2023

Sequence Checking and Deduplication for Existing Fingerprint Databases

Tahsin Islam Sakif
West Virginia University, ts0056@mix.wvu.edu

Sequence Checking and Deduplication for Existing Fingerprint Databases

Tahsin Islam Sakif

Thesis submitted to the

Benjamin M. Statler College of Engineering and Mineral Resources

At West Virginia University

In partial fulfillment of the requirements for the degree of

Master of Science in

Computer Science

Jeremy Dawson, Ph.D., Chair

Donald Adjeroh, Ph.D

Katerina Goseva-Popstojanova, Ph.D

Lane Department of Computer Science and Electrical Engineering

Morgantown, WV

2023

Keywords: Machine Learning, Fingerprints, Neural Networks, Deep Learning

ABSTRACT

Sequence Checking and Deduplication for Existing Fingerprint Databases

Tahsin Islam Sakif

Biometric technology is a rapidly evolving field with applications that range from access to devices to border crossing and entry/exit processes. Large-scale applications to collect biometric data, such as border crossings result in multimodal biometric databases containing thousands of identities. However, due to human operator error, these databases often contain many instances of image labeling and classification; this is due to the lack of training and throughput pressure that comes with human error. Multiple entries from the same individual may be assigned to a different identity. Rolled fingerprints may be labeled as flat images, a face image entered into a fingerprint field or images entered in incorrect orientation (such as rotated face images, left or right iris, etc.) are common errors found large database records. Ultimately, these enrollment errors make it impossible to identify that individual upon subsequent identification encounters. Sorting through hundreds of images to check for classification errors is a tedious and time-consuming task, especially when several biometric databases are combined. Our goal is to correctly identify misclassified fingerprints using controlled embeddings and thresholds.

This work provides a new perspective on image sorting as it focuses not on the traditional aspects of increasing accuracy metrics but provides a look into multiple factors through various embeddings and thresholds to provide a tool that can be used to scour large datasets with ease to provide what percentage of the images need manual correction. The proposed network provides various metric scores which allowed for analysis on the most effective embedding and thresholds to use, resulting in a proof-of-concept to be used for practical purposes in the real world.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

I would like to express my deepest appreciation to my research and academic advisor Dr. Jeremy Dawson. He was an invaluable help in my academic journey here at West Virginia University as an international student and allowed me to complete this research which enhanced me into an engineer, researcher and student. I would also like to extend my gratitude to Dr. Donald Adjeroh and Dr. Katerina Goseva-Popstojanova as they provided me with valuable information on neural networks and software engineering. I am also grateful to receive assistance from my colleagues and friends at the WVU Biometrics Lab to procure the data and tools to complete this work. I am indebted to my friends and family members who have provided support over my long journey in West Virginia University, and without their kind words of encouragement I would not be where I am today. Thank you all for everything.

Introduction

1.1 The World of Biometrics

Over the decades, “biometrics” as a definition has evolved throughout literature as advancements have been made throughout science and technology. The word comes from the Greek words “bio” (life) and metric (to measure) [1]. Biometrics as a science is recognizing the identity of a person based on the unique physical or behavioral attributes of the individual such as face, fingerprints, and iris [2]. Biometrics as a technology is a rapidly evolving field with applications that range from access to devices to border crossing and entry/exit processes. These biometric modalities are categories in a biometric system which depends on what the human input is in the system, which typically include face, iris, fingerprint and hand geometry [3]. As there are many modalities, there is no single modality that is considered the best, as there are different factors to consider such as the location, device, existing data and users. As with any type of data, the more information that is available, the more reliable the system is. Large-scale applications to collect biometric data, such as border crossings result in multimodal biometric databases containing thousands of identities.

Large multimodal biometric datasets have become a cornerstone in homeland security and identification. The US Department of Homeland Security (DHS) expects to have over 259 million face, fingerprint and iris scans in their biometric databases in 2022 [4]. In contrast, the database has increased by 40 million since 2017.

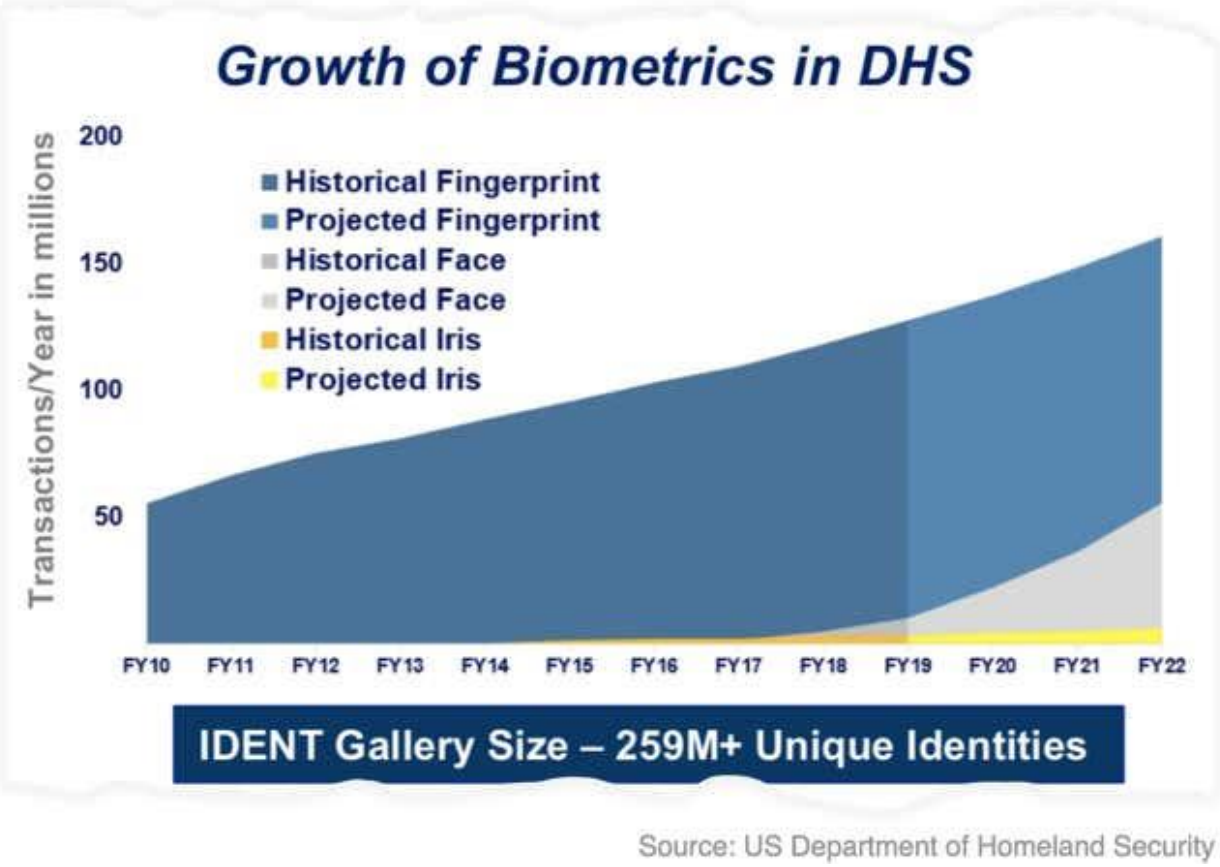


Figure 1 Growth of Biometrics in Department of Homeland Security

DHS's database is also the world's second largest database; India's biometric network that spans the entire country is even more impressive. Introduced in 2009, "Aadhaar" was a method to identify a resident and their identity easily, especially for the rural poor who lack the education and financial means to have self-identification [5]. Aadhaar profiles stores fingerprints, iris and face images and produces a unique 12-digit number which is stored in their centralized database, where other third-party may access for identity inspection [6]. This profile allows the residents to perform almost anything; creation of bank accounts, food rations or even new SIM cards [7]. The reasoning for such a system was for both security and fraud prevention. These digital biometric systems are constantly growing worldwide amongst governments as they are relatively

low cost and easy method to track residents. However, these databases are not perfect. Aadhaar as an experimental system has had issues with authentication due to human tampering, resulting in people unable to sign up for schools or access food because of errors in the dataset [8]. DHS's new Homeland Advanced Recognition Technology (HART) database containing millions of modalities of biometrics of face, iris, fingerprint and even DNA of both citizens and foreigners in the United States faces scrutiny as it contains inaccurate data. Their tests of their own systems falsely rejected 1 in 25 travelers; a report recently noted that "DHS error-prone scanning system could cause 1632 passengers to be wrongfully delayed or denied boarding every day at New York's John F. Kennedy (JFK) International Airport alone" [9]. The Federal Bureau of Investigation (FBI) admitted that it's Next Generation Identification database "may not be sufficiently reliable to accurately locate other photos of the same identity, resulting in an increased percentage of misidentifications". Other foreign governments who rely on such a database reported false positive rates almost as high as 98%. If the data in these large datasets are not sequence checked for image labeling and classification errors, the consequences can be catastrophic as innocent people may be labeled as suspects or unable to prove their identity under the mercy of their biometric data under the government.

1.2 Problem Statement

As these biometric modalities result in large amounts of information, datasets are created in order to analyze and store biometric information. However, with any human endeavor, human factors can affect even the most robust systems. Large biometric datasets contain image labeling and classification errors due to human operators. This is due to a variety of factors. Lack of training is the most common as these large operations require a large personnel to handle them, and without proper training common errors can occur such as multiple entries from the same individual being assigned to a different identity. Throughput pressure may be another factor. Electronic Biometric Transmission Specification (EBTS) and Electronic Fingerprint Transmission (EFT) files that are used by law enforcement agencies and other organizations may contain mislabeled fingerprint images or images that are out of sequence. Multiple entries from the same individual may be assigned to a different identity [10]. Rolled fingerprints may be labeled as flat images, a face image entered into a fingerprint field or images entered in incorrect orientation (such as rotated face images, left or right iris, etc.) are common errors found in standard EBTS records. Such errors can result in an entire dataset being unreliable and not fit for use, as they are the datasets used for research on the basis that they are correctly obtained and checked [11].

As we cannot ascertain that the reliability of these automated biometric systems is completely reliable due to the risk of being tampered due to human error, there needs to be a more reliable method to process and remove duplicate identities in a dataset. The process to detect and remove duplicate identities in a database is commonly referred to as de-duplication [12]. Currently, there is no standardized deduplication scheme for

large scale databases which contain face, iris and fingerprint images [13]. Ultimately, these enrollment errors make it impossible to identify that individual upon subsequent identification encounters. Sorting through hundreds of images to check for classification errors is a tedious and time-consuming task, especially when several biometric databases are combined.

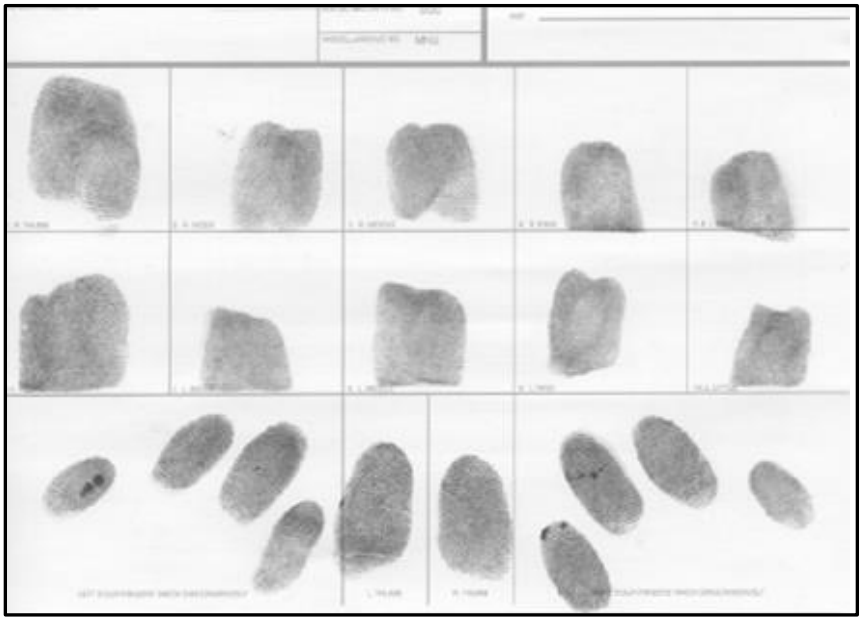


Figure 2. Misclassified Prints Due to Manual Error

1.3 The Role of Neural Networks

Utilizing the nature of neural networks is the solution to sort through the datasets at an efficient rate. Object and image classification is a well-known application area of convolutional neural network (CNN) architectures. Neural network approaches are especially suitable for fingerprint databases [14]. Fingerprints form a specific set of patterns of minutiae and ridges that allow for statistical characteristics. Neural networks tend to work well on seeking patterns, and fingerprints have a specific composition that allows for human beings to be similar but also distinct enough to use them as unique

identifiers [15]. Neural networks can also avoid issues with conventional approaches to matching minutia, which are sensitive to noise (inked fingerprints resulting in smudging which distort current minutia) and is also expensive (basically a graph matching problem) [14]. They are also trainable from samples as large datasets of fingerprints are easier to obtain from forensic sources that allow for millions of prints to be analyzed and trained on [16].

1.4 Previous Work

The concept for the biometric data classification was originally designed for multiple modalities used for both face, iris and fingerprint images. A preprocessing module was created to enable comparison of the different modalities in a common format compatible with the neural network architecture. WSQ files obtained from data collections were reshaped into 180X180X3 shape and loaded into a trained “ImageNet” dataset. The weights of this model were fine tuned in order to be suitable for biometric sorting purposes, using 5 epochs. Both inter-modality and intra-modality activities were used on the data collected through numerous collections at the West Virginia University Biometrics Lab following IRB protocols.

For face images, the inter-modality classifications were frontal, profile and other seen in Figure 4. In the case of iris images, they were classified as left or right, and for fingerprint images, the classes were flat and rolled, with finger type in consideration such as index, middle, ring, thumb. A compilation of results can be seen in Figure 3.

.

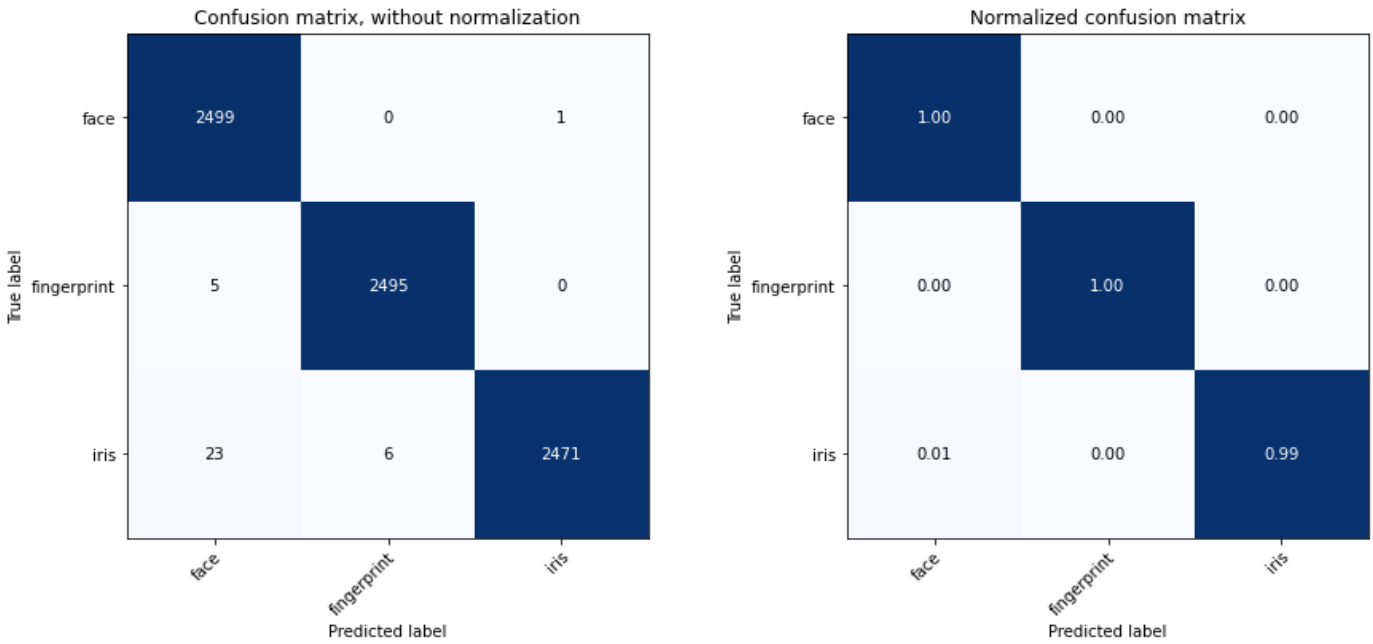
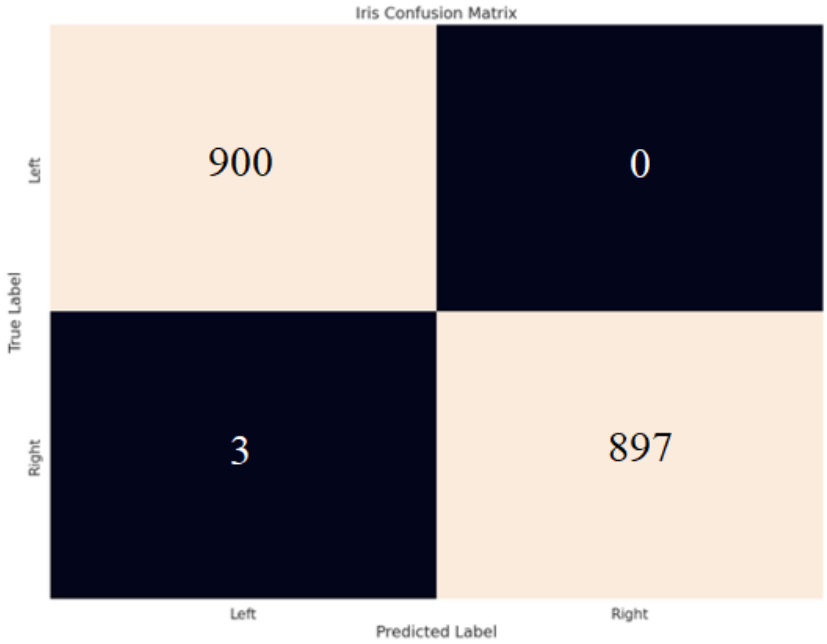


Figure 3 Face, Fingerprint and Iris Classification

Figure 4 Frontal, Profile and Other Specifications for Face



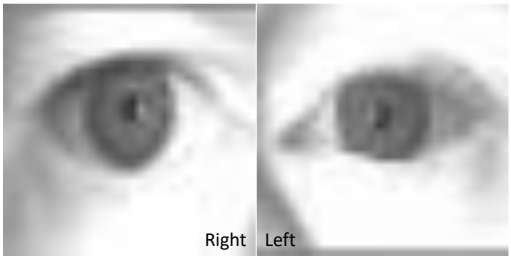


Figure 5. Iris Classification Results

Inter-modality classification accuracy of face/fingerprint/iris sorting obtained a classification accuracy of 99%. The network was also able to easily distinguish between left and right iris sorting along with face pose sorting, both with a classification accuracy of 99% as seen in Figure 5. Fingerprint classification reached an accuracy of 84%. The fingerprint dataset caused confusion as the system could not recognize differences between left/right index and right/left thumbs causing huge drops in accuracy. Since the fingerprint classification was lacking in accuracy, the aim was to use a different neural network architecture to improve the efficacy of fingerprint detection.

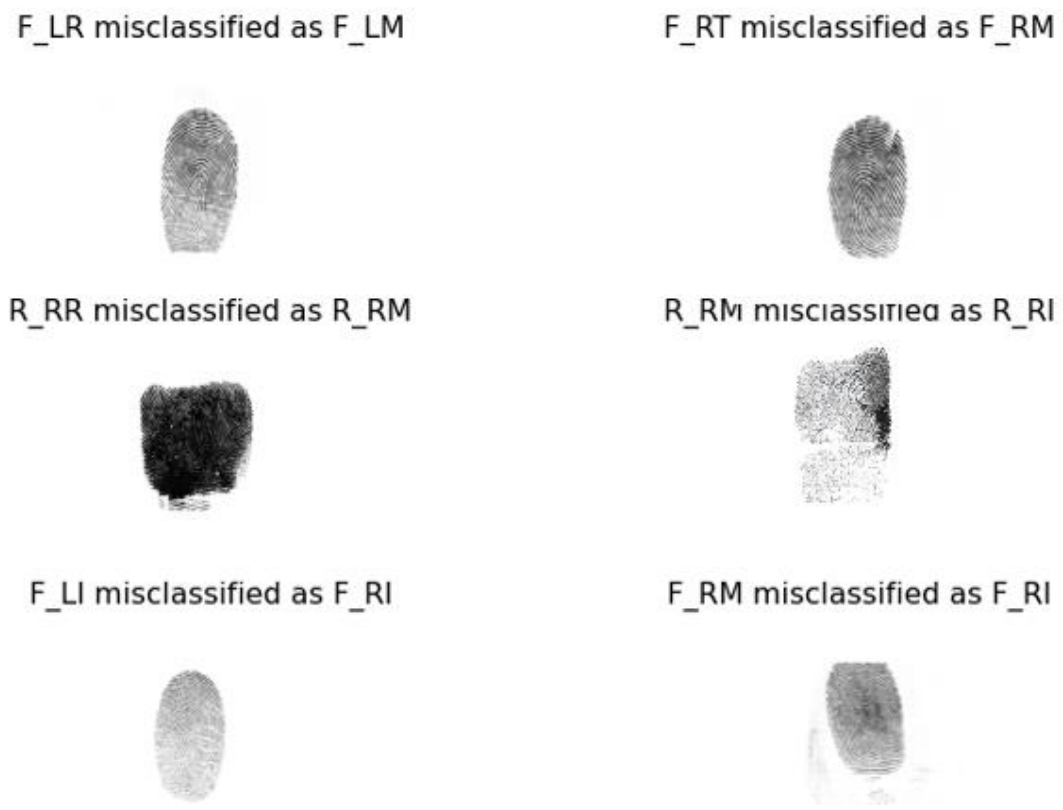


Figure 6. Misclassified Fingerprints from Initial Testing

As seen in Figure 6, Misclassified Fingerprints from Initial Testing revealed that some fingerprints were prone to misclassification as the network could not distinguish one from the other if they were of a similar type.

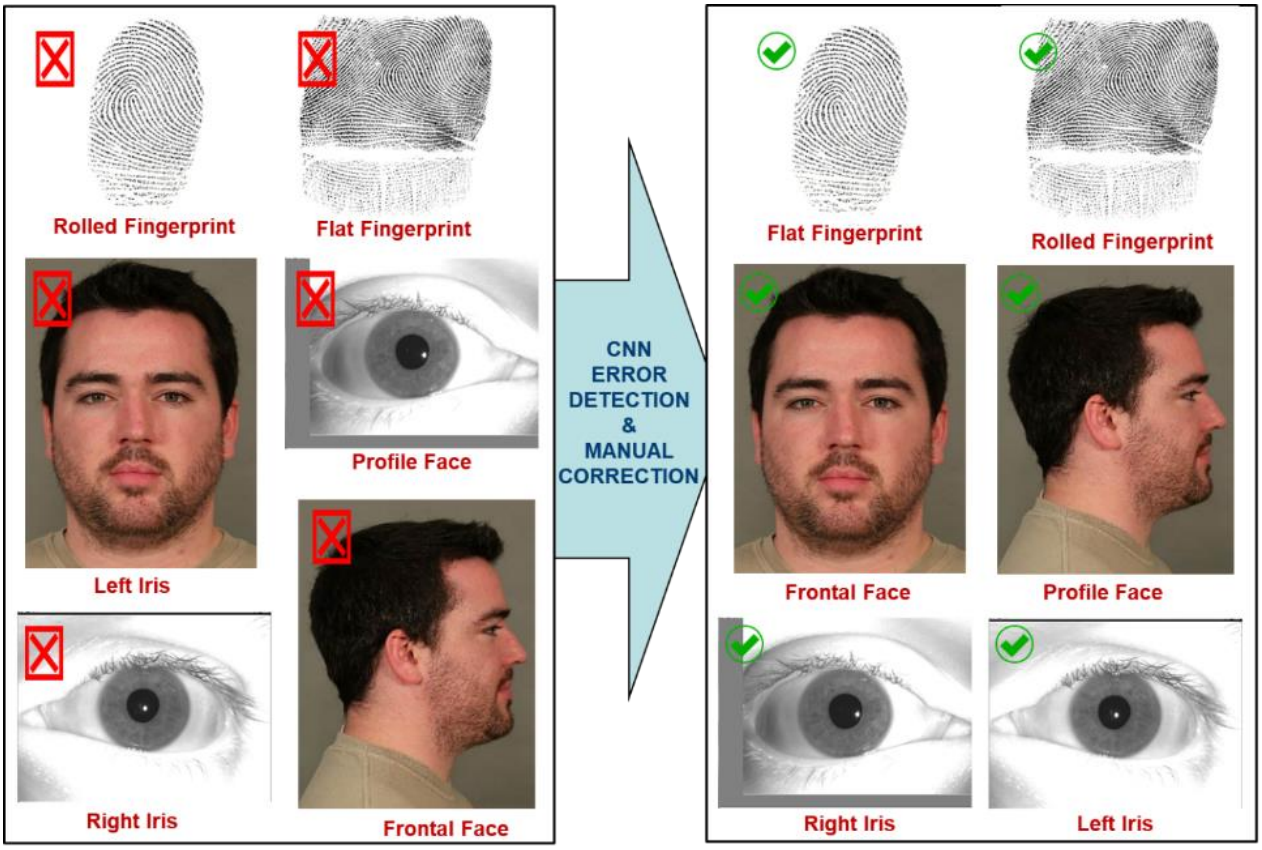


Figure 7. Classification Tool Concept

The result was to bring the concept of a classification tool to reality, as feeding it a dataset will provide a method to classify images into their correct labels or show where the errors lie for manual correction as shown in Figure 7.

1.5 Goals and Objectives

The goal of this research effort is to apply deep learning techniques to identify misclassified biometric images in multibiometric datasets and label the suspected errors for examination and manual correction. Inter-modality classification methods will be used to sort face, fingerprint, and iris data from large scale datasets, while intra-modality classification will be used to sort mistakes in left vs right iris, rolled vs flat fingerprints and frontal vs profile/other-angle face. In order to accomplish these tasks, a modified version of the Residual Network architecture will be used, known as Resnet. The main focus will be on fingerprint analysis, and of how to minimize the risk of errors in large datasets to ease manual error correction.

The objectives are as follows:

1. Preprocess images into a viable dataset to use to test the neural network on. Images will have to be sorted into classes that correspond to their specific finger (index, middle, ring, little, thumb).
2. Design a network architecture that uses the ideas and structure of the previous ImageNet neural network and fine tune it for fingerprint classification.
3. Testing of the new neural network architecture to test various options such as different embeddings, weights, etc.
4. Optimization of neural network in order to achieve a high classification score to be used on different datasets.

1.6 Impact

If this research is successful, it will impact the field of biometrics in providing a proof-of-concept tool that will allow for finding classification errors in large datasets that may be compromised due to human or machine error. Large datasets that are no longer usable due to errors may be recovered and the data saved for further use. As a new concept, this research will provide an idea for a more specialized tool that will become the standard for preprocessing of large amounts of images before being finalized into a dataset and allow for far less manual correction.

This research, however, is focused on image sorting, but not in the traditional sense of obtaining the highest statistical value. The use cases are:

- Large datasets that we know have errors to quickly sort through and identify misclassified images in a practical sense. The ideal would be for such a network to be run on any dataset to quickly find the ideal manual classification value to reduce the amount of time needed to go through the whole dataset.
- Feedback to operator in real time as this network can be used in areas or situations without large computational ability or before the dataset is shipped for finalization.
- Show the different metrics that impact such a network, as most research has mostly focused on improving the accuracy only without a wide perspective on the other factors.

The clarification needs to be made that fingerprint research has been well documented and published, but the practical aspect of identifying misclassifications and identifying the factors that impact the results is a new territory yet to be explored.

This research is organized into the background required to understand the various concepts of neural networks, fingerprints and image classification and lead into the metrics used for understanding the neural network's results. Afterwards an overview of the dataset used and the results obtained from them to accomplish the goals of the research.

Chapter 2: Theory and Background

2.1 Neural Networks Background

Machine learning is one of today's most rapidly growing technical fields in both research and industry. At its fundamentals it is both artificial intelligence and data science, within the realms of both computer science and statistics. It is based on the question of how we can create computer systems that improve automatically through experience. The field has progressed quite dramatically over the past decade, used in both laboratories and commercial use [16]. Neural networks are a specific type of machine learning model which functions similarly to the human mind; interconnected nodes known as neurons work together via an algorithm to understand data being fed into it. Similarly, to our own brain's neural system, the way we perceive with our eyes, mouth and noise and process that information is similarly how a neural network being fed the proper data is able to perceive and understand patterns to come to the correct conclusion. This idea was introduced as Hebbian learning, where D.O. Hebb, a renowned neuropsychologist explained that a network of neurons learning tasks repeatedly reinforces learned behavior into a stronger entity [17]. With the introduction of the perception in the 1950s by Frank Rosenblatt, Hebb's concept was brought to reality in computing. However, the field had a dull recession as researchers believed that the effectiveness of neural networks is limited. It was only when in the 1980s where gradient descent, backpropagation and other innovative ideas brought back the topic of neural networks with interest [18]. As with Moore's law as time went on the computing power of the world has substantially increased, along with new ideas and algorithms neural networks

evolved into practical applications. In the current era, neural networks have been applied everywhere from research, industry and any other trainable task.

2.2 Background on Fingerprints

One of the most common methods of identification in biometric systems is fingerprints. Fingerprints are made from a sequence of ridges and furrows on the surface of the finger, with a central core. There are a unique set of patterns of loops and swirls results in each human's print that makes up a fingerprint. An arch is a pattern where ridges enter from one side of the finger to another, while rising in the center to form an arc. A loop pattern is when ridges form a curve and exit the same side they enter while whorl ridges form in a circular shape around the middle of the finger [19]. The irregularities that these ridges are characterized by are known as minutiae. Minutiae are characteristics of ridges, such as when they end (ridge ending) and splits into more ridges (ridge bifurcation) [20]. These minutiae and its patterns allow us to analyze fingerprints and identify them, as no two fingerprints are the same [21]. Fingerprint classification by feature type of loop, whorl and etc. is mature and well understood [22]. Lin et. al were able to create a fingerprint classification algorithm with a robust feature extractor which was able to extract salient features from input images [15].

A study by R. A. Searston demonstrated that over time humans can tell the difference between finger types based on fingerprints given sufficient experience and training [23]. Trainee examiners were tested over their first 12 months of working in a fingerprint unit and were eventually able to distinguish different fingerprints, obtaining various index scores. As expert latent print examiners are able to sort fingerprints by type, we can use

this concept in this research to train the neural network to identify and correctly classify prints based on their unique characteristics [24].

2.3 Image Classification

As neural networks are applicable in any modern task, it is of utmost importance in the field of image classification which is deeply ingrained in this paper's work. For analyzing and deciphering the data obtainable from images, the typical network that is used is known as a convolutional neural network [25]. A convolutional neural network, often abbreviated as CNN, is an algorithm in deep learning that essentially takes in an image and assigns characteristics to it, known as weights and biases. The main goal of this is to be able to understand the difference between one image and the next. Traditional network architecture was known as the multi-layer feed forward network, which allows us to receive binary scores on predicting classes [26]. However, a convolutional neural network differs in that it is able to have high accuracy in complex images with enormous pixel densities and is able to fit the dataset by reducing the amount of parameters required [27]. As an example, if an image was of an immense size of 16K resolution, which consists of a pixel count of 15360X8640. With its separate color planes of red, green and blue (RGB) it will be very difficult computationally to directly feed this as an input image into a feed forward network. However, with a convolutional neural network it is possible to pick the most crucial features in order to achieve the desired prediction, while being computationally easier to process [28]. This allows us to use them on massive datasets, consisting of thousands of images with ease.

These neural networks have an output known as a feature map. LeCun et al. stated that “At the output, each feature map represents a particular feature extracted at all locations on the input” [25]. This factor allows for convolutional neural networks to optimize and learn from the images inputted, while factoring in the different weights, scales and size of the network. The main purpose of the CNN is to essentially extract these high-level features from the input image [29].

After the convolutional layer in a CNN, there is also another feature known as the pooling layer. This layer’s main purpose is to factor in the computational power of the neural network, and to optimize it to the best of its ability [30]. This is done through an action known as dimensionality reduction, which as its name implies down samples the provided input image for efficient learning. There are two methods of pooling, which are known as Max Pooling and Average Pooling. Max Pooling is obtained when in a specific kernel, obtaining the maximum combined value of the input image is achieved. Average Pooling gives the average value of all the values in a specified kernel.

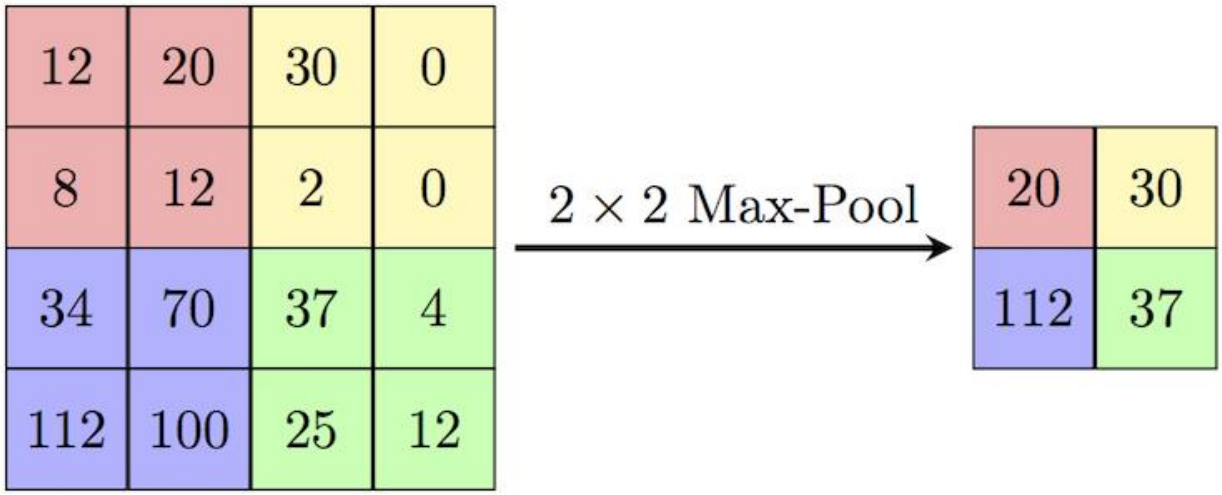


Figure 8 Max Pooling Demonstration [31]

With these architectures built into a convolutional neural network, image classification has achieved another level of advancement [32]. For this current work, it will be focused on one of the most recognized CNNs which is ResNet.

2.4 ResNet50

Resnet50 is a residual network that won the ImageNet Large Scale Visual Recognition Challenge for its image classification efficiency with large datasets [28]. Resnet allows for networks to be hundreds to thousands of layers deep and still achieve great accuracy. This neural network architecture can be used as a deduplication and finger sequence checking method to sort through massive datasets with ease. Presented in this paper is an implementation of the network to classify fingerprints in a large dataset and label misclassifications efficiently.

The Resnet50 architecture was one of the biggest surprises of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The competition involved the classification of 1000 separate object categories. Resnet stands for Residual Network which was introduced by Microsoft Research for deep residual learning for image recognition. Deep convolutional neural networks function with many layers, each of the layers performing a specific task. However, adding more layers does not result in better accuracy, as it is an inverse relationship (the higher the layers, the lower the accuracy). Degradation occurs as deeper networks start to converge, resulting in a large problem, as not all systems are the same resulting in difficult optimization unique to each [33]. Residual networks resolve this problem through the usage of skip connections. Instead of going through layer by layer, skip connections allow for an alternate shortcut for

gradients to pass through and also learn identify functions. Through this “skipping,” it allows for higher layers of the model to function as well as the lower models. This skipping adds outputs from previous layers to outputs of the stacked layers resulting in more efficiency and deeper networks.

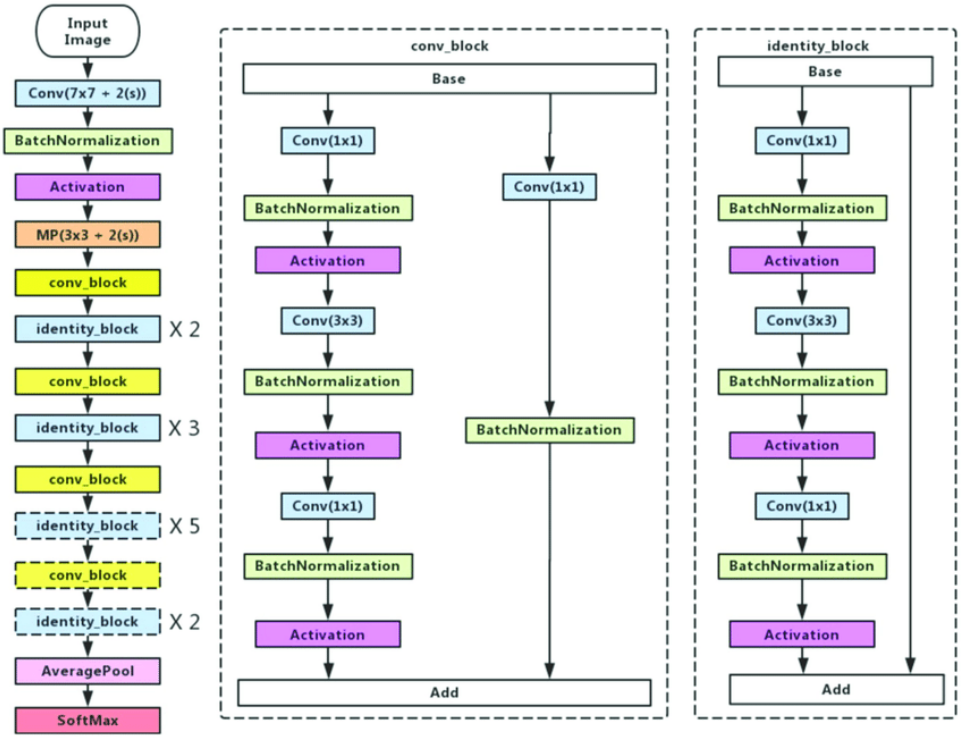


Figure 9. Resnet50 Architecture [33]

2.5 Siamese Network

According to the Merriam-Webster dictionary, the word “Siamese” is an adjective meaning something that exhibits great resemblance, or very like. A neural network of this concept was introduced in the 90s by Bromley [34]. It is a class of neural architectures which consists of two or more identical subnetworks, more specifically the networks being a complete duplicate of one another with the same settings and weights. Both the networks generate feature vectors for each input and compares them.

Siamese networks have been used for face recognition, which is the problem of identifying a specific individual. The networks have been used to see whether a person in an input image pair is the same person. In terms of this paper's work, the same concept can be used for fingerprints [35]. By feeding the network positive and negative images, it is possible to use the dual nature of Siamese networks to come to conclusions regarding an image. These positive and negative images are known as anchors. These are an image of the same class (positive) or an image that is not of the same class (negative). The anchor essentially functions as a ”representative” for that specific class. As an example, feeding the network an image of a left index fingerprint as a positive anchor teaches the network that this print is a left index fingerprint, and allows it to compare it with another left index print for confirmation. Now comparing the anchor, which is the representative with all other prints in a dataset the network can provide an estimation of how close or far the other prints are to the representative image, which in this case is the left index print.

2.6 Outcome Metrics

Outcome metrics are a means to quantitatively assess the results of the experiment. In this work, various outcome metrics were used to assess the usability of the neural network.

Accuracy

Accuracy is a metric that allows to evaluate our effectiveness for classification models. It is the number of correct predictions over the total number of predictions [36].

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Equation 1. Accuracy Explanation

Accuracy is calculated by positive and negative terms. There is True Positive (TP) where the model predicts the positive class correctly. True Negative (TN) is where the model predicts the negative class correctly. False Positive (FP) is when the model predicts a class to be correct when it is not. False Negative (FN) is when the model predicts a class to be incorrect when it is not. By combining these terms, it is possible to obtain a numerical value for accuracy:

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

Equation 2. Accuracy Equation

In the case of this research, accuracy defines the number of fingerprints correctly classified over the total number of fingerprints in the dataset.

Conditional and Unconditional Accuracy

In the case of this research, there are different ways of calculating the accuracy as it is not simple as the case of only positives and negatives. In order to provide a full picture, we calculated conditional and unconditional accuracy. Conditional accuracy only considers the images that are classified as positive or negative but does not include the images that are uncertain. Unconditional accuracy on the other hand considers all images/factors.

$$Conditional\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 3. Conditional Accuracy Equation

$$Unconditional\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN + Unconditional\ Positive + Unconditional\ Negative}$$

Equation 4. Unconditional Accuracy Equation

Uncertain

Uncertain are predictions that fall in the range of threshold and 1-threshold. These are predictions for which there is not enough sufficient confidence to mark them as positive or negative. These can be images that have been tampered with (blurred or lower quality) or have some sort of error that the neural network is unable to classify accurately.

Precision

Precision determines which proportion of positive identifications were accurately predicted to be correct. It is when True Positive is over the combination of True Positive and False Positive [37].

$$Precision = \frac{TP}{TP + FP}$$

Equation 5. Precision Equation

In the case of this research, precision defines the network's ability to correctly predict and classify specific classes of the fingerprints, such as front left index and other classes.

Recall

Recall determines which proportion of actual positive identifications were correct. It is calculated using True Positive over the combination of True Positive and False Negative [37].

$$Recall = \frac{TP}{TP + FN}$$

Equation 6. Recall Equation

Conditional and Unconditional Recall

In the case of this research, there are different ways of calculating the recall as it is not simple as the case of only positives and negatives. In order to provide a full picture, we calculated conditional and unconditional recall. Conditional recall only considers the

images that are classified as positive or negative but does not include the images that are uncertain. Unconditional recall on the other hand considers all images.

$$Conditional\ Recall = \frac{TP}{TP + FN}$$

Equation 7. Conditional Recall

$$Unconditional\ Recall = \frac{TP}{TP + FN + Uncondtional\ Positive}$$

Equation 8. Unconditional Recall

F1 Score

The F1 Score, also known as the F-measure, is a metric which is based on error. It measures the neural network model's performance by calculating the harmonic mean of precision and recall for the minority positive class [38]. It is one of the most commonly used metrics for classification models as it provides easy to understand results for balanced and imbalanced datasets factoring in the precision and recall values.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Equation 9. F1 Score Formula

To interpret the score, F1 provides an overall model performance from 0 to 1, with 1 being the best possible score. It shows the model's ability to detect positive cases in recall and accurately classified cases in precision. In the scope of this research, there will be Conditional F1 and Unconditional F1, with one considering the uncertain factor while the other does not.

F1 Score	Interpretation of Score
Greater than 0.9	Excellent
0.8 – 0.9	Great
0.5 – 0.8	Average/Medium
Less than 0.5	Low/Undesirable

Table 1. F1 Score Distribution

Chapter 3: Datasets, Networks and Experiment

3.1 Dataset Description

For the biometric data classification effort described here, a preprocessing module was developed to enable comparison of the different modalities in a common format compatible with the neural network architecture. WSQ files were reshaped into an 8-bit PNG using the reshape function and convert them into 128X128X3 shape. Network weights were loaded from the previously trained “ImageNet” dataset used for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) which was involved in the classification of 1000 separate object categories. Transfer learning was then used to fine tune the weights of this model for biometric sorting purposes, resulting in 5 epochs to train to its current level. Inter-modality classification was trained and tested on a small subset of operational data comprised of 7500 images (2500 for each modality) and the intra-modality activities were used on data collected by the West Virginia University Biometrics Lab via approved IRB protocols, with 40353 images used for training and 10088 used for testing. The datasets were sorted into both intramodality and inter-modality specifications. The fingerprint images were labeled and organized into 20 categories and in greyscale format. There are 20 classes, one for each finger categorized by flat or rolled based on how they were taken.

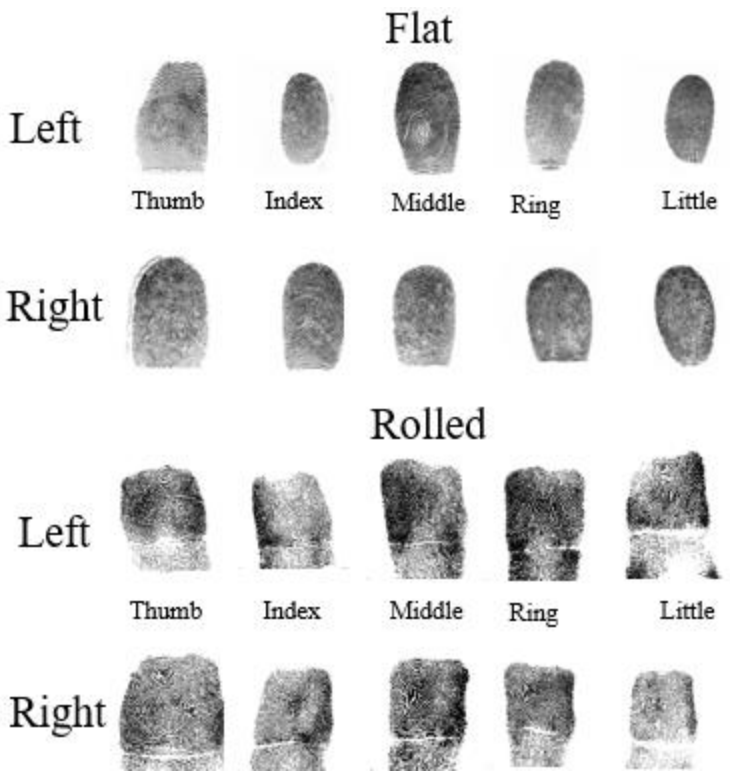


Figure 10. 20 Classes for Each Respective Fingerprint for Flat and Rolled Prints

3.2 Finger Sequence Network

The first iteration of the finger sequencing neural network was made using the Resnet50 convolutional neural network and shown in the figure below:

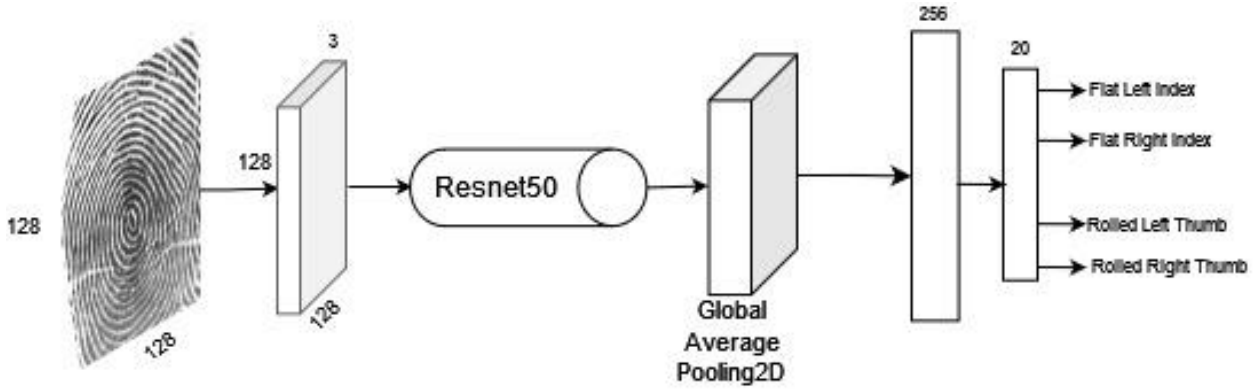


Figure 11. FingerSequence Network Initial Architecture

An image of size 128X128X3 was fed into the ResNet50 network and passed through the Global Average Pooling layer. This layer serves as a feature extractor. The extraction is achieved by calculating the average of the values in the pooling window. The network then passes through a dense layer of 256 neurons which represents an embedding of the image. The final layer consists of 20 neurons each representing one of the possible 20 classes. Only one of the output neurons is set to 1. All the other neurons are set to 0. An example would be feeding it a flat left index image it will try to correspond the entry to one specific class.

The accuracy of this network was 87%. Classification accuracy needed to be improved in order to reliably scour through datasets and provide accurate results.

3.3 Siamese Training

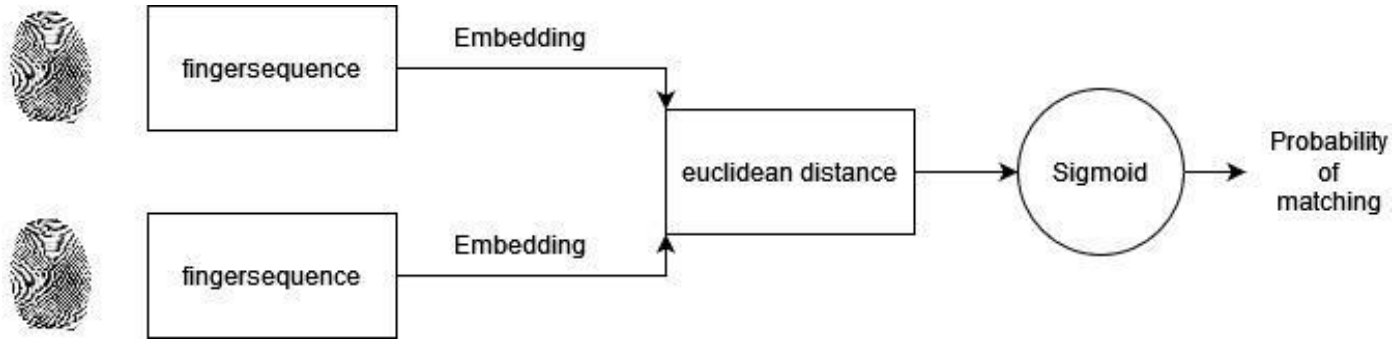


Figure 12. Siamese Network Architecture

The updated network was the previous network converted into a Siamese network. It consists of two finger sequence networks modified to produce an embedding vector of dimensions 256 for each image. The purpose of the embeddings is to reduce the dimensionality of the images and are vector representations. They are then passed through to the Euclidean Distance layer. This is a distance measure which allows for the summarization of the relative difference between two objects in a domain. By calculating the distance between two vectors, this metric allows to measure the similarity between two observations. The Euclidean distance of the embedding vector is then fed into a sigmoid function. This produces the probability of matching between 0 and 1, with 0 meaning the images are the same and 1 meaning the images are different.

The original network always outputs a vector of size 20, where only one value is set to 1. Each entry corresponded with one class. In the new network, the output is a probability of 0 or 1, indicating whether the two images are the same. This is done by feeding the network 2 images. The network outputs a 1 or 0 if the images are the same or not instead of directly classifying the image.

The new network is trained with pairs. There is a representative image for a specific class that is fed to the network, which functions as the anchor. This allows the network to learn that positive images are 1, and negative images are 0. Once pair is positive with an image that is of the same class as the anchor image, and the other pair is negative, which is an image that is not of the same class as the anchor image. A positive pair can be labeled as the genuine pair, while the negative pair can be labeled as an impostor pair.

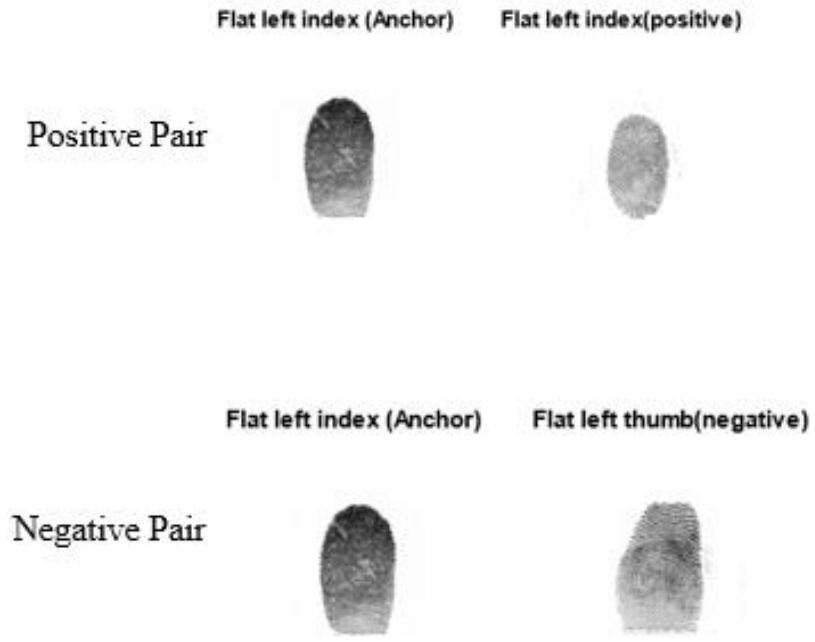


Figure 13. A Flat Left Index compared with a Flat Left Thumb

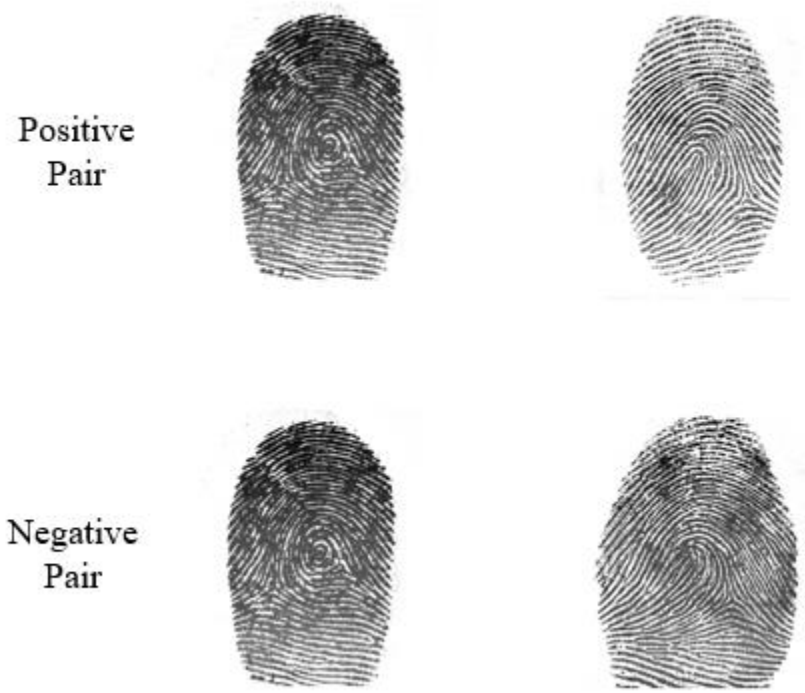


Figure 14. Example of Positive Pair (Index and Index) with Negative Pair (Index and Thumb)

3.4 Design of Experiments

The design of experiments is using statistics to plan, conduct and analyze controlled tests in order to understand various factors of parameters. It is possible to manipulate our inputs to achieve our desired outputs to identify key explanations and results [39]. These inputs are known as independent variables, and the outputs the dependent variables. By establishing a design of experiments, it is possible to obtain validity, replicability and reliability in research.

A factor is an independent variable whose settings are set by the experimenter. In this work, the factors are the number of neurons in the embedding, and the value of the thresholds used. A key idea was to see if more embeddings result in higher classification accuracy.

In this paper we are using 3*4 full factorial design to evaluate the impact of the factors on the precision, accuracy and uncertainty of different thresholds. This allows us to examine all the combinations to see how the embeddings and thresholds affect the network.

The embedding levels used are 64, 128, 256 and 512 while the threshold levels were 0.1, 0.2 and 0.5. The independent variables are the embeddings and the threshold values. The dependent variables are the accuracy, precision and uncertain values obtained from running the network. There is an addition of the unconditional variants of accuracy and recall to see how much the uncertain factor affects the network. Each experiment was replicated 10 times. Results were then stored in a spreadsheet for analysis. The goal of

the research was to see how changing the embeddings (64, 128, 256, 512) and the thresholds (0.1, 0.2, 0.5) affected the classification accuracy of the Siamese network.

Factor 1: Embedding	Factor 2: Threshold
64	0.1
64	0.2
64	0.5
128	0.1
128	0.2
128	0.5
256	0.1
256	0.2
256	0.5
512	0.1
512	0.2
512	0.5

Table 2. Embeddings and Thresholds Outline

Chapter 4: Results

4.1 Design of Experiment Results: Embedding

It was hypothesized that by using two different embeddings of 64, 128, 256, 512 the classification accuracy, precision and recall would improve. The uncertain factor (number of images that need to be manually classified) is factored into the conditional versus unconditional factors of accuracy and recall view its impact.

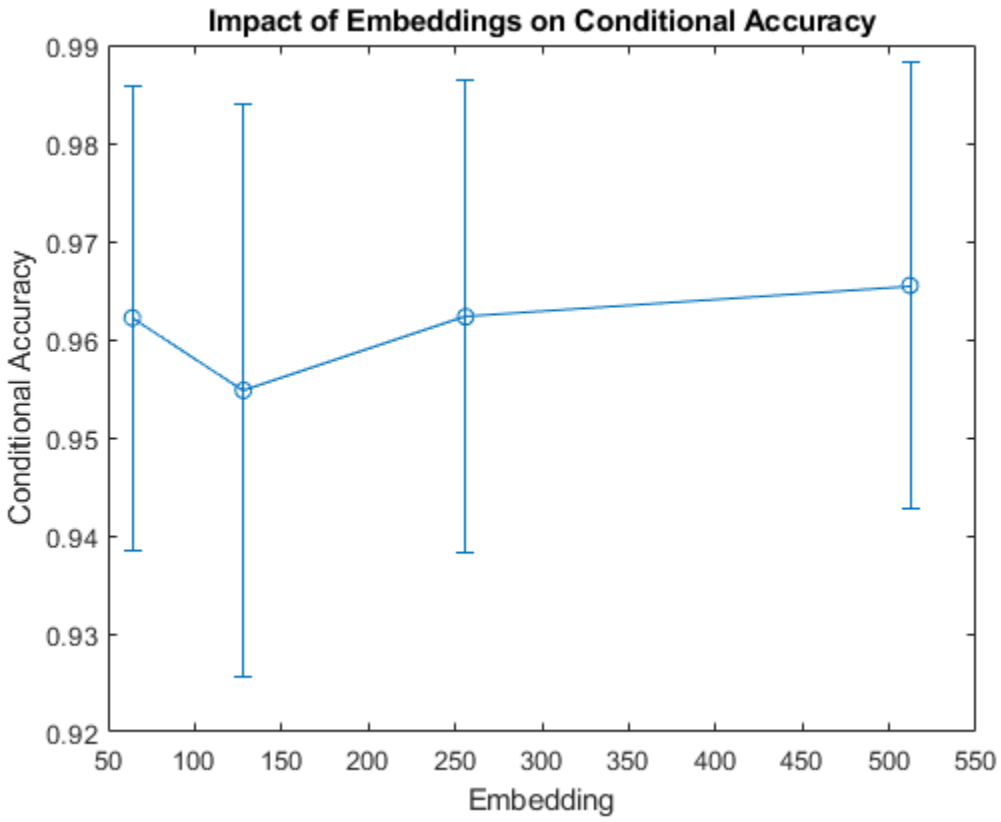


Figure 15. Impact of Embeddings on Conditional Accuracy

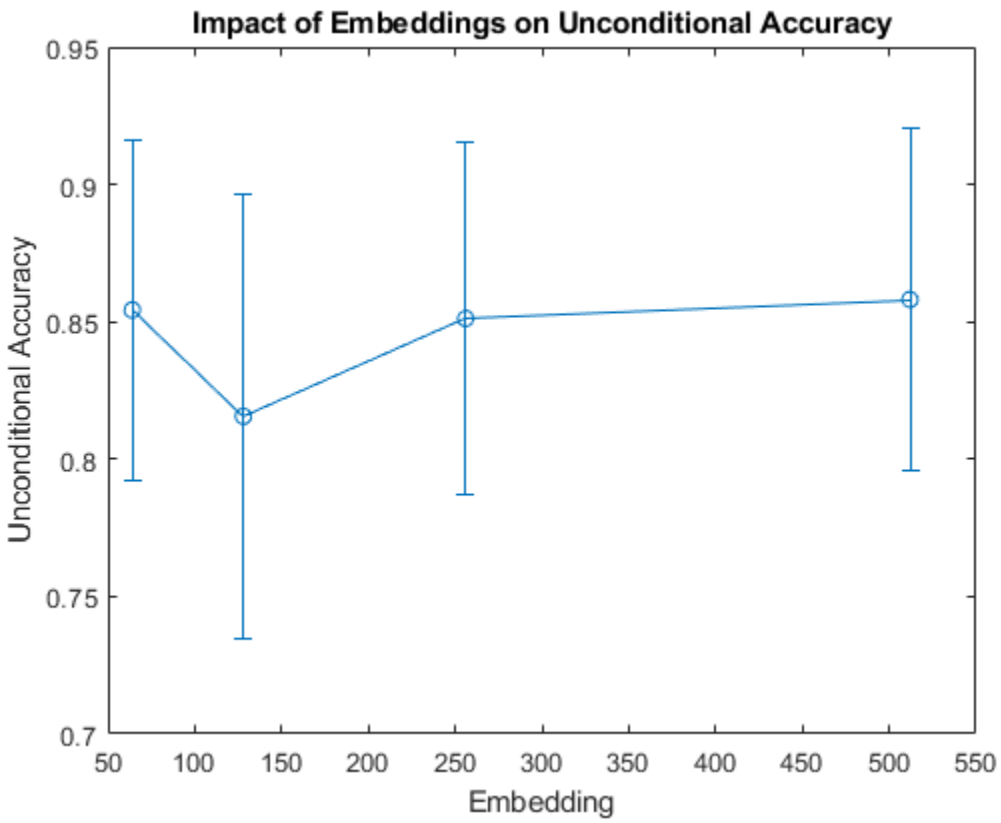


Figure 16. Impact of Embeddings on Unconditional Accuracy

As we compare both Figure 14 and 15, it is possible to see that the impact of embeddings is not significantly affecting the classification network. The error bars show the standard deviation of the accuracy at the specified level. Higher amounts of embedding seem to not significantly impact unconditional or conditional accuracy. However, the effect of the uncertain images are evident as the values for unconditional accuracy is much lower than the conditional accuracy values. The uncertain factor decreases the classification accuracy significantly. Surprisingly, 128 embeddings seem to perform badly compared to the other embeddings.

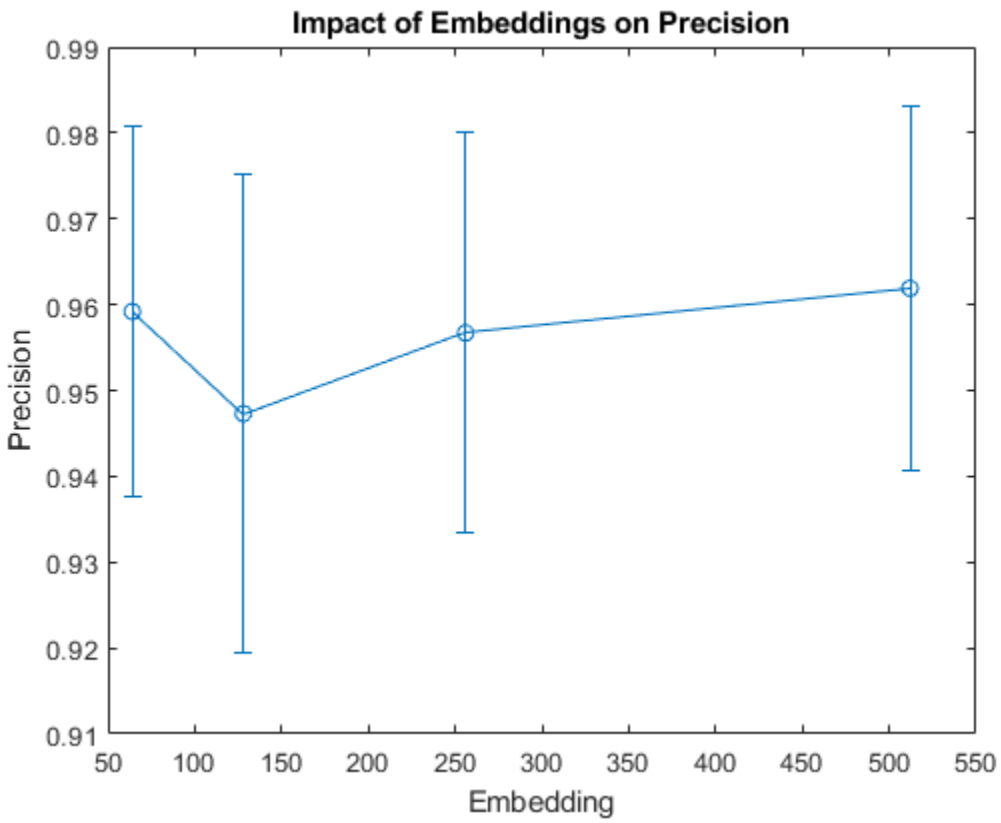


Figure 17. Impact of Embeddings on Precision

Similarly, to classification accuracy, the impact of embeddings seem to not affect the precision in a significant manner. While the outlier of 128 embedding performing slightly worse in average compared to the others, the figure overall shows more processing power does not necessarily mean better precision scores.

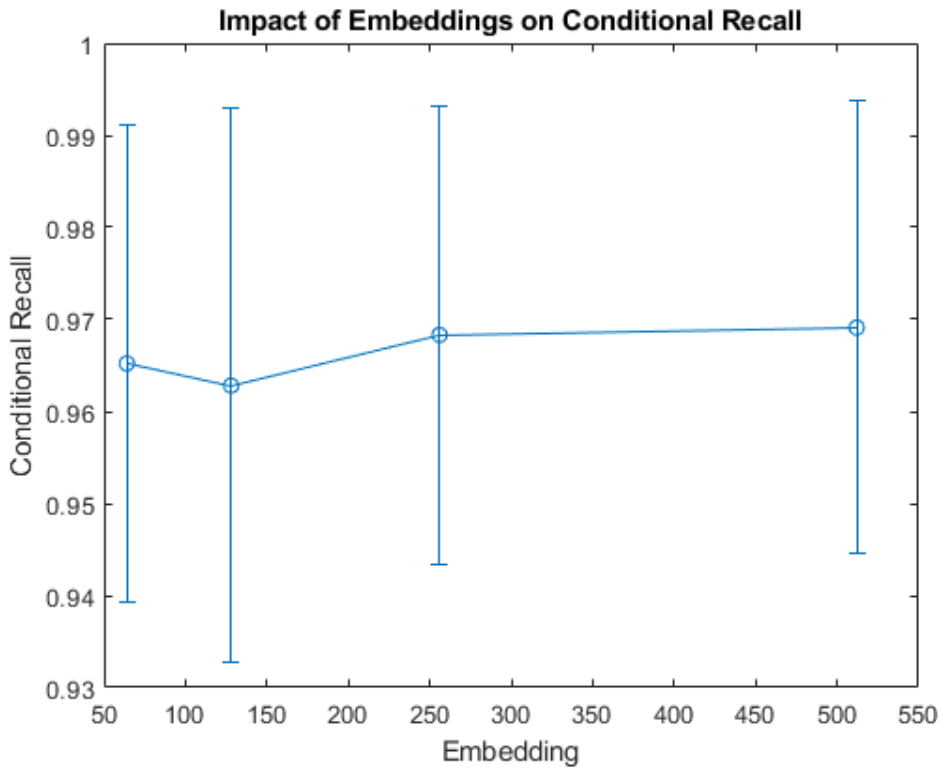


Figure 18. Impact of Embeddings on Conditional Recall

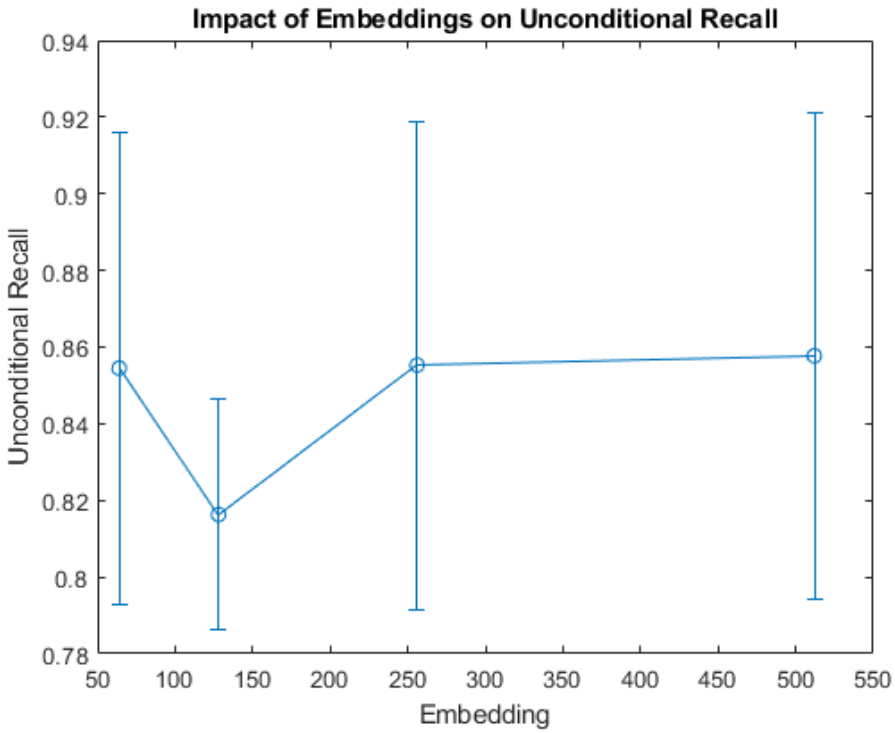


Figure 19. Impact of Embeddings on Unconditional Recall

Similarly to conditional accuracy and unconditional accuracy, recall seems to display the same behaviors of not having much variance aside from the unconditional recall performing much worse in average. A notable drop in recall score can be found in the 128 embedding, as it seems to perform poorer than the other embeddings, perhaps to due overfitting or the Resnet architecture causing conflict with that embedding.

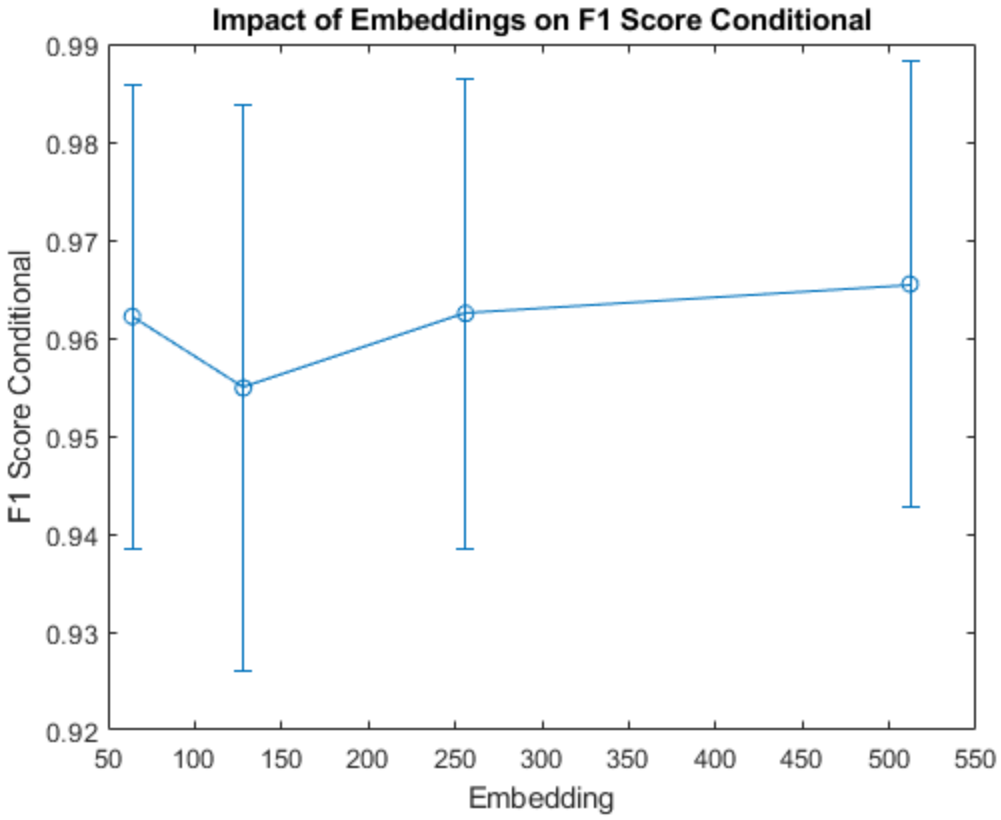


Figure 20. Impact of Embeddings on F1 Score Conditional

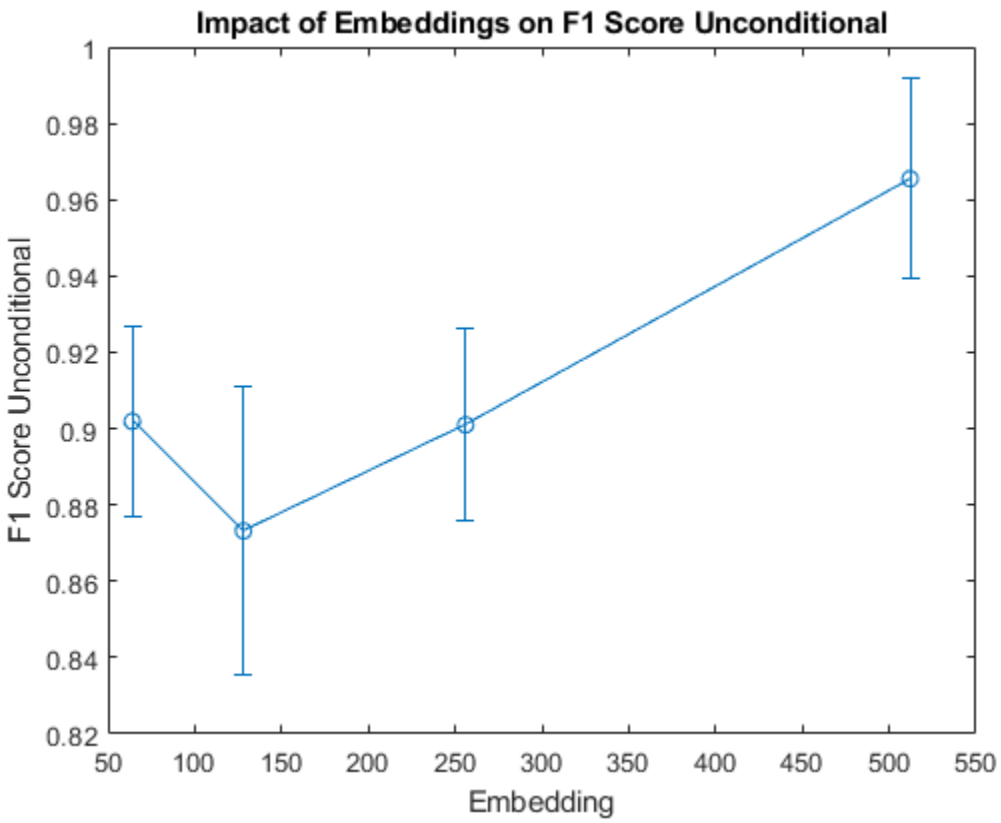


Figure 21. Impact of Embeddings on F1 Score Unconditional

The F1 scores for both conditional and unconditional show that the impact of uncertain images are definitely providing a larger difference in higher embeddings. 128 embedding being an anomaly, the overall trend is that higher embeddings seem to provide a better F1 score.

Interaction Effects

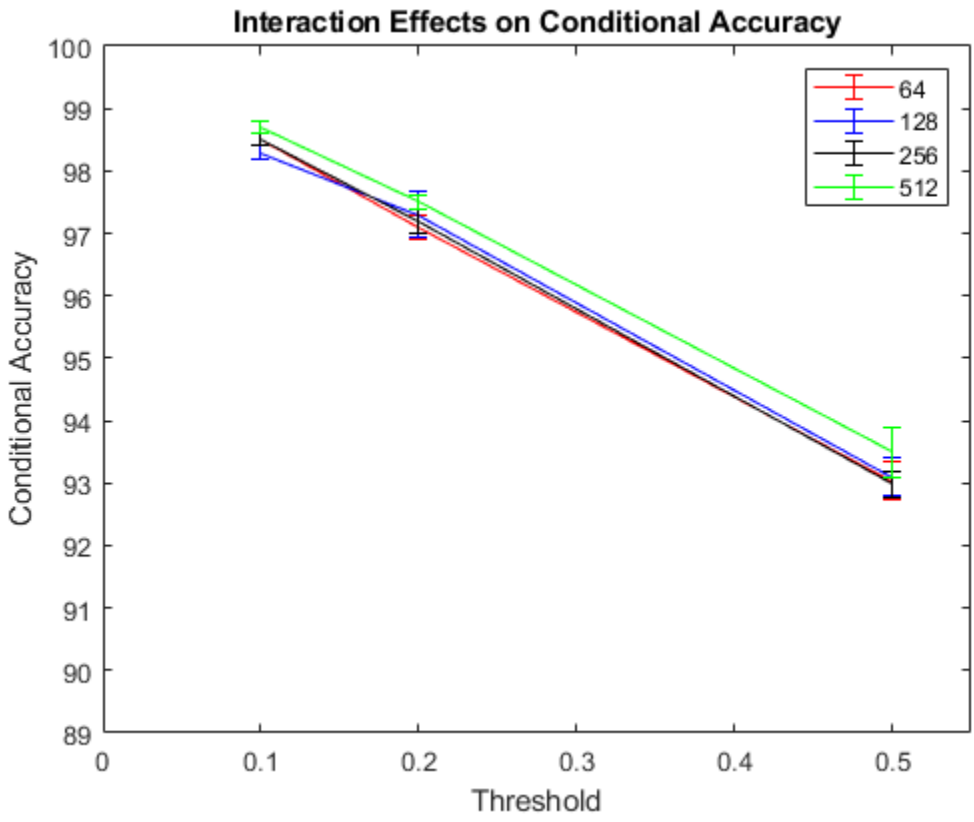


Figure 22. Interaction Effects on Conditional Accuracy of 0.1,0.2 and 0.5 thresholds

As we can see in Figure 21, the interaction effects of accuracy on 0.1,0.2, 0.5 thresholds shows that accuracy is generally higher on 512 embedding over 64, 128, 256 embedding, however since the two lines are almost parallel, there is no significant interaction.

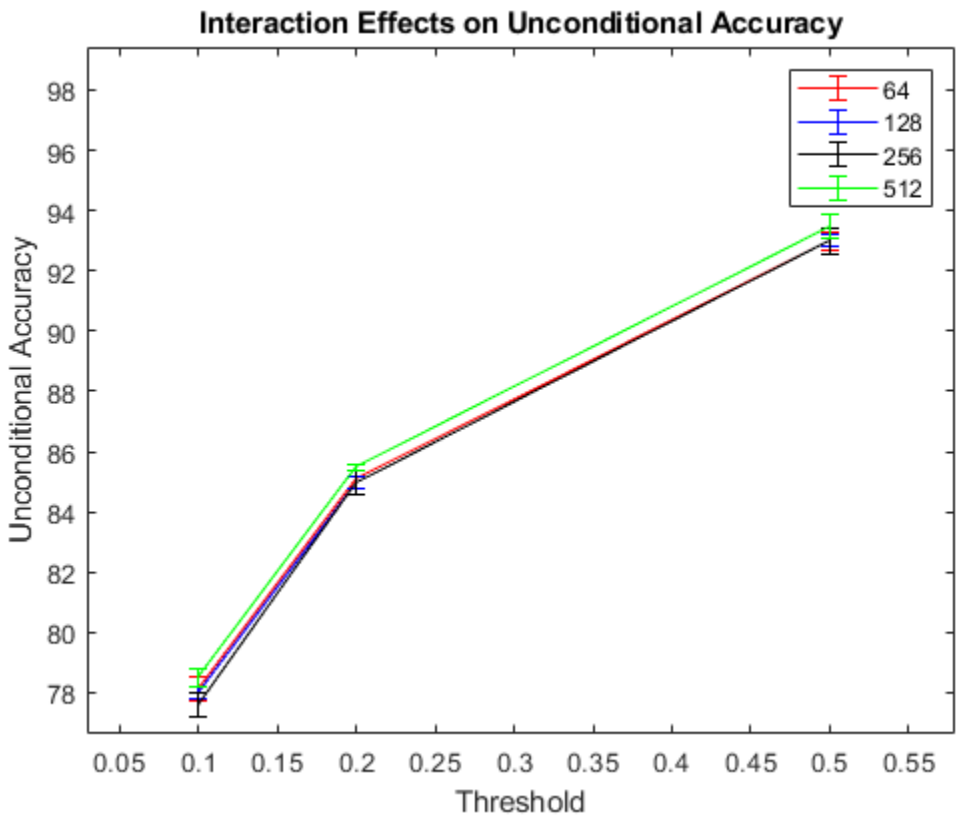


Figure 23. Interaction Effects on Unconditional Accuracy

On unconditional accuracy we see that a higher threshold results in higher unconditional accuracy. This is due to 0.5 threshold not having any uncertain values, resulting in the inverse graph from conditional accuracy. Similarly, since the lines are almost parallel, there is no significant interaction.

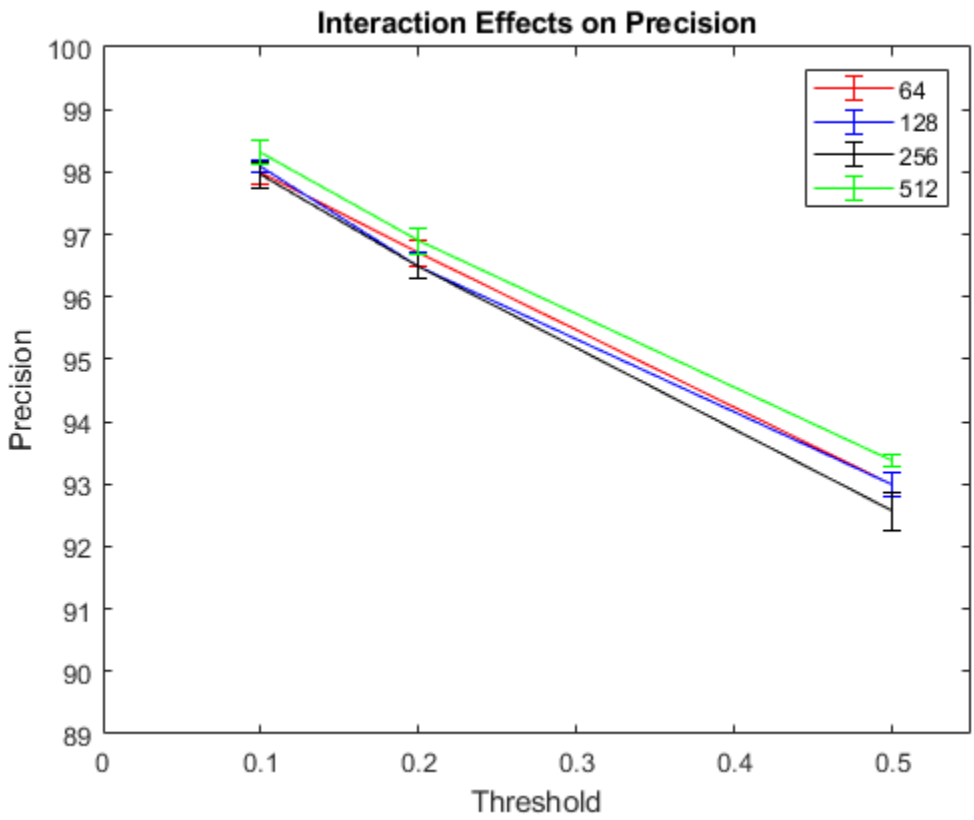


Figure 24. Interaction Effects on Precision of 0.1, 0.2, 0.5 thresholds

There is a minor interaction between the threshold and precision. The lower threshold value yields higher precision regardless of the embedding. Compared to accuracy, there is slightly more importance on choosing the correct embedding.

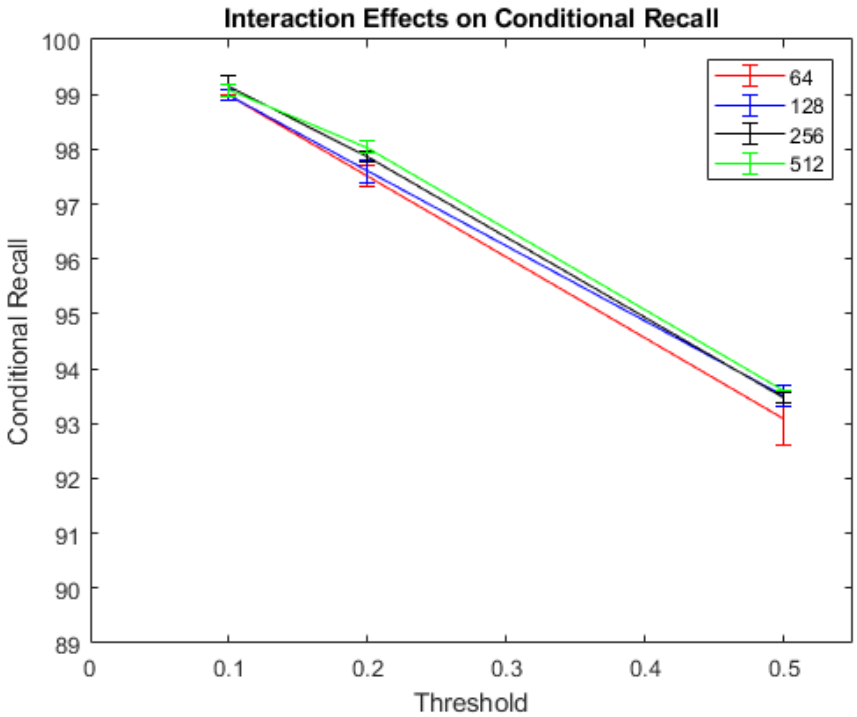


Figure 25. Interaction Effects on Conditional Recall

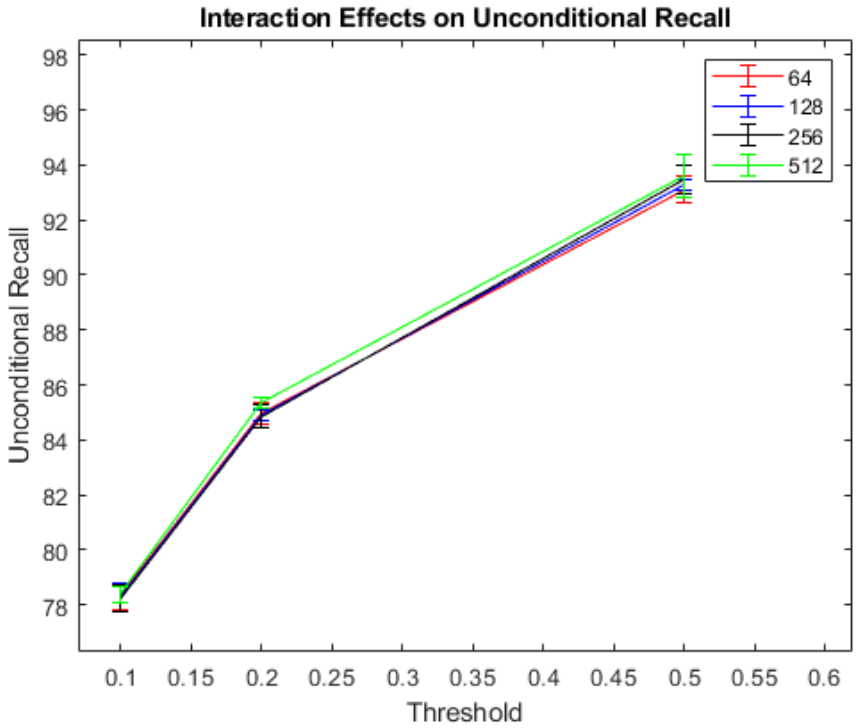


Figure 26. Interaction Effects on Unconditional Recall

Similarly to the results on accuracy, we can see the inverse relationship between conditional recall and unconditional recall due to the factor of uncertain images (images that could not be classified). The thresholds of 64, 128, 256 and 512 all behave in the same manner, and since the lines are parallel there does not seem to be any significant interaction between them.

4.2 Design of Experiments Results: Threshold

Since embeddings did not seem to affect the metrics for the network effectively, the tuning of the thresholds was focused on. Since the Siamese network outputs a probability of 0 or 1 indicating whether the two images are the same, values below the threshold are considered negatives (image does not match the anchor) and values above the 1-threshold are considered positives (image matches anchor image).

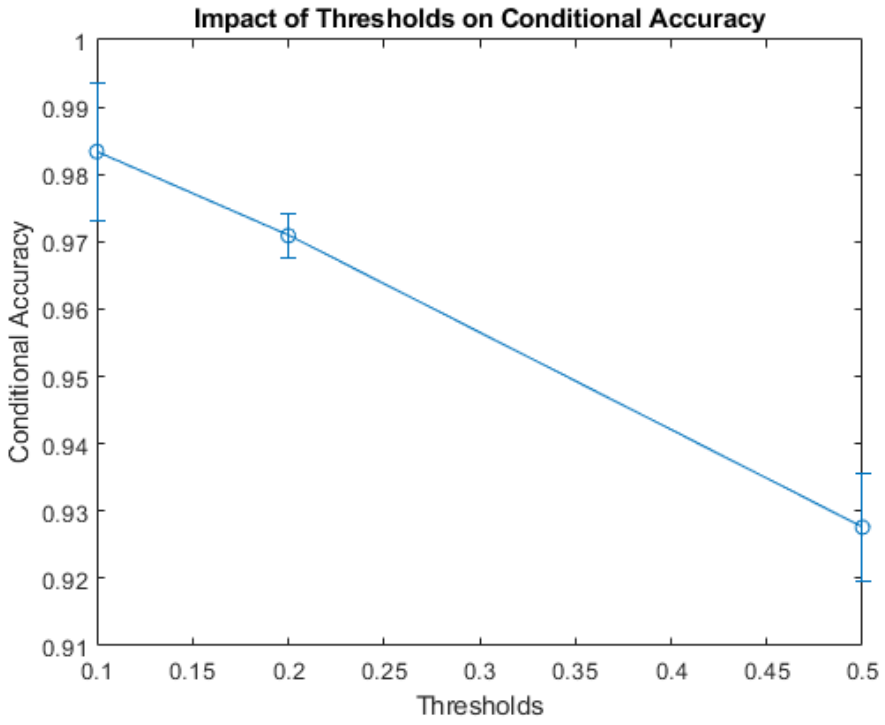


Figure 27. Impact of Threshold on Conditional Accuracy

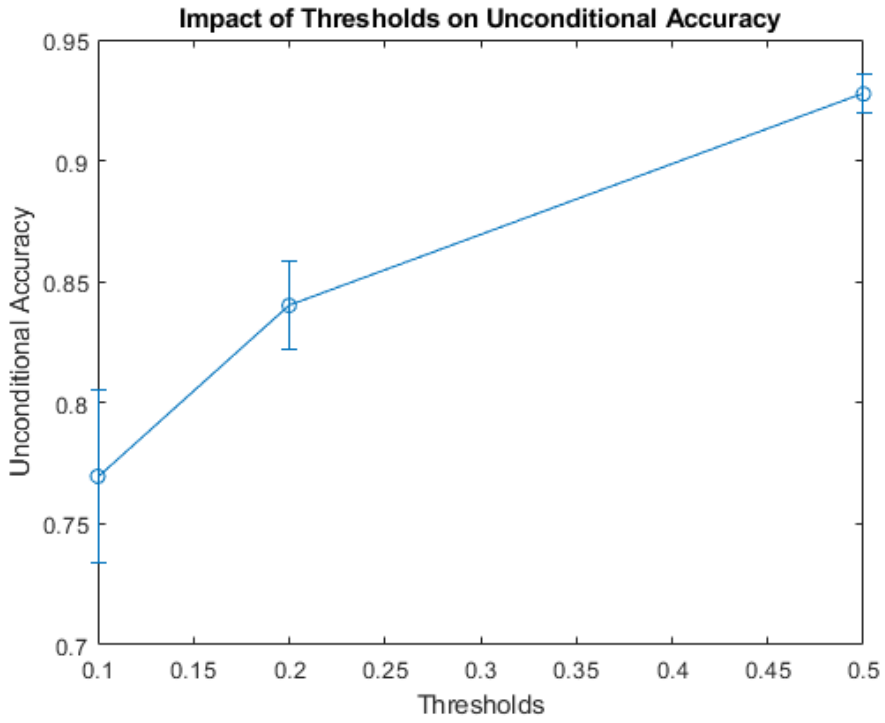


Figure 28. Impact of Threshold on Unconditional Accuracy

The threshold values of 0.1, 0.2 and 0.5 show that narrower values of the threshold result in higher accuracy scores. In terms of this research the tighter the threshold value, the more accurate the system becomes in recognizing fingerprints for what they are (classifying a right thumb as a right thumb). We can see the impact of the uncertain images in the inverse relationship between the conditional accuracy and unconditional accuracy results. Since 0.5 threshold does not have any uncertain values, the accuracy for that threshold is the highest in unconditional accuracy. Narrowing the threshold seems to have an adverse effect on the uncertain value. As we tighten the threshold values, the value of the uncertainty increases in conditional accuracy. This results in a higher portion of fingerprint images that remain unclassified or unrecognized by the neural network to belong to any specific class.

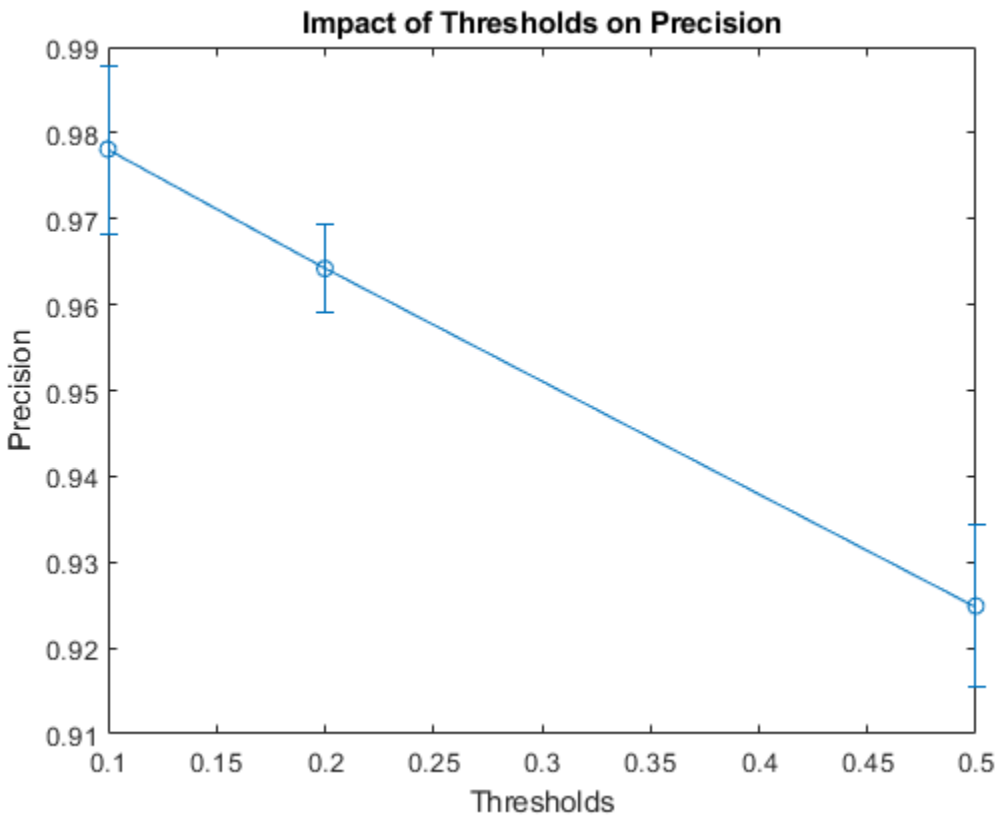


Figure 29. Impact of Threshold on Precision

The effect of thresholds is clearly seen in the results of precision, as the narrower the threshold, the better the precision score.

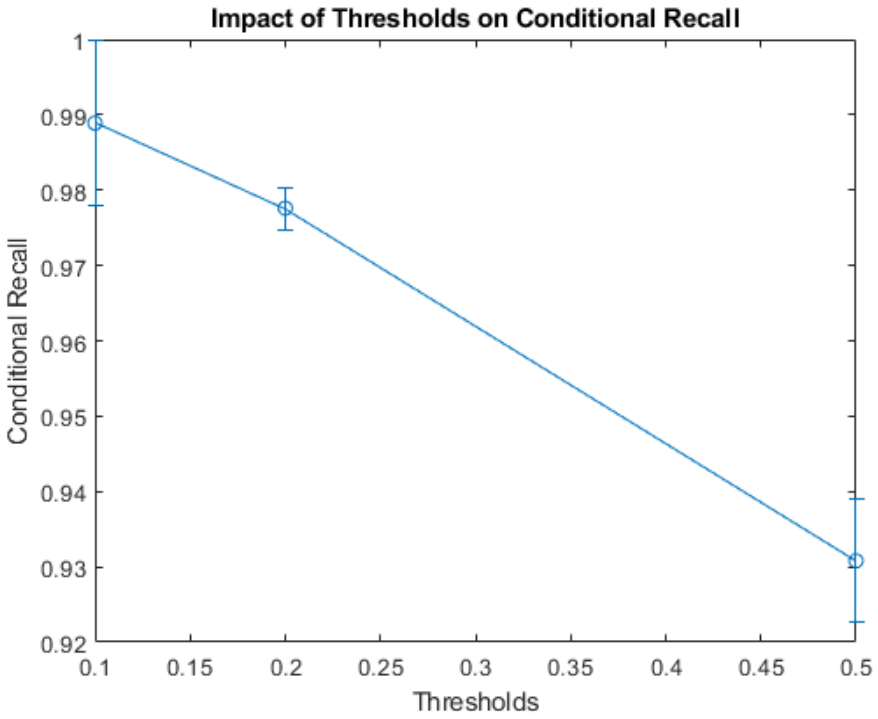


Figure 30. Impact of Threshold on Conditional Recall

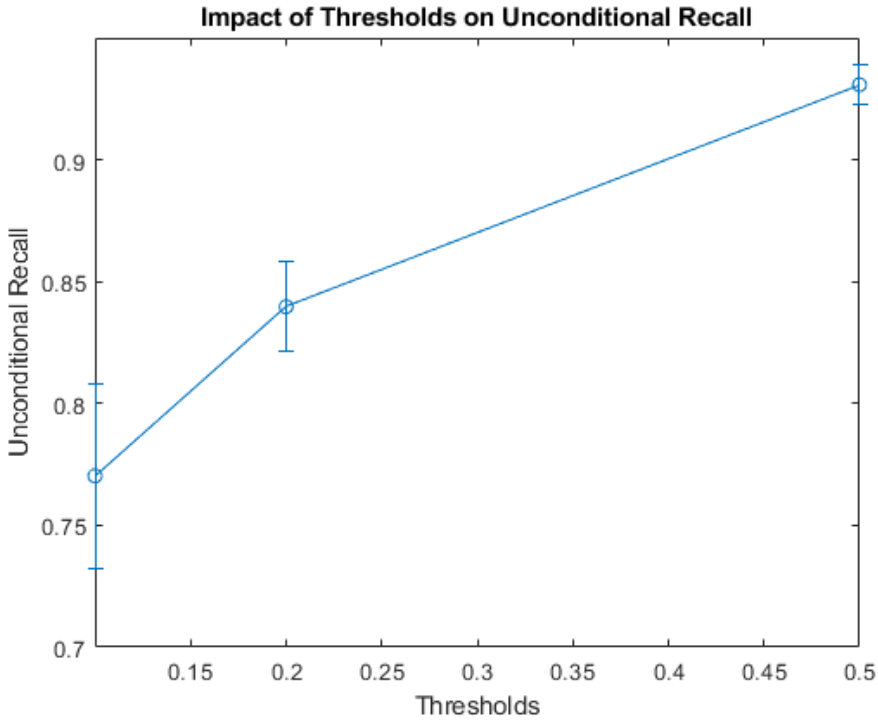


Figure 31. Impact of Threshold on Unconditional Recall

Recall attempts to answer the question of what proportion of actual positive images were classified correctly in the dataset. It has a similar relationship with the accuracy values as the lower threshold values provide a higher recall score. This shows how the neural network was correctly able to identify the fingerprint images that were actually of the class they were from. Similarly to the conditional/unconditional relationship of accuracy, the narrower the threshold the higher the accuracy in conditional recall, while a higher threshold value results in higher unconditional recall due to no uncertain images in the 0.5 threshold value.

Narrowing the thresholds produced a more significant improvement in achieving higher accuracy, precision, and recall/true positive rate. However, we need to consider the effects of conditional and unconditional factors as on one of them the uncertain images are considered.

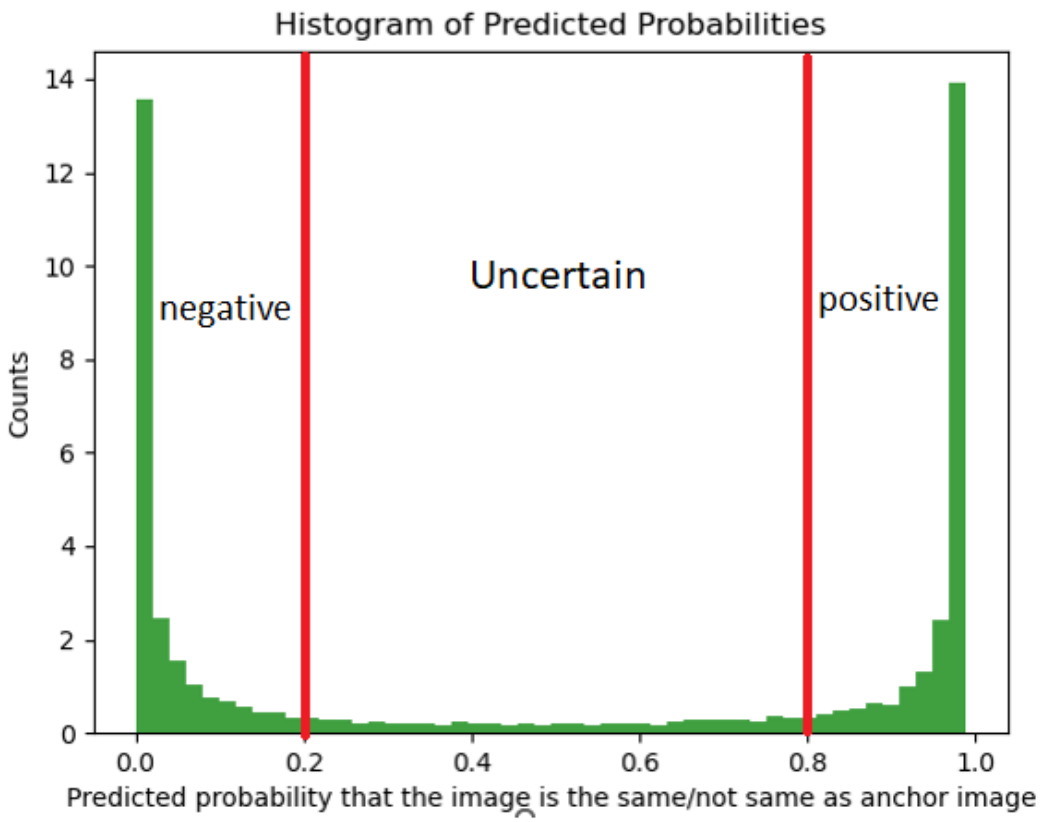


Figure 32. Combined Histogram of Predicted Probabilities of Results

Threshold	Accuracy	Uncertain	Recall/True Positive Rate
0.5	92.5%	0%	92.9%
0.2	96.8%	14%	96.4%
0.1	98.4%	24%	98.0%

The new model was able to achieve an accuracy of up to 98.4% compared to the old Finger Sequence network of 84% by implementing it into a Siamese network and fine tuning the threshold values.

4.5 Discussion

Changing the values of the embeddings from 64, 128 and 512 did not have a significant impact on classification accuracy and other metrics in the network. However, changing the threshold values into three subsets of 0.1, 0.2 and 0.5 displayed significant increase in accuracy when the values were narrowed. This in return increased the uncertain value from 0% to 24% from 0.5 threshold to 0.1 threshold. The uncertain factor is the number of images that fall in between the range between the threshold and 1-threshold. These are the predictions for which we do not have sufficient confidence to mark them as “positive” or “negative” resulting in them being “uncertain”. It is clearly seen on the conditional versus unconditional metrics of accuracy and recall as the uncertain factor drastically decreases the overall accuracy/recall. Smaller values for the threshold seem to result in higher accuracy and recall, but increasing the number of images that fall under “uncertain”. The tradeoff is between increasing the accuracy and precision while not labeling too many images as uncertain, as those images will have to be manually inspected or run through more sophisticated algorithms to classify them. 24% uncertainty may not seem a high value, but in a database which may have thousands or millions of fingerprints, that value is a high number of prints to manually inspect and resolve. While using this proof-of-concept network, in order to trade for higher accuracy, there needs to be an equal tradeoff in having a higher uncertain value.

Chapter 5: Conclusion

In the current modern era, fingerprints have become an invaluable tool for identification purposes in the field of biometrics. Classification of fingerprints have become a growing concern as large datasets in Department of Defense or other institutions have datasets where one misclassified image can render a whole archive to be useless. Advancements in neural networks have allowed for efficient scouring of such datasets for misclassified images and if needed manual inspection on those deemed to be unfit for testing purposes. The goal of this study was to use a Siamese network created to correctly identify labeled fingerprints and test how different embeddings, namely 64, 128, 256 and 512 and different thresholds affected the classification accuracy of the network to provide more information on how these factors play a large role in finding the most efficient network possible to reduce the need for manual inspection. The results obtained were accuracy, precision, recall and conditional/unconditional variants (factoring in uncertain which are images that could not be classified). The most notable results were that having a lower embedding produced better results, while having a lower threshold resulted in higher accuracy and precision at the cost of a higher amount of uncertainty. This is a tradeoff that will need to be considered if such a network was to be used for a commercial scale, as one may need to balance higher accuracy with a higher case of more unclassified images, resulting in those images processed through a different network or manually inspected in person. 98% accuracy may be an ideal percentage to chase for, but a 24% uncertain rate is an immense number of images if the dataset contains millions of images.

5.1 Future Work

This proof-of-concept network is a steppingstone in using neural networks to improve the classification accuracy for biometric databases to keep records of fingerprints accurate and legitimate. This idea is not limited to just fingerprints, as it can be used for face, iris and other modalities. In this work, 64,128, 256 and 512 embeddings along with 0.5, 0.2 and 0.1 threshold values were used. For future work a wide range of thresholds, embeddings and different datasets may be used to gauge the best settings to obtain the highest classification scores possible while keeping the uncertain score low. The final product will result in a lightweight, efficient tool that can be used for any dataset as a standard procedure before a dataset is finalized, being an invaluable tool in saving both time and costs of large data extraction operations. This work can be the steppingstone into commercial off the shelf sequence checking tools that become the industrial standard in maintaining dataset accuracy.

References

[1] S. Mayhew, “History of Biometrics,” *Biometric Update*, Jul. 20, 2018. https://www.biometricupdate.com/201802/history-of-biometrics-2

[2] A. Jain, R. Bolle, and S. Pankanti, “Introduction to Biometrics,” *Biometrics*, pp. 1–41, 1996, doi: https://doi.org/10.1007/0-306-47044-6_1.

[3] Jain, Anil K, P. J. Flynn, and A. A. Ross, *Handbook of Biometrics.* Gardners Books, 2010.

[4] J. Rohrlich, “Homeland Security will soon have biometric data on nearly 260 million people,” *Quartz*, Nov. 07, 2019. https://qz.com/1744400/dhs-expected-to-have-biometrics-on-260-million-people-by-2022 (accessed Mar. 21, 2023).

[5] U. Rao and V. Nair, “Aadhaar: Governing with Biometrics,” *South Asia: Journal of South Asian Studies*, vol. 42, no. 3, pp. 469–481, May 2019, doi: https://doi.org/10.1080/00856401.2019.1595343.

[6] J. Lynch, “HART: Homeland Security’s Massive New Database Will Include Face Recognition, DNA, and Peoples’ ‘Non-Obvious Relationships,’” *Electronic Frontier Foundation*, Jun. 07, 2018. https://www.eff.org/deeplinks/2018/06/hart-homeland-securitys-massive-new-database-will-include-face-recognition-dna-and

[7] R. Padmavathi, K.M Mohammed Azeezulla, P. Venkatesh, K. S. Mahato, and G. Nithin, “Digitalized Aadhar enabled ration distribution using smart card,” *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*, May 2017, doi: https://doi.org/10.1109/rteict.2017.8256670.

[8] A. Siddiqui, “IITM Journal of Management and IT SOUVENIR National Conference

on Emerging Trends in Information Technology- Cyber Security: A Panoramic View," National Conference on Emerging Trends in Information Technology. Accessed: Mar. 12, 2023. [Online]. Available: https://iitmjp.ac.in/wp-content/uploads/2017/06/IT-Conference-2015.pdf#page=40

[9] H. Rudolph, L. Moy, and A. Bedoya, "Airport Face Scans: An Investigation," *www.airportfacescans.com*, Dec. 17, 2017. https://www.airportfacescans.com/ (accessed Mar. 21, 2023).

[10] P. Higgins, "Standards for the electronic submission of fingerprint cards to the FBI," *Journal of Forenseic Identification*, vol. 45, no. 4, pp. 409–418, 1995.

[11] K. E. Wertheim, "Human Factors in Large-Scale Biometric Systems: A Study of the Human Factors Related to Errors in Semiautomatic Fingerprint Biometrics," *IEEE Systems Journal*, vol. 4, no. 2, pp. 138–146, Jun. 2010, doi: https://doi.org/10.1109/jsyst.2010.2049878.

[12] "What Is Data Deduplication? | Benefits & Use Cases | NetApp," *www.netapp.com*. https://www.netapp.com/data-management/what-is-data-deduplication/

[13] P. S. Sudhish, A. K. Jain, and K. Cao, "Adaptive fusion of biometric and biographic information for identity de-duplication," *Pattern Recognition Letters*, vol. 84, pp. 199–207, Dec. 2016, doi: https://doi.org/10.1016/j.patrec.2016.10.011.

[14] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: https://doi.org/10.1126/science.aaa8415.

[15] H. Lin and A. Jain, "Classification of fingerprint images," vol. 2, pp. 665–672, 1999, Available: https://www.researchgate.net/profile/Lin-Hong-

12/publication/2929466_Classification_of_Fingerprint_Images/links/54733a5a0cf216f8cfaeb8b7/Classification-of-Fingerprint-Images.pdf

[16] K. Fukushima, “A neural network for visual pattern recognition,” *Computer*, vol. 21, no. 3, pp. 65–75, Mar. 1988, doi: https://doi.org/10.1109/2.32.

[17] D. O. Hebb, *The organization of behavior : a neuropsychological theory*. New York: Routledge, Taylor & Francis Group, 2012.

[18] F. Rosenblatt, *The Perceptron*. Ithaca, New York: Cornell Aeronautical Laboratory, 1958, pp. 85–4601.

[19] Q. Zhang and H. Yan, “Fingerprint classification based on extraction and analysis of singularities and pseudo ridges,” *Pattern Recognition*, vol. 37, no. 11, pp. 2233–2243, Nov. 2004, doi: https://doi.org/10.1016/j.patcog.2003.12.020.

[20] D. Michelsanti, A.-D. Ene, Y. Guichi, R. Stef, K. Nasrollahi, and T. B. Moeslund, “Fast Fingerprint Classification with Deep Neural Networks,” *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2017, doi: https://doi.org/10.5220/0006116502020209.

[21] R. Mary. Lourde and D. Khosla, “Fingerprint Identification in Biometric SecuritySystems,” *International Journal of Computer and Electrical Engineering*, pp. 852–855, 2010, doi: https://doi.org/10.7763/ijcee.2010.v2.239.

[22] B. Pandya, G. Cosma, A. A. Alani, A. Taherkhani, V. Bharadi, and T. M. McGinnity, “Fingerprint classification using a deep convolutional neural network,” *2018 4th International Conference on Information Management (ICIM)*, pp. 86–91, May 2018, doi: https://doi.org/10.1109/infoman.2018.8392815.

[23] R. A. Searston and J. M. Tangen, “The Emergence of Perceptual Expertise with

Fingerprints Over Time," *Journal of Applied Research in Memory and Cognition*, vol. 6, no. 4, pp. 442–451, Dec. 2017, doi: https://doi.org/10.1016/j.jarmac.2017.08.006.

[24] H. C. Lee and R. E. Gaensslen, *Advances in fingerprint technology*. Boca Raton, Fla.: Crc Press, 2001.

[25] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pp. 253–256, May 2010, doi: https://doi.org/10.1109/iscas.2010.5537907.

[26] A. K. Agrawal and Y. N. Singh, "Evaluation of Face Recognition Methods in Unconstrained Environments," *Procedia Computer Science*, vol. 48, pp. 644–651, 2015, doi: https://doi.org/10.1016/j.procs.2015.04.147.

[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv.org*, 2014. https://arxiv.org/abs/1409.1556

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Jun. 2016, doi: https://doi.org/10.1109/cvpr.2016.90.

[29] H. Zheng, J. Fu, T. Mei, and J. Luo, "Learning Multi-attention Convolutional Neural Network for Fine-Grained Image Recognition," *IEEE Xplore*, Oct. 01, 2017. https://ieeexplore.ieee.org/document/8237819 (accessed Apr. 07, 2023).

[30] V. Meshram and K. Patil, "Border-Square net: a robust multi-grade fruit classification in IoT smart agriculture using feature extraction based Deep Maxout network," *Multimedia Tools and Applications*, May 2022, doi: https://doi.org/10.1007/s11042-022-12855-7.

[31] "Max-pooling / Pooling - Computer Science Wiki," *Computersciencewiki.org*, 2018.

https://computersciencewiki.org/index.php/Max-pooling_/_Pooling

[32] Q. Zheng, M. Zhang, and J. Yang, “A Bilinear Multi-Scale Convolutional Neural Network for Fine-grained Object Classification,” *IAENG International Journal of Computer Science*, vol. 45, no. 2, 2018, Available: https://www.iaeng.org/IJCS/issues_v45/issue_2/IJCS_45_2_12.pdf

[33] K. Sai Prasad and D. Pasupathy, “DEEP CONVOLUTIONAL NEURAL NETWORK IMPLEMENTATIONS FOR EFFICIENT BRAIN STROKE DETECTION USING MRI SCANS,” *Journal of Theoretical and Applied Sciences*, vol. 100, no. 17, 2022, Accessed: Apr. 24, 2023. [Online]. Available: http://www.jatit.org/volumes/Vol100No17/14Vol100No17.pdf

[34] J. Bromley *et al.*, “SIGNATURE VERIFICATION USING A ‘SIAMESE’ TIME DELAY NEURAL NETWORK,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 07, no. 04, pp. 669–688, Aug. 1993, doi: https://doi.org/10.1142/s0218001493000339.

[35] S. K. Roy, M. Harandi, R. Nock, and R. Hartley, “Siamese Networks: The Tale of Two Manifolds,” *openaccess.thecvf.com*, 2019. https://openaccess.thecvf.com/content_ICCV_2019/html/Roy_Siamese_Networks_The_Tale_of_Two_Manifolds_ICCV_2019_paper.html (accessed Apr. 24, 2023).

[36] Google, “Classification: Accuracy | Machine Learning Crash Course,” *Google Developers*, 2019. https://developers.google.com/machine-learning/crash-course/classification/accuracy

[37] Google, “Classification: Precision and Recall | Machine Learning Crash Course | Google Developers,” *Google Developers*, Mar. 05, 2019.

https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

[38] J. Korstanje, “The F1 score,” *Medium*, Aug. 31, 2021. https://towardsdatascience.com/the-f1-score-bec2bbc38aa6

[39] K. M. Bower, “What Is Design of Experiments (DOE)? | ASQ,” *Asq.org*, 2019. https://asq.org/quality-resources/design-of-experiments