

MULTIMODAL PRE-TRAINED MODEL TO GENERATE FEATURE FUSION
FOR ENHANCING ONLINE BANKING SECURITY

TEH HUNG WEI

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA

**DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND
COPYRIGHT**

Author's full name : TEH HUNG WEI

Date of Birth : 29 – 09 – 1997

Title : MULTIMODAL PRE-TRAINED MODEL TO GENERATE FEATURE
FUSION FOR ENHANCING ONLINE BANKING SECURITY

Academic Session :

I declare that this thesis is classified as:

☐**CONFIDENTIAL** (Contains confidential information under the
Official Secret Act 1972)*☐**RESTRICTED** (Contains restricted information as specified by
the organization where research was done)*☒**OPEN ACCESS** I agree that my thesis to be published as online
open access (full text)

1. I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:
2. The thesis is the property of Universiti Teknologi Malaysia
3. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
4. The Library has the right to make copies of the thesis for academic exchange.



SIGNATURE OF STUDENT

MCS212011

MATRIC NUMBER

Certified by:



SIGNATURE OF SUPERVISOR

DR. NUR EILYAH @WONG YEE
LENG


NAME OF SUPERVISOR

Date: 12 OCTOBER 2023

Date: 12 OCTOBER 2023

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

“I hereby declare that I have read this thesis and in my
opinion this thesis is sufficient in term of scope and quality for the
award of the degree of Master of Computer Science

Signature : 

Name of Supervisor I : DR. NUR ELIYAH @ WONG YEE LENG

Date : 12 OCTOBER 2023

MULTIMODAL PRE-TRAINED MODEL TO GENERATE FEATURE FUSION
FOR ENHANCING ONLINE BANKING SECURITY

TEH HUNG WEI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Computer Science

Faculty of Computing
Universiti Teknologi Malaysia

OCTOBER 2023

DECLARATION

I declare that this thesis entitled “*Multimodal Pre-trained Model To Generate Feature Fusion For Enhancing Online Banking Security*” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature

:



.....

Name

: TEH HUNG WEI

Date

: 12 OCTOBER 2023

DEDICATION

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr. Nur Eiliyah @ Wong Yee Leng, for encouragement, guidance, critics and friendship. Without her continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Teknologi Malaysia (UTM) for funding my Master study. Librarians at UTM also deserve special thanks for their assistance in supplying the relevant literatures.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

ABSTRACT

Nowadays, the trend of e-commerce become more significant. One of the biggest benefits is it provides fast and convenient transaction among users at any time and location. For instance, in Malaysia, there are several platforms, such as Touch n Go eWallet, CIMB clicks, Maybank2U and etc. By knowing the other user phone number or bank account number, money can be simply transferred to other users in just several clicks. However, hackers tend to bypass the security systems or hack users account by knowing the private information of the account user. In the worst scenario, sometimes user's assets even stolen through online without any notification. This had brought great loss of assets throughout these years. Traditional approaches such as password, security tokens, two-authentication method had slowly become ineffective against these security threats. More methods such as unimodal biometric authentication method and machine learning were applied in detecting fraudulent activities and user recognition system. Although the results were better as compared to traditional approaches, but still, it is not robust enough against these security threats. Nowadays, recognition system was using multimodal biometrics as it provides higher performance. In this research work, the idea of applying transfer learning in building robust user recognition model was proposed. Furthermore, the datasets were obtained from CASIA biometric database and the selected multimodal biometrics for this research work were iris, face and palmprint. Next, these multimodalities were combined using feature fusion technique. Lately, two classification models were trained by adding fully-connected layer before the output layer. The first and second model were trained to classify 100 and 300 users respectively. Lastly, the proposed models showed high accuracy rate of 99.60 %, and 97.60 % respectively.

ABSTRAK

Pada masa kini, trend e-dagang menjadi semakin signifikan. Salah satu manfaat terbesar adalah ia menyediakan urusan niaga yang pantas dan mudah di antara pengguna pada bila-bila masa dan di mana-mana sahaja. Contohnya, di Malaysia, terdapat beberapa platform seperti Touch n Go eWallet, CIMB clicks, Maybank2U, dan sebagainya. Dengan mengetahui nombor telefon atau nombor akaun bank pengguna lain, wang boleh dipindahkan dengan mudah ke pengguna lain hanya dalam beberapa klik. Walau bagaimanapun, penjenayah cenderung untuk mengelakkan sistem keselamatan atau menggodam akaun pengguna dengan mengetahui maklumat peribadi pengguna akaun. Dalam skenario terburuk, terkadang aset pengguna dicuri secara dalam talian tanpa sebarang pemberitahuan. Ini telah membawa kepada kerugian besar aset selama bertahun-tahun. Pendekatan tradisional seperti kata laluan, token keselamatan, dan kaedah autentikasi dua faktor secara perlahan-lahan menjadi tidak berkesan terhadap ancaman keselamatan ini. Lebih banyak kaedah seperti kaedah autentikasi biometrik unimodal dan pembelajaran mesin digunakan dalam pengesanan aktiviti menipu dan sistem pengenalan pengguna. Walaupun hasilnya lebih baik berbanding pendekatan tradisional, namun ia masih belum cukup kuat menghadapi ancaman keselamatan ini. Pada masa kini, sistem pengenalan menggunakan biometrik multimodal kerana ia memberikan prestasi yang lebih tinggi. Dalam kerja penyelidikan ini, gagasan penggunaan pembelajaran pemindahan dalam membina model pengenalan pengguna yang kuat dicadangkan. Selanjutnya, dataset diperolehi dari pangkalan data biometrik CASIA dan biometrik multimodal yang dipilih untuk kerja penyelidikan ini adalah iris, wajah, dan cap jari tangan. Kemudian, multimodal ini digabungkan menggunakan teknik perpaduan ciri. Akhirnya, dua model klasifikasi dilatih dengan menambahkan lapisan sepenuhnya disambungkan sebelum lapisan output. Model pertama dan kedua dilatih untuk mengklasifikasikan 100 dan 300 pengguna masing-masing. Akhirnya, model yang dicadangkan menunjukkan kadar ketepatan yang tinggi, iaitu 99.60 % dan 97.60 % masing-masing.

TABLE OF CONTENTS

	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST of tables	ix
	LIST OF FIGURES	xi
	LIST OF ABBREVIATIONS	xiii
	LIST OF APPENDICES	xv
CHAPTER 1	INTRODUCTION	1
1.1	Introduction	1
1.2	Problem Background	1
1.3	Problem Statement	5
1.4	Research Purpose	6
1.5	Research Objectives	6
1.6	Research Scope	7
1.7	Research Significance	7
1.8	Research Contributions	8
1.9	Organization of the Thesis	8
CHAPTER 2	LITERATURE REVIEW	9
2.1	Introduction	9
2.2	Overview of Online Banking Security	11
2.3	Traditional Approaches for Online Banking Security	12
2.3.1	Password	12
2.3.2	Security Tokens	13

2.3.3	Two-factor Authentication	13
2.4	Biometric Authentication	14
2.5	Multiple Biometric Authentication	16
2.6	Machine Learning	20
2.7	Deep Learning	21
2.8	Transfer Learning	22
2.8.1	Feature Extraction	24
2.8.2	Fully Connected Layer	27
2.8.3	State-of-the-Art Architectures	27
2.8.3.1	VGG16 Model	29
2.8.3.2	ResNet-50 Model	31
2.8.3.3	InceptionResNetV1 Model	34
2.9	Modality Fusion	37
2.10	Related Work	41
2.10.1	Related Work for Banking Security	41
2.10.2	Related Work for Multimodal Biometric Authentication	44
2.10.3	Overview of Biometric Authentication Approach	51
2.10.4	Summary	53
2.11	Research Gaps	53
2.12	Chapter Summary	54
CHAPTER 3	RESEARCH METHODOLOGY	56
3.1	Introduction	56
3.2	Research Workflow	56
3.2.1	Phase 1: Problem Background Analysis	57
3.2.2	Phase 2: Literature Review	58
3.2.3	Phase 3: Experiment and Evaluation	59
3.2.3.1	Data Collection	60
3.2.3.2	Image Pre-Processing	61
3.2.3.3	Transfer Learning	62
3.2.3.4	Modality Fusion	66

3.2.3.5	Classification	67
3.2.3.6	Model Evaluation	67
3.2.4	Phase 4: Result Documentation	72
3.3	Chapter Summary	73
CHAPTER 4	RESEARCH IMPLEMENTATION	74
4.1	Introduction	74
4.2	Preparation Stage	74
4.3	Proposed Method & Algorithms	76
4.3.1	Data Pre-processing & Augmentation	77
4.3.2	Transfer Learning	81
4.3.3	Modality Fusion & Classification stage	82
4.3.4	Model Evaluation	84
4.4	Chapter Summary	85
CHAPTER 5	RESEARCH RESULTS AND DISCUSSION	86
5.1	Introduction	86
5.2	Preparation Stage Results	86
5.2.1	Iris Image Processing	87
5.2.2	Face Image Processing	88
5.2.3	Palmprint Image Processing	89
5.2.4	Data Augmentation	89
5.3	Model Evaluation	93
5.4	Models Comparison	101
5.5	Chapter Summary	102
CHAPTER 6	CONCLUSION	103
6.1	Introduction	103
6.2	Achievement of Research Objectives	103
6.3	Summary of Contribution	104
6.4	Suggestions for Improvement for Future Works	105
REFERENCES		106

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Comparisons Between Iris Recognition, Face Recognition, and Palmprint Recognition in terms of Pros, Cons, and Challenges	15
Table 2.2	Existing Multimodal Biometric Authentication Systems	18
Table 2.3	Comparisons of Different Machine Learning Based Feature Extraction Techniques Based on Pros and Cons	24
Table 2.4	Summary of VGG16 and the Model Configurations	30
Table 2.5	Drawbacks for Early Fusion, Feature Fusion, and Score Fusion Respectively	40
Table 2.6	Related Work for Banking Security	43
Table 2.7	Comparison of Models	45
Table 3.1	Overall Summary	62
Table 3.2	VGG16 Model Layers and Configurations	63
Table 3.3	ResNet101V2 Model Layers and Configurations	64
Table 3.4	InceptionResNetV2 Model Layers and Configurations	65
Table 3.5	Configurations at Fully Connected Layers	66
Table 3.6	Confusion Matrix	68
Table 5.1	Summarized Results for Training, Validation, and Testing Datasets for Model_100	91
Table 5.2	Summarized Results for Training, Validation, and Testing Datasets for Model_300	92
Table 5.3	Results for Model_100	94
Table 5.4	Results for Model_300	97
Table 5.5	Comparison Results between Model_100 and Model_300	100
Table 5.6	Comparison Between Different Models Applied related to Banking Security Domain with Proposed Model	101

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	Literature Map	10
Figure 2.2	Generic Process Flow of Transfer Learning	23
Figure 2.3	Typical Example of CNN Feature Extraction Layer (<i>Convolutional Neural Network: An Overview</i> , 2022)	26
Figure 2.4	Typical Example of Fully Connected Layer (<i>Convolutional Neural Network: An Overview</i> , 2022)	27
Figure 2.5	Summary of ResNet-50 Architecture	32
Figure 2.6	ResNet-50 Architecture (a) Overall Structure (b) Description of Blocks and lines (c) Connector for Identity Block (d) Connector for Convolutional Block	33
Figure 2.7	Overall Structure of InceptionResNet_V1	35
Figure 2.8	Different modules within InceptionResnet_V1 (a) Stem Module (b) Inception-Resnet-A Module (c) Reduction-A Module (d) Inception-Resnet-B Module (e) Reduction-B Module (f) Inception-ResNet-C Module (Szegedy et al., 2017)	37
Figure 2.9	Early Fusion Process	38
Figure 2.10	Feature Fusion Process	38
Figure 2.11	Score Fusion Process	39
Figure 3.1	Research Framework	57
Figure 3.2	Research Methodology	59
Figure 3.3	Image Preprocessing Steps for Iris, Face and Palmprint Dataset	61
Figure 3.4	Example of ROC curve (a) Ideal Case (b) Bad Case (c) Worst Case (<i>Understanding AUC - ROC Curve / by Sarang Narkhede / Towards Data Science</i> , n.d.)	70
Figure 4.1	Device Specifications (CPU)	75
Figure 4.2	Google Colab Runtime Settings	75
Figure 4.3	Proposed Model	76
Figure 5.1	Iris Image Processing Result (i) Before (ii) After	87

Figure 5.2	Face Image Processing Result (i) Before (ii) After	88
Figure 5.3	Palmprint Image Processing Result(i) Before (ii) After	89
Figure 5.4	Model_100 Fitting Result Using Callbacks	93
Figure 5.5	Model_100 Finalized Result	93
Figure 5.6	Model_300 Fitting Result Using Callbacks	96
Figure 5.7	Model_300 Finalized Result	96
Figure 6.1	Gantt Chart for Research Proposal	113
Figure 6.2	Gantt Chart for Research Thesis	113

LIST OF ABBREVIATIONS

2FA	-	Two Factor Authentication
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
AUC	-	Area Under Curve
CTC	-	Connectionist Temporal Classification
CM	-	Confusion Matrix
CNN	-	Convolutional Neural Network
DCNN	-	Deep Convolutional Neural Network
DL	-	Deep Learning
DLCNN	-	Deep Learning Convolutional Neural Network
DT	-	Decision Tree
EBC	-	Extended Borda-Count
EER	-	Equal Error Rate
ELM	-	Extreme Learning Machine
FAR	-	False Acceptance Rate
FC	-	Fully-Connected
FN	-	False Negative
FNR	-	False Negative Rate
FP	-	False Positive
FPR	-	False Positive Rate
GLCM	-	Grey Level Co-occurrence Matrix
HAF	-	Hybrid Adaptive Features
HMSB	-	Histogram-based Multi-Scale Binary Pattern
IDA	-	Independent Discriminant Analysis
ILSVRC	-	ImageNet Large Scale Visual Recognition Challenge
IPCA	-	Improved Principal Component Analysis
KNN	-	K-Nearest Neighbors
LBP	-	Local Binary Pattern
LDA	-	Linear Discriminant Analysis
LBPH	-	Long Short-Term Memory

ML	-	Machine Learning
MRG	-	Modified Region Growing
NB	-	Naïve Bayes
NN	-	Neural Network
OGWO	-	Gray-wolf Optimization
OTP	-	One-Time Password
PCA	-	Principal Component Analysis
RNN	-	Recurrent Neural Network
ROC	-	Receiver Operating Curve
ROI	-	Region of Interest
RF	-	Random Forest
SIFT	-	Scale-invariant Feature Transform
SVM	-	Support Vector Machine
TL	-	Transfer Learning
TP	-	True Positive
TPR	-	True Positive Rate
TN	-	True Negative

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Link for Code	113
Appendix B	Gantt Chart	113

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter firstly began with the problem background for the research work. Next, the problem statement, research purpose and objectives will be shown accordingly. Afterwards, the research scope, research significance, and research contribution will be clearly explained. Lastly, the organization of the thesis will be discussed in brief.

1.2 Problem Background

Nowadays, internet has become inseparable from human life (Mallet et al., 2022). Mostly everyone is having their own electrical device to gain access to the internet. At the same time, the trend of online banking has become crucial to everyone. Everyone tends to pay with cashless method, as it provides a convenient way to access their financial accounts (Malinka et al., 2022). Moreover, people are able to make transaction anytime from anywhere through internet. In the future, it was believed the trends would continue to grow and they predicted that the transaction platform would be completely performed in cashless environment.

However, as online banking become more popular, it also become more vulnerable to security attacks at the same time (Dhoot et al., 2020). These attacks such as phishing attacks, malware, identity theft, fraud, and unauthorized access can compromise the confidentiality, integrity, and availability of financial information (Buckner et al., 2016). In which it leads to great financial losses and damage to reputation.

Throughout these years, various methods have been implemented to resolve these security threats, such as passwords, security tokens, and two-factor authentication (Sengar et al., 2020). These methods can be considered as the most basic approach in addressing these security threats. Well, all of these methods were efficient at the beginning, but the efficiency is reducing with time.

For password, it was one of the general methods to be used to protect user account. However, it is quite hard to create a strong password yet memorable one (Buckley & Nurse, 2019). If the password too complex, people tend to forget. If the password too simple, it will be very easy to get bypassed by the other parties. According to study by SplashData (*100 Most Common Passwords Of 2022. Can You Spot Your Password?*, 2022), ‘123456’, ‘123456789’, ‘qwerty’, and ‘password’ are still the most commonly used passwords in 2022. These common passwords are easily misused by attackers to steal valuable asset from users.

Security tokens are physical devices that generate one-time passcode that must be entered along with a password to access an account. It is stronger and more secure than passwords alone. However, it can be lost or stolen. Sometimes, it will not be convenient to carry them, and it can be annoying for the users (Khanaa et al., 2014).

Two-factor authentication, in which user provides two different authentication factors to verify their identity. This is more secure to use compared to other traditional methods (Khanaa et al., 2014). It requires a secondary device, such as phone or security token, to gain second form of authentication in addition to their password. However, this is also the main drawback of two-factor authentication method, in which it will also be inconvenient for users, if they do not have secondary device with them.

Besides, biometric authentication method is another promising approach to be mentioned. In recent years, various security domain has start to enhance their security system using biometric authentication (Akhtar et al., 2018; Albalawi et al., 2022; Chawla et al., 2021; Dhoot et al., 2020; Tse & Hung, 2020). Generally, biometric authentication is an approach of identifying user based on their unique physical or behavioural characteristics, for instance, Iris, face recognition, fingerprint, voice and

etc. Most of the time, characteristic of human is unable to be changed throughout the time. Moreover, it is hard for attackers to replicate. Besides, it can be more efficient as it does not require user to remember or carry any secondary device in order to gain access to their account (Buckner et al., 2016; Mallet et al., 2022; Selvakumar et al., 2022).

However, there are also few vulnerabilities or potential limitations to use biometric authentication. Firstly, this method is still not fully accepted by some people, as it violates people privacy which made them feel uncomfortable (Akhtar et al., 2018; Buckley & Nurse, 2019). Secondly, single biometric authentication method (unimodal) sometimes will produce false positives or false negatives due to several reasons, such as limited samples, aging, variation or low quality of biometric data. It means that this method sometimes failed to recognize the biometric data or mistakenly identify unauthorized user as the correct user (Sengar et al., 2020). Thirdly, although it is more secure than traditional password-based methods, it is still not foolproof (Akhtar et al., 2018). As the biometric data can be stolen, replicated, or even the systems can be deceived by unauthorized user.

Fortunately, some of these issues can be improved or even resolved by combining multiple biometric modalities into one. Multimodal biometric authentication method can further improve the accuracy of the authentication process (Tse & Hung, 2020). Besides, it made the unauthorized user harder to steal or replicate all of the biometric data, which enhanced the security level. Moreover, using multimodal biometric authentication allows for greater accessibility and flexibility in the authentication process, as it enable user to use any combination of biometric factors that are available (Mustafa et al., 2020), hence the range of users is widen.

At the same time, introducing multimodal biometric modalities also lead to greater challenges. Firstly, it might be costly for combining different biometric data, as all of them have different feature representations, classifiers and evaluation metrics (Choudhary & Naik, 2019). Secondly, once again the process of combining multiple biometric data is more complex than unimodal biometric data. Thirdly, the design of the fusion strategy is also one of the great challenges, in which it determines how the

outputs of the different biometric systems are combined to make a final decision (Ghayoumi, 2015). There is various type of fusion strategies that have been proposed in the literature, such as early fusion, feature fusion, and score fusion. Where each of them has own pros and cons. Lastly, usage of multimodalities indicate that the implementation stage would be more difficult, as well as greater challenge in obtaining multimodal biometric dataset.

In an effective recognition system, the selection of a classifier is crucial. Traditionally, machine learning has been popular for building recognition models. However, in the era of big data, datasets have become more complex and non-linear, causing machine learning to face limitations. Fortunately, machine learning classifiers can still achieve high performance, albeit with significant effort in dataset preparation. In computer vision, datasets are often highly complex, requiring image processing and feature extraction techniques to enhance classifier performance. Consequently, training becomes time-consuming, and the computational cost hinders real-time recognition systems.

Deep learning has made significant advancements in recent years, surpassing the performance of traditional machine learning classifiers. It achieves this by introducing non-linearity through hidden layers and neurons, allowing the model to learn the dataset's nature and make predictions, much like the human brain. Deep learning architectures, such as convolutional neural networks (CNN), are particularly effective for feature extraction. Additionally, deep learning models often require less pre-processing effort for input data, resulting in superior performance. However, training a robust recognition model using DL requires huge amount of dataset. Where most of the time, the number of training samples in biometric dataset for an user are very limited. Additionally, training DL model from scratch for biometric authentication purpose was time consuming as the complexity of biometric dataset are high.

Therefore, transfer learning has gained significant attention among researchers in recent years. It is an extension of deep learning that leverages pre-trained models to build new recognition models. Surprisingly, even when trained on small datasets, these

new models exhibit impressive performance. The pre-trained models are typically trained by other researchers using abundant resources. By utilizing a pre-trained model and only training the classification layer (fully connected layer), researchers can construct new recognition models without starting from scratch. This approach offers a valuable opportunity to overcome the limitations of limited training datasets in biometric recognition systems. Notably, there are several state-of-the-art architectures available, such as VGG16, ResNet, and InceptionResNet. These models have been trained on large datasets to classify a wide range of 1000 classes, showcasing their superior performance. Consequently, applying transfer learning to train biometric recognition systems becomes a promising and beneficial strategy.

In this research, a transfer learning approach to enhance the security of online banking with hybridization of multimodal biometric authentication methods has been proposed. In this research work, iris, face, and palmprint biometrics has been used. Feature extraction has been performed using pre-trained model. Next, feature fusion has been performed to merge these multimodalities. Lastly, softmax layer has been trained to perform multiclass classification for the biometric recognition model. In summary, this research aims to create a model in which it is more robust and secure than single biometric method alone. Additionally, the model trained can adapt to more variations and improve over time.

1.3 Problem Statement

Currently, online banking faces persistent vulnerabilities to security attacks and threats, including identity theft and fraud. Traditional methods like password-based systems, security tokens, and two-factor authentication have been explored, but they often prove inconvenient and insecure for users when confronted with evolving security threats. Additionally, relying solely on a single biometric authentication method using machine learning falls short in providing sufficient security. Consequently, there is a pressing need to develop a more robust and secure authentication system for online banking during the user authentication stage. This

calls for the adoption of a multimodal authentication method that incorporates transfer learning to achieve enhanced security.

The following are the research questions that will be addressed:

1. What are the machine learning techniques and biometric modalities used in recognition system related to online banking security?
2. How to effectively perform feature extraction and integrate multiple biometric modalities for better recognition rate.
3. How to evaluate performance for the recognition models which trained to recognize 100 and 300 classes?

1.4 Research Purpose

The aim of this research is to enhance the security of online banking through the implementation of transfer learning model that hybridized with different biometric modalities (iris, face, palmprint) to provide a more robust and accurate user recognition system.

1.5 Research Objectives

The objectives of this research are :

1. To study the existing machine learning technique and biometric modalities used in recognition system related to online banking security.
2. To propose new recognition model with the knowledge of transfer learning using pre-trained models (VGG16, ResNet50, and InceptionResNet V2) and integrate different feature representation through feature level fusion.

3. To evaluate the effectiveness of proposed models and compare the recognition accuracy with other models applied which related to banking security domain.

1.6 Research Scope

In this research, the scopes and assumptions are listed below:

1. The dataset for iris, face, and palmprint used in this research is obtained from CASIA database (*BIT*, 2005).
2. This research work only uses Python programming language.
3. The performance of the model will be judged mainly on accuracy score.

1.7 Research Significance

It is important to conduct this research due to several reasons.

1. There is a need to simplify the image pre-processing process and feature extraction through the use of pretrained state-of-the-art models.
2. There is a need for constructing a recognition model which could be more effective, stable and robust to be applied in online banking security domain.
3. This research will contributes to the broader view of transfer learning adoption and multimodal fusion.

1.8 Research Contributions

In this section, the research contributions would be listed as below:

1. Simplify the image pre-processing stage and extract features using pre-trained model, as this will significantly reduce the overall training time while preserving the model performance.
2. Develop an effective multimodal biometric recognition model that combines iris, face, and palmprint through feature level fusion and with assist of transfer learning. The model trained could be more robust and effective against more variations or security threats as compared to other approaches.
3. Evaluate and compare the model difference between researchers work. This could help in identifying the effectiveness of different models.

1.9 Organization of the Thesis

In this research work, there were total 5 chapters. In chapter 1, it described the research introduction. Next, chapter 2 was literature review, where it provides fundamental background studies, and also some related researchers work. This research work continued with Chapter 3, which was the research methodology, where it described the entire flow about this research work. Then this thesis report continued with Chapter 4, which would be the research design and implementation, where all of steps for the proposed method would be clearly explained. Afterward, Chapter 5 would be the research results, analysis and discussion. Finally, Chapter 5 was about the conclusion for this research work, which includes the achievement of research objectives, research contributions and some suggestions for the future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this chapter, the focus is on providing an overview of online banking and security. It begins by discussing traditional approaches utilized in this context, highlighting the drawbacks associated with each method. Following that, the chapter explores the concept of biometric authentication and its potential advantages. Additionally, the chapter introduces a more robust recognition approach known as multimodal biometrics authentication. This approach involves combining multiple biometric traits to enhance the accuracy and reliability of authentication systems. To facilitate a better understanding, brief explanations about machine learning and deep learning are provided. Moreover, the chapter delves into the concept of transfer learning and its relevance in biometric recognition. State-of-the-art architectures, such as VGG16, ResNet50, and InceptionResNetV2, are discussed to showcase their applications in this field. Furthermore, the chapter provides clear explanations of modality fusion techniques such as early fusion, feature fusion, and score fusion. Each technique is described in detail. Towards the end of the chapter, a review of other researchers' work is presented which is aimed to identify existing research gaps in the field of multimodal biometric authentication. These gaps serve as the basis for further investigation and development in subsequent chapters.

The literature map for this chapter will be shown as below (Figure2.1), which describes the flow for this chapter.

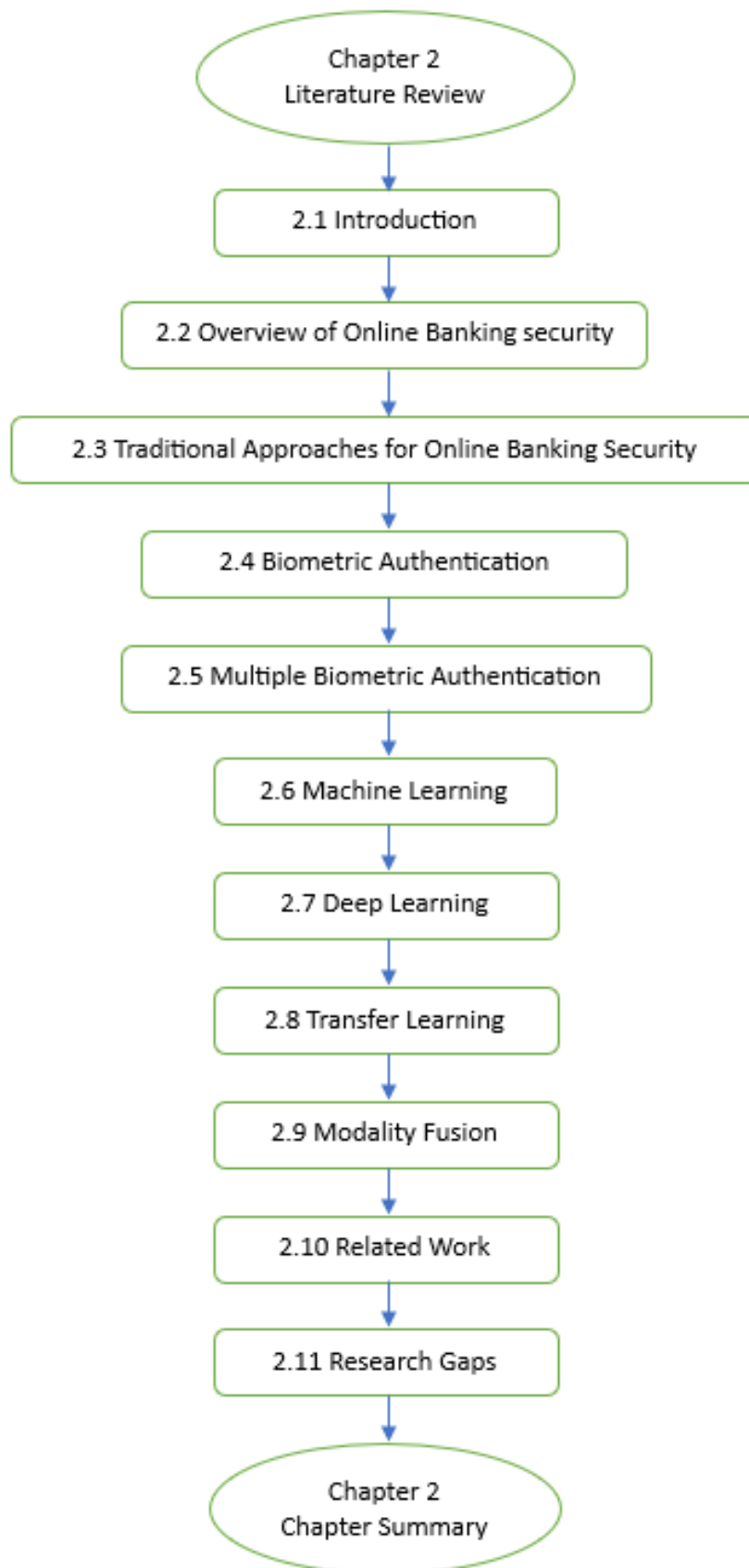


Figure 2.1 Literature Map

2.2 Overview of Online Banking Security

Online banking, sometimes also known as e-banking, internet banking, or web banking is the service provides by bank, which enable user to access their financial account without physically present in the bank. By accessing the Internet connection using electronic devices such as mobile phone or computer, user can conduct financial transactions through the website of the bank. It has become more popular to be used for everyone, due to its convenience, speed and accessibility (Keskar & Pandey, 2018).

The history of online banking can be traced back to 1980s, this is the time where banks began to provide online banking services using certain devices, such as keyboards, terminals, and monitors (Stewart, 2009). Due to the fast development Internet and in the speed of microprocessor, as in 1990s, the process of online banking is evolving fast and banks began to provide more advanced services, such as account transfers and loan applications through the internet (Vegas, 2005). Nowadays, it became even more advanced, in which all of the banks have their own banking application to provide various services to the users, such as account management, bill payment, investment, insurance and more.

At the same time, the fast development online banking also poses several security risks and challenges throughout the years. These security issues such as phishing attack, malware, identity theft, and unauthorized access, could be threaten to the confidentiality, integrity, and availability of financial information (Stewart, 2009). Well, various methods have been used to resolve these problems, such as password-based systems, security tokens, and two-factor authentication. These methods are considered as the traditional approaches to address security issues. However, none of them are completely secure. Moreover, they are inconvenient to be used sometimes. Therefore, there is a need for more robust and adaptable approaches in addressing online banking security issue.

2.3 Traditional Approaches for Online Banking Security

2.3.1 Password

The use of passwords is a fundamental method for securing user accounts, including those in online banking. However, there are several challenges and misconceptions that users often encounter when it comes to password security.

- (a) Firstly, users are confronted with the need to remember multiple passwords for different accounts, leading many to resort to using the same password across various accounts (Aloul & Zahidi, 2009). This practice significantly increases the risk of being hacked, as compromising one account can potentially grant access to other accounts using the same password. Therefore, it is crucial for users to establish unique passwords for each of their online banking accounts to mitigate the impact of potential security breaches.
- (b) Secondly, creating a strong password can be difficult, and remembering it can be equally challenging. Research such as the analysis of the 100 most common passwords in 2022 (*100 Most Common Passwords Of 2022. Can You Spot Your Password?*, 2022) has shown that many users still opt for easily guessable passwords like '123456' or 'abcde,' leaving their accounts vulnerable to attacks. It is essential for users to understand the importance of constructing strong and complex passwords that combine a mixture of uppercase and lowercase letters, numbers, and special characters to enhance their online banking security.
- (c) Thirdly, there are misconceptions among users regarding password security. For example, some individuals mistakenly believe that appending a single character, such as an exclamation mark ('!'), at the end of their password will significantly strengthen its security (Mayer & Volkamer, 2018). However, in reality, such an approach does not substantially enhance the password's strength. It is crucial to educate users about effective password security practices, emphasizing the need for robust and unpredictable passwords that are not easily guessable or derived from personal information.

2.3.2 Security Tokens

Security tokens are physical devices that generate one-time passcodes and are commonly utilized in conjunction with passwords to access online banking accounts. While security tokens provide an additional layer of security compared to single-factor authentication method such as the password, they have some inherent drawbacks. Firstly, security tokens can be easily lost or stolen (Buckley & Nurse, 2019), leading to unauthorized access if obtained by malicious individuals. Secondly, users often find it inconvenient to carry around a physical token, potentially leading to situations where they choose not to use it, compromising the additional security it offers. Additionally, the maintenance and distribution of security tokens by financial institutions can be costly and time-consuming.

2.3.3 Two-factor Authentication

Two-factor authentication is another approach to address security threats. Generally, two-factor authentication (2FA) is an effective approach as compared to single-factor authentication methods for enhancing online banking security and mitigating security threats (Aloul & Zahidi, 2009). This method involves requiring users to provide two distinct authentication factors to access their accounts, adding an extra layer of protection. 2FA can be implemented using various methods, including SMS, email, passwords, or security tokens. However, they does have limitations. Firstly, 2FA can be inconvenient for users who do not have immediate access to a secondary device or reliable internet connectivity required for receiving authentication codes. This can lead to frustration and potential delays in accessing online banking services. Secondly, 2FA methods relying on passwords or one-time codes sent via SMS or email are still susceptible to various attacks, including phishing and SIM swapping. Furthermore, managing multiple authentication factors can be burdensome for users who struggle to remember or keep track of them.

2.4 Biometric Authentication

Another highly effective authentication method, surpassing traditional approaches, is biometric authentication. This method utilizes individual physical characteristics or behavioural traits to accurately identify the authentic user (Albalawi et al., 2022). With the uniqueness of every individual, biometric authentication is considered the most secure way to safeguard users' financial accounts against threats in online banking.

Moreover, biometric authentication offers several advantages over traditional methods. Firstly, it presents significant challenges for replicating or stealing one's biometric data, making it a difficult target for malicious actors. Secondly, this method eliminates the need for users to remember or carry extra devices, providing a more convenient user experience.

Various types of biometric authentication methods exist, such as iris recognition, face recognition, and palm vein recognition. Each of these methods captures distinct features from different parts of the human body, resulting in different feature representations, classifiers, and evaluation metrics (Akhtar et al., 2018). Iris recognition involves capturing an image of the user's iris, while face recognition requires an image of the user's face. Palmprint recognition, on the other hand, relies on an image of the veins in the user's palm (Israa, 2015).

Additionally, Table 2.1 presents a comprehensive overview of the pros, cons, and challenges associated with these biometric authentication methods (Iris, Face, and Palmprint Recognition).

Table 2.1 Comparisons Between Iris Recognition, Face Recognition, and Palmprint Recognition in terms of Pros, Cons, and Challenges

	Iris Recognition	Face Recognition	Palmprint Recognition
Pros	<ul style="list-style-type: none"> • Highly accurate and reliable • Non-invasive • Resistant to changes over time • Easy to use 	<ul style="list-style-type: none"> • Widely used as it is easy to implement • Non-invasive • Wide range of application 	<ul style="list-style-type: none"> • Highly accurate and reliable • Non-invasive • Resistant to changes over time • Wide range of application
Cons	<ul style="list-style-type: none"> • Expensive to implement • May not suitable for individuals with certain medical conditions 	<ul style="list-style-type: none"> • Less accurate than other biometric methods • Result changes with user's appearance 	<ul style="list-style-type: none"> • Require special hardware (higher cost) • Might not suitable for users with certain medical conditions
Challenge	<ul style="list-style-type: none"> • Sensitive to image quality • Sensitive to environmental factors 	<ul style="list-style-type: none"> • Sensitive to image quality • Sensitive to environmental factors • Affected by cultural differences 	<ul style="list-style-type: none"> • Sensitive to environmental factors • Sensitive to position of the hand

In addition, there are also some drawbacks for this method. Nowadays, there are still many people rejecting this method, as they feel uncomfortable in a way that this method has violated their privacy (Akhtar et al., 2018; Buckley & Nurse, 2019).

Therefore, the difficulty to conduct this research increased as the dataset is rare and hard to obtain. Lastly, a single biometric authentication method sometimes produces significant false positive or false negative predictions, in which this method sometimes mistakenly identified an unauthorized user as the correct one (Sengar et al., 2020), or the model fail to recognize the biometric data of the actual user.

However, it is important to acknowledge some drawbacks associated with biometric authentication in the context of online banking security. Firstly, there is a notable segment of individuals who reject this method due to concerns about privacy violations (Akhtar et al., 2018; Buckley & Nurse, 2019). This resistance from users adds complexity to conducting research in this area, as obtaining rare and elusive datasets becomes more challenging.

Furthermore, a single biometric authentication method can occasionally yield significant false positive or false negative predictions. In certain cases, the system may mistakenly identify an unauthorized user as the correct one (Sengar et al., 2020), or it might fail to recognize the biometric data of the actual user. These limitations highlight the need for continuous improvement and refinement of biometric authentication systems in online banking security.

2.5 Multiple Biometric Authentication

Multimodal biometric authentication, also referred to as multiple biometric authentication, is an advanced method that combines multiple biometric data points to verify a user's identity. This approach enhances the robustness and security of the authentication system, providing heightened protection against security threats (Tse & Hung, 2020).

While compared to unimodal biometric authentication methods, multimodal biometric authentication demonstrates superior performance. It offers the advantage of backup options in case one biometric modality fails or becomes unavailable, resulting in a more convenient user experience (Mustafa et al., 2020). Moreover, it

significantly raises the bar for unauthorized users attempting to attack an account since replicating or stealing multiple biometric characteristics becomes much more challenging. The combination of multiple biometric modalities leads to improved accuracy, reliability, and reduced rates of false positive or false negative predictions during the authentication process.

In Table 2.2, it shows the existing multimodal biometric authentication systems. The biometrics and type of fusion applied by the authors were clearly shown as well as their result findings.

Table 2.2 Existing Multimodal Biometric Authentication Systems

No.	Author	Biometric			Type of Fusion	Results (%)
		1	2	3 (if any)		
1	(Hammad et al., 2019)	ECG	Fingerprint	-	Decision Level	EER: 0.14
2	(Regouid et al., 2019)	ECG	Ear	Iris	Feature Level	EER: 0.5
3	(Gunasekaran et al., 2019)	Face	Fingerprint	Iris	Score Level	Accuracy: 82
4	(Alay & Al-Baity, 2020)	Iris	Face	Finger Vein	Feature Level	Accuracy: 99.39
5	(Mustafa et al., 2020)	Iris	Fingerprint	-	Decision Level	Accuracy: 95
6	(Haider et al., 2020)	Pulse Response	Hand Geometry	Finger Vein	Score Level	Accuracy: 92
7	(Daas et al., 2020)	Finger Knuckle Print	Finger Vein	-	Score Level	Accuracy: 99.89 EER: 0.05
8	(Ma et al., 2020)	Face	Ear	-	-	-
9	(El_Rahman, 2020)	ECG	Fingerprint	-	Decision Level	-
10	(Ammour et al., 2020)	Face	Iris	-	Decision Level	Accuracy: 99.13

No.	Author	Biometric			Type of Fusion	Results (%)
		1	2	3 (if any)		
11	(Zhou et al., 2020)	Palm Vein	Finger Vein	Iris	Hybrid: Feature Level & Decision Level	Accuracy: 99.33
12	(Purohit & Ajmera, 2021)	Palmprint	Ear	Fingerprint	Feature Level	Accuracy: 91.67
13	(Safavipour et al., 2022)	Fingerprint	Iris	Face	Feature Level	Accuracy: 100
14	(Medjahed et al., 2022)	Face	Palmprint	-	Score Level	Accuracy: 100
15	(Rajasekar et al., 2022)	Iris	Fingerprint	-	Score Level	Accuracy: 99.89 EER: 0.18

According to the table shown above, the study of multimodal biometrics had started to advance since many years ago. In recent years, the most popular biometric modalities applied in authentication system were face and iris as they were easily obtainable traits for most people. Besides, the biometric traits around hand (such as finger vein, finger print, and palmprint) were also popular choices in multimodal biometric authentication systems. Additionally, these multimodal biometric authentication systems showed superior performance.

However, the application of multimodal biometric authentication system faced two significant challenges, which were the limited availability of multimodal biometric datasets and also the high complexity involved in processing such datasets. Over time, technological advancements have addressed these challenges, making multimodal biometric authentication system a viable option in various domains. In the subsequent section, several popular techniques that used for fusing different biometric modalities had been explored and clearly explained.

2.6 Machine Learning

Machine learning plays a vital role in enhancing online banking security by enabling intelligent analysis of vast amounts of data and identifying patterns indicative of fraudulent activities or unauthorized access attempts. With its ability to learn from historical data, machine learning algorithms can adapt to evolving security threats and make accurate predictions in real-time.

In the context of online banking security, machine learning offers several key advantages. One of the primary applications is fraud detection. By training models on labelled data, machine learning algorithms can learn to recognize patterns associated with fraudulent transactions. These models can then identify suspicious activities and flag them for further investigation, helping to protect users from financial losses and unauthorized access.

Another important application of machine learning in online banking security is user behaviour analysis. By analyzing user activity, machine learning models can establish normal patterns of behaviour for individual users. Any deviations from these patterns can be flagged as potential security risks, enabling proactive measures to prevent unauthorized access or identity theft.

Furthermore, machine learning algorithms can assist in anomaly detection, where unusual or abnormal activities are identified within the online banking system. By learning from patterns in historical data, machine learning models can detect deviations that may indicate security breaches or malicious activities, allowing for timely intervention and mitigation.

2.7 Deep Learning

Generally, deep learning is a subset of machine learning, has emerged as a powerful tool in various domains, including online banking security. It utilizes artificial neural networks with multiple hidden layers to learn complex patterns and make accurate predictions.

Besides, deep learning offers several advantages. One notable advantage is its ability to handle and analyze large-scale and unstructured datasets, such as transaction logs, user behaviour data, or network traffic. By employing deep learning models, financial institutions can uncover hidden patterns and correlations that traditional machine learning algorithms may struggle to detect.

Deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are particularly effective in analyzing complex data types. CNNs excel at image recognition tasks, allowing for advanced biometric authentication systems that leverage facial or fingerprint recognition. RNNs, on the other hand, are well-suited for sequential data analysis, enabling accurate predictions and detection of patterns in time-dependent online banking activities.

Training deep learning models requires significant computational resources and large volumes of labelled data. However, advancements in transfer learning have made it possible to leverage pre-trained deep learning models and adapt them to specific online banking security tasks. This approach reduces the training time and dataset requirements, while still achieving high performance.

In summary, deep learning offers advanced capabilities for analyzing complex and unstructured data, enabling accurate fraud detection, user behaviour analysis, and anomaly detection in online banking security. By leveraging deep learning techniques, financial institutions can enhance their security measures and protect users from evolving cyber threats.

2.8 Transfer Learning

Generally, transfer learning is a technique that utilizes knowledge gained from a source task to enhance the performance of a target task, building upon previous learning experiences (Chan et al., 2023). In simpler words, it allows researchers to construct a new model by leveraging an existing model that was trained on a different task. By modifying only the output layer, it is able to construct a new classification model for a different task, while retaining and freezing the pre-learned weights from the source task. Additionally, transfer learning enables the training of a high-performing model without requiring an extensive amount of data (Gunawan et al., 2023).

Moreover, transfer learning facilitates feature extraction by enabling researchers to utilize a pre-trained model. Typically, the pre-trained model includes layers responsible for learning how to extract features from input images. As a result, these pre-trained layers can be directly employed, eliminating the need for training from scratch.

In Figure 2.2, the generic process flow of transfer learning is illustrated. According to the figure, the source model undergoes pre-training using a vast amount

of data and associated labels. The convolutional layers within the source model are specially crafted to extract vital features from the input data, while the fully-connected layers are employed to learn these extracted features and make predictions through the final output layer. The model's parameters, including weights and biases at each layer, are meticulously trained to ensure accurate predictions.

Subsequently, these learned parameters are frozen and extracted from the source model, ready to be transferred to a target model for a different task. This transfer allows the target model to focus solely on training the fully-connected layers specific to its unique task. As a result, significant reductions in effort, such as training data and time, are achieved when training the target model. Furthermore, experimental evidence supports the notion that the performance of the target model remains robust in this transfer process. In the present day, a range of popular pre-trained models exists, including VGG16, ResNet-50, and InceptionResNet. Each of these models will be displayed and provided a detailed explanation in the subsequent sections.

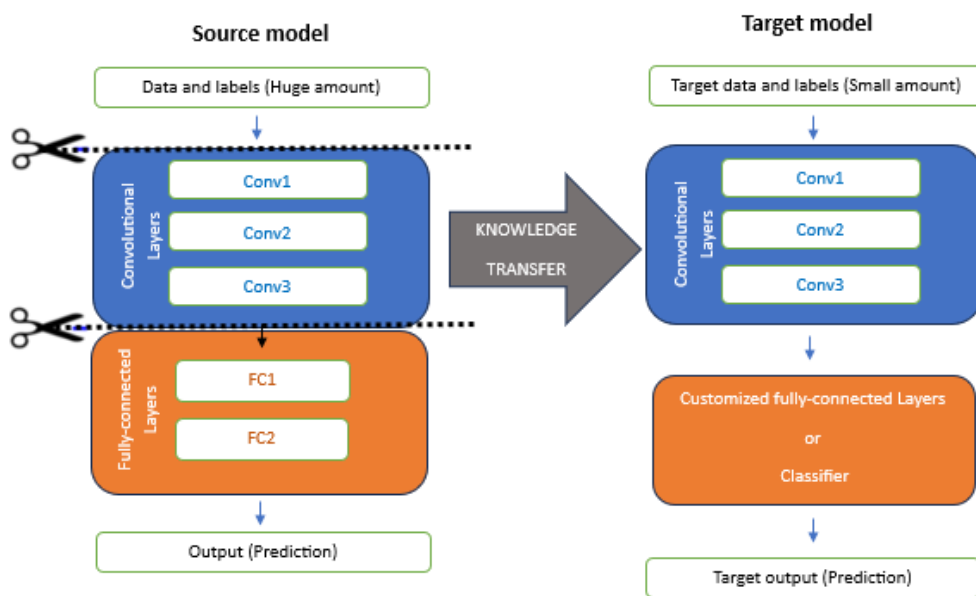


Figure 2.2 Generic Process Flow of Transfer Learning

2.8.1 Feature Extraction

Feature extraction is an important process in image processing. This process extracts relevant features or information from the images (Zebari et al., 2020). The extracted features would then be used as the input dataset fed to the classifier. Generally, there are various types of machine learning based feature extraction techniques which are commonly used, such as principal component analysis (PCA), linear discriminant analysis (LDA), independent discriminant analysis (IDA), local binary patterns (LBP), histograms, gradient-based method, and scale-invariant feature transform (SIFT). Each of these techniques has pros and cons. They are all tabulated as shown in Table 2.3.

Table 2.3 Comparisons of Different Machine Learning Based Feature Extraction Techniques Based on Pros and Cons

Technique	Pros	Cons
PCA	<ul style="list-style-type: none">- Simple and effective for dimensionality reduction- Can handle large amounts of data, missing data	<ul style="list-style-type: none">- Sensitive to the scaling of data- Information might lost in the process
LDA	<ul style="list-style-type: none">- Simple and effective for classification problem- Can handle large amounts of data	<ul style="list-style-type: none">- Assume linear relationships between the features- Sensitive to the scaling of the data
IDA	<ul style="list-style-type: none">- Can handle non-linear relationships between features- Can handle large amounts of data	<ul style="list-style-type: none">- Sensitive to scaling of the data- Computationally expensive
LBP	<ul style="list-style-type: none">- Robust to changes or small variation within the image- Can handle large amounts of data	<ul style="list-style-type: none">- Sensitive to changes in the scale and orientation of the image

Technique	Pros	Cons
Histogram	<ul style="list-style-type: none"> - Simple to implement - Can handle large amounts of data 	<ul style="list-style-type: none"> - Sensitive to changes in the scale and orientation of the image - Sensitive to the choice of bin sizes
Gradient-based method	<ul style="list-style-type: none"> - Can capture the shape and structure of the image - Robust to changes and small variations within the image 	<ul style="list-style-type: none"> - Sensitive to changes in the scale and orientation of the image - Computationally expensive
SIFT	<ul style="list-style-type: none"> - Robust to changes in the scale and orientation of the image - Can handle large amounts of data 	<ul style="list-style-type: none"> - Computationally expensive

Based on the findings in Table 2.3, several important insights can be summarized. Firstly, it was observed that many machine learning-based feature extraction methods tended to be computationally expensive, posing challenges in terms of processing time and resource requirements. Additionally, these methods were found to be sensitive to variances within input images, potentially impacting the accuracy and reliability of feature extraction. Another noteworthy observation is the difficulty in determining the optimal algorithm parameters. This selection process requires careful consideration and experimentation to achieve the best results. Furthermore, it was noted that some feature extraction techniques may inadvertently result in information loss, potentially affecting the overall effectiveness of the extracted features. These findings highlight the need for further research and development in the field of feature extraction for online banking security. Overcoming the computational complexity, addressing sensitivity to input variances, and refining the parameter selection process are key areas of focus to improve the efficiency and performance of machine learning-based feature extraction methods in online banking security systems.

In recent years, there is one feature extraction technique has become very popular to be used in the context of online banking security, which is Convolutional Neural Network (CNN). CNN architecture consists of multiple layers and parameters that play a crucial role in enhancing the security of online banking transactions. The convolutional layers within the CNN facilitate feature extraction process, allowing the model to identify and capture important patterns from the input data. By analyzing the structure and characteristics of the data, the CNN can identify potential threats or anomalies that may pose risks to online banking security. These layers generate a feature map, which represents the learned features.

Additionally, the pooling layer in the CNN contributes further to online banking security by reducing the dimensionality of the feature map. This reduction in dimensionality not only enhances the efficiency of the model but also helps in minimizing computational requirements (Benradi et al., 2023).

In Figure 2.3, it shows the typical example of CNN feature extraction layers for a better understanding.

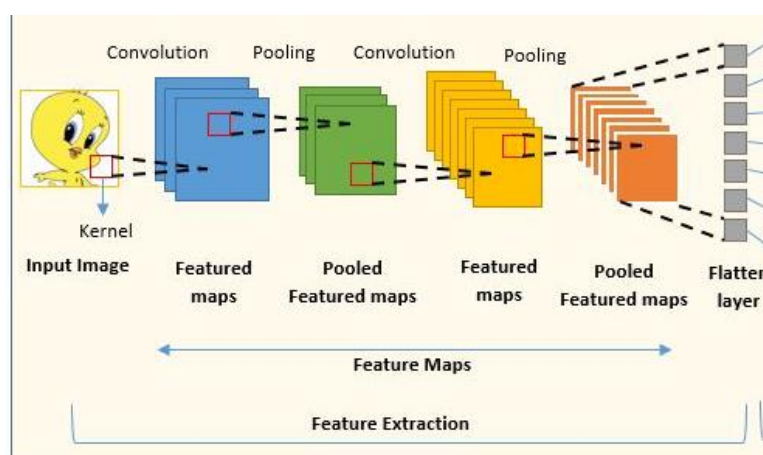


Figure 2.3 Typical Example of CNN Feature Extraction Layer (*Convolutional Neural Network: An Overview*, 2022)

2.8.2 Fully Connected Layer

In Figure 2.4, it shows the fully connected layer in a pre-trained model. In short, fully connected layer (FC layer) is used to learn complex pattern or feature within the training data by performing linear data transformation followed by applying non-linear activation function to the transformed output. From Figure 2.4, there are two hidden layers and 1 classification (output) layer. It has same architecture concepts as in Artificial Neural Network (ANN). In transfer learning, modification can be done on fully connected layer to accommodate specific task. This approach allows researchers to efficiently train only the parameters within the fully connected layer, enabling the construction of a new model tailored to the desired task.

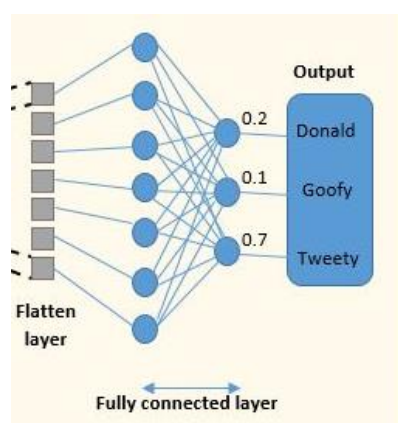


Figure 2.4 Typical Example of Fully Connected Layer (*Convolutional Neural Network: An Overview*, 2022)

2.8.3 State-of-the-Art Architectures

In the present era, pre-trained models have become extensively utilized across diverse domains. These models have been meticulously trained using significant computational resources, leading to remarkable performance and establishing them as the State-of-the-Art Architecture. This section presents a glimpse of notable architectures such as VGG16, ResNet 50, and InceptionResnet, each serving specific purposes such as iris recognition, facial recognition, and palmprint recognition,

respectively. Furthermore, the subsequent subtopics offer a thorough exploration of the structures and conceptual ideas behind these architectures, providing clear and comprehensive explanations.

In the field of iris recognition, Singh et al. (2022) proposed a system that utilized a pre-trained VGG16 model, a convolutional neural network (CNN) architecture (A. Singh et al., 2022). By applying the VGG16 model for feature extraction, their research work achieved significant progress. The system was trained and tested using the CASIA-1000 database, resulting in an impressive accuracy rate of 96%.

Another application of pre-trained convolutional neural networks can be seen in Li and Lima's (2021) research on human emotion recognition based on facial expressions (Li & Lima, 2021). They employed a pre-trained ResNet-50 model for feature extraction in their system. The performance of their resulting model surpassed that of other state-of-the-art methods, achieving an impressive accuracy score of 95.39%.

Similarly, Zhang et al. (2018) proposed a palmprint recognition system based on a Deep Convolutional Neural Network (DCNN) (Zhang et al., 2018). In their research, they utilized the InceptionResnet model, a pre-trained model, as a feature extractor. The model performed exceptionally well, with accuracy, precision, and recall reaching 100%, and an equal error rate (EER) of only 2.30%.

In summary, these research works demonstrate the effectiveness of pre-trained convolutional neural networks in computer vision. By utilizing models such as VGG16, ResNet-50, and InceptionResnet, researchers have achieved remarkable results by showing superior accuracy and performance of these systems.

2.8.3.1 VGG16 Model

In 2014, VGG16, also known as ConvNet had won the 1st and 2nd place in object detection and classification in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) event. The model was created by Karen Simonyan and Andrew Zisserman. In the following year, they had shared their model findings by publishing their paper titled “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION” (Simonyan & Zisserman, 2015). In their paper, they discussed about the VGG16 configurations and investigate the effectiveness of the VGG16 for large-scale image classification. It was found out that their model performance was able to outperform other complex recognition model.

Nowadays, VGG16 remains one of the most popular artificial neural networks used for computer vision tasks. Originally trained on the imagenet dataset, which contains a vast number of training images spanning 1000 different classes, the VGG16 model achieved an impressive accuracy score of 92.7%. Table 2.4 provides a summary of the pre-trained VGG16 model.

Table 2.4 Summary of VGG16 and the Model Configurations

Layers	Kernel Configuration	Number of Parameters	Dimension
Input	-	-	224 x 224 x 3
Convolution Block_1	Stride 1, Padding = Same		
• Conv1_1	3 x 3 x 64	1,792	224 x 224 x 64
• Conv1_2	3 x 3 x 64	36,928	224 x 224 x 64
Pooling Layer_1	Stride 2		
- Max Pooling 1	2 x 2 x 64	0	112 x 112 x 64
Convolution Block_2	Stride 1, Padding = Same		
• Conv2_1	3 x 3 x 128	73,856	112 x 112 x 128
• Conv2_2	3 x 3 x 128	147,584	112 x 112 x 128
Pooling Layer_2	Stride 2		
- Max Pooling 2	2 x 2 x 128	0	56 x 56 x 128
Convolution Block_3	Stride 1, Padding = Same		
• Conv3_1	3 x 3 x 256	295,168	56 x 56 x 256
• Conv3_2	3 x 3 x 256	590,080	56 x 56 x 256
• Conv3_3	3 x 3 x 256	590,080	56 x 56 x 256
Pooling Layer_3	Stride 2		
- Max Pooling 3	2 x 2 x 256	0	28 x 28 x 256
Convolution Block_4	Stride 1, Padding = Same		
• Conv4_1	3 x 3 x 512	1,180,160	28 x 28 x 512
• Conv4_2	3 x 3 x 512	2,359,808	28 x 28 x 512
• Conv4_3	3 x 3 x 512	2,359,808	28 x 28 x 512
Pooling Layer_4	Stride 2		
- Max Pooling 4	2 x 2 x 512	0	14 x 14 x 512
Convolution Block_5	Stride 1, Padding = Same		
• Conv5_1	3 x 3 x 512	2,359,808	14 x 14 x 512
• Conv5_2	3 x 3 x 512	2,359,808	14 x 14 x 512
• Conv5_3	3 x 3 x 512	2,359,808	14 x 14 x 512
Pooling Layer_5	Stride 2		
- Max Pooling 5	2 x 2 x 512	0	7 x 7 x 512
Flatten Layer	25088	0	1 x 1 25088
Fully Connected Layer			
• FC6 (Dense)	4096	102,764,544	1 x 1 x 4096
• FC7 (Dense)	4096	16,781,312	1 x 1 x 4096
• FC8 (Prediction)	1000	8,194	1 x 1 x 1000
TOTAL	-	134,268,738	-

The VGG16 model consists of 16 layers with learnable parameters (weights and biases). Of these layers, 13 are convolutional layers, and 3 are dense layers. Additionally, the model includes 5 max pooling layers, which play a crucial role in extracting important features and reducing the dimensions of the data. In total, the VGG16 model has 134,268,738 learnable parameters.

Throughout the VGG16 model, a filter configuration of 3x3 with a stride of 1 and same padding is applied to all convolution layers. For the max pooling layers, a filter configuration of 2x2 with a stride of 2 is used. The model is composed of 5

convolution blocks, with the number of channels set to 64, 128, 256, 512, and 512 respectively.

After the relevant information is extracted from these blocks, it is flattened before being passed to the fully-connected (fc) layers. The fc layers consist of 3 dense layers. The first two layers contain 4096 neurons each, while the last layer is responsible for recognizing 1000 classes and has 1000 neurons.

2.8.3.2 ResNet-50 Model

Generally, ResNet was known as the Residual Network. It proposed the idea of using ‘Skip Connections’ within the deep neural network. The name ResNet-50 indicates that the network consists of 50 layers with learnable parameters. Specifically, it includes one convolution layer at the beginning stage, 48 convolution layers in the middle stage, and one fully connected (fc) layer at the end.

One of the main challenges in the complex structure of Deep Learning Convolution Neural Networks (DLCNN) was the occurrence of the vanishing gradient problem during back-propagation. This problem arises due to the repeated multiplication of small numbers during the application of the chain rule, resulting in numerical instability of the network's parameters. Consequently, the training process becomes extremely slow as the weights and biases vary infinitesimally. ResNet addresses this issue by introducing skip connections, which act as gradient highways. These connections enable parameters to change with significant magnitudes, thereby accelerating the model training process and improving performance stability.

Figure 2.5 illustrates the overall structure of ResNet-50, along with the input dimensions at each stage. In summary, ResNet-50 is easier to interpret and construct compared to other artificial neural networks because of its repeated modular blocks. This can be observed further in Figure 2.6(a), which displays three repetitive Conv Block Stage 1 modular blocks (green), four repetitive Conv Block Stage 2 modular

blocks (grey), six repetitive Conv Block Stage 3 modular blocks (light blue), and three repetitive Conv Block Stage 4 modular blocks (light red).

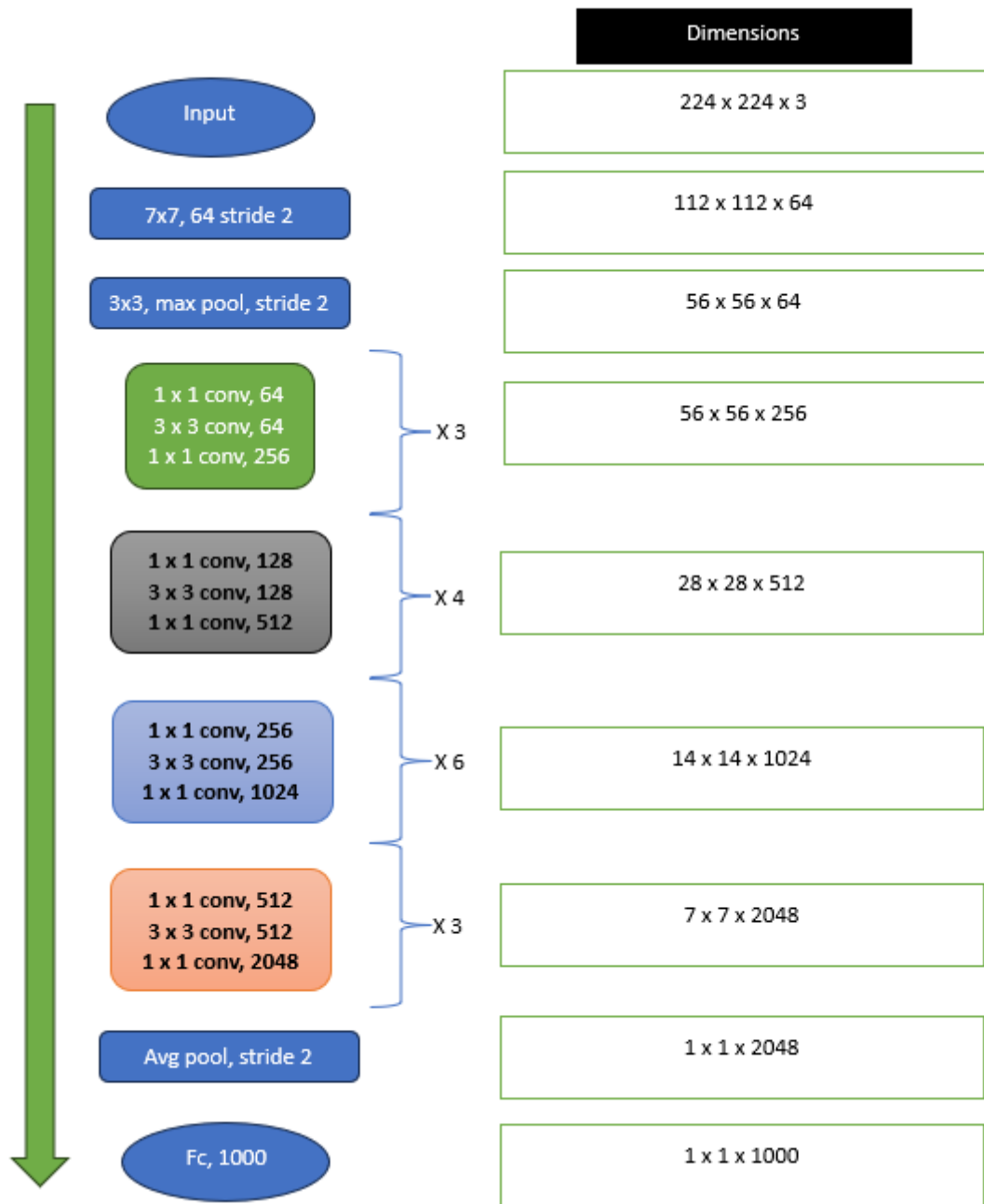


Figure 2.5 Summary of ResNet-50 Architecture

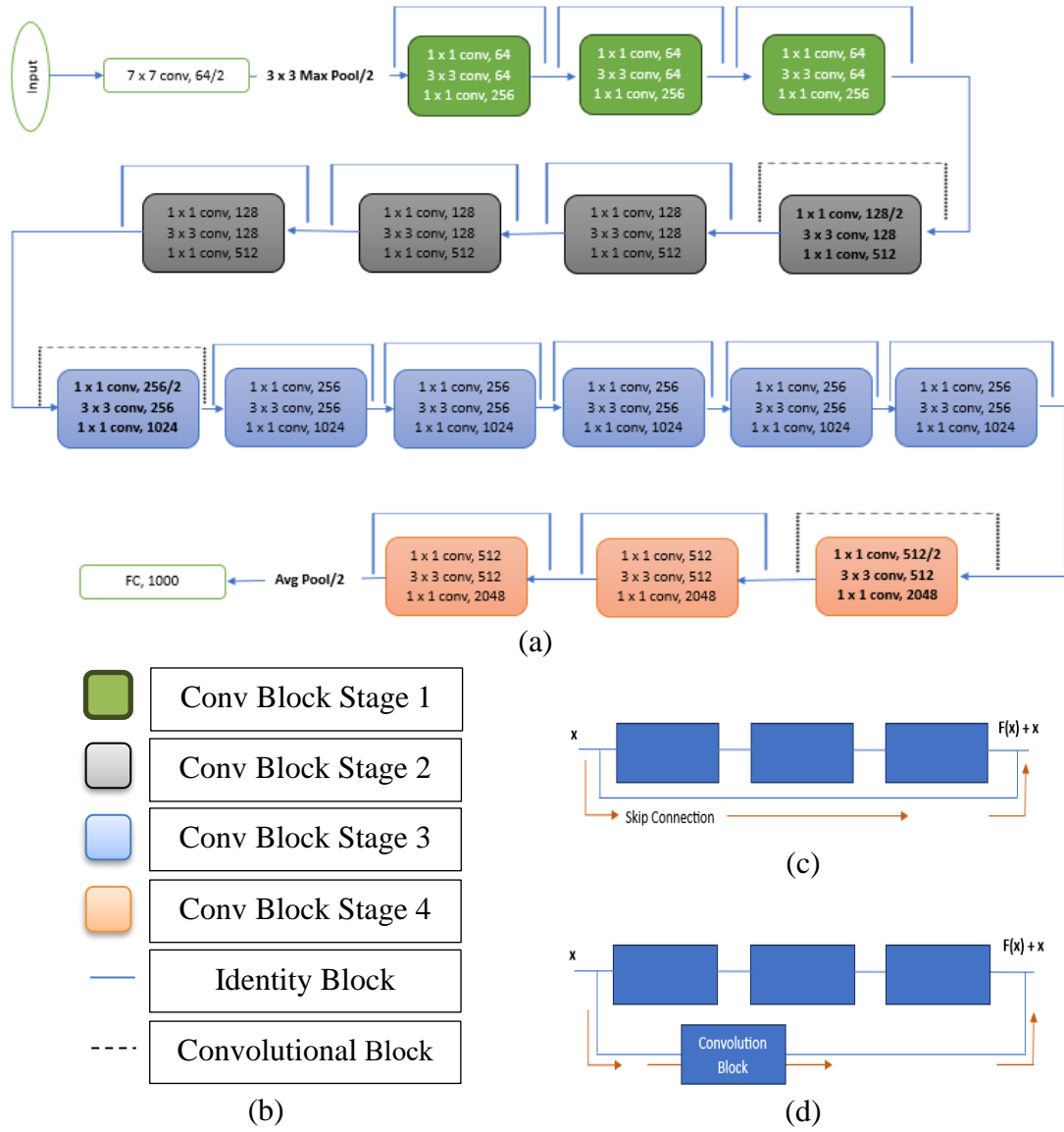


Figure 2.6 ResNet-50 Architecture (a) Overall Structure (b) Description of Blocks and lines (c) Connector for Identity Block (d) Connector for Convolutional Block

Furthermore, Figure 2.6 (c) and (d) illustrate the connectors used for the identity block and convolutional block, respectively. Referring to Figure 2.6 (a), connector (c) is applied when the skip connection line is a solid line, while connector (d) is applied when the skip connection line is a dotted line. These connectors serve the purpose of ensuring that the input size matches the output size at every stage of the convolutional block. Essentially, the input value "x" is added to the output layer only if the input size is equal to the output size. If the sizes do not match, a convolutional block is introduced in the skip connection to align the input size with the output size.

2.8.3.3 InceptionResNetV1 Model

The previous section provided a clear explanation of the residual network concept, which aims to enhance model stability by addressing the issue of vanishing gradients. In contrast, the inception architecture, starting with Inception V1, aimed to construct a robust model capable of handling variations in the target object within an image. It achieved this by introducing multiple filters with different sizes on the same level, expanding the width of the network instead of increasing its depth.

In subsequent iterations, such as Inception V3, the inception architecture was further refined. To reduce computational training parameters within the filters, a 1x1 convolutional layer was added before each convolutional layer. Additionally, a dimension reduction module was introduced, which involved factorization into smaller convolutions. For example, a 5x5 convolution was transformed into two 3x3 convolution layers, resulting in a 28% reduction in computational parameters. Furthermore, the 3x3 convolutional block was factorized into 3x1 and 1x3 convolutional blocks, further reducing computation cost by 33%.

The InceptionResNet model emerged as a combination of the Inception and ResNet architectures, leveraging the strengths of both to create a high-performance model. Notably, research conducted by Szegedy et al. validated the superior recognition performance and reduced computational costs of the InceptionResNet architecture, akin to Inception V3 (Szegedy et al., 2017).

Figure 2.7 provides an overview of the InceptionResNet_V1 schema, while Figures 2.8 (a)-(f) depict the different modules within InceptionResNet_V1. The input passes through the stem module before being introduced to the Inception blocks. The overall structure includes three main Inception modules: Inception-Resnet-A (b), Inception-Resnet-B (d), and Inception-Resnet-C (f). Residual connections replace the pooling operation within these modules, and 1x1 convolutions are employed to match the input and output dimensions. Lastly, figure 2.8 (c) and (e) illustrate the reduction modules, namely Reduction-A and Reduction-B, specifically designed to capture different variations within an image using various kernel sizes.

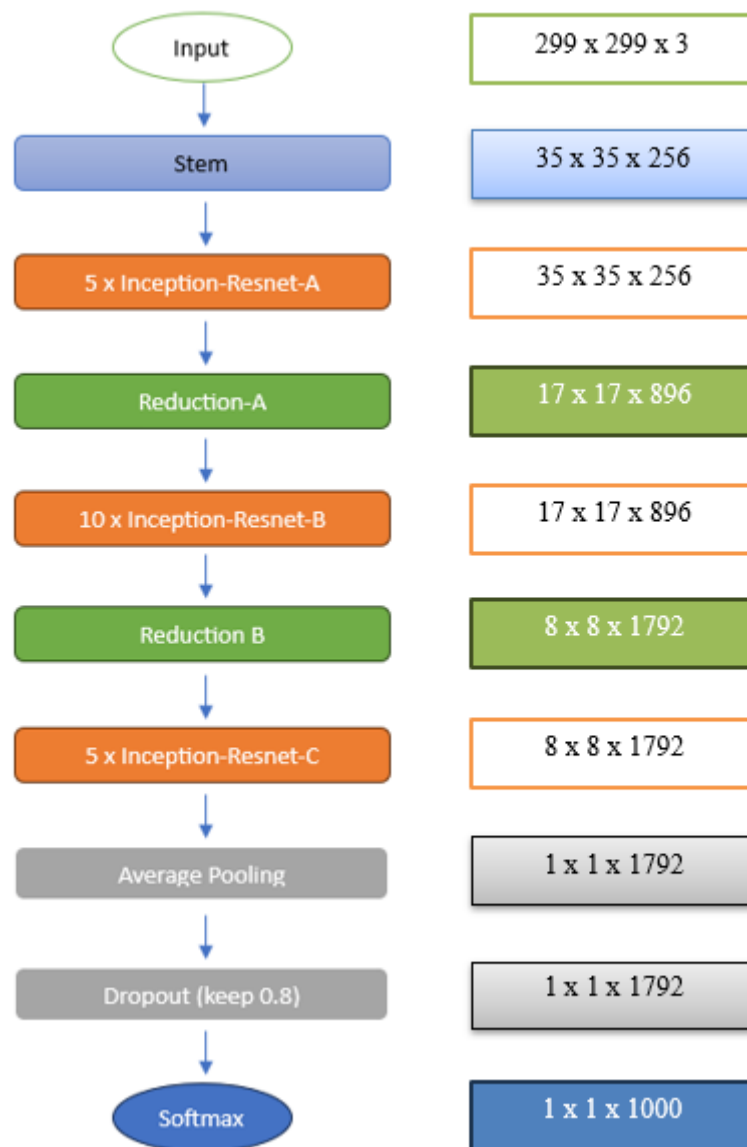
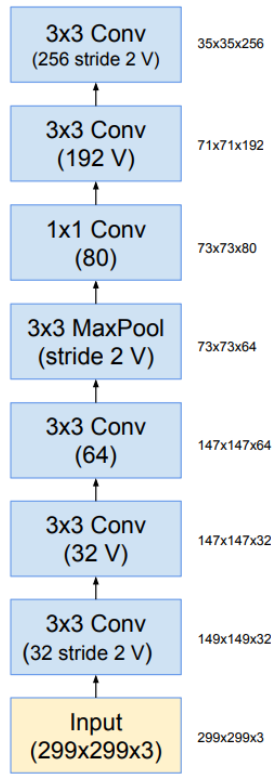
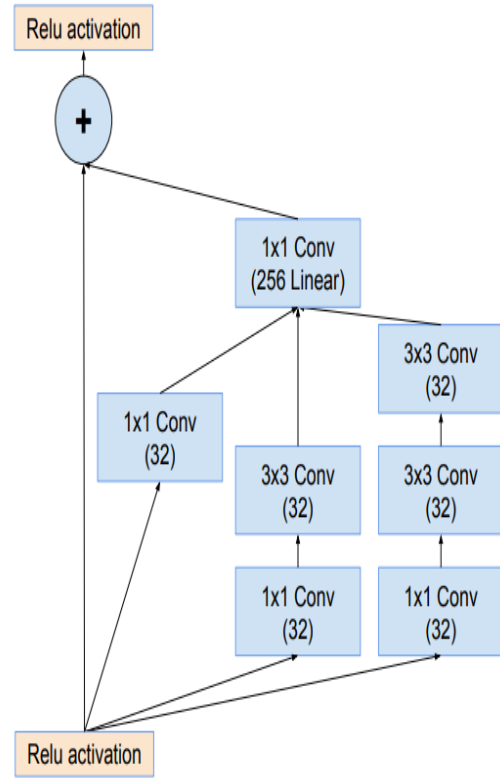


Figure 2.7 Overall Structure of InceptionResNet_V1

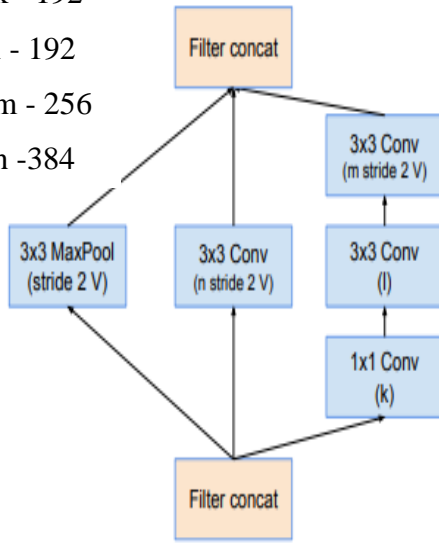


(a)

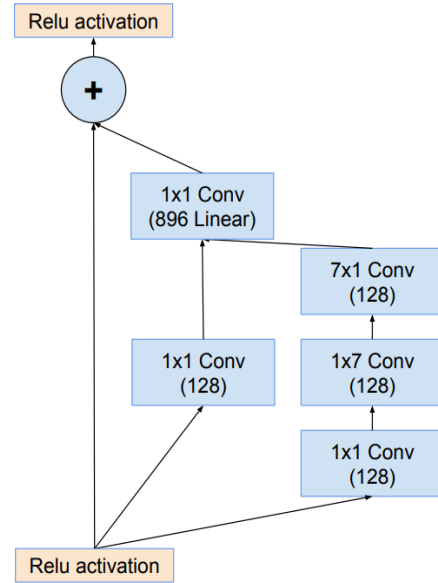


(b)

k - 192
l - 192
m - 256
n - 384



(c)



(d)

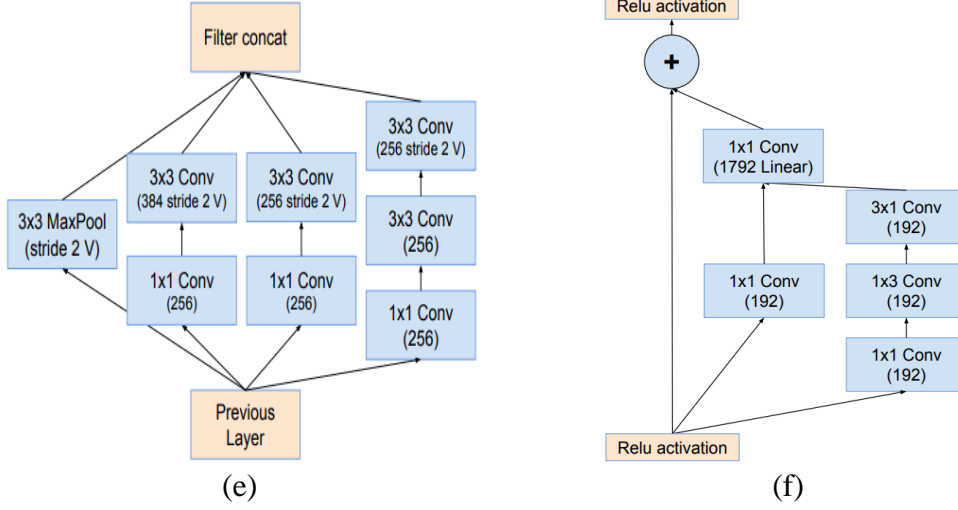


Figure 2.8 Different modules within InceptionResnet_V1 (a) Stem Module (b) Inception-Resnet-A Module (c) Reduction-A Module (d) Inception-Resnet-B Module (e) Reduction-B Module (f) Inception-ResNet-C Module (Szegedy et al., 2017)

2.9 Modality Fusion

Modality Fusion is the process of combining multiple biometric modalities into one (M. Singh et al., 2019). Generally, the performance of the authentication model will be improved as multimodal biometric offered more robust feature representations. There are some common approaches in performing modality fusion, such as early fusion, feature fusion, and score fusion.

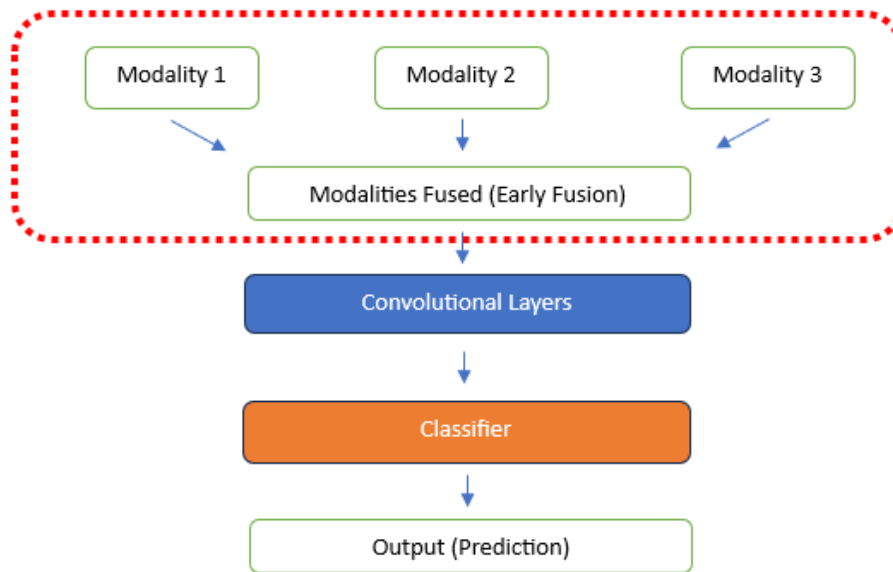


Figure 2.9 Early Fusion Process

Early fusion occurs at the early stage of processing as shown in Figure 2.9, where the data from different modalities is merged into a single unified representation. This approach aims to capture the relationships between modalities at the raw data level, providing a comprehensive understanding of the combined information.

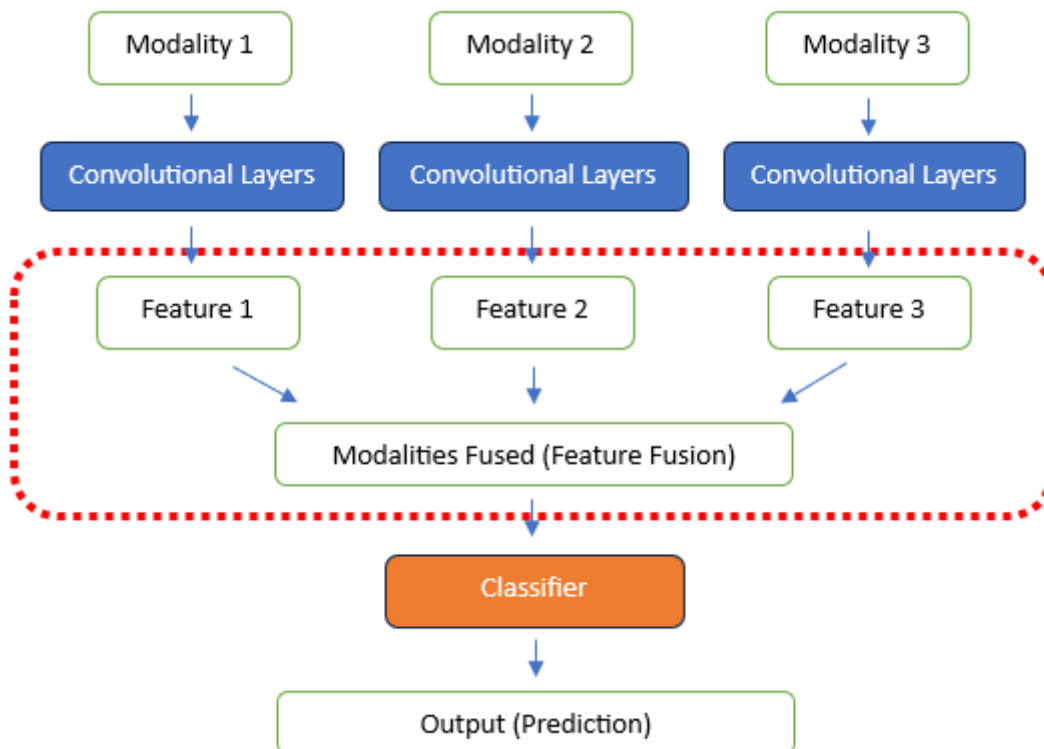


Figure 2.10 Feature Fusion Process

In Figure 2.10, it shows the process of feature fusion, where the processed or extracted features from multiple modalities are fused. This fusion takes place after the feature extraction stage, where the feature representations of different modalities are concatenated, averaged, or weighted to form a single fused feature representation. This fused dataset then becomes the input for the classifier.

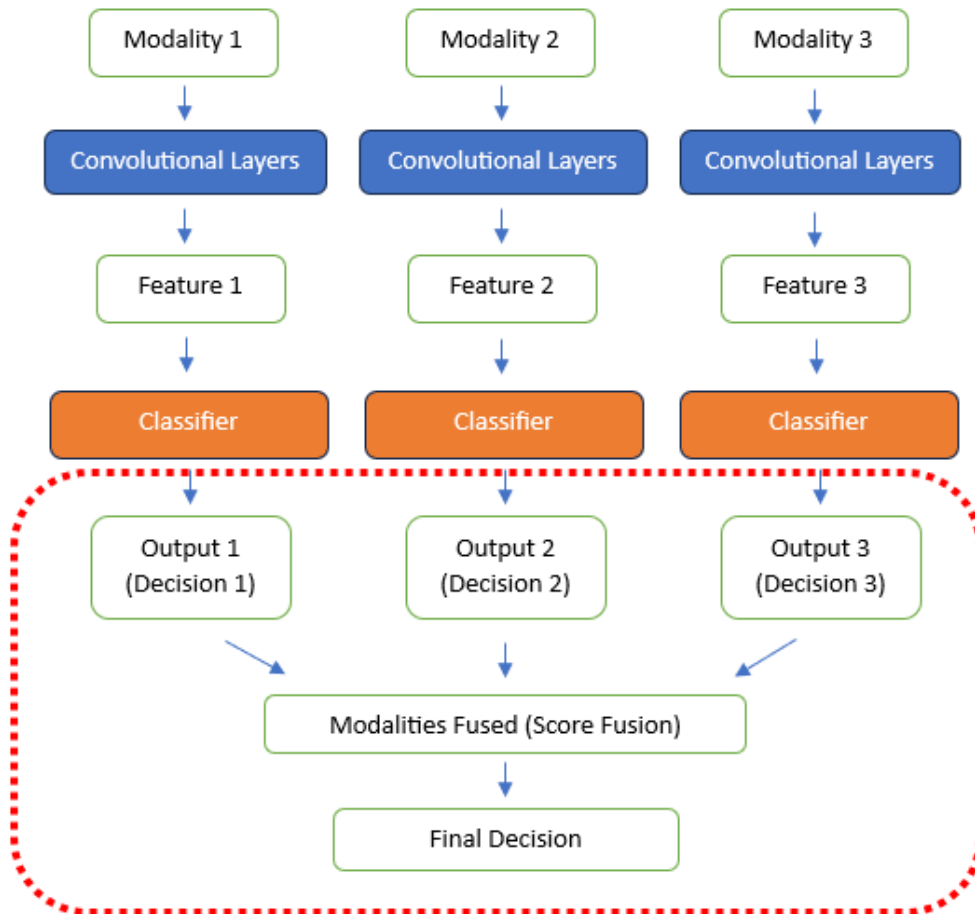


Figure 2.11 Score Fusion Process

Score fusion, also known as decision-level fusion, takes place at the decision outputs layer. From Figure 2.11, the output scores from different modalities are combined to make a final decision for the prediction. Score fusion leverages the complementary information captured by each modality to improve the overall prediction performance.

While these fusion methods offer advantages, they also come with certain drawbacks, as summarized in Table 2.5. Early fusion increases the complexity of the dataset, requiring equal quality for all modalities to achieve good performance. Feature fusion requires compatible dimensions of extracted features from different modalities and may introduce irrelevant or redundant information during merging. Score fusion, on the other hand, risks losing information about individual modalities due to fusion at a late stage.

Table 2.5 Drawbacks for Early Fusion, Feature Fusion, and Score Fusion Respectively

Technique	Cons
Early Fusion	<ul style="list-style-type: none"> - Increased dimensionality - Sensitive to quality variations
Feature Fusion	<ul style="list-style-type: none"> - Feature incompatibility - Feature redundancy
Score Fusion	<ul style="list-style-type: none"> - Loss of fine-grained information - Selection of fusion strategy

In summary, modality fusion techniques provide opportunities to enhance biometric systems by integrating multiple modalities. However, careful consideration of the strengths and limitations of each fusion method is essential to achieve optimal performance in biometric authentication applications.

2.10 Related Work

In this section, various research articles will be reviewed, which were firstly some articles related to biometric authentication system in bank security domain, followed by other biometric authentication research work in different domain, lastly some of the overview about the application of biometric in authentication system. These research articles were collected and studied to provide a solid fundamental background for this research work.

2.10.1 Related Work for Banking Security

In 2019, Sonal Kataria and Md Tabrez Nafis published their research work titled "Internet Banking Fraud Detection Using Deep Learning Based on Decision Tree and Multilayer Perceptron." They noted that detecting fraud had become more challenging due to the emergence of new fraud techniques. In their paper, they discussed the importance of fraud detection. Additionally, they trained a three-layer perceptron artificial neural network to detect fraudulent activities, achieving a satisfactory result of 84%.

In 2020, Ashwini M. Zinjurde and Vilas B. Kamble conducted research titled "Credit Card Fraud Detection and Prevention by Face Recognition" as internet banking became the most commonly used form of banking transactions. In their research work, they introduced a new approach to prevent fraudulent activities during online transactions, which involved a two-step authentication mechanism. By integrating OTP verification and face recognition, they developed a credit card fraud detection model with a user recognition accuracy of 87%. In the same year, Jiawei Ge, Zhou Fang, and Qiping Tao published their research work titled "Bank Card Number Recognition Based on Deep Learning" to enhance mobile Internet services. Their model used bank card images as input and integrated CNN, LSTM, and CTC to optimize recognition results. Their model achieved recall and precision scores of 75.2% and 86.32% respectively.

In 2022, Abdul Shareef Pallivalappil and his partners published their work titled "Evaluation of Digital Wallet Transaction Accuracy Using Machine Learning." Their research aimed to enhance the performance of digital wallet transactions. They observed that many studies in this domain could not address real-world scenarios due to the rapidly changing environment or limited scope. Abdul Shareef Pallivalappil et al. proposed a high-performance model, characterized by fast recognition speed, high accuracy, and a low error rate. In the same year, Taraknath Paul conducted related research work titled "Internet of Things and Machine Learning Banking Business Model Using Neuro Fuzzy Technique." In his paper, he discussed various attacks that could potentially affect online transactions. He trained an IoT and ML model using XGBoost and Logistic Regression, achieving accuracy scores of 73.5% and 74.6% respectively. Another research work by Oguntimilehin et al. focused on enhancing the security of mobile banking applications through a facial recognition model (Oguntimilehin et al., 2022). In their approach, they employed a CNN model, utilizing the collaboration of two computer vision models. Specifically, the Firebase Machine Learning vision model undertook the crucial tasks of face detection and image preprocessing, while the MobileFaceNet model handled user recognition and the transfer of data structures into the database. During the registration process, a dataset was meticulously compiled by capturing four user images. Noteworthy findings emerged, indicating a remarkable 100% recognition rate for face-only scenarios. However, challenges arose when users wore glasses, caps, or both, causing the recognition rate to significantly drop from 90% to 0%.

In 2023, Samiyeh Khosravi and her partners published their research work titled "Using Supervised Machine Learning Approaches to Detect Fraud in the Banking Transaction Network." In their paper, they mentioned that financial fraud activities have gradually increased, resulting in losses of personal and company properties. Moreover, fraud is becoming a growing problem in the banking sector, and detection has become more challenging. Hence, they trained various models using different ML algorithms. Based on their findings, the top two best algorithms were random forest and XGBoost.

Table 2.6 Related Work for Banking Security

No.	Author	Method	Result (%)
1	(Kataria & Nafis, 2019)	Dataset: Internet Banking Fraud Dataset Classifier: Multilayer Perceptron Algorithm	Accuracy: 84
2	(Zinjurde & Kamble, 2020)	Dataset: Real-time Face Detection Classifier: LBPH + CNN	Accuracy: 87
3	(Jiawei et al., 2020)	Dataset: Card Number Recognition Training Classifier: CNN + LSTM + CTC	Recall: 75.2 Precision: 86.32
4	(Pallivalappil et al., 2022)	Dataset: Digital Wallet Transaction Classifier: -	-
5	(Paul, 2022)	Dataset: Random Data Classifier: Logistic Regression & XGBoost	Accuracy (LR): 74.61 Accuracy (XGBoost): 73.5
6	(Oguntimilehin et al., 2022)	Dataset: Real-time Face Detection Classifier: CNN	Accuracy: 0-100
7	(Khosravi et al., 2023)	Dataset: Customer Transactions From Iran's Bank Classifier: CNN	Accuracy (XGBoost): 96 Accuracy (Random Forest): 96

Table 2.6 illustrates that there is still significant room for improvement in banking security with respect to recognition performance. It is important to note that the results are not flawless, and it is crucial to remember that failure to detect fraudulent activities can result in substantial losses. Additionally, some of these authors have also emphasized the significance of adaptability and sustainability of trained models over time. Over the course of time, model performance tends to degrade. However, utilizing biometric datasets in the recognition process offers a superior solution, as human biometric traits remain consistent over time, contributing to the stability of model performance. Therefore, multimodal biometrics have been confirmed to deliver even better results by incorporating multiple human biometric traits in the recognition process.

The subsequent subsection explores an examination of multimodal biometric authentication techniques employed in recent years. These studies are particularly significant as they offer valuable insights and serve as a fundamental reference for the implementation of robust multimodal biometric recognition systems.

2.10.2 Related Work for Multimodal Biometric Authentication

Next, several existing multimodal biometric recognition system had been examined. Additionally, some information from the research works were extracted and compiled in Table 2.7.

Table 2.7 Comparison of Models

No.	Author	Method	Result (%)
1	(Zhou et al., 2020)	Biometric: iris, palm vein, finger vein Dataset: CASIA, PolyU, SDU Algorithms involved: Gabor filter, LBP, LDA, PCA Type of Fusion: Hybrid fusion (Feature level + Decision level)	Accuracy: 99.33
2	(Gavisiddappa et al., 2020)	Biometric: iris, face, fingerprint Dataset: CASIA Algorithms involved: Histogram of oriented gradient, Local binary pattern, Gray-level co-occurrence matrix features like cluster prominence, inverse difference moment normalized, autocorrelation, minutiae feature extraction Type of Fusion: Feature level + Feature selection	Accuracy: 97.09
3	(Daas et al., 2020)	Biometric: finger knuckle print, finger vein Dataset: FV Algorithms involved: - Type of Fusion: Feature level, Score level	Accuracy: 98.84 (Feature level); 99.89 (Score level)
4	(Purohit & Ajmera, 2021)	Biometric: palm, fingerprint, ear Dataset: IIT Delhi , CASIA	Accuracy: 91.67

No.	Author	Method	Result (%)
		Algorithms involved: Gabor feature, HMSB operator, MRG Type of Fusion: Feature level + Feature Selection	
5	(Wang et al., 2022)	Biometric: face, finger vein Dataset: CASIA-WebFace, SDUMLA-FV, FV-USM Algorithms involved: Prewitt edge detection operator, Contrast-limited adaptive histogram equalization, Gabor Filter Type of Fusion: Feature level	Accuracy: >98.4
6	(A. Gona & Subramoniam, 2022a)	Biometric: face, ear, palmprint, fingerprint Dataset: ORL, USTB, NCUT, FVC 2002 Algorithms involved: Gabor filter, IPCA Type of Fusion: Feature level + Feature ranking	Accuracy: 98.54
7	(Rajasekar et al., 2022)	Biometric: iris, fingerprint Dataset: CASIA V3 iris, FVC 2006 fingerprint Algorithms involved: Wildes iris localization, Daugman's rubber sheet, Log gabor filter, Gabor filter, Line based Type of Fusion: Score level + Optimized Fuzzy Genetic Algorithm	Accuracy: 99.83 EER: 0.18

No.	Author	Method	Result (%)
8	(A. K. Gona & Subramoniam, 2022a)	Biometric: face, iris, fingerprint Dataset: ORL, MMU-IRIS, FVC 2002 Algorithms involved: Gaussian filter, GLCM, PCA Type of Fusion: -	Accuracy: 97.97 (iris); 94.02 (fingerprint); 94.01 (face)

From the summarized results in Table 2.7, all of the model performances exceed at least 91%, demonstrating that the usage of multimodal approaches is superior to the other methods discussed in earlier sections. However, in these research works, various algorithms were employed to preprocess and enhance image quality. While these techniques were effective, they were also time-consuming processes. In the upcoming paragraphs, each of the research works listed in Table 2.7 will be explained in more detail.

In a study conducted by Zhou et al., a hybrid fusion model combining iris, palm vein, and finger vein was proposed for a multi-biometric recognition system (Zhou et al., 2020). The researchers utilized datasets from CASIA, PolyU, and SDU databases, consisting of 100 classification classes. By employing PCA for feature extraction from iris and palm vein, and PCA with Gabor filter for finger vein, the authors achieved significant results. Through the use of score fusion (weighted voting strategy based on confidence factor) and feature fusion (multi-set discriminative correlation analysis method), the proposed model attained an impressive recognition rate of 99.33%.

Furthermore, Gavisiddappa, Mahadevappa, and Patil proposed a multimodal biometric authentication system utilizing Modified ReliefF Feature Selection and Multi Support Vector Machine (Gavisiddappa et al., 2020). The authors obtained biometric datasets from the CASIA database, utilizing 89 classes for training the authentication system. They employed various algorithms for feature extraction from iris and face images, including histogram of oriented gradient, local binary pattern, and Gray-level co-occurrence matrix features. For fingerprint features, they utilized minutiae feature extraction technique. The researchers performed feature fusion to combine the extracted features, followed by feature selection using the Modified Relief algorithm. Finally, a multi-class support vector machine was employed for subject classification, resulting in an accuracy score of 97.09%.

Another notable study by Daas et al. focused on a deep learning-based multimodal biometric recognition system that fused finger vein and finger knuckle print modalities (Daas et al., 2020). The dataset used was obtained from the FV database, with 106 classes for classifier training. The researchers applied image pre-

processing steps such as cropping ROI, data augmentation, and resizing. They employed pre-trained models (AlexNet, VGG16, and ResNet50) for feature extraction. By utilizing feature fusion techniques (concatenation and adding) and score fusion techniques (weighted sum, weighted product, and Bayesian rule), the researchers achieved highly accurate models, with the feature fusion (addition) model reaching an accuracy score of 98.84% and the weighted sum model achieving an outstanding accuracy score of 99.89%.

Similarly, Purohit and Ajmera proposed an optimal feature level fusion approach for secured human authentication in a multimodal biometric system (Purohit & Ajmera, 2021). The researchers utilized datasets from the IIT Delhi ear database version 1.0, CASIA palmprint, and CASIA-FingerprintV5. The modalities employed in this study were ear, palmprint, and fingerprint, with 121 classes used for training the multimodal classifier. The authors applied manual image pre-processing techniques to enhance image quality. Gabor filter was used for palmprint feature extraction, while Histogram-based Multi-Scale Binary Pattern (HMSB) was employed for feature extraction from ear and fingerprint images. The feature representations from different modalities were fused using the optimal feature level fusion gray wolf optimization (OGWO) technique. A multi-kernel SVM was employed for class recognition, resulting in an accuracy score of 91.67%.

Next, Wang, Shi, and Zhou presented a Convolutional Neural Network approach for a multimodal biometric recognition system (Wang et al., 2022). Their study focused on fusing face and finger vein features through feature fusion (concatenation). The face images were obtained from the CASIA-WebFace database, while finger vein images were sourced from SDUMLA-FV and FV-USM databases. By following a series of image pre-processing steps such as resize, normalization, crop ROI, and denoise, the researchers extracted features using pre-trained models, which were AlexNet and VGG-19 networks. The recognition rates achieved by both models were more than 98.4%.

In another study by A. Gona and Subramoniam, a robust multi-modal biometric system was proposed, integrating Convolutional Neural Network with improved

feature ranking (A. Gona & Subramoniam, 2022). The researchers gathered datasets from various databases, including ORL (Face), USTB (Ear), NCUT (Palmprint), and FVC-2002 (Fingerprint). Through a series of image pre-processing steps, such as normalization, ROI determination, and denoising using Gabor Filter, the raw images were prepared. Feature extraction was performed using the Improved Principal Component Analysis (IPCA) method. Subsequently, feature fusion was executed to merge different modalities, enhancing classifier performance through the Extended Borda-Count (EBC) method. The final step involved using a DLCNN to predict the output classes, achieving an accuracy rate of 98.54%.

Likewise, Rajasekar et al. aimed to enhance multimodal biometric recognition through an optimized fuzzy genetic algorithm (Rajasekar et al., 2022). The researchers utilized datasets from CASIA V3 iris and FVCC2006 fingerprint databases. For iris images, pre-processing steps included Wildes Iris Localization, Daugman's rubber sheet Iris Normalization, and Log Gabor Filter feature extraction. Additionally, fingerprint images underwent Gabor Filter image enhancement, core points ROI selection, and line-based feature extraction. The authors developed their own matching module to obtain matching scores from a database containing 700 different classes. Finally, score fusion was achieved using an optimized fuzzy genetic algorithm, resulting in an accuracy score of 99.83%.

Lastly, A. K. Gona and Subramoniam proposed a multimodal biometric recognition system using a Deep Learning Convolutional Neural Network (DLCNN) (A. K. Gona & Subramoniam, 2022b). The datasets used were obtained from the ORL (face), MMU-IRIS (iris), and FVC-2002 (fingerprint) databases. The number of classes varied across modalities, with 40 classes for face, 46 for iris, and 120 for fingers. The images underwent resizing and denoising using a Gaussian filter. Feature extraction was performed using a grey-level co-occurrence matrix (GLCM), followed by feature reduction using principal component analysis (PCA). The features from different modalities were fused through feature fusion, and the classifier employed was the designed DLCNN architecture. However, the authors provided accuracy scores only for individual biometric modalities, with the iris, fingerprint, and face models achieving scores of 97.97%, 94.02%, and 94.01%, respectively.

2.10.3 Overview of Biometric Authentication Approach

In this section, several overview papers related to biometric authentication were selected to study. This has contributed to the understanding of the technology and recent advancements. Basically, these recent overview articles were aimed to provide comprehensive understanding of the field.

Akhtar et al. delve into the captivating realm of biometrics, aiming to provide a comprehensive overview of the technology and its recent advancements (Akhtar et al., 2018). Their study traces the evolutionary path of biometrics, exploring the emergence of novel traits and research areas, as well as their wide-ranging applications. Highlighting the advantages of biometrics over traditional tokens or passwords, the authors emphasize its potential to prevent fraudulent activity. However, they also address the challenges posed by incorrect matches, shedding light on areas for further improvement. Ultimately, the paper underscores the profound impact that biometrics is expected to have on various aspects of life in the coming decade, emphasizing the significance of this technology for the future.

In the realm of recognition systems, Ma et al. embark on an exploration of multimodal biometrics that harmoniously integrate facial and ear features (Ma et al., 2020). Recognizing the power of combining these two modalities rather than relying on a single source, the authors elucidate the benefits of this approach. Categorizing existing multimodal biometric systems based on their fusion methods, the study unveils a promising finding: the amalgamation of multiple biometric sources leads to enhanced accuracy and robustness in recognition. Within this context, the authors shed light on the exceptional advantages offered by the ear as a biometric modality, thanks to its stable 3D structure and non-invasive capture method. To address potential data degradation challenges, the study proposes an adaptive biometric quality-based fusion approach that prioritizes high-quality samples. By concluding with an adaptive multimodal identification system that seamlessly integrates face and ear through sparse representation and leverages a general biometric quality assessment method, the authors lay the groundwork for future advancements in multimodal biometrics.

In their comprehensive review, Alsellami et al. survey the vast landscape of biometric traits, with a specific focus on multimodal biometrics (Alsellami et al., 2021). By exploring authentication systems that harness multiple biometric characteristics to identify individuals, the authors transcend the boundaries of unimodal approaches. Within this framework, they delineate biometric traits into two categories: physiological, including fingerprints, irises, and DNA, and behavioural, encompassing voice recognition and signature verification. Highlighting the limitations of relying solely on a single trait, the study illuminates the issues related to noisy data, intra-class variations, and vulnerability to deception. As the paper concludes, it becomes evident that multimodal biometrics emerge as a beacon of hope, offering more accurate and efficient results compared to their unimodal counterparts, thus surpassing previous benchmarks in accuracy and effectiveness.

Within the realm of biometric authentication systems, Albalawi et al. set out on a pioneering journey, exploring the integration of artificial intelligence (AI) (Albalawi et al., 2022). Their study presents a comprehensive examination of various biometric techniques utilized for user identification, categorizing them based on physiological and behavioural traits. Methodically comparing different algorithms, the authors astutely dissect the strengths and weaknesses of each approach. Moreover, the paper offers a groundbreaking proposition—an innovative system that harmonizes AI algorithms with iris recognition, accompanied by a meticulously outlined design. By summarizing the current state of biometric authentication technologies, the study unveils the immense potential of AI in enhancing accuracy and efficiency. As the authors present a hybrid approach that synergizes AI with support vector machines (SVM) for advanced iris recognition, they usher in a new era of biometric authentication systems.

In short, these studies collectively contribute to the understanding and advancement of biometric authentication systems, covering topics such as the benefits and challenges of biometrics, the integration of multiple modalities, and the potential of AI. Through their pioneering work, these researchers pave the way for a future where secure identification and access control are fortified by cutting-edge advancements in biometric technology.

2.10.4 Summary

Upon reviewing these literature articles, several research gaps can be identified. Firstly, there is a limited number of studies specifically focusing on multimodal biometric authentication systems in the banking domain, indicating a potential research gap in this specific area. Furthermore, the optimal procedures and techniques for achieving optimized performance remain unclear. Additionally, there is a need to investigate the impact of different factors, such as wearing accessories, on the performance of biometric authentication systems. Addressing these research gaps can contribute to the development of robust and effective biometric recognition models for secure authentication systems, particularly in the online banking sector.

2.11 Research Gaps

After a thorough analysis of the literature articles, it becomes evident that there exists a significant research gap when it comes to biometric authentication systems in the context of online banking security. While biometrics have gained considerable attention in various domains, including access control and identification, there is a dearth of studies specifically focusing on their implementation and effectiveness in the banking sector.

Online banking security is a critical concern in today's digital landscape, with cyber threats and identity theft posing substantial risks. Biometric authentication systems offer a promising solution to enhance security and protect sensitive financial information. However, the limited number of studies exploring the application of multimodal biometrics in online banking suggests an underexplored research area.

Moreover, the optimal procedures and techniques for achieving optimal performance in the context of online banking security. Each biometric modality, such as fingerprints, iris recognition, or voice recognition, has its strengths and weaknesses. Understanding the most effective fusion methods and algorithms for combining these

modalities in online banking authentication systems is a critical research gap that needs to be addressed.

Additionally, the impact of various factors on the performance of multimodal biometric authentication systems in online banking security warrants further investigation. For instance, the influence of wearing accessories, such as glasses or jewelry, on the accuracy and reliability of biometric systems requires empirical analysis. Such factors can introduce variations in biometric data, affecting the system's performance and posing challenges in real-world scenarios. Exploring and understanding these factors can help develop more robust and adaptable biometric authentication systems tailored specifically for online banking security.

By addressing these research gaps and focusing on the development of robust and effective multimodal biometric recognition models in the context of online banking security, researchers can contribute to enhancing the security, reliability, and user experience of online banking systems. This research would enable the creation of advanced authentication mechanisms that provide a seamless and secure banking experience for users, mitigating the risks associated with cyber threats and safeguarding financial transactions.

2.12 Chapter Summary

In this chapter, the focus is on providing an in-depth understanding of online banking and security, starting with an introduction to the topic. The discussion progresses to highlight the limitations of traditional security approaches, such as passwords, security tokens, and two-factor authentication methods. To overcome these limitations, the significance of biometric authentication methods in the field of security is emphasized. While unimodal biometric authentication methods offer potential advantages, they also have their own limitations. To address these limitations, researchers have proposed multimodal biometric authentication methods that combine multiple biometric traits to achieve improved performance results. Additionally, the concept of transfer learning is introduced, showcasing the effectiveness of leveraging

pre-trained state-of-the-art architectures to build robust biometric recognition models. The chapter further explores three common techniques for modality fusion: early fusion, feature fusion, and score fusion. Each technique is explained in detail to provide a comprehensive understanding of their applications and benefits. In addition, recent related work in the field is reviewed, revealing existing research gaps that require further exploration. The subsequent chapter will delve into the research methodology, presenting a clear and detailed explanation of the entire research flow. This will enable a comprehensive understanding of the methods employed in this study and how they contribute to addressing the research gaps identified in the previous chapter.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, the research methodology would be shown clearly. First of all, the research design framework would be discussed, followed by the brief explanation from phase to phase. Lastly, addition information about the dataset collected and performance evaluation metrics would be explained.

3.2 Research Workflow

In this research study, it would be separated into four different phases, which started from Phase 1 Problem Background Analysis, Phase 2 Literature Review, Phase 3 Experiment and Evaluation and Phase 4 Result Documentation. More details were showed in Figure 3.1 as below. In short, Phase 1 had conducted the selection of research title, study of problem background on online banking issues and formulation of research purpose, objectives and scope. Next, the framework continued with Phase 2 Literature Review, which mainly focused on the important topics which relevant to this research title, such as the online banking authentication, image classification techniques, unimodal and multimodal authentication systems and concept of transfer learning. Subsequently, the Phase 3 Experimental and Evaluation was the phase where the research experimental part be shown, starting off with the workflow design, data collection, data pre-processing, feature extraction, training of classification layer, and lastly the model evaluation process. In last phase, it was the result documentation part, where the comparisons, interpretation, and result discussion would be shown in documented format.

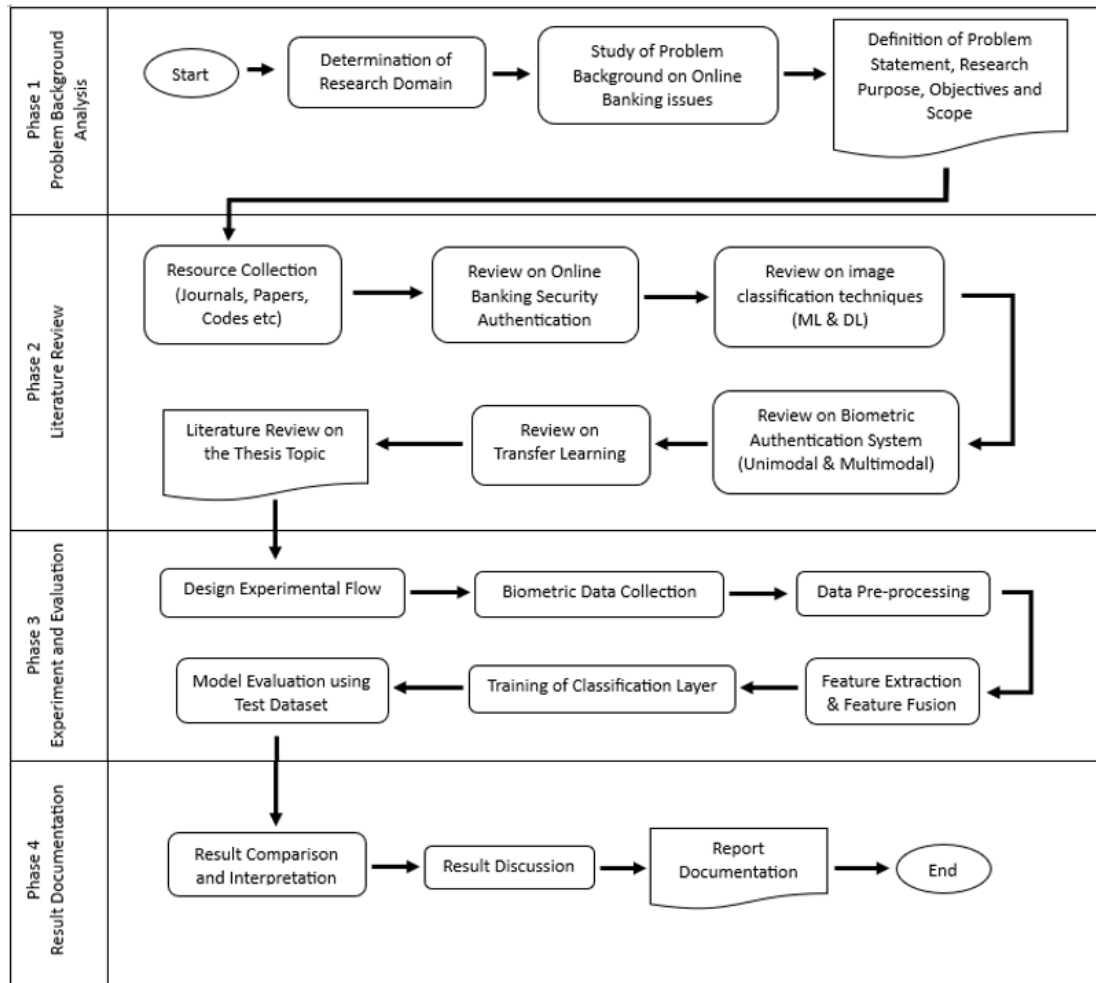


Figure 3.1 Research Framework

3.2.1 Phase 1: Problem Background Analysis

In the initial phase of this research, the problem background revolves around the pervasive occurrence of security attacks and threats. These issues persist due to the absence of a robust and efficient authentication method, hindering effective mitigation of such risks. Consequently, the need arises for the development of an enhanced online banking security solution through the utilization of transfer learning and the hybridization of multiple biometric modalities. To address this challenge, the research aims to train an effective authentication model using transfer learning techniques. By leveraging the knowledge acquired from pre-trained models in diverse domains, the proposed approach seeks to enhance online banking security significantly.

Furthermore, hybridization of multimodal biometric features, such as iris, face, and palmprint is applied to create a robust and accurate authentication system. In summary, Chapter 1 provides a comprehensive overview of the research purpose, objectives, and scope. It serves as a clear roadmap for the study, outlining the specific goals to be achieved and the extent of the research investigation.

3.2.2 Phase 2: Literature Review

Next, the second phase of the research would be an extensive literature review was conducted to gather valuable insights on key topics relevant to the research domain. Multiple sources, including journals, articles, and videos, were utilized to compile the necessary information. The literature review encompassed several important aspects, starting with a comprehensive examination of online banking security authentication. This involved studying various techniques such as machine learning and deep learning algorithms for image classification.

Furthermore, the review delved into the intricacies of both unimodal and multimodal biometric authentication systems. Understanding the strengths and limitations of these systems is crucial for the development of an effective authentication model. Lastly, the concept of transfer learning was thoroughly studied, as it plays a significant role in leveraging pre-existing knowledge for improving online banking security. This phase holds great importance in the research process, as it ensures a solid foundation of concepts, knowledge, and information before proceeding to the next phase.

Moreover, the most commonly available human biometrics were chosen for this research work, such as face, iris, and palmprint. The features representing these three modalities will be extracted and combined to form a unified feature set that encompasses all three biometric characteristics.

3.2.3 Phase 3: Experiment and Evaluation

The focus shifted towards the experiment and evaluation process in the third phase of this research study. This phase involved designing the experimental workflow to ensure a systematic approach. The subsequent steps included were explained as below referring to Figure 3.2.

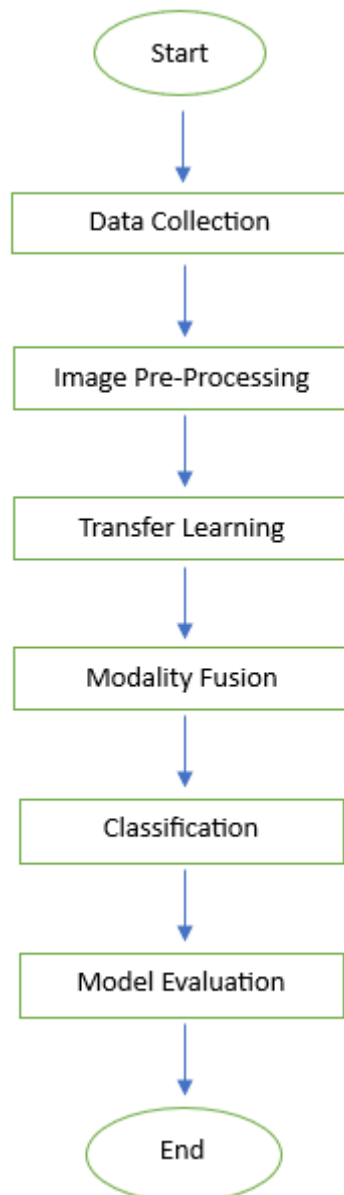


Figure 3.2 Research Methodology

3.2.3.1 Data Collection

The CASIA dataset was obtained from BIT website (*BIT*, 2005). In this research, the selected biometrics were iris, face, and palmprint. The brief description would be provided as below:

CASIS-Iris_Thousand contains 20,000 iris images from 1000 classes. They are collected using IKEMB-100 camera, it was dual-eye iris camera which provides effect of “What You See Is What You Get”. The left and right iris pictures were captured for every classes. Besides, all images have dimension of 640x480 pixel, and were in gray level JPEG images. Since it is the first publicly available Iris dataset with one thousand classes, it is well suited to be applied in iris recognition research.

CASIA-FaceV5 contains 2,500 color facial images of 500 classes. These images were captured using Logitech USB camera in one session. The volunteers in this datasets include graduate students, workers, waiters, etc. All face images are 16 bit color BMP files, with dimension of 640x480 pixel. In the dataset, typical intra-class variations were found, such as illumination, pose, expression, eye-glasses, and imaging distance.

CASIA Palmprint Image Database contains total number of 5,502 palmprint images from 312 classes. For each subject, left and right palmprint images were collected. These pictures were all in 8 bit gray level JPEG, with dimension of 640x480 pixel. Images were captured using CMOS camera which placed on top of their device. In their device, it ensures that the illumination was evenly distributed when capturing palmprint image.

In this research project, the goal is to perform modality fusion by combining iris, face, and palmprint images to represent individuals or classes. The challenge at hand is that the originally downloaded datasets for these biometric modalities do not pertain to the same set of individuals. To address this challenge, a new dataset is being created. For example, this new dataset will comprise data from 100 unique individuals for each modality, resulting in 100 iris users, 100 face users, and 100 palmprint users.

The next step involves fusing the biometric data from these three modalities, assuming they belong to the same individuals or classes. The aim is to create a unified representation for each individual or class using the combined information from the three biometric datasets. Subsequently, a recognition model will be trained to recognize and classify these individuals or classes. Given the 100 users in each modality, this research constitutes a 100-class recognition problem.

3.2.3.2 Image Pre-Processing

Next, the collected data then underwent a data preprocessing step, where the biometric images were processed and transformed into the desired format, preparing them for model training. Furthermore, data augmentation was performed to increase the size of dataset. Further details regarding this process will be shown in Figure 3.3 as below.

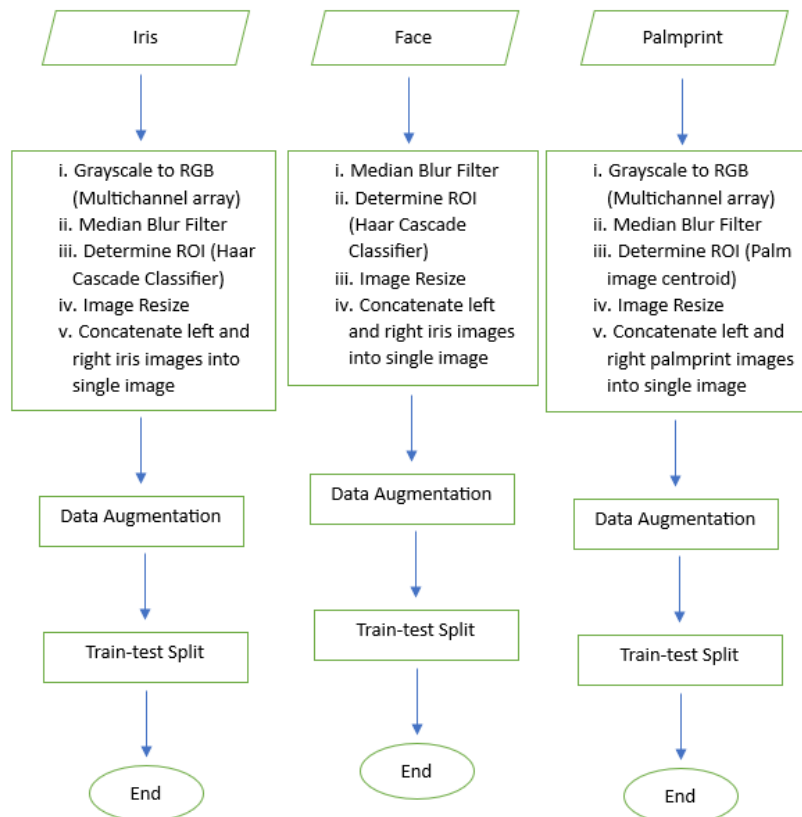


Figure 3.3 Image Preprocessing Steps for Iris, Face and Palmprint Dataset

3.2.3.3 Transfer Learning

In this section, some of the pre-trained models were selected, which were VGG16, ResNet101V2, and InceptionResnetV2 models. These models were selected as they were previously been applied in unimodal recognition system, which were already discussed in Chapter 2. In this research work, these pretrained models would be applied for extracting features from iris, face and palmprint images respectively. Therefore, only the CNN layers would be preserved. These models summary will be shown as Table 3.1. Besides of that, the models layers and configurations had been tabularized with clearer information. Where Table 3.2 shows the information about model VGG16, Table 3.3 and Table 3.4 shows the details about model ResNet101V2 and InceptionResNetV2 respectively.

Table 3.1 Overall Summary

Biometrics	Model	Input Size	Number of Parameters
Iris	VGG16	256 x 128 x 3	14,714,688
Face	ResNet101V2	128 x 128 x 3	42,626,560
Palmprint	InceptionResnetV2	256 x 128 x 3	54,336,736

Additionally, these pretrained models were obtained from Tensor Flow-Keras Applications. All of them were welly trained using Imagenet data, where it consists of 1000 classes of images, 1,281,167 training images, 50,000 validation images, and 100,000 test images.

Table 3.2 VGG16 Model Layers and Configurations

Layers	Dimensions	Non -Trainable Parameters
• Input_3	256 x 128 x 3	0
• Block1_conv1	256 x 128 x 64	1,792
• Block1_conv2	256 x 128 x 64	36,928
- Block1_pool (MaxPooling2D)	128 x 64 x 64	0
• Block2_conv1	128 x 64 x 128	73,856
• Block2_conv2	128 x 64 x 128	147,584
- Block2_pool (MaxPooling2D)	64 x 32 x 128	0
• Block3_conv1	64 x 32 x 256	295,168
• Block3_conv2	64 x 32 x 256	590,080
• Block3_conv3	64 x 32 x 256	590,080
- Block2_pool (MaxPooling2D)	32 x 16 x 512	0
• Block4_conv1	32 x 16 x 512	1,180,160
• Block4_conv2	32 x 16 x 512	2,359,808
• Block4_conv3	32 x 16 x 512	2,359,808
- Block2_pool (MaxPooling2D)	16 x 8 x 512	0
• Block5_conv1	16 x 8 x 512	2,359,808
• Block5_conv2	16 x 8 x 512	2,359,808
• Block5_conv3	16 x 8 x 512	2,359,808
- Block2_pool (MaxPooling2D)	8 x 4 x 512	0
Total	-	14,714,688

Table 3.3 ResNet101V2 Model Layers and Configurations

Layers	Dimensions	Non -Trainable Parameters
• Input_1	128 x 128 x 3	0
• Conv1_conv	64 x 64 x 64	9472
- Pool1_pool (MaxPooling2D)	32 x 32 x 64	0
Conv Block stage 1 • Conv2_block1_1 ↓ • Conv2_block3_3	16 x 16 x 256	217,856
Conv Block stage 2 • Conv3_block1_1 ↓ • Conv3_block4_3	8 x 8 x 512	1,078,272
Conv Block stage 3 • Conv4_block1_1 ↓ • Conv4_block23_3	4 x 4 x 1,024	26,329,600
Conv Block stage 4 • Conv5_block1_1 ↓ • Conv5_block3_3	4 x 4 x 2,048	14,991,360
Total	-	42,626,560

Table 3.4 InceptionResNetV2 Model Layers and Configurations

Layers	Dimensions	Non -Trainable Parameters
Input_2	256 x 128 x 3	0
Stem Module	29 x 13 x 320	441,920
10 x Inception-Resnet-A Module	29 x 13 x 320	1,234,080
Reduction-A Module	14 x 6 x 1088	2,666,240
20 x Inception-Resnet-B Module	14 x 6 x 1088	22,549,120
Reduction-B Module	6 x 2 x 2080	3,883,008
10 x Inception-Resnet-C Module	6 x 2 x 2080	20,362,880
Conv_7b	6 x 2 x 1536	3,199,488
Total	-	54,336,736

3.2.3.4 Modality Fusion

In this research work, feature level fusion was applied to combine multiple biometric modalities. Where the feature extracted from iris, face, and palmprint images were merged through concatenation layer. Feature level fusion was chosen as it was generally more effective as compared to score level and decision level fusion for multimodal biometric recognition (Purohit & Ajmera, 2021; Safavipour et al., 2022). In short, feature level fusion would produce the best vector for the system to make best decision as the vector space produced would have maximum distinction and minimum dimensions. In Table 3.5, the information of the feature level fusion for this research work has been listed down in Table 3.5.

Table 3.5 Configurations at Fully Connected Layers

Layers	Dimensions	Trainable Parameters
Flatten (ResNet101V2)	1 x 1 x 32,768	0
Flatten_1 (InceptionResNetV2)	1 x 1 x 18,432	0
Flatten_2 (VGG16)	1 x 1 x 16,384	0
Concatenate	1 x 1 x 67,584	0
Dense (Softmax Layer)	Model_100: 1 x 1 x 100 Model_300: 1 x 1 x 300	6,758,500 20,275,500

3.2.3.5 Classification

Moving forward, the training of the classification layer took place, utilizing a 1 hidden layer of softmax classification method. Softmax activation function was applied for the proposed model as it was multiclass recognition model. Following the training phase, the performance of the model was then tested using a separate test and validate dataset.

3.2.3.6 Model Evaluation

During the model evaluation process, various crucial evaluation metrics were measured to assess the performance of the proposed model. Each of these metrics is thoroughly explained as subsequent subtopics, providing a comprehensive understanding of the model's performance and its ability to accurately identify and authenticate individuals in the online banking biometric authentication system.

(a) Confusion Matrix

Confusion Matrix, is an essential report or table in which it indicates the classification result for the ML model. It consists of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Referring to Table 3.1, TP and TN are the cases where class prediction is correct for actual positive and actual negative classes. FP is the case where the false prediction is made when the actual class is negative. FN is the case where the false prediction is made when the actual class is positive.

In online bank security systems, TP means that the system able to recognized the authorized user correctly, TN means that the system able to detect unauthorized user correctly. Besides, FP means that the system has recognized the unauthorized user as the correct one. For FN, it means that the system has failed to recognized the authorized user.

Next, by using the value obtained from the confusion matrix, other performance metrics such as accuracy, precision, recall and ROC could be measured.

Table 3.6 Confusion Matrix

		Predict	
		Class	
Actual	Positive(1)	True Positive (TP)	False Negative (FN)
	Negative(0)	False Positive (FP)	True Negative (TN)

(b) Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

Equation 3.1 presents the formula for calculating the accuracy of the model. The numerator represents the sum of correct predictions, while the denominator represents the total number of predictions made. This calculation allows us to determine the overall accuracy of the model by quantifying the proportion of correct predictions. As the accuracy value approaches 1, it indicates a higher level of accuracy, indicating that the model has successfully made a greater number of correct predictions.

(c) Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.2)$$

Equation 3.2 illustrates the precision of the model. Precision specifically focuses on the proportion of false positive (FP) predictions made by the model. As the precision value approaches 1, it indicates that the model has made fewer FP predictions, meaning that it has a higher ability to accurately identify true positive

instances. In other words, a precision value closer to 1 signifies a lower rate of incorrectly identifying instances as positive when they are actually negative, thus indicating better precision performance by the model.

(d) Recall / Sensitivity / True Positive Rate

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.3)$$

Equation 3.3 presents the formula for calculating the recall of the model. The recall metric quantifies the proportion of false negative (FN) predictions made by the model. By examining the formula, it provides insight into the number of FN predictions. As the recall value approaches 1, it indicates that the model has made fewer FN predictions, implying that it has a higher capability to accurately identify true positive instances. In simpler terms, a recall value closer to 1 indicates a lower rate of incorrectly classifying positive instances as negative, thus highlighting better recall performance by the model.

(e) Receiver Operating Curve (ROC)

In summary, the Receiver Operating Characteristic (ROC) curve is plotted using True Positive Rate (TPR) and False Positive Rate (FPR) values. Figure 3.2 illustrates different scenarios for the ROC curve, including the (a) ideal case, (b) bad case, and (c) worst case. In the ideal case, the TPR value is 1, and the FPR value is 0, indicating accurate classification of all data. Conversely, the bad case represents a scenario where the TPR value is 0 and the FPR value is 1, indicating no correct classifications and numerous false predictions made by the model. The worst case scenario signifies a model that completely fails, lacking the ability to differentiate between positive and negative classes. Furthermore, the Area Under the Curve (AUC) represents the area under the ROC curve and ranges between 0 and 1. A higher AUC value, closer to 1, indicates better model performance.

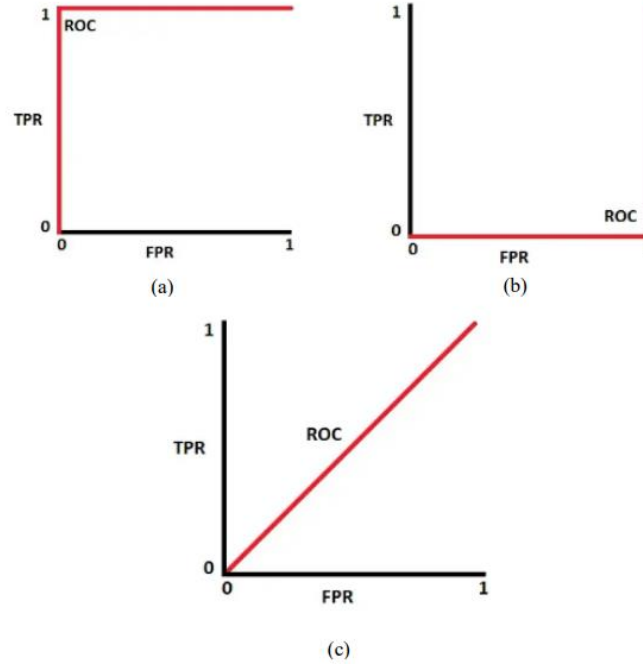


Figure 3.4 Example of ROC curve (a) Ideal Case (b) Bad Case (c) Worst Case (*Understanding AUC - ROC Curve* / by Sarang Narkhede / *Towards Data Science*, n.d.)

(f) Specificity / True Negative Rate

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3.4)$$

Equation 3.5 illustrates the formula for calculating specificity, which is a vital parameter for evaluating model performance. Specificity measures how effectively the model makes correct predictions for actual negative classes. It provides insight into the model's ability to accurately identify and classify negative instances. A higher specificity value indicates that the model has a lower rate of incorrectly classifying negative instances as positive, thus highlighting its capacity to distinguish between positive and negative classes effectively.

(g) False Positive Rate (FPR)

$$FPR = \frac{FP}{TN + FP} = 1 - \textit{specificity} \quad (3.5)$$

Equation 3.7 provides the formula for calculating the False Positive Rate (FPR), which can be obtained by subtracting the specificity value from 1. The FPR quantifies the rate of false positive predictions made by the model when predicting the negative class. It indicates the proportion of negative instances that are incorrectly classified as positive. By understanding the FPR, we can evaluate how well the model distinguishes between negative and positive classes, specifically in terms of minimizing false positive predictions.

(h) False Accepting Rate (FAR)

$$FAR = \frac{FP}{FP + TN} \quad (3.6)$$

Equation 3.8 provides the formula for calculating the False Acceptance Rate (FAR). This metric is crucial in assessing the model's capability to correctly identify and recognize actual positive classes during the authentication process. The FAR measurement helps determine the extent to which the model may mistakenly accept unauthorized users or false positives as genuine, which is a critical consideration in ensuring the security and reliability of online banking systems. By analyzing the FAR, we can evaluate the model's ability to effectively distinguish between legitimate users and potential impostors, thereby enhancing the overall security and trustworthiness of online banking transactions.

(i) False Rejection Rate (FRR)

$$FRR = \frac{FN}{FN + TP} \quad (3.7)$$

In equation 3.9, the formula for calculating the False Rejection Rate (FRR) is presented. This metric provides insight into whether the model is capable of correctly recognizing actual positive classes during the authentication process. The FRR measures the rate at which the model erroneously rejects genuine positive instances, indicating the proportion of false negative predictions made by the model. Evaluating the FRR allows us to assess the model's ability to accurately identify and authenticate legitimate users in online banking security.

3.2.4 Phase 4: Result Documentation

The final phase of this research study focuses on result documentation, where all the findings, discussions, interpretations, and analyses are compiled into a well-structured thesis format. This documentation ensures that all the relevant information is presented and reported coherently for future reference and to contribute to the body of knowledge in the field.

In addition to presenting the results, this phase also includes a comparison between the proposed model developed in this research and other models proposed by previous studies. This comparative analysis enhances the quality of future work by providing insights into the strengths and limitations of different approaches.

By completing this phase, the research study concludes with a comprehensive and informative thesis, which serves as a valuable resource for researchers and practitioners interested in the field of online banking security and feature fusion.

3.3 Chapter Summary

Throughout this chapter, the steps for the planning process from problem background analysis to result documentation of the thesis had been covered. Next, the four different phases were clearly discussed. In the next coming chapter, the research work continued with the research implementation.

CHAPTER 4

RESEARCH IMPLEMENTATION

4.1 Introduction

In this chapter, the flow of the proposed model will be clearly explained. First and foremost, the preparation stage for the implementation process of this experimental work will be clearly presented. Lastly, the proposed model, along with the algorithms for each subsequent stage, is demonstrated as shown below.

4.2 Preparation Stage

First of all, all programs were run using Laptop. The computer specs used would be AMD Ryzen 5 4600H with Radeon Graphics 3.00 GHz (CPU) with 16GB usable RAM. Then the programming language used would be Python 3.8.10. Besides, the environment used were Jupyter Notebook. These specs were used especially for debugging and image processing as shown in Figure 4.1. As the size of dataset is too heavy for the modelling training process, data augmentation and model training were performed using Google Colab Pro. Where it provides very powerful GPU for us to execute the codes. In Figure 4.2, the GPU settings were shown, the selected GPU type was A100, and the runtime shape was set to High-Ram. These enable us to have much higher computational resources to process huge and complex dataset.

Device specifications

IdeaPad Gaming 3 15ARH05

Device name	LAPTOP-GLE7GHSB
Processor	AMD Ryzen 5 4600H with Radeon Graphics 3.00 GHz
Installed RAM	16.0 GB (15.4 GB usable)

Figure 4.1 Device Specifications (CPU)

Notebook settings

Hardware accelerator
GPU ▼ ⓘ

GPU type
A100 ▼

Runtime shape
High-RAM ▼

☐ Omit code cell output when saving this notebook

Cancel Save

Figure 4.2 Google Colab Runtime Settings

4.3 Proposed Method & Algorithms

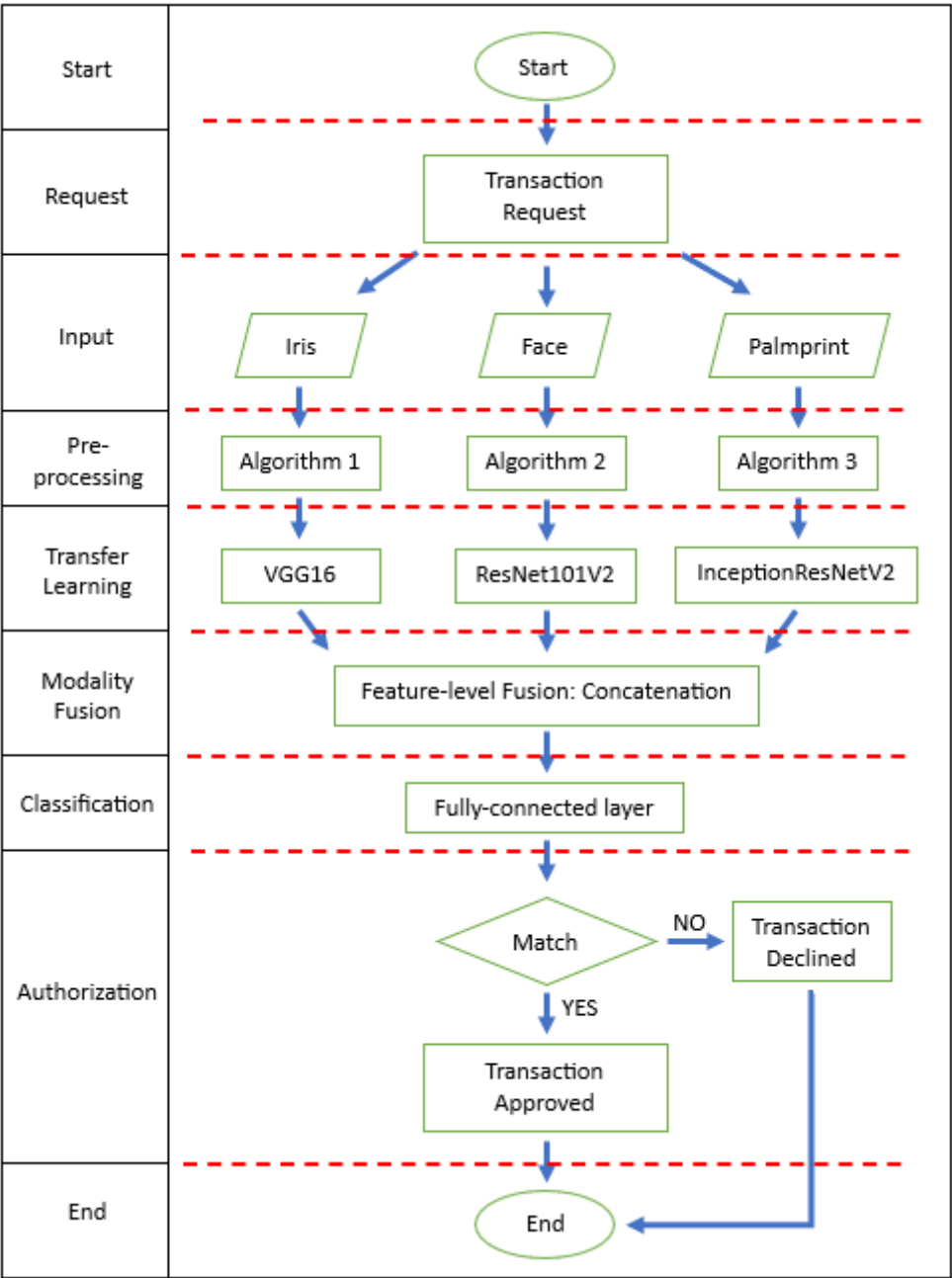


Figure 4.3 Proposed Model

In Figure 4.3, an overview of the proposed model has been displayed. In essence, the recognition model initiates when a user requests approval for their online transaction. The model collects three distinct biometric data types from the user: iris, face, and palmprint. Subsequently, a series of image pre-processing steps are employed to ensure that the data conforms to the desired format before being fed into the model

network. Within the model, transfer learning techniques are applied to extract features from these multiple biometrics. These features are combined to create a feature bank. This is achieved through a process known as concatenation in feature-level fusion. In the classification stage, the feature bank passes through a fully-connected layer, consisting of a single hidden layer with softmax activation function. Finally, the model produces a probabilistic result. If this result matches the data stored in the dataset, the transaction is approved; otherwise, it is rejected. This method ensures a robust and secure approval process for online transactions.

From the paragraph above, it outlines the general flow of the recognition model. Before that, model training must be completed and will be presented in the upcoming subtopics. Additionally, specific algorithms for each consequential process are also outlined.

4.3.1 Data Pre-processing & Augmentation

In this section, a clear listing of algorithms 1-3 is provided, with each algorithm describing the coding steps for preprocessing iris, face, and palmprint data. Furthermore, algorithm 4 is presented to detail the data augmentation process below:

(a) Iris

ALGORITHM 1: PRE-PROCESSING FOR IRIS

```

1  Start
2  Initialize empty lists datasetL and datasetR to store image data for left and
   right eyes, respectively.
3  Load the Haar Cascade Classifier for eye detection using OpenCV.
4  Iterate over the user directories in the current path:
5      For each user directory, iterate over the subdirectories "L" and "R"
      (presumably corresponding to left and right eyes):
6          For each subdirectory:
7              Create a path to the image folder.
8              Retrieve image filenames from the subdirectory.
9              Randomly select 5 images from the subdirectory.
10             Separate the selected images from the remaining images.
11             Process each selected image:
12                 Load the image in grayscale.
```

ALGORITHM 1: PRE-PROCESSING FOR IRIS

```
13   If eyes are detected:
14       Draw rectangles around the detected eyes.
15       Crop and resize the region containing the eyes to 128x128
        pixels.
16       Apply median blur to the image.
17       Expand the image dimensions to have three color channels.
18       Append the processed image and the user index to datasetL
        (or datasetR as appropriate).
19   end if
20   If no eyes are detected, print "no eyes detected"
21   end process
22 end for
23 end iterate
24 Initialize empty lists x_L, y_L, x_R, and y_R to store features and labels for
    left and right eyes, respectively.
25 Populate x_L, y_L, x_R, and y_R by extracting features and labels from
    datasetL and datasetR.
26 Concatenate the feature vectors for left and right eyes into a single feature
    matrix x_concat.
27 Convert the labels into a numpy array y.
28 Save x_concat and y as pickle files (binary serialization) for further use.
29 End
```

(b) Face

ALGORITHM 2: PRE-PROCESSING FOR FACE

```
1   Start
2   Initialize an empty list dataset to store image data along with user labels.
3   Load the Haar Cascade Classifier for frontal face detection using OpenCV.
4   Iterate over the user directories in the current path:
5       Create a path to the user's image folder.
6       Change the current working directory to the user's image folder.
7       Iterate over image files in the user's image folder:
8           Create the full path to the image file (face_imag).
9           Read the image in color (BGR format).
10          Use the Haar Cascade Classifier to detect faces.
11          For each detected face:
12              Crop and extract the region of interest (ROI) containing the
                face.
13              Resize the ROI to 128x128 pixels.
14              Append the processed face image and the user index to the
                dataset list.
15          end for
16      end iterate
17 end iterate
18 Initialize empty lists x and y to store features and labels, respectively
```

ALGORITHM 2: PRE-PROCESSING FOR FACE

- 19 *Populate x and y by extracting features (face images) and labels (user indices) from dataset.*
 - 20 *Reshape the x array to have dimensions $(-1, 128, 128, 3)$ to ensure a consistent format for image data.*
 - 21 *Reshape the y array to have dimensions $(-1, 1)$.*
 - 22 *Save the feature data (X) and labels (y) as pickle files for later use.*
 - 23 **End**
-

(c) Palmprint

ALGORITHM 3: PRE-PROCESSING FOR PALMPRINT

- 1 **Start**
- 2 *Initialize three empty lists: dataset, datasetL, and datasetR. dataset stores image data with user labels, and datasetL and datasetR store left and right palm images, respectively.*
- 3 *Define substring identifiers for left (substring_L) and right (substring_R) palm images and specify the number of images to select (num_to_select).*
- 4 *Iterate over user directories in the current path (cur_path):*
 - 5 *Create a path to the user's image folder.*
 - 6 *Change the current working directory to the user's image folder.*
 - 7 *Initialize two empty lists: left and right to store left and right palm image filenames.*
 - 8 *Iterate over image files in the user's image folder:*
 - 9 *If the filename contains the substring "l," add it to the left list.*
 - 10 *If the filename contains the substring "r," add it to the right list.*
 - 11 **end iterate**
 - 12 *Shuffle the left and right lists to randomize the selection of palm images.*
 - 13 *Select the first num_to_select images from both the left and right lists.*
 - 14 *Iterate over the selected left and right palm image filenames and process them:*
 - 15 *Create full paths for left (L_img) and right (R_img) palm images.*
 - 16 *Read the left image (L_img) in grayscale.*
 - 17 *Calculate the centroid of the left palm image.*
 - 18 *Determine the top-left corner of the Region of Interest (ROI) and its size.*
 - 19 *Extract the ROI from the left image, resize it to 128x128 pixels, and expand its dimensions to have three color channels.*
 - 20 *Read the right image (R_img) in grayscale.*
 - 21 *Calculate the centroid of the right palm image.*
 - 22 *Determine the top-left corner of the ROI and its size.*
 - 23 *Extract the ROI from the right image, resize it to 128x128 pixels, and expand its dimensions to have three color channels.*
 - 24 *Append the processed left and right palm images along with the user index to datasetL and datasetR.*
 - 25 **end iterate**
 - 26 **end iterate**

ALGORITHM 3: PRE-PROCESSING FOR PALMPRINT

27 *Initialize* empty lists x_L , y_L , x_R , and y_R to store features and labels for left and right palm images, respectively.
28 *Populate* x_L , y_L , x_R , and y_R by extracting features (processed palm images) and labels (user indices) from datasetL and datasetR.
29 *Reshape* the label array (y_L) to have dimensions (-1, 1).
30 *Concatenate* the feature vectors for left and right palm images into a single feature matrix (x_{concat}).
31 *Save* the feature data (x_{concat}) and labels (y) as pickle files for later use.
32 *End*

(d) Data Augmentation

ALGORITHM 4: DATA AUGMENTATION

1 *Start*
2 *Load* the face image data and labels from pickle files and save into X_{face} and y_{face} .
3 *Load* the palm image data and labels from pickle files and save into X_{palm} and y_{palm} .
4 *Load* the eye image data and labels from pickle files and save into X_{eye} and y_{eye} .
5 *Initialize* empty lists to store processed data and labels for face, palm, and eye images:
6 $face_train_dataset$ and $face_train_label$ for face training data and labels.
7 $face_vt_dataset$ and $face_vt_label$ for face validation/test data and labels.
8 $palm_train_dataset$ and $palm_train_label$ for palm training data and labels.
9 $palm_vt_dataset$ and $palm_vt_label$ for palm validation/test data and labels.
10 $eye_train_dataset$ and eye_train_label for eye training data and labels.
11 $eye_vt_dataset$ and eye_vt_label for eye validation/test data and labels.
12 *end initialize*
13 *For* face image data processing:
14 *For* each face image in X_{face} :
15 *Reshape* the image to have a batch size of 1.
16 *Apply* data augmentation using $datagen_train$ for training data and $datagen_validation$ for validation/test data.
17 *Append* the augmented images to the respective datasets ($face_train_dataset$ and $face_vt_dataset$) and labels ($face_train_label$ and $face_vt_label$).
18 *Control* the number of augmented samples (10 for training and 2 for validation/test).
19 *end for*
20 *end for*
21 *For* palm image data processing:

ALGORITHM 4: DATA AUGMENTATION

```
22   Follow the same processing steps as in face image data processing, but
    apply them to X_palm and the corresponding datasets and labels
    (palm_train_dataset, palm_train_label, palm_vt_dataset, and
    palm_vt_label).
23   end for
24   For eye image data processing:
25       Follow the same processing steps as in face image data processing, but
        apply them to X_eye and the corresponding datasets and labels
        (eye_train_dataset, eye_train_label, eye_vt_dataset, and eye_vt_label).
26   end for
27   After processing, convert the datasets and labels to NumPy arrays:
28       For face data, create face_train and label_train arrays.
29       For palm data, create palm_train and palm_train_label arrays.
30       For eye data, create eye_train and eye_train_label arrays.
31   end convert
32   Split the validation and test data for face, palm, and eye images using
    train_test_split from scikit-learn:
33       For face images, create x_valid_face, X_test_face, y_valid_face, and
        y_test_face.
34       For palm images, create x_valid_palm, X_test_palm, y_valid_palm,
        and y_test_palm.
35       For eye images, create x_valid_eye, X_test_eye, y_valid_eye, and
        y_test_eye.
36       Split the data into validation and test sets, specifying a test size of 50%
        and ensuring that class proportions are maintained (stratified split).
37   End
```

4.3.2 Transfer Learning

Algorithm 5 outlines the creation of a pre-trained models for iris, face, and palmprint for feature extraction in this research work as follows:

ALGORITHM 5: DEFINING PRE-TRAINED MODEL (VGG16, RESNET101V2, INCEPTIONRESNETV2)

```
1   Start
2   Define the image sizes for different image types:
3       image_size1: 128x128 pixels for face images.
4       image_size2: 256x128 pixels for iris and palm image.
5   end define
6   Define the Eye Model:
7       Initialize the VGG16 model (VGG16) with the following settings:
8       | include_top=False: Exclude the final classification layers.
```

ALGORITHM 5: DEFINING PRE-TRAINED MODEL (VGG16, RESNET101V2, INCEPTIONRESNETV2)

```
9      weights="imagenet": Initialize the model with pre-trained weights
      from ImageNet.
10     input_shape: Set the input shape to 256x128 pixels with 3 color
      channels (RGB).
11   end initialize
12   Freeze all layers in the VGG16 model by setting layer.trainable to False
      for each layer.
13   Flatten the output of the VGG16 model using Flatten() to create x_eye.
14 end define
15 Define the Palm Model:
16   Initialize the InceptionResNetV2 model (InceptionResNetV2) with the
      following settings:
17     include_top=False: Exclude the final classification layers.
18     weights="imagenet": Initialize the model with pre-trained weights
      from ImageNet.
19     input_shape: Set the input shape to 256x128 pixels with 3 color
      channels (RGB).
20   end initialize
21   Freeze all layers in the InceptionResNetV2 model by setting
      layer.trainable to False for each layer.
22   Flatten the output of the InceptionResNetV2 model using Flatten() to
      create x_palm.
23 end define
24 Define the Face Model:
25   Initialize the ResNet101V2 model (resnet) with the following settings:
26     include_top=False: Exclude the final classification layers.
27     weights="imagenet": Initialize the model with pre-trained weights
      from ImageNet.
28     input_shape: Set the input shape to 128x128 pixels with 3 color
      channels (RGB).
29   end initialize
30   Freeze all layers in the resnet model by setting layer.trainable to False
      for each layer.
31   Flatten the output of the resnet model using Flatten() to create x_face.
32 End
```

4.3.3 Modality Fusion & Classification stage

In this stage, Algorithm 6 illustrates the procedures for performing feature-level modality fusion through concatenation. It also covers the training process for the classification layer with a single hidden layer.

ALGORITHM 6: FEATURE LEVEL FUSION & MODEL TRAINING PROCESS

```
1  Start
2  Concatenate the outputs of the three pre-trained models (face, palm, and
   eye) to create a combined feature vector:
3      Use the concatenate function to concatenate  $x_{face}$ ,  $x_{palm}$ , and  $x_{eye}$ 
   along axis=1.
4  end concatenate
5  Define a Dense Layer:
6      Create a dense layer  $z$  with 300 units and a softmax activation function,
   which is applied to the combined feature vector (100 units for 100
   recognition classes problem).
7  end define
8  Create the Combined Model:
9      Build the combined model (model) with the following settings:
10         Inputs: Provide the input layers of the face (resnet.input), palm
   (InceptionResNetV2.input), and eye (VGG16.input) models.
11         Outputs: Set the output layer to  $z$ , which is the dense layer.
12     end build
13 end create
14 Configure early stopping:
15     Initialize the early_stopping callback with the following settings:
16         Monitor: "val_loss" to monitor the validation loss.
17         Min delta: 0.0001 as the minimum change to qualify as an
   improvement.
18         Patience: 20, the number of epochs with no improvement after
   which training will stop.
19         Verbose: 1 for displaying messages.
20         Mode: "auto" to automatically determine the direction of
   improvement.
21         Restore best weights: True to restore the model's weights to the best
   epoch.
22     end initialize
23 end configure
24 Configure the optimizer:
25     Initialize the Adam optimizer (opt) with a learning rate of 0.0001.
26 end configure
27 Compile the model:
28     Compile model with the following settings:
29         Optimizer: 'adam' or the previously defined opt.
30         Loss function: Sparse categorical cross-entropy with logits
   (tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)).
31         Metrics: Include accuracy as a metric.
32     end compile
33 end compile
34 Measure the start time (st) using a timer (e.g., time.time()).
35 Train the model:
36     Fit the model to the training data with the following parameters:
37         Input data: [face_train, palm_train, eye_train] for face, palm, and eye
   images.
38         Labels: label_train.
```

ALGORITHM 6: FEATURE LEVEL FUSION & MODEL TRAINING PROCESS

39 *Number of epochs: 100.*
40 *Batch size: 64.*
41 *Validation data: [x_valid_face, x_valid_palm, x_valid_eye] for*
 validation images and y_valid_eye for labels.
42 *Callbacks: Include the early_stopping callback for monitoring and early*
 stopping.
43 ***end train***
44 *Measure the end time (et) using a timer.*
45 *Calculate the elapsed time (elapsed_time) as the difference between the end*
 time and start time.
46 *Print the training time in seconds as "training time = elapsed_time seconds".*
47 ***End***

4.3.4 Model Evaluation

Algorithm 7 demonstrates the algorithm for model evaluation process. This part is essential as it produced results that could used to judge the performance of the proposed model.

ALGORITHM 7: MODEL EVALUATION ALGORITHM

1 ***Start***
2 ***Compute*** *confusion matrix:*
3 *| Store values for TP, TN, FP, and FN.*
4 ***end compute***
5 *Compute accuracy.*
6 *Compute precision.*
7 *Compute recall.*
8 *Compute ROC AUC Score*
9 *Compute specificity*
10 *Compute FPR*
11 *Compute FAR*
12 *Compute FRR*
13 ***End***

4.4 Chapter Summary

In this chapter, the machine used for program execution in this research is introduced. Section 4.3 provides a comprehensive overview of the proposed model. Additionally, the algorithms for each subsequent stage are presented. The next chapter will feature the research results and engage in in-depth discussions.

CHAPTER 5

RESEARCH RESULTS AND DISCUSSION

5.1 Introduction

In this chapter, the result for the research work will be clearly presented. Firstly, the result for image processing for different biometrics (iris, face, and palmprint) would be shown. Afterward, the dimensions for train, validate, and test datasets after performing data augmentation would be presented. Later on, the result for model classification would be analyzed and discussed. Lastly, the results for this research work would be compared with other researchers work.

5.2 Preparation Stage Results

In this part, the result for image processing of different biometrics modality (iris, face, palmprint) will be shown. In short, the dimensions for the final output in this stage were (256,128,3) for iris and palmprint, and (128,128,3) for face. Besides, although iris and palmprint images were originally black-scaled (single channeled) image, they were convert into multichannel image, so that they could be fed to the pre-trained model in the next model training stage. This was done by duplicating the single dimension image array with 2 more identical channels.

5.2.1 Iris Image Processing

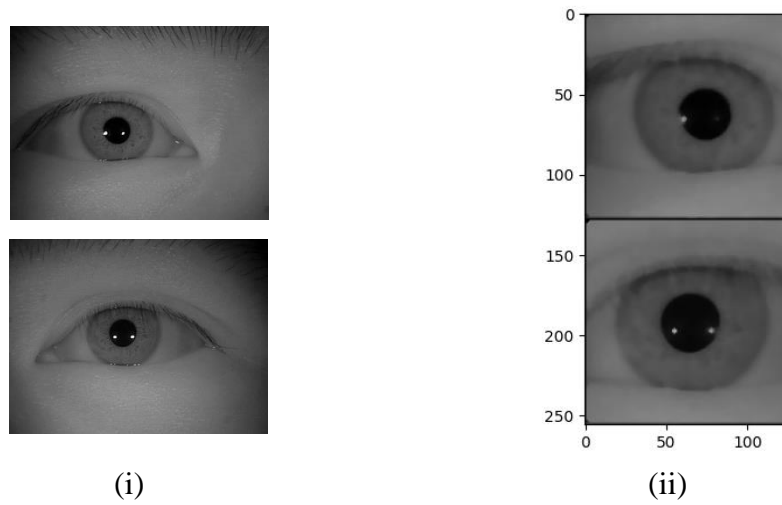


Figure 5.1 Iris Image Processing Result (i) Before (ii) After

In Figure 5.1(ii), it represents the processed iris image. From the result, the left and right iris images were concatenated by stacking them up in vertical direction. This was aimed to make the classifier to learn and differentiate differences between left and right iris within the same picture, which reduced the computational cost. Additionally, the iris area had been measured using Haar Cascade classifier. Then the ROI area was cropped. Originally, the left and right iris images had size of 640x480 pixels, then they were resized to (128,128,3), and lastly the dimension became (256,128,3) as they concatenated. In addition, the Median Blur filter was applied to denoise and enhance the input image, which could boost the model performance.

5.2.2 Face Image Processing

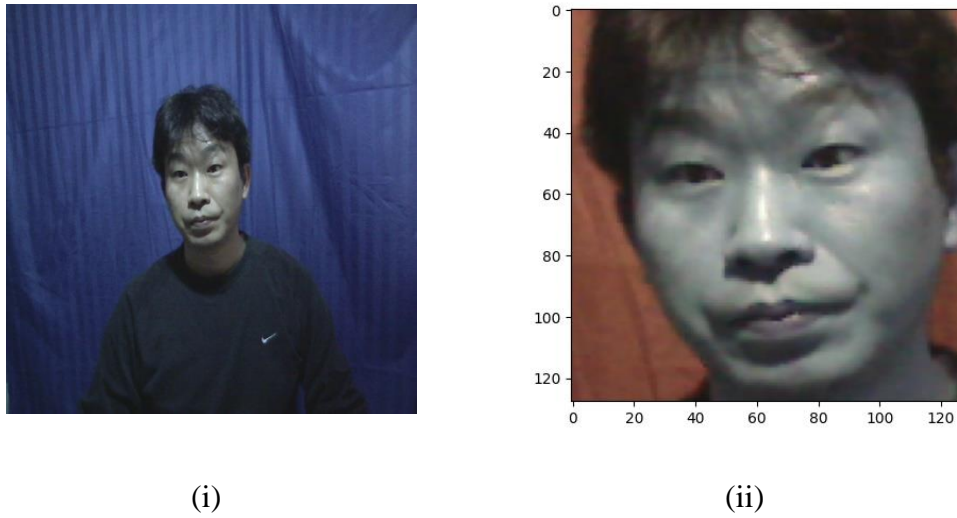


Figure 5.2 Face Image Processing Result (i) Before (ii) After

In Figure 5.2, it shows the result for face image processing. Initially, the picture was obtained in RGB format. Therefore, the face image would be read using RGB scale command. As the color could be useful as it enabled the model to learn how to extract features based on the colors. The face image was read and rescaled from 640x480 pixels to (128,128,3). Similarly, Haar Cascade algorithm was applied to crop the ROI. This helped to reduce the training computational cost as the unnecessary regions were removed. Lastly, Median Blur filter was applied to face image too to enhance the picture quality.

5.2.3 Palmprint Image Processing

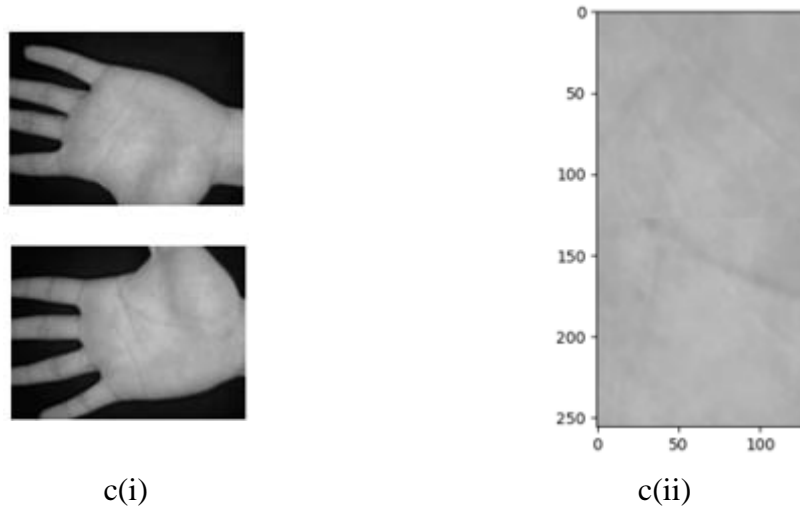


Figure 5.3 Palmprint Image Processing Result(i) Before (ii) After

In Figure 5.3, image processing result for palmprint is shown. First of all, left and right palm images were read. Then the centroid of the image was determined and a size of 128x128 pixels was cropped around the centroid. The cropped image would be the ROI for palmprint which was sufficiently to be used for model training. Next, Median Blur filter was applied on the image to improve the image quality. Finally, the cropped left and right palmprint images were concatenated by stacking them up vertically for better performance. The dimension of the final image was (256,128,3).

5.2.4 Data Augmentation

In this section, the dimension for dataset after performing data augmentation would be shown. Based on the results obtained from previous section, they were saved in different pickle files with respect to iris, face, and palmprint. In this research work, 2 classifiers (named model_100 and model_300) would be trained. Where model_100 was trained with of 100 classes and model_300 was trained with 300 classes. Generally, each classes had 5 pictures. In short, every biometric datasets had 500 original images, and 1500 original images for model_100 and model_300 respectively.

In order to create a robust and generalized classifier model, data augmentation was performed to increase the size of dataset. For model_100, each image was augmented to produce 20 new images. Hence the resultant training dataset was increased to 10,000 (500×20). Besides, validate dataset was created using different data augmentation function (defined with different parameter settings), where each image was augmented to produce 4 new images. Therefore producing validate dataset with size of 2000 (500×4). Lastly, the validate dataset was then split into halves, where half of it became the test dataset.

Next, for model_300, the number of augmented image had been decreased from 20 to 10 and 4 to 2 for train and validate dataset respectively. This is aimed to investigate the recognition strength of the proposed model. As decreasing number of training images per class and increasing number of class from 100 to 300 would definitely increased the difficulty of the recognition task for the model.

Lastly, the information regarding to the dataset dimension had been tabularized as shown in Table 5.1 and Table 5.2 as below.

Table 5.1 Summarized Results for Training, Validation, and Testing Datasets for Model_100

Model_100 (100 classes)	Iris	Face	Palmprint
Original dataset	(500,256,128,3)	(500,128,128,3)	(500,256,128,3)
Number of augment image per class	20		
Training dataset	(10000,256,128,3)	(10000,128,128,3)	(10000,256,128,3)
Train-Test Split Ratio	10:1:1		
Validation dataset	(1000,256,128,3)	(1000,128,128,3)	(1000,256,128,3)
Test dataset	(1000,256,128,3)	(1000,128,128,3)	(1000,256,128,3)

Table 5.2 Summarized Results for Training, Validation, and Testing Datasets for Model_300

Model_300 (300 classes)	Iris	Face	Palmprint
Original dataset	(1500,256,128,3)	(1500,128,128,3)	(1500,256,128,3)
Number of augment image per class	10		
Training dataset	(15000,256,128,3)	(15000,128,128,3)	(15000,256,128,3)
Train-Test Split Ratio	10:1:1		
Validation dataset	(1500,256,128,3)	(1500,128,128,3)	(1500,256,128,3)
Test dataset	(1500,256,128,3)	(1500,128,128,3)	(1500,256,128,3)

5.3 Model Evaluation

In this section, the results for model 100 and model 300 is presented. Generally model 100 is trained to recognize 100 classes and model 300 is trained to recognize 300 classes. Therefore, there is 100 and 300 units in the classification layer for both model, respectively with the softmax activation function set. Furthermore, the batch size and epoch is set to 64 and 100 respectively and Callbacks is applied in the model fitting process (patience = 20, restore_best_weights = True). As the training process found the minimum validation loss, it would stop training if there is no other minimum detected after 20 more fitting process. Then, the model would stop training and restore to its best parameters determined.

In Figure 5.4, it shows the proposed model_100 stopped training at epoch 61, and the model_100 parameters restored to that in epoch 41 as shown in Figure 5.5. Additionally, Table 5.3 has presented as below to provide more detailed information.

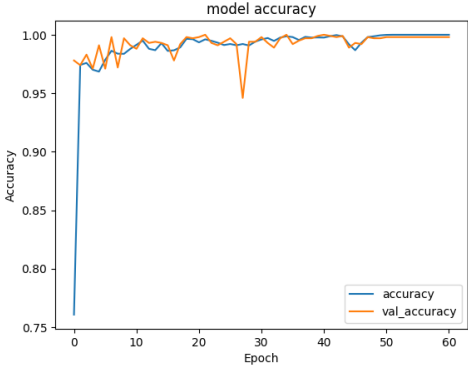
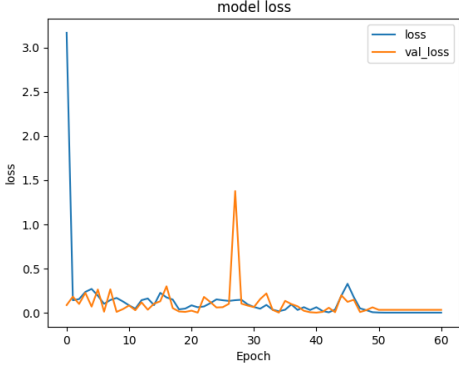
```
Epoch 61/100
156/157 [=====>.] - ETA: 0s - loss: 5.1342e-10 - accuracy: 1.0000Restoring model weights from the end of the best epoch: 41.
157/157 [=====] - 15s 96ms/step - loss: 5.1260e-10 - accuracy: 1.0000 - val_loss: 0.0315 - val_accuracy: 0.9980
Epoch 61: early stopping
training time = 924.6336717605591 seconds
```

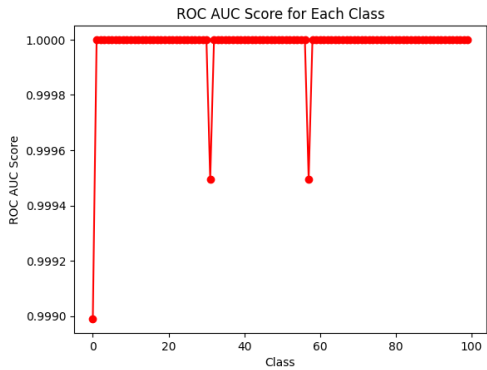
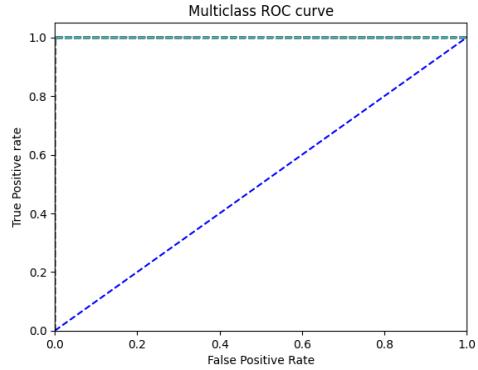
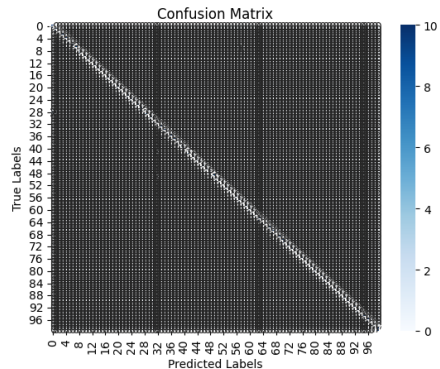
Figure 5.4 Model_100 Fitting Result Using Callbacks

```
Epoch 41/100
157/157 [=====] - 15s 94ms/step - loss: 0.0609 - accuracy: 0.9976 - val_loss: 1.0681e-05 - val_accuracy: 1.0000
```

Figure 5.5 Model_100 Finalized Result

Table 5.3 Results for Model_100

Evaluation Metrics	Results
Training & Validate Accuracy	99.76% & 100% 
Training time	924.63 seconds
Training & Validate Loss	0.0609 & 0 
ROC AUC Score	1

Evaluation Metrics	Results
	 
Confusion Matrix	

In Table 5.3, it shows the results for `model_100`. From the accuracy and loss curves shown, the model trained performed well on the training dataset as well as for validation dataset. One inference could be made toward the model was there were no overfitting or underfitting issues happened.

Additionally, the ROC_AUC_SCORE for each class and multiclass ROC curve had been plotted using “One VS All” approach. According to the figures, model_100 performed really well in classifying all of the classes. The average ROC_AUC_SCORE had been calculated as 1. Lastly, the average sensitivity, specificity, TPR, FPR, FAR, FRR scores were 0.995, 1, 99.6 %, 0.004 %, 0, 0.005 respectively. These results were then tabulated as shown in Table 5.3 and Table 5.4 for better model analyzation.

Besides, the confusion matrix has been displayed in Table 5.3 as well. The black region in the image were filled with too many zeros. The white region was filled by non-zero integer values. Therefore, this indicates that the model performed very well. Next, the precision and recall scores also been measured, which were 0.996, and 0.995 respectively.

Next, the results for Model_300 would be shown as below, likewise, the training process stopped at epoch 28, and the parameters restored to that In epoch 8 as shown in Figure 5.6 and Figure 5.7. Furthermore, Table 5.4 is presented for more detailed information.

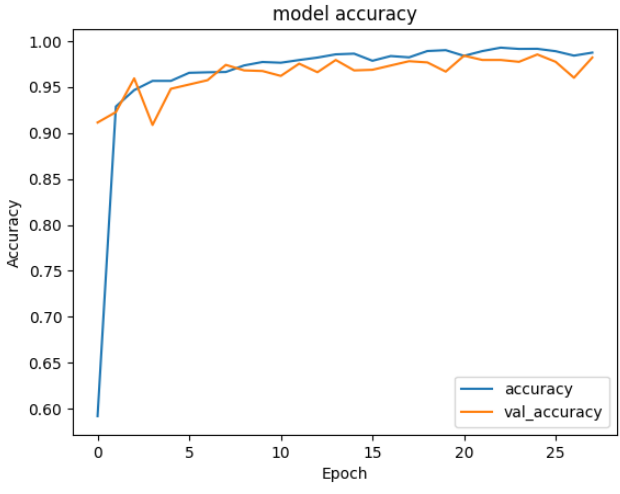
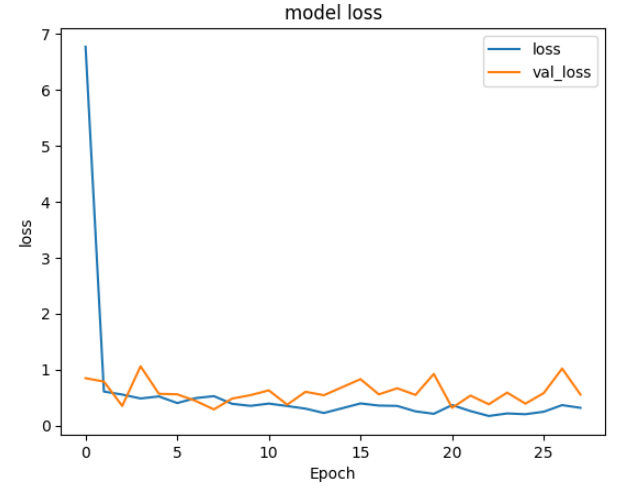
```
Epoch 28/100
234/235 [=====>.] - ETA: 0s - loss: 0.3176 - accuracy: 0.9873Restoring model weights from the end of the best epoch: 8.
235/235 [=====] - 22s 94ms/step - loss: 0.3171 - accuracy: 0.9873 - val_loss: 0.5552 - val_accuracy: 0.9820
Epoch 28: early stopping
training time = 652.0651087760925 seconds
```

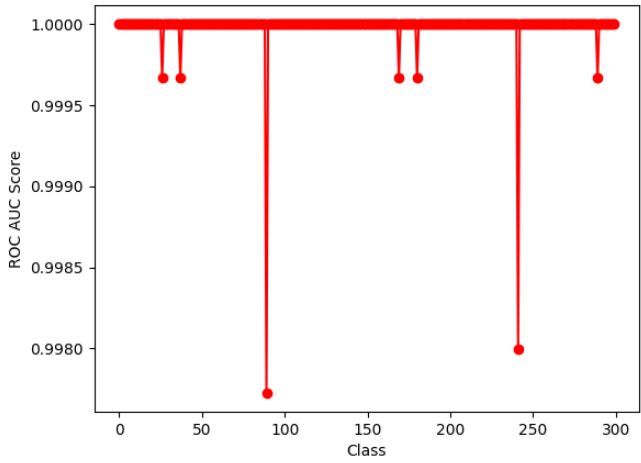
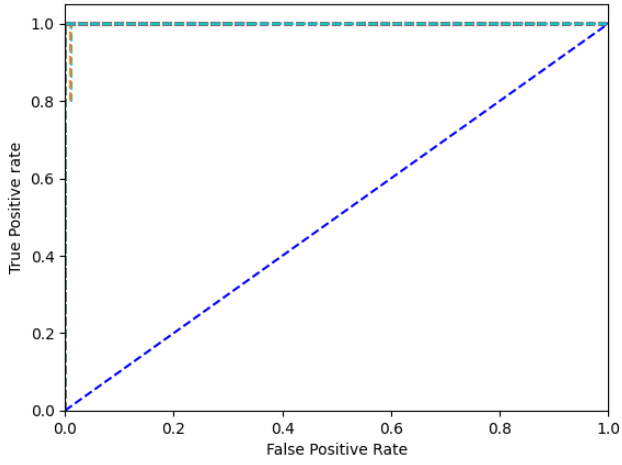
Figure 5.6 Model_300 Fitting Result Using Callbacks

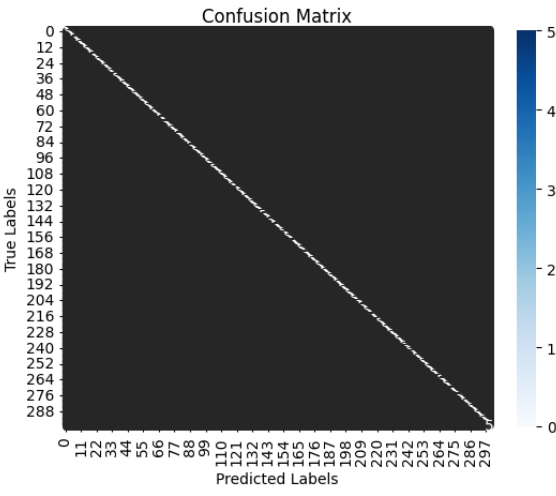
```
Epoch 8/100
235/235 [=====] - 22s 94ms/step - loss: 0.5259 - accuracy: 0.9663 - val_loss: 0.2893 - val_accuracy: 0.9740
```

Figure 5.7 Model_300 Finalized Result

Table 5.4 Results for Model_300

Evaluation Metrics	Results
Training & Validate Accuracy	<div>96.63% & 97.40%</div> <div></div>
Training time	652.07 seconds
Training & Validate Loss	<div>0.5259 & 0.2893</div> <div></div>

Evaluation Metrics	Results																						
ROC AUC Score	<div>1</div> <div><p>ROC AUC Score for Each Class</p><table border="1"><caption>Approximate data for ROC AUC Score for Each Class</caption><thead><tr><th>Class</th><th>ROC AUC Score</th></tr></thead><tbody><tr><td>0</td><td>1.0000</td></tr><tr><td>25</td><td>1.0000</td></tr><tr><td>30</td><td>0.9997</td></tr><tr><td>35</td><td>0.9997</td></tr><tr><td>90</td><td>0.9980</td></tr><tr><td>165</td><td>0.9997</td></tr><tr><td>170</td><td>0.9997</td></tr><tr><td>240</td><td>0.9980</td></tr><tr><td>290</td><td>0.9997</td></tr><tr><td>300</td><td>1.0000</td></tr></tbody></table></div> <div><p>Multiclass ROC curve</p></div>	Class	ROC AUC Score	0	1.0000	25	1.0000	30	0.9997	35	0.9997	90	0.9980	165	0.9997	170	0.9997	240	0.9980	290	0.9997	300	1.0000
Class	ROC AUC Score																						
0	1.0000																						
25	1.0000																						
30	0.9997																						
35	0.9997																						
90	0.9980																						
165	0.9997																						
170	0.9997																						
240	0.9980																						
290	0.9997																						
300	1.0000																						

Evaluation Metrics	Results
Confusion Matrix	

In Table 5.4, it shows the results for model 300. From these accuracy and loss curves observed, it could be determined that model_300 performed quite well too. As the validate and train accuracy curve did not deviate much. This indicated that the model was not having overfitting or underfitting issues.

Furthermore, the ROC_AUC_SCORE for each class and the multiclass ROC curve for model_300 is presented. In short, the average ROC value was measured as 1, which further validated that this model performed really well.

Moreover, the confusion matrix is presented as well. Similarly, the black regions were filled by many zeros, and the white region were filled by non-zeros number. Hence, it could be further ensured that the model was performing well as there were many correctly predicted classes based on the confusion matrix.

In the next section, both models (model_100 and model_300) performance would be compared with other researcher works.

Table 5.5 Comparison Results between Model_100 and Model_300

Evaluation Metrics	Model_100	Model_300
Test Accuracy	99.60 %	97.60 %
Average Precision	0.997	0.980
Average Recall	0.996	0.976
ROC_AUC_SCORE	1.000	1.000
Average Specificity	1.000	1.000
FPR	0	0
FAR	0	0
FRR	0.005	0.024

In the final phase of this study, both models are verified with their own test datasets. The outcomes of these tests are summarized in Table 5.5, providing valuable insights into the recognition capabilities of the proposed model within the context of online banking security.

Notably, the primary distinction between model_100 and model_300 lies in the size of the training dataset per class, which decreased from 20 to 10 images, while the number of classes for model training increased from 100 to 300. These adjustments were made to simulate real-world scenarios where a broader range of classes is encountered, ultimately elevating the recognition complexity for model_300. Besides, the size decrement of training dataset per class is performed to demonstrate the recognition power of the proposed model. Upon analyzing the test accuracy presented in Table 5.5, it is evident that model_300 achieved an accuracy rate of 97.60%, reflecting a 2% reduction compared to model_100, which achieved an accuracy rate of 99.60%. This slight reduction in accuracy could be attributed to the increased recognition complexity introduced by the larger number of classes and lesser number of training images per class.

It is important to highlight that both models exhibit robust performance across various metrics. This observation underscores the model's resilience in handling both False Positives (FP) and False Negatives (FN) predictions. Of particular significance are the remarkably low values achieved in terms of False Acceptance Rate (FAR), False Positive Rate (FPR), and False Rejection Rate (FRR) for both models. The low FAR signifies a reduced risk of unauthorized access, as the system effectively rejects

impostor attempts. Simultaneously, the low FPR ensures a minimal rate of legitimate access denials, thus enhancing the user experience and convenience. Moreover, the low FRR underscores the model's efficacy in minimizing the chances of false rejections when legitimate users seek access. In the context of online banking security, this level of robustness is especially critical as it reflects the model's capacity to reliably distinguish between legitimate and unauthorized access attempts. Ultimately, this bolsters the overall security posture of online banking systems

5.4 Models Comparison

Table 5.6 Comparison Between Different Models Applied related to Banking Security Domain with Proposed Model

No.	Models		Accuracy (%)
1.	(Kataria & Nafis, 2019)		84
2.	(Zinjurde & Kamble, 2020)		87
3	(Paul, 2022)	LR	74.61
		XGBoost	73.5
4.	(Khosravi et al., 2023)	XGBoost	96
		Random Forest	96
6.	Proposed Model	Model_100	99.6
		Model_300	97.6

In Table 5.6, comparisons are made between different models with the proposed model. All of them are related to online bank security domain, and their findings are important to compare with the proposed model in this research work. As the ultimate goal of this study is to develop a robust multimodal authentication model that can be applied for online banking security systems. In summary, the accuracy achieved by the proposed model is the highest as compared to their model. Indicating that the pre-trained architecture has a high potential to be applied in image recognition task system especially online banking security domain.

5.5 Chapter Summary

Throughout this chapter, the results for this research work had been clearly shown. In the first section, the image processing results for iris, face and palmprint were shown. Next, the dimension for the train, validate, and test dataset used after performing data augmentation was clearly stated. In addition, all the performance metrics discussed earlier were measured for evaluating the models (Model_100 and Model_300). Lastly, the result for proposed methods were compared with other researchers work related to banking security domain. In the next chapter, it would be the end for this research work, where the conclusion would be drawn and some future research directions or suggestion would be discussed.

CHAPTER 6

CONCLUSION

6.1 Introduction

In this chapter, conclusion about this research work would be made. Firstly, the achievement of research objectives had been listed, and then research contributions were clear stated. Finally, this research work ended with some suggestions and recommendations for future studies.

6.2 Achievement of Research Objectives

Throughout this research work, the objectives set out were successfully achieved and were summarized in the following paragraph. Besides, the research purpose throughout this research work can be concluded as a success.

To begin with, a comprehensive understanding of machine learning technique and biometric modalities applied in recognition system related to online banking security has been reviewed. In addition, more topics such as image pre-processing, feature extraction, and feature fusion techniques in the realm of multimodal biometrics was obtained through an extensive literature study. This exploration involved reviewing relevant research articles and identifying the primary challenges faced in developing robust multimodal biometric recognition models, particularly the limited availability and high complexity of the datasets.

Moreover, since training a reliable recognition model demands substantial resources, such as time and RAM, the concept of transfer learning emerged as a valuable solution. By leveraging pre-trained models, namely VGG16, ResNet101V2,

and InceptionResNetV2, the training process was significantly improved and facilitated. This enable the biometric images conduct simpler image pre-processing steps and then performing feature extraction using the selected pre-trained models.

Furthermore, feature fusion was seamlessly achieved through concatenation, followed by training a fully connected (fc) layer with 100 or 300 units with softmax as activation function for classification purposes. Additionally, the entire training process for the proposed models were implemented using the Python programming language. Finally, the proposed method's performance was rigorously evaluated and compared with other models applied in banking security system in terms of the model's recognition accuracy. The results demonstrated the exceptional performance of the proposed method, reinforcing its potential for integration into online banking systems to enhance user recognition processes.

In conclusion, the successful implementation and evaluation of the proposed method underscore its viability and effectiveness in the realm of online banking security. By harnessing the power of multimodal biometrics and incorporating transfer learning techniques, this approach presents promising prospects for elevating the security and reliability of user authentication in online banking systems.

6.3 Summary of Contribution

Throughout this research work, a novel approach utilizing multiple pre-trained models (VGG16, ResNet101V2, InceptionResNetV2) for feature extraction has been proposed. These pre-trained models consist of deep CNN layers that enable the extraction of features in a more profound manner. Leveraging the pre-trained parameters facilitates better learning within the recognition model. Consequently, the need for training a recognition model from scratch is significantly reduced, leading to enhanced performance.

Additionally, two multimodal recognition models, namely model_100 and model_300, were trained using softmax classification layer. Model_100 was trained to

classify 100 classes, while model_300 was trained for 300 classes. Despite the dataset being expanded to accommodate 300 classes and a decrease in the number of training images per class from 20 to 10, model_300 still showed impressive performance as compared to model_100. In short, both models demonstrated superior performance with high accuracy, low false rejection rate (FRR), and low false acceptance rate (FAR). The low FRR and FAR achieved by these models are crucial factors in ensuring the integrity and reliability of online banking security systems.

Importantly, these proposed models did not undergo any fine-tuning steps. This indicates the potential for further performance improvement if fine-tuning were implemented. These findings highlight the high potential of the proposed method for further enhancement and application in real-life recognition systems, such as online banking security. The low FRR and FAR achieved by these models are crucial factors in ensuring the integrity and reliability of online banking security systems.

6.4 Suggestions for Improvement for Future Works

After doing this research work, there were some thoughts and recommendations provided. Firstly, the proposed model can be test using different dataset in the future. Secondly, try to execute feature extraction using different pre-trained models. Thirdly, further optimize the model performance by performing fine tuning. Fourthly, try different modalities fusion strategy (such as score fusion), as it was very popular to apply in various multimodal domains. Lastly, at the classification layer, it can be trained using machine learning approaches.

REFERENCES

- 100 Most Common Passwords Of 2022. Can You Spot Your Password?* (2022).
<https://techcult.com/most-common-passwords/>
- Akhtar, Z., Hadid, A., Nixon, M. S., Tistarelli, M., Dugelay, J. L., & Marcel, S. (2018). Biometrics: In search of identity and security (Q A). *IEEE Multimedia*, 25(3), 22–35. <https://doi.org/10.1109/MMUL.2018.2873494>
- Alay, N., & Al-Baity, H. H. (2020). Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors (Switzerland)*, 20(19), 1–17. <https://doi.org/10.3390/s20195523>
- Albalawi, S., Alshahrani, L., Albalawi, N., Kilabi, R., & Alhakamy, A. (2022). A Comprehensive Overview on Biometric Authentication Systems using Artificial Intelligence Techniques. *International Journal of Advanced Computer Science and Applications*, 13(4), 782–791. <https://doi.org/10.14569/IJACSA.2022.0130491>
- Aloul, F., & Zahidi, S. (2009). *Two factor authentication using mobile phones*. 641–644.
- Alsellami, B., Deshmukh, P. D., Ahmed, Z. A. T., Tawfik, M., & Al-Madani, A. M. (2021). Overview of Biometric Traits. *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, 807–813. <https://doi.org/10.1109/ICIRCA51532.2021.9545069>
- Ammour, B., Boubchir, L., Bouden, T., & Ramdani, M. (2020). Face–iris multimodal biometric identification system. *Electronics (Switzerland)*, 9(1). <https://doi.org/10.3390/electronics9010085>
- Benradi, H., Chater, A., & Lasfar, A. (2023). A hybrid approach for face recognition using a convolutional neural network combined with feature extraction techniques. *IAES International Journal of Artificial Intelligence*, 12(2), 627–640. <https://doi.org/10.11591/ijai.v12.i2.pp627-640>
- BIT*. (2005). <http://biometrics.idealtest.org/dbDetailForUser.do?id=5#/>
- Buckley, O., & Nurse, J. R. C. (2019). The language of biometrics: Analysing public perceptions. *Journal of Information Security and Applications*, 47, 112–119. <https://doi.org/10.1016/j.jisa.2019.05.001>

- Buckner, C. A., Lafrenie, R. M., Dénomée, J. A., Caswell, J. M., Want, D. A., Gan, G. G., Leong, Y. C., Bee, P. C., Chin, E., Teh, A. K. H., Picco, S., Villegas, L., Tonelli, F., Merlo, M., Rigau, J., Diaz, D., Masuelli, M., Korrapati, S., Kurra, P., ... Mathijssen, R. H. J. (2016). A Survey of Machine Learning Techniques for Behavioral-Based Biometric User Authentication. *Intech, 11*(tourism), 13. <https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics>
- Chan, J. Y. Le, Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2023). State of the art: a review of sentiment analysis based on sequential transfer learning. In *Artificial Intelligence Review* (Vol. 56, Issue 1). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10183-8>
- Chawla, B., Tyagi, S., Jain, R., Talegaonkar, A., & Srivastava, S. (2021). Finger Vein Recognition Using Deep Learning. *Advances in Intelligent Systems and Computing, 1164*(1), 69–78. https://doi.org/10.1007/978-981-15-4992-2_7
- Choudhary, S. K., & Naik, A. K. (2019). Multimodal biometric authentication with secured templates – A review. *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019, 2019-April(Icoei)*, 1062–1069. <https://doi.org/10.1109/icoei.2019.8862563>
- Convolutional Neural Network: An Overview.* (2022). <https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/>
- Daas, S., Yahi, A., Bakir, T., Sedhane, M., Boughazi, M., & Bourennane, E. B. (2020). Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion. *IET Image Processing, 14*(15), 3859–3868. <https://doi.org/10.1049/iet-ipr.2020.0491>
- Dhoot, A., Nazarov, A. N., & Koupaei, A. N. A. (2020). A Security Risk Model for Online Banking System. *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, 2020–2023. <https://doi.org/10.1109/IEEECONF48371.2020.9078655>
- El_Rahman, S. A. (2020). Multimodal biometric systems based on different fusion levels of ECG and fingerprint using different classifiers. In *Soft Computing* (Vol. 24, Issue 16). Springer Berlin Heidelberg. <https://doi.org/10.1007/s00500-020-04700-6>
- Gavisiddappa, G., Mahadevappa, S., & Patil, C. M. (2020). Multimodal biometric

- authentication system using modified relief feature selection and multi support vector machine. *International Journal of Intelligent Engineering and Systems*, 13(1), 1–12. <https://doi.org/10.22266/ijies2020.0229.01>
- Ghayoumi, M. (2015). A review of multimodal biometric systems: Fusion methods and their applications. *2015 IEEE/ACIS 14th International Conference on Computer and Information Science, ICIS 2015 - Proceedings*, 131–136. <https://doi.org/10.1109/ICIS.2015.7166582>
- Gona, A. K., & Subramoniam, M. (2022a). Multimodal Biometric Reorganization System using Deep Learning Convolutional Neural Network. *International Conference on Edge Computing and Applications, ICECAA 2022 - Proceedings, Icecaa*, 1282–1286. <https://doi.org/10.1109/ICECAA55415.2022.9936398>
- Gona, A. K., & Subramoniam, M. (2022b). *Multimodal Biometric Reorganization System using Deep Learning Convolutional Neural Network. Icecaa*, 1282–1286. <https://doi.org/10.1109/icecaa55415.2022.9936398>
- Gona, A., & Subramoniam, M. (2022a). Convolutional neural network with improved feature ranking for robust multi-modal biometric system. *Computers and Electrical Engineering*, 101(May), 108096. <https://doi.org/10.1016/j.compeleceng.2022.108096>
- Gona, A., & Subramoniam, M. (2022b). Convolutional neural network with improved feature ranking for robust multi-modal biometric system. *Computers and Electrical Engineering*, 101(November 2021), 108096. <https://doi.org/10.1016/j.compeleceng.2022.108096>
- Gunasekaran, K., Raja, J., & Pitchai, R. (2019). Deep multimodal biometric recognition using contourlet derivative weighted rank fusion with human face, fingerprint and iris images. *Automatika*, 60(3), 253–265. <https://doi.org/10.1080/00051144.2019.1565681>
- Gunawan, K. W., Hidayat, A. A., Cenggoro, T. W., & Pardamean, B. (2023). Repurposing transfer learning strategy of computer vision for owl sound classification. *Procedia Computer Science*, 216(2022), 424–430. <https://doi.org/10.1016/j.procs.2022.12.154>
- Haider, S. A., Rehman, Y., & Usman Ali, S. M. (2020). Enhanced multimodal biometric recognition based upon intrinsic hand biometrics. *Electronics (Switzerland)*, 9(11), 1–20. <https://doi.org/10.3390/electronics9111916>
- Hammad, M., Liu, Y., & Wang, K. (2019). Multimodal biometric authentication

- systems using convolution neural network based on different level fusion of ECG and fingerprint. *IEEE Access*, 7, 25527–25542. <https://doi.org/10.1109/ACCESS.2018.2886573>
- Israa, A. (2015). Physiological Biometric Authentication Systems, Advantages, Disadvantages And Future Development: A Review. *International Journal of Scientific & Technology Research.*, 1(1), 7.
- Jiawei, G., Zhou, F., & Qiping, T. (2020). Bank card number recognition system based on deep learning. *ACM International Conference Proceeding Series, Itnec*, 745–749. <https://doi.org/10.1145/3443467.3443847>
- Kataria, S., & Nafis, M. T. (2019). Internet banking fraud detection using deep learning based on decision tree and multilayer perceptron. *Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom 2019*, 1298–1302.
- Keskar, M. Y., & Pandey, N. (2018). Internet banking: a review (2002–2016). *Journal of Internet Commerce*, 17(3), 310–323. <https://doi.org/10.1080/15332861.2018.1451969>
- Khanaa, V., Thooyamani, K. P., & Udayakumar, R. (2014). Two factor authentication using mobile phones. *World Applied Sciences Journal*, 29(14), 208–213. <https://doi.org/10.5829/idosi.wasj.2014.29.csea.2268>
- Khosravi, S., Kargari, M., Teimourpour, B., Eshghi, A., & Aliabdi, A. (2023). Using Supervised Machine Learning Approaches to Detect Fraud in the Banking Transaction Network. *2023 9th International Conference on Web Research, ICWR 2023*, 115–119. <https://doi.org/10.1109/ICWR57742.2023.10139083>
- Li, B., & Lima, D. (2021). Facial expression recognition via ResNet-50. *International Journal of Cognitive Computing in Engineering*, 2(February), 57–64. <https://doi.org/10.1016/j.ijcce.2021.02.002>
- Ma, Y., Huang, Z., Wang, X., & Huang, K. (2020). An Overview of Multimodal Biometrics Using the Face and Ear. *Mathematical Problems in Engineering*, 2020. <https://doi.org/10.1155/2020/6802905>
- Malinka, K., Hujňák, O., Hanáček, P., & Hellebrandt, L. (2022). *E-Banking Security Study — 10 Years Later*.
- Mallet, J., Pryor, L., Dave, R., Seliya, N., Vanamala, M., & Sowell-Boone, E. (2022). Hold On and Swipe: A Touch-Movement Based Continuous Authentication Schema based on Machine Learning. *Proceedings - 2022 Asia Conference on*

- Algorithms, Computing and Machine Learning, CACML 2022*, 442–447.
<https://doi.org/10.1109/CACML55074.2022.00081>
- Mayer, P., & Volkamer, M. (2018). Addressing misconceptions about password security effectively. *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/3167996.3167998>
- Medjahed, C., Rahmoun, A., Charrier, C., & Mezzoudj, F. (2022). A deep learning-based multimodal biometric system using score fusion. *IAES International Journal of Artificial Intelligence*, 11(1), 65–80.
<https://doi.org/10.11591/ijai.v11.i1.pp65-80>
- Mustafa, A. S., Abdulelah, A. J., Ahmed, A. K., & Shamil Mustafa, A. (2020). Multimodal Biometric System Iris and Fingerprint Recognition Based on Fusion Technique. *International Journal of Advanced Science and Technology*, 29(03), 7423–7432. <https://www.researchgate.net/publication/340511996>
- Oguntimilehin, A., Akukwe, M. L., Olatunji, K. A., Abiola, O. B., Adeyemo, O. A., & Abiodun, I. A. (2022). Mobile Banking Transaction Authentication using Deep Learning. *Proceedings of the 5th International Conference on Information Technology for Education and Development: Changing the Narratives Through Building a Secure Society with Disruptive Technologies, ITED 2022*, 1–7.
<https://doi.org/10.1109/ITED56637.2022.10051553>
- Pallivalappil, A. S., Sinha, D., Kumari, P. L., Katti, A., Sakthi, G., & Narkhede, M. R. (2022). Evaluation of Digital Wallet Transaction Accuracy using Machine Learning. *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, 239–244.
<https://doi.org/10.1109/ICACITE53722.2022.9823436>
- Paul, T. (2022). Internet of Things and Machine Learning Banking Business Model Using Neuro Fuzzy Technique. *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability, ICITISEE 2022*, 189–194.
<https://doi.org/10.1109/ICITISEE57756.2022.10057612>
- Prabu, S., Lakshmanan, M., & Mohammed, V. N. (2019). A Multimodal Authentication for Biometric Recognition System using Intelligent Hybrid Fusion Techniques. *Journal of Medical Systems*, 43(8).
<https://doi.org/10.1007/s10916-019-1391-5>

- Purohit, H., & Ajmera, P. K. (2021). Optimal feature level fusion for secured human authentication in multimodal biometric system. *Machine Vision and Applications*, 32(1), 1–12. <https://doi.org/10.1007/s00138-020-01146-6>
- Rajasekar, V., Predić, B., Saracevic, M., Elhoseny, M., Karabasevic, D., Stanujkic, D., & Jayapaul, P. (2022). Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm. *Scientific Reports*, 12(1), 1–11. <https://doi.org/10.1038/s41598-021-04652-3>
- Regouid, M., Touahria, M., Benouis, M., & Costen, N. (2019). Multimodal biometric system for ECG, ear and iris recognition based on local descriptors. *Multimedia Tools and Applications*, 78(16), 22509–22535. <https://doi.org/10.1007/s11042-019-7467-x>
- Safavipour, M. H., Doostari, M. A., & Sadjedi, H. (2022). A hybrid approach to multimodal biometric recognition based on feature-level fusion of face, two irises, and both thumbprints. *Journal of Medical Signals and Sensors*, 12(3), 177–191. https://doi.org/10.4103/jmss.jmss_103_21
- Selvakumar, R., Logesh, S., Maha Vishnu, S., Maniraj, S., & Praveen Kumar, A. (2022). Face Biometric Authentication System for ATM using Deep Learning. *3rd International Conference on Electronics and Sustainable Communication Systems, ICESC 2022 - Proceedings, Icesc*, 647–655. <https://doi.org/10.1109/ICESC54411.2022.9885334>
- Sengar, S. S., Hariharan, U., & Rajkumar, K. (2020). Multimodal Biometric Authentication System using Deep Learning Method. *2020 International Conference on Emerging Smart Computing and Informatics, ESCI 2020*, 309–312. <https://doi.org/10.1109/ESCI48226.2020.9167512>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.
- Singh, A., Pandey, A., Rakhra, M., Singh, D., Singh, G., & Dahiya, O. (2022). An Iris Recognition System Using CNN & VGG16 Technique. *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022*, 1–6. <https://doi.org/10.1109/ICRITO56286.2022.9965172>
- Singh, M., Singh, R., & Ross, A. (2019). A comprehensive overview of biometric fusion. *Information Fusion*, 52(December 2018), 187–205.

- <https://doi.org/10.1016/j.inffus.2018.12.003>
- Stewart, D. (2009). Mobile banking. *ABA Bank Marketing*, 41(5), 56–60.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 4278–4284. <https://doi.org/10.1609/aaai.v31i1.11231>
- Tse, K. W., & Hung, K. (2020). User Behavioral Biometrics Identification on Mobile Platform using Multimodal Fusion of Keystroke and Swipe Dynamics and Recurrent Neural Network. *ISCAIE 2020 - IEEE 10th Symposium on Computer Applications and Industrial Electronics*, 262–267. <https://doi.org/10.1109/ISCAIE47305.2020.9108839>
- Understanding AUC - ROC Curve | by Sarang Narkhede | Towards Data Science.* (n.d.). Retrieved 23 December 2022, from <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
- Vegas, L. (2005). *IABE-2005 Annual Conference*. I(1).
- Wang, Y., Shi, D., & Zhou, W. (2022). Convolutional Neural Network Approach Based on Multimodal Biometric System with Fusion of Face and Finger Vein Features. *Sensors*, 22(16), 1–15. <https://doi.org/10.3390/s22166039>
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56–70. <https://doi.org/10.38094/jastt1224>
- Zhang, L., Cheng, Z., Shen, Y., & Wang, D. (2018). Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset. *Symmetry*, 10(4), 1–15. <https://doi.org/10.3390/sym10040078>
- Zhou, C., Huang, J., Yang, F., & Liu, Y. (2020). A hybrid fusion model of iris, palm vein and finger vein for multi-biometric recognition system. *Multimedia Tools and Applications*, 79(39–40), 29021–29042. <https://doi.org/10.1007/s11042-020-08914-6>
- Zinjurde, A. M., & Kamble, V. B. (2020). Credit Card Fraud Detection and Prevention by Face Recognition. *Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, ICSIDEMPC 2020*, 86–90. <https://doi.org/10.1109/ICSIDEMPC49020.2020.9299587>

Appendix A Link for Code

GitHub link for execution code:

https://github.com/juxue97/Research_Code.git

Appendix B Gantt Chart

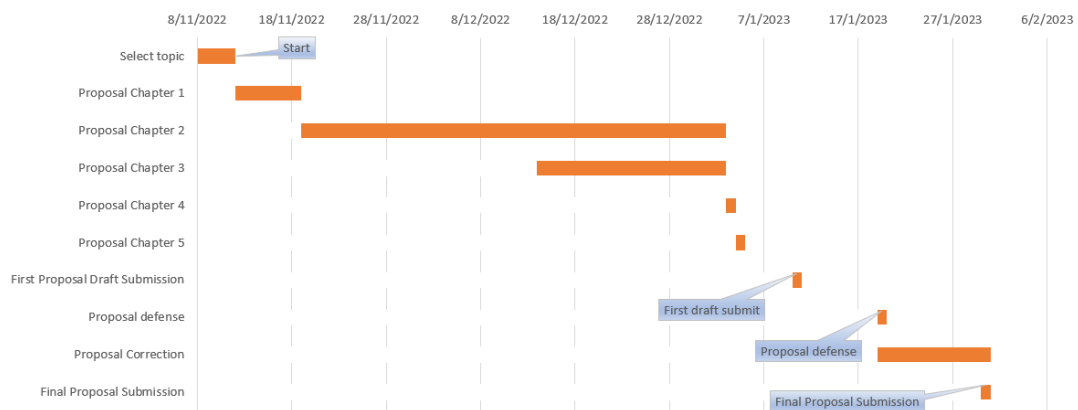


Figure 6.1 Gantt Chart for Research Proposal

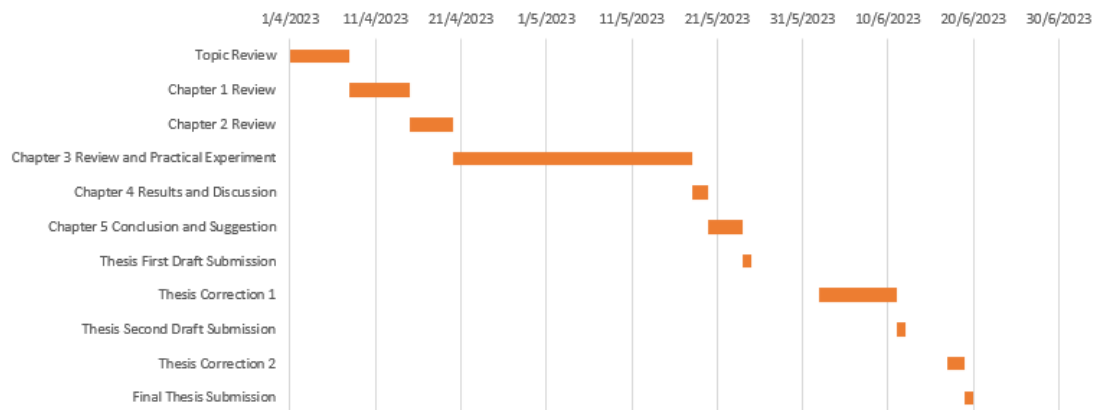


Figure 6.2 Gantt Chart for Research Thesis