIC

[42] T. Ojala, M. PietikÃ¤inen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996. DOI: 10.1016/0031-3203(95)00067-4.

[43] K. K. Verma, B. M. Singh, and A. Dixit. A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. *International Journal of Information Technology*, 14:397–410, 2019. DOI: 10.1007/s41870-019-00364-0.

[44] M. H. Ali. Gait and Locomotion Analysis for Tribological Applications. PhD thesis, ProQuest LLC, East Eisenhower Parkway, 2013.

[45] N. Safuan, M. Ismail, and N. Jamil. Feature extraction technique for human gait video analysis. *Journal of Engineering and Applied Sciences*, 12:534–541, 2017. DOI: 10.3923/jeasci.2017.534.541.

[46] J. Wu, J. Wang, and L. Liu. Feature extraction via KPCA for classification of gait patterns. *Human Movement Science*, 26(3):393–411, 2007. DOI: https://doi.org/10.1016/j.humov.2007.01.015. URL: https://www.sciencedirect.com/science/article/pii/S0167945707000139.

[47] G. Luo, Y. Zhu, R. Wang, Y. Tong, W. Lu, and H. Wang. Random forest-based classification and analysis of hemiplegia gait using low-cost depth cameras. *Medical & Biological Engineering & Computing*, 58(2):373–382, 2020. DOI: 10.1007/s11517-019-02079-7. URL: https://doi.org/10.1007/s11517-019-02079-7.

[48] Predict. Random Forest Intuition: Understanding the Algorithm with Python. 2020. URL: https://medium.com/predict/

random-forest-intuition-understanding-the-algorithm-with-python-b127 ~:text=Random%20forest%20is%20an%20ensemble%20learning%20algorithm%20that,averaging%20the%20results%20to%20make%20the%20final%20prediction. Accessed: 2024-08-23.

[49] K. Nishimura, A. Ishida, J. Murakami, and N. Yamamoto. Feature extraction of gait data via PCA and personal identification using neural networks. In *2024 9th International Conference on Big Data Analytics (ICBDA)*, pages 258–263, 2024. DOI: 10.1109/ICBDA61153.2024.10607330.

[50] JavaTpoint. Machine Learning Decision Tree Classification Algorithm. 2024. URL: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm. Accessed: 2024-08-23.

[51] GeeksforGeeks. DenseNet Explained. 2024. URL: https://www.geeksforgeeks.org/densenet-explained/. Accessed: 2024-08-23.

[52] Q. Ji, J. Huang, W. He, and Y. Sun. Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images. *Algorithms*, 12:51, 2019. DOI: 10.3390/a12030051.

[53] GeeksforGeeks. k-Nearest Neighbours. 2024. URL: https://www.geeksforgeeks.org/k-nearest-neighbours/. Accessed: 2024-08-23.

[54] B. Sharma. ML from Scratch: k-Nearest Neighbors Classifier. 2019. URL: https://towardsdatascience.com/ml-from-scratch-k-nearest-neighbors-classifier-3fc51438346b. Accessed: 2024-08-23.

[55] Roboflow. What is ResNet-50?. 2024. URL: https://blog.roboflow.com/what-is-resnet-50/. Accessed: 2024-08-23.

[56] N. Kundu. Exploring ResNet50: An In-Depth Look at the Model Architecture and Code Implementation. 2023. URL: https://medium.com/@nitishkundu1993/

exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-co
Accessed: 2024-08-23.

[57] ProgrammingR. KNN Hyperparameters: A Friendly Guide to Optimization. 2024. URL: https://www.programmingr.com/knn-hyperparameters-a-friendly-guide-to-optimization/. Accessed: 2024-08-23.

[58] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. In *IEEE Transactions on Image Processing*, volume 18, pages 1905–1910, August 2009. DOI: 10.1109/TIP.2009.2020535.

[59] Author from Towards Data Science. Machine Learning Part 20: Dropout Keras Layers Explained. In *Towards Data Science*, 2024. Accessed: 2024-09-12. Available at: https://towardsdatascience.com/machine-learning-part-20-dropout-keras-layers-explained-8c9f6dc4c9ab