Binary Gender Classification of African Fingerprints using CNN

MSc Research Project
Data Analytics

John Maruthukunnel Jacob
Student ID: 21138494

School of Computing
National College of Ireland

Supervisor: Dr Cristina Muntean

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	John Maruthukunnel Jacob
Student ID:	21138494
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Dr Cristina Muntean
Submission Due Date:	15/12/2022
Project Title:	Binary Gender Classification of African Fingerprints using CNN
Word Count:	XXX
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Binary Gender Classification of African Fingerprints using CNN

John Maruthukunnel Jacob
21138494

Abstract

Fingerprints, one of the most popular biometric authenticators, can distinguish between genders. The primary difficulty in fingerprint classification is time-effective models. In this paper, the author compares the binary gender classification performance of VGG-19, VGG-16, InceptionV3, and ResNet-50 for fingerprints. Traditional deep networks perform poorly and take longer to train because of their narrow depth range. In addition, they have to deal with model overfitting because there is a lack of fingerprint data. To solve the issues with insufficient fingerprint data, data augmentation techniques rotate, zoom, and flip is implemented. Transfer learning is used to pre-train the four CNNs to accelerate model training. Evaluation of the proposed model is done by testing, training accuracy, and loss. The VGG-19 model performed the best with a testing accuracy of 71.9% along with VGG-16 with 72.3% in comparison to InceptionV3 with 67.3%, and ResNet-50 with the lowest accuracy of 60.8%.

1 Introduction

Since more than a century ago, fingerprint identification has been a regular practice. Fingerprints have a number of distinguishing features that make it easy to recognize them, classify them, and store them for later use. Since fingerprints contain so many ridges and furrows, they constitute a distinguishing mark. It is now established that everyone has a distinct fingerprint. Ridge width and interridge spacing are two factors that affect epidermal ridge density. Before birth, the human fetus begins to form ridge patterns, which remain the same for the rest of a person's life. It is possible to identify not just criminals but also amnesiacs and unidentified deceased when there is proof of fingerprints.

Fingerprints are one of the most crucial forensic tools for analyzing criminal evidence because of their uniqueness and endurance. Evidence such as DNA, hair, fibers, and footprints can all be highly useful when reviewing a crime scene. One of these is the fingerprint, which has been utilized in criminal investigations during the past century to identify individuals, including witnesses, victims, and suspects. The distinctness of the fingerprint characteristics can speed up suspect identification and assist with gender identification problems. This expedites the identification of unidentified suspects since these particular fingerprint traits can be used to differentiate people depending on gender. A person can be recognized by a variety of traits, such as their body type, voice, stature, and face. The basic traits that set people apart include their gender.

Additionally, several studies have shown that a person's fingerprints can reveal their gender, which is useful for reducing the list of suspects. Comparing a suspect's prints with the countless other probable matches in the fingerprint databases can also aid forensic investigators in precisely identifying a suspect. It takes a lot of time during an investigation to search through numerous fingerprint databases to uncover and identify unfamiliar fingerprints. Finding the offender in a huge fingerprint database will take much less time if the gender, hands, or fingers are known. The investigators will often focus more tightly and advance the investigation more quickly if the gender of a suspect can be established at the outset of a forensic probe. Any government's crime investigation divisions can make use of fingerprint classification to efficiently and quickly solve crimes. Such technology will aid in gender detection and differentiation at border controls. This is because some people may impersonate another gender to avoid detection. In addition to physical examination, the issue of gender-based criminal impersonation calls for the use of biometric fingerprints. Despite the security measures put in place, technically proficient criminals may be able to fake credentials such as foreign passports, certifications, and driver's licenses. Counterfeiting risk will be reduced by biometric characteristics like fingerprints.

In the previous years of research, fingerprint classification into males and females has shown promising results. The machine learning models to classify fingerprint ridges employed by (Ceyhan and Sagiroglu; 2015) have achieved accuracies over 90%. The classification of fingerprints using CNN models has been carried out in the past 5 years. The ResNet-34 model used to classify the SOCOFing dataset achieved an accuracy of 75.2% (Shehu et al.; 2019). The best models achieved from previous work are VGG-16, InceptionV3, and ResNet-50 with accuracies of 79.2%, 77.1%, and 79.1% respectively (Hsiao et al.; 2022). This research aims to compare a new model VGG-19 with the previously mentioned models.

1.1 Research Question and Objectives

How do the three CNN algorithms ResNet-50, VGG-16, and InceptionV3 compare with the novel architecture VGG-19 in classifying fingerprints of the African population into male and female classes?

The primary goal of this proposal is to compare a new model VGG-19 with ResNet-50, VGG16, and InceptionV3 in classifying male and female classes in the SOCOFing dataset. The secondary objective is to investigate how transfer learning and data augmentation will improve the accuracy and training times of these models.

1.2 Plan of Paper

Section 1 discusses the background and motivation of the Research topic selected for this project; Section 2 discusses the relevant literature that inspired this research; Section 3 describes the research methodology chosen and the stages involved; Section 4 outlines the design specification of the research. Section 5 discusses the implementation of the four models selected. Section 6 outlines the evaluation and inference of the models. Section 7 illustrates the discussion and Section 8 discusses the conclusion.

2 Related Work

Numerous studies on gender determination have been undertaken to date and are based on the three-dimensional domain examination of ridges. Previous studies have employed the breadth, finger imprint patterns, and thickness-to-valley ratios as ridge-associated criteria for gender identification (Shehu et al.; 2019). Females have a thicker epidermal ridge because they have better and more clear ridge features than males, according to multiple experimental findings published in studies done on the African population (Ahmed and Osman; 2016). Women have significantly finer ridges than men do because of their significantly higher ridge density, according to the Spanish Caucasian population sample (Gutiérrez-Redomero et al.; 2008). It has been shown that there are topological differences in ridge thickness. In the distal area of the fingers, these variations between males and females are statistically significant, but not in the proximal region, according to (Gutiérrez-Redomero et al.; 2008). The thumb and index finger exhibit lower ridge density than the other three fingers, according to research on differences in ridge densities between the fingers (Gutiérrez-Redomero et al.; 2008).

At maturity, there are still ridge density disparities between the sexes, and they invariably manifest as females having thinner ridges than males of the same age bracket. The finger breadth (but not the proximal area) of adults' radial and ulnar epidermal ridges tends to increase with age and hand size (Sánchez-Andrés et al.; 2018). The authors of the study (Nayak et al.; 2010), claim that among the Chinese population, fingerprints with a mean ridge density of 12 or less/25 mm2 are more likely to be of male origin. Those who have a mean ridge density of 13 or more/25 mm2 on the other hand are more likely to be female in origin. When compared to fingerprints having a mean ridge density of 11 ridges/25 mm2 or less, fingerprints with a mean ridge density of more than 12 ridges/25 mm2 are more likely to be produced by women in the Malaysian population (Nayak et al.; 2010). The study supports earlier findings and demonstrates gender variations in fingerprint ridge density.

2.1 Data Augmentation and Transfer Learning

Conventional data augmentation involves applying routine geometric changes to image input, such as contrast adjustment, scaling, rotation, and translation (Chen and Cao; 2019). Recently, the use of GANs (Generative Adversarial Networks) for data augmentation has become more widespread (Chen and Cao; 2019). The model used in this research paper consists of two networks, one of which continuously distinguishes between real and false images, and the other of which generates false images (Chen and Cao; 2019). When there is a lack of data, CNN always uses common data augmentation techniques including rotation, cropping, and flipping. However, these approaches have limitations and may even degrade classification performance in other circumstances, such as small images and scattered object characteristics (Zhang et al.; 2020). To address this problem, the model devised a data augmentation method called circular shift that provides variations for CNN-based models without jeopardizing too much data (Zhang et al.; 2020). Three popular picture datasets are utilized to evaluate the suggested procedure, and the results of the experiment show consistent improvement on different CNN-based classifiers (Zhang et al.; 2020).

A new effective learning model may require several hours of training, plenty of data, and powerful processors. Real-world applications require time and money to gather and analyze enormous volumes of domain-specific data, and deep learning model implementation is highly difficult. The researchers are certain that prior object knowledge helps to overcome this challenge because of their similarity and significance to new things. Some research suggests that deep learning models that have been pre-trained for a classification problem can be effective for classification problems (Al-Qerem et al.; 2021). Transfer learning is the process by which CNN models that have been trained on a certain dataset can be modified for a new task applicable to a different domain. Transfer learning methods considerably perform better in classification-based data augmentation when compared to Generative Adversarial Networks (GANs) when used on the same dataset (Al-Qerem et al.; 2021). Classification Based GAN Augmentation (CBGA), three transfer learning methods (TL-VGG, TL-INC, and Resnet 50), and a dataset of medical pictures were utilized to test the effectiveness of the classifiers (Al-Qerem et al.; 2021). The study shows that for data augmentation in classification performance, all transfer learning strategies consistently beat CNNs built on generative adversarial networks (Al-Qerem et al.; 2021).

2.2 Fingerprint Analysis using Deep Learning Models

In this work (Alam et al.; 2019), the authors classified the gender of fingerprints using discrete wavelet transforms (DWT) and singular value decomposition (SVD) methods using Euclidean distance and the KNN classifier for classification. The model classifier achieved 91.25% accuracy for men, compared to 88.96% for women (Alam et al.; 2019). It was found that small fingers have higher accuracy for finger-by-finger classification. The most important characteristics, such as RTVTR and Ridge density, were taken out of the database and used by the authors(Arun and Sarath; 2011) to identify whether a given fingerprint belongs to a male or female. Using an SVM classifier trained 150 male images and 125 female images, they were able to produce a classification function feature vector patterns belonging to each class(Arun and Sarath; 2011). These features are then used to train the classifier. This paper (Lee et al.; 2020) uses two well-known machine learning techniques, namely Naïve Bayes (NB) and Classification and Regression Trees (CART) algorithms, to detect gender based on ridge counts. The bootstrapping without replacement method has been used to assess the performance of the prediction models with ethnicity and finger digit. The ethnicity-specific models of Indian and Malaysian individuals performed more than the global model(Lee et al.; 2020).

To classify the ridges in the 25 mm2 section of a fingerprint, the authors of the study (Ceyhan and Sagiroglu; 2015) employed Naive Bayes (NB), k-Nearest Neighbors (kNN), Decision Tree (DT), and Support Vector Machine (SVM) classification algorithms. The success rate was found to be 95.3% with the Naive Bayes algorithm, 94.3% with the decision tree algorithm, and 93.8% with the support vector machine algorithm with training data comprising 66% of the initial dataset and 34% as testing data. The database contains information from 600 Turks (300 males and 300 women)(Ceyhan and Sagiroglu; 2015). After the second testing procedure (using the 10-fold cross-validation technique), the Naive Bayes method, decision tree algorithm, and support vector machine algorithm all had success rates of 94.2% (Ceyhan and Sagiroglu; 2015).

Since convolutional neural networks display accuracy close to 100% at a large learning cost, even with high-performance computation, they have lately been employed as a viable substitute. CNN-based models have consistently excelled at image classification tasks. As demonstrated by an earlier study, CNNs may be used for fingerprint categorization with outstanding results (Jian et al.; 2020). By using comprehensive fingerprint classification with CNNs, classifications of gender, hand, and fingers will be more precise. Hard-coded fingerprint characteristics were once used to categorize fingerprints by gender. With a CNN, the model is improved since it takes into account all of the fingerprint data and is trained on the traits to distinguish between genders. (Shehu et al.; 2019) proposed CNN to identify the gender from fingerprint images; specifically, the ResNet-34 and ResNet-18 CNN models were trained to make use of the transfer learning technique. 1,230 images of both male and female fingerprints from the Sokoto Coventry Fingerprint Dataset used as the foundation for the base learning of the ResNet model (Shehu et al.; 2019). The classes for men and women were then changed. The method produced accuracy rates of 73.98%, 76.42%, and 75.20% for male, female, and overall individuals, respectively (Shehu et al.; 2019). The study (Hsiao et al.; 2022) used three commonly used CNNs to conduct biometric prediction on 1000 samples (VGG16, Inception-v3, and Resnet50). The outcomes showed that employing the ring finger of male participants, VGG16 had the highest accuracy in classifying gender (79.2%), left- and right-hand fingerprints (94.4%), finger position (84.8%), and height range (69.8%). (Hsiao et al.; 2022). In this study (Abuared et al.; 2020), the CNN model based on VGG19 and Transfer Learning is used to classify two cancer types and one non-cancer type from the HAM10000 dataset. Calculating the network's accuracy and loss explains, tests, and evaluates the training process. After training, 600 images were used to assess the network's accuracy and loss. Both training and testing accuracy were 0.985 (Abuared et al.; 2020). Training and testing losses were 0.100 and 0.119. Table I presents training and validation outcomes. The discrepancy between training and testing outputs isn't large, indicating the network isn't overfitting. In terms of classification accuracy and speed, the CNN model outperformed experts. Additionally, it was a useful guide resource for fingerprints that were difficult to manually recognize (Hsiao et al.; 2022).

2.3 Literature Review Summary

The research and studies covered in the literature review served as inspiration for this study. Examples of data augmentation and transfer learning approaches were provided in Section 3.1. It was discovered that data augmentation might be used to get better performance in the case of small datasets. The transfer learning method shows the capacity to train models more quickly and generate better results when faced with time restrictions. As described in section 3.2, fingerprint classification results using machine learning techniques have been significant. Even though it is still in its early phases, research into deep learning methods for fingerprint-based gender classification has produced intriguing findings. The VGG-19 model has shown significant accuracy for classification problems. However, because of a lack of data augmentation and insufficient data, researchers were unable to fully evaluate their potential. From the literature review done for this research project, it was observed that the fingerprint ridge density for male and females from different ethnicities varied over marginal differences. The ridge densities of males of African and Chinese males varies slightly when comparing the fingerprints. This shows that the African origin matters for the classification done in this research. The overall pattern observed was that the females have a greater number of ridges than males.

3 Methodology

The modified Knowledge Discovery in Databases (KDD) process is used in this study to accomplish its objectives. The main stages involved in classifying the fingerprints as male and female are shown below in Figure 1.

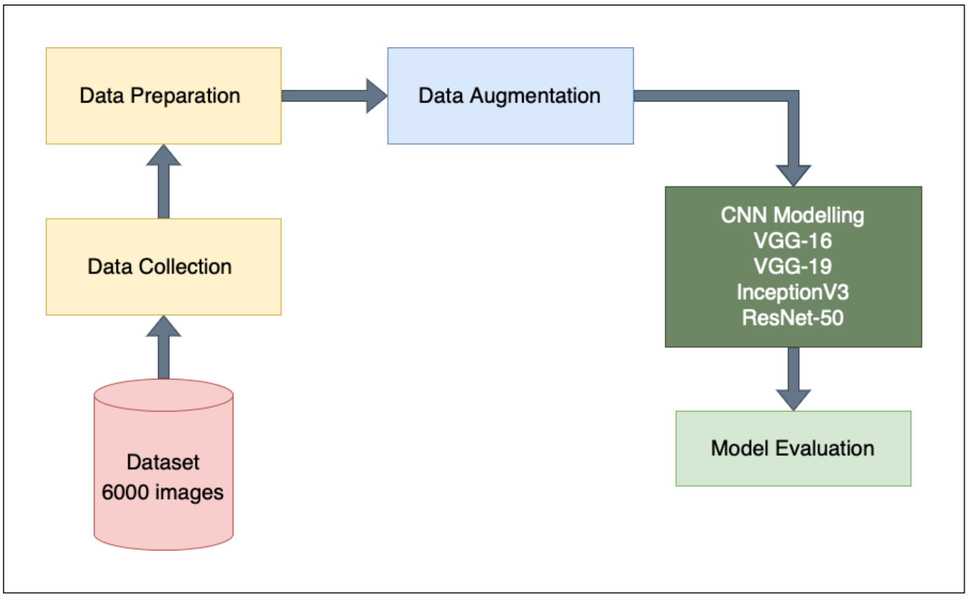

Figure 1: Research methodology

3.1 Data Collection

The Sokoto Coventry Fingerprint Dataset (SOCOFing), a biometric fingerprint database developed with academic research in mind, is used in this research. The SOCOFing database contains 6,000 fingerprint images from 600 people in Africa, together with labels for the individuals' genders, hand names, and finger names saved in BMP format (.BMP). The folder named 'Real' is only used for this research. The dataset is accessible to the public on Kaggle. Figure 2 shows an example of male and female fingerprint images stored in the dataset.

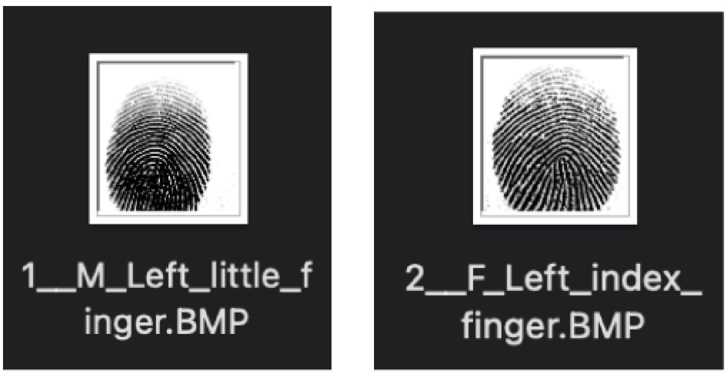

Figure 2: Male and Female fingeprint image file

3.2 Data Preparation

This stage is essential for giving the data the correct organization, precision, and format for the model. Its label is indicated by the image filename and folder name saved in the file system. The dataset is downloaded from Kaggle and saved as Real. The photos in the dataset are all stored as BMP files. All of the photos used in this study were converted to JPEG files (.jpg), which is more effective for training models. The images are found to be 96*103 pixels in dimension, which is further converted to 96*96 as the model requires square images. The initial dataset is split into two classes male and female using the labels of the images as shown in Figure 2.

3.3 Exploratory Data Analysis

The dataset used in this study contains fingerprints belonging to two classes, namely male and female. The bar plots for the dataset are done to identify the total number of images belonging to each class. It was found that the male class contains 4770 fingerprint images and the female class contains 1230 fingerprint images. From the bar plots shown in Figure 3, we can see that the dataset is imbalanced. To balance the dataset data augmentation is employed which is explained in the next subsection.

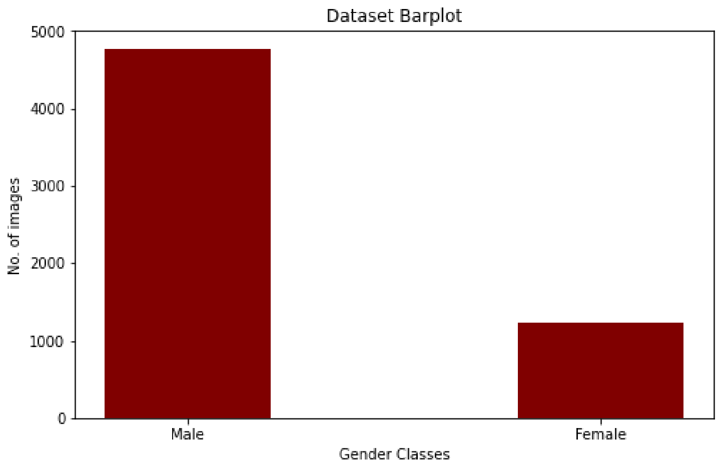

Figure 3: Barplot of dataset

3.4 Data Augmentation

From the Figure 3, it is observable the dataset is imbalanced and may result in overfitting in modeling. 2000 male fingerprints are randomly selected into the final dataset used for modeling. All of the female fingerprints available are selected. Data augmentations techniques such as rotate, skew, flip, shear, and zoom are used to increase the class sizes to 2000 each respectively. The augmentation is done using the Augmentor python package. The images in the final dataset are converted into 96*96 dimensions for modeling. This image size was chosen to not lose image quality.

3.5 Modelling

In this stage, the selected models ResNet-50, InceptionV3, VGG-19, and VGG-16 are performing feature extraction and classification on the fingerprint dataset.

3.5.1 VGG-19

VGG19 is an improved version of VGG16. VGG19 is a deep CNN with convolutional and max pooling feature extractor layers. Classifier follows these levels. CNN's architecture determines the size and number of convolutional and fully linked layers. The 19 layers of the VGG19 model are used to classify images. The VGG19 model implements several 3-by-3 filters per layer in place of the huge 11*11 filters used in Alexnet. Three fully connected layers, five pooling layers, sixteen convolution layers, and one softmax layer make up the VGG 19 model. In both variants of VGGNet, two completely connected layers with a total of 4096 channels each are utilized, followed by a second fully connected layer with a total of 1000 channels to predict a total of 1000 labels, and lastly a softmax layer for classification in the final fully connected layer. In this research sigmoid function is used in place of softmax to gain better accuracy.

3.5.2 VGG-16

The vgg16 network's convolution layers all have the same configuration, which is a convolution core size of 3X3, a step size of 1, and a maximum of five pooling layers, each of which is a 2x2 layer with a step size of 2. Three full connection layers are included, the first two of which have 4096 channels each, and the third of which has 1000 channels and 1000 label categories. The softmax layer is the bottom layer. After all hidden layers, the ReLu nonlinear activation function is used. In this research sigmoid function is used instead of softmax to improve accuracy.

3.5.3 InceptionV3

Inception refutes the notion that adding more convolutional layers improves neural networks. Inception uses dense components to approximate sparse nodes. Inception is distinct from prior convolutional neural networks since the network builder doesn't need to organize the convolutional and pooling layers. The network model picks and applies convolutional layer filters. Insert all possible values. The network chooses learning parameters and model filter combinations when all outputs are connected. The InceptionV3 network model employs three inception modules to extract visual features on various scales. Convolution kernels reduce dimensionality. 1x1 convolution compresses the input layer into a bottleneck intermediate layer. It minimizes presentation layer size without influencing network or computer costs. Inception decomposes a filter's convolution to reduce calculation. The 5x5 convolution is broken into two 3x3 convolutions with the same receptive field but better representation. Decomposing into two convolutions increases linearity. In this research, the inceptionv3 uses the sigmoid function as the bottom layer to increase accuracy.

3.5.4 ResNet-50

The ResNet-50 base model is a member of a family of models, where "50" denotes the number of parameter layers in the network design. There are different ResNet topologies with different parameter levels, including 18 layers, 34 layers, 101 layers, and 152 layers. The foundation of the residual network is a simple network with intermediary shortcut links. In place of the 2-layer block in the 34-layer net, the ResNet-50 architecture uses a 3-layer bottleneck block. In this research, the model uses the sigmoid function as the last layer to improve accuracy in comparison to the softmax layer.

3.6 Evaluation

In this stage, the model is evaluated using performance metrics such as training accuracy and loss, and testing accuracy and loss. The testing and training accuracy graphs are plotted to evaluate the accuracy trends of the loss function in equal intervals of epochs. The model loss graphs for training and testing are also plotted to infer the trends. The model performance increases with increasing test accuracy and decreasing loss.

4 Design Specification

The deep learning models ResNet-50, VGG-16, and InceptionV3 are implemented using colab notebook to do feature extraction and classification. After data pre-processing the data is uploaded to google colab using google drive. Pre-trained models of the above-mentioned CNNs are used for classification. The pre-trained models are used to reduce the model training time. The design of the model is shown in Figure 4. The dataset is initially collected and EDA is performed on it. Data augmentation is performed on the male and female classes if the data is imbalanced. After data augmentation the data transformation includes image resize and format conversion. Transfer learning is used to pre-train the models vgg16, inceptionv3, and resnet-50 in the 'Imagenet' dataset. These models are then used to classify the data that was split into training and testing data in an 80:20 ratio. The models are then evaluated using the accuracy and loss function. In the final stage, the models implemented are compared with each other and models from previous literature work.

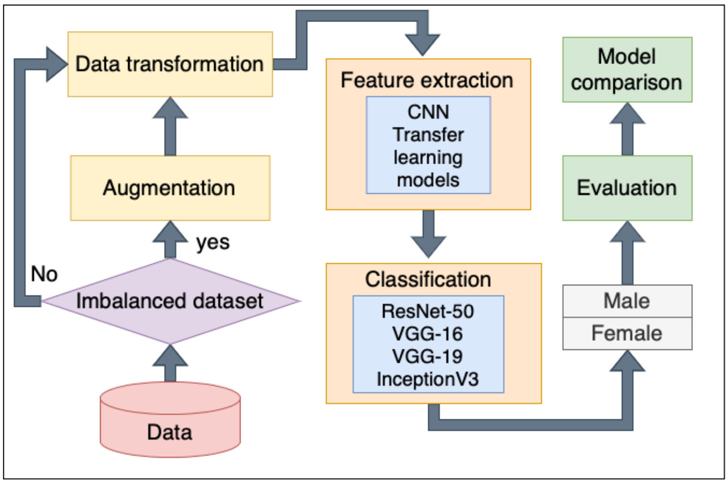

Figure 4: Model design

5 Implementation

5.1 Environmental Setup

The author's system is a MacBook air m1 with an 8-core CPU (central processing unit) with 4 performance cores and 4 efficiency cores and an 8-core integrated GPU (graphics processing unit). The data pre-processing, exploratory data analysis, and data augmentation is done in google colab pro. The augmented dataset is zipped and copied to google drive and mounted in python over google colab. Models are created using Keras API imported from TensorFlow. All the models are optimized using adam optimizer. The loss function is defined as categorical cross-entropy. The sigmoid layer is used as the last classification layer.

5.2 Implementation of VGG-16

The pre-trained VGG-16 model on the 'imagenet' dataset is used for feature extraction of the model. The classifier's last fully linked layers are not included. This is done so that we may extend the VGG-16 model with our own fully connected layers for our task-specific categorization. Setting trainable to "False" freezes the model's weights. This prevents any modifications to the previously trained weights during training. The pre-trained model is downloaded and the last layer is frozen to perform training using the selected dataset. This pre-trained model is used for model creation. Different parameters were tuned to produce the best results and the final values are loss='categorical_crossentropy', adam optimizer, dropout value of 0.2, 512 feature selection in a dense layer, and, sigmoid activation as the final layer. Two fully connected layers are used for this model. The model is trained using the training dataset which contains 3200 images. The model history is created using a fit generator with batch size = 512 and 50 epochs. The model test and training loss and accuracy are produced for each epoch in training. The model training took 8.6 minutes to run a total of 50 epochs.

5.3 Implementation of VGG-19

The pre-trained VGG-19 model on the 'imagenet' dataset is used for feature extraction of the model. The classifier's last fully linked layers are not included. This is done so that we may extend the VGG-19 model with our own fully connected layers for our task-specific categorization. Setting trainable to "False" freezes the model's weights. This prevents any modifications to the previously trained weights during training. In this research, two fully connected layers are employed in the VGG-19 model. This pre-trained model is used for model creation. Different parameters were tuned to produce the best results and the final values are loss='categorical_crossentropy', adam optimizer, dropout value of 0.5, 512 feature selection in a dense layer, and, sigmoid activation as the final layer. The model is trained using the training dataset which contains 3200 images. The model history is created using a fit generator with batch size = 512 and 50 epochs. The model test and training loss and accuracy are produced for each epoch in training. The model training took 8.4 minutes to run a total of 50 epochs.

5.4 Implementation of InceptionV3

The pre-trained InceptionV3 model on the 'imagenet' dataset is used for feature extraction of the model. The classifier's last fully linked layers are not included. This is done so that we may extend the model with our own fully connected layers for our task-specific categorization. Setting trainable to "False" freezes the model's weights. This prevents any modifications to the previously trained weights during training. The architecture in this study uses two fully connected layers. The last layer of the model that was previously trained using the 'imagenet' dataset is frozen, and the model is then trained using the selected dataset. This pre-trained model is used in the model's development. The model is trained using a training dataset made up of 3200 images. Different parameters were tuned to produce the best results and the final values are loss='categorical_crossentropy', adam optimizer, dropout value of 0.2, 512 feature selection in a dense layer, and, sigmoid activation as the final layer. The model history is created using a fit generator with a batch size of 512 and 50 epochs. The model test, training loss, and accuracy are generated for each training period. The model training took 50 minutes to run a total of 50 epochs.

5.5 Implementation of ResNet-50

The pre-trained ResNet-50 model on the 'imagenet' dataset is used for feature extraction of the model. The classifier's last fully linked layers are not included. This is done so that we may extend the ResNet-50 model with our own fully connected layers for our task-specific categorization. Setting trainable to "False" freezes the model's weights. This prevents any modifications to the previously trained weights during training. Two fully connected layers are used in the architecture of this study. The final layer of the pre-trained model using the 'imagenet' dataset is frozen, and training is then carried out using the chosen dataset. The development of the model uses this pre-trained model. The training dataset, which contains 3200 images, is used to train the model. Different parameters were tuned to produce the best results and the final values are loss='categorical_crossentropy', adam optimizer, dropout value of 0.2, 512 feature selection in a dense layer, and, sigmoid activation as the final layer. Using a fit generator with a batch size of 512 and 50 epochs, the model history is produced. For each training epoch, the model test and training loss and accuracy are generated. The model training took 7.7 minutes to run a total of 50 epochs.

6 Evaluation

In this stage the four models VGG-16, VGG-19, InceptionV3, and ResNet-50 are evaluated using testing accuracy and loss. The training accuracy and loss of the models are also plotted to infer the trends of the values.

6.1 Experiment 1: Evaluation of VGG-16

The model achieved a training accuracy of 78% and a testing accuracy of 72.3%. The training loss was 0.45 and the testing loss was 0.52 in comparison. From the graph, we can see that the training and testing accuracy improves hand in hand to 10 epochs. After 10 epochs the testing accuracy averages around 70% and slowly increases. The training and testing loss decreases rapidly for the first 5 epochs and steadily decreases to their final values after 50 epochs. We can observe that the training and testing loss decreases with values close to each other. The model accuracy and loss graphs are shown in Figure 6.

```
7/7 [==============================] - 8s 1s/step - loss: 0.4548 - accuracy: 0.7791
test_loss, test accuracy [0.45477351546287537, 0.7790625095367432]

2/2 [==============================] - 1s 265ms/step - loss: 0.5287 - accuracy: 0.7237
test_loss, test accuracy [0.5286616683006287, 0.7237499952316284]
```

Figure 5: VGG-16 training and testing accuracy and loss

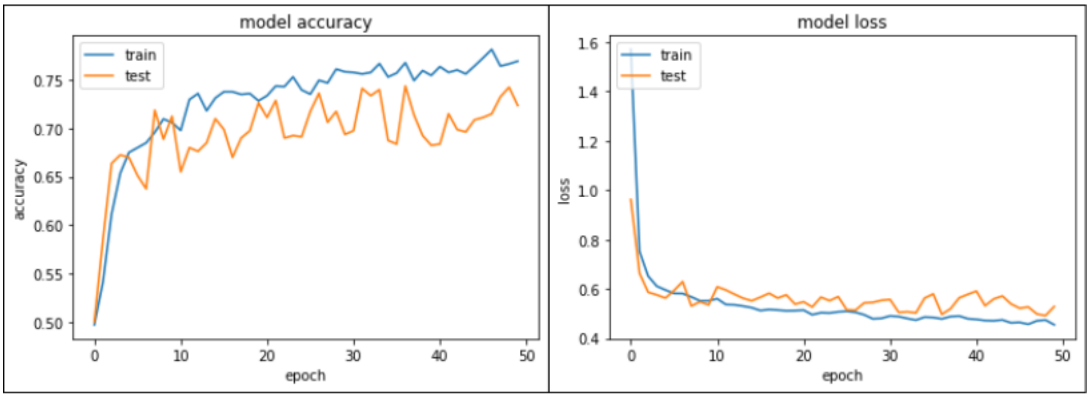

Figure 6: VGG-16 accuracy and loss graph

6.2 Experiment 2: Evaluation of VGG-19

The model achieved a training accuracy of 74% and a testing accuracy of 71.9%. The training loss was 0.51 and the testing loss was 0.52 in comparison. From the graph, we can see that the training and testing accuracy improves hand in hand to 10 epochs. After 10 epochs the testing accuracy averages below the training accuracy and slowly increases. The training and testing loss decreases rapidly for the first 5 epochs and steadily decreases to their final values after 50 epochs. We can observe that the training and testing loss decreases to values similar. The model accuracy and loss graphs are shown in Figure 8.

```
7/7 [==============================] - 9s 1s/step - loss: 0.5118 - accuracy: 0.7400
training loss, training accuracy [0.5117575526237488, 0.7400000095367432]

2/2 [==============================] - 1s 372ms/step - loss: 0.5285 - accuracy: 0.7188
test_loss, test accuracy [0.5285354852676392, 0.71875]
```

Figure 7: VGG-16 training and testing accuracy and loss

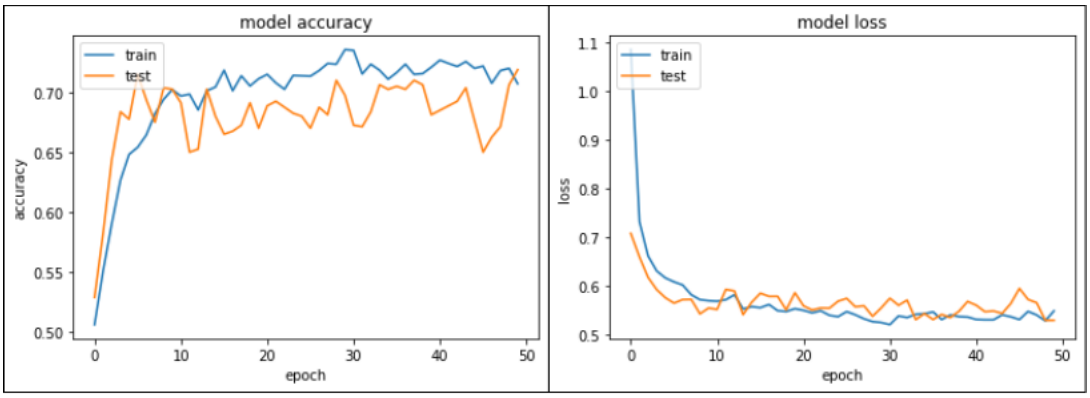

Figure 8: VGG-16 accuracy and loss graph

6.3 Experiment 3: Evaluation of InceptionV3

The model achieved a training accuracy of 71.7% and a testing accuracy of 67.4%. The training loss was 0.53 and the testing loss was 0.58 in comparison. From the graph, we can see that the testing accuracy goes above training accuracy till 10 epochs. After 10 epochs the testing accuracy averages below the training accuracy and slowly increases. The training and testing loss decreases rapidly for the first 5 epochs and steadily decreases hand in hand with their final values after 50 epochs. The model accuracy and loss graphs are shown in Figure 10.

```
7/7 [==============================] - 52s 7s/step - loss: 0.5386 - accuracy: 0.7169
training loss, training accuracy [0.5386039614677429, 0.7168750166893005]

2/2 [==============================] - 11s 4s/step - loss: 0.5858 - accuracy: 0.6737
test_loss, test accuracy [0.5857769250869751, 0.6737499833106995]
```

Figure 9: InceptionV3 training and testing accuracy and loss

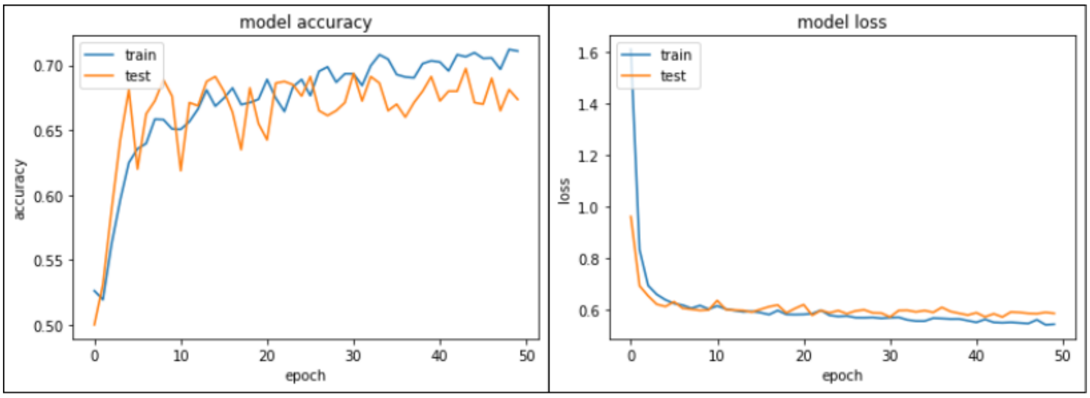

Figure 10: InceptionV3 accuracy and loss graph

6.4 Experiment 4: Evaluation of ResNet-50

The model achieved a training accuracy of 68.3% and a testing accuracy of 60.8%. The training loss was 0.57 and the testing loss was 0.7 in comparison. From the graph, we can see that the testing accuracy increases below training accuracy till 10 epochs. After 10 epochs the testing accuracy averages below the training accuracy. The training loss decreases rapidly for the first 5 epochs and testing loss averages around 0.7 for 50 epochs. The model accuracy and loss graphs are shown in Figure 12.

```
7/7 [==============================] - 8s 1s/step - loss: 0.5779 - accuracy: 0.6837
train loss, train accuracy [0.5778746604919434, 0.6837499737739563]

2/2 [==============================] - 1s 186ms/step - loss: 0.7030 - accuracy: 0.6087
test loss, test accuracy [0.7029893398284912, 0.6087499856948853]
```

Figure 11: ResNet-50 training and testing accuracy and loss

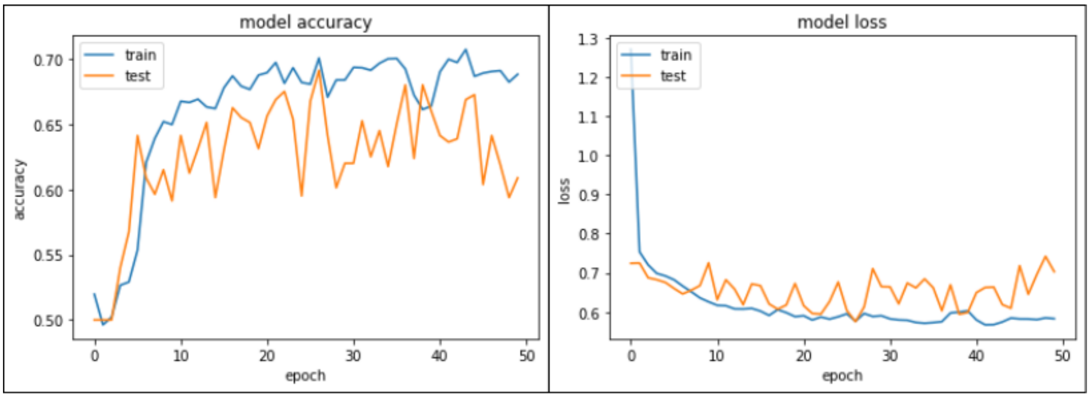

Figure 12: ResNet-50 accuracy and loss graph

6.5 Discussion

Table 1 compares the testing accuracy and loss of the four selected models. The VGG-16 and VGG-19 models performed the best out of the four models after 50 epochs. ResNet-50 produced the lowest accuracy out of all. The testing accuracy reflects the lack of image data to achieve better accuracies. Table 2 compares the models implemented to the models used in previous research. From table 2 we can see that the models used in previous literature have performed better using a different dataset. The VGG-16 has performed best for this research and in comparison to previous research. The new model VGG-19 chosen for this research has shown equivalent results. In this research, the four models were tested over a different number of epochs. The same model history was generated for 10,20,50 and 100 epochs. The models showed the best results for 50 epochs. When the number of epochs was 10 or 20 the models converged quickly but the accuracies were lower compared to 50 epochs. When the model was run for 100 epochs the final accuracies were no better than for 50 epochs and the total runtime was large for InceptionV3 especially. 50 epochs were found to be best in terms of accuracy and total runtime. This research makes use of data with a slight imbalance between the two classes male and female. This can cause the model to overfit the male class. The findings of this research apply only to the African population as the dataset consists of African-origin fingerprints. The results of this research can only classify the sex of the person by birth and not distinguish people by their identified gender. The lack of a large dataset and samples of different ethnicities limits the real-world applications of this research.

Table 1: Model Comparison

Model Name	Testing Accuracy	Testing Loss
VGG-19	71.9%	0.52
VGG-16	72.3%	0.52
InceptionV3	67.3%	0.58
ResNet-50	60.8%	0.7

Table 2: Model Comparison with Previous Work

Author	Dataset-same/different	Model Name	Accuracy
Shehu et al; (2019)	same	ResNet-34	75.2%
Hsiao et al. (2022)	different	VGG-16	79.2%
Hsiao et al. (2022)	different	InceptionV3	77.1%
Hsiao et al. (2022)	different	ResNet-50	79.1%
This research	same	InceptionV3	67.3%
This research	same	ResNet-50	60.8%
This research	same	VGG-19	71.9%
This research	same	VGG-16	72.3%

7 Conclusion

This research aimed to compare a new model VGG-19 in comparison to three best models from previous research VGG-16, InceptionV3, and ResNet-50. The VGG-19 model achieved an accuracy of 71.8% which is comparable to the accuracy VGG-16 achieved of 72.3%. The data augmentation helped to balance the dataset and achieve better accuracy. The pre-training of models using transfer learning has improved the training time from more than an hour to within 20 mins. The models have shown significant accuracy in classifying males and females using fingerprints. This research has shown that the VGG-19 model can perform as well as the previous best models used for this specific classification. The model can be further developed using better data augmentation such as GANs to increase dataset size. The implementation of this study can be further extended by training models using fingerprint data from different ethnicities to widen the real-life application.

References

Abuared, N., Panthakkan, A., Al-Saad, M., Amin, S. A. and Mansoor, W. (2020). Skin cancer classification model based on vgg 19 and transfer learning, *2020 3rd International Conference on Signal Processing and Information Security (ICSPIS)*, pp. 1–4.

Ahmed, A. A. and Osman, S. (2016). Topological variability and sex differences in fingerprint ridge density in a sample of the sudanese population, *Journal of Forensic and Legal Medicine* **42**: 25–32.

Al-Qerem, A., Salem, A. A., Jebreen, I., Nabot, A. and Samhan, A. (2021). Comparison between transfer learning and data augmentation on medical images classification, *2021 22nd International Arab Conference on Information Technology (ACIT)*, IEEE, pp. 1–7.

Alam, S., Dua, M., Gupta, A. et al. (2019). A comparative study of gender classification using fingerprints, *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, pp. 880–884.

Arun, K. S. and Sarath, K. S. (2011). A machine learning approach for fingerprint based gender identification, pp. 163–167.

Ceyhan, E. B. and Sagiroglu, S. (2015). Gender inference within turkish population by using only fingerprint feature vectors, Vol. 2015-January, IEEE Computer Society, pp. 146–150.

Chen, H. and Cao, P. (2019). Deep learning based data augmentation and classification for limited medical data learning, *2019 IEEE international conference on power, intelligent computing and systems (ICPICS)*, IEEE, pp. 300–303.

Gutiérrez-Redomero, E., Alonso, C., Romero, E. and Galera, V. (2008). Variability of fingerprint ridge density in a sample of spanish caucasians and its application to sex determination, *Forensic Science International* **180**: 17–22.

Hsiao, C.-T., Lin, C.-Y., Wang, P.-S. and Wu, Y.-T. (2022). Application of convolutional neural network for fingerprint-based prediction of gender, finger position, and height, *Entropy* **24**(4): 475.

Jian, W., Zhou, Y. and Liu, H. (2020). Lightweight convolutional neural network based on singularity roi for fingerprint classification, *IEEE Access* **8**: 54554–54563.

Lee, L. C., Adam, M. Y., Abd Sanih, S. N. and Bohari, N. I. (2020). Forensic gender discrimination in malaysian population using machine learning methods, *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, IEEE, pp. 1–5.

Nayak, V. C., Rastogi, P., Kanchan, T., Yoganarasimha, K., Kumar, G. P. and Menezes, R. G. (2010). Sex differences from fingerprint ridge density in chinese and malaysian population, *Forensic Science International* **197**: 67–69.

Shehu, Y. I., Ruiz-Garcia, A., Palade, V. and James, A. (2019). Detailed identification of fingerprints using convolutional neural networks, Institute of Electrical and Electronics Engineers Inc., pp. 1161–1165.

Sánchez-Andrés, A., Barea, J. A., Rivaldería, N., Alonso-Rodríguez, C. and Gutiérrez-Redomero, E. (2018). Impact of aging on fingerprint ridge density: Anthropometry and forensic implications in sex inference, *Science and Justice* **58**: 323–334.

Zhang, K., Cao, Z. and Wu, J. (2020). Circular shift: an effective data augmentation method for convolutional neural network on image classification, *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 1676–1680.