



UNIVERSITY  
OF  
JOHANNESBURG

## COPYRIGHT AND CITATION CONSIDERATIONS FOR THIS THESIS/ DISSERTATION



- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- NonCommercial — You may not use the material for commercial purposes.
- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

### How to cite this thesis

Surname, Initial(s). (2012). Title of the thesis or dissertation (Doctoral Thesis / Master's Dissertation). Johannesburg: University of Johannesburg. Available from: <http://hdl.handle.net/102000/0002> (Accessed: 22 August 2017).

---

# Lip Print Based Authentication In Physical Access Control Environments

by

**WARDAH FARRUKH**

**DISSERTATION**

submitted in fulfilment of the requirements for the degree

**MASTER OF SCIENCE**

in the

**FACULTY OF SCIENCE**

at the

**UNIVERSITY OF JOHANNESBURG**

JOHANNESBURG

Supervisor: Professor D.T. van der Haar

November 2021

---

# Acknowledgements

Firstly, I want to express my gratitude to my parents, particularly my mother, who has always supported me with love, patience, and understanding. Without you, I could never have reached this milestone. It is your unconditional love that motivates me to set higher goals.

I would like to express my sincere gratitude to my supervisor, Professor Dustin van der Haar, for his dedicated support and guidance throughout this research project. The example Prof has set as a mentor and leader is one, I aspire to be if I am ever given the opportunity. Prof, I thank you for guiding me throughout this journey, for encouraging me try out my strengths, for giving me the opportunity to fail at times and for your constant support. I hope to learn a lot more from you in the future.

Lastly, I would like to thank the Academy of Computer Science and Software Engineering for giving me the opportunity to work as a teaching assistant while completing my research. It has been a true privilege to learn and work with you.



# Abstract

In modern society, there is an ever-growing need to determine the identity of a person in many applications including computer security, financial transactions, borders, and forensics. Early automated methods of authentication relied mostly on possessions and knowledge. Notably these authentication methods such as passwords and access cards are based on properties that can be lost, stolen, forgotten, or disclosed. Fortunately, biometric recognition provides an elegant solution to these shortcomings by identifying a person based on their physiological or behavioural characteristics. However, due to the diverse nature of biometric applications (e.g., unlocking a mobile phone to cross an international border), no biometric trait is likely to be *ideal* and satisfy the criteria for all applications. Therefore, it is necessary to investigate novel biometric modalities to establish the identity of individuals on occasions where techniques such as fingerprint or face recognition are unavailable. One such modality that has gained much attention in recent years which originates from forensic practices is the lip.

This research study considers the use of computer vision methods to recognise different lip prints for achieving the task of identification. To determine whether the research problem of the study is valid, a literature review is conducted which helps identify the problem areas and the different computer vision methods that can be used for achieving lip print recognition. Accordingly, the study builds on these areas and proposes lip print identification experiments with varying methods which identifies individuals solely based on their lip prints and provides guidelines for the implementation of the proposed system. Ultimately, the experiments encapsulate the broad categories of methods for achieving lip print identification.

The implemented computer vision pipelines contain different stages including data augmentation, lip detection, pre-processing, feature extraction, feature representation and classification. Three pipelines were implemented from the proposed model which include a traditional machine learning pipeline, a deep learning-based pipeline and a deep hybrid-learning based pipeline. Different metrics reported in literature are used to assess the performance of the prototype such as IoU, mAP, accuracy, precision, recall, F1 score, EER, ROC curve, PR curve, accuracy and loss curves. The first pipeline of the current study is a classical pipeline which employs a facial landmark detector (One Millisecond Face Alignment

algorithm) to detect the lip, SURF for feature extraction, BoVW for feature representation and an SVM or K-NN classifier. The second pipeline makes use of the facial landmark detector and a VGG16 or ResNet50 architecture. The findings reveal that the ResNet50 is the best performing method for lip print identification for the current study. The third pipeline also employs the facial landmark detector, the ResNet50 architecture for feature extraction with an SVM classifier.

The development of the experiments is validated and benchmarked to determine the extent or performance at which it can achieve lip print identification. The results of the benchmark for the prototype, indicate that the study accomplishes the objective of identifying individuals based on their lip prints using computer vision methods. The results also determine that the use of deep learning architectures such as ResNet50 yield promising results.



# Contents

|           |  |    |
|-----------|--|----|
| Chapter 1 | Introduction .....                         | 1  |
| 1.1       | Introduction and Background .....          | 1  |
| 1.2       | Research Problem .....                     | 2  |
| 1.3       | Aims and Objectives .....                  | 4  |
| 1.4       | Assumptions and Constraints .....          | 4  |
| 1.5       | Research Methodology .....                 | 6  |
| 1.6       | Contributions .....                        | 7  |
| 1.7       | Dissertation Outline .....                 | 7  |
| 1.8       | Conclusion .....                           | 9  |
| Chapter 2 | Problem Background .....                   | 12 |
| 2.1       | Introduction .....                         | 12 |
| 2.2       | Environment .....                          | 13 |
| 2.3       | Biometrics .....                           | 16 |
| 2.4       | Lip Biometrics .....                       | 23 |
| 2.5       | Conclusion .....                           | 28 |
| Chapter 3 | Algorithm Background .....                 | 29 |
| 3.1       | Introduction .....                         | 29 |
| 3.2       | Traditional Machine Learning Methods ..... | 29 |
| 3.3       | Deep Learning Methods .....                | 37 |
| 3.4       | Similar Work .....                         | 42 |
| 3.5       | Literature Trends and Gaps .....           | 46 |
| Chapter 4 | Research Methodology .....                 | 49 |
| 4.1       | Introduction .....                         | 49 |
| 4.2       | Research Design .....                      | 49 |
| 4.3       | Research Paradigm .....                    | 50 |
| 4.4       | Research Methods .....                     | 51 |
| 4.5       | Justification .....                        | 53 |
| 4.6       | Population and Data Sampling .....         | 54 |
| 4.7       | Data Analysis .....                        | 55 |
| 4.8       | Reliability and Validity .....             | 55 |
| 4.9       | Ethical and Legal Considerations .....     | 58 |
| 4.10      | Risks .....                                | 60 |

|                      |  |     |
|----------------------|--|-----|
| 4.1.1                | Conclusion .....   | 60  |
| Chapter 5            | Methods .....  | 63  |
| 5.1                  | Introduction .....   | 63  |
| 5.2                  | Model.....   | 65  |
| 5.3                  | Prototype .....  | 67  |
| 5.3.1                | Capturing and Data Augmentation .....                          | 67  |
| 5.3.2                | Lip Detection.....   | 68  |
| 5.3.3                | Traditional Machine Learning Pipeline .....                    | 69  |
| 5.3.4                | Deep Learning Pipeline.....                                    | 72  |
| 5.3.5                | Deep Hybrid Learning Pipeline.....                             | 76  |
| 5.4                  | Benchmark .....  | 77  |
| 5.4.1                | Functional Requirements.....                                   | 78  |
| 5.4.2                | Non-functional Requirements.....                               | 78  |
| 5.5                  | Conclusion.....  | 82  |
| Chapter 6            | Results .....  | 84  |
| 6.1                  | Introduction .....   | 84  |
| 6.2                  | Operational Results .....                                      | 84  |
| 6.2.1                | Dataset Selection.....   | 84  |
| 6.2.2                | Lip Detection.....   | 86  |
| 6.2.3                | Lip Print Recognition.....                                     | 89  |
| 6.3                  | Metrics .....  | 93  |
| 6.3.1                | Intersection over Union Performance for Object detection ..... | 94  |
| 6.3.2                | Traditional Pipeline with SVM.....                             | 96  |
| 6.3.3                | Traditional Pipeline with K-NN .....                           | 97  |
| 6.3.4                | Deep Learning Pipeline with VGG16 .....                        | 99  |
| 6.3.5                | Deep Learning Pipeline with ResNet50 .....                     | 101 |
| 6.3.6                | Deep Hybrid Learning Pipeline.....                             | 104 |
| 6.4                  | Discussion on Findings .....                                   | 105 |
| 6.5                  | Comparison with Similar Systems .....                          | 107 |
| 6.6                  | Conclusion .....   | 109 |
| Chapter 7            | Conclusion .....   | 111 |
| 7.1                  | Introduction.....  | 111 |
| 7.2                  | Objectives of Study .....                                      | 111 |
| Research Objective 1 | .....  | 112 |
| Research Objective 2 | .....  | 112 |

|                            |     |
|----------------------------|-----|
| Research Objective 3 ..... | 112 |
| Research Objective 4 ..... | 113 |
| Research Objective 5 ..... | 113 |
| 7.3 Summary .....          | 113 |
| 7.4 Findings.....          | 115 |
| 7.4.1 Limitations.....     | 115 |
| 7.4.2 Trends.....          | 116 |
| 7.5 Impact .....           | 117 |
| 7.6 Future Work.....       | 118 |
| 7.7 Conclusion.....        | 119 |
| References .....           | 121 |



UNIVERSITY  
OF  
JOHANNESBURG



# List of Figures

|   |    |
|---|----|
| Figure 1.1: Roadmap of the chapters in this dissertation .....                                | 9  |
| Figure 2.1: Various biometric modalities that have been proposed for person recognition ..... | 18 |
| Figure 2.2: A typical biometric system .....  | 19 |
| Figure 2.3: Significant breakthroughs in the history of automated face recognition .....      | 21 |
| Figure 2.4: Anatomy of the lip .....  | 24 |
| Figure 2.5: Suzuki and Tsuchihashi classification of lip prints .....                         | 24 |
| Figure 2.6: The difference between physiological and behavioural lips .....                   | 27 |
| Figure 3.1: 68 facial landmark coordinates surrounding each facial structure .....            | 31 |
| Figure 3.2: A basic CNN architecture .....  | 38 |
| Figure 3.3: Visual representation of LeNet-5 architecture .....                               | 39 |
| Figure 3.4: Visual representation of AlexNet architecture .....                               | 39 |
| Figure 3.5: Visual representation of VGG16 and VGG19 architectures .....                      | 40 |
| Figure 3.6: Residual Block .....  | 41 |
| Figure 5.1: Proposed model for lip print recognition .....                                    | 64 |
| Figure 5.2: Traditional Machine Learning Pipeline .....                                       | 69 |
| Figure 5.3: A representation of SURF features over the original image .....                   | 71 |
| Figure 5.4: Deep Learning Pipeline .....  | 72 |
| Figure 5.5: Architecture of VGG16 .....   | 73 |
| Figure 5.6: Architecture of ResNet50 .....  | 75 |
| Figure 5.7: Deep Hybrid Learning Pipeline .....   | 75 |
| Figure 5.8: ResNet50 with SVM classifier .....  | 76 |
| Figure 5.9: Example of ROC Curve obtained from .....  | 80 |
| Figure 5.10: Example of Precision-Recall Curve obtained from .....                            | 81 |
| Figure 5.11: Accuracy and Loss curves for CNN model obtained from .....                       | 82 |
| Figure 6.1: Sample images from LFW .....  | 84 |
| Figure 6.2: Sample images from CFD .....  | 85 |
| Figure 6.3: Visual representation of the bounding box for the lip .....                       | 85 |
| Figure 6.4: Instances of the Haar cascade technique .....                                     | 86 |
| Figure 6.5: Instances of the One Millisecond Face Alignment algorithm .....                   | 87 |

|   |     |
|---|-----|
| Figure 6.6: Traditional pipeline confusion matrices.....  | 89  |
| Figure 6.7: Shared misclassifications from traditional pipeline .....   | 90  |
| Figure 6.8: Deep learning pipeline confusion matrices .....   | 90  |
| Figure 6.9: Samples of misclassified lips from deep learning pipeline .....                                     | 91  |
| Figure 6.10: Deep hybrid learning confusion matrix .....  | 91  |
| Figure 6.11: Samples of misclassified lips from deep hybrid learning pipeline                                   | 91  |
| Figure 6.12: Samples of IoU score, ground truth region and predicted region of<br>different targets .....       | 93  |
| Figure 6.13: AP of samples from the dataset.....  | 94  |
| Figure 6.14: Samples of IoU score, ground truth region and predicted region of<br>Haar cascade classifier ..... | 94  |
| Figure 6.15: SVM Pipeline ROC Curve.....  | 95  |
| Figure 6.16: SVM Pipeline PR Curve .....  | 96  |
| Figure 6.17: K-NN Pipeline ROC Curve .....  | 97  |
| Figure 6.18: K-NN Pipeline PR Curve.....  | 97  |
| Figure 6.19: VGG16 ROC Curve .....  | 98  |
| Figure 6.20: VGG16 PR Curve.....  | 99  |
| Figure 6.21: VGG16 Loss Curve.....  | 99  |
| Figure 6.22: VGG16 Accuracy Curve .....   | 99  |
| Figure 6.23: ResNet50 ROC Curve.....  | 101 |
| Figure 6.24: ResNet50 PR Curve .....  | 101 |
| Figure 6.25: ResNet50 Loss Curve .....  | 101 |
| Figure 6.26: ResNet50 Accuracy Curve.....   | 102 |
| Figure 6.27: Deep Hybrid Learning Pipeline ROC Curve.....   | 103 |
| Figure 6.28: Deep Hybrid Learning Pipeline PR Curve .....   | 103 |

## List of Tables

|  |     |
|--|-----|
| Table 2.1: A brief history on the study of lips.....                     | 25  |
| Table 6.1: Summary of Misclassifications from Pipelines .....            | 89  |
| Table 6.2: SVM Pipeline Results .....                                    | 95  |
| Table 6.3: K-NN Pipeline Results .....                                   | 97  |
| Table 6.4: VGG16 Pipeline Results .....                                  | 98  |
| Table 6.5: ResNet50 Pipeline Results .....                               | 100 |
| Table 6.6: Deep Hybrid Learning Pipeline Results .....                   | 102 |
| Table 6.7: Summary of results obtained.....                              | 105 |
| Table 6.8: Comparison of similar systems for lip print recognition ..... | 106 |



# Chapter 1 Introduction

## 1.1 Introduction and Background

Selecting and deciding upon a suitable identification and authentication method is an important task in designing a security system. Over the years, identification systems have mostly relied on traditional approaches to either identify or authenticate users. These methods include access cards (ID cards, smart cards) for physical access control environments and passwords for virtual access control environments [3, 122]. Due to the rapid advances in technology, more devices are consequently being connected to the Internet. Authentication methods such as passwords are usually left to the end user's responsibility, hence exposing them to vulnerabilities of being forgotten, not being updated and easily predictable. Alternatively, access cards are vulnerable to loss, theft, and replication which further increases the risk of security breaches and unauthorised individuals [5, 122]. Unfortunately, although end users are constantly reminded and advised to strengthen and update passwords, hackers also constantly derive new and advanced methods to crack passwords with intentions to either illegally access sensitive information or to cause damage to data [122]. This unfortunate reality (data leaks and damage to data) has led to the development of biometric systems as a means to mitigate the risks associated with traditional approaches to identification and authentication.

Biometric identification is defined as the means of recognising a person based on acquired and measured physiological or behavioural characteristics [3, 9]. Fingerprints, for example qualify as a physiological characteristic while keystrokes are considered as a behavioural characteristic. The use of biometric methods for recognition has gained popularity over past few years due to levels of security they have to offer. Fingerprint, face, iris, palm print, retina, hand geometry and gait are various forms of biometric modalities. Most of these biometric modalities such as face, and fingerprint have and are being thoroughly researched. However, not all biometric modalities have been extensively researched in literature such as lip biometrics. This dissertation explores the extent of using the lip as a biometric modality.

The chapter begins with a concrete definition of the research problem in section 1.2 which justifies the need for the current study along with its hypothesis. Section 1.3 discusses the

aims and objectives that have been derived from the research problem. The particular constraints and assumptions of the current study are outlined in section 1.4. Thereafter, an overview of the research methodology is highlighted in section 1.5. Published work is discussed in section 1.6 and an outline of the dissertation is provided in section 1.7. The chapter concludes with a conclusion in section 1.8.

## 1.2 Research Problem

Today, employing appropriate security measures has become a concern in access control environments because sensitive information ending up in the wrong hands can be catastrophic. As discussed in section 1.1, traditional methods of authentication are not very secure as they are easily transferable and can be obtained by unauthorised users. Biometric methods deal with these disadvantages since users are identified by who they are and not by something they have to remember or possess. The reason why biometric methods have the upper hand is because they are more difficult to imitate [7, 9, 122]. However, current biometric systems are also lacking in certain areas with respect to accuracy, deployment, and template revocation, which are commonplace in access control environments [123]. As society faces increased threats from sophisticated attacks such as sensor spoofing and obfuscation attacks [127], new methods for identification have become a concern in access control environments [124].

Presently, there are many well-known biometric modalities that can be used for person recognition. Of these modalities, the most famous and commonly used modality is the fingerprint, followed by the face [127]. Choras [116] states that even well-established biometric modalities with a False Nonmatch Rate of 2% can cause challenges in real life security systems when deployed at a large scale. For instance, if there are 100 000 users per day, a 2% error rate results in 2000 false rejection per day [116]. Furthermore, concerns such as usability, user privacy and integration with end applications have not been adequately addressed [127].

More importantly, not all users can use any given biometric system. For example, an amputee who had their hand amputated cannot use a fingerprint or hand-based system. Similarly, a visually impaired person might have difficulties using an iris or retina-based

system. Therefore, to cater for these individuals, new biometric modalities should be appropriately researched to diversify modality offerings for biometric system designers.

Up until today, there is no optimal biometric modality that satisfies all the requirements for any application. However, it is fair to claim that due to the Coronavirus pandemic contactless methods of biometric recognition will soon become the norm [125]. Fingerprints have been the primary technique for biometric recognition. Nonetheless, individuals are now more aware than ever of the dangers of coming into contact with infected surfaces [126]. Nonetheless, it is worthwhile to uncover and investigate new knowledge and biometric technologies that can be used to address this issue. Thus, considering a smaller part of the face for recognition can be an effective way to solve this problem. One such modality that has the potential of recognising individuals which can be used in conjunction with facial recognition or speaker recognition to increase its overall accuracy and reliability is the lip. The introduction of lips is most welcome to the field of biometrics as they are unique to each individual and hold discriminative power (as seen in chapter 2). However, lip biometrics has not been extensively researched and it is still in its emerging stage. Furthermore, there is a lack of research and approaches that determine which algorithms can be used to effectively achieve lip print identification. Thus, this leads to the research problem of the current study:

***There is a lack of research and approaches, particularly deep learning-based methods, for using lip prints to achieve lip print identification.***

With the research problem identified, the hypothesis of the current study can now be determined. The hypothesis will serve as a way to measure whether the study has achieved what it set out to do. Thus, the determined hypothesis can be stated as:

***Deep learning-based methods can be employed to achieve lip print identification in an effective manner.***

Due to the lack of research in the realm of lip print recognition it is worthwhile to explore and investigate different computer vision methods that can adequately recognise an individual based on their lip print. Computer vision includes methods for acquiring, processing, and analysing not only images but videos as well to produce information [10]. Biometric identification or authentication deals with the recognition of individuals based on their physiological or behavioural characteristics. Thus, combining computer vision with

knowledge of human physiology and behaviour opens up the opportunity to explore new technology approaches for person recognition.

In the next section, the research aims, and objectives will be addressed based on the research problem. Each of the objectives will be operationalised progressing further into the study.

### 1.3 Aims and Objectives

In order to address the research problem, the study has been initiated to highlight the problem areas found and to provide a solution to them. The current study aims to create a lip print identification model that will be used to recognise individuals solely based on their lip prints. The model should consist of different computer vision methods and intends to support developments in the area of lip print recognition. This will be achieved using the following research objectives:

- **RO1** – Conduct a literature review within the research domain to identify the problem areas and relevant computer vision methods which can be used to achieve lip print identification along with appropriate datasets that can be employed.
- **RO2** – Adopt a high-resolution face dataset which will allow for discriminatory lip features to be extracted.
- **RO3** – Create experiments based on existing literature and findings by the author that can be used to achieve lip print identification.
- **RO4** – Implement a prototype based on the experiments which can recognise individuals based on their lip prints by employing computer vision methods from the designed model.
- **RO5** – Validate the performance of the prototype to determine its feasibility and report on these results in research articles and the dissertation.

### 1.4 Assumptions and Constraints

When pursuing research there are inherent assumptions and constraints found that the research is built on. These assumptions and constraints should be taken into consideration

in order to achieve the objectives of the study. Each of these are briefly described in the following subsections.

#### 1.4.1 Assumptions

There are specific assumptions that are made in the study that affect the applicability of the study and should remain true throughout the study:

- 1 For the current study, it is assumed that lip prints are the patterns and grooves that exist on the surface of the lip and are unique to each individual.
- 2 Collection and testing of algorithms are only performed on images assumed to be free of any lip diseases.
- 3 The ground truth labels that encapsulate the lip created by the author are correct for the study.
- 4 The equipment used is appropriate for capturing samples with an appropriate fidelity that can be used for identification.
- 5 The right protocols and considerations were followed for the subjects in the adapted face dataset.

#### 1.4.2 Constraints

The constraints of a study are a set of boundaries that need to be in place for the study to be conducted. These constraints may hinder research efforts and in order to successfully address the study objectives some of the constraints should be dealt with. The initial research performed outlines the following constraints:

- 1 Currently, there are a limited number of lip datasets that exist within the research domain.
- 2 Compiling a new high-resolution lip dataset is a challenging task due to lack of resources and other crucial factors such as the COVID-19 pandemic.
- 3 Images may exhibit inherent "noise" which may impact the overall prediction outcome.
- 4 Privacy and ethical concerns arise when using a secondary dataset such as data anonymity. The data should be kept anonymous at all times and should not harm the participants in any way.



- 5 The limited amount of research within the area of lip print recognition may hinder the process of designing experiments.
- 6 Limited computational resources were available to train certain deep learning models.

## 1.5 Research Methodology

The research methodology that will be followed for this study will be outlined in this section and covered thoroughly in chapter 4. In order to address the objectives in an effective manner, a comprehensive research methodology should be posed which will define the research methods used throughout the study. A quantitative research approach will be adopted for the current study. A quantitative research approach determines a hypothesis which is then tested by implementing certain methods to collect measurable outcomes or results. It is based on collecting and converting data into a numerical form so that results can be made, and conclusions can be drawn [2]. This approach follows deductive reasoning, meaning that it starts out with a hypothesis and examines all possibilities to reach a conclusion [2].

The study will make use of a positivist research paradigm approach. This paradigm gives validity and objectivity to research, and it is based on precise methods. Research in this paradigm relies heavily on deductive logic, formulating hypotheses, testing hypotheses, and implementing certain methods to derive conclusions [5]. Since objectivity is the main principle, it is important to maintain an objective stance throughout the study in order to analyse the results accurately without any bias and prove the hypothesis to be either true or false.

The research methods that will be used for the study will include: a literature review, a model that outlines the designed experiments, and a prototype containing the implemented experiments . The problem background will give insight on what has been attempted in the particular domain thereby justifying the need for the study. Certain areas that will be addressed include the environment and the problem domain the study resides in. A critical assessment of existing work that has been performed within the problem domain will also be investigated. The areas mentioned above will contribute to the formulation of the model that defines the algorithms that can be used for lip print identification. Thereafter, a prototype system for lip print recognition will then be implemented from the proposed model

to prove that the model is feasible. The prototype system will produce results which will be analysed and compared to validate the hypothesis.

## 1.6 Contributions

During the course of the study, a manuscript titled "A comparative analysis of computer vision methods for lip print identification" was submitted to IET Biometrics which is currently under the revision stage and is included in Appendix A of this dissertation. The main objective of the work was to compare relevant methods that can be used for lip print identification. The article compared traditional and deep learning computer vision methods and how they performed on a common dataset for lip print recognition. The first pipeline is a traditional method with Speeded Up Robust Features (SURF) bagged features with either an SVM or K-NN machine learning classifier. The second pipeline compared the performance of the VGG16 and VGG19 deep-learning architectures. The paper provided a starting point for the advancement of the research necessary to produce this dissertation.

Another manuscript will be submitted to International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI). This article investigates the effectiveness of employing end-to-end object detection deep learning architectures to detect the lips and process lip prints for biometric identification purposes. The main aim of the paper is to propel the field of lip print recognition forward by applying a deep learning image segmentation technique to improve the performance of lip print identification. The major contribution of the study is the proposal of a modified YOLOR model on a publicly available face dataset to achieve lip print identification. The manuscript is included as Appendix B in this dissertation.

## 1.7 Dissertation Outline

The following section will provide a summary of the chapters that will be discussed in this dissertation. Figure 1.1 is a representation of the chapters contained within this dissertation. These chapters are as follows:

**Chapter 2 : Problem Background** – Research is conducted within the domain of biometrics and access control. This chapter will place the current topic within a contemporary context

and demonstrate knowledge on the area of focus, compares different methods and techniques while also revealing certain trends and gaps that are prevalent in existing literature.

**Chapter 3 : Algorithm Background** – Chapter 3 will analyse and compare different traditional and deep learning-based methods that can potentially be used to achieve a lip print recognition. This chapter also thoroughly investigates previous work that has been attempted in this domain thus far.

**Chapter 4 : Research Methodology** – The research methodology utilised for the current study will be thoroughly discussed in this chapter. From the knowledge gathered in the problem background (chapter 2) and the algorithm background (chapter 3), the methodology of the overall study is defined in this chapter. It outlines how the study will be facilitated to ensure smooth sailing of the current research.

**Chapter 5 : Methods** – Chapter 5 will investigate the proposed model containing the designed experiments, the prototype and the benchmark which will be used for the study. The model in section 5.2 discusses the different methods that can be used to develop the research prototype. Thereafter, a concrete implementation of the experiments is discussed in section 5.3. The benchmark section defines how the model could be measured along with which metrics will be derived for the study.

**Chapter 6 : Results** – This is the results chapter, which contains the performance outcomes gathered on the prototype. Chapter 6 will be based on the benchmark defined in chapter 5. It also highlights the important areas that need to be investigated further.

**Chapter 7 : Conclusion** – The study is finally concluded with a conclusion chapter. This chapter will highlight the significant findings of the study as well as the potential future work that may come from the study.

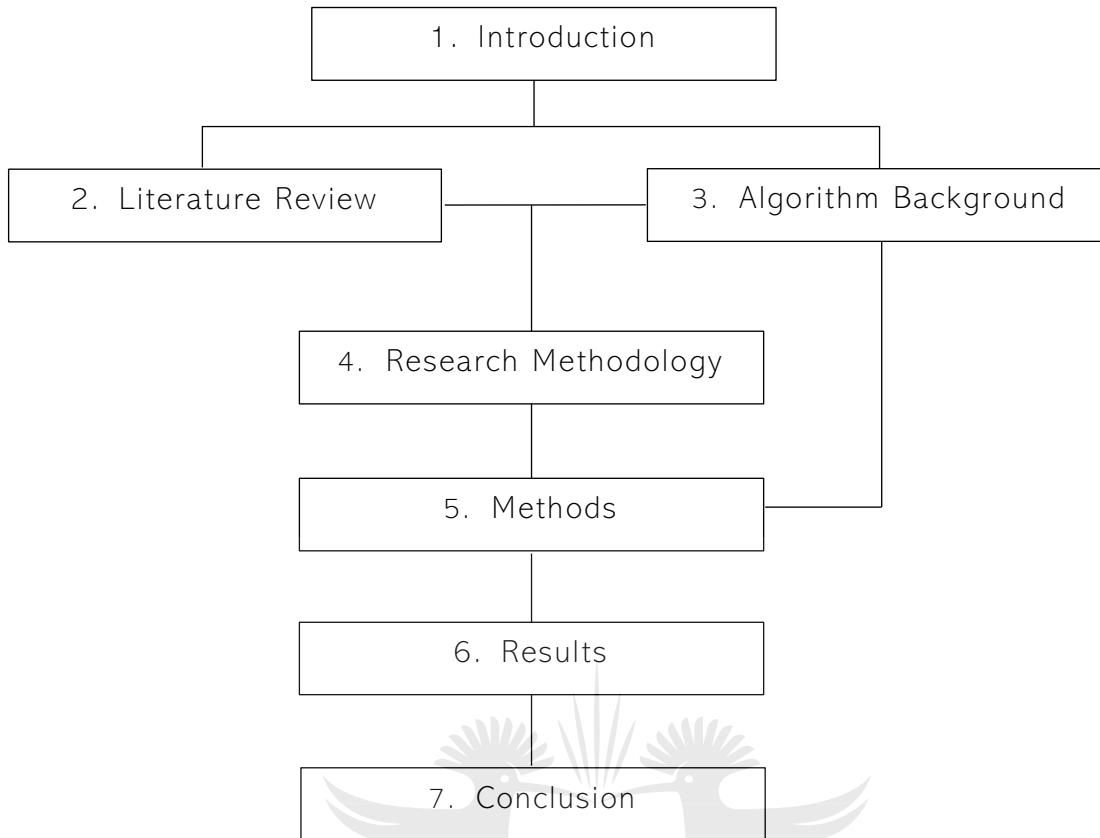


Figure 1.1: Roadmap of the chapters in this dissertation

## 1.8 Conclusion

This chapter begins by outlining the research problem the study will attempt to address throughout the research study. The research problem posed in the study shows that there is a lack of research and approaches for using lip prints to achieve lip print recognition in an effective manner. The study has defined a set of objectives in order to address the research problem. These objectives will be operationalised progressing further into the study. The main aim of the study is to create a lip print recognition model using deep learning computer vision methods.

During the course of the study the assumptions and constraints which have been identified for the current study were considered in order to achieve the objectives of the study adequately. These assumptions and constraints outline the appropriate scope and boundaries of the study and suggests the research methodology which can be adopted. As a result, the research methodology chosen for the current study is a positivist-based

quantitative approach. The methods used will include a literature review, a model, and a prototype. The literature review will give an insight on what has been attempted in the particular domain thereby justifying the need for a lip print recognition model. The model will formulate a solution to how the problem can be addressed. Thereafter, based on the designed model, a prototype will be developed.

Since the necessary foundation of the study has been established, the study may be pursued further, according to the outline specified in section 1.7. In order to determine suitable methods and techniques to use in the study, three areas of research will be investigated. These areas include access control, biometrics and lip print recognition using computer vision. The next chapter starts the literature review by having an in-depth look at the environment and domain of the research study.



PART I  
LITERATURE REVIEW



UNIVERSITY  
OF  
JOHANNESBURG

# Chapter 2      Problem Background

## 2.1 Introduction

A literature review is a systematic way of examining and incorporating existing literature to a specific topic [45]. It critically examines, assesses, and integrates research findings, theories and practices that are related to a current study. By conducting an effective literature review, the researcher should demonstrate a comprehensive and accurate understanding of the current knowledge, compare different theories and methods, and reveal trends and gaps in existing literature. In the previous chapter the necessary foundation for the current study was established. The research problem and the hypothesis were defined, and a set of objectives were derived which will be operationalised progressing further into the study. The purpose of a literature review is therefore, to place the current topic within a contemporary context and demonstrate knowledge on the area of focus, compare different methods and techniques while also revealing certain trends and gaps that are prevalent in existing literature.

Traditionally in access control, there is a trade-off in convenience when switching to more secure technologies. Finding a good balance between security and user acceptance is an exhausting task. However, biometrics are one of the few examples of a security upgrade that also increases convenience. Because traditional access control systems such as a password-based access control system have a critical vulnerability, biometrics is a useful solution to overcome some of the challenges posed by traditional access control systems.

In this chapter the various elements that facilitate access control systems and biometrics are explored. The literature review begins with a discussion on the environment the study resides in (access control), by covering the applicability and potential work that resides in the area. Section 2.2 is followed by a discussion on the domain of the current study (2.3). The current study resides within the domain of biometrics and computer vision. Section 2.3 discusses the different biometric categories, biometric properties, and the basic architecture of a biometric system. Furthermore, it discusses how computer vision plays a key role in biometric recognition. The next section, section 2.4, discusses the history of lip biometrics

and how the lip can be used as a biometric modality. The chapter is concluded in section 2.5.

## 2.2 Environment

Due to the exponential growth of the digital world, our lives are increasingly becoming intertwined with countless digital applications and systems. In many cases, access to these systems needs to be secure and authenticated. The security of these systems is a top priority because the service must only be delivered to an authenticated user. In order to provide access control, the user needs to be authenticated before gaining access to the environment. Once the user has gained access using some kind of authenticator their actions are restricted based on their permission levels. These concepts, namely, access control, identification, authentication and the different types of authenticators will be discussed further in the subsections that follow.

### 2.2.1 Access Control

One of the fundamental concepts of security, access control, restricts or regulates access to sensitive data [1]. At its essence, access control restricts the flow of information and determines how the user or system can interact with the computing environment and without access control there is no security. Therefore, failure to restrict access to certain individuals could have adverse effects on the organisation.

Access control can be split into two groups or environments: physical access control and logical access control [2]. Physical access control limits access to an area such as campuses, buildings, and rooms with valuable property. It is a key element in securing critical infrastructure such as transportation hubs, ports, and military infrastructure [2]. Logical access control limits access to a virtual entry point such as system files and computer networks. In both cases, before gaining access to these environments, the user must be identified or authenticated. A typical access control system may include advanced locks, access control cards or biometric recognition to allow access to individuals.

The core underlying idea of access control is to protect the confidentiality, integrity, and availability of data [3]. This means that the data should be protected from unauthorised



access, maintain its original state and it should be highly secure but easily accessible. This selective restriction of an access control mechanism typically consists of four fundamental components: identification, authentication, authorisation, and accounting [3]. The components applicable to the current study will be discussed in the section below.

### 2.2.2 Identification and Authentication

In the physical world, face-to-face identification, or the human ability to identify individuals is effortless. However, in a virtual environment, confirming ones identity can be quite challenging. Before the user can gain entry to the specific environment, it is important to determine the true identity of that particular person. In these cases, there are two scenarios that can occur: identification and authentication.

Identification, as described by [3] is identifying a user by analysing an identifier or a unique key to the access control system without the presence of a claim. This unique identifier can for example be a username or an email address. During this process, the identifier is provided to the access control system which then performs a search in the database and grants access if it exists in the enrolment database. However, a simple identifier on its own is not sufficient to keep a system secure because it leaves the system vulnerable to attacks. Therefore, more complex identifiers which are difficult to produce in an unauthorised manner should be used.

Authentication is commonly defined as the ability to confirm the identity of the user when the user also provides a claim to be an authorised user [3]. Therefore, for the user to be granted the permission to have access and use the resources the user must first provide a claim associated with an enrolled user. Thereafter, their identity must be proved by the system. The system commonly asks for evidence from the user attempting to do the authentication. This is achieved through the use of various authenticators such as a password, PIN, or a representation of a body part. These authenticators will be discussed further below in section 2.2.3.

Generally, a typical approach to manage information security is to mitigate it through available security mechanisms. In this case, access control, especially identification and authentication play a vital role in granting access to physical and logical resources within the environment [3]. Within the context of biometrics, identification captures biometric

information from a user and cross-references it against other users in the system. In doing so the system determines who the user is. However, authentication aims to determine if a person is who they claim to be by comparing their characteristics to a reference biometric template saved in the database. The current study will focus on identification.

### 2.2.3 Authenticators

An authenticator is the means used to confirm the identity of a user in access control. They are used to prove that a user is really who they claim to be. This subsection gives a description of the different types of authenticators. These are something you know, something you possess and something you are [3].

The first authenticator, something you know, is the most common type of authenticator. It usually requires users to enter something stored in their memory to prove their identity [4]. An example of this type of authenticator is a password or a PIN. During the authentication stage, the user's password is compared against the previously stored hashed password in the database. If the hash of the password received matches with the stored hash, the user is granted access to the system and its resources. However, if the passwords do not match access is denied. Limitations associated with this authenticator include forgotten and stolen passwords [4]. In the case of forgotten passwords, the user is denied access even though they are authorised to do so. However, when the password is stolen the user can be oblivious that their password is stolen even though access is granted because there is no direct relationship between the user and the password.

The next authenticator, something you possess, generally requires the user to have a physical item in their possession. This physical item is a token, and it is typically scanned using a receiver at access control points. Examples of these tokens include smart cards, smart phones, or SIM cards [59]. One disadvantage of this authenticator is that it is susceptible to theft [5]. Therefore, whoever is in possession of the token can gain access to the system and there is no direct relationship between the user and token.

The last authenticator, something you are, deals with the characteristics of the user. These characteristics are body measurements called biometrics [6]. In simple terms, this authenticator requires biometric information from the user to prove their identity. Examples of biometric recognition include fingerprint, face, and iris recognition. Generally, biometric

recognition works by comparing the user's biometric data with the referenced biometric data stored in the database [59]. One issue regarding biometric recognition is data privacy. Some users are reluctant to biometric recognition due to the irreversible link between the biometric trait and the personal information about a person [59].

However, when assessing the authenticators against their limitations, one authenticator stands out- something you are i.e., biometrics, due to the direct link between the user and their biometric trait [3]. The next subsection will unpack the core properties of biometrics along with its applications and future opportunities.

## 2.3 Biometrics

From the previous subsection it was highlighted that traditional authentication (such as passwords, smart cards) methods have certain limitations and disadvantages. While biometric systems have their own limitations, they deal with the limitations present in other authenticators [7]. In other words, it relieves the user from the predicament of remembering multiple passwords and the biometric data is not reversible to its original sample. Biometrics is defined as the discipline of recognising a user based on their physical and behavioural characteristics [1, 59]. These characteristics are inherently linked with the user and cannot be separated from them. For example, the fingerprints of a user belong only to that specific user and cannot be detached from them. Therefore, the identity claimed by the user can be verified reliably.

### 2.3.1 Biometric Categories

Biometric systems aim to mimic the recognition process of human beings. The human ability to recognise a face, voice or walking pattern of individuals is an incredible pattern recognition process that captures and stores certain characteristics about the observed individual. However, the field of biometrics has gone further in the art of recognition; it not only speeds up the recognition process, but it also introduces new modalities such as iris, veins, and lip prints. Biometrics can be classified into two main categories: physiological and behavioural [7].

Physiological characteristics of a person refer to the consistent traits of the human body that do not change during their lifetime [8]. They are based on hereditary characteristics and therefore morphological traits are unique even amongst twins. Physiological characteristics can either be genetic or phenotypic. Phenotypic traits mainly consist of fingerprints, iris, retina, and palm veins whereas genetic traits consist of DNA [7, 8].

Behavioural characteristics are based on the pattern of actions by a person [3, 7]. Therefore, it deals with the personal behaviour of a person. While physiological characteristics are hereditary, behavioural characteristics is based on acquired actions. Therefore, unlike physiological-based characteristics, which is always present, behavioural-based characteristics, only exists when the person performs an action in the presence of a sensor. Behavioural characteristics are learned and acquired over time. Examples of these characteristics include voice, signature, gait, and keystroke dynamics [8].

### 2.3.2 Biometric Properties

A biometric modality is described as biometric information that can differentiate one individual from another [3]. In simple terms it is a type of trait, either physiological or behavioural. Irrespective of the trait, there are a few properties that can help determine which biometric modalities can be used in real-life scenarios. These properties are [3, 7, 9]:

1. **Universality:** This property observes how frequent the biometric trait is in the general population. Therefore, every individual using the system should have that trait.
2. **Uniqueness:** The underlying trait must be adequately different across individuals.
3. **Permanence:** The biometric trait must be resistant to change over a period of time. The trait should not change overtime otherwise it is not useful.
4. **Measurability:** It must be possible to digitise the raw data and process it to extract certain features.
5. **Performance:** The system that uses the biometric trait should have a high degree of performance which includes the recognition accuracy and the resources used.
6. **Acceptability:** The general population using the system should willingly present their trait to the system.
7. **Circumvention:** This measures the extent to which the system can be deceived by using artifacts such as fake fingers or 3D face masks.

It is difficult to find the “best” biometric trait that satisfies all the conditions for any application. In other words, *“no biometric is ideal but most of them are admissible”* [9]. To put it another way, each biometric has certain properties that may be useful to an application or system depending on the nature of its requirements.

A number of biometric modalities have been presented for biometric recognition (which can be seen in figure 2.1). However, the most popular biometric modalities used today are fingerprint, face, and iris [127]. Large databases, such as driving license and immigration databases, are available, which is one of the key reasons for the popularity of fingerprint and facial recognition. Law enforcement and other government agencies from all across the world have gathered these [127]. Recently, iris recognition has been increasingly adopted for large-scale recognition due to its high accuracy, however, there are relatively fewer iris databases. A large number of academic datasets contain face and fingerprint samples that are publicly available for research purposes. These datasets serve as a basis for research in biometric recognition and they are a basis for a number of evaluations and benchmarks.

Other biometric modalities, such as palm print and deoxyribonucleic acid (DNA), are increasingly being used in forensic and law enforcement applications [127]. Voice, signature, hand geometry, and vascular pattern recognition are being used in commercial applications for recognition. Unfortunately, their use thus far is limited. Modalities such as ear, gait, keystroke dynamics, retina/sclera, electrocardiogram (ECG), and electroencephalogram (EEG) have been proposed for biometric recognition. However, they have not attained a sufficient level of sophistication and acceptance [127].

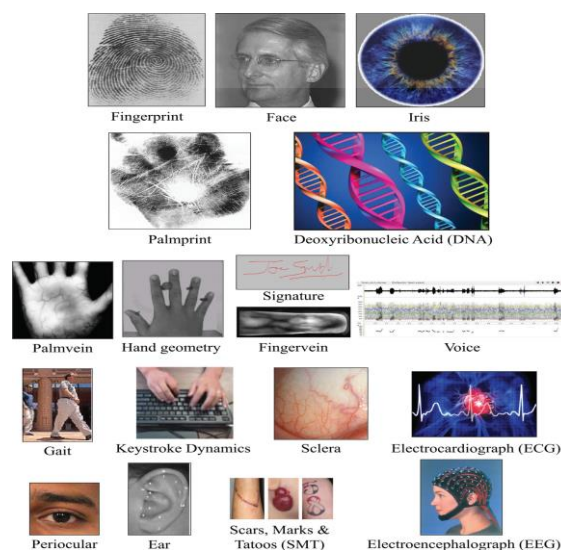


Figure 2.1: Various biometric modalities that have been proposed for person recognition [127]

### 2.3.3 Biometric System Architecture

Fundamentally, a biometric system is a pattern recognition system that collects biometric data from a person, extracts particular discriminatory features from the data, compares these features to the stored reference biometric template in a database, and makes a conclusion based on the results. Therefore, a generic biometric system consists of four components: sensor, feature extractor, matcher and database [9]. The architecture of a typical biometric system is represented in figure 2.2

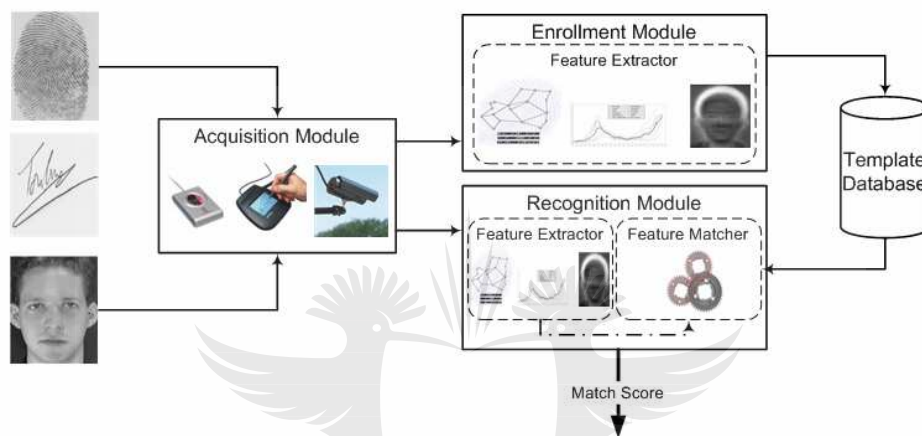


Figure 2.2: A typical biometric system [96]

The sensor is responsible for capturing biometric data from the individual. During this phase, the biometric trait is presented and captured using contact and non-contact-based approaches such as a camera, scanner, or microphone. The acquired trait is converted into a digital sample or representation which is then transferred to the subsequent module [7]. However, the quality of the acquired raw data depends heavily on the technology or the biometric sensor being utilised since it is responsible for transforming a real phenomenon such as a fingerprint to a digital form [7]. Therefore, these sensors should reduce the amount of added noise from the equipment and increase the quality of the captured trait.

Once the biometric characteristic has been captured it is sent to the feature extraction module. However, before certain salient features can be extracted, the data is first pre-processed to enhance its quality [9]. Thereafter, a set of salient features are extracted to represent the biometric trait. Since not all features are comparable a feature representation step is needed to make the features comparable (a requirement for a template). These features are then represented in a reference biometric template. The generated reference

biometric template is then sent for storage the database for future biometric match processes. The matching module is responsible for recognising the person. Therefore, once the biometric characteristic has been captured by the sensor, the feature extraction module proceeds to extract features from the acquired biometric characteristic [7]. The stored biometric reference template in the database is then used to compare these query features [7]. This comparison aims to confirm that the extracted features as found in the biometric data and the stored reference biometric template originate from the same individual.

The database module is used to store the generated template of the individual as well as their biographic information such as name and address. One interesting factor to consider regarding this module is the security of the stored template. If a template is compromised, it is impossible to reconstruct the original biometric trait. However, it is possible to replay the referenced biometric data if intercepted, which does not pose a threat to the system and the individual.

#### 2.3.4 Computer Vision for Biometric Recognition

As we are entering the age of Artificial Intelligence (AI), computers are becoming increasingly intelligent that it is now possible to interpret and understand the visual world. Machines can identify and locate an object that they “see” using various means such as cameras and videos. At its core, computer vision technology contributes significantly to developing highly intelligent systems. Computer vision has always been occupied with a variety of fields such as science, engineering, and pure art [10].

Computer vision is an important component of biometrics since it is concerned with the extraction and analysis of an image. To achieve a high-level recognition and robust performance, computer vision technology with state-of-the-art techniques can be applied to biometric systems. The integration between biometrics and computer vision is directly linked with various applications such as cyber security, facial recognition, and iris recognition.

#### Facial Recognition

Identification, access control, and forensics are just a few of the applications of facial recognition, which is one of the most popular research disciplines in computer vision. The

general process of an automated face recognition system generally consists of components for face detection, feature extraction and classification [11]. In the past algorithms such as Haar Cascades were used for object detection and Principal Component Analysis (PCA) for feature extraction [19, 104]. Over the years, computer vision methods and algorithms have been used to study constrained and unconstrained facial recognition [11]. 2D approaches have reached a certain level of maturity with a high performance. However, 3D approaches are being used as an alternative solution to face recognition because it is invariant to pose and lighting conditions [9]. This is achieved through state-of-the-art computer vision methods such as deep learning-based models. More recently, computer vision and facial recognition have been integrated into smartphones, which enables the technology to identify their users. Figure 2.3 depicts a brief summary of the milestones in the development of face recognition algorithms.

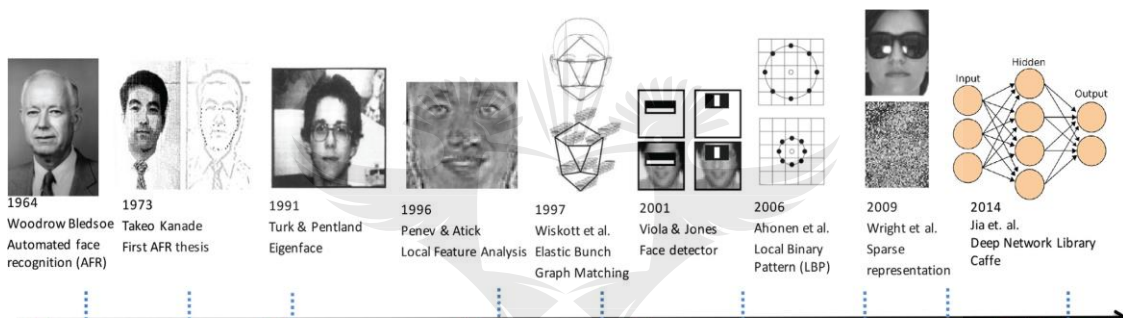


Figure 2.3: Significant breakthroughs in the history of automated face recognition [127]

## Iris Recognition

Iris recognition is an important application for computer vision because it provides a useful cue for human face analysis: it can be used for determining gaze direction, and it can be used for identification [12]. Firstly, if the head has been located accurately, it can determine a region of interest, inside which the iris can be sought. Once the iris has been located, it can be pre-processed using methods such as grayscaling and histogram equalization. Thereafter, certain feature extraction methods such as a hough transform can be applied on the pre-processed image. Lastly certain machine learning classifiers such as Support Vector Machines (SVM) can be used for matching. Furthermore, with the help of computer vision methods it is also possible to estimate the direction of the gaze with a reasonable accuracy [12]. However, more sophisticated computer vision methods such as deep learning-based



models significantly improves the performance of these systems over traditional approaches [115].

### 2.3.5 Applications, Challenges, and Opportunities

The notion of biometrics has been around for over a century. It was first used for the purpose of criminal identification. However, in 1924, it progressed to the identification of both criminal and police personnel [3]. Since then, biometric identification has been the subject of intense research and development, with ground-breaking results. Biometrics is now adopted in government applications such as border control, e-voting and biometric passports [7, 59]. The most commonly used passports today are second generation passports, which store two fingerprints and a photograph of the passport holder. Biometrics is also used in access control environments such as time-attendance systems or accessing remote information systems [7]. Many organisations have adopted this technology to prevent unauthorised persons from accessing their premises or information. Recent smartphones are equipped with biometric technologies which help users to safely unlock their mobile devices or access a mobile application. With recent updates it is now possible to select a form of biometric recognition such as fingerprint recognition and facial recognition. Other applications of biometrics include forensics for criminal investigation and in the military to identify enemies on the battlefield [59].

It needs to be understood that biometric attributes are not without its concerns. A biometric is not a secret, especially with physiological biometric data. One of the concerns regarding biometrics is its security since biometric databases have and can be compromised [1]. In this case, a breached biometric database poses more of a threat than a breached password database. The reason is that users can always change their passwords. However, when biometric data is leaked it is problematic and, in some cases, impossible, to change a user's physiological characteristics such as their fingerprint or iris. Fortunately, recent advances in cancellable biometrics, which is an active area of research, can be used to overcome these challenges [7].

The genuine acceptance rate of biometric systems has increased due to the adoption of computer vision and machine learning algorithms. However, more research is still needed, particularly in real-world applications such as surveillance systems. As mentioned previously,

it is difficult to find a biometric characteristic that satisfies all the requirements for any application. Therefore, many other physiological characteristics such as ear biometrics and lip-based biometrics and behavioural characteristics such as gait analysis and keystroke dynamics are being further explored and investigated so that modalities that can be better suited to some situations can be available.

## 2.4 Lip Biometrics

Human lip recognition is one of the most intriguing and emerging approaches of human recognition, originating in criminal and forensic professions. It can be considered as a new type of biometric measurement. In this section, various aspects of lips and lip recognition will be explored to provide information that is relevant to the current study.

### 2.4.1 History and Background

From an anatomical point of view, human lips can be described as two sensitive mucocutaneous folds comprised of skin, muscles, mucous membranes, and sebaceous glands [13]. The meeting point of the two mucocutaneous folds is termed as “labial cord” which is a white wavy line known as the vermillion border [95]. For personal identification, the area of interest is the mucosal area called vermillion or Klein’s zone [95] which can be seen in figure 2.4. This area (upper and lower lip vermillion) is covered by lines and patterns called lip prints. Lip prints can be defined as normal wrinkles or grooves present between the inner labial mucosa and outer skin of the lips [14]. These patterns are identifiable within six weeks of intrauterine life. This biological phenomenon was first noticed by anthropologist R. Fischer in 1902 [14]. While criminologist Edmond Locard originally proposed utilizing lip prints for personal identification and criminalization in France in 1932, Le Moyne Snyder was the first to suggest using lip prints for identification purposes. He introduced a case where lip prints helped crime investigators in 1950 and proved that the grooves present on the lips are as distinctive as fingerprints [15].

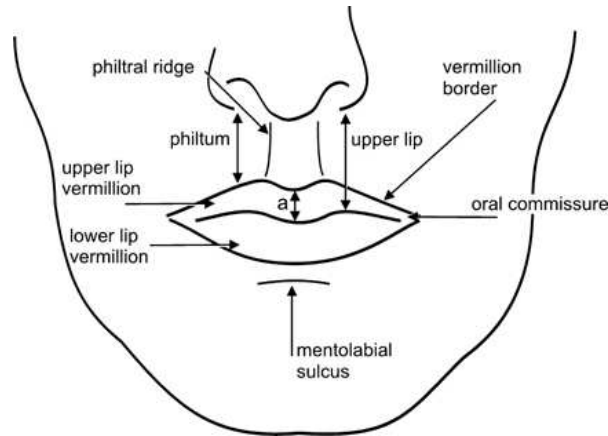


Figure 2.4: Anatomy of the lip [94]

Until 1950, the existence of grooves or patterns was postulated but no practical application was suggested. It was not until around 1950 that considerable research in this field began. Santos was the first to establish a classification system for lip grooves in 1967. Straight lines, curved lines, angled lines, and sine-shaped lines are all examples of this classification. Suzuki and Tsuchihashi investigated 1364 people at the Department of Forensic Odontology between 1960 and 1971 [15]. The pattern of grooves visible on the lips is unique to each human being, according to their research. Their research also devised a new classification method of lip prints which are: Type I; vertical groove, Type I'; partial length groove of Type I, Type II; branched groove, Type III; intersected groove, Type IV; reticular pattern and Type V; other patterns such as ellipses, triangles and ovals [16]. These patterns are shown in figure 2.5. However, Renaud's Classification was the first systematic way of achieving more consistent consensus of a potential match. This classification divides the upper and lower lip into right and left with four quadrants [13]. This classification has an addition of three types of lip prints namely, incomplete bifurcated grooves, grooves in X and horizontal grooves.

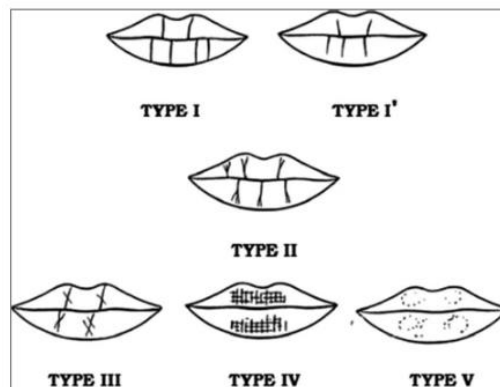


Figure 2.5: Suzuki and Tsuchihashi classification of lip prints [16]

In the year 1966, Poland took an interest in the study of lip prints when a lip print was revealed on a glass window at the scene of a crime [15]. Extensive research was carried out and its results were comparable to previous work. The research, however, did not allow for practical applications as yet. It was only in 1982, when the Forensic Institute of Warsaw University Criminal Law Department collected lip prints of 1500 individuals from different cities around the country and it established that identification based on lip prints is possible. By 1999 lip prints were generally accepted by the forensic community as a form of identification and were considered unique [17]. Since then, lip prints have been used for identification and criminalisation purposes especially in forensic practices. A brief history of the study of lips is summarised in table 2.1.

Table 2.1: A brief history on the study of lips [95]

| Year | Important landmarks  |
|------|--|
| 1902 | Fisher described the physiological features of lip   |
| 1930 | A study was designed which led to lip print use in criminology   |
| 1932 | The importance of cheiloscopy was acknowledged by Edmond Locard  |
| 1950 | The feasibility of using lip prints for person recognition was introduced                                |
| 1960 | The grooves and criss-cross lines on the lips were recommended to be classified into separate categories |
| 1972 | The uniqueness of lip prints was proven after studying 4000 lip prints                                   |
| 1974 | Another study was conducted which resulted in a new classification for lip prints                        |
| 1981 | The inspection of a person's lip print is used as a technique for determining their identity             |
| 2000 | To emphasize the importance of cheiloscopy in forensic science identification, lip patterns were studied |

Forensic odontology is defined as the study of dental applications in legal activities and the inspection as well as the assessment of dental evidence which can be presented in the court of law [17]. The study of lip prints in the field of forensic odontology is known as *cheiloscopy* which originates from the Greek word *cheilos* meaning lips [17]. The importance of

cheiloscopy is linked to the fact that lip prints are unique to each individual and permanent even after death. In forensic odontology, the mouth, particularly lip prints can be used for identification. In this case, lip prints in crime scenes can play a major role in criminal investigations since latent or visible lip prints can be found on various objects such as glasses, cups, clothes, or cigarette butts. Experts within this field can record lip prints found on these objects and compare the patterns to a suspect's lip print. These patterns rarely change which means that they are resistant to external factors such as a physical wound or subjection to hot and cold stimuli [14]. This enables it to be a reliable tool in investigations. However, an important factor to consider is that a well-defined lip print pattern depends on whether the mouth is open or closed [15]. When the mouth is closed the lip displays distinct grooves whereas when the mouth is open the grooves can be unclear.

#### 2.4.2 Physiological and Behavioural Characteristics of Lips

In the previous subsection (2.4.1), it was mentioned that the grooves that appear on the surface of the lips are unique. Therefore, it can be treated as a biometric measure and qualifies the criteria of being a recognition system. Compared to other biometric modalities such as fingerprint, face or iris, lip features contain both physiological and behavioural characteristics [18]. Physiologically, individuals have distinctive lips since it is hereditary. Alternatively, individuals can also be differentiated by the way they talk even when speaking the same utterance [18]. Therefore, this makes lip recognition an interesting method of human recognition.

Physiological features of the lips are defined as static features which can be extracted from lip images. Physiological lip features can be categorised into two types: geometric shape and texture features [18, 97]. Geometric shapes focus on the static shape of the lip while texture features determine the static texture of the lip [18]. Behavioural features of the lips can be defined as the movement of the lip during utterance, These features are dynamic and can be extracted from sequences of lip images and are useful for speaker identification and verification. The difference between physiological and behavioural lip features is illustrated in figure 2.6. The current study will adopt physiological features of the lips for human recognition due its discriminative power. It should be noted that it would be difficult to retrofit an existing database for the dynamic case.

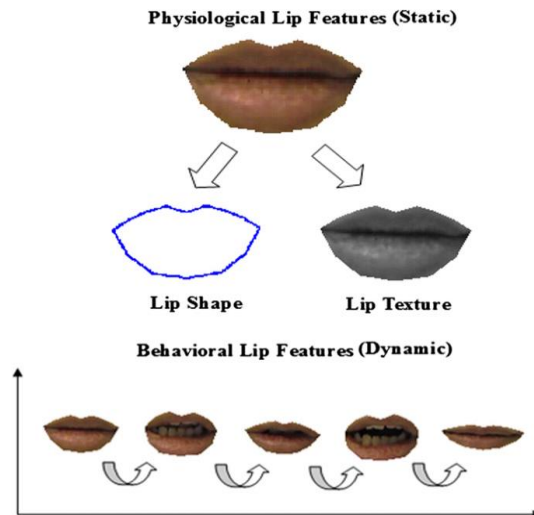


Figure 2.6: The difference between physiological and behavioural lip features [18]

Both physiological and behavioural lip features have the potential to be used as a biometric modality. These features can be implemented in different scenarios such as speech recognition, speaker identification, multi-modal audio-video speech recognition and identification, lip-reading and finally, lip print recognition based on static face images [97].

Choras states in his lip recognition proposal paper that using the lip as a biometric modality has the following advantages [97]:

1. User interaction is not needed because lip biometrics is regarded as passive biometrics.
2. Images of the user may be obtained from a distance without the knowledge of the user.
3. Better results can be expected for lip biometrics than behavioural biometrics.
4. Lips can be implemented in hybrid systems such as face-lips biometric systems.

Lip biometrics can also be utilized in multimodal systems to improve the effectiveness of well-known biometrics such as face recognition [97]. Due to the limitations and drawbacks of traditional biometric systems, the concept of multimodal or hybrid systems has lately garnered attention. As a result, lip biometrics appear to be a natural alternative for supporting well-known biometric systems in developing applications including access control, border control, and biometric recognition. Similar work that has been conducted in the field of lip print recognition will be discussed further in section 3.4, thereby validating the lack of research in this area.

## 2.5 Conclusion

This chapter started with an overview on access control and areas within access control such as access control environments, identification and authentication, and the different types of authenticators. It was determined that using biometrics as an authenticator has many advantages over traditional authentication methods. Thereafter, a detailed discussion on biometrics was discussed including the role of computer vision in biometric recognition. Computer vision technology contributes significantly to developing highly intelligent systems and therefore can also be used in developing biometric systems. The history of lip biometrics was then discussed, and it was determined that lip prints hold discriminative power since they are unique to each individual.

Biometric identification has gained much attention recently. The main reason for this tendency is because biometric systems deal with the limitations present in traditional methods of authentication. While there are many well-established biometric systems, they each have their own drawbacks and limitations. Therefore, novel, and innovative solutions are still needed. This chapter has established that using the lip can be used as a biometric modality with its own advantages due to its discriminative power. Therefore, it would be worthwhile to investigate the different computer vision methods and techniques that can be adopted to achieve lip print recognition. The next chapter will have a look at traditional and deep learning methods that can potentially be used in the realm of lip print recognition.

# Chapter 3      Algorithm Background

## 3.1 Introduction

To conduct an effective literature review, analysing and comparing different methods that can potentially be used to achieve a specific outcome should be thoroughly investigated. In the previous chapter, the current topic was placed within a contemporary context, and it demonstrated knowledge on the area of focus which is biometrics and lip print recognition. The purpose of this chapter is, therefore, to provide a detailed discussion on various methods and algorithms applicable to lip print recognition. These algorithms are based on studies conducted by researchers who have done extensive research in this domain. The algorithms can be divided into two categories: traditional machine learning methods and deep learning methods. Currently, there is more traditional machine learning research for lip print recognition rather than deep learning methods. By illuminating these methods, a model for lip print recognition can be designed.

This chapter begins with a discussion on traditional machine learning methods which can be used for lip print recognition (3.2). These methods are divided into respective pattern recognition steps such as pre-processing, feature extraction and classification. Section 3.3 investigates the various deep learning methods that can be used in the realm of lip print recognition. In section 3.4 similar work that has been conducted in lip print recognition is investigated. Thereafter, in section 3.5, certain literature trends and gaps are revealed. This chapter is concluded in section 3.6 with a conclusion.

## 3.2 Traditional Machine Learning Methods

The way traditional machine-based systems are designed is well researched and works well when applied to lip print recognition. The steps involved in traditional biometric system usually include object detection, pre-processing, feature extraction, feature selection and classification. Therefore, the traditional biometric system route for lip print recognition usually follows the same steps mentioned above.



The first step is object detection which detects an object or the region of interest within the image frame followed by pre-processing the acquired region. Thereafter, features are extracted from the pre-processed data followed by feature representation and classification. The next subsections will discuss different methods used to achieve object detection pre-processing, feature extraction, feature representation and classification, respectively.

### 3.2.1 Object Detection (Lip Detection)

Object detection is a computer vision technique which identifies and locates the object of interest in an image or video with a bounding box [98]. For the case of lip print recognition, appropriate methods or approaches include making use of the Haar cascade approach and the One Millisecond Face Alignment with an Ensemble of Regression Trees approach. More detail on these methods is provided in the subsections below.

#### Haar Cascades

Haar cascades is a machine learning technique which uses a cascade function to train multiple positive and negative images [19]. The core basis for Haar cascades is the Haar-like features which are essentially rectangular regions that are calculated on the pixels in the image. There are three types of Haar features which include edge features for edge detection, line features for line detection and four-rectangle features used to detect sloped lines [19]. Haar cascades are one of the many algorithms that are currently being used for object detection. Consequently, it can be used to locate faces and facial features in an image.

#### One Millisecond Face Alignment with an Ensemble of Regression Trees

The algorithm presented by Kazemi et al [20] is an implementation of the One Millisecond Face Alignment with an Ensemble of Regression Trees (ERT). The algorithm is a cascade of linear regressors where each regressor returns an update of the current estimate of landmark positions. Therefore, an ensemble of regression trees is used to approximate landmark points from pixel intensities. The initial position estimates are computed based on a given face box. A landmark is a keypoint for a specific facial structure such as the eyes

and mouth. Specific (x, y)-coordinates are specified surrounding each facial structure. With these coordinates, an ensemble of regression trees is used to approximate the facial landmark positions with high quality predictions. The 68 coordinates surrounding each facial structure is visualised in the figure 3.1.

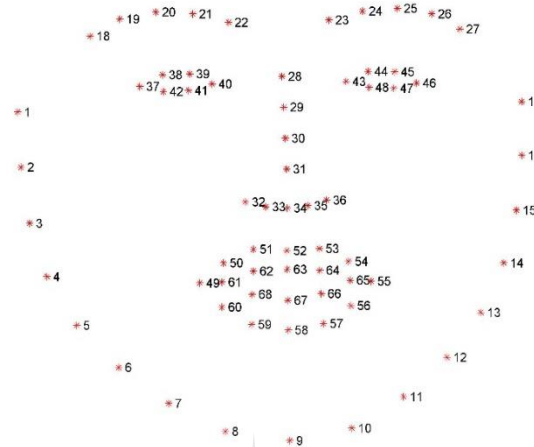


Figure 3.1: 68 facial landmark coordinates surrounding each facial structure used in One Millisecond Face Alignment

As a result of object detection, the sections of the image frame which are not relevant to the problem domain are discarded. Thus, the focus is placed on the object of interest, which enables pre-processing algorithms to further enhance the features on the object of interest.

### 3.2.2 Pre-processing

The pre-processing step usually involves simplifying the image data in such a way to make the feature extraction phase more effective. Hence, pre-processing is a crucial step because it suppresses noise and enhances important features. The pre-processing algorithms that will be investigated include grayscaling, histogram equalization, blurring and edge enhancement.

#### Grayscaling

Grayscaling is most commonly used in image processing to reduce computational complexity. It involves converting a complex colour space such as RGB (Red, Green, Blue) to a simpler colour space. Therefore, a grayscale image is simply an image with shades of gray with the luminance ranging from 0 to 255 [21]. Grayscaling removes all colour

information, leaving only the luminance of each pixel which provides the benefit of reducing computational complexity.

## Histogram Equalisation

Histogram Equalisation is an image processing technique that uses its intensity distribution to adjust the contrast of the image. In simple terms, it is used to enhance contrast in images. To enhance the image's contrast, it stretches out the most frequent pixel intensity of the image. Therefore, this allows the lower contrast areas in the image to gain a higher contrast. This method is more suitable for grayscale images rather than RGB images because using a complex colour space leads to dramatic changes in the image's colour balance [21]. Histogram equalisation has proven to be an approach that positively impacts various feature extraction methods further down the line [99].

## Blurring

Blurring is often used to remove or reduce unnecessary noise and details from the image, and it is also effective at smoothing images. An example of a method used for blurring is known as Gaussian blurring. This technique uses a Gaussian function to transform each pixel value in the image to a smooth blur [21]. Blurring is an effective technique since it removes noise around areas of interest before feature extraction. Blurring is a useful pre-processing technique because it removes noise around areas of potential key-points.

## Edge Enhancement

Edge enhancement is an image pre-processing filter that enhances the contrast of an image to improve its acutance. Edge enhancement filters are divided into two groups namely Gradient edge enhancement [117] and Laplacian edge enhancement [117]. These filters work by increasing the contrast of the pixels around specific edges ensuring that the edges are prominently visible [22]. This can be an effective pre-processing technique for lip print recognition to ensure that the lines found on the lips are visible for feature extraction to take place.

The various pre-processing approaches described in this section indicate how the acquired object of interest might be prepared for feature extraction. The pre-processing methods achieve a certain level of enhancement and noise reduction. These adjustments allow for feature extraction to take place effectively.

### 3.2.3 Feature Extraction

The next phase, feature extraction, is an important phase in any pattern recognition pipeline. Here, features refer to those discriminative attributes that are similar for objects in the same class. The main goal is to acquire the most appropriate information from the pre-processed data. The algorithms that will be discussed include Hough Transform, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Edge Detection, Local Binary Patterns (LBP) and Gabor Wavelets. Each of these will be discussed in the subsequent sections.

#### Hough Transform

A Hough transform is a feature extraction method which can be used to separate shapes in images. In particular, it has been used to extract shapes such as lines, circles, and ellipses. The basic idea behind a Hough transform is that every point in the edge map is converted to a line in Hough space. The areas where the most Hough space lines traverse are considered as true lines. The main advantage of this technique is that it is relatively unaffected by noise, and it is tolerant of gaps in feature boundary descriptions [23].

#### SIFT and SURF

Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) are two spectra detectors and descriptors which calculate features only in areas that are of interest in the image. SIFT makes use of gradient information calculated at each pixel to detect keypoints whereas [99] SURF adopts combinations of HAAR-like features of integrated pixel values instead of gradients [99]. Although both methods are robust, SURF is considered more beneficial because it extracts relevant features faster than SIFT [23]. However, it is

often useful to apply pre-processing methods such as histogram equalization on the images before extracting SIFT or SURF features [99].

## Edge Detection

Edges usually occur on the border between two different regions in an image where most of the information of an image is confined in edges. Edge detection is essentially the concept of detecting significant local changes in an image. An edge detector enhances the gradients in an image which can be represented as an edge, contour, or line. There are various types of edge detection methods such as Sobel edge detection [133], Canny edge [133] detection and Prewitt edge detection [23]. The Sobel method is concerned with detecting the magnitude and direction of gradients in an image while the Canny method adds a post-processing operation in order to produce edges that are more connected [99]. Edge detection effectively determines where certain objects start and end in an image.

## Local Binary Patterns

Ojala et al [24] was the first to develop Local Binary Patterns (LBP) in 1994 by. This method encodes patterns and contrast to define texture in an image. LBP can be used for many applications such as texture matrix, feature descriptor and image processing operators. The LBP uses a set of histograms of the feature descriptor neighbourhood surrounding each pixel to create descriptors and considers the result as a binary number [23]. Therefore, this makes it a simple yet effective method for feature extraction. It is interesting to note that when LBP is integrated with the histogram of gradients (HOG) descriptor, it improves detection performance for some applications [100].

## Gabor Wavelets

In 1946 a Hungarian electrical engineer, Dennis Gabor, proposed the Gabor function [101]. Presently, Gabor functions are more commonly used for feature extraction, especially in texture-based image analysis and more practically in face recognition [101]. Gabor wavelets are created from one particular atom by dilation. These wavelets provide a complete image

representation. This feature-based method aims to find important local features in an image and represent the corresponding information in an efficient way [101].

Extracting the correct features that relate to the problem space is an important phase. Therefore, the feature extraction methods selected for the study should highlight discriminative characteristics within the image so that they can be used to classify different objects.

### 3.2.4 Feature Representation

The next phase in the traditional machine learning pipeline is feature representation. In some cases, redundant or irrelevant features could overpower the crucial features for classification. Therefore, to overcome this challenge, feature selection methods can be adopted. Feature selection is the process of selecting a subset of the original features to train the model, thus reducing model complexity and the number of irrelevant features [103, 104]. The appropriate feature representation methods that will be discussed include a Bag of Words approach or in this case a Bag of Visual Words (BoVW) approach and Principal Component Analysis (PCA).

#### Bag of Visual Words

After discriminative features have been extracted using a feature extraction method, the K-means clustering algorithm is used to split these features into a number of clusters [102] where each cluster has features with similar descriptors. In general, the BoVW approach creates supervised classifiers based on visual words taken from labelled images to predict an unseen image. Therefore, the clustering algorithm creates a vocabulary of visual words to describe different local patterns in the image [102]. The number of clusters defines the size of the vocabulary which can vary from depending on the keypoints used.

#### Principal Component Analysis

Principal Component Analysis or PCA is an unsupervised feature reduction method for projecting data points from a high dimensional representation to a lower dimensional

representation while preserving the relevant linear structure [104]. This method produces principal components. Each component is a linear combination of the original variable and the components as whole form an orthogonal basis for the space of the data [104]. Therefore, PCA can be used to identify critical features without much loss of information.

After relevant features have been selected using a feature representation method, classification can now take place with a reasonably high accuracy. The next subsection will discuss the different classification techniques which can potentially be used for lip print recognition.

### 3.2.5 Classification

The classification phase of the traditional machine learning pipeline may take a supervised or an unsupervised approach. However, based on current literature, only supervised approaches have been used for lip print recognition such as Support Vector Machines (SVMs), K-Nearest Neighbours (K-NNs) and Random Forest. These classifiers are trained based on prior samples provided with ground-truth labels. These methods will be discussed further below.

#### Support Vector Machine

A Support Vector Machine or SVM is a supervised machine learning classifier that can be used for classification and regression purposes. However, they are most commonly used for classification tasks. The algorithm is based on the idea of creating a hyperplane that best divides the dataset into two classes [25]. SVMs make use of kernels which determine how data will be separated. A linear kernel is used for linear problems while a Radial Basis Function (RBF) kernel is used to solve non-linear problems [25].

#### K-Nearest Neighbour

K-Nearest Neighbour, also known as K-NN is a supervised and pattern classification machine learning classifier which helps determine which class the new test value belongs to when the nearest  $k$  instances are chosen, and the distance is calculated between them [25]. The  $k$  instances in the training set that are closest to the test value are identified and the distance

between those categories is calculated [25]. There are different distance measures than be used to calculate the distance such as Euclidean and Manhattan distance. K-NNs are widely used because of its uncomplicated nature.

## Random Forest

The Random Forest classifier which is an ensemble learning models consists of many decision trees which are trained with the bagging method. The bagging method is an ensemble of learning methods that improves the accuracy of machine learning algorithms. In simpler terms, to get a precise prediction the random forest classifier builds many decision trees and combines them together [25].

Classification is the final stage in the traditional machine learning pipeline. The classification phase determines whether the implemented pipeline has achieved recognition effectively. Now, that the traditional machine learning methods to achieve lip print recognition have been discussed, deep learning methods that can potentially be used to achieve lip print recognition will be investigated.

## 3.3 Deep Learning Methods

Deep learning approaches have recently gained much popularity and are being utilised in various fields such as computer vision and biometric recognition. Different from traditional machine learning, where the original image requires separate pre-processing and feature extraction steps, deep learning methods automatically pre-process and extract features from the image achieved in different layers. Therefore, deep learning methods learn high-level features from the data and eliminate the need for hard core feature extraction. The methods that will be discussed include different object detection and classification approaches such as Region-Based Convolutional Neural Networks (R-CNNs), You Only Live Once (YOLO) as well as different architectures of Convolutional Neural Networks (CNNs) including VGGNet, ResNet, LeNet-5 and AlexNet



### 3.3.1 Convolutional Neural Networks

The term deep neural networks refer to Artificial Neural Networks (ANNs). ANNs aim to simulate biological systems, corresponding to the human brain with neurons connecting to one another via synapses [26]. This analogy is retained in ANNs. Convolutional Neural Networks or CNNs are one of the most popular deep neural networks. They have shown exceptional performances in computer vision problems. A CNN takes an image as an input and assigns priority to different components of the image, allowing them to be distinguished from one another. They are designed to deal with data in the form of numerous arrays. [106]. CNNs use convolutional filters to extract valuable information from images. Some layers detect edges while other layers can detect parts of objects or complete objects such as a face or other complex shapes [106]. These layers include the convolutional layer, pooling layer, and the fully connected layer [25] which can be seen in figure 3.2. CNNs take advantage of the fact that they are distinctively designed for image inputs [105]. The different CNN architectures will be described in the subsections below.

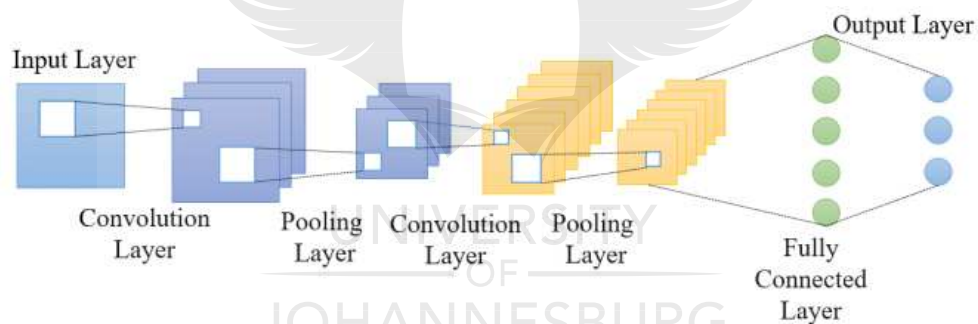


Figure 3.2: A basic CNN architecture [107]

#### LeNet-5 Architecture

In the year 1998, LeCun et al. introduced the first CNN, LeNet-5. [106]. Because of the restrictions at the time, this network is made up of only a few layers and filters. The architecture contains two convolution layers, two average pooling layers, two fully connected layers, and an output layer with a Gaussian connection, as illustrated in figure 3.3. The architecture takes a  $32 \times 32$  grayscale image as input, which is processed by the first convolutional layer, which includes six feature maps and a  $5 \times 5$  kernel. The second layer is an average pooling layer or sub-sampling layer with a filter size of  $2 \times 2$ . The third layer is a second convolutional layer with 16 feature maps of size  $5 \times 5$ . The fourth layer is an average

pooling of size  $2 \times 2$  for each feature map. Finally, the last layer which is a fully connected layer 120 filters and  $5 \times 5$  kernel for each map [106].

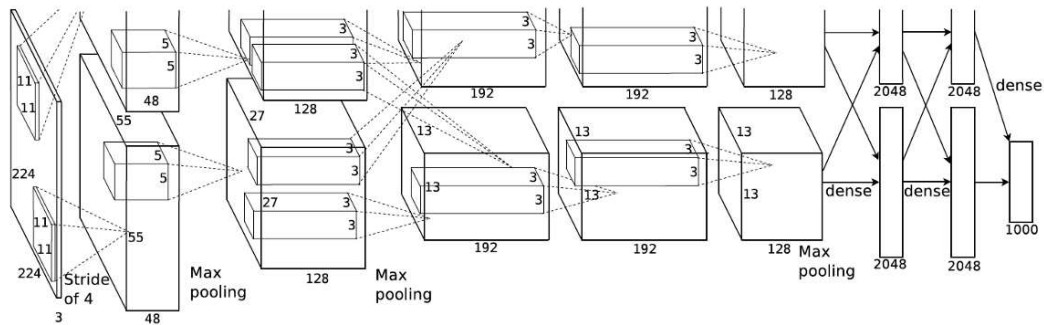


Figure 3.3: Visual representation of LeNet-5 architecture [106]

## AlexNet Architecture

The AlexNet architecture has a very similar architecture to LeNet-5 except that it is deeper, in the sense that it features convolutional layers stacked on top of each other [105]. The AlexNet architecture consists of five convolutional layers, three  $2 \times 2$  max-pooling layers and two fully connected layers [106]. Figure 3.4 illustrates the different layers within the architecture. The input for the architecture is  $224 \times 224$  image which passes through the first convolutional layer of where the filter size is  $11 \times 11 \times 3$  with a 4-pixel stride. The AlexNet architecture has a total of 60 million parameters. Therefore, it has previously suffered from overfitting. However, two methods have been used to reduce overfitting, which is data augmentation and dropout, which is a regularisation technique [105, 106].

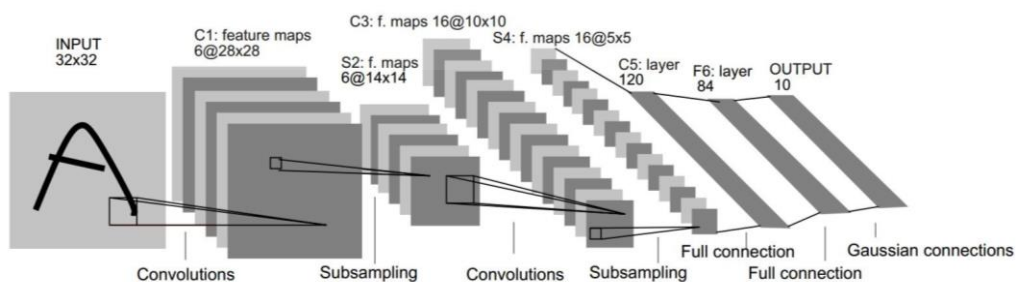


Figure 3.4: Visual representation of AlexNet architecture [105]

## VGGNet Architecture

The VGG network architecture was originally proposed by Simonyan and Zisserman [108] which secured first and second places in the ImageNet Challenge. Its main contribution was in showing that the depth of a network is a critical component to achieve a better classification accuracy in CNNs [108]. However, one major drawback which stems from the VGGNet architecture is that it is slow to train due to number of layers it contains. The architecture uses  $3 \times 3$  convolutional layers which are stacked on top of each other, thereby increasing its depth. Figure 3.5 illustrates the different layers within the VGG16 and VGG19 architectures. Notably, there are 13 convolutional layers, 5 max-pooling layers and 3 dense layers within the VGG16 architecture while the VGG19 architecture has 16 convolutional layers with 5 max-pooling layers and 3 fully connected layers [108].

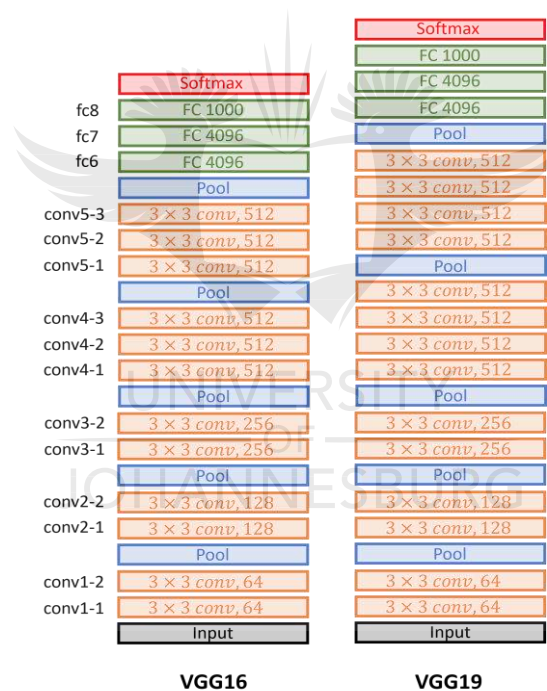


Figure 3.5: Visual representation of VGG16 and VGG19 architectures [108]

## ResNet Architecture

Residual network (ResNet), which is a modern CNN, was the winner of ILSVRC in 2015. This architecture has a total of 152 layers. ResNet is made up of residual blocks, as seen in figure 3.6, that are stacked on top of each other. There are two  $3 \times 3$  convolution layers in each residual block. ResNet uses batch normalization [75, 76] after each convolution

layer and has specific skip connections. Because deeper models are more difficult to tune, skip connections are used to take activation from one layer and feed it to another. This technique trains very deep networks and avoids the vanishing gradient problem [75, 76]. The good performance of ResNet on image recognition problems shows that deeper networks may be beneficial for many image recognition tasks [109].

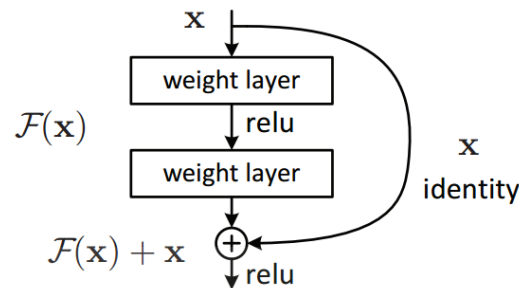


Figure 3.6: Residual Block [76]

Various experiments have proven that state-of-the-art deep learning architectures such as VGG and ResNet can perform reasonably well on challenging image recognition and object detection tasks [109]. Consequently, most of the famous object detection and segmentation architectures such as Single Shot Detection (SSD), R-CNN, Faster R-CNN and Mask R-CNN are built on the backbones based on architectures such as VGG and ResNet [109].

### 3.3.2 Region-Based Convolutional Neural Networks (R-CNNs)

To deal with the difficulty of object detection, Ross Girshick et al. [110] introduced an architecture called R-CNN in 2013. The R-CNN architecture extracts 2000 region proposals from an image using a selective search method. The CNN architecture is then provided with these 2000 region proposals for feature extraction [110]. Thereafter, the features are fed into an SVM model, which classifies the object in the region proposal. A region proposal network (RPN) takes an image as input and produces rectangular region proposals with a confidence score for each proposal. One disadvantage of the R-CNN approach is that it takes a long time to train the network because, on average, there are around 2000 candidate proposals. Additionally, the CNN architecture, SVM model and bounding box regressor are trained separately and therefore, this makes it slow to implement [111]. To overcome these challenges, a faster object detection technique, was developed to address these issues known as Faster R-CNN. The method is comparable to

the R-CNN technique, except that the convolution operation is performed just once per image, and a feature map is generated as a result [111]. Region proposals, on the other hand, become bottlenecks in the Fast R-CNN algorithm, lowering its performance. Faster R-CNNs, one of the latest occurrences of the method, performs object localisation at a high speed. This is due to the removal of selective search algorithm, and instead uses a regional proposal network. The latest iteration of the method, called Mask R-CNN, extends Faster R-CNN by using anchor boxes to detect multiple objects, objects of different sizes as well as overlapping objects in an image [114]. Mask R-CNN is also simpler to train and implement.

Object detection techniques in the past used regions to locate the object within the image. The network examines a section of the image that has a high chance of containing the object. Object detection methods such as YOLO (You Only Live Once) and Single Shot Detection (SSD) do not use region proposals. Instead, a single convolutional network predicts the bounding boxes. Although these approaches are faster, one limitation is that small objects are particularly difficult to localise within the image [112, 113].

## 3.4 Similar Work

Lip print recognition has not been extensively researched and therefore this biometric modality has not been widely used. The most common feature used in this context is fingerprint due to its accuracy [27]. Furthermore, fingerprint recognition has a variety of scanners in the market and is easy to install. Lip print sensors are not yet developed and research in this area is still in its emerging stage. However, there has been an increase in research in this area. This chapter will survey previous work in lip print recognition. Therefore, the goal is to describe the related research areas. By clearly describing the previous work current trends and limitations that exist within the realm of lip print recognition will be revealed.

### 3.4.1 Traditional Machine Learning Research

Before discussing previous work that has been conducted in this domain thus far it is first necessary to highlight how the data sets were created. Thus far, lip prints have been collected using traditional approaches such as fingerprinting roller, magnetic powder with a

magna brush, lipstick and cellophane tape. These methods require the participant to either press their lips onto a surface such as paper or cellophane tape or press their lips against a suitable surface and then process the prints with a fingerprint developing powder or magnetic powder. The acquired lip prints are then digitised and used to effectively create lip print data sets for lip print recognition. Previous work on how these data sets were used to achieve lip print recognition will be discussed below.

Gomez et al [128] presented a lip identification system in 2002 based on lip morphology. They gathered 500 photos of 50 subjects' faces for their investigation. The shape of the lip was extracted using image transform in this study. The polar coordinates of the lip envelope were used to extract the first set of features, and samples of the lip envelope's height and width were used to extract the second set of features. The study achieved a recognition rate of 96.9% and an equal error rate (EER) of 1.5%. One disadvantage that can be noted is that the study only focused on the geometric feature of the lip and the lip patterns were completely ignored.

Choras [116] developed a lip print recognition system in 2007 that uses the extracted lip shape and colour information to identify a person. Colour characteristics for the masked lips were computed in this study. Thereafter, shape features of the binarized lip were merged with the colour features of the masked lip. A 76% accuracy was reported on the best selected features of the lip. Two years later, in 2009, Choras [97] devised a new method for lip print recognition. In this work, the lips were first segmented, binarized and normalised. Thereafter, geometrical parameters of the binarized lips were calculated. This method achieved a recognition rate of 82%.

In 2010, Smacki et al [28] proposed a method for recognising lip prints using hough transform based on section comparisons. Sections refer to the lines and patterns found on the lips. An algorithm was devised to compare these sections. It was found that the sections with a length greater than 30 pixels resulted in high error rates. The FAR achieved for identification was 30% while the FAR achieved for verification was 17%. One disadvantage that can be noted is that the hough transform was used to detect only straight lines. However, lip prints have much more complex patterns such as ellipses and ovals [13].

Wrobel et al [29] proposed using Bifurcation analysis for lip recognition in 2011. The steps in the proposed solution were: pre-processing, feature extraction, and identification. Linear

contrast stretching was used to pre-process the image. The black pixels were utilized to detect bifurcations for feature extraction. The resulting bifurcation matrices were then used to compare the lip print to other lip prints in the database. It was discovered that obtaining bifurcations from lip prints was a difficult operation. With a 23% error rate, the best outcome was reached. The author also states that some pixels imitated bifurcations that did not exist in the lip print. Therefore, it had a negative impact on identification.

Another solution proposed by Wrobel and Froelich [30] used fuzzy c-means clustering for lip recognition. First, a hough transform was used to extract features from the lip prints. Thereafter, fuzzy c-means clustering was used to cluster those features. The representatives of clusters were then used to compare the images instead of comparing all pairs of individual characteristic features. The best results were obtained with the number of clusters equal to 80. The overall accuracy achieved was 82,2%.

Smacki et al. [31] proposed a method for lip recognition using the Dynamic Time Warping (DTW) algorithm. The first step in this method was pre-processing the lip print images. The second stage was to extract features from the pre-processed images by creating horizontal, vertical, and oblique (45° and 135°) lip pattern image projections. The DTW algorithm was then used to compare the projections against each other. Lip prints from 30 people were used to test the process. The findings were discovered to be suitable for use in forensic laboratories. The author also claims that modifying the DTW algorithm can improve accuracy.

Porwik et al. [32] proposed a method of comparing and recognising lips using DTW and the Copeland vote counting approach. The lip print was first rendered on a durable surface using a fingerprint powder and then converted into a digital image. It was then normalised, and patterns were extracted from the image paired with the Copeland voting approach to refine the accuracy.

Another solution proposed by Bandyopadhyay et al. [33] first used a Gaussian filter to pre-process the acquired image. Sobel edge detection and Canny edge detection was used to detect vertical and horizontal groove patterns in the lips for feature extraction. It was found that using Sobel and Canny edge detection for the extracting lip patterns yielded satisfactory results. The solution, however, did not match the lip print to other lip prints because it was not included in the scope of the study.

Bakshi et al [34] proposed a solution using local feature extraction methods (SIFT and SURF). The paper proposes that grayscale lip images comprise of local features. This claim was then experimentally proven by extracting and matching SIFT and SURF features from 23 grayscale images of 10 different subjects. The results indicate the viability of the approach with an accuracy of 93.99% and 94.09% respectively. Although the results received were satisfactory, the proposed solution used both SIFT and SURF for feature extraction and matching.

A new approach proposed by Wrobel et al [35] uses lip print furrow-based patterns. In this method a lip pattern is created for each person. Several lip patterns are taken from each individual and then appropriately prepared. The prints are then divided into upper and lower lips and the furrows are then made visible. The lip furrows are then parametrized. The lip print pattern comprises of furrows and are stored in the database. This approach achieved an accuracy of 92.73%.

In 2017, Wrobel et al [36] proposed a lip recognition system using Probabilistic Neural Networks (PNN). In the first step, the Region of Interest (ROI) of the lips is detected using an ensemble of HOG-based and Haar-based classifiers. Thereafter, the contour of the lip is extracted which are then passed as input data of the neural network. Three different databases were used in their experiment; Multi-PIE Face Database, PUT database and a local database consisting of 50 images from 5 subjects. Feature extraction was based on lip contours and facial landmarks. Classification accuracies of 86.95%, 87.14% and 87.26% were obtained, respectively. One issue pointed out in this system is that it extracts the contour of the lips and completely ignores the surface of the lips which are more discriminative than lip contours.

More recently in 2021, Sandhya et al [37] compared machine learning algorithms for lip-based identification. The main goal of this project was to create a reliable computer system that could recognize people based on their lip prints and work across a variety of datasets. To extract characteristics from the segmented upper and lower lip, local binary patterns are applied. As part of this process, shape-related characteristics are also extracted. Following that, other classifiers such as SVM and K-NN, Ensemble classifiers and ANN are used for classification. This approach achieved accuracies of 81.84%, 80%, 97% and 85.81% for



each of the classifiers respectively on a local dataset with captured lip prints. These lip prints were acquired with either a magna brush or magnetic powder.

### 3.4.2 Deep Learning Research

Recognition based on lip prints is a relatively new biometric modality. Notably, work that has been conducted in the realm of lip print recognition is still in its emerging stage and only traditional machine learning approaches have been used to achieve lip print recognition. Therefore, as confirmed by Shaheed et al. [115], currently there are no deep learning implementations for lip print biometrics.

## 3.5 Literature Trends and Gaps

The various studies in lip print recognition have generally been inspired by its uniqueness and its ability to identify individuals. It is also worth noting that the area of lip print recognition is still in its emerging stage with potential for future research. This subsection will discuss the various trends and gaps in literature.

When reviewing the literature on lip print recognition, it is evident that there are a significant number of traditional machine learning methods that has been investigated even though biometric recognition has shifted from traditional machine learning to deep learning. However, this is mainly due to the lack of investigation and research which is still at a very early stage in this field. A significant trend that can be noted from the existing work is that all the proposed solutions used pattern recognition steps which are pre-processing, feature extraction and classification to identify individuals based on their lip prints. Another significant and interesting trend that can be noted is that the previous studies acquired lip prints or employed datasets that acquired lip prints using traditional methods such as magnetic powder or lipstick. Therefore, based on recent trends there is an opportunity to expand on the previous work done by adopting state-of-the-art methods such as deep learning approaches.

While there are significant trends when reviewing the literature, certain gaps are also apparent. The major gap in the field of lip print recognition is the lack of research undertaken using deep learning methods. Deep learning methods have been successful in achieving

state-of-art results in biometric recognition for fingerprint, face, iris, ear, palmprint and gait since the paradigm shift in 2012 [38]. Therefore, although, lip print recognition is still in its early stage, deep learning methods for lip print recognition can produce promising results. Another gap that can be noted is the lack of datasets that can be adopted for lip print recognition. Currently the datasets that are available for lip print recognition have been acquired using traditional methods such as fingerprinting roller, magnetic powder, and lipstick. Furthermore, not many freely accessible databases of lip-print images are made available for research purposes.

Generally, any physiological or behavioural characteristic can be a biometric modality. In this case lip print biometrics is proposed due to the environmental challenges encountered with facial recognition. The primary issue of facial recognition stems from environmental factors such as lighting, background, head pose and facial hair [39]. Unfortunately, facial occlusions such as glasses, cosmetics or scarves are another critical factor that affects the performance of facial recognition [36]. Broadly speaking, one way of approaching this problem is to find features of the face that are not affected by these conditions such as the eyes, nose, or mouth. Therefore, despite the shortcoming that exist in current work, lip print recognition has potential to contribute to biometric identification by providing an alternative opportunity to identify an individual in the presence of many occlusions.

### 3.6 Conclusion

This chapter explored the various methods that would be appropriate for the pipelines of the particular study. Firstly, the traditional machine learning methods that could possibly be employed for object detection were discussed. These methods include the Haar cascade classifier and the One Millisecond Face Alignment algorithm. The Haar cascade classifier is good at detecting edges and lines and therefore, this makes it especially effective in face detection. The One Millisecond Face Alignment algorithm uses pixel intensities to estimate landmark positions on the face. This algorithm has the ability to make high quality predictions. Secondly, traditional methods that could be used to achieve pre-processing were highlighted which are grayscaling, histogram equalisation, blurring and edge enhancement. These methods, when applied to the region of interest can illuminate the patterns on the lips, thus making it suitable for feature extraction. The feature extraction

methods which were discussed included hough transform, SIFT and SURF, edge detection, local binary patterns and gabor wavelets. Thereafter, feature representation methods such as BoVW and PCA were highlighted. These methods select a subset of crucial features which can be used to train the model for classification. The possible classification methods which can be employed are SVM, K-NN and random forest.

The next set of methods which were discussed involved deep learning methods. The paradigm of deep learning employ representation learning does not require separate pre-processing and feature extraction steps that are necessary for traditional machine learning. Deep learning methods automatically pre-process and extract features from the image. The appropriate methods which were discussed are CNN architectures such as VGGNet, AlexNet, LeNet-5 and ResNet. Thereafter, object detection and segmentation methods were highlighted such as R-CNNs, Mask R-CNN, YOLO and SSD. Deep learning methods have been successful in achieving state-of-art results in biometric recognition for modalities such as fingerprint, face, and eyes. Therefore, lip print recognition also has the potential for employing deep learning methods.

Thereafter, recent trends and gaps missing in literature were identified. Based on existing work, it is evident that much more work has been done using traditional machine learning methods. Another trend that can be noted is that the current work employed datasets that acquired lip prints using traditional methods such as magnetic powder or lipstick or created a local dataset following the same methods. The major gap revealed for lip print recognition is lack of deep learning methods investigated. Another gap that was identified is the lack of datasets that can be used for lip print recognition.

Now that different traditional and deep learning methods have been discussed, and an investigation on similar work within the domain of lip print recognition has been conducted, it is necessary to uncover methodologies and techniques that can be used to further develop the study and address its research problem. The next chapter will discuss the research methodology for this study.

# Chapter 4      Research Methodology

## 4.1 Introduction

Research is defined as the process of collecting and analysing data for a specific purpose. A research methodology is a systematic plan used to yield data on a particular research problem and address the objectives of the study [40]. As a result, the research methodology used in a study can be seen as a tool that answers the research problem and addresses its objectives. In the previous chapter, the literature review of the study was discussed and what remains is how the solution of the research problem will be pursued. A literature review assists in uncovering methodologies that helps design the current study. The purpose of this chapter is, therefore, to set out the research design, the research paradigm and research methods followed by the study to effectively address the research problem and its objectives.

The sections in this chapter present the different research methodology approaches. It provides a discussion concerning the approach that was undertaken for the study as well as a justification for the approach used. In section 4.2 the research design is discussed, followed by the research paradigm in section 4.3. Section 4.4 discusses the research methods which includes a literature review, model, and prototype. The research methodology chosen for the study is then justified in section 4.5. The population and data sampling process are discussed in section 4.6. The different performance metrics that the current will use is unpacked in section 4.7. The validity and reliability of the study is laid out in section 4.8. Ethical and legal considerations (4.9) and the potential risks (4.10) are then discussed. Lastly the chapter is concluded in section 4.11.

## 4.2 Research Design

A research design is needed to ensure a smooth sailing of the different research steps involved, thereby making the research coherent. The three most common designs to conducting research are qualitative, quantitative, and mixed methods [40].

A qualitative research approach aims at discovering the motives of human behaviour. It involves collecting non-numerical data to uncover the deeper meaning of human concepts

and their experiences [41]. The approach is inductive rather than deductive, meaning that a theory is derived based on the collected data. Data collection is carried out in several ways such as interviews, surveys, observations and focus groups. The data is not converted to a numerical form or statistically analysed. Therefore, a qualitative research approach focuses mainly on the significance of human behaviour and their experiences, including beliefs and emotions [41].

A quantitative research approach is different from the approach discussed above. It is based on collecting and transforming data into a numerical form so that results can be made, and conclusions can be drawn [41]. The data usually needs to be processed before it can be analysed. This approach usually has one or more hypotheses which are then tested by implementing certain methods to collect results. The results drawn from this approach can be generalised to broader populations. This approach follows deductive reasoning, meaning that it starts out with a hypothesis and examines all possibilities to reach a conclusion [41].

A mixed methods research approach or a pragmatic approach is a combination of both qualitative and quantitative research. The purpose of this approach is to build upon both qualitative and quantitative approaches to understand the research problem more clearly. Depending on what measures have been used to collect data, it is analysed in an appropriate manner [42].

Based on the analysis of the research approaches discussed, a quantitative research approach is adopted for this study. As outlined above, a quantitative research approach has a hypothesis that is formulated as a model, is tested by implementing certain methods to collect measurable outcomes or results. Therefore, based on the model, a prototype will be implemented to gather results. The hypothesis outlines that deep learning-based methods can be employed to achieve lip print identification in an effective manner will then be tested, which can either be accepted or rejected based on the results and a conclusion will be drawn from it.

### 4.3 Research Paradigm

A research paradigm is an approach or perspective to research that is used to understand and address research problems [43]. A large number of research paradigms have been

proposed throughout the years. Therefore, the four main research paradigms will be discussed, namely positivism, interpretivism, critical theory and pragmatism.

The positivist paradigm is known as the scientific method of investigation. This paradigm gives validity and objectivity to research, and it is based on precise methods. Research in this paradigm relies heavily on deductive logic, formulating hypotheses, testing hypotheses, and implementing certain methods to derive conclusions [44].

The interpretivist or constructivist paradigm is subjective rather than objective. This paradigm is used to understand human behaviour and their experiences using non-scientific methods. The point of research is to understand the viewpoint of the respondent being observed. It sacrifices reliability for validity [44].

Unlike the positivist paradigm, which is guided by the principles of objectivity, critical theory poses that knowledge can never be truly objective. Furthermore, this paradigm assumes that a scientific investigation should be conducted with the goal of social change in mind [45].

Pragmatism assumes that there are many different ways or methods to solve a research problem and not just one. This is achieved by an integration of multiple research paradigms encompassing positivism, interpretivism and critical theory. Through this integration, pragmatism assumes that a better understanding of the research can be gained [44].

Since quantitative research is generally associated with the positivist paradigm and based on the discussion of the research paradigms outlined above, this study will adopt the positivist paradigm. Since objectivity is the main principle, it is important to maintain an objective stance throughout the study in order to analyse the results accurately without any bias and to prove the hypothesis to be either true or false. This will determine to which extent the results can be generalised.

## 4.4 Research Methods

After identifying the research problem and research methodologies that will be used, appropriate research methods are needed to approach the problem in order to give direction to the study. The appropriate research methods adopted for this study will include: a secondary method which is the literature review and a primary method which is the model.

The prototype that is based on the model will also be investigated. Each of these methods will be discussed further in the subsections that follow.

#### 4.4.1 Literature Review

Conducting research and relating it to existing work is the core foundation of any academic research activity. A literature review as described by [46] is a way of collecting and incorporating previous research to uncover areas in which more research is needed, which is a key component for creating frameworks and models. A well-conducted literature review provides an overview of methods or techniques best suited for the research. This is why a literature review as a research method is fundamental. For this study, all the existing work in the area relating to lip print recognition has been investigated and discussed in chapters 2 and 3. The existing work indicates what has been attempted in the domain thus far and which methods and algorithms have been used. These methods will play a crucial role in model formation.

In order to conduct an effective review, specific areas that relate to this study have been addressed as well. These areas include access control (section 2.2), biometrics (section 2.3), and computer vision (section 2.3). Access control is explored to gain insight on different access control environments as well as the different types of authenticators applicable in this domain. The next area that was addressed is biometrics. Biometrics is a key area that this study will make use of. The different biometric attributes have been investigated as well as which biometric traits are currently used in access control. Additionally, the study also addressed how lips can potentially be a biometric trait and how these two domains (access control and biometrics) incorporate. The next area that this study addressed is computer vision. Computer vision is a key component in this study since its methods will be investigated and potentially be used for lip print recognition.

#### 4.4.2 Model

A model can be considered as a type of Information Technology (IT) artefact. An artefact, which is an abstract representation, can be converted into material existence such as an artificial object or process [47]. In simple terms, an artefact is anything that is created in order to develop a prototype. This includes things such as models, diagrams, or

mathematical formulas. Artefacts are designed and created based on literature and innovation. Designing, creating, and evaluating artefacts are typical IT research activities [48].

For this study, the model will provide a representation of how the research problem will be addressed. Based on the literature study, the model will give an outline of the different methods and algorithms that can potentially be used to address the research problem. It will consist of different stages, from localisation to classification. Research will be conducted on each stage to determine the methods appropriate for that particular stage. Essentially, the model is a representation of investigated methods that can potentially be used to implement the prototype along with its pipelines. The model is, therefore, a primary research method and a primary outcome for the study.

#### 4.4.3 Prototype

The aim of a prototype is to demonstrate the feasibility of a model. A prototype is an implementation of methods that are selected from the designed model. The methods for implementation may vary depending on the model. The implementation of the prototype will yield measurable results from different pipelines which will be compared and analysed. It will be tested against a set of benchmarks to determine whether it accomplishes its task. The prototype is necessary since it will determine whether the methods chosen for this study can be used for lip print recognition. Ultimately, the prototype will determine if the model is feasible.

### 4.5 Justification

As discussed in sections 4.3 and 4.4 the research design adopted for this study is a positivist-based quantitative approach. Since this approach is deductive, deductive reasoning will be used to formulate a hypothesis. The hypothesis will then be tested on an implementation of the model which will either be accepted or rejected. Because this approach makes use of scientific methods for data collection and analysis, generalisation is made possible since the research findings are not a mere coincidence [49]. It can, therefore, potentially reflect the broader population. A positivist-based quantitative approach relies on



hypotheses testing which follows specific guidelines and objectives and the research is conducted in a reproducible fashion. Therefore, it can be repeated at any time and still get the same results [49]. Backed by the reasons discussed above and research found in literature, the positivist-based quantitative approach is adopted and is a good fit for the study.

As discussed in section 4.4 there are 3 research methods used for the study: literature review model and prototype. The literature review will give an insight on what has been attempted in the particular domain thereby justifying the need for the study. Once analysed, the literature review can be used to formulate a hypothesis. The hypothesis is a key component in the model. The model will be used to test the hypothesis by investigating the extent to which these two factors relate within the given hypothesis. The model, which is an abstract representation of methods that can be used to address the research problem, will put the hypothesis into action, thus either accepting or rejecting it. The prototype will determine feasibility and empirical evidence of the methods applied to this use case.

## 4.6 Population and Data Sampling

In order to evaluate the model, a population and sample set is needed in order to attain repeatable results at any given time. The population and sample set chosen for the study will be discussed further below.

The population of a study refers to the broader group of people to whom the researcher intends to generalise the results of the study [41]. Importantly, the population should only include the people to whom the results will apply. Therefore, in the current study, the population chosen are users that require permission in physical access control environments.

The sample refers to the selected participants chosen for participation in a study while the sample frame are the participants who could possibly participate in the study. Therefore, based on the above definitions the sample frame for the study includes all individuals that can present the unique biometric trait, called lip prints. The sample set chosen for the study include high-resolution samples from the Chicago Face Database (CFD) [50]. The CFD is a publicly available secondary data set. Secondary data is the use of an existing piece of research data to address a research problem different from the original work [51]. The CFD

was developed at the University of Chicago by Debbie S. Ma, Joshua Correll and Bernd Wittenbrink. It is intended for use in scientific research. The CFD provides high-resolution ( $2444 \times 1718$ ) images of 597 male and female targets of different ethnicities. Each target is represented with a neutral expression. The CFD will be used to validate the model by calculating performance metrics that can be compared against the different implemented pipelines.

## 4.7 Data Analysis

Once the hypothesis has been assumed and realised, a benchmark will then be performed to assess its potential in lip print recognition. The benchmark will determine how well the implemented model performs when it is tested. Various performance metrics will be used to properly assess the model. These metrics include accuracy, precision, recall, f1 score, equal error rate (ERR) receiver operating characteristic (ROC) curve, precision-recall curve, accuracy and loss curves, intersection over union (IoU) and mean average precision (mAP). These performance metrics will be unpacked further in Chapter 5. The benchmark also allows for the solution to be compared against current systems. Therefore, the results will then be compared against other approaches in literature.

## 4.8 Reliability and Validity

Research conducted via the use of scientific methods yield observations and data. However, for this data to be of any use, the study must have certain properties such as validity and reliability. Validity and reliability are two important concepts used to approve and evaluate the quality of a study. Each of these concepts will be discussed further in the subsections that follow.

### 4.8.1 Reliability

Reliability of a research is defined as the consistency and reproducibility of measurement over time [52]. The research is said to be reliable if it repeatedly produces similar results

under the same conditions. The better the reliability, the more accurate the results, which enables for more correct decisions to be made in the study.

Reliability is important to any research design and considering the use of secondary data, as is the case with the current study, it is important to determine to which extent it relates to the research problem or how reliable it is to use. As discussed in section 4.6, the CFD [50] will be used for the current study. The goal of the CFD was to determine whether the participants used were appropriate for research purposes such as facial recognition. Researchers within the field of social psychology were invited to view and rate the CFD participants. The reliability analysis revealed high consistency among expert ratings. Therefore, this makes the CFD reliable to use.

Due to the large external factors that influence the credibility of a research, many research studies can be unreliable [53]. Therefore, to confirm that the literature review used for the study is reliable, the study will use objective sources as much as possible where the author of these sources has little or no commercial interest. The study will specifically pay attention to this when selecting literature to ensure its reliability. Many objective academic papers in the areas relating to this study will be investigated. The possible keywords that will be used during the research process will include: "access control", "identification", "biometrics", "lip print recognition", "computer vision" and "pattern recognition",

It is important reduce the influence of external factors that might create variations in the result. Therefore, in the experimental setup, the data will be tested under the same conditions. This means that each image in the data set will go through the same computer vision methods involved in the different stages of the pipelines from object detection to classification. Consequently, this will ensure that the prototype is reliable.

Other ways of improving reliability of the current study will be test it to through repetition and through the design experiment. Re-tests will be conducted to determine how similar the results are for each pipeline implementation. Methods needed for the designed model will be thoroughly investigated using literature works. Thereafter, during the prototype phase implementing a method that will improve the overall accuracy of the pipeline will be selected to improve the reliability of the study.

Lastly, when analysing the results, some researchers naturally look for results that confirm their hypothesis [54], consequently creating analysis bias. Thus, for this study, the author

will discuss the results and the associated parameters with their supervisor to ensure that reliability is preserved at every stage of the process.

#### 4.8.2 Validity

Validity of a research is defined as how accurately a method measures what it is intended to measure [52]. In simple terms, it determines if the results support the objectives of the study. It is seen as a compulsory requirement for all types of studies. There are different measures to ensure validity of study such as an appropriate methodology adopted for the study, suitable sample selected, or appropriate secondary data used.

As discussed above (in section 4.3 and 4.4), the study makes use of the positivist-based quantitative approach. This approach is appropriate for the current study since a hypothesis based on current literature will be formulated. The hypothesis will then be tested based on the implementation of the designed model. It will then be tested against a set of benchmarks to determine whether it accomplishes its task. Throughout this entire process it will be important to maintain an objective stance because it will determine the extent to which the results can be generalized which is crucial for identification in access control environments. Therefore, this approach is a good fit for the current study, and it is valid.

In an ideal situation, the entire population should be studied in order to reach a conclusion. However, this is almost impossible and therefore, a sample representation of a population is used. It is important to have a sufficient sample size to achieve reliable results. An insufficient sample size is most likely to produce more false negatives [55]. However, a very large sample size can be difficult to manage, and it can be a hindrance if the result can be found accurately from a smaller sample [55]. Sekaran [56] determines that a sample size greater than 30 and less than 500 is appropriate for most studies. Therefore, for the current study a sample size of 500 will be used mainly due to less computational power. The data set includes self-identified Asian, Black, Latino and White ethnicities of different ages. Therefore, there is very little bias towards gender, age, and race. Due to the diverse nature of the data set, the results obtained by the study can potentially reflect the broader population. The data set consists of high-resolution images of each individual. A high-resolution image is needed since the pattern of grooves have to be visible on the lips. Therefore, this makes the data set valid to use.

Other ways to improve validity include clearly defining the aim and objectives of the study, comparing the results with the hypothesis of the study, and comparing the results with existing work in literature [57]. The aim and objectives of the study have been clearly defined in Chapter 1. Each of these objectives will be operationalised progressing further into the study. The study will also determine whether the objectives have been met following the results and interpretation. The results and the hypothesis will then be evaluated, effectively determining whether the hypothesis has been accepted or rejected. To further increase the validity of the study, the results will also be compared against other approaches in literature.

## 4.9 Ethical and Legal Considerations

Ethical and legal considerations form a crucial element in a study. It should be considered in every aspect of the study which includes the participants of the study, data collection methods, and utilisation of the data. Following ethical guidelines ensures the study's validity and promotes its contribution to scientific study. The ethical and legal considerations of the secondary data as well as ethical considerations concerning biometrics in general will be discussed further in the subsections below.

### 4.9.1 Secondary Data

Given the importance of ethics in conducting research, particularly within the domain of access control and biometrics, it is important that certain ethical and legal considerations be considered throughout the study. It is usually believed that the use of secondary data alleviates the researcher from the burden of taking ethics into consideration [58]. However, the entire research process involves ethical and legal considerations, whether the data is primary or secondary. Although the CFD is highly ethical, it is an open access database. Therefore, the study must meet certain ethical and legal conditions itself to ensure that is conducted in an appropriate manner.

Firstly, the data will be kept secure from any destruction or damage. The data will not be used for any reason other than the one specified in the study. It is important to note that the data has been appropriately anonymised, and the researcher cannot reverse it, thus greatly reducing the risk of harm. Secondly, no harm will be done to the participants through

the reuse of their images in any way. The images of the participants will not be used for unethical or illegal purposes and the anonymity of the participants will be maintained throughout the study. One of the terms of use of the CFD is that the database must be used for research purposes only. Thus, the current study will abide by this term, using the CFD for non-commercial research purposes. Lastly, the use of this data will not bring about any damage or distress to the original authors, participants, or the public in any way.

#### 4.9.2 Social and Legal Issues

A biometric system recognises the physiological and behavioural characteristics of individuals. Therefore, it is mandatory for those who design and deploy these systems to consider the social, and legal contexts of these systems. Failing to address the cultural and social contexts of these systems greatly affects the efficacy of these systems that shape the way individuals interact with them.

The main social issue surrounding biometrics is the irreversible link between the biometric trait and the personal information about a person [59]. Due to this reason individuals question about whether to engage with system or not. For instance, some individuals may choose not to place their finger on a fingerprint scanner for fear of contracting a disease such as the novel Coronavirus. Similarly, some individuals may avoid having their photograph from being captured for lip print recognition because of concerns over how these images will be used. Others may avoid this due to religious reasons such as coverings of the face. In these cases, the performance of the system is compromised. Equivalently, legal considerations such as reliability and privacy are also crucial factors in achieving the effectiveness of a biometric system. In the long run, the biometric system should accurately identify individuals while also protecting their data. By doing so, opinions about different techniques can change. Society then becomes familiar with the technique and generally view it as trustworthy such as fingerprint recognition [59].

Although biometric systems are beneficial, they raise social, and legal concerns. These concerns greatly affect the system's performance, its acceptance by society or the decision on whether to use it in the first place. Therefore, these concerns should be considered when designing, developing, and deploying biometric recognition systems.

## 4.10 Risks

Risk is defined as the potential for harm. It is an indication of what could cause harm or what could be harmed. In a human subject research, risk is categorised in 2 categories namely minimal risk and greater than minimal risk [60]. A study is considered minimal risk when the risks of the study are not greater than those experienced in normal daily life. Thus, the current study is categorised as minimal risk since no harm is encountered by the participants or the environment in any way. However, the possible risks that the study may encounter are:

1. The methods chosen from the design model for the prototype can be irrelevant
2. Challenges in implementation can arise
3. Human error resulting in misinterpretation of results

To mitigate these risks, the study will thoroughly investigate the methods appropriate for the current study using existing literature. The prototype will be tested extensively to ensure that the implementation is correct. Although human error is possible, the results of the study will be thoroughly analysed to minimise any shortcomings.

## 4.11 Conclusion

The chapter begins by outlining the research design and research paradigm that the current study will use to address its research problem and objectives. It was determined that the study will adopt a positivist-based quantitative approach. Thereafter, the research methods that will be used for the study were discussed. The research methods include a primary method; model, secondary method; literature review and a prototype based on the designed model. These approaches were then justified. The population to whom the study is relevant, the appropriate data sample set, and sample frame chosen for the study were then outlined. It was determined that the appropriate population for the current study are users who require permission in physical access control environments. The sample set chosen are individuals with lip prints and the sample set used will be the CFD.

Thereafter, the different performance metrics were briefly outlined which will be used to assess the implementation of the model were highlighted. These include accuracy, recall, precision, f1 score, EER, ROC curve, loss curve, accuracy curve and IoU. The reliability of

the research as well as its validity was discussed. Different approaches such as re-tests, experiment design, appropriate data sample set, and appropriate methodology adopted will be used to increase the reliability and validity of the study. Thereafter, the ethical and legal issues arising from this study were determined. Lastly, the potential risks were identified.

Now that the research methodology; a tool that addresses the research objectives of the study has been established, the next part of the study, methods, can be considered. The next chapter will consider the model along with the possible computer vision methods that can be used for lip print recognition, the prototype implementation with its chosen methods and its benchmark including functional and non-functional requirements.





PART II  
MODEL AND BENCHMARK



UNIVERSITY  
OF  
JOHANNESBURG

# Chapter 5      Methods

## 5.1      Introduction

Research can be purely theoretical, or it can use models as the object of study. In this case, the production of a model is a core aspect in the research process. A model, as described in section 4.4.2 can be considered as a type of IT artefact or experiments which is essentially an abstract representation that can be converted into material existence [47]. However, the blueprint alone, without implementation, is not enough to address the research problem. Primarily, a prototype is needed to determine the feasibility of the designed model and provide empirical evidence of the methods applied. Therefore, the purpose of this chapter, is to conceptualise a model based on existing literature, specify a prototype system in order to address the research problem and determine the relevant benchmark that can be used to evaluate the performance of the prototype.

The hypothesis of the current study postulates that deep learning-based methods can be employed to achieve lip print identification in an effective manner. Therefore, the model will put the hypothesis into action and the prototype will determine if the methods chosen can potentially be used for lip print recognition. Ultimately, the prototype will be tested against a set of benchmarks to determine whether it accomplishes its task.

The chapter commences by highlighting the overall structure of the lip print recognition model, followed by the model's main components in section 5.2. Each of these components will be outlined further. The current study will make use of these components to implement three pipelines. Section 5.3 discusses, in-depth, the implementation of each pipeline based on the designed model. Thereafter, the benchmark that will be used to assess the performance of the prototype will be outlined in section 5.4. The benchmark will discuss both functional and non-functional requirements. Lastly, the chapter will conclude with section 5.5.

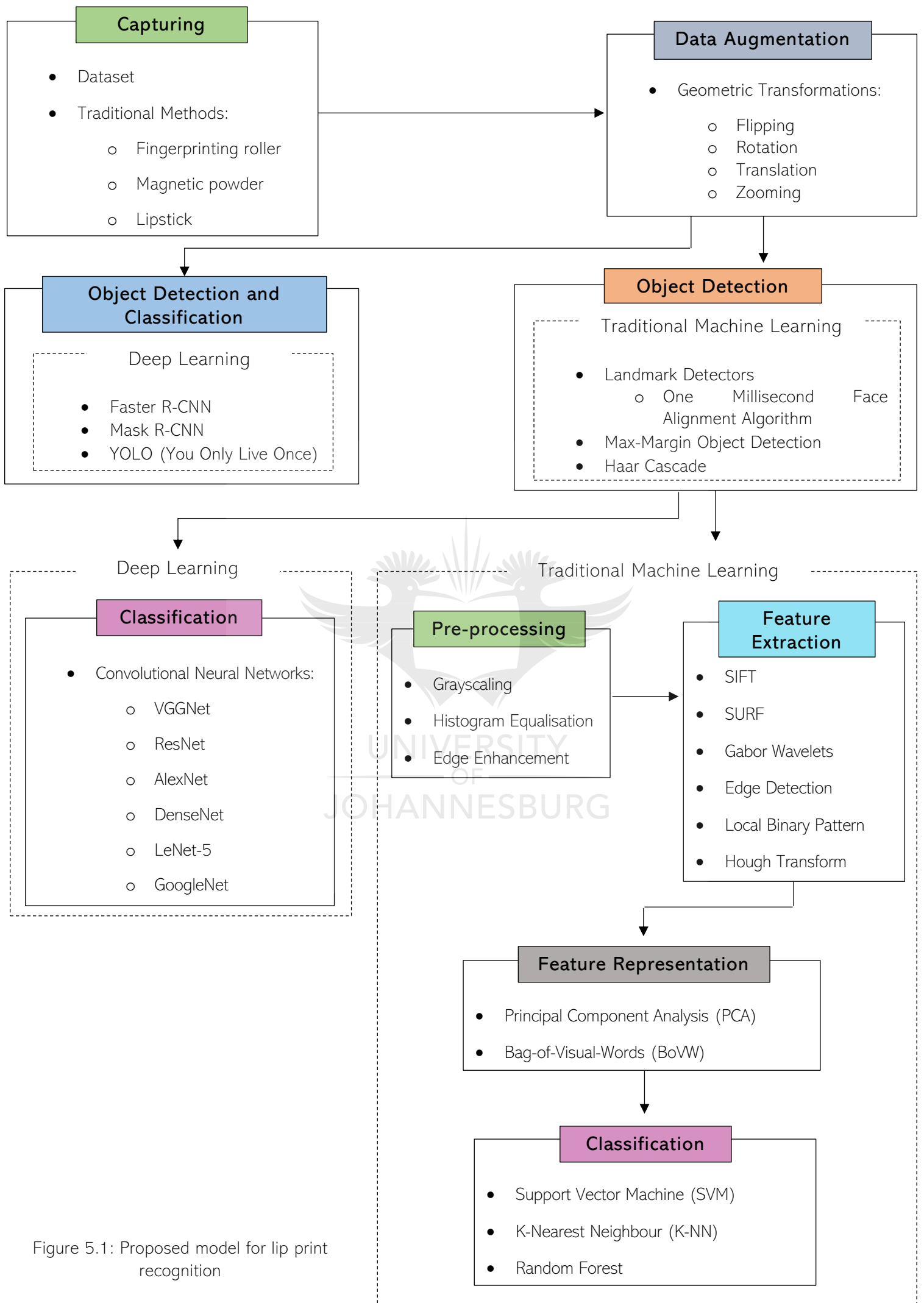


Figure 5.1: Proposed model for lip print recognition

## 5.2 Model

The model for lip print recognition is presented visually in figure 5.1. As depicted in the model, there are a number of structural components involved in the construction of the model which consists of capturing, data augmentation, object detection, pre-processing, feature extraction and classification. These components will be discussed further to understand how lip print recognition can be achieved.

The first structural component in the model as presented in figure 5.1 is the capturing component. This component is concerned with the source of the data for the model. There are two sources which can be adopted for the model; datasets and using traditional methods to acquire lip prints which is essentially a live capture. Based on existing work many of the datasets were created using traditional methods such as fingerprint rolling, magnetic powder or lipstick [28, 29, 30, 31, 32, 33, 34, 37]. Participants were required to disclose their lips on a durable surface such as a cellophane tape or a piece of paper, which were then converted to a digital form. The next source, dataset, is a collection of visual data such as images captured by a sensor device or camera. For lip print recognition, the literature review has determined that high-resolution face datasets can be used because the grooves present on the surface of the lips are visible.

After capturing the dataset, the next component in the model is data augmentation. Data augmentation is a process in which the dataset is expanded by applying methods such as flipping, rotating, cropping, and translating [87]. This process has become a significant tool in achieving state-of-the-art results in modern machine learning pipelines [87, 89].

The next component in the model is object detection. The process of object detection deals with isolating a region in an image where the feature resides. In this case, the lip region is the area of interest where the features are present. Since, this study utilises high-resolution face dataset samples, object detection is a crucial component in the model. The appropriate object detection methods for this model include traditional machine learning approaches such as Haar cascades and Face landmark detectors and deep learning approaches such as Faster R-CNNs, Mask R-CNN and YOLO discussed in sections (3.2.1 and 3.3.2). After object detection, the focus is placed on the isolated region which ensures effective pre-

processing, feature extraction and classification to take place. These steps will be discussed further below using a traditional machine learning approach and deep learning approach.

### 5.2.1 Traditional Machine Learning Approach

The pre-processing component in the model is adopted to process the data in such a way that any unnecessary noise is suppressed, difference of illumination is reduced, and important features are enhanced. This is mainly because the data before pre-processing may contain noise and it is difficult to extract features from a noisy image [61], thus greatly affecting the rate of recognition. The relevant methods for pre-processing which have been unpacked in section (3.2.2) include grayscaling, histogram equalisation, edge enhancement and blurring.

The next component, feature extraction, plays a crucial role in a computer vision model. After pre-processing, feature extraction techniques are applied on the pre-processed data. The main goal of feature extraction is to obtain the most relevant information from the data [62]. The appropriate feature extraction methods for this model include SIFT, SURF, hough transform, edge detection, gabor wavelets and local binary patterns which were discussed in section (3.2.3). Feature extraction makes the task of classifying the pattern adequately. However, the quality of the features passed determines whether or not the classifier can make accurate predictions.

Feature representation is the next component in the traditional machine learning approach. After feature extraction, feature representation techniques are applied on the extracted data. This is done in order to reduce the number of irrelevant features that could potentially overpower crucial features for classification. The appropriate feature representation techniques for this model include a Bag of Visual Words approach and Principal Component Analysis.

The last component in the lip print recognition model is classification. The main purpose of classification is to take the data from the features extracted previously, train it on a machine learning classifier and accurately predict the target class for each case in the data. Typically, a classification component consists of a train and test set. The model learns from the data during training to make predictions on unseen data during testing. Classification requires

the task of machine learning classifiers such as SVM, K-NN or Random Forest discussed in section (3.2.5).

## 5.2.2 Deep Learning Approach

Traditional pre-processors and feature extractors can be replaced by convolutional neural networks (CNNs). This is mainly due to the fact the CNNs have the stronger ability to extract complex features that express the image in more detail and are therefore, much more efficient. The convolutional layers are the most crucial layers of the CNN model. These layers learn complex features by building on top of each other. The first layers detect edges, while the next layers combine them to form detect shapes and the following layers merge this information to infer the object in the image. Thereafter, classification can take place based on these features. The appropriate architectures for this model include CNN models such as VGGNet, ResNet and AlexNet discussed in section (3.3.1).

Now that the model has been discussed and established, the prototype system can be defined. The next section will provide details about the different pipelines or experiments that have been chosen for the prototype as well as justify the choices made for the methods used.

## 5.3 Prototype

In order to prove that the lip print recognition model is feasible, a concrete implementation of the model is needed i.e., the prototype. There are three pipelines: traditional machine learning-based pipeline, a deep learning-based pipeline, and a deep hybrid learning-based pipeline. Each of these will be discussed further in the subsections that follow.

### 5.3.1 Capturing and Data Augmentation

The first component in the model is the capturing component which uses a dataset as its source. The dataset adopted for the current study; the CFD, has been discussed and justified in section (4.5). The next component after capturing is data augmentation. For the current model, data augmentation is necessary due to the limited data. Therefore, in order to tackle

this problem, data augmentation is used to artificially increase the amount of data in the dataset. Augmenting the data beforehand rather than performing transformations in mini batches is known as offline augmentation [88]. This method is preferred for limited datasets as the size of the dataset is increased by a factor equal to the number of transformations performed. However, expanding a dataset with data augmentation is not only helpful for increasing the diversity of a training set. It is also significant in reducing overfitting and improving the generalisation of a model [88].

The specific data augmentation techniques that can be used for a dataset must be chosen carefully, and within the context of the training set and knowledge of the problem domain. Therefore, the methods used for data augmentation include the following modifications to the original image: horizontal flip, rotation and an increase and decrease in brightness. Geometric transformations are helpful for positional biases present in the dataset [89]. If positional biases are present, such as in a face dataset where every face is centred in the frame and where localisation is crucial, it would require the model to be tested on perfectly positioned lip images constantly. Therefore, geometric transformations are an excellent solution for positional biases [87, 89]. Furthermore, altering the brightness of the original image is important because in real-life scenarios dramatic image variations arise from changes in illumination. Therefore, it is important to take these factors into consideration because it will determine how well the prototype can recognise the lips under different conditions. Furthermore, these modifications will determine how well the chosen object detection algorithm performs in detecting the lips. Lastly, the dataset is split into its respective sets: 80% for training and 20%. This split will be used for all pipeline implementations.

### 5.3.2 Lip Detection

The next component in the model is object detection. Object detection is important because the lip region, where the crucial features reside, will be used further down the different pipeline stages. Initially the haar cascade method was employed for object detection. However, due to the large number of false positives a different approach was utilised. Because the lip or in this case the mouth is a facial component, detecting it is a subset of the shape prediction problem. Given an image, a shape predictor attempts to localise key

points of interest along the shape. In the context of the mouth facial landmark, the goal is to detect this facial structure on the face using shape prediction methods [63].

The object detection method chosen for the different pipelines is the facial landmark detector included in the dlib library known as One Millisecond Face Alignment with an Ensemble of Regression Trees. An advantage of this method is that it can detect facial landmarks in real-time as well with precise facial landmark predictions [21, 84]. This method uses an ensemble of regression trees to estimate landmark positions from pixel intensities. There are 68 (x, y)-coordinates that map to facial components on the face. The coordinates specified for the mouth are located in between (49-68) which can be seen in figure 3.1. However, before the lips can be isolated the face should be detected first. Therefore, the pre-trained (hog-based) face detector is used to detect the face in the image. The classic dlib Histogram of Gradients (HOG) face detector uses a vector of features with an SVM classifier to detect faces. After the face has been detected, the mouth region is detected using the coordinates specified above and a bounding box is created around the detected region.

The pipeline discussions in the next section assume that the capturing, data augmentation and object detection components remain the same for each pipeline implementation.

### 5.3.3 Traditional Machine Learning Pipeline

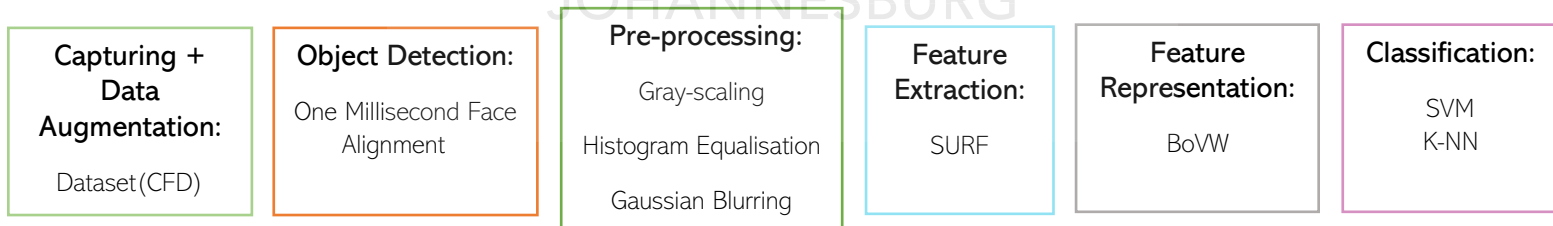


Figure 5.2: Traditional Machine Learning Pipeline

The traditional machine learning pipeline consists of 7 stages namely, capturing, data augmentation, localisation, pre-processing, feature extraction, feature representation and classification as visualised in figure 5.2. There are two classification variants for this pipeline: SVM and K-NN. The methods from stage four (pre-processing) will be described and justified below.



## Pre-processing

The next stage in the traditional pipeline after lip detection is pre-processing. As mentioned in section 3.2.2, pre-processing is important because it reduces noise effect and illuminates key features which is necessary for feature extraction and classification. The pre-processing methods adopted for this pipeline is grayscaling, histogram equalisation and gaussian blurring. Grayscale converts the image to a simpler colour space. This method is important because a grayscale image helps identify important edges or other prominent features. Although colour information can be useful to identify objects of known hue, in this case “colour” is seen as noise and therefore grayscaling the image is necessary. Thereafter, the histogram equalisation method enhances the contrast in the image, thus illuminating the pattern of grooves on the lips. After histogram equalisation is completed, smoothing is done on the images using Gaussian blurring to remove grooves which are small or highly insignificant [34]. The image is passed through a 5×5 Gaussian kernel. These methods allowed for consistent feature extraction to take place.

## Feature Extraction

The next stage in the pipeline is feature extraction. Feature extraction is important because it will obtain the most relevant information from the pre-processed data. The traditional pipeline will use local features for extraction rather than global features. Local features are used for feature extraction because global features have certain limitations such as sensitivity to noise, illumination variation and failure to detect important features in the image [64]. Using local features which encapsulate local information can potentially obtain better details of the image [64]. Lip images are rich in distinct local features. Therefore, the method chosen for this stage is the Speeded Up Robust Features (SURF) method. SURF is the most popular local feature extraction method which has proved to be promising due to its high performance [64, 65].

SURF adopts combinations of HAAR-like features of integrated pixel values, and it detects keypoints using the Hessian matrix. SURF was chosen as the feature extraction method because local, discriminative features can effectively be computed, and it can describe lip local properties effectively. The hessian threshold is set to 500 with a descriptor size of 128-dimensions. 500 is an optimal hessian value because a consistent number of features

are extracted from the lips by the algorithm for all the images. SURF-128 is used because it is more distinctive and much faster to compute [65]. An example of SURF features over the original image is illustrated in figure 5.3.

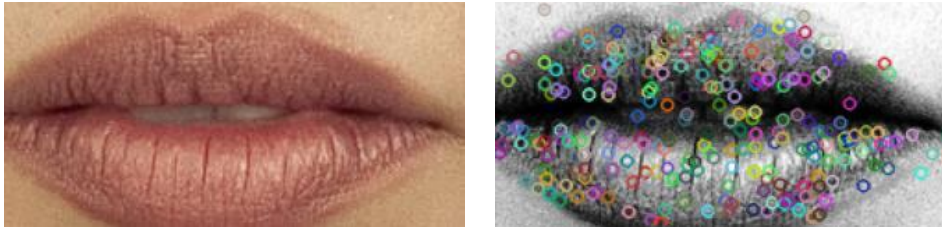


Figure 5.3: A representation of SURF features over the original image as shown by the author

## Feature Representation

The next stage in the pipeline after feature extraction is feature representation. Due to the enormous number of local characteristics for each image, it is necessary to quantize the feature vectors into visual words before training the relevant classifiers. As a result, the model Bag of Visual Words (BoVW) is utilized to solve this shortcoming. Because of its effectiveness and performance, this approach has gained popularity in recent years [65, 66]. Local descriptors are encoded into a histogram representation using the BoVW technique, which uses the k-means algorithm to cluster the feature descriptors [66]. Thereafter, each image is represented by a k-bins histogram. The number of clusters that best suits the pipeline is 80. These features are then used to train the classifiers.

## Classification

The final stage in the traditional pipeline is classification. The methods chosen for this stage include the Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN) classifiers. A Support Vector Machine was chosen as one of the classifiers for this pipeline because it has proven to be successful when used for pattern classification problems and its ability to achieve accurate results [67, 68]. Although SVMs work well with default parameter values, the performance of an SVM can be improved significantly using parameter optimisation. Therefore, a grid search is used to optimise the parameter values taking parameters such as regularisation, kernel and gamma into consideration. The advantage of using grid search

results in a greater learning accuracy [69]. The regularisation parameter, which determines the tolerance of misclassified points is set to 100. The kernel parameter, which transforms the data into its required form is set to radial basis function (rbf) with a gamma value of 0.001. Since the current study is a multi-class classification problem, a one-vs-rest classifier is used. This strategy splits a multi-class classification into a binary classification problem per class. Lastly, the probability parameter is set to True in order to compute probability estimates for each prediction.

The next classifier, K-Nearest Neighbour, is chosen because it is generally easy to implement, the training is very fast and it is robust to noisy training data. K-NNs have been used for pattern recognition since the 1970's [70]. As with SVMs, K-NNs also work well with default parameter values. However, its performance can also significantly be improved using parameter optimisation. Thus, the grid search algorithm is used for this classifier as well. The " $k$ " parameter or the number of nearest neighbours, which determines how many neighbours should be checked when data is being classified, is the most fundamental parameter. After trying different values for  $k$  (ranging from 1 to 19), this value is set to  $k = 1$ . The weights parameter which determines how weight should be distributed between the neighbouring values is set to distance. This means that closer neighbours will have a higher weight. The metric parameter which determines how the distance of the neighbouring points is chosen is set to euclidean. Lastly, a one-vs-rest classifier is used for the multi-class classification problem.

### 5.3.4 Deep Learning Pipeline

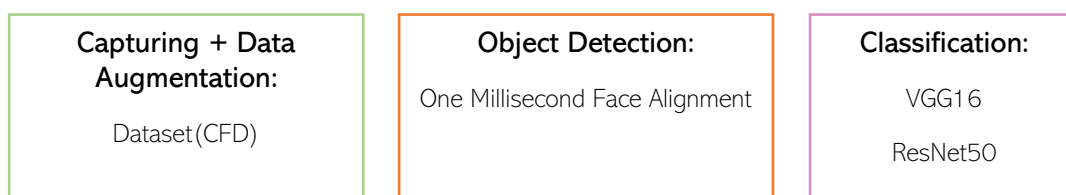


Figure 5.4: Deep Learning Pipeline

In traditional machine learning problems, a good amount of time is spent in manual feature selection and extraction. However, in deep learning the network manually extracts features and learns their importance using weights. The diagram for the deep learning pipeline is

represented in figure 5.4. The first approach in this pipeline implements a VGG16 convolutional neural network and the second approach implements a ResNet50 convolutional neural network. These approaches will be described and justified below.

## VGG16 Convolutional Neural Network

VGG16 is a deep learning architecture with many different layers. The VGG16 architecture was considered for the current study because it is a good architecture for benchmarking on a specific task and it is an excellent vision model architecture. The architecture consists of convolutional layers, pooling layers and fully connected layers which can be seen in figure 5.5. A convolutional layer is tasked with filtering the input with useful information. It is equipped with parameters so that the filters can automatically extract useful information from the input [71]. This layer also introduces the Rectified Linear Unit (ReLU), a piecewise linear function that increases non-linearity in the image for a better performance of a certain task [71, 72]. There are other non-linear functions such as Sigmoid, however, ReLU performs better in most situations [71]. Pooling layers are responsible for reducing the dimensionality of each feature map thus making the neural network more robust and it retains the most crucial information [72, 73]. There are different methods for pooling such as max-pooling, average-pooling and probabilistic pooling. Before passing the input to the next layer it is flattened into a column vector. The next layer is the fully connected layer which takes all the neurons from the previous layer and connects it to every neuron it has [71, 73]. Classification then takes place by using the softmax activation function [71].

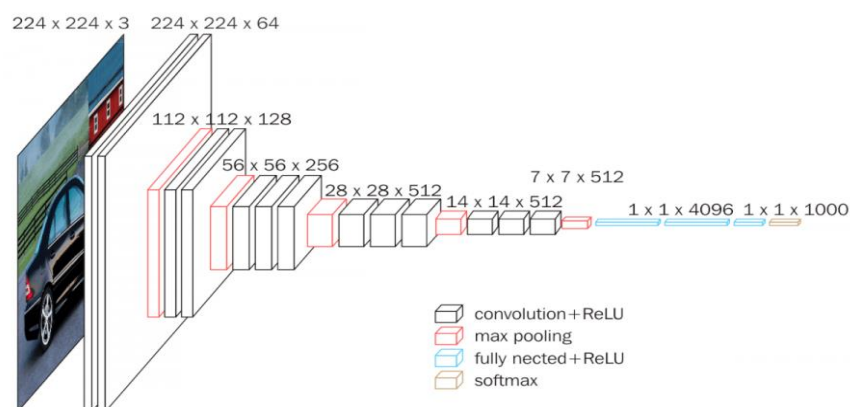


Figure 5.5: Architecture of VGG16 [75]

For the current study the VGG16 deep learning pipeline adopts a transfer learning approach using ImageNet weights. In the context of deep learning, most models need plenty of data. Gathering a great amount of labelled data for a range of different domains can be difficult, considering the time it takes to label the data. This forms the basis of transfer learning, which uses knowledge from pre-trained models to solve new problems [74]. Since transfer learning is the fine tuning of a pre-trained model, the fully connected layer which is responsible for classification. is replaced with a new layer relevant to the problem domain. Therefore, the output layer of the VGG16 model is replaced with a new layer that outputs the number of classes in the dataset with a softmax activation function since the number of classes is greater than two.

Before training the model, the images were first resized. The width and height of the images were computed, and an average dimension was determined. The images are resized to a width of 260 and a height of 224. The chosen parameters for the actual compilation of the model by using random search to optimise the parameter values include Adam as the learning optimiser function with a learning rate of 0.001. The loss function used is set to categorical cross-entropy due to the multiclass classification task. The number of epochs chosen to train the VGG16 model is set to 100. A random search was initiated a grid of hyperparameter values and selects random combination of values to train the model [69]. Lastly, to ensure a high accuracy in classification, data augmentation is used which assists in preventing overfitting [75]. Overfitting occurs when the model learns the details and noise in the training to the extent that it negatively affects the performance of the model. The technique employed for data augmentation includes a 1.3 zoom range. The parameters chosen for this model showed to be reasonable and yielded promising results. Lastly, the metrics used to assess the performance of the model are accuracy and loss.

## ResNet50 Convolutional Neural Network

The ResNet50 model is a 50-layer deep neural network. It is a variant of the ResNet model which has 48 convolutional layers, 1 max-pool layer and 1 average-pool layer. The only difference with a residual network as compared to other networks is the identity block between the layers. The ResNet50 model consists of 5 stages. From stage 2 to stage 5, each stage contains a convolution block and identity block as depicted in figure 5.6. Each

convolutional block and identity block contains 3 convolutional layers. The concept of skip connection also known as residual connections was first introduced by ResNet. As the name suggests, skip connections skip some layer in the network and feeds the output of one layer as the input to the next [75, 76]. This technique avoids information loss during training of deep networks and boosts the performance of a model [75, 76]. Therefore, this makes ResNets faster to train and are computationally less expensive than other CNNs.

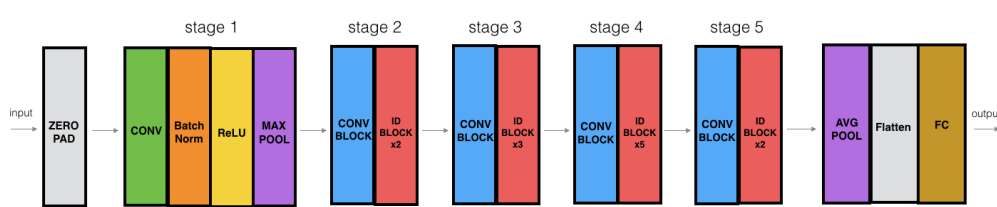


Figure 5.6: Architecture of ResNet50 [75]

The transfer learning technique is applied to the ResNet50 pre-trained architecture using ImageNet weights. The fully connected layer of the ResNet50 is removed and the convolutional layer of this pre-trained architecture is used as the base model. The fully connected layer is replaced with a new layer that outputs the number of classes in the dataset with a softmax activation function to classify the images into their respective classes.

The images are resized to a width of 260 and a height of 224. The input resolution has a significant impact on both accuracy and training time. For the current study, resizing the images is done to have faster training. The parameters chosen for the compilation of the model include Adam as the optimiser function with a learning rate of 0.001. The loss is set to categorical cross-entropy. The number of epochs chosen to train the ResNet50 model is set to 100. Subject to overfitting, data augmentation is used to alleviate this problem. The technique used includes a 1.3 zoom range. Lastly, the metrics used to assess the performance of the model are accuracy and loss. The parameters chosen for this pipeline yielded promising results.

CNNs are used for image classification and recognition due to its high accuracy. Essentially, these networks form the basis of computer vision methods. Therefore, it is suitable to apply this approach to the prototype implementation of the current study.

### 5.3.5 Deep Hybrid Learning Pipeline

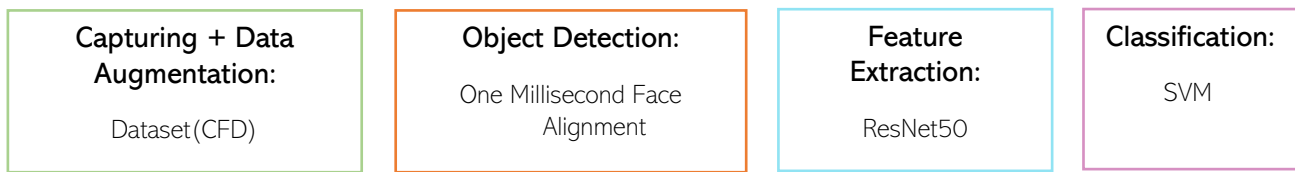


Figure 5.7: Deep Hybrid Learning Pipeline

Deep hybrid learning is a fusion approach in which a conventional deep learning model is combined with other algorithms. It is essentially converting a deep neural network architecture and fusing it with a learning algorithm, typically a machine learning algorithm [78]. This approach can have multiple forms such as early fusion where fusion happens before feature extraction or late fusion where fusion happens after feature extraction. The diagram for the deep hybrid learning pipeline is illustrated in figure 5.7.

The approach employed for this pipeline is a late fusion approach. ResNet50 is used as the feature extractor and SVM is used as the classifier as seen in figure 5.8. After feature extraction and feature vectors have formed using ResNet50, SVM is used to classify the data. The motivating thought behind using a CNN (ResNet50) as a feature extractor is the ability of deep neural networks to capture fine and precise high dimensional features which is a key component for classification [77, 78]. Using traditional machine learning algorithms to select the right set of features has been a challenging task [77] since feature selection and extraction takes place manually. Therefore, it is better to adopt a CNN architecture for feature extraction. Usually, a CNN architecture when used for classification, generally takes a long time to run. Furthermore, when fine tuning CNN parameters to obtain satisfactory results, computational complexity, and execution time increases [77]. Therefore, it is better to employ a machine learning algorithm for classification. Using SVMs for classification with CNN features has proven to produce admirable results [77, 78]. Lastly, CNNs suffer from the concern of vanishing gradients during training. This can be avoided by using an SVM as a classifier.

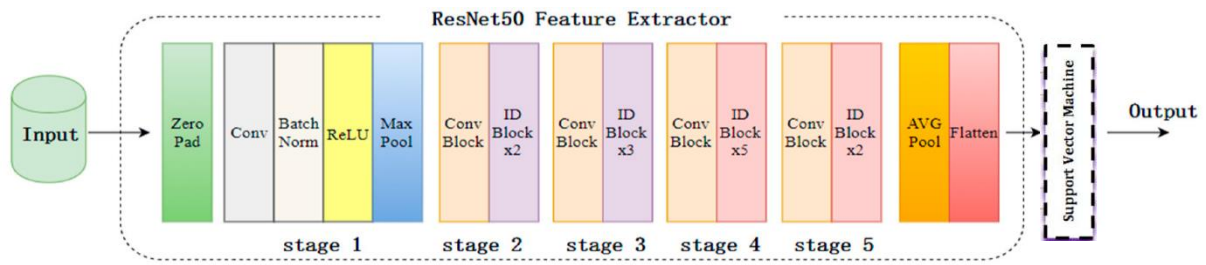


Figure 5.8: ResNet50 with SVM classifier

The ResNet model saved in the previous deep learning pipeline is used as the input for this pipeline with the fully connected layer removed. ResNet50 is chosen as the feature extractor because it can capture discriminative features efficiently. Thereafter, the outputs from the last layer are taken in by the SVM as a feature vector for the training process. Finally, classification on the test set is performed with the extracted features. To ensure that appropriate parameters are used for the SVM, a grid search is performed. The regularisation parameter is set to 1000. The kernel parameter is set to radial basis function (rbf) with a gamma value of 0.1 and a one-vs-rest classifier is used. Lastly, the probability parameter is set to True in order to compute probability estimates for each prediction.

The model and prototype have been thoroughly discussed in the sections above. It is now necessary to determine relevant benchmarks that can be used to evaluate the performance of the prototype. Therefore, the next section will discuss the appropriate benchmarks that can be employed.

## 5.4 Benchmark

“The goal is to turn data into information and, information into insight” – Carly Fiorina. It is paramount to critically analyse the implemented model since insights can be gained into the strengths of each pipeline but more importantly the limitations of each pipeline can also be revealed, thus paving way for future research opportunities. The benchmark consists of functional and non-functional requirements which will be assessed. These requirements are discussed further below.



### 5.4.1 Functional Requirements

A functional requirement can be defined as a description of tasks or functions that a system must perform [79]. In order to determine the functional requirements of the current study, it is beneficial to revisit its objectives:

1. Conduct a literature review within the research domain to identify the problem areas and relevant computer vision methods which can be used to achieve lip print identification along with appropriate datasets that can be employed.
2. Adopt a high-resolution face dataset which will allow for discriminatory lip features to be extracted.
3. Create experiments based on existing literature and findings by the author that can be used to achieve lip print identification.
4. Implement a prototype based on the experiments which can recognise individuals based on their lip prints by employing computer vision methods from the designed model.
5. Validate the performance of the prototype to determine its feasibility and report on these results in research articles and the dissertation.

After revisiting the objectives of the current study, a set of functional requirements are derived:

1. Select a secondary face dataset of a high-resolution.
2. Detect the mouth region (lips) within the image frame.
3. Identify the particular individual based on the extracted features.

### 5.4.2 Non-functional Requirements

Non-functional requirements are defined as a set of standards or metrics used to judge or evaluate a system [79]. Therefore, in order for the functional requirements to be met, certain non-functional requirements need to be addressed. These are:

#### Confusion Matrix

When evaluating the performance of a classification model on a set of test data for which the true values are known, a confusion matrix is commonly utilized [82]. True Positive (TP),

True Negative (TN), False Positive (FP), and False Negative (FN) are the most important parts of a confusion matrix (FN) [82].

- True Positive- the predicted positive value matches the actual positive value.
- True Negative- the predicted negative value matches the actual negative value.
- False Positive- the predicted value is falsely predicted as a positive value.
- False Negative- the predicted value is falsely predicted as a negative value.

For the current study a prediction represents whether the lip sample has been matched correctly. Each class represents a subject from the data set; therefore, a misclassification means that a lip of one subject gets mismatched with another lip sample. Furthermore, the confusion matrix will visualise the performance of the different algorithms and provide insight on the misclassifications.

### Accuracy

Accuracy compares how close a measured value is to its true value. Therefore, it is the ratio of the number of correct instances to all the number of instances [81]. The formula is as follows:

$$\text{Accuracy} = (TP + TN)/(N + P) \quad (5.1)$$

### Precision and Recall

Precision measures the number of correct instances out of all the total number of instances (true positives and false positives) [81]. The formula is as follows:

$$\text{Precision} = (TP)/(TP + FP) \quad (5.2)$$

Recall determines what portion of actual positives are identified correctly divided by all the correct instances [81] which can be seen in the formula below.

$$\text{Recall} = (TP)/(TP + FN) \quad (5.3)$$

## F1 Score

F1 Score is an estimate of the accuracy of the model. It is the weighted average of both precision and recall [81]. The formula is as follows:

$$\mathbf{F1\ Score} = 2 \times ((\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})) \quad \mathbf{(5.4)}$$

## Intersection over Union

Intersection over Union (IoU) is the most popular evaluation metrics for measuring the overlap between two bounding boxes [75, 80]. Given an image, The IoU is a metric that compares how close the anticipated and ground truth regions are. It is calculated by dividing the intersection's size by the union of the two zones [80].

$$\mathbf{IoU} = (\text{Area of Overlap}) / (\text{Area of Union}) \quad \mathbf{(5.5)}$$

## Mean Average Precision

Mean Average Precision (mAP) calculates the mean AP (AP is calculated for each class) over all the classes or the overall IoU thresholds [75]. mAP results in an average performance score as the final measure of performance. A high mAP value indicates a better performance.

## Equal Error Rate

The Equal Error Rate (EER) is the most crucial performance measure to evaluate the performance of a biometric recognition system because it describes the overall accuracy of the biometric system [7]. It describes the point where the false acceptance rate (FAR) and the false rejection rate (FRR) are equal [9]. A low EER value indicates a better performance.

## Receiver Operating Characteristic

FAR and FRR are traded off in a Receiver Operating Characteristic (ROC) curve [9]. As shown in figure 5.9, it is a probability curve that compares the true positive rate (TPR)

against the false positive rate (FPR) at various thresholds. The goal of a ROC curve is to maximize the area beneath the curve (AUC). The AUC indicates how well the model distinguishes between positive and negative classes. The greater the AUC, the better.

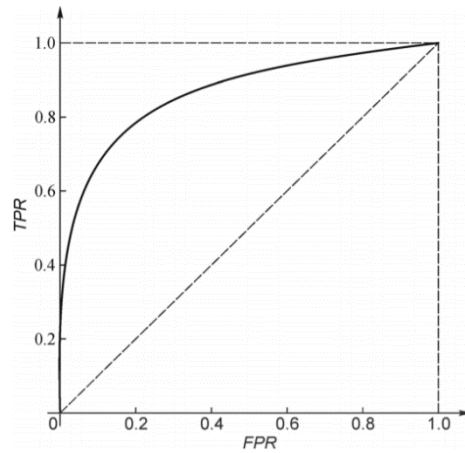


Figure 5.9: Example of ROC Curve obtained from [82]

### Precision-Recall Curve

Precision-Recall Curve: is a trade-off between precision and recall for different threshold values [82]. Ideally, a precision-recall curve will show a curve which is as close as possible to the lower right-hand portion of the graph. A high area under the curve indicates both a high recall and a high precision [82]. This ensures that the classifier is predicting accurate results and returning majority of all positive results. Figure 5.10 is an example of a precision-recall curve.

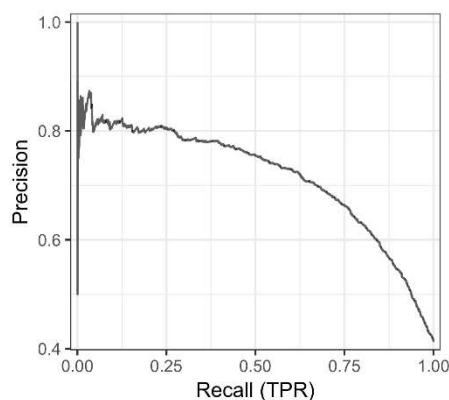


Figure 5.10: Example of Precision-Recall Curve obtained from [83]

## Learning Curves

Learning Curves: during training of a neural network, the current state of the model at each step of the training process can be evaluated on a training dataset and validation dataset to give an idea of how well the model is “learning” or “generalising” [85]. In this case two learning curves can be created: accuracy and loss. These will be further explained below.

- Loss is the summation of the errors made during training. During the training process the goal is to minimise the loss function. The loss curve gives insight about the learning rate and the model’s behaviour such as overfitting or underfitting [85].
- Accuracy is used to measure the model’s performance. Essentially, it is the number of predictions where the predicted value is equal to the true value. It is graphed and observed during the training process. The final value is typically associated with the model’s overall accuracy [85]. Figure 5.1.1 illustrates these learning curves.

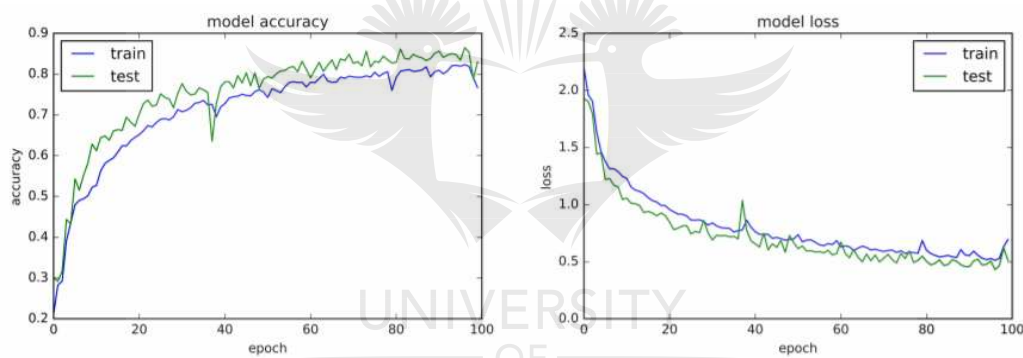


Figure 5.1.1: Accuracy and Loss curves for CNN model obtained from [86]

## 5.5 Conclusion

This chapter started with an illustration and detailed discussion of the designed model that would be appropriate for the current study. The different components of the model (capturing, data augmentation, object detection, pre-processing, feature extraction, feature representation and classification) together with their potential methods have been laid out. The designed model addresses research objective **RO3** in section (1.3) which involves the development of a lip print recognition model. Therefore, objective 3 of the current study has been met. The lip print recognition model contributes significantly to proposing a variety of computer vision methods which can potentially be used to solve the problem.

Thereafter, each of the different pipelines (traditional machine learning, deep learning, and deep hybrid learning) along with their respective diagrams depicting its architecture and methods have been discussed and implemented. Furthermore, the implementation of the model is another step closer to achieving the objectives of the current study. **RO4** which involves developing a prototype based on the designed model has now been addressed. Thereafter, the chapter discussed the functional and non-functional requirements that will be used to benchmark the implemented model. The benchmark allows for a critical analysis of the implemented model, thereby giving insight into the strengths and limitations of each pipeline.

Now that a firm understanding of the model, prototype and benchmark has been posed, the hypothesis can be tested which will either be accepted or rejected based on the results gathered for the benchmark tests. The next chapter, therefore, will aim to determine whether the objectives of the current study have been met and accomplished in addressing the research problem.



# Chapter 6 Results

## 6.1 Introduction

The previous chapter discussed and outlined the experiment of the current study along with the prototype and its benchmark. The implemented prototype discussed previously yielded a set of results which can be used to provide insights on how well the research problem can be solved. Therefore, in this chapter the results of the current study are presented and discussed with reference to the hypothesis of the study, which postulates that employing deep learning-based methods can achieve lip print recognition in an effective manner.

The analysis and interpretation of the results is carried out in two phases: operational results (functional requirements) and metrics (non-functional requirements). Section 6.2 will analyse and discuss the results based on the functional requirements highlighted in the previous chapter. The next section, section 6.3 will discuss the non-functional metrics that were achieved for each of the implemented pipelines. Thereafter, in section 6.4 the results are comprehensively discussed. In section 6.5 the current study is compared with similar systems. The chapter concludes in section 6.6.

## 6.2 Operational Results

The functional requirements discussed in the previous chapter include adopting a high-resolution face dataset, detecting the region of interest (lips) within the image frame, and identifying the particular individual based on their lip print. In this section, the approaches used to achieve the above-mentioned functional requirements will be discussed.

### 6.2.1 Dataset Selection

The selection of an appropriate dataset is crucial to the development of any research study. The selected dataset must contain features relevant to the problem domain with sufficient predictive power that enables the training process to learn from it [90]. In order to extract visual features related to lip print identification, an accurate extraction of the lip is essential

with its patterns visible. Large public databases are available for many facial analysis problems. These datasets are divided into two categories: datasets produced within a controlled environment and datasets produced in an uncontrolled environment [90]. Images from social networks such as Facebook and image search engines such as Google Images have contributed to the production of large datasets, comprising of facial images in different uncontrolled situations, more commonly referred to as “in-the-wild” [90]. However, these “in-the-wild” datasets are not applicable to the problem domain of the current study because it would not adequately address the research problem. Figure 6.1 is an example of an “in-the-wild-dataset” called Labelled Faces in the Wild (LFW). Notably, the figure depicts that the images are of a low resolution and consequently, the lip patterns found on the lip are not as visible and can be left for future work. Furthermore, these datasets are subject to poor lighting, extreme pose, and strong occlusions because they are produced in an uncontrolled environment. As a result, the lip patterns, which are crucial for lip print identification, would be difficult to extract due to these factors. Therefore, datasets produced in a controlled environment are more ideal for the current study. However, there is little consistency amongst the different sets. Some sets have a very low resolution, and the lip patterns are not visible while other sets have minimal subjects with minimal subject diversity and the results from these datasets would not be effective. Therefore, the CFD, a dataset which consists of various ethnicities and a high resolution of captured images in which the grooves on the lips are visible, was used for the current study. Figure 6.2 shows some sample images from the CFD dataset.



Figure 6.1: Sample images from LFW





Figure 6.2: Sample images from CFD

### 6.2.2 Lip Detection

Before discussing the appropriate methods used to perform object detection it is important to determine how the ground truth labels were created manually by the author. The approach adopted was to find an approximate bounding box within which the lip must lie. Since the current study is only concerned with the lip facial landmark, the bounding box should only include the lip region to ensure that accurate pre-processing and feature extraction takes place and only features from within the lip are highlighted and used. Therefore, the bounding box coordinates were created by considering the left and right corner of the lip as well as the upper and lower lip which is visualised in figure 6.3. These bounding box coordinates produced a set of 2 (x, y) coordinates with a width and height which were consequently used as ground truth labels.

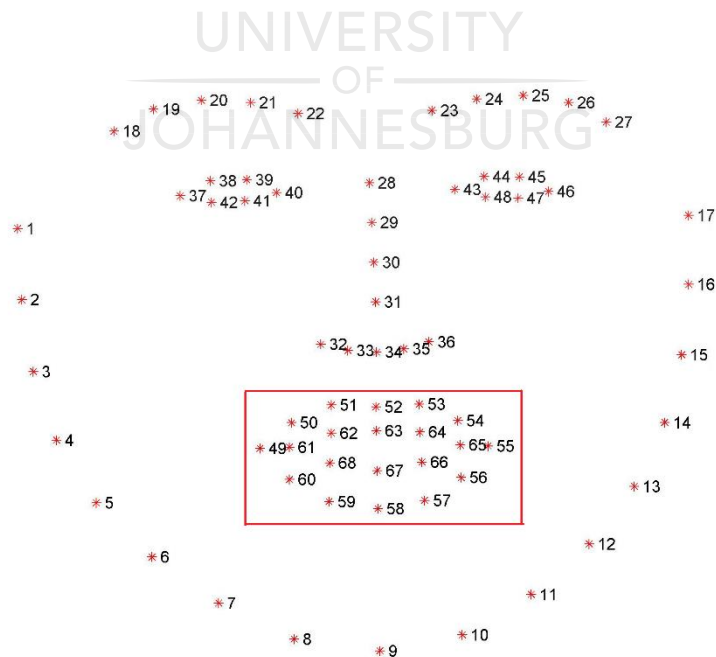


Figure 6.3: Visual representation of the bounding box for the lip

The methods used to perform object detection include Haar cascade and the One Millisecond Face Alignment algorithm. Initially, the Haar cascade classifier based on the Viola-Jones algorithm was used to detect the lip region. However, although the algorithm was successful in detecting the face, detecting the lip region was unsuccessful as excessive false positives were achieved as shown below. Therefore, the One Millisecond Face Alignment algorithm was used for lip detection. However, the results obtained from both these methods will be discussed further in this section which will provide an experimental comparison of these methods.

Figure 6.4 shows instances where the Haar cascade, Viola-Jones algorithm was unable to detect the lip within the image frame. The predicted images show how the algorithm performed in detecting the lip from the face for each image in the dataset. Notably, the algorithm either falsely classified regions in the image as a lip or it detected the lip but with several false positives. Subject 4, for example, in figure 6.4, detected the lip but with a false positive. Therefore, although the algorithm succeeded in detecting the face of each image in the dataset, it failed to ideally detect the lip or the lip was missed entirely even though the images were taken in a highly controlled environment with consistent lighting. This is mainly because Haar cascades are prone to false positives [91]. The Viola-Jones algorithm can also report a face in an image when no face is present. Therefore, in cases where some false positives detections can be tolerated, using the Haar cascade technique is beneficial due its fast computation [91]. However, for the current study, false positives are not condoned and consequently, a different technique was adopted.

|           | Subject 1   | Subject 2   | Subject 3  | Subject 4   |
|-----------|---|---|--|---|
| Actual    |  |  |  |  |
| Predicted |  |  |  |  |

Figure 6.4: Haar cascade bounding box predictions compared to the ground truth labels

Figure 6.5 shows the performance of the One Millisecond Face Alignment algorithm. The predicted images depict how the algorithm performed in detecting the Region of Interest (ROI) i.e., the lips. The algorithm correctly detected the lips of each image in the dataset with no false positives. Because the lip is a facial component and it is crucial to know where it is located on the face for the current study, the One Millisecond Face Alignment algorithm uses an ensemble of regression trees to estimate that landmark position from pixel intensities for object detection. Therefore, it has the potential to make high-quality predictions and in real-time as well.

The number of samples used for the current study is 500 targets out of a total 597 from the CFD with variations in transformations, illumination and rotations. As mentioned in section 4.8.2, due to the diverse nature of the data set, the results obtained by the study can potentially generalise better. The accurate detection of the lip within the image frame is an important step in the completion of the different pipelines. The Haar cascade technique failed to ideally detect the lip with many false positives. However, the One Millisecond Face Alignment algorithm accurately detected the lip region within the image frame. The main reason for the high accuracy is because the images in the dataset were produced in a highly controlled environment with consistent lighting. Johnston et al. [90] indicate in their facial landmark identification paper that face detectors perform best when images are taken in a controlled environment and perform poorly when they are taken in uncontrolled environments. The results show that the performance of the One Millisecond Face Alignment algorithm is viable and outperformed the Haar cascade method with no false positive detections. The IoU and mAP will be discussed in the non-functional requirements results section.








|           | Subject 1   | Subject 2   | Subject 3  | Subject 4   |
|-----------|---|---|--|---|
| Actual    |  |  |  |  |
| Predicted |  |  |  |  |

Figure 6.5: One Millisecond Face Alignment bounding box predictions compared to the ground truth labels

The field of computer vision has experienced substantial progress, owing largely to advances in deep learning. However, classifying and detecting an unknown number of objects within an image was considered a challenge only a few years ago [129]. Although, it is now possible to perform both object detection and classification using approaches such as Faster R-CNN and YOLO. However, small object detection is still a challenging task in this field [129, 130]. Apart from small representations of an object, the diversity of input images also makes the task difficult [130]. For example, YOLOv3 allows images of various resolutions but a high-resolution image requires more processing time which greatly affects its speed at which it can make predictions [129]. More importantly, class imbalance proves to be an issue for small object detection [129]. When considering the current study, the lip is the main object, and the remainder of the image is filled with the background. Therefore, it becomes challenging for approaches such as R-CNN to detect small objects because most regions will not contain objects and are considered negatives, which also greatly affects its processing time [129]. For these reasons and limited computational processing resources, a deep learning approach for object detection was not adopted. However, the results for an end-to-end deep learning approach can be found in the second manuscript (Appendix B).

### 6.2.3 Lip Print Recognition

The main pipelines that were implemented for lip print recognition include a traditional machine learning pipeline (with a SURF-BoVW feature space and SVM or K-NN classifier), a deep learning pipeline (VGG16 and ResNet50) and a deep hybrid learning pipeline (ResNet50 with an SVM classifier). It should be noted that this is a multi-class classification in line with subject identification where each class is a subject in the dataset. An overview of the classification of the lips will be provided in this section. It will discuss which lips were classified correctly and which were misclassified and comment on shared misclassifications across the pipelines.

Figure 6.6 shows the confusion matrices of the traditional machine learning pipelines, deep learning pipelines and deep hybrid learning pipeline, where along the x-axis, the true class labels are listed and along the y-axis the class predictions are listed. The diagonal shows the number of correct classifications whereas all the other entries not in the diagonal are

the misclassifications in the test set that has a total of 500 individuals. These misclassifications are summarised in table 6.1.

Table 6.1: Summary of Misclassifications from Pipelines

| Pipeline Variation | Misclassifications       |                        | Total Number of Misclassifications |
|--------------------|--------------------------|------------------------|------------------------------------|
|                    | Number of Female Targets | Number of Male Targets |                                    |
| SURF (BoVW) + SVM  | 4                        | 9                      | 13                                 |
| SURF (BoVW) + K-NN | 9                        | 12                     | 21                                 |
| VGG16              | 8                        | 11                     | 19                                 |
| ResNet50           | 0                        | 1                      | 1                                  |
| ResNet50 + SVM     | 3                        | 1                      | 4                                  |

It can be seen from the confusion matrix in figure 6.6 (A) and table 6.1 that there were 13 misclassifications for the SVM pipeline while there were 21 misclassifications for the K-NN pipeline which can be seen in figure 6.6 (B). The lips that were misclassified for the SVM variant belonged to 4 female targets and 9 male targets while the K-NN variant misclassified 9 female targets and 12 male targets. It is evident that the SVM and K-NN variants struggled to identify male targets, particularly, those of African descent. Shared misclassifications for this pipeline can be seen in figure 6.7 where both variants failed to identify the same male targets.

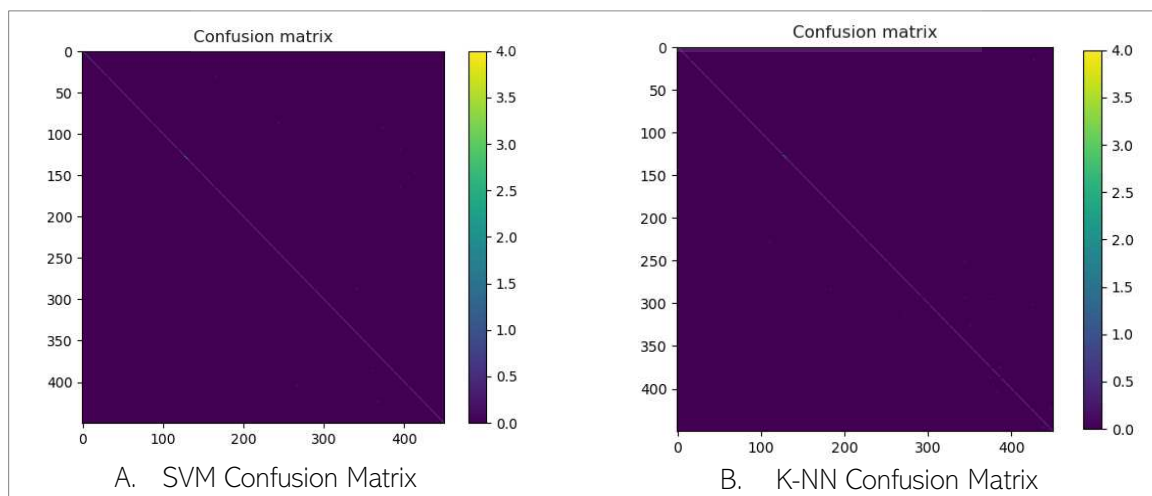


Figure 6.6: Traditional pipeline confusion matrices




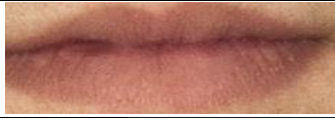

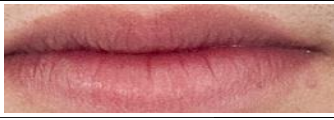
| Shared Misclassifications   | SVM Predicted Class   | K-NN Predicted Class   |
|---|---|--|
| BM-247  | AF-208  | BF-003   |
|  |  |  |
| WM-231  | WF-002  | WF-038   |
|  |  |  |

Figure 6.7: Shared misclassifications from the SVM and K-NN pipeline

Figures 6.8 (A) and 6.8 (B) show the confusion matrices of the deep learning-based pipeline with VGG16 and ResNet50, respectively. Table 6.1 and the confusion matrices indicate that the VGG16 deep learning pipeline had a total 19 misclassifications whereas the ResNet50 deep learning pipeline had a total of only one misclassification. Similar to the previous pipeline where the highest number of misclassifications belonged to males, the VGG16 variant also misclassified 11 male target lips while misclassifying a total of 8 female target lips. The ResNet50 variant misclassified 1 male target lip. Samples of the misclassifications for the deep learning pipeline can be seen in figure 6.9. An interesting fact that can be noticed amongst the traditional and deep learning-based pipelines is that target WM-231 was particularly difficult to identify.

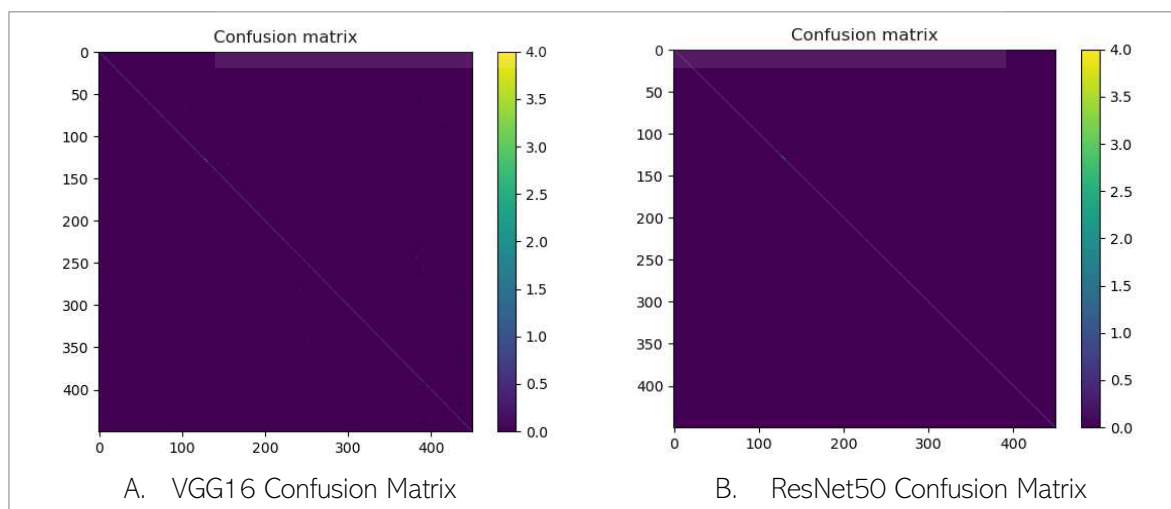


Figure 6.8: Deep learning pipeline confusion matrices

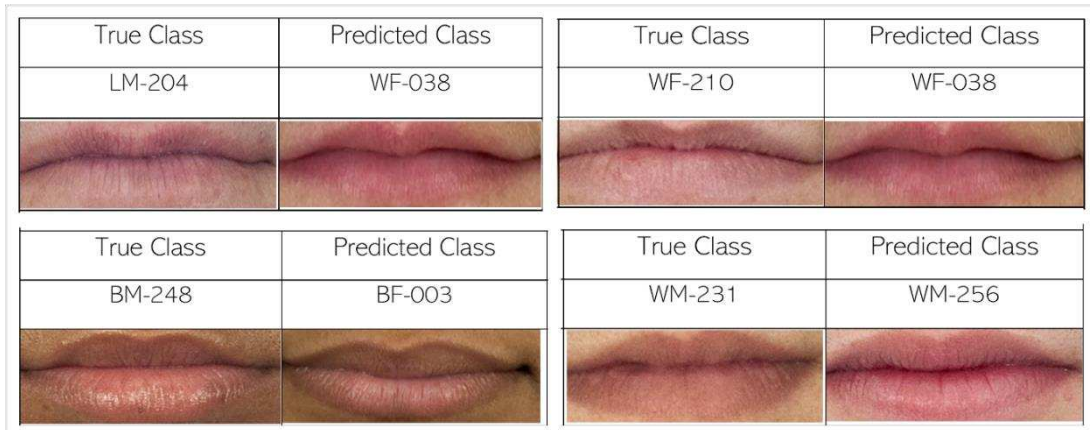


Figure 6.9: Samples of misclassified lips from the VGG16 pipeline

Figure 6.10 shows the confusion matrix of the deep hybrid learning-based pipeline. It is evident from Table 6.1 and the confusion matrix represented below that there was a total of 4 misclassifications for this pipeline where 3 females and one male were incorrectly identified. Samples of the misclassifications can be seen in figure 6.11.

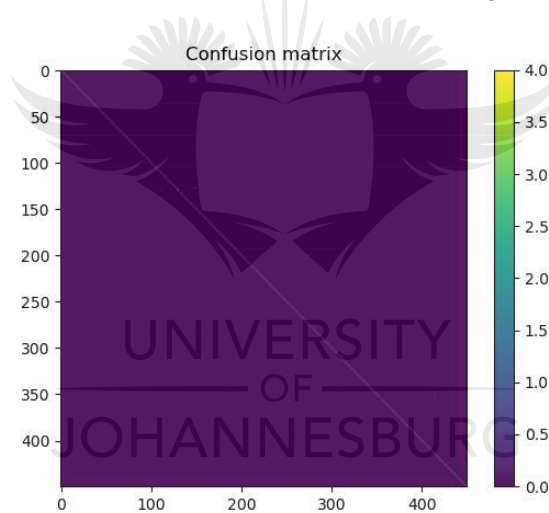


Figure 6.10: Deep hybrid learning confusion matrix

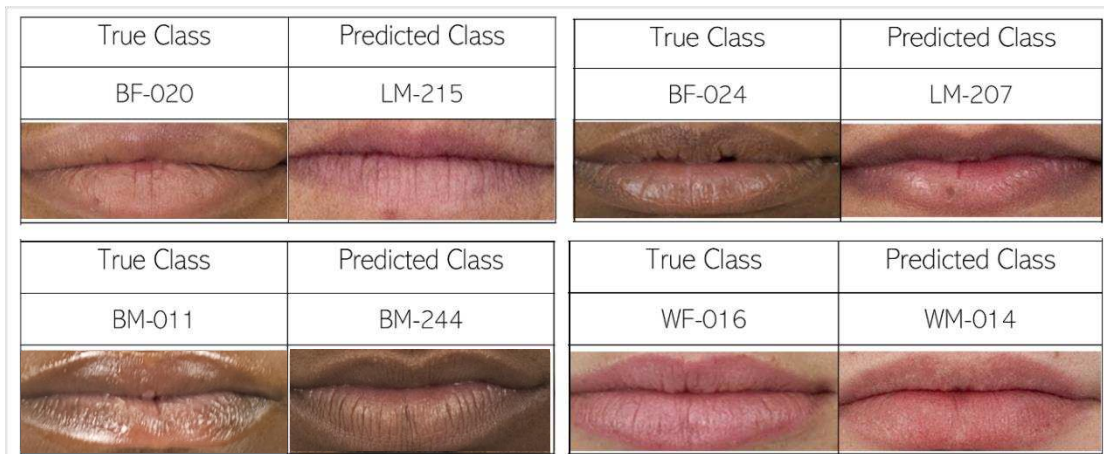


Figure 6.11: Samples of misclassified lips from deep hybrid learning pipeline

One observation that can be made from the misclassifications is that the misclassified lips were similar in size and shape. Target WM-231 from the traditional pipeline, for example, was misclassified with a lip that looks similar in both shape and size. Similarly, target LM-204 and target WF-210 from the deep learning pipeline for example were also misclassified with a lip that looks similar in size and shape. Furthermore, it is apparent, that target BM-011 and target WF-016 from the deep hybrid learning pipeline were misclassified with a lip similar in shape and size as well. Therefore, by considering the misclassifications from the different pipelines it is evident that if the lip was thin or if the outline of the lip was similar, it was misclassified with a lip equivalently thin in size and similar in shape. However, targets BF-020 and BF-024 from the deep hybrid learning pipeline were misclassified with a lip that looks different from the actual lip. Two potential reasons for these misclassifications could be due to the lack of specificity (grooves visible on the lips) or it could be caused by the facial landmark detector when detecting the lips. It is evident that some lips were not adequately detected by the facial landmark detector and therefore the grooves on the lips might not have been visible which caused misclassifications.

Overall, the results for lip print identification based on the confusion matrices and misclassifications indicate that it is possible to achieve lip print identification. The pipeline with the highest number of misclassifications was the K-NN variation with 21 misclassifications, followed by the VGG16 architecture with 19 misclassifications. The SVM variation achieved a total of 13 misclassifications. The pipelines with the lowest misclassifications were the deep hybrid learning pipeline with 4 misclassifications and the ResNet50 variation with only 1 misclassification.

### 6.3 Metrics

In addition to the operational results, metrics were gathered for each individual pipeline where multiple runs were performed to increase robustness. These results will be used to judge or evaluate the performance of the pipelines. The IoU and mAP of the One Millisecond Face Alignment algorithm and Haar cascade classifier will be discussed followed by the performance of the different pipelines which include the traditional machine learning pipeline with a SURF-BoVW feature space and an SVM or K-NN classifier, the deep learning-based



pipeline with VGG16 and ResNet50 architectures and the deep hybrid learning pipeline with a ResNet50 architecture (without a fully connected layer) and an SVM classifier.

### 6.3.1 Intersection over Union Performance for Object detection

The One Millisecond Face Alignment algorithm successfully detected the lip region of each individual in the dataset with no false positive detections. However, that is not enough to assess the performance of the algorithm. Therefore, the IoU of each detected lip within the image frame was determined, followed by the AP for each sample. Thereafter, the mAP was computed based on the AP from each sample. As discussed in the previous chapter (section 5.4), IoU measures the overlapping area between the ground truth region and the predicted region by dividing the area of union between them. For the current study, the IoU was given a threshold of 0.8. By comparing the IoU with a given threshold a detected lip can be predicted as correct or incorrect. If the prediction is greater than the threshold it is correct whereas if it is less than the threshold it is incorrect. Figure 6.12 is an illustration of the ground truth region, predicted region and the IoU metric of some targets from the dataset.



Figure 6.12: Samples of IoU score, ground truth region and predicted region of different targets

When the IoU threshold of 0.8 can be considered, it is apparent from figure 6.9 that most predictions received an accuracy of 0.90 or higher. However, one observation and disadvantage that can be noted is that the lips that were physically larger in size achieved an IoU score of less than 0.90. One potential reason for this could be due to erroneous detections (inaccurate landmark detection). Kim et al. [131] state that erroneous detections can be classified into three types: (I) landmarks are not detected at all; (II) some of the

facial landmarks are incorrectly detected and their coordinates are slightly off the correct position; (III) non-facial regions are detected as facial landmarks. Based on the results it is notable that the lips which achieved an IoU score of less than 0.90 were incorrectly detected and slightly off the correct position. Therefore, when a target had lips larger than the average it possibly became more difficult for the algorithm to ideally detect the entire lip.

Figure 6.13 is a representation of the Average Precision (AP) achieved for the samples in the dataset. Overall, the results are satisfying and indicate that the algorithm is highly viable and achieved a mAP of 93%. The One Millisecond Face Alignment algorithm correctly located lips of different shapes and sizes with a good performance. According to Rosebrock [132] dlib's 68-point facial landmark detector is the most popular facial landmark detector in the computer vision field due to the speed and reliability of the dlib library.

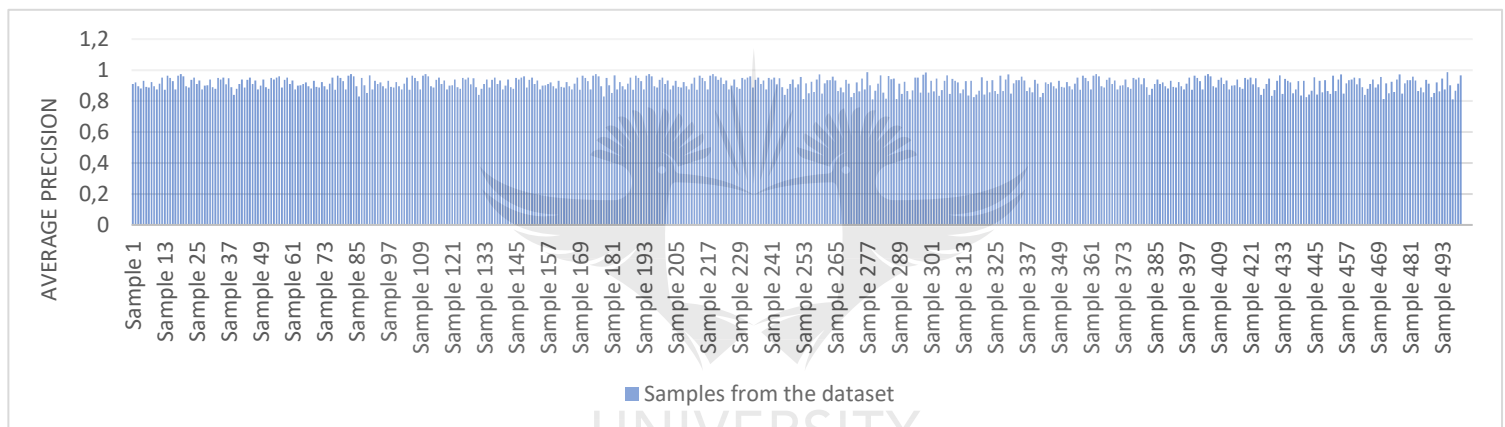


Figure 6.13: Average precision of samples from the dataset using the One Milli Second Face Alignment

Unfortunately, the Haar cascade classifier performed poorly. Not only were multiple false positives detected but the IoU for each sample did not yield satisfactory results either which can be seen in figure 6.14. When the IoU threshold of 0.8 can be considered, it is apparent from figure 6.13 that most predictions received an accuracy of 0.52 or lower. A mAP of 49% was achieved which suggests that the Haar cascade classifier is not a feasible approach for lip print recognition.

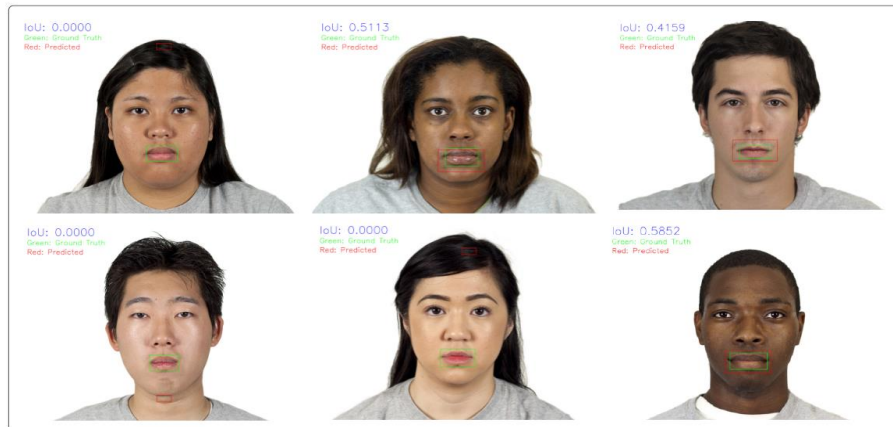


Figure 6.14: Samples of IoU score, ground truth region and predicted region of Haar cascade classifier

### 6.3.2 Traditional Pipeline with SVM

The first variation of the traditional machine learning pipeline made use of the One Millisecond Face Alignment algorithm for object detection, SURF for feature extraction, a bag-of-words approach for feature representation and SVM as the classifier. The metrics presented for the traditional pipeline with SVM indicates that this approach yielded promising results with an accuracy of 95.83%. A precision score of 97.56% and recall score of 95.93%, demonstrates that the model achieved a high percentage of accurate results. Furthermore, an f1 score of 96.74% was achieved. A low EER rate of 1.42% indicates a very good performance of the model. These results can be seen in table 6.2.

When the ROC curve with a micro-average of 98% and a macro-average of 99%, in figure 6.15 can be considered, it is apparent that the curve closely draws towards the upper left corner of the figure, indicating stable predictions of the proposed model. The PR curve in figure 6.16, depicts that it is close to the lower right-hand portion of the graph which determines that the classifier is predicting accurate results i.e., accurately predicting the lips. Overall, the results from this variation indicates that it achieves lip print identification with a very good performance.

Table 6.2: SVM Pipeline Results

| Accuracy | Precision | Recall | F1 Score | EER   |
|----------|-----------|--------|----------|-------|
| 95.83%   | 97.56%    | 95.93% | 96.74%   | 1.42% |

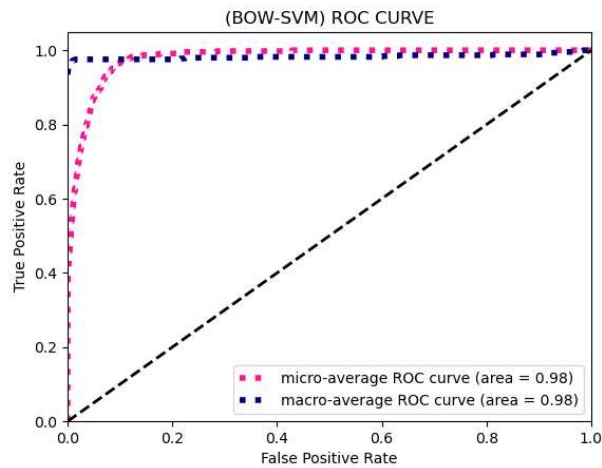


Figure 6.15: SVM Pipeline ROC Curve

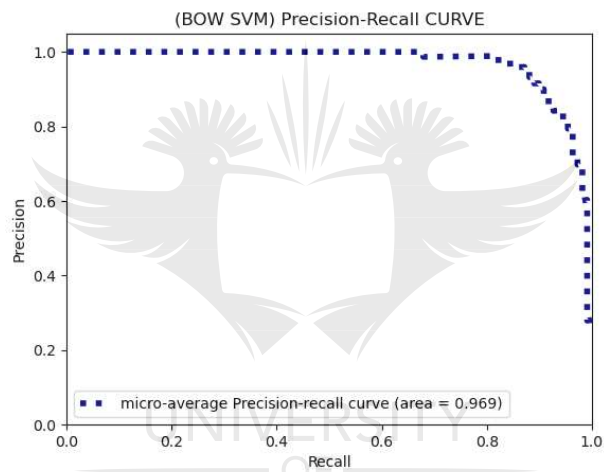


Figure 6.16: SVM Pipeline PR Curve

### 6.3.3 Traditional Pipeline with K-NN

The second variation of the traditional machine learning pipeline made use of the One Millisecond Face Alignment algorithm for object detection, SURF for feature extraction, a bag-of-words approach for feature representation and K-NN as the classifier where  $k$  is equal to 1. The metrics presented for this variation indicates that this approach also yielded promising results. The overall accuracy achieved for this variation was 93.75%. This indicates that there were a high number of correct classifications out of the total classifications that were made. A precision score of 96.34% was achieved which hints at the classifier's ability to make stable and accurate predictions. The recall rate was calculated to be 94.71%

which determines that a large number of actual positives were identified correctly. The EER achieved for this variation was 2.64% with an f1 score of 95.52%. The metrics for this variation can be seen in table 6.3.

The ROC curve for this variation can be seen in figure 6.17. It can be seen from the ROC curve with a micro-average of 98% and macro-average average of 98% indicates that the model performed well in identifying the lips. The PR curve illustrated in figure 6.18 with a micro-average of 94%. The metrics discussed for this variation indicate that it is also a viable approach for lip print identification.

Table 6.3: K-NN Pipeline Results

| Accuracy | Precision | Recall | F1 Score | EER   |
|----------|-----------|--------|----------|-------|
| 93.75%   | 96.34%    | 94.71% | 95.52%   | 2.64% |

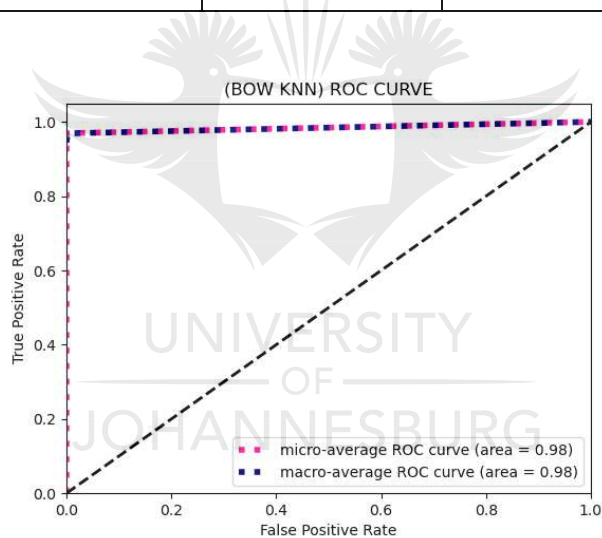


Figure 6.17: K-NN Pipeline ROC Curve

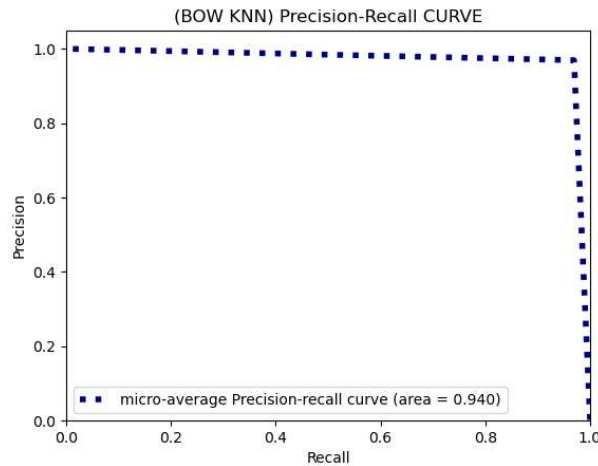


Figure 6.18: K-NN Pipeline PR Curve

### 6.3.4 Deep Learning Pipeline with VGG16

The first variation of the deep learning-based pipeline made use of the One Millisecond Face Alignment algorithm with a VGG16 architecture as the classifier. In section 6.2.3 it was discussed that this variation had the highest number of misclassifications with a total of 20 misclassifications. For the accuracy, the VGG16 model achieved an overall accuracy 94.51% which stipulates that it can make fairly accurate predictions. The precision score was 92% and the recall score was 94%, which are both adequate results for accuracy in terms of relevant occurrences. An f1 score of 93% was calculated with an EER rate of 2.14%. The results are represented in table 6.4.

The ROC curve for the VGG16 variation is shown in figure 6.19. with a micro-average of 89% and macro-average of 94%. The lack of accurate predictions for the VGG16 model is revealed in the curve. As discussed previously, a total of 20 misclassifications were obtained for this variation. The PR curve which can be seen in figure 6.20 achieved a micro-average of 89.9%. This indicates that there is still room for improvement.

The loss curve of the VGG16 model, in figure 6.21 shows that overfitting did not take place for this variation. The training and validation curves decrease to a point of stability. The gap between the validation and accuracy curves grows closer together as the number of epochs increases which is a favourable result which can be seen in figure 6.22.

Table 6.4: VGG16 Pipeline Results

| Accuracy | Precision | Recall | F1 Score | EER   |
|----------|-----------|--------|----------|-------|
| 94.51%   | 92%       | 94%    | 93%      | 2.14% |

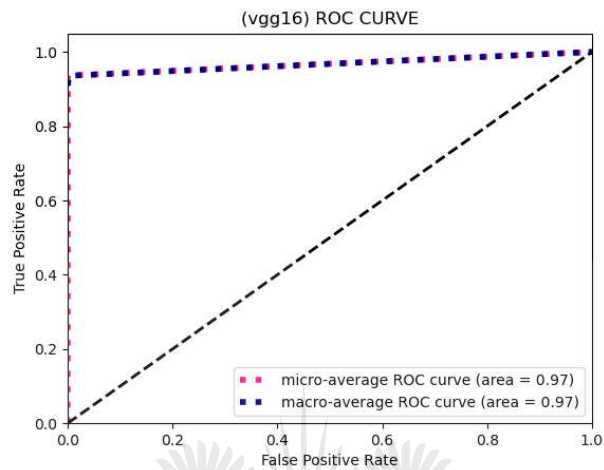


Figure 6.19: VGG16 ROC Curve

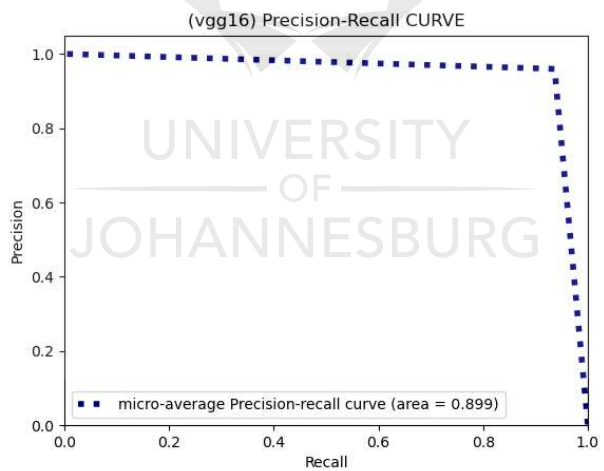


Figure 6.20: VGG16 PR Curve

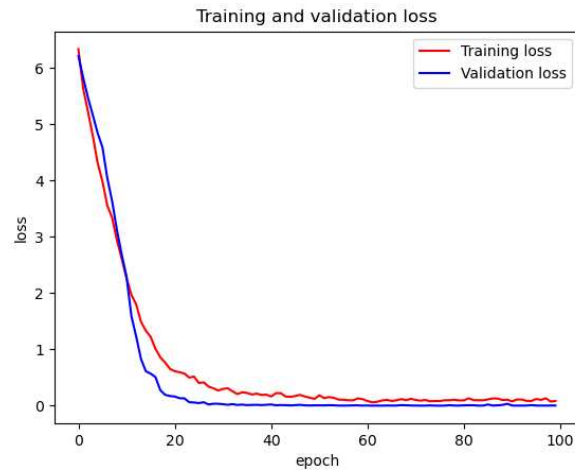


Figure 6.21: VGG16 Loss Curve

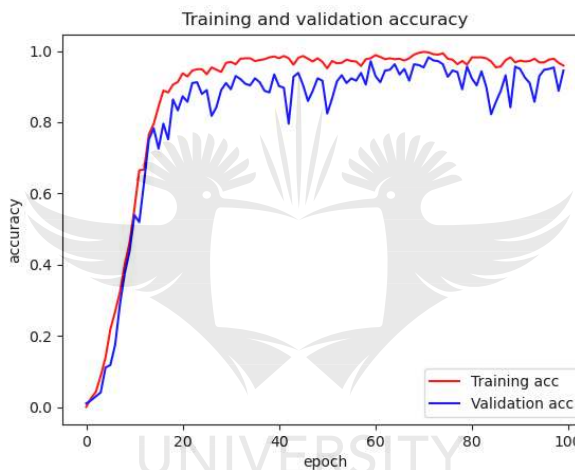


Figure 6.22: VGG16 Accuracy Curve

### 6.3.5 Deep Learning Pipeline with ResNet50

The second variation of the deep learning-based pipeline made use of the One Millisecond Face Alignment algorithm with a ResNet50 architecture as the classifier. In section 6.2.3 it was discussed that this variation had the lowest number of misclassifications with a total of one. The metrics presented for the deep learning pipeline with ResNet50 indicates that this approach yielded promising results with an accuracy of 99.78%. A precision score of 99.66% and recall score of 99.77% demonstrates that the model achieved a high percentage of relevant results, and a large number of actual positives were identified correctly. Additionally, the f1 score was calculated to be 99.70%. A low EER of 0.21%



suggests that the model's performance is excellent. The metrics for this variation can be seen in table 6.5.

The ROC curve for the deep learning ResNet50 pipeline can be seen in figure 6.23. Notably, the curve shows a favourable ROC curve representation. A micro-average of 99% and macro-average of 100% was achieved. Additionally, the PR curve which can be seen in figure 6.24 indicates that the model is very accurate at distinguishing between positive and negative classes.

The loss curve of the ResNet50 model is shown in figure 6.25. Notably, the training and validation loss curves indicate that overfitting does not take place since both curves decrease to a point of stability. However, as the number of training epochs increase both the curves move towards a minimal value. It decreases to a point of stability with a minimal gap between the two final values. The accuracy curve of the ResNet50 model is illustrated in figure 6.26. It is apparent that the gap between the validation and accuracy curves grows closer together as the number of epochs increases which is a desirable result.

Table 6.5: ResNet50 Pipeline Results

| Accuracy | Precision | Recall | F1 Score | EER   |
|----------|-----------|--------|----------|-------|
| 99.78%   | 99.66%    | 99.77% | 99.70%   | 0.21% |

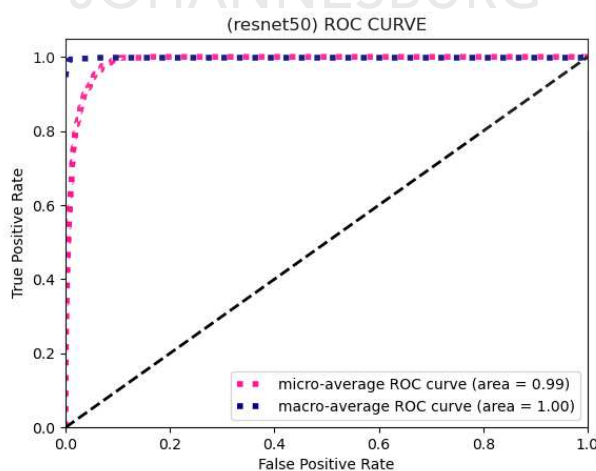


Figure 6.23: ResNet50 ROC Curve

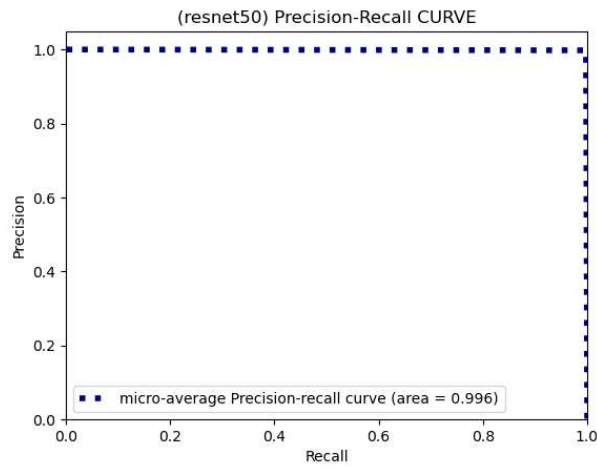


Figure 6.24: ResNet50 PR Curve

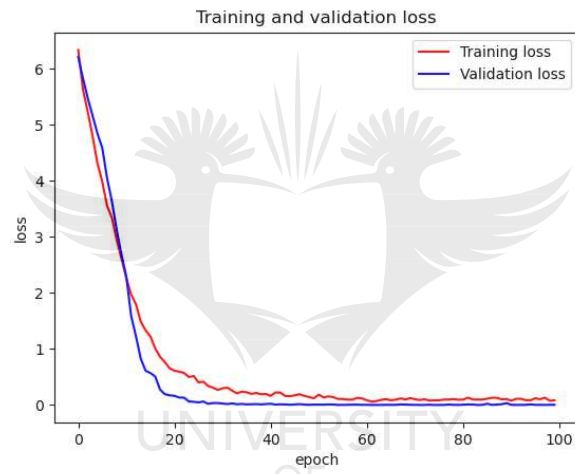


Figure 6.25: ResNet50 Loss Curve

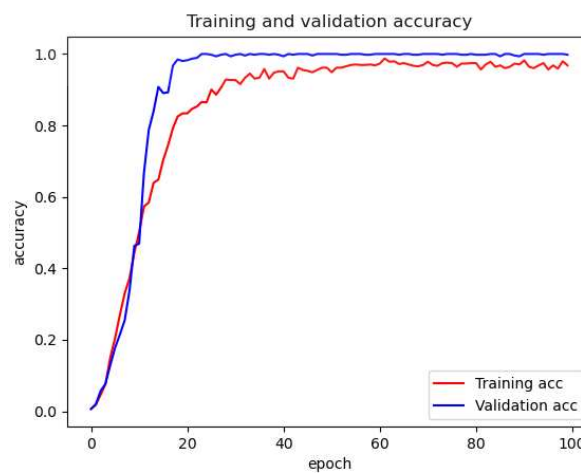


Figure 6.26: ResNet50 Accuracy Curve

### 6.3.6 Deep Hybrid Learning Pipeline

The deep hybrid learning pipeline made use of the One Millisecond Face Alignment algorithm, ResNet50 as the feature extractor and SVM for classification. The metrics presented for the deep hybrid learning pipeline with SVM indicates that this approach yielded promising results with an accuracy of 98.90%. A precision score of 98.85% and recall score of 99.22% demonstrates that the model achieved a high percentage of relevant results. Furthermore, an f1 score of 99.03% was achieved which reveals that the classifier was able to make accurate predictions in terms of relevant instances. A low EER rate of 0.79% was produced which suggests that the model is able to make accurate. These results can be seen in table 6.6.

The ROC curve for the deep hybrid learning pipeline is plotted in figure 6.27 which shows a favourable ROC curve representation. A micro-average and macro-average score of 97% was achieved. The PR curve which can be seen in figure 6.28 achieved a micro-average of 98.2% indicates that the classifier is returning accurate results (high precision) and returning majority of all positive results (high recall).

Table 6.6: Deep Hybrid Learning Pipeline Results

| Accuracy | Precision | Recall | F1 Score | EER   |
|----------|-----------|--------|----------|-------|
| 98.90%   | 98.85%    | 99.22% | 99.03%   | 0.79% |

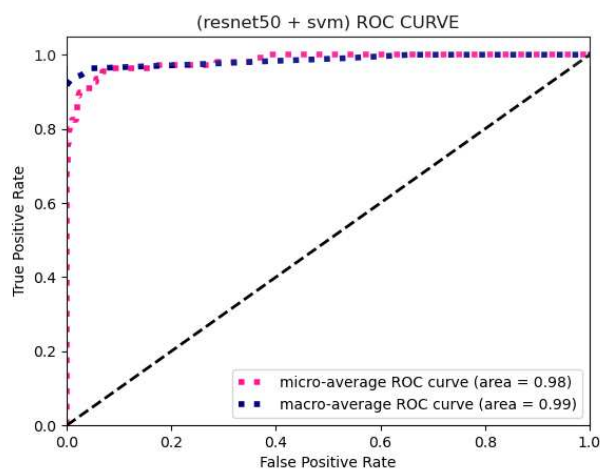


Figure 6.27: Deep Hybrid Learning Pipeline ROC Curve

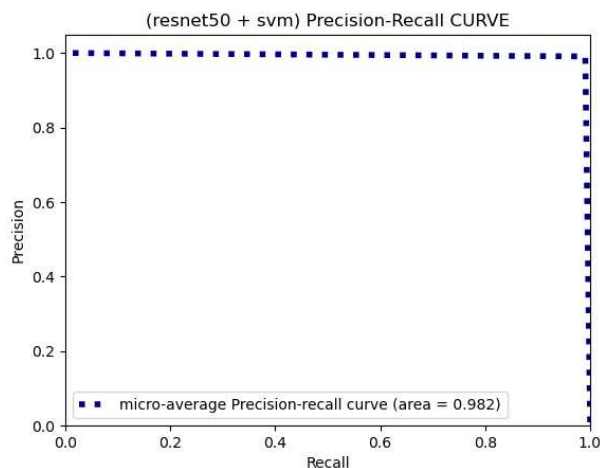


Figure 6.28: Deep Hybrid Learning Pipeline PR Curve

## 6.4 Discussion on Findings

By observing the results discussed above, several interesting findings can be extracted. Firstly, two methods were tested for object detection: Haar cascade and the One Millisecond Face Alignment algorithm. In this case the concept of object detection was to detect or localise the lip within the image frame. Since this was the most crucial step in the different pipelines, which have been discussed chapter 5, the method with the least number of false positive detections was adopted. The Haar cascade method detected every face in the image, however it failed to ideally detect the lip from the face. An ideal detection of both the face and the lip was almost completely rare. The best results that were achieved comprised of a detected lip with one or two false positives. At times, the hair was misclassified as a lip while at other times the eyebrow or the top corner of the nose was misclassified as a lip. While the Haar cascade method achieved fewer ideal outcomes the One Millisecond Face Alignment algorithm achieved desirable results. Kopp et al. [92] state in their facial landmark detection paper that this method performs reasonably well in terms of accuracy. However, one interesting finding that should be noted is that the algorithm did not perfectly detect the lips that were physically larger in size (larger than the average) as compared to the other detections.

Secondly, different computer vision methods were used in each pipeline to achieve lip print identification. Another interesting finding that can be noticed from the different pipelines is

that the lips that were incorrectly predicted were often misclassified with the opposite gender. Generally, the lips that were mismatched were in some way similar in size and shape. Thus, if a male lip was similar in size and shape it was misclassified with a female lip equivalently small in size and shape. One reason for this could be because of the lack of grooves or patterns visible. Although the images were high in resolution, all the grooves or patterns on the lips might not have been easily visible and therefore, were misclassified.

For the broader category of the traditional pipeline, two potential reasons for misclassification could be due to the lack of specificity (lack of grooves visible on the lips) or it could be caused by the facial landmark detector when detecting the lips. Despite some misclassifications, the results obtained by applying SURF are substantial enough to achieve lip print identification. The metric for the SVM pipeline were quite promising. The Support Vector Machine variation achieved a recognition accuracy of 95.83% and the K-Nearest Neighbour variation achieved an accuracy of 93.75%. The K-NN variation had the highest number of inaccurate predictions across all the experiments. A summary of the results is shown in table 6.7.

A more modern pipeline that made use of the convolutional neural network approach was VGG16 variation and the ResNet50 variation. The VGG16 classifier had the second highest number of inaccurate predictions across all the experiments. As discussed above, some lips were similar in size and shape. It is therefore significant that the CNN would have trouble distinguishing between these lips because it may have struggled to extract features that were distinct to those lips. The VGG16 variation achieved a recognition accuracy of 94.51% which is a good result, but it is lower than the accuracy of the other classifiers. The ResNet50 deep learning pipeline generally classified most lips accurately; with only one misclassification. The results of the ResNet50 deep learning pipeline were the most desirable. This approach outperformed the other pipelines with an accuracy score of 99.78%. The EER for this classifier was very low which suggest that the model is accurate and that is possible to achieve lip print identification by adopting this approach.

The effectiveness of the ResNet50 architecture with regards to lip print identification can also be seen in the deep hybrid learning pipeline where it acts as a feature extractor for the pipeline before classification. A Support Vector Machine was used as the classifier for this pipeline. The deep hybrid learning pipeline achieved the best results after the ResNet50 deep learning pipeline with a total of only four misclassifications. A good recognition

accuracy of 98.90% was achieved, which suggests that a hybrid model has potential in the realm of lip print recognition.

Table 6.7: Summary of results obtained

| Pipeline Variation | Accuracy      | Precision     | Recall        | F1 Score      | EER          |
|--------------------|---------------|---------------|---------------|---------------|--------------|
| Traditional SVM    | 95.83%        | 97.56%        | 95.93%        | 96.74%        | 1.42%        |
| Traditional K-NN   | 93.75%        | 96.34%        | 94.71%        | 95.52%        | 2.23%        |
| VGG16              | 94.51%        | 92%           | 94%           | 93%           | 2.14%        |
| <b>ResNet50</b>    | <b>99.78%</b> | <b>99.66%</b> | <b>99.77%</b> | <b>99.70%</b> | <b>0.21%</b> |
| ResNet50 + SVM     | 98.90%        | 98.85%        | 99.22%        | 99.03%        | 0.79%        |

Based on the results discussed above, it can be concluded that the ResNet50 deep learning pipeline with an accuracy of 99.78% and the deep hybrid learning pipeline (ResNet50 feature extractor and SVM classifier) with an accuracy of 98.90%, performed the best in terms of accuracy out of all the pipelines which can be seen in table 6.7. These pipeline variations also obtained the lowest EER of 0.21% and 0.79%, respectively which is extremely promising. The next section will compare the results of the current study with the performance of other similar systems.

## 6.5 Comparison with Similar Systems

Table 6.8 shows three similar systems that dealt with lip print recognition and the overall accuracy scores achieved by these systems. The first system proposed in 2017 used facial landmark points to extract contours of the lips from three different datasets (Multi-PIE, PUT and a local database). In the second step, features extracted from the ROI were modelled by image processing algorithms. The extracted features were then used the input to the neural network. The second system did not make use of machine learning algorithms. Instead, a new approach, based on lip print furrow-based patterns was used to create different patterns on the lips. The prints were divided into upper and lower lips and the

furrows were made visible which were then parametrized and stored in a database. The third system which is a more recent work in lip print recognition, used local binary patterns and various classifiers such as SVM, K-NN and ANN to recognize lips. The overall accuracies for this system are shown in the table below. It is evident that the results discussed in section 6.4 indicate that higher accuracy scores were achieved by this study in comparison with similar systems shown in table 6.8.

Table 6.8: Comparison of similar systems for lip print recognition

| Similar Systems  | Method Description   | Dataset and Sample Count   | Accuracy  |
|--|--|--|---|
| Wrobel et al. [36]<br>Probabilistic Neural<br>Network for lip print<br>recognition<br>2017         | Three different datasets were used (Multi-PIE Face, PUT and local database). Facial landmark points were used to extract the contours of the lips. PNN was then used for classification. | Multi-PIE – 337 samples<br>PUT – 100 samples<br>Local – 50 samples | 1. Multi-PIE: 86.95%<br>2. PUT: 87.14%<br>3. Local: 87.26%                      |
| Wrobel et al. [35]<br>Lip print furrow-based<br>patterns 2018                                      | A new approach which creates lip patterns for each individual using furrow-based patterns  | Local – 50 samples   | 92.73%  |
| Sandhya et al [37]<br>Machine learning<br>algorithms for lip print<br>based identification<br>2021 | Local binary patterns and shape related features are used to extract features. SVM, K-NN, ensemble classifiers and ANN used for classification.  | University of Silesia<br>database – 15 samples                     | 1. SVM: 81.84%<br>2. K-NN: 80%<br>3. Ensemble classifier: 97%<br>4. ANN: 85.81% |

When lip print recognition is compared to well established biometric modalities such as fingerprint and face recognition, it is evident there is room for improvement. In 2014 Kaur et al. [93] compared the performance of different biometric systems such as fingerprint, face, iris, and hand geometry. The EER of these systems were compared. Fingerprint recognition in 2014 had an EER of 2%. Iris recognition had an EER of 0.01% while hand geometry recognition had an EER of 1%. Because these systems are used in a large-scale real-life environment and have been around for quite some time, the EER of these systems are very low. Consequently, lip print recognition at this stage does not give results comparable with face or fingerprint systems because it has not been used in a large-scale

environment and has only recently gained attention. Therefore, there is still much needed within the realm of lip print recognition. However, new biometric modalities should still be investigated to improve the accuracy and reliability of hybrid systems.

## 6.6 Conclusion

The chapter begins by discussing the results of the functional requirements which were outlined in chapter 5. The functional requirements included selecting an appropriate dataset, detecting the lip within an image frame and identifying the target based on their lip print. The selection of an appropriate dataset was thoroughly discussed and highlighted that those datasets produced in a controlled environment would be better suited for the current study rather than those produced in uncontrolled environments. The results of the two methods used to perform lip detection (One Millisecond Face Alignment and the Haar cascade algorithm) were then discussed and the results established that the performance of the One Millisecond Face Alignment algorithm outperformed the Haar cascade method with a greater accuracy in detecting the lips. Thereafter, the performance of identifying the lips and the number of misclassifications obtained were outlined which led to the results of the non-functional requirements.

The results of the non-functional requirements addressed the performance of the different implementations of the prototype, and it demonstrated how they performed in achieving the objectives and benchmarks set out for the current study. Each set of results obtained were discussed using the different metrics discussed in chapter 5. The different implementations of the prototype (traditional, deep learning and deep hybrid-based learning) yielded promising results. The ResNet50 architecture achieved the highest accuracy of 99.78%, followed by the hybrid architecture with an accuracy of 98.90%. Ultimately, the prototype is able to use both traditional and deep learning-based methods to achieve lip print identification with promising results.

Since the results of the current study have suggested that the implemented system is able to perform lip print identification, it is fair to claim that **RO4** and **RO5** of the research study has been accomplished. Therefore, it is worthwhile to investigate the findings, limitations as well as future work for the current study. The next chapter will therefore investigate the significant findings that came out of this research as well as the impact and potential



improvements that could be made to the system. These insights will draw the final conclusions of the current study and determine whether the objectives of the study have been met.



# Chapter 7 Conclusion

## 7.1 Introduction

The previous chapter presented and discussed the results of the lip print recognition experiments, by evaluating the prototype using a set of benchmarks, thereby allowing the final conclusions of the study to be drawn. The potential of using lip prints to identify individuals using computer vision methods is realised and can be analysed further to gain further insight.

This chapter highlights the crux of the study by briefly describing the objectives, summary of the study along with its findings. Furthermore, the objectives of the study can be aligned with the results obtained in chapter 5 and the advantages as well as the limitations of the model can be analysed. More so, the most significant findings and insights will be explored to determine the value of the implemented model. This chapter will also consider what the future work for this domain entails.

The final chapter begins by revisiting the objectives of the study stated in chapter 1 (section 7.2), followed by a summary of the study in section 7.3. Thereafter, the important findings of the study are highlighted in section 10.4 and the overall impact of the study is discussed in section 7.5. Future work of the study is conducted in section 7.6. The study concludes with an overall conclusion in section 7.7.

## 7.2 Objectives of Study

Setting objectives and determining whether these objectives have been achieved is one of the most crucial aspects of the study. This is because objectives determine the scope, depth and overall, the direction of the study. This chapter will outline the objectives of the current study and expand on how the study has addressed each objective.

## Research Objective 1

Research Objective 1 (**RO1**) – Conduct a literature review within the research domain to identify the problem areas and relevant computer vision methods which can be used to achieve lip print identification along with appropriate datasets that can be employed.

This objective is achieved by conducting a literature review, which highlights the research problem as lip print recognition. The literature review discussed various elements that facilitate access control and biometrics (chapter 2). By illuminating the environment and domain of the current study, the applicability and potential work that resides in the area is determined. It also emphasized on the importance of using the lip as a biometric modality. Chapter 3 investigated the different methods that could be used to achieve lip print recognition as well as similar work that has been conducted in this domain. Therefore, the literature review has placed the current topic within a contemporary context and demonstrated knowledge on the area of focus, compared different methods and techniques. Furthermore, it revealed certain trends and gaps that were prevalent in existing literature.

## Research Objective 2

Research Objective 2 (**RO2**) – Adopt a high-resolution face dataset which will allow for discriminatory lip features to be extracted.

This objective is achieved by employing the Chicago Face Database (CFD) for the current study. The CFD provides high-resolution ( $2444 \times 1718$ ) images of male and female targets of different ethnicities. Due to its high-resolution, the patterns on the lips are visible and therefore, it was employed for the current study. The dataset allowed the study to proceed and enabled the development of the model. The results also confirmed the validity of the choice in the dataset.

## Research Objective 3

Research Objective 3 (**RO3**) – Create experiments based on existing literature and findings by the author that can be used to achieve lip print identification.

The model developed for the current study was unpacked in chapter 5. Effective methods, based on current literature, were determined. These methods were grouped into traditional machine learning and deep learning approaches. Consequently, suitable computer vision pipelines were implemented. Hence, the proposed lip print recognition model indicates that the study has accomplished Research Objective **RO3**.

## Research Objective 4

Research Objective 4 (**RO4**) – Implement a prototype based on the experiments which can recognise individuals based on their lip prints by employing computer vision methods from the designed model.

The model is implemented in chapter 5 (section 5.3). For the current study three pipeline approaches were implemented namely, a traditional machine learning-based approach, a deep learning-based approach, and a deep hybrid learning-based approach. Therefore, this study has consequently achieved **RO4**.

## Research Objective 5

Research Objective 5 (**RO5**) – Validate the performance of the prototype to determine its feasibility and report on these results in research articles and the dissertation.

The implemented prototype of the study is evaluated in chapter 6 based on a set of benchmarks outlined in chapter 5. The system yields promising results which indicates that the implemented model can accomplish lip print recognition. Chapter 5 effectively demonstrates the manner in which the prototype is able to validate the model's feasibility. Therefore, Research Objective **RO5** has been achieved.

## 7.3 Summary

This research aimed to determine whether traditional and deep learning methods are conducive to lip print recognition. By doing so the researcher adopted a positivist-based quantitative approach to address the research problem and achieve the objectives of the study. As a result, this study has considered various aspects of access control, biometrics,

and lip biometrics. The applicability of identifying the lips using traditional and deep learning methods was investigated, and a research problem was defined.

In chapter one, the research problem of this study was defined as the lack of research and approaches, particularly deep learning-based methods, for consolidating lip prints to achieve lip print identification. Therefore, it is for this reason that the current study was conducted. Furthermore, a hypothesis and objectives were devised following the research problem. A set of assumptions and constraints were established for the study to ensure consistency in the data for the quantitative based research approach used to address the research problem.

The lack of research in the domain of lip print recognition was investigated, which revealed a wide array of applicable approaches to address the research problem. The literature review of this study considered access control, biometrics, computer vision and lip biometrics to arrive at a feasible model for lip print recognition. The literature review allowed for a comprehensive and accurate understanding of the environment and domain of the current study (chapter 2) as well as the discovery of appropriate computer vision methods to address the research problem. Chapter 3 described the various traditional and deep learning approaches that could potentially be employed for the study. Consequently, these methods were used to construct a model.

The model of the current study was created as a representation of methods that are appropriate for lip print recognition. The methods selected for the implementation of the prototype were discussed in detail in chapter 5. The different approaches for lip print recognition pipelines were therefore clearly laid out by the model. Thereafter, the study progressed with the implementation of the prototype where each pipeline was discussed thoroughly. The chapter also outlined a set of benchmarks. Because the methodology for this study follows a positivist-based quantitative approach, the hypothesis had to be tested. In order to adequately test the hypothesis of the current study, a set of benchmark tests were necessary which were used to gather results from the model.

The last part of the study, chapter 6, presented the results and metrics of the implemented system. The benchmark consisted of functional and non-functional requirements tests which were assessed. The results critically analysed the implemented model and valuable insights were gained into the strengths of each pipeline but more importantly the limitations of each pipeline were also revealed. The study successfully addressed research objective **RO5**,

which led to the validation of the hypothesis. The significant findings from these results are highlighted in the next section.

## 7.4 Findings

When observing the results presented in chapter 6, several interesting findings can be extracted. However, before discussing the significant findings of the study, the limitations of the study should be considered as well. Therefore, this section will highlight the findings of the study along with its limitations.

### 7.4.1 Limitations

Several problems arose during data collection. An insufficient number of datasets was one of the more notable limitations. Due to this reason, finding a suitable dataset that would address the research problem adequately was a challenging task. The literature studies highlighted the limited number of high-resolution public face datasets available in the research domain. The Yale Face Database was one consideration before finally deciding to employ the Chicago Face Database. One reason for this was because the Yale Face Database only contains 165 images of 15 subjects. Therefore, the limited number of study samples would not be an effective representation of the broader population. Other face datasets which were considered for the study included Faces95 and the SUT-Lips-DB database. Faces95 has a total of 75 samples and the quality of the images are poor. Therefore, it would be difficult to achieve lip recognition using this dataset. The SUT-Lips-DB database contains images of acquired lip prints using a traditional approach (lip stick or colouring agents). Thus, this dataset is not suited for the current study because in real-life scenarios it would be more ideal for a sensor to capture the lips from a specific distance and perform identification in a non-invasive manner rather than having the individual press their lips on cellophane tape or a piece of paper.

The next notable limitation was the lack of prior research studies on the topic. It is useful to investigate previous work that has been conducted in the research area because it helps identify what has been achieved in the research area thus far. Since lip print recognition is still in its early stage, deep learning methods have not been introduced to the field as yet.

Therefore, addressing deep neural architectures was a challenge. Consequently, the results achieved using the deep learning-based and deep hybrid learning-based approaches cannot adequately be used to compare the results of previous work. Another limitation which can be noted is the lack of facial hair on the male targets. It is important to analyse how a facial landmark detector will perform in the presence of various backgrounds such as a moustache because it makes the system robust to these changes.

Despite some limitations, the study was able to adequately address the research problem with promising results. The results showed various trends that will be discussed below.

#### 7.4.2 Trends

In terms of object detection, the One Millisecond Face Alignment algorithm adequately detected the lips within the image frame compared to the Haar cascade classifier. The main reason for its high accuracy and mAP of 93% is because the Chicago Face Database contains images that were produced in a highly controlled environment with consistent lighting. Therefore, these images made it suitable for the detector to detect the lip of each individual in the dataset. The CFD is also diverse in nature and therefore the results can potentially generalise better. The Haar cascade classifier however, performed poorly despite the quality of the images. In some instances, the lip was not detected at all. Thus, an accurate detection of the lip was rare.

From the results of the study, the pipeline that stands out in terms of its accuracy is the ResNet50 architecture approach which achieved a 99.78% recognition accuracy and an EER of 0.21%. The next best pipeline in terms of its accuracy is the ResNet50 architecture with an SVM approach. This pipeline achieved an EER of 0.79% with an accuracy of 98.90%. The traditional approach which used SURF as the feature extractor and SVM as the classifier yielded promising results with a recognition accuracy of 95.83% and an EER of 1.42%. The VGG16 architecture achieved an accuracy of 94.51% with an EER of 2.14% which are also promising results. The pipeline that produced the lowest accuracy was the traditional K-NN approach which produced a 93.75% recognition accuracy. These results suggest that there is value in using deep learning methods for lip print recognition.

The most accurate pipeline of this study made use of the ResNet50 architecture. This study has, therefore, successfully utilised deep learning approaches for the domain of lip print

recognition. Based on current literature, there are no deep learning-based implementations for lip print recognition. This study contributes uniquely by applying deep learning approaches to the research area. The successful results of the ResNet50 pipeline indicate that it is feasible for lip print recognition. The next best pipeline, which used the ResNet50 architecture to extract features and an SVM classifier, was a deep hybrid learning-based approach. The results of this pipeline indicate that a combination of deep learning for feature extraction with traditional classifiers is also a feasible approach for lip print recognition.

Having highlighted the trends and limitations of the current study, the potential impact and contributions can now be realised. These will be discussed in the section to follow.

## 7.5 Impact

The results obtained from the study demonstrate the potential benefits and contributions of the domain and its related fields. The number of biometric modalities that are available is rather large. Some of these include fingerprints, palmprints, face, iris, retinal, hand geometry, gait, keystroke, and signature [9]. These modalities are an integral part of the human body, meaning that they are subject to change over time. For example, a fingerprint may deteriorate, a voice may be altered by a person's state of health and the face changes over time [118]. Consequently, the presented modality may be refused, even if the individual is who they claim to be. Therefore, new modalities should be investigated in order to mitigate these challenges. Hence, this study can be used to alleviate some of the challenges mentioned above by employing the lip as a biometric modality. The results of the study demonstrate that the lip has the potential of being treated as an alternative biometric modality. Lip biometrics can be used to enhance the effectiveness of well-known biometrics such as face recognition, by its implementation in multimodal or hybrid systems [97].

It has been noted that there is a lack of research that determines which algorithms are conducive to lip print recognition. The literature review has confirmed that deep learning methods in the realm of lip print recognition has not been explored. With biometric recognition, shifting from traditional methods to deep learning methods, the results obtained demonstrate the feasibility of employing deep learning methods such as convolutional neural networks for lip print recognition. Deep learning methods have been successful in achieving state-of-art results in biometric recognition for fingerprint, face, iris, ear, palmprint and gait



[119]. Therefore, although, lip print recognition is still in its early stage, deep learning methods for lip print recognition can produce promising results as exhibited by this study.

In addition to the provision of some future direction for future research, this study has also contributed to the literature on lip print recognition. The research of this study in lip print recognition, which makes use of computer vision techniques, provides a good basis for further explorations that aim to link and investigate the fields of biometrics, computer vision and lip print recognition.

## 7.6 Future Work

Currently, this study employed a high-resolution face dataset to achieve lip print recognition. This work can be further extended to incorporate low quality and blurred images to enhance the reliability of the designed model. New developments in the field, such as lip print sensors, along with current avenues of research related to the study and testing algorithms on recently released contact-based datasets can be used to further improve the model, thus paving way for lip print biometric systems. Unlike other well established biometric modalities, there are no standard techniques for lip print recognition. The absence of research in this area, particularly deep learning-based research, has inadvertently contributed to the limited acceptance of lip prints despite the probative potential that it can be used as a biometric modality [121]. Hence, future research should be targeted towards addressing these limitations.

On the technical side of this study, improvements can be made in terms of the object detection approaches used. Sandhya et al [120] hinted at the fact that a standard and uniform methodology is required to acquire lip prints. Because manual methods are prone to human error [120], automated methods for the acquisition of lip prints should be investigated. This study has proved that using the One Millisecond Face Alignment algorithm can be used to adequately detect the lip within the image. However, more sophisticated methods worth exploring are deep learning methods such as YOLO, R-CNNs, SSD and semantic segmentation-based methods. These approaches have produced state-of-the-art results in recent research works. The manuscripts have provided a starting point for the advancement of deep learning research for lip print identification. Additionally, more experiments for detailed ablation studies can be conducted such as modifying the One

Millisecond Face Alignment algorithm to detect lips physically larger in size or examining the impact of the proposed model on additional face datasets since the CFD is relatively small. Another improvement that can be made is to detect the entire face region first before detecting the lips.

Many approaches could be utilised in the future work for the improvement and expansion of the current solution, thereby making the lip print recognition system in this study a basis for further discoveries and future work within the field. The lip biometric has fewer research compared to other popular biometrics. It is relatively new and hence offers many research possibilities. Therefore, there are still many gaps in this area that can be researched and explored. One of the evident gaps within this domain is unconstrained lip recognition. Lip print recognition has not been explored in the unconstrained environment or in the wild. Therefore, there is a need to explore the power of deep learning-based methods to develop effective and efficient lip recognition methods in real-life scenarios.

## 7.7 Conclusion

In many ways biometrics is an ideal way to identify and authenticate human beings. When comparing biometric recognition with other identification and authentication methods, biometric recognition deals efficiently with the limitations and disadvantages associated with these methods such as a forgotten password or a lost token. Today, fingerprint, face, iris but also DNA, retina, palm-vein, and body odour can be used in biometric recognition. However, there is still much room for improvement with respect to new modalities for biometric recognition. In a time where individuals demand flawless security measures that are convenient, yet secure. Biometric recognition systems have carved a niche for itself, and the current research trends have shown the prospects of using lip prints as a measure for human identification.

Lip prints, a unique physiological characteristic of the human body, which can be used as a biometric modality has not been extensively researched. Therefore, this study has aimed to address this by applying different computer vision methods to achieve lip print identification and more importantly this study has pushed the field forward by employing deep learning architectures to lip print identification. This dissertation successfully demonstrates the

applicability of using deep learning architectures for lip print identification. Furthermore, the study was able to identify a problem area as well as a research gap within the research domain and consequently, implemented computer vision methods that can identify different lip prints. The field of biometrics has come a long way since its early adoption, but there is still much room for growth. Should lip print systems become more prevalent, the developed solution has the potential to positively affect the future of lip biometrics.

*"Biometrics is a slow-moving train, but a train, nonetheless. This may accelerate it" – Peter Browne*



## References

- [1] N. B. Sukhai, "Access control & biometrics," in *InfoSecCD '04*, 2004, pp. 124-127.
- [2] J. Vacca, "Access Controls," in *Cyber Security and IT Infrastructure Protection*, Elsevier Inc., 2014, pp. 269-280.
- [3] S. Boonkrong, *Authentication and Access Control: Practical Cryptography Methods and Tools*, Apress, 2021, pp. 45-70.
- [4] J. M. Kizza, "Authentication," in *Computer Network Security*, Springer, Boston, MA, 2005, pp. 233-256.
- [5] H. Vallabh, "Authentication using Finger-Vein Recognition," MSc Dissertation, 2012.
- [6] N. Lal, S. Prasad and M. Farik, "A Review Of Authentication Methods," *International Journal of Scientific & Technology Research*, vol. 5, no. 11, pp. 246-249, 2016.
- [7] F. Belhadj, "Biometric system for identification and authentication," PhD Thesis, 2017.
- [8] I. Alsaadi, "Physiological Biometric Authentication Systems, Advantages, Disadvantages And Future Development: A Review," *International Journal of Scientific & Technology Research*, vol. 4, no. 12, pp. 285-289, 2015.
- [9] A. K. Jain, P. Flynn and A. A. Ross, *Handbook of Biometrics*, Springer, 2008.
- [10] J. Yang, Y. Chen, C. Zhang, D. S. Park and S. Yoon, "Introductory Chapter: Machine Learning and Biometrics," in *Machine Learning and Biometrics*, IntechOpen, 2018.
- [11] I. Adjabi, A. Ouahabi, A. Benzaoui and A. Taleb-Ahmed, "Past, Present, and Future of Face Recognition: A Review," *Electronics*, vol. 9, no. 8, 2020.
- [12] E. R. Davies, "Circle and Ellipse Detection," in *Computer and Machine Vision: Theory, Algorithms, Practicalities*, Academic Press, 2012, pp. 303-331.
- [13] A. Kaushal and M. Pal, "Cheiloscopy: A Vital Tool in Forensic Investigation for Personal Identification in Living and Dead Individuals," *International Journal of Forensic Odontology*, vol. 5, no. 2, pp. 71-74, 2020.
- [14] S. A. Ahmed, H. E. Salem and M. M. Fawzy, "Forensic dissection of lip print as an investigative tool in a mixed Egyptian population," *Alexandria Journal of Medicine*, vol. 54, no. 3, pp. 235-239, 2018.
- [15] L. V. K. Reddy, "Lip prints: An Overview in Forensic Dentistry," *Journal of Advanced Dental Research*, vol. 2, no. 1, pp. 17-20, 2011.
- [16] N. Ganapathi, J. Dineshshankar, Y. Ragnathan, A. Ravi, M. S. Kumar and R. Aravindhan, "Lip prints: Role in forensic odontology," *Journal of Pharmacy and Bioallied Sciences*, vol. 5, no. 1, pp. S95-S97, 2011.

- [17] N. Ishaq, E. Ullah, I. Jawaad, A. Ikram and A. Rasheed, "Cheiloscopy: A Tool For Sex Determination," *The Professional Medical Journal*, vol. 21, no. 5, pp. 883-887, 2014.
- [18] S.-L. Wang and A. Wee-ChungLiew, "Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power," *Pattern Recognition*, vol. 45, p. 3328–3335, 2012.
- [19] P. I. Wilson and J. Fernandez, "Facial Feature Detection Using Haar Classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127-133, 2006.
- [20] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867-1874.
- [21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson, 2018.
- [22] L. G. Shapiro and G. C. Stockman, "Filtering and Enhancing Images," in *Computer Vision*, Pearson, 2001.
- [23] S. Krig, *Computer Vision Metrics: Survey, Taxonomy, and Analysis*, Apress, 2014.
- [24] T. Ojala, M. Pietikäinen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.
- [25] C. C. Aggarwal, *Data Classification: Algorithms and Applications*, New York: CRC Press, 2014.
- [26] M. Nielsen, *Neural Networks and Deep Learning*, 2019.
- [27] P. Sharma, S. Deo, S. Venkateshan and A. Vaish, "Lip Print Recognition for Security Systems: An Up-Coming Biometric Solution," *Intelligent Interactive Multimedia Systems and Services*, vol. 11, pp. 347-359, 2011.
- [28] Ę. Smacki, P. Porwik, K. Tomaszycy and S. Kwarciańska, "The lip print recognition using Hough Transform," *Journal of Medical Informatics and Technologies*, 2010.
- [29] K. Wróbel, R. Doroz and M. Palys, "Lip Print Recognition Using Bifurcation Analysis," *Lecture Notes in Computer Science*, vol. 2, pp. 72-81, 2015.
- [30] K. Wrobel and W. Froelich, "Recognition of Lip Prints Using Fuzzy C-Means Clustering," *Journal of Medical Informatics & Technologies*, vol. 24, pp. 67-74, 2015.
- [31] L. Smacki, K. Wrobel and P. Porwik, "Lip print recognition based on DTW algorithm," *Third World Congress on Nature and Biologically Inspired Computing*, pp. 594-599, 2011.
- [32] P. Porwik and T. Orczyk, "DTW and Voting-Based Lip Print Recognition System," *11th IFIP TC 8 International Conference on Computer Information Systems and Industrial Management*, pp. 191-202, 2012.
- [33] S. K. Bandyopadhyay, S. Arunkumar and S. Bhattacharjee, "Feature Extraction of Human Lip Prints," *Journal of Current Computer Science and Technology*, vol. 2, no. 1, pp. 1-8, 2012.

- [34] S. Bakshi, R. Raman and P. Sa, "Lip print recognition based on local feature extraction," 2011.
- [35] K. Wrobel, R. Doroz, P. Porwik and M. Bernas, "Personal Identification utilizing lip print furrow based patterns. A new approach," *Pattern Recognition*, vol. 81, pp. 585-600, 2018.
- [36] K. Wrobel, R. Doroz, P. Porwik, J. Naruniec and M. Kowalski, "Using a Probabilistic Neural Network for lip-based biometric verification," *Engineering Applications of Artificial Intelligence*, vol. 64, no. C, pp. 112-127, 2017.
- [37] S. Sandhya, R. Fernandes, S. Sapna and A. Rodrigues, "Comparative analysis of machine learning algorithms for Lip print based person identification," *Evolutionary Intelligence*, 2021.
- [38] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun and D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," 2019.
- [39] W. Zhao and R. Chellappa, *Face Processing Advanced Modeling and Methods*, Academic Press, 2005.
- [40] C. Williams, "Research Methods," *Journal of Business & Economic Research*, vol. 5, no. 3, pp. 65-72, 2007.
- [41] C. R. Kothari, *Research Methodology Methods and Techniques*, New Delhi: New Age International Publishers.
- [42] L. Doyle, A.-M. Brady and G. Byrne, "An overview of mixed method research," *Journal of Research in Nursing*, vol. 14, no. 2, pp. 175-185, 2009.
- [43] T. S. Kuhn, *The Structure of Scientific Revolutions (2nd Edition)*, University of Chicago Press, 1970.
- [44] C. Kivunja and A. B. Kuyini, "Understanding and Applying Research Paradigms in Educational Contexts," *International Journal of Higher Education*, vol. 6, no. 5, pp. 26-41, 2017.
- [45] M. DeCarlo, "Paradigms, theories, and how they shape a researcher's approach," in *Scientific Inquiry in Social Work*, Open Social Work Education, 2018.
- [46] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333-339, 2019.
- [47] S. Gregor and A. R. Henver, "Positioning And Presenting Design Science Research For Maximum Impact," *MIS Quarterly*, vol. 37, no. 2, pp. 337-355, 2013.
- [48] S. Gilliland, "Towards a framework for managing enterprise architecture acceptance," 2014.
- [49] E. Daniel, "The Usefulness of Qualitative and Quantitative Approaches and Methods in Researching Problem-Solving Ability in Science Education Curriculum," *Journal of Education and Practice*, vol. 7, no. 15, pp. 91-100, 2016.
- [50] D. S. Ma, J. Correll and B. Wittenbrink, "The Chicago face database: A free stimulus set of faces and norming data," *Behavior Research Methods*, vol. 47, pp. 1122-1135, 2015.

- [51] J. P. Tripathy, "Secondary Data Analysis: Ethical Issues and Challenges," *Iranian Journal of Public Health*, vol. 42, no. 12, pp. 1478-1479, 2013.
- [52] E. A. Drost, "Validity and Reliability in Social Science Research," *Education Research and Perspectives*, vol. 38, no. 1, pp. 105-124, 2011.
- [53] J. Ioannidis, "Why most published research findings are false," *PLoS Med*, vol. 2, no. 8, pp. 0696-0701, 2005.
- [54] J. Smith and H. Noble, "Bias in research," *Evidence Based Nursing*, vol. 17, no. 4, pp. 100-101, 2014.
- [55] B. K. Nayak, "Understanding the relevance of sample size calculation," *Indian J Ophthalmol*, vol. 58, pp. 469-470, 2010.
- [56] U. Sekaran, "Sampling," in *Research Methods for Business 4th Edition*, Hermitage Publishing Services, 1984, pp. 263-298.
- [57] H. Mohajan, "Two Criteria for Good Measurements in Research: Validity and Reliability," *Annals of Spiru Haret University*, vol. 17, no. 3, pp. 58-82, 2017.
- [58] P. Tubaro, "Data Big and Small," 15 10 2015. [Online]. Available: <https://databigandsmall.com/2015/10/18/research-ethics-in-secondary-data-what-issues/>. [Accessed 20 02 2021].
- [59] J. N. Pato and L. I. Millet, "Cultural, Social and Legal Considerations," in *Biometric Recognition: Challenges and Opportunities*, Washington, D.C, The National Academies Press, 2010, pp. 85-115.
- [61] D. Krishna, F. Talukdar and R. Laskar, "Improving Face Recognition Rate with Image Preprocessing," *Indian Journal of Science and Technology*, vol. 7, no. 8, pp. 1170-1175, 2014.
- [62] G. Kumar and P. K. Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems," in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, Rohtak, India, 2014, pp. 5-12.
- [63] Y. Wu and Q. Ji, "Facial Landmark Detection: a Literature Survey," *International Journal on Computer Vision*, vol. 127, p. 115–142, 2019.
- [64] L. Kabbai, M. Abdellaoui and A. Douik, "Image classification by combining local and global features," *The Visual Computer*, vol. 35, p. 679–693, 2019.
- [65] H. Bay, T. Tuytelaars and L. V. Gool, "Surf: Speeded up robust features," in *European Conference on Computer Vision*, 2006, pp. 404--417.
- [66] Y. Zhang, R. Jin and Z. H. Zhou, "Understanding bag-of-words model: A statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1, pp. 43-52, 2010.
- [67] L. Auria and R. Moro, "Support Vector Machines (SVM) as a technique for solvency analysis," *DIW Discussion Papers*, vol. 811, 2008.

- [68] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," 2011.
- [69] I. Syarif, A. Prugel-Bennett and G. Wills, "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance," *Telkomnika*, vol. 14, no. 4, pp. 1502-1509, 2016.
- [70] C. M. Ma, W. S. Yang and B. W. Cheng, "How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset," *Journal of Applied Sciences*, vol. 14, pp. 171-176, 2014.
- [71] M. Coskun, O. Yildirim, A. Ucar and Y. Demir, "An Overview of Popular Deep Learning Methods," *European Journal of Technic*, vol. 7, no. 2, pp. 165-176, 2017.
- [72] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Foundations and Trends in Signal Processing*, p. 197–387, June 2014.
- [73] A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.
- [74] M. Burugupalli, *Image Classification Using Transfer Learning and Convolutional Neural Networks*, 2020.
- [75] Z. Chen, *Automatic Detection of Photographing or Filming*, 2020.
- [76] M. Talo, *Convolutional Neural Networks for Multi-class Histopathology Image Classification*.
- [77] P. Simon and U. Vijayasundaram, "Deep Learning based Feature Extraction for Texture Classification," *Procedia Computer Science*, vol. 17, p. 1680–1687, 2020.
- [78] S. Almabdy and L. Elrefaei, "Deep Convolutional Neural Network-Based Approaches for Face Recognition," *Applied Sciences*, vol. 9, no. 20, 2019.
- [79] R. Sharma and K. K. Biswas, "Functional requirements categorization Grounded Theory approach," in *2015 International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)*, Barcelona, 2015.
- [80] H. Rezaatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*, 2019.
- [81] H. Dalianis, "Evaluation Metrics and Evaluation," in *Clinical Text Mining*, Springer, Cham, 2018, pp. 45-53.
- [82] Y. Jiao and P. Du, "Performance measures in evaluating machine learning based bioinformatics predictors for classifications," *Quantitative Biology*, vol. 4, no. 4, p. 320–330, 2016.
- [83] A. Beger, "Precision-Recall Curves," 2016. Available:  
[https://www.andybeger.com/papers/Beger\\_2016\\_PrecisionRecallCurves.pdf](https://www.andybeger.com/papers/Beger_2016_PrecisionRecallCurves.pdf)



- [84] B. Johnston and P. d. Chazal, "A review of image-based automatic facial landmark identification techniques," *Journal on Image and Video Processing*, vol. 86, 2018.
- [85] K. Horak, "Introduction to Learning Curves". Available:  
[http://147.229.71.91/STU/lectures/KH\\_MachineLearning\\_LearningCurves.pdf](http://147.229.71.91/STU/lectures/KH_MachineLearning_LearningCurves.pdf)
- [86] Y. Lu, "Food Image Recognition by Using Convolutional Neural Networks (CNNs)," 2016.
- [87] M. Kuchnik and V. Smith, "Efficient Augmentation via Data Subsampling," in *The International Conference on Learning Representations*, 2019.
- [88] A. Gandhi, "Data Augmentation- How to use Deep Learning when you have Limited Data—Part 2," 2021. [Online]. Available:  
<https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/>
- [89] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 60, 2019.
- [90] B. Johnston and P. d. Chazal, "A review of image-based automatic facial landmark identification techniques," *EURASIP Journal on Image and Video Processing*, 2018.
- [91] A. Rosebrock, "OpenCV Face detection with Haar cascades," pyimagesearch, 21 4 2021. [Online]. Available: <https://www.pyimagesearch.com/2021/04/05/opencv-face-detection-with-haar-cascades/>.
- [92] P. Kopp, D. Bradley, T. Beeler and M. Gross, "Analysis and Improvement of Facial Landmark Detection," 2019.
- [93] G. Kaur and C. K. Verma, "Comparative Analysis of Biometric Modalities," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 4, pp. 603-613, 2014.
- [94] "Anatomy of the Lips, Mouth, and Oral Region", *Online Research Resources Developed at NHGRI*, 2021. [Online]. Available: <https://elementsofmorphology.nih.gov/anatomy-oral.shtml>.
- [95] L. Adamu, "78-Lip Prints: An Emerging Tool for Personal Identification", *Bayero Journal of Biomedical Sciences*, vol. 1, no. 1, pp. 78-87, 2016.
- [96] F. Minhas, "Fingerprint Based Person Identification and Verification", Ph.D, Pakistan Institute of Engineering and Applied Sciences, 2005.
- [97] M. Choras, "Lips Recognition for Biometrics", *Advances in Biometrics*, pp. 1260-1269, 2009. Available: 10.1007/978-3-642-01793-3\_127
- [98] C. Anusha and P. Avadhani, "Object Detection using Deep Learning", *International Journal of Computer Applications*, vol. 182, no. 32, pp. 18-22, 2018. Available: 10.5120/ijca2018918235.
- [99] Scott Krig. *Computer vision metrics: Survey, taxonomy, and analysis*. Apress, 2014.

- [100] T. Lindahl, "Study of Local Binary Patterns", Linkoping University, 2007.
- [101] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition", *Pattern Analysis and Applications*, vol. 9, no. 2-3, pp. 273-292, 2006.
- [102] A. Karim and R. Sameer, "Image Classification Using Bag of Visual Words (BoVW)", *Al-Nahrain Journal of Science*, vol. 21, no. 4, pp. 76-82, 2018.
- [103] A. Kumar, "Machine Learning - Feature Selection vs Feature Extraction - Data Analytics", *Data Analytics*, 2021. [Online]. Available: <https://vitalflux.com/machine-learning-feature-selection-feature-extraction/>.
- [104] A. Parveen, H. Inbarani and E. Kumar, "Performance analysis of unsupervised feature selection methods", *2012 International Conference on Computing, Communication and Applications*, 2012. Available: 10.1109/iccca.2012.6179181.
- [105] S. Kaul, "Region Based Convolutional Neural Networks For Object Detection And In Adas Application Recognition", University of Texas, 2017.
- [106] T. Bezdan and N. Bacanin, "Convolutional Neural Network Layers and Architectures", in *International Scientific Conference On Information Technology And Data Related Research*, 2019, pp. 445-451.
- [107] H. Gu, Y. Wang, S. Hong and G. Gui, "Blind Channel Identification Aided Generalized Automatic Modulation Recognition Based on Deep Learning", *IEEE Access*, vol. 7, pp. 110722-110729, 2019. Available: 10.1109/access.2019.2934354.
- [108] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014.
- [109] A. Khan, A. Sohail, U. Zahoora and A. Qureshi, "A survey of the recent architectures of deep convolutional neural networks", *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455-5516, 2020.
- [110] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Region-based convolutional networks for accurate object detection and segmentation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 142–158, 2015.
- [111] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.
- [112] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. Available: 10.1109/cvpr.2016.91
- [113] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C-Y Fu, and A. C Berg, "SSD: Single Shot MultiBox Detector", *Computer Vision – ECCV 2016*, pp. 21-37, 2016. Available: 10.1007/978-3-319-46448-0\_2
- [114] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN", 2017

- [115] Shaheed, K., Mao, A., Qureshi, I. *et al.*, "A Systematic Review on Physiological-Based Biometric Recognition Systems: Current and Future Trends", *Archives of Computational Methods in Engineering*, 2021
- [116] M. Choras, "Human Lips Recognition", *Advances in Soft Computing*, pp. 838-843, 2007. Available: [10.1007/978-3-540-75175-5\\_104](https://doi.org/10.1007/978-3-540-75175-5_104)
- [117] P. Mlsna and J. Rodríguez, "Gradient and Laplacian Edge Detection", *The Essential Guide to Image Processing*, pp. 495-524, 2009. Available: 10.1016/b978-0-12-374457-9.00019-6
- [118] C. Kiennert, S. Bouzefrane and P. Thoniel, "Authentication Systems", *Digital Identity Management*, pp. 95-135, 2015
- [119] B. Kieffer, M. Babaie, S. Kalra and H. Tizhoosh, "Convolutional Neural Networks for Histopathology Image Classification: Training vs. Using Pre-Trained Networks," *2017 Seventh International Conference on Image Processing Theory, Tools and Applications*, pp. 1-6, 2017.
- [120] S. Sandhya and R. Fernandes, "Lip Print: An Emerging Biometrics Technology - A Review", *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, 2017.
- [121] M. Abedi, C. Afoakwah and D. Bonsu, "Lip print enhancement: review", *Forensic Sciences Research*, pp. 1-5, 2020.
- [122] G. Mabazu Hocquet, "Reconnaissance and Assessment of Iris Features towards Human Iris Classification", Ph.D, University of Johannesburg, 2018.
- [123] Reshmi M. P, V. J Arul Karthick, "Biometric Identification System using Lips", *International Journal of Science and Research (IJSR)*, vol. 2, no. 4, pp. 304 – 307, 2013
- [124] L. V. K. Reddy, "Lip Prints: An Overview in Forensic Dentistry," *J. Adv Dental Research*, vol. 1, pp. 17-20, 2013
- [125] "Why choose Biometrics now in the midst of Covid-19?", *EURONOVATE*, 2021. [Online]. Available: <https://www.euronovate.com/why-choose-biometrics-now-in-the-midst-of-covid-19/>.
- [126] "Non-contact Biometric Identification and Authentication | Vilmate", *Nearshore Software Development Company in Ukraine - VILMATE*, 2021. [Online]. Available: <https://vilmate.com/blog/contactless-biometric-identification/>.
- [127] A. Jain, K. Nandakumar and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities", *Pattern Recognition Letters*, vol. 79, pp. 80-105, 2016. Available: 10.1016/j.patrec.2015.12.013.
- [128] E. Gomez, C. Travieso, J. Briceno and M. Ferrer, "Biometric identification system by lip shape", *Proceedings. 36th Annual 2002 International Carnahan Conference on Security Technology*. Available: 10.1109/ccst.2002.1049223
- [129] K. Fessel, "5 Significant Object Detection Challenges and Solutions", *Medium*, 2021. [Online]. Available: <https://towardsdatascience.com/5-significant-object-detection-challenges-and-solutions-924cb09de9dd>.

- [130] N. Nguyen, T. Do, T. Ngo and D. Le, "An Evaluation of Deep Learning Methods for Small Object Detection", *Journal of Electrical and Computer Engineering*, vol. 2020, pp. 1-18, 2020. Available: 10.1155/2020/3189691
- [131] H. Kim, H. Kim, S. Rho and E. Hwang, "Augmented EMTCNN: A Fast and Accurate Facial Landmark Detection Network", *Applied Sciences*, vol. 10, no. 7, p. 2253, 2020
- [132] A. Rosebrock, "Facial landmarks with dlib, OpenCV, and Python - PyImageSearch", *PyImageSearch*, 2021. [Online]. Available: <https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>
- [133] Mohapatra, B., 2019. Image edge detection techniques. ACCENTS Transactions on Image Processing and Computer Vision, 5(15), pp.15-19.



# APPENDICES



UNIVERSITY  
OF  
JOHANNESBURG

# Appendix A

A manuscript is submitted to IET Biometrics "A comparative analysis of computer vision methods for lip print identification" which is currently under the revision stage. The paper provided a starting point for the advancement of the research necessary to produce this dissertation.



# A comparative analysis of computer vision methods for lip-based identification

Wardah Farrukh, Dustin van der Haar

Cnr University Road and Kingsway Avenue, Academy of Computer Science and Software Engineering, University of Johannesburg, APK Campus, Johannesburg

---

**Abstract.** The concept of biometric identification is centered around the theory that every individual is unique and has distinct characteristics. Various metrics such as fingerprint, face, iris, or retina are adopted for this purpose. Nonetheless, new alternatives are needed to establish the identity of individuals on occasions where the above techniques are unavailable. One emerging method of human recognition is lip-based identification. It can be treated as a new kind of biometric measure. The patterns found on the human lip are permanent unless subjected to alternations or trauma. Therefore, lip prints can serve the purpose of confirming an individual's identity. The main objective of this work is to design robust experiments using computer vision methods that can recognise an individual solely based on their lip prints. This article compares traditional and deep learning computer vision methods and how they perform on a common dataset for lip-based identification. The first pipeline is a traditional method with Speeded Up Robust Features (SURF) features with either an SVM or K-NN machine learning classifier, which achieved an accuracy of 95.45% and 94.31%, respectively. A second pipeline compares the performance of the VGG16 and VGG19 deep-learning architectures. This approach obtained an accuracy of 91.53% and 93.22%, respectively.

*Keywords:* Access Control, Biometrics, Lip Print Identification

---

## 1. Introduction

The concept of biometrics has been around for over a century. It was first used for the purpose of criminal identification. However, it later progressed to the identification of both criminal and police personnel a little later in 1924 [2]. Since then, biometric identification has come under research and development and has made ground-breaking progress. Today, attributes such as fingerprint, face, iris but also DNA, retina, palm-vein, and body odor can be used in biometric identification. Various biometric modalities have and are being investigated due to the need for security at access control environments such as borders, buildings, airports, and online transactions. Therefore, there is still much room for improvement with respect to new modalities for biometric identification.

One of the most interesting and emerging methods of human identification, which originates from criminal and forensic practices, is human lip identification. The human lips are two highly sensitive mucocutaneous folds comprised of skin, muscles, mucous membranes, and sebaceous glands [3]. These lips are covered by lines and patterns called

lip prints. Lip prints can be defined as normal wrinkles or grooves present between the inner labial mucosa and outer skin of the lips [4]. These patterns are identifiable within six weeks of intrauterine life. The fact that lip prints are unique has been confirmed by Tsuchihashi and Suzuki [5]. In their paper, the authors studied lip prints of various subjects and established that each lip print was unique. Based on the patterns found on the lips, they devised a classification method of lip-prints which can be seen in Figure 1.

The study of lip prints in the field of forensic odontology is known as Cheiloscopy which originates from the Greek word “cheilos” meaning lips [6]. The importance of cheiloscopy is linked to the fact that lip prints are unique to each individual and permanent even after death. Therefore, it has the potential to be treated as a biometric measure and qualifies the criteria for being a recognition system. Compared to other biometric modalities such as fingerprint, face or iris, lip biometric features contain both physiological and behavioural characteristics [7]. Physiologically, individuals have distinctive lips since it is hereditary. Alternatively, individuals can also be differentiated by the way they talk even when articulating the same utterance. Therefore, this makes lip-based identification one of the most interesting methods of human identification. However, the fact that physiological or behavioural lip features could be more discriminative than traditional biometric features has not been comprehensively studied.

Lip prints can be recorded using various methods. Some of these methods include applying colouring agents like applying lipstick on the individual’s lip and having them pressed on cellophane tape or a piece of paper [29]. Other methods of recording lip prints include using a finger printer, preferably a roller finger printer, applying conventional fingerprint developing powder or using magna brush with a magnetic powder [29]. One critical issue that should be noted from these methods is the fact that the lip prints are acquired using a contact-based approach where it is disclosed on a durable surface. However, manual methods are mostly error prone [29]. Research that has been conducted in lip-based identification thus far have used manual methods to acquire the lips which were then digitized. Therefore, this research works towards the generation of fully automating the process of lip detection within an image frame and classifying them depending on the patterns extracted from the Region of Interest (ROI) using various classifiers such as Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), VGG16 and VGG19. The current study focuses on static images rather than a series of videos because the physiological characteristic of the lip (lip prints) is considered rather than the behavioural characteristic. In particular the major contribution of the study is employing an object detection technique to efficiently detect the lip area and introduce a deep learning approach in the realm lip-based recognition since there are very few works in this research area.

The remainder of the paper is organized as follows. Section 2 conducts a detailed literature review. Section 3 represents the proposed architecture along with the main contributions of this research. Section 4 outlines the experimental setup of the model along with its different pipelines. Section 5 discusses the results which were obtained, followed by Section 6 which concludes the paper.



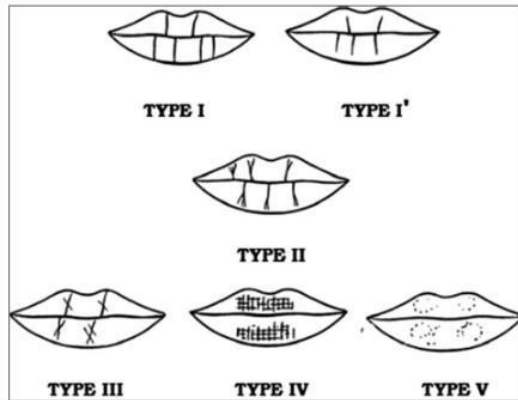


Figure 1: Suzuki and Tsuchihashi classification of lip [5]

## 2. Literature Study

### 2.1. Background

In the field of physical security and information security, access control regulates what the user can view in an environment [8]. To provide access control, the user needs to be authenticated before gaining access to the environment. There are three approaches to authentication [9]: something the user knows (such as passwords), something the user possesses (such as smart cards) and something the user is (such as fingerprint). These authenticators are used in many access control systems along with their environment. According to Rouhani et al. [10] there are two types of environments for access control systems: physical and logical (virtual). Physical environments limit access to an area such as campuses or buildings. A virtual environment limits access to a virtual entry point such as system files and data

Lease [11] states that when assessing the authentication methods against their disadvantages, one method stands out- biometrics. The three different biometric attributes are phenotypic (fingerprint, iris), behavioural (keystroke, voice, signature) and genetic (eye colour, DNA) [9]. A generic biometric system consists of the following components: sensor, pre-processing, feature extraction, template generator, matcher, and application device [9]. One disadvantage of these systems is that it is not viable if the target area being used to identify the individual is covered. For example, if a given person wears glasses or has a beard, the face recognition process can be unsettled when other face images of the same person are taken without glasses or facial hair. Therefore, an effective way to solve this problem can be by considering a smaller part of the face thereby eliminating the concerns with facial analysis. Another example worth mentioning is that some individuals may choose not to place their finger on a fingerprint scanner for fear of contracting a disease such as the novel Coronavirus. Therefore, new alternatives are needed to establish the identify of individuals on occasions where the above techniques are unavailable.

## 2.2. Similar Work

Sharma et al [1] states that work done in lip print technology for recognition is minimal. The most common feature used for authentication is fingerprint due to its accuracy [1]. Fingerprint recognition has a variety of scanners in the market and are easy to install while lip print sensors are not yet developed and research in this area is still in its emerging stage. In recent years there has been an increase in research in this area. Various methods have been used for lip-based identification which are shared by fingerprint recognition such as hough transform, bifurcation analysis, dynamic time warping, similarity coefficients, edge detection and local feature extraction methods.

Smacki et al [12] proposed a method for recognising lip prints using Hough transform based on section comparison. Sections refer to the lines and patterns found on the lips. An algorithm was devised to compare sections. It was found that the sections with length greater than 30 pixels resulted in high error rates.

Wrobel et al [13] proposed using bifurcation analysis for lip recognition. The proposed solution contained pre-processing, feature extraction and identification. The image was pre-processed using linear contrast stretching. For feature extraction, the black pixels were used to find bifurcations. The lip print was then compared to other lip prints in the database using the bifurcation matrices obtained. It was found that extracting bifurcations was a challenging task. The best result that was achieved had an error rate of 23%. The author also states that some pixels imitated bifurcations that did not exist in the lip print. Therefore, it had a negative impact on identification.

Another solution proposed by Wrobel and Froelich [14] uses fuzzy c-means clustering for lip recognition. First, a hough transform is used to extract features from the lips. Thereafter, fuzzy c-means clustering is used to cluster the features. The representatives of clusters are then used to compare the images. The best results were obtained with the number of clusters equal to 80. The overall accuracy achieved was 82,2%.

Smacki et al. [15] proposed a method for lip recognition using the Dynamic Time Warping (DTW) algorithm. The first step in this method was pre-processing the lip print images. The second step was extracting features from the pre-processed images. The method was tested using lip prints of 30 individuals. It was found that the results obtained could possibly be used in forensic labs. The author also stated that modifying the DTW algorithm can result in a higher accuracy.

Porwik et al. [16] proposed a method of comparing and recognising lips using DTW and the Copeland vote counting approach. The lip print was first rendered on a durable surface using a fingerprint powder and then converted into a digital image. It was then normalised, and patterns were extracted from the image. The DTW algorithm was used to calculate the similarity between the lip patterns, and it was paired with the Copeland voting approach to refine the accuracy.

Another solution proposed by Bandyopadhyay et al [17] first used a Gaussian filter to pre-process the acquired image. Sobel edge detection and Canny edge detection was used to detect vertical and horizontal groove patterns in the lips for feature extraction. It was found that using Sobel and Canny edge detection for the extracting lip patterns yielded satisfactory results.

Bakshi et al [18] proposed a solution using local feature extraction methods (SIFT and SURF) on grayscale lip images. Extracting and matching SIFT and SURF features from 23 grayscale images of 10 different subjects worked well with an accuracy of 93.99% and 94.09%, respectively.

In 2014, Travieso et al [38] developed a lip biometric approach based on shape information using a Discrete Hidden Markov Model (DHMMK). The lips are described by shape features (geometrical and sequential) which are then modelled by a DHMMK and classified using an SVM. The experiments were carried out using ten-fold cross validation on 3 datasets: GPDS-ULPGC Face Dataset, PIE Face Dataset and RaFD Face Dataset. The approach achieved an average classification accuracy of 99.8%, 97.13%, and 98.10% respectively.

Wrobel et al [19] proposed a lip recognition system using Probabilistic Neural Network (PNN). Three different databases were used in their experiment; Multi-PIE Face Database, PUT database and a local database consisting of 50 images from 5 subjects. Feature extraction was based on lip contours and facial landmarks is used to achieve classification accuracies of 86.95%, 87.14% and 87.26%, respectively.

Another solution proposed by Wrobel et al [28] is a new approach which uses lip print furrow-based patterns. In this method a lip pattern is created for each person. Several lip patterns are taken from each individual and then appropriately prepared. The prints are then divided into upper and lower lips and the furrows are then made visible. The lip furrows are then parametrized. The lip print pattern comprises of furrows and are stored in the database. This approach did not make use of machine learning techniques and achieved an accuracy of 92.73%.

Das et al [37] proposed an efficient lip biometric framework using SIFT for feature extraction and matching, and a spatial steganographic algorithm to ensure minimum distortion along with hiding the identity of the lip images. The proposed framework was validated on NITRLipV1 and NITRLipV2 which achieved an overall accuracy of 92.7%.

Dela Cruz et al [36] proposed a lip authentication system using Viola-Jones and Active Appearance Based Model (AAM). Here the Viola-Jones algorithm was used to detect the face because of its speed and accuracy. The AAM was chosen for its dimensionality reduction by extracting the relevant feature. It was able to extract the location of points on the lips. The study achieved an accuracy of 87.5% in verifying the identity of a person.

More recently, Sandhya et al [29] compared machine learning algorithms for lip-based identification. Local binary patterns are used to extract features from the segmented upper and lower lip. Shape related features are also extracted. Thereafter, various classifiers such as SVM, K-NN, Ensemble classifiers and ANN are used for classification. This approach achieved accuracies of 81.84%, 80%, 97% and 85.81% respectively.

Although lip print recognition methods might not give results comparable to other methods such as fingerprint and facial recognition, there is value in researching different modalities that can be better suited to situations where a traditional modality is not available. From the published literature discussed above, it can be seen that lip-based recognition is not a well-researched area, which is surprising given the interest in its acoustic equivalent-speaker verification. More importantly, it can be seen that most datasets were created using a contact-based approach and the lip images were then digitized. Therefore, in these cases lip detection techniques were not used (ASI manual intervention for the segmentation of lips is still required). Lip detection is an ongoing and important area of research because it is a basic step in many applications such as automatic lip-reading applications, face recognition systems,

facial expression recognition [39] and is crucial in a standalone lip-based biometric system [39]. Furthermore, it is evident that deep learning architectures have not been introduced in the realm of lip-based identification. Since biometric recognition has shifted from classical models to deep learning models and achieved state-of-the-art results, it is worthwhile to explore how lip-based identification will perform using different deep learning architectures. Taking the aforementioned aspects into account, the main motive of the work is to consolidate the research on lip biometrics by automating the process of object detection and identification using a facial landmark detector to localize and detect the lips and applying classical and deep learning methods on a publicly available dataset.

### 3. Proposed Experiment

The basic outline of the experiments performed is represented in Figure 2. As a pattern recognition problem, the proposed system consists of different steps such as capturing a dataset, object detection, pre-processing the ROI, feature extraction and classification based on the extracted features. The figure depicts that currently two pipelines namely, a traditional machine learning pipeline and a deep learning pipeline.

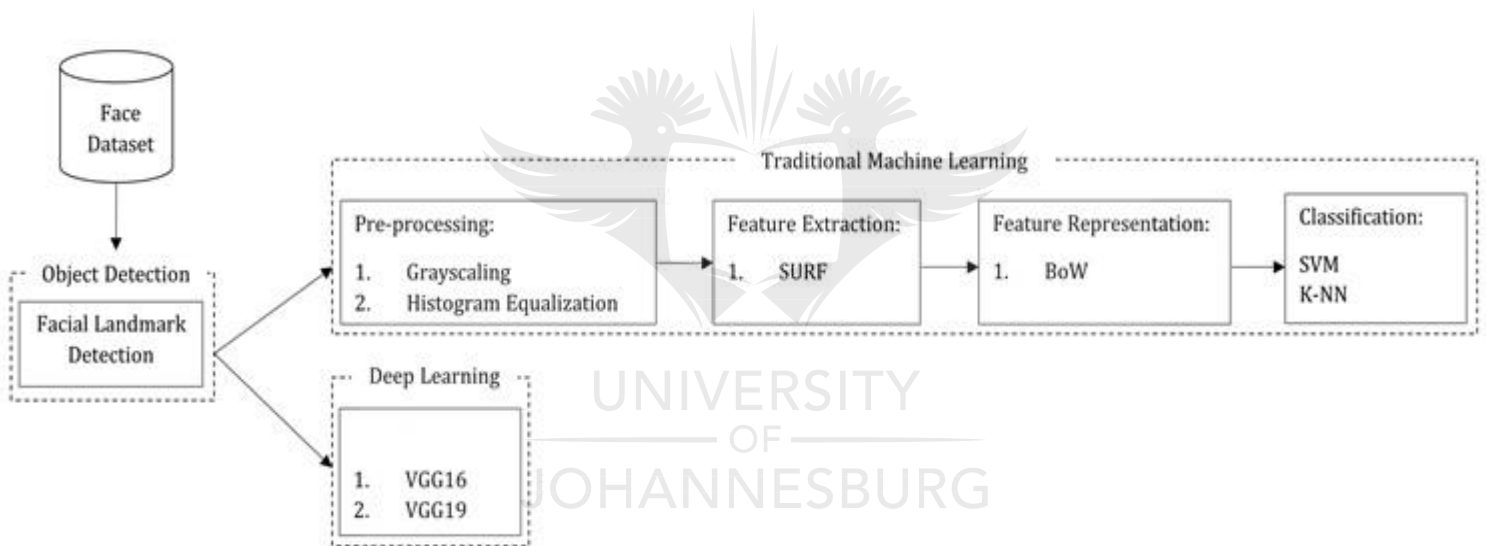


Figure 2: Architecture of Proposed System

The main steps and contributions of this research are described as follows:

1. The lips are automatically detected within the static image frame using a facial landmark detector.
2. The acquired ROIs are pre-processed to emphasize and consider every pattern on the lips.
3. Features are extracted to account for the uniqueness in the shape and patterns found on the lips.
4. The feature vectors are quantized into visual words using a Bag-of-Words (BoW) approach. It should be noted that this approach is used only for the traditional machine learning pipeline.

5. The extracted features are trained using various classification models. Consequently, these models are analysed to determine their performance in lip-based identification.
6. Manual work is eliminated by completely automating the process of objection detection, pre-processing, feature extraction and classification.
7. A deep-learning approach for lip-print identification is introduced.

## 4. Experiment Setup

### 4.1. Dataset

The selection of an appropriate dataset is crucial to the development of any research study. The selected dataset must contain features relevant to the problem domain with sufficient predictive power that enables the training process to learn from it [25]. Large public databases are available for many facial analysis problems. These datasets are divided into two categories: datasets produced within a controlled environment and datasets produced in an uncontrolled environment [25]. Images from social networks such as Facebook and image search engines such as Google Images have contributed to the production of large datasets, comprising of facial images in different uncontrolled situations [25]. However, these “in-the-wild” datasets are not applicable to the problem domain of the current study and datasets produced in a controlled environment will be used. However, there is little consistency amongst the different face sets. Some sets have a very low resolution with minimal subjects while some have a moderate resolution with many subjects but the grooves on the lips are not visible.

After reviewing several face datasets such as NITRLipV1 [41], the Chicago Face Dataset [20] (CFD) has been selected for the proposed study. The NITRLipV1 database was not considered for this study because the images are not of a good enough quality. More importantly, the NITRLipV1 database lacks ethnic diversity, where the subjects are of South Asian descent. Therefore, the results obtained from these would be biased. It is important to consider different ethnicities to achieve reliable results. The goal of the CFD was to determine whether the participants used were appropriate for research purposes such as facial recognition.

The CFD consists of high-resolution images (2444 pixels × 1718 pixels) of 597 male and female targets of varying ethnicities. Each target is represented with a neutral expression image. For a subset of 158 targets, the dataset also includes images with happy (mouth open and closed), fearful and angry facial expressions. The CFD is particularly suited since it includes self-identified Asian, Black, Latino and White ethnicities of different ages. Therefore, there is very bias towards gender, age, or race.

Currently, the CFD has limited data in terms of the number of samples per individual. Therefore, to overcome this drawback data augmentation is applied to artificially increase the data in the dataset. The methods used for data augmentation include the following modifications to the original image: horizontal flip, a rotation and an increase and decrease in brightness. Geometric transformations are helpful for positional biases present in the dataset [35]. Furthermore, altering the brightness of the original image is important because in real-life scenarios dramatic image variations arise from changes in illumination. Therefore, it is important to take these factors into consideration because it will determine how well the proposed experiments can recognise the lips under different conditions.

## 4.2. Traditional Machine Learning Pipeline

The traditional machine learning pipeline for lip-based identification involves detecting the region of interest, pre-processing the ROI, extracting features, and classifying the lip prints based on the extracted features. This section will give an outline of the algorithms used to achieve lip-based identification.

### 4.2.1. Object Detection

The Face Landmark Detection algorithm offered by the Dlib library is used for object detection. The algorithm presented by Kazemi et al [21] is an implementation of the One Millisecond Face Alignment with an Ensemble of Regression Trees (ERT). This technique uses an ensemble of regression trees to estimate landmark positions from pixel intensities. A landmark is a keypoint for a specific facial structure such as the eyes, nose, mouth, and chin. For this study, the region of interest is the mouth area. There are 68 (x, y)-coordinates that map to facial components on the face. The coordinates specified for the mouth is from 49-68. However, before the lips can be isolated the face should be detected first. Therefore, the pre-trained (hog-based) face detector is used to detect the face in the image. The dlib face detector uses HOG features with a linear classifier to detect faces. After the face has been detected, the mouth region is detected using the coordinates specified above and a bounding box is created around the detected region.

### 4.2.2. Pre-processing

The detected lips are pre-processed to enhance the ROI. Pre-processing is an important phase because it suppresses noise or enhances important features [22]. The first method used for pre-processing is grayscaling. This method is important because a grayscale image helps identify important edges or other prominent features. The second method used is histogram equalization. Histogram equalization is important because it adjusts intensities to enhance contrast within the image, thus illuminating the pattern of grooves on the lips.

### 4.2.3. Feature Extraction

The next phase, feature extraction, is a core component of the traditional machine learning pipeline. The main goal is to obtain the most relevant information from the preprocessed data. For the current study, local features will be used for feature extraction because global features have certain limitations such as sensitivity to noise, illumination variation and failure to detect important features in the image [30]. These disadvantages can be resolved by using local features which encapsulate local information to obtain better details of the image [30]. The relevant feature extraction method for this pipeline is Speeded Up Robust Features (SURF). SURF is the most popular local feature extraction method which has proved to be most promising due to its high performance [30]. The hessian threshold is set to 500 with a descriptor size of 128-dimensions. 500 is an optimal hessian value because a consistent number of features are extracted from the lips by the algorithm for all the images. Lastly, SURF-128 is used because it is more distinctive and much faster to compute. An example of SURF features over the original image is illustrated in figure 3.

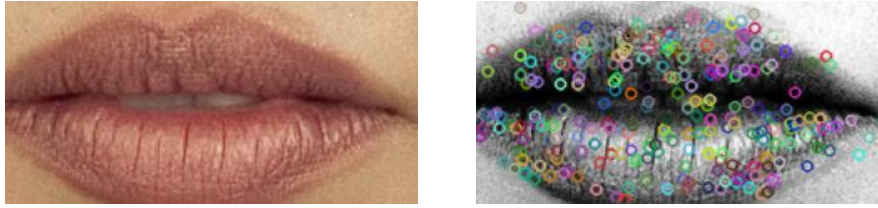


Figure 3: A Representation of SURF Features over the Original Image

#### 4.2.4. Feature Representation

Before, the respective classifiers are trained, it is important to quantize the feature vectors into visual words due to the large number of local features obtained for each image. Therefore, to overcome this drawback, the model, Bag-of-Words (BoW) is used. This model has become popular in recent years due to its effectiveness and performance [31]. By using the BoW approach, local descriptors are encoded into a histogram representation using the k-means algorithm to cluster the feature descriptors [31]. Thereafter, each image is represented by a k-bins histogram. The number of clusters that best suits the pipeline is 80. These features are then used to train the classifiers.

#### 4.2.5. Classification

An SVM was chosen as one of the classifiers for this pipeline because it has proven to be successful when used for pattern classification problems and its ability to achieve accurate results [32]. In this work, SVM is used to classify the unknown lip-prints and identify to which person it belongs to. A linear kernel with a C value of 100 and a gamma value of 0.001 is used along with a one-vs-all classifier. 80% of the data is used for training the classifier and 20% is used for testing.

The next classifier, K-NN, is chosen because it is generally easy to implement, the training is very fast, and it is robust to noisy training data. K-NNs have been used for pattern recognition since the 1970's [33]. The "k" parameter or the number of nearest neighbours, which determines how many neighbours should be checked when data is being classified, is the most fundamental parameter. This value is set to  $k = 1$  along with a euclidean distance and a one-vs-all classifier. Here 80% of the data is used for training the classifier and 20% is used for testing.

### 4.3. Deep Learning Pipeline

The deep learning pipeline for lip identification involves detecting the relevant region of interest and using a Convolutional Neural Network (CNN) classifier. There are many CNN architectures that have been developed such as AlexNet, LeNet, VGG16, VGG19 and ResNet. For this study, the VGG16 and VGG19 architectures are used because it is a good architecture for benchmarking on a specific task. The object detection phase follows the same step as the previous pipeline by using the Face Landmark Detection algorithm offered by the Dlib library to detect the lip region.

A deep learning approach can play a great role to develop a biometric system for lip print identification. Traditional machine learning algorithms can lead to a lack of extracted features [35]. To overcome these challenges, deep neural networks have attracted the attention of researchers for their capability of complex pattern recognition. Researchers have proposed

different CNN models trained on natural images for object recognition [40]. A transfer learning or fine tuning of these models has shown a remarkable performance for similar tasks [40]. According to [40], for scarce datasets, such as ear biometrics, the transfer learning or fine-tuning of models has achieved a better performance than when training a model from scratch. Therefore, it is worthwhile to explore how a transfer learning model will perform for lip-based identification. Inspired by [40], this work employs VGG network architectures for robust feature extraction and classification.

The VGG network architecture was originally proposed by Simonyan and Zisserman [23] which secured first and second places in the ImageNet Challenge. For this pipeline, a transfer learning approach using ImageNet weights is applied using VGG16 and VGG19 architectures as the base model. The weights of the pre-trained models are used for training.

Before training the respective models, the images were first resized. The width and height of the images were computed, and an average dimension was determined. The images are resized to a width of 260 and a height of 224. To ensure efficient training of the network, the chosen parameters for the actual compilation of the models include Adam as the learning optimizer function with a learning rate of 0.001 and a batch size of 32. The loss is set to categorical cross-entropy and the metrics used are accuracy and loss. The number of epochs chosen to train the VGG16 and VGG19 model is set to 100. The training and test set used for this pipeline was 80% for training and 20% for testing. Lastly, to ensure a high accuracy in classification, data augmentation is used which assists in preventing overfitting [34]. To prevent overfitting the parameter used for data augmentation is the horizontal flip.

#### 4.4. Performance Evaluation Parameters

A lip-based identification system is judged by the quality of lip detection and identification modules. The benchmark parameters used to assess the performance of the proposed model are discussed below:

- a. **Intersection over Union (IoU):** According to [26] IoU is the most popular evaluation metrics for measuring the overlap between two bounding boxes. Given an image, the IoU measures the similarity between the predicted region and the ground truth region using Eq. (1). It is defined as the size of the intersection divided by the union of the two regions [26]. The lower the IoU, the worse the prediction result is. The highest prediction result that can be achieved is 1.

$$IoU = \frac{\text{Area of Overlap } (X \cap Y)}{\text{Area of Union } (X \cup Y)} \quad (1)$$

- b. **Precision:** It measures the number of correct instances out of all the total number of instances (true positives and false positives). The formula is as follows:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- c. **Recall:** It determines what portion of actual positives are identified correctly divided by all the correct instances which can be seen in Eq. (3):



$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- d. **F1 Score:** It is an estimate of the accuracy of the model. It is the weighted average of both precision and recall. The equation is as follows:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

- e. **Accuracy:** Accuracy compares how close a measured value is to its true value. Therefore, it is the ratio of the number of correct instances to all the number of instances which can be seen in Eq. (5):

$$Accuracy = \frac{TP+TN}{N+P} \quad (5)$$

- f. **Equal Error Rate (EER):** EER is the most important indicator to evaluate the performance of a recognition system because it describes the overall accuracy of the biometric system. It describes the point where the false acceptance rate (FAR) and the false rejection rate (FRR) are equal. A low EER value indicates a better performance.
- g. **Receiver Operating Characteristic (ROC) curve:** A ROC curve is a trade-off between FAR and FRR. It is a probability curve that plots the true positive rate (TPR) against the false positive rate (FPR) at various thresholds. A ROC curve aims to maximise the area under the curve (AUC). The higher the AUC, the better the model is at distinguishing between positive and negative classes.

#### 4. Results

A set of results are obtained from the implemented pipelines discussed and outlined in the previous section. For each class a recall and precision score were calculated which were then averaged to produce an overall macro precision and recall average for the classifiers. A macro average was used since all the classes need to be treated equally to evaluate the overall performance of the classifiers. These results give a valuable insight into the performance of the pipelines and will be discussed in this section.

##### 5.1. Intersection over Union Performance Measure for Object Detection

The One Millisecond Face Alignment algorithm successfully localized the lip region of each target in the dataset with no false positive detections. However, that is not enough to assess the performance of the algorithm. Therefore, the IoU of each detected lip within the image frame was determined, followed by the average precision (AP) for each class. It is apparent from figure 4 that most classifications received an accuracy of 0.90 or higher. Therefore, all the classifications were correct. The ground truth labels that contain the bounding box

coordinates for this study were created by the authors for the region of interest in the images. These coordinates were then compared with the predicted coordinates to compute the segmentation metrics. Consequently, an average was computed across all the samples and a performance result of 0.93 was achieved. Hence, the proposed approach for object detection (One Millisecond Face Alignment algorithm) has worked well for the current study.



Figure 4: Samples of IoU score, ground truth region and predicted region of different targets

### 5.2. Traditional Pipeline with SVM

The first variation of the traditional machine learning pipeline made use of the One Millisecond Face Alignment algorithm for object detection, grayscaling and histogram equalization for pre-processing, SURF for feature extraction with a bag-of-words approach and an SVM classifier.

The metrics presented for this pipeline yielded promising results with an accuracy of 95.45%. A precision score of 94.87% and recall score of 93.59% was obtained which demonstrates that the model achieved a high percentage of relevant results and its ability to classify unknown lip prints. The EER was calculated to be 2.27%. The lower the EER, the higher the accuracy of the model. Furthermore, an f1 score of 94.22% was achieved as shown in Table 1.

Figure 5 shows the confusion matrix of the traditional machine learning pipeline with SVM. The diagonal shows the number of correct classifications whereas all the other entries not in the diagonal are the misclassifications. Notably, there were 2 instances where the model incorrectly predicted the individual. The 2 individuals that were misclassified were both females. These lips were misclassified with the same gender. A ROC curve visualizes the performance of the classification problem. The ROC curve in Figure 6 leans closely towards the upper left corner of the diagram, which is a desirable result.

Table 1. Traditional Pipeline with SVM results

| Accuracy | Precision | Recall | F1 Score | Misclassifications |
|----------|-----------|--------|----------|--------------------|
| 95.45%   | 94.87%    | 93.59% | 94.22%   | 2                  |

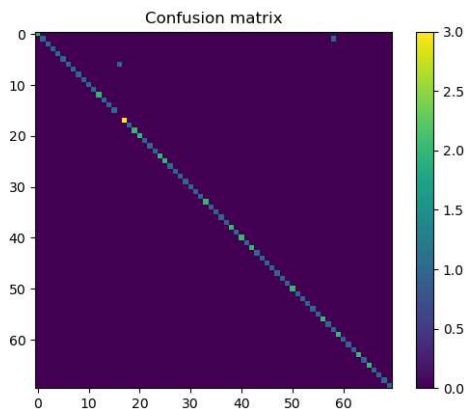


Figure 5: SVM Confusion Matrix

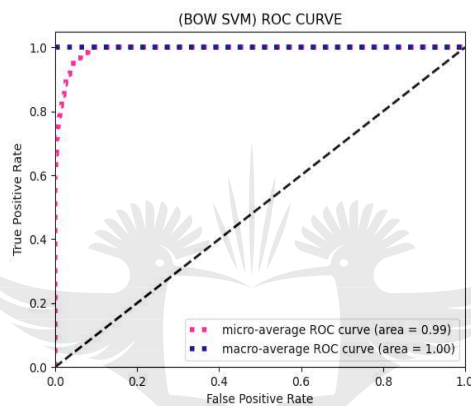


Figure 6: SVM ROC Curve

### 5.3. Traditional Pipeline with K-NN

The second variation of the traditional machine learning pipeline made use of the One Millisecond Face Alignment algorithm for object detection, grayscaling and histogram equalization for pre-processing, SURF for feature extraction with a bag-of-words approach and a K-NN classifier.

The metrics presented for this pipeline also yielded promising results with an accuracy of 94.31%. This shows that there were a high number of correct unknown lip print classifications. The classifier achieved a precision score of 91.05% and a recall score of 92.68% as shown in Table 2. The EER achieved for this variant was 2.84%. Finally, the f1 score was calculated to be 91.85%.

The confusion of this variant is shown in figure 7. The diagonal shows the number of correct classifications whereas all the other entries not in the diagonal are the misclassifications. The confusion matrix depicts that there were 4 misclassifications for this pipeline. The individual that was misclassified in this pipeline was also misclassified in the previous pipeline. The x-axis on the confusion matrix lists the true labels while the y-axis lists the predicted labels. The ROC curve which is represented in figure 8 maximizes the area under the curve which is a desirable result.

Table 2. Traditional Pipeline with K-NN results

| Accuracy | Precision | Recall | F1 Score | Misclassifications |
|----------|-----------|--------|----------|--------------------|
| 94.31%   | 91.05%    | 92.68% | 91.85%   | 4                  |

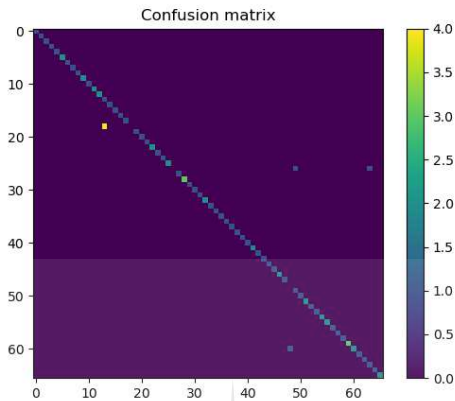


Figure 7: K-NN Confusion Matrix

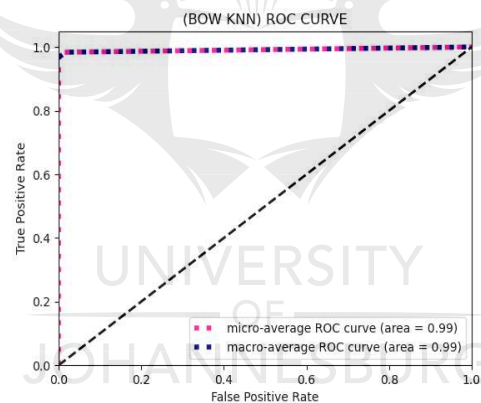


Figure 8: K-NN ROC Curve

#### 5.4. Deep Learning Pipeline with VGG16

The first variation of the deep learning pipeline made use of the One Millisecond Face Alignment algorithm and the VGG16 model. This pipeline achieved promising results which suggest that is worthwhile for lip print identification. The VGG16 model obtained a validation accuracy of 91.53% which indicates that it can make reasonably accurate classifications. The recall score was 91% and the precision score was 95% as shown in Table 3. The EER achieved for this pipeline was 4.23%. The f1 score was calculated to be 92.95%. The number of epochs the dataset was trained for was 100. One issue involved with this variation is that it received the highest number of misclassifications which can be seen in the confusion matrix in Figure 9. The entries not in the diagonal depict the number of misclassifications. The ROC curve of the VGG16 classifier is shown in Figure 10 and the accuracy and loss curves are shown in Figures 11 and 12.

The gap between the training and validation accuracy curve in Figure 10 is a clear indication that it is minimal. Therefore, this is a desirable result since a higher gap indicates higher overfitting. The loss curve in Figure 11 indicates a good fit since training and validation loss decrease with a minimal gap between the two final loss values.

Table 3. Deep Learning Pipeline with VGG16 results

| Accuracy | Precision | Recall | F1 Score | Misclassifications |
|----------|-----------|--------|----------|--------------------|
| 91.53%   | 95%       | 91%    | 92.95%   | 10                 |

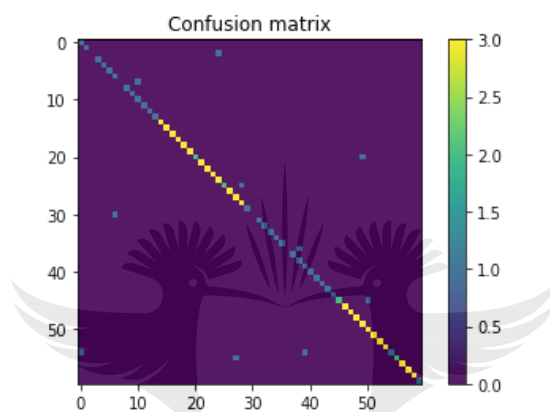


Figure 9: VGG16 Confusion Matrix

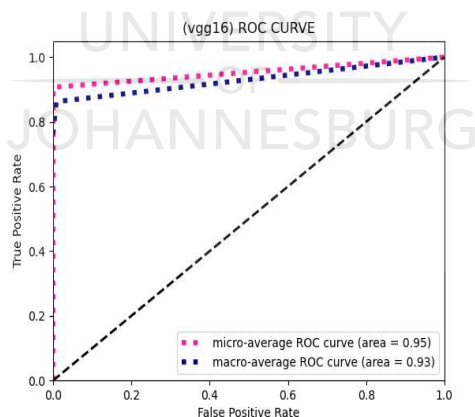


Figure 10: VGG16 ROC Curve

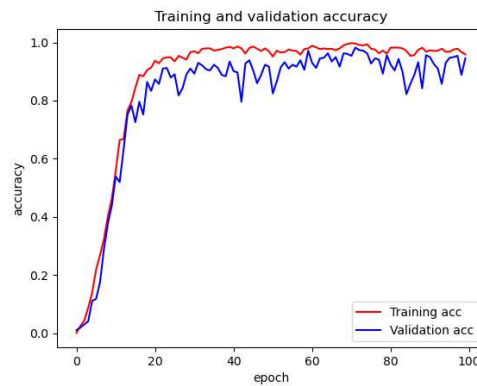


Figure 11: VGG16 Accuracy Curve

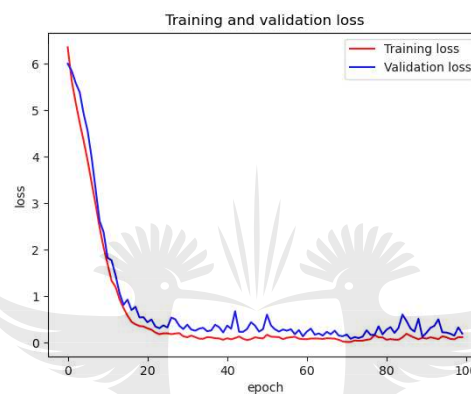


Figure 12: VGG16 Loss Curve

### 5.5. Deep Learning Pipeline with VGG19

The second variation of the deep learning pipeline made use of the One Millisecond Face Alignment algorithm and the VGG19 model. This pipeline achieved better results than the previous variation in terms of its accuracy. Despite some misclassifications, the validation accuracy achieved by the classifier was 93.22%. The recall score and precision score were 95% and 90% respectively as shown in Table 4. The EER was calculated to be 3.39%. Furthermore, an f1 score of 92.43% was calculated. The confusion matrix and roc curve are illustrated in Figures 13 and 14, respectively. The confusion matrix depicts the number of misclassifications as well as the lip prints which were classified correctly. The training and accuracy curves shown in Figure 15 move closer towards a higher number as the number of epochs increase which is a favourable result. The training and validation loss curves shown in Figure 16 move towards a smaller value as the number of epochs increase. It decreases to a point of stability with no gap between the two final values.

Table 4. Deep Learning Pipeline with VGG19 results

| Accuracy | Precision | Recall | F1 Score | Misclassifications |
|----------|-----------|--------|----------|--------------------|
| 93.22%   | 95%       | 90%    | 92.43%   | 6                  |

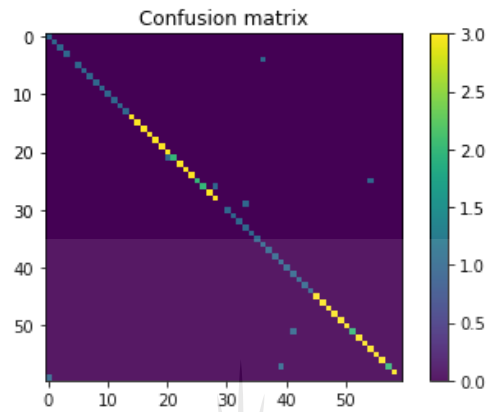


Figure 13: VGG19 Confusion Matrix

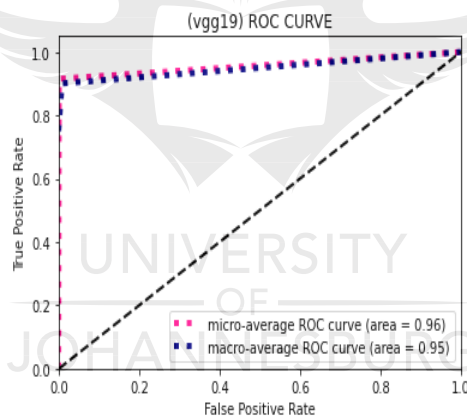


Figure 14: VGG19 ROC Curve

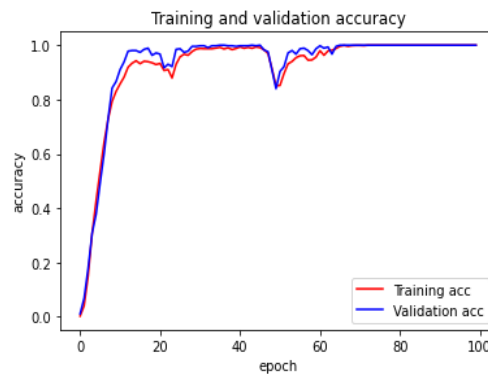


Figure 15: VGG19 Accuracy Curve

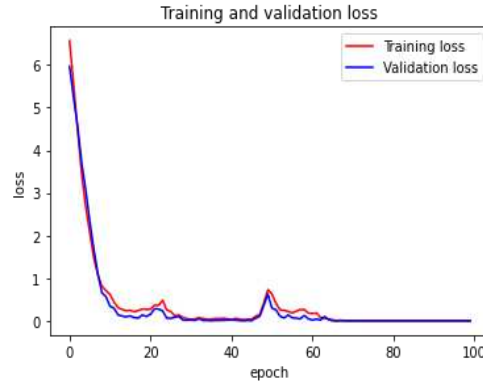


Figure 16: VGG19 Loss Curve

## 5. Discussion

The metrics for the traditional machine learning pipeline generally yielded promising results. The SVM variation achieved the highest accuracy of 95.45%. The lowest accuracy received for this pipeline was 94.31%. Therefore, it can be said that the traditional machine learning pipeline produced a high-performance metric for lip-based identification. A more current approach that made use of the VGG16 and VGG19 models also yielded promising results. The highest accuracy achieved in this pipeline was 93.22%. A summary of the results can be seen in Table 5.

Table 5. Summary of the results obtained

| Pipeline                 | Accuracy | Precision | Recall | F1 Score |
|--------------------------|----------|-----------|--------|----------|
| Traditional with SVM     | 95.45%   | 94.87%    | 93.59% | 94.22%   |
| Traditional with K-NN    | 94.31%   | 91.05%    | 92.68% | 91.85%   |
| Deep Learning with VGG16 | 91.53%   | 95%       | 91%    | 92.95%   |
| Deep Learning with VGG19 | 93.22%   | 90%       | 95%    | 92.43%   |

The lip-based identification methods proposed by this study indicate that there is no bias towards different skin colours and the proposed methods can adequately detect and perform lip print identification due to the diversity of the dataset. More importantly this study made use of an object detection algorithm to detect the lips as compared to previous systems which used traditional methods to acquire lip prints as described in Section 2.2. This study also introduced deep-learning architectures which is an apparent gap in the field of lip-based identification. Based on the results discussed above, it can be concluded that the SVM pipeline performed the best in terms of accuracy. However, as the results indicate, by using a deep learning approach it can be concluded that it is possible to achieve lip-based identification with a good performance. The results demonstrate that further improvements can be made by building new models for lip-based identification to achieve a better performance and



commercial deployment.

Currently, this study employed a high-resolution face dataset to achieve lip-based identification. This work can be further extended to incorporate low quality and blurred images to enhance the reliability of the designed model. On the technical side of this study, improvements can be made in terms of object detection approaches used. Sandhya et al [29] hinted at the fact that a standard and uniform methodology is required to acquire lip prints. This study has proved that using the One Millisecond Face Alignment algorithm can be used to adequately detect the lip within the image. However, more sophisticated methods worth exploring are deep learning methods such as YOLO and R-CNNs. Hence, future research should be targeted towards addressing these limitations.

## **6. Conclusion**

This study compared traditional machine learning and deep learning methods for lip-based identification and exhibited promising results. The first pipeline which is a traditional machine learning pipeline with an SVM and K-NN classifier achieved an accuracy of 95.45% and 94.31%. The second pipeline which is a deep learning pipeline with VGG16 and VGG19 architectures achieved an accuracy of 91.53% and 93.22%. An overview of the study indicates that it has potential for lip-based identification in applications such as handheld devices. One of the main challenges encountered was the availability of datasets. Although there are many face datasets, a dataset with a high-resolution of lips along with its pattern of grooves is limited.

The current state-of-the-art works in the field of lip print identification have been presented in the Similar Work section 2.2. Although, similar techniques were used such as preprocessing, feature extraction and classification, this study exhibited promising results compared to previous work. This study also utilized object detection algorithms to obtain the region of interest and made use of deep learning methods as well. The results indicate that the traditional machine learning pipeline achieved the best accuracy, however, it also indicates that deep learning architectures have potential in the realm of lip-based identification.

Biometric identification has shifted from traditional methods to deep learning methods. Deep learning methods have been successful in achieving state-of-art results in biometric identification for fingerprint, face, iris, ear, palmprint and gait since the paradigm shift in 2012 [24]. Therefore, although, lip print identification is still in its early stage, deep learning methods for lip print identification can produce promising results as exhibited by this study. This work can further be extended by incorporating low quality images to enhance the reliability of the system and exploring more sophisticated deep learning methods to achieve state-of-the-art results. In future studies, we aim to expand this work by employing state-of-art deep learning approaches in the field of lip-based identification.

## References

- [1] P. Sharma, S. Deo, S. Venkateshan and A. Vaish, "Lip Print Recognition for Security Systems: An Up-Coming Biometric Solution," *Intelligent Interactive Multimedia Systems and Services*, vol. 11, pp. 347-359, 2011.
- [2] S. Boonkrong, *Authentication and Access Control: Practical Cryptography Methods and Tools*, Apress, 2021, pp. 45-70.
- [3] A. Kaushal and M. Pal, "Cheiloscopy: A Vital Tool in Forensic Investigation for Personal Identification in Living and Dead Individuals," *International Journal of Forensic Odontology*, vol. 5, no. 2, pp. 71-74, 2020.
- [4] S. A. Ahmed, H. E. Salem, and M. M. Fawzy, "Forensic dissection of lip print as an investigative tool in a mixed Egyptian population," *Alexandria Journal of Medicine*, vol. 54, no. 3, pp. 235-239, 2018.
- [5] Y. Tsuchihashi, "Studies on personal identification by means of lip prints," *Forensic Science*, vol. 3, pp. 233-248, 1974.
- [6] N. Ishaq, E. Ullah, I. Jawaad, A. Ikram and A. Rasheed, "Cheiloscopy: A Tool for Sex Determination," *The Professional Medical Journal*, vol. 21, no. 5, pp. 883-887, 2014.
- [7] S.L. Wang and A. Wee-ChungLiew, "Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power," *Pattern Recognition*, vol. 45, pp. 3328-3335, 2012.
- [8] R. Anderson, "Access Control," in *Security Engineering: A Guide to Building Dependable Distributed Systems*, John Wiley & Sons, 2010, pp. 51-71.
- [9] H. Vallabh, *Authentication using Finger-Vein Recognition*, 2012.
- [10] S. Rouhani, V. Pourheidari and R. Deters, "Physical Access Control Management System Based on Permissioned Blockchain," *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 1078-1083, 2018.
- [11] D. Lease, *Factors influencing the adoption of biometric security technologies by decision making information technology and security managers*, 2005.
- [12] Ę. Smacki, P. Porwik, K. Tomaszycycki and S. Kwarciańska, "The lip print recognition using Hough transform," *Journal of Medical Informatics and Technologies*, 2010.
- [13] K. Wróbel, R. Doroz and M. Palys, "Lip Print Recognition Method Using Bifurcations Analysis," *Lecture Notes in Computer Science*, vol. 2, pp. 72-81, 2015.
- [14] K. Wrobel and W. Froelich, "Recognition of Lip Prints Using Fuzzy C-Means Clustering," *Journal of Medical Informatics & Technologies*, vol. 24, pp. 67-74, 2015.
- [15] L. Smacki, K. Wrobel and P. Porwik, "Lip print recognition based on DTW algorithm," *2011 Third World Congress on Nature and Biologically Inspired Computing*, pp. 594-599, 2011.
- [16] P. Porwik and T. Orczyk, "DTW and Voting-Based Lip Print Recognition System," *11th IFIP TC 8 international conference on Computer Information Systems and Industrial Management*, pp. 191-202, 2012.
- [17] S. K. Bandyopadhyay, S. A. and S. B. , "Feature Extraction of Human Lip Prints," *Journal of Current Computer Science and Technology*, vol. 2, no. 1, pp. 1-8, 2012.
- [18] S. Bakshi, R. Raman and P. Sa, "Lip pattern recognition based on local feature

extraction," *2011 Annual IEEE India Conference (INDICON)*, India, 2011.

- [19] K. Wrobel, R. Doroz, P. Porwik, J. Naruniec and M. Kowalski, "Using a Probabilistic Neural Network for lip-based biometric verification," *Engineering Applications of Artificial Intelligence*, vol. 64, no. C, p. 112-127, 2017.
- [20] D. Ma and B. Wittenbrink, "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data," *Behavior Research Methods*, vol. 47, pp. 1122-1135, 2015.
- [21] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874, 2014.
- [22] M. Sonka, V. Hlavac and R. Boyle, "Image Pre-Processing," in *Image Processing, Analysis and Machine Vision*, 1993, pp. 56-111.
- [23] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2014.
- [24] B. Kieffer, M. Babaie, S. Kalra and H. Tizhoosh, "Convolutional Neural Networks for Histopathology Image Classification: Training vs. Using Pre-Trained Networks," *2017 Seventh International Conference on Image Processing Theory, Tools and Applications*, pp. 1-6, 2017.
- [25] B. Johnston and P. d. Chazal, "A review of image-based automatic facial landmark identification techniques," *EURASIP Journal on Image and Video Processing*, no. 86 (2018), 2018.
- [26] H. Rezatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*, 2019.
- [27] J. O. Kim, W. Lee, J. Hwang, C. H. Chung and K. S. Baik, "Lip print recognition for security systems by multi-resolution architecture," *Future Generation Computer Systems*, vol. 20, no. 2, pp. 295-301, 2004.
- [28] K. Wrobel, R. Doroz, P. Porwik and M. Bernas, "Personal identification utilizing lip print furrow based patterns. A new approach," *Pattern Recognition*, vol. 81, pp. 585-600, 2018.
- [29] S. Sandhya, R. Fernandes, S. Sapna and A. Rodrigues, "Segmentation of Lip Print Images Using Clustering and Thresholding Techniques," in *Advances in Artificial Intelligence and Data Engineering*, Singapore, Springer, 2020, pp. 1023-1034.
- [30] L. Kabbai, M. Abdellaoui and A. Douik, "Image classification by combining local and global features," *The Visual Computer*, vol. 35, p. 679-693, 2019.
- [31] Y. Zhang, R. Jin and Z. H. Zhou, "Understanding bag-of-words model: A statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1, pp. 43-52, 2010.
- [32] L. Auria and R. Moro, "Support Vector Machines (SVM) as a technique for solvency analysis," *DIW Discussion Papers*, vol. 811, 2008.
- [33] C. M. Ma, W. S. Yang and B. W. Cheng, "How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset," *Journal of Applied Sciences*, vol. 14, pp. 171-176, 2014.
- [34] M. Burugupalli, *Image Classification Using Transfer Learning and Convolutional Neural Networks*, 2020.

- [35] Nur-A-Alam, M. Ahsan, M.A. Based, J. Haider and M. Kowalski, "An intelligent system for automatic fingerprint identification using feature fusion by Gabor filter and deep learning," *Computers & Electrical Engineering*, vol. 95, 2021.
- [36] J. C. Dela Cruz, R. G. Garcia, S. M. M. Go, J. L. Regala and L. J. T. Yano, "Lip Biometric Authentication Using Viola-Jones and Appearance Based Model (AAM) System," *2020 35th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2020, pp. 372-377.
- [37] Das, S., Muhammad, K., Bakshi, S., Mukherjee, I., K Sa, P., Sangaiah, A. and Bruno, A., 2019. Lip biometric template security framework using spatial steganography. *Pattern Recognition Letters*, 126, pp.102-110.
- [38] Travieso, C., Zhang, J., Miller, P. and Alonso, J., 2014. Using a Discrete Hidden Markov Model Kernel for lip-based biometric identification. *Image and Vision Computing*, 32(12), pp.1080-1089.
- [39] A. Hassanat, M. Alkasassbeh, M. Al-awadi and E. Alhasanat, "Colour-based lips segmentation method using artificial neural networks", *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015.
- [40] Kamboj, A., Rani, R. and Nigam, A., 2021. A comprehensive survey and deep learning-based approach for human recognition using ear biometric. *The Visual Computer*.
- [41] S. Bakshi, R. Raman and P. Sa, "NITRLipV1: a constrained lip database captures in visible spectrum", *ACM SIGBioinformatics Record*, vol. 6, no. 1, pp. 1-1, 2016.

# Appendix B

A manuscript will be submitted to International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI). This article investigates the effectiveness of employing end-to-end object detection deep learning architectures to detect the lips and process lip prints for biometric identification purposes.



# A Deep Learning Model for Lip-based Identification using YOLOR

**Abstract.** In recent years, the exploitation of lip prints for biometric identification has gained much attention from the research community, with most of the efforts devoted to establishing that the physiological characteristic of the lip is discriminative and unique to each individual. Up until now research in this area has employed more traditional feature engineering-based approaches and results that have been achieved are still not comparable with those yielded by more commonly used biometric characteristics. Furthermore, the field of lip detection is still an ongoing topic of research due to its many challenges which hinders the success of lip detection techniques. Within this regard, this work will determine the viability of newer methods on the task of lip detection and identification through the application of newer deep learning methods which is an apparent gap in this area. In this study YOLOR is applied on samples of faces from the CFD dataset to effectively achieve lip detection and identification. The results obtained are promising with a mAP of 99.5% and a precision and recall score of 67% and 99%, respectively.

**Keywords:** Deep Learning, YOLOR, Lip Detection, Lip Identification

## 1 Introduction

It is widely acknowledged that traditional methods of authentication such as passwords are not the most effective means of authentication, leading to an increased attention of biometric recognition [1]. Replacing passwords or access cards with a biometric trait has many advantages; it cannot be lost, forgotten, stolen, or disclosed [1]. Currently, biometric recognition systems are applied in many real-life scenarios ranging from border control to unlocking mobile devices. Although various biometric traits have achieved state-of-the-art results, such as fingerprint, face, and iris, research is still carried out in this area to design novel biometric systems, based on other biometric traits, which may possess useful properties not available in mainstream solutions [2]. One such trait that has gained considerable attention recently is the human lip. Its uniqueness has been confirmed by Tsuchihashi and Suzuki [3] due to the wrinkles or grooves present on the surface of the lip. Furthermore, previous research that has been undertaken in this field has proven that lip prints can be used to recognise individuals [4]. Choras states in his lip recognition paper that when using the lip as a biometric trait user interaction is not needed because lip biometrics is regarded as a passive biometric and images of the user may be acquired from a distance without the knowledge of the user [5]. Furthermore, they mention better results can be expected for lip biometrics than behavioural biometrics [5] and lastly, the lips can be implemented in hybrid systems such as face-lips biometric systems. These highlight the potential of using the lip as a biometric. However, research within this area is still in its early stages and has

relied on traditional machine learning methods thus far to achieve recognition. Biometric recognition has now shifted from hand-crafted features to deep learning architectures and have achieved state-of-art results for fingerprint, face, and iris recognition [6] as well as many others. To date lip-based recognition has not gained much deep learning-based research, especially compared to other biometric traits. More importantly, previous research into lip-based recognition is inconsistent with results reported on small or private datasets, and different metrics making comparison difficult [7]. Therefore, novel approaches are required for lip-based recognition, in order to evaluate whether the recognition rates are comparable with those achieved through other well-established biometric traits.

With the rapid development of artificial intelligence (AI), deep learning architectures have become popular for object detection because they can identify and process data efficiently. Therefore, it is worthwhile to explore how these architectures will perform in the realm of lip-based recognition. Furthermore, lip detection remains an ongoing topic of research which originates from numerous applications such as speaker verification or standalone lip-based biometric systems. Automating the process of lip detection and recognition is quite a difficult task in computer vision due to the variation amongst humans and environmental conditions [8]. Taking the above aspects into account, this paper investigates the effectiveness of employing deep learning techniques to detect the lips and process lip prints for biometric identification, in order to measure whether their usage could achieve a high-level recognition performance. The aim of this paper is to advance the field forward by applying a deep learning image segmentation technique to improve the performance of lip-based identification. The major contribution of the current study is the proposal of a modified YOLOR model on a publicly available face dataset to achieve lip-based identification.

The remainder of the paper is organised as follows. Section 2 discusses the literature study with a problem background in section 2.1 and similar work conducted in section 2.2. The architecture adopted and the experiment setup are outlined in section 3. The results of the current study are highlighted in section 4, followed by a discussion of the obtained results in section 5. Conclusions are eventually drawn in section 6.

## **2 Literature Study**

### **2.1 Problem Background**

From published literature [4, 6], it can be seen that lip-based recognition is not a well-researched area. More importantly, lip-based recognition works are mostly based on hand-crafted features. Many of the hand-crafted features are based on edges (SIFT [9], LBP [10], Sobel and Canny [11]), or are derived from the transfer domain such as Gabor wavelets [12]. Once the feature or feature representation is extracted, it is fed into the classifier to perform recognition. According to Minaee et al [6] numerous challenges arise in hand-crafted features for biometric recognition. For instance, it would take a considerable number of experiments to effectively find the best classical method for a certain biometric. Furthermore, many of the traditional methods are based on a multi-

class Support Vector Machine (SVM) trained in a one-vs-one fashion, which does not scale well when the number of classes is extensive [6]. Therefore, more sophisticated, and state-of-the-art methods are needed to consolidate the research on lip biometrics. Since biometric recognition has shifted from classical models to deep learning models, it is worthwhile to explore how lip-based identification will perform using state-of-the-art methods. This would entail automating the process of lip detection and identification. According to Hassanat et al [8], lip detection is an automatic process used to localize the lips in digital images. This is a crucial area of research because it is a fundamental step in many applications such as lip-reading applications, face recognition, and facial expression recognition [8]. Moreover, this area of research is crucial in a lip-based biometric system [8]. However, there are still many challenges to overcome in the field of lip detection [7, 8]. The lips comprise a very small part of the face and therefore, algorithms such as Faster R-CNN and YOLO require more processing time which greatly affects its speed at which it can make predictions [13]. Datasets are also often private [7] and do not have a balance of ethnicities, thus making it difficult to compare results. Additionally, low colour contrast between the lip and face regions, diversity of lip shape and the presence of facial hair and poor lighting conditions hinder the success of algorithms for lip detection [8]. Although lip detection techniques have been proposed, there is a significant room for further improvement in lip detection and identification.

## 2.2 Similar Work

In past decades, numerous image segmentation techniques have been proposed. However, only a few of these techniques have been applied to the task of lip detection due to the low chromatic and luminance contrast between the lips and the skin [14]. Therefore, a range of methods have been proposed to solve the problem of lip detection in coloured images which can be divided into three categories i.e., colour-based, model-based and hybrid-based techniques [14, 15].

### Colour-based Approach

Colour-based approaches use a preset colour filter that is able to differentiate between lip and non-lip pixels within a specific colour space [15] and it can segment the lip with good results where there is a high colour contrast. For example, Chang et al [16] segmented the lips by identifying skin regions in the image using chromaticity thresholding. The lips were detected by thresholding the RGB components of a bounding box below the nostrils and above the chin. The lip was then detected using the bounding box. Sadeghi et al [17] proposed a Gaussian mixture model of RGB values using a modified version of the predictive validation technique. Model parameters are chosen that allows the use of full covariance matrices. Thereafter, a Bayesian rule is used to label each pixel as lips or non-lips. Another traditional approach proposed by Shemshaki et al [18] uses chromatic and  $YCbCr$  colour spaces to detect the skin and lip pixel values. Based on the results this approach was accurate even in dealing with complex images and facial rotation. However, the approach adopted by Xinjun et al



[19] is slightly more flexible where they design colour filters consisting of both RGB and  $YUV$  components. These filters effectively threshold the pixels based on a colour ratio. Four different filters are cascaded to increase the robustness of lip pixel detection. Other approaches include clustering to perform colour-based lip detection. Beaumesnil and Luthon [20] propose using k-means clustering on the  $U$  channel from the  $LUX$  colour space to classify pixels as either lips or face. Skodras et al [21] improve this method by using k-means clustering to automatically adapt the number of clusters. Rohani et al [22] use fuzzy c-means clustering with preassigned number of clusters. FCM clustering is applied to each transformed image along with  $b$  components of the  $CIELAB$  colour space. Cheung et al [23] build on this by initialising an excess number of clusters, which are then reduced by merging clusters with coincident centroids. In more recent approaches, the power of neural networks is exploited to perform lip detection. Hassanat et al [8] use colour spaces to classify pixels as lips or non-lips using artificial neural networks. A novel method for fusing existing colour spaces was proposed which produced better results than individual colour spaces. More recently, Guan et al [24] proposed a new fuzzy deep neural network that integrates fuzzy units and traditional convolutional units. The convolutional units are used to extract salient features while the fuzzy logic modules are used to provide robust segmentation results with an end-to-end training scheme.

Although colour-based approaches are computationally inexpensive and allow rapid detection of the target region, Wang et al [25] discourage approaches that rely solely on colour information due to the low contrast between the lips and skin. Additionally, colour-based techniques are highly sensitive to variations in illumination and camera orientations, which alter all the pixel values [14]. Furthermore, Gritzman [14] states that some authors have also expressed concerns that the resulting segmentation is often noisy.

### Model-based Approach

Model-based approaches use previous knowledge of the lip contour to build a lip model and can be quite robust [15, 24]. The mostly widely used lip models are the active contour model (ACM), the active shape model (ASM), and the active appearance model (AAM). Model-based techniques are usually invariant to rotation, transformation, and illumination [15]. Liew et al [26] used a deformable geometric model to identify the lip shape. The model enables prior knowledge about the lips' expected shape, and it describes different shape variations. A stochastic cost function was proposed to describe the probability of the lip and non-lip regions was proposed. The results determined that nearly all the lip contours from around 2000 lips taken from 20 speakers could be extracted. In 2014, Sangve and Mule [27] did not only use colour spaces for lip detection like the aforementioned methods, but also introduced person recognition based on the lips. They did this by combining the region of interest with the AAM, ASM, and point distribution model (PDM). More recently in 2018, Lu and Liu [28] proposed a localised active contour model-based method using two initial contours in a combined color space to segment the lip region. Illumination equalisation is applied to the original images thereby decreasing the interference of uneven illumination.

Lip models provide geometric constraints for the final lip shape and reduce the influence caused by false edges to a certain extent. However, research [29] has shown that lip-model-based approaches depend on good initialisation and are still sensitive to noise boundaries caused by various backgrounds such as mustache and teeth.

### Hybrid Approach

Hybrid approaches are often combined with colour-based techniques and model-based techniques. By utilising a hybrid approach, the computational complexity of model-based techniques is reduced by using colour-based techniques to obtain a quick estimation of the lip region [14, 15]. Sensitivity to illumination and noisy segmentation of colour-based techniques is reduced by the smoothness and shape constraints of model-based techniques [14]. For example, Werda et al [30] propose a hybrid technique for lip Point Of Interest localisation using colour information to locate the mouth in the first stage, and a geometric model to extract the shape of the lip in the second stage. First a colour transform is applied to reduce the effect of lighting, then the horizontal and vertical projections are analysed to detect the corners of the mouth, and finally a geometric lip model is applied. Similarly, Mok et al [31] propose to segment the lips by first transforming the RGB image to the *CIELAB* colour space, thereafter, applying a fuzzy clustering method incorporating a shape function to obtain a rough estimation. An ASM is matched to the lips for accurate detection of lip contours. Tian et al [32] present a lip tracking method by combining lip shape, colour and motion information. The shape of the lip is modelled using two parabolas. The colour information of the lips and the skin is modelled as a Gaussian mixture and the motion of the lip is obtained by modified Lucas-Kanade tracking.

Based on previous works discussed above, the existing lip detection techniques achieve good segmentation results to a certain extent. Research demonstrates that state-of-the-art deep learning-based approaches are not yet introduced to the field of lip detection and recognition. With the rapid development of AI, deep neural networks have excelled in image processing and computer vision [6], providing a promising direction for solving the difficulties in lip detection. Therefore, the main objective of this work is to consolidate the research on lip biometric, in which a state-of-the-art deep learning-based approach is applied to a publicly available dataset.

## 3 Experiment Setup

### 3.1 Dataset Selection

Extracting visual features related to lip-based identification requires an accurate extraction of the lip. Therefore, selecting an appropriate dataset is crucial to the development of the study. Although large public databases such as “in-the-wild” datasets [33] are available for many facial analysis problems, they are not applicable to the problem domain of the current study because it would not adequately address the task at hand. Furthermore, they are subject to poor lighting, extreme pose, and strong

occlusions. Therefore, datasets produced in a controlled environment will be used. However, there is little consistency amongst the different face sets. Some sets have a very low resolution with poor visibility of lip patterns, while other sets have few ethnicities and the results from these datasets would not be effective. Therefore, the Chicago Face Database (CFD) [34], is employed for the current study. The CFD consists of high-resolution images of 597 male and female samples of different ethnicities. Each sample is represented with a neutral expression image. The CFD is particularly suited since it includes self-identified Asian, Black, Latino and White ethnicities of different ages. Currently, the CFD has limited data in terms of the number of samples per individual. Therefore, to overcome this drawback data augmentation is applied to artificially increase the data in the dataset. The methods used for data augmentation include the following modifications to the original image: horizontal flip, a rotation and an increase and decrease in brightness. Geometric transformations are helpful for positional biases present in the dataset [33]. Furthermore, altering the brightness of the original image is important because in real-life scenarios dramatic image variations arise from changes in illumination.

### 3.2 Proposed Architecture

YOLO is an effective state-of-the-art object recognition algorithm [35] which uses convolutional neural networks (CNNs) to detect objects. As the name suggests, the algorithm only requires a single forward propagation through the neural network to detect objects. Therefore, the prediction is done in a single algorithm run in the entire image. The original YOLO network architecture consists of 24 convolutional layers and 2 fully connected layers. The convolutional layers extract features while the fully connected layers calculate bounding box predictions and probabilities [36]. Over the last 5 years, the YOLO algorithm has progressed and updated to various variants such as YOLOv3, YOLOv4, YOLOv5 and the most recent, YOLOR.

YOLOR (“You Only Learn One Representation”) is the latest evolution of the YOLO models introduced by Wang et al [37]. YOLOR pretrains an implicit knowledge network with all of the tasks present in the dataset, namely object detection, instance segmentation, panoptic segmentation, keypoint detection, image caption, multi-label image classification, and long tail object recognition [37]. The concept of YOLOR is based on using implicit and explicit knowledge in conjunction with each other. According to Wang et al [37] explicit knowledge is provided to neural networks by providing thoroughly annotated image datasets or clear metadata. For neural networks, however, implicit knowledge is obtained by features in the deeper layers. The authors have combined both explicit and implicit learning in a unified network. This unified network creates a unified representation to serve a variety of tasks (mentioned above) at once. There are three processes by which this architecture is made functional: space reduction, kernel space alignment and more functions [37]. According to the results obtained, when implicit knowledge is introduced to the neural network that was already trained with explicit knowledge, the network benefits the performance of various tasks [37]. This novel approach propels YOLOR to the state-of-the-art for object detection

in the speed/accuracy tradeoff landscape. Therefore, it is employed for the current study.

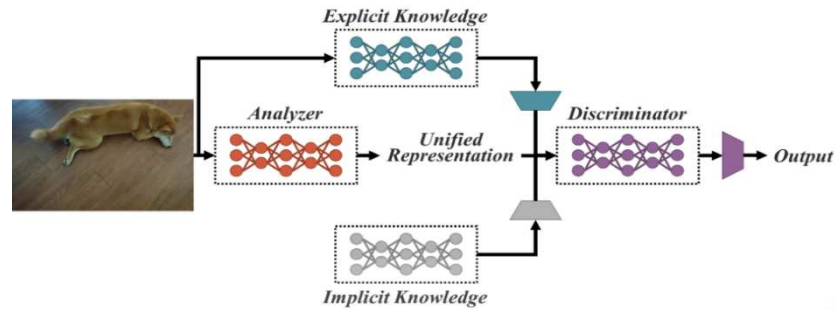


Fig 1. YOLOR concept with implicit and explicit knowledge [37]

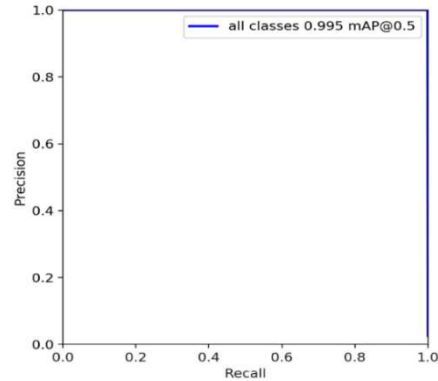
For the current study, the classification head for lip recognition has been modified by keeping only one fully connected layer with neurons equal to the number of samples in the training database. Since all YOLO anchor boxes are auto-learned in YOLOv5 (anchor box auto-learning has been integrated), the anchor boxes are modified using the YOLOv5 anchor box parameters which can be seen below:

Anchor box: [10,13, 16,30, 33,23] [30,61, 62,45, 59,119] [116,90, 156,198, 373,326]

Furthermore, the learning rate is set to 0.001 with Adam as the learning optimizer. The size of the input images has been modified to 448×448. The momentum is set to 0.937 with a decay of 0.0005. Lastly, the number of epochs chosen to train the model is set to 100 with a batch size of 8. Lastly, the benchmark parameters that will be used to assess the performance of the proposed model include mAP, Precision, Recall and a Precision × Recall (PR) Curve.

## 4 Results

Essentially, the training procedure consisted of 100 epochs where the dataset was split into 80% for training, 10% for validation and 10% for training. The fundamental and first step of the lip biometric system is to detect the lip within the image frame. This step is crucial and must be accurate as this affects the system's overall performance. Using all the predictions for detecting the lips within the images a smooth precision × recall curve was built. The PR curve for the model during testing is shown Figure 2 and the mAP graph is represented in figure 4. From figure 2 and figure 4 it is evident that the model achieved excellent results in terms of detecting the lip region with a mAP of 99.5%. Figure 4 shows the results of mAP for the different epochs obtained on the dataset with the YOLOR model.



**Fig 2.** Precision × Recall Curve of YOLOR model with a mAP of 0.995

Precision is a ratio between the number of positive results and the number of positive results predicted by the classifier. Considering the precision of the model, it is evident that it increases and decreases as the number of epochs progress until it finally stabilises. The overall precision achieved is 67% while the overall recall obtained is 99% which is represented in figure 4. Based on these results, the model has a slightly lower precision in comparison to the total recall. This indicates that although the model is classifying most of the samples correctly, it has a few false positives and some prediction noise. This could be due to the changes in illumination and rotations of the images.

After the models were trained and tested, they were also tested against images that had not been used during training. Figure 3 represents the detections made by the YOLOR model after training. It is evident that those images which were augmented by applying brightness and darkness transformations with rotations received a lower classification score than the original images. The original images received a classification score of 0.85 or higher which indicates the model's potential in making correct predictions. Although the results achieved for the augmented images are low in comparison to the original images, the brightness and darkness transformations are beneficial for the model to adapt to different illumination situations. Similarly, the rotation transformations are quite favourable for improving the robustness of the model. It is also worth noting that the developed model works fairly well on the artificially developed data.



**Fig 3.** Lip predictions made by the YOLOR model

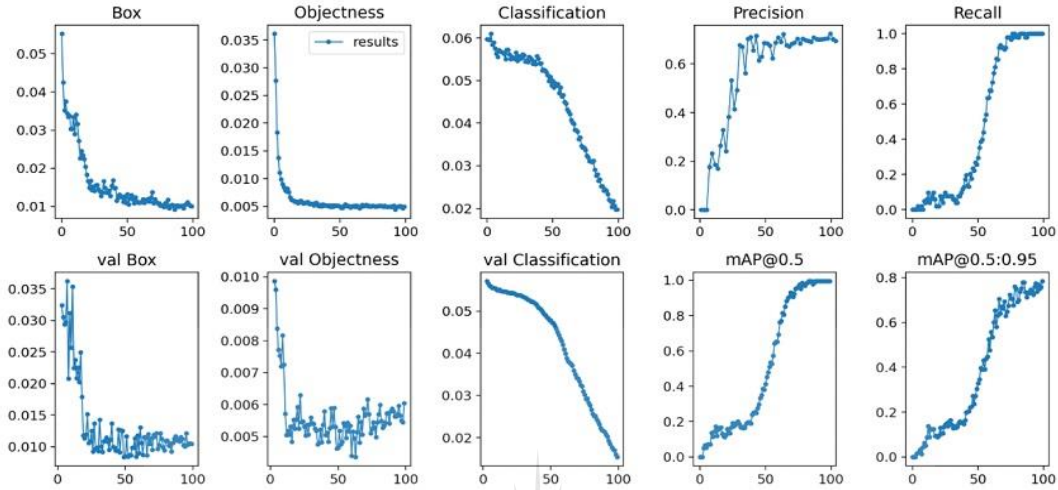


Fig 4. Lip Recognition results obtained including loss, classification loss, precision, recall and mAP for 100 epochs

## 5 Discussion

This work presented a deep learning architecture to detect lips from within an image frame. Based on the results discussed above, this model, as demonstrated in Figures 4 and 6, was well balanced between precision and recall and had a good confidence rate in its predictions. Additionally, this model also ensured a precision rate higher than 60% and a recall rate near 100%. Furthermore, the model was able to ideally detect the lips with an average mAP of 99.5%, thus highlighting its potential in the field of lip-based recognition. The results of this experiment further demonstrate that the YOLOR model can provide a high detection accuracy. The results can be compared to similar work that has used the CFD for iris segmentation which can be seen in Table 1. Based on the results it is noticeable that lip-based recognition in a short stretch of time will achieve recognition rates comparable to well-established modalities.

Table 1. Results of similar work reported on CFD

|                    | Modality | Method  | Dice | Precision | Recall |
|--------------------|----------|---|------|-----------|--------|
| Hurtado et al [38] | Iris     | A hybrid method for iris segmentation using facial landmark detection | 0.89 | 0.88      | 0.92   |

It is evident from previous works that more traditional feature engineering-based approaches have been used for the task of lip detection. Although good segmentation results were achieved, these approaches are discouraged [25] and have certain limitations. Deep learning-based approaches in this context are advantageous since they do not rely solely on colour information and are not sensitive to noise boundaries due

to the annotation of images. Furthermore, it provides an end-to-end learning scheme thereby completely automating the process of region selection, feature extraction and classification. Due to its powerful learning ability and advantages in dealing with occlusions, transformations and backgrounds, deep learning-based object detection has the potential to contribute to the field of lip detection and recognition.

## 6 Conclusion

The goal of the proposed work was to train a dataset using a deep learning architecture that can perform end-to-end lip-based identification. This study has employed a state-of-the-art deep learning architecture i.e., YOLOR to perform object detection and exhibited promising results. The average mAP obtained is 99.5% which indicates the model's potential in efficiently detecting the lips. A precision score of 67% and a recall score 99% is achieved by the respective model. The results of this experiment further demonstrate that the YOLOR model can provide high classification accuracy. An overview of the study indicates that deep learning architectures have potential for lip-based recognition. Previous work has demonstrated that traditional methods were mostly used to detect the lips and it highlighted that very little work has been done on lip-based recognition using deep learning architectures. This study has therefore fulfilled the apparent gap in this domain by not only using a deep learning architecture but one of the most recent architectures. The lip biometric has fewer research compared to other popular biometrics. It is relatively new and hence offers many research possibilities. Therefore, there are still many gaps in this area that can be researched and explored. One of the evident gaps within this domain is unconstrained lip recognition. Lip-based recognition has not been explored in the unconstrained environment. Therefore, there is a need to explore the power of deep learning-based methods to develop effective and efficient lip recognition methods in real-life scenarios.

## References

- [1] S. Boonkrong, *Authentication and Access Control: Practical Cryptography Methods and Tools*, vol. 11844, Apress, 2021, pp. 405-417.
- [2] A. Kamboj, R. Rani and A. Nigam, "A comprehensive survey and deep learning-based approach for human recognition using ear biometric," *The Visual Computer*, 2021.
- [3] Y. Tsuchihashi and T. Suzuki, "Studies on personal identification by means of lip prints," *Forensic Science*, vol. 3, pp. 233-248, 1974.
- [4] S. Sandhya and R. Fernandes, "Lip Print: An Emerging Biometrics Technology - A Review," *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIIC)*, pp. 1-5, 2017.

- [5] M. Choras, "Lips Recognition for Biometrics," *Advances in Biometrics*, pp. 1260-1269, 2009.
- [6] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun and D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," 2019.
- [7] C. Wright and D. Stewart, "One-Shot-Learning for Visual Lip-Based Biometric Authentication," *Advances in Visual Computing. ISVC 2019. Lecture Notes in Computer Science*, vol. 11844, 2019.
- [8] A. Hassanat, M. Alkasassbeh, M. Al-awadi and E. Alhasanat, "Colour-based Lips Segmentation Method using Artificial Neural Networks," in *IEEE Information and Communication Systems (ICICS)*, Amman, 2015.
- [9] S. Bakshi, R. Raman and P. Sa, "Lip pattern recognition based on local feature extraction," in *2011 Annual IEEE India Conference (INDICON)*, India, 2011.
- [10] S. Sandhya, R. Fernandes, S. Sapna and A. Rodrigues, "Comparative analysis of machine learning algorithms for Lip print," *Evolutionary Intelligence*, 2021.
- [11] S. K. Bandyopadhyay, S. Arunkumar and S. Bhattacharjee, "Feature Extraction of Human Lip Prints," *Journal of Current Computer Science and Technology*, vol. 2, no. 1, pp. 1-8, 2012.
- [12] B. Niu, J. Sun and Y. Ding, "Lip Print Recognition Using Gabor and LBP Features," *DEStech Transactions on Computer Science and Engineering*, 2017.
- [13] K. Fessel, "5 Significant Object Detection Challenges and Solutions," Medium, 2021. [Online]. Available: <https://towardsdatascience.com/5-significant-object-detection-challenges-and-solutions-924cb09de9dd>.
- [14] A. D. Gritzman, "Adaptive Threshold Optimisation for Colour-based Lip Segmentation in Automatic Lip-Reading Systems," University of the Witwatersrand, Johannesburg, 2016.
- [15] U. Saeed and J. L. Dugelay, "Combining Edge Detection and Region Segmentation for Lip Contour Extraction," in *Articulated Motion and Deformable Objects, 6th International Conference*, Port d'Andratx, 2010.
- [16] T. Chang, T. Huang and C. Novak, "Facial feature extraction from color images," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Computer Vision and Image Processing*, Israel, 1994.
- [17] M. Sadeghi, J. Kittler and K. Messer, "Modelling and segmentation of lip area in face images," *Proceedings of IEE Conference on Vision, Image and Signal Processing*, vol. 149, no. 3, p. 179–184, 2002.
- [18] M. Shemshaki and R. Amjadifard, "Lip Segmentation Using Geometrical Model of Color Distribution," *7th IEEE Iranian Machine Vision and Image Processing*, 2011.



- [19] M. A. Xinjun and Z. Hongqiao, "Lip Segmentation Algorithm Based on Bicolor Space," in *Proceedings of the 34th Chinese Control Conference*, Hangzhou, 2015.
- [20] B. Beaumesnil and F. Luthon, "Real time tracking for 3d realistic lip animation," *Pattern Recognition*, vol. 1, pp. 219-222, 2006.
- [21] E. Skodras and N. Fakotakis, "An unconstrained method for lip detection in color images," *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1013-1016, 2011.
- [22] R. Rohani, S. Alizadeh, F. Sobhanmanesh and R. Boostani, "Lip segmentation in color images," *Innovations in Information Technology*, pp. 747-750, 2008.
- [23] Y. Cheung, M. Li, X. Cao and X. You, "Lip segmentation under map-mrf framework with automatic selection of local observation scale and number of segments," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3397-3411, 2014.
- [24] C. Guan, S. Wang and A. W.-C. Liew, "Lip Image Segmentation Based on a Fuzzy Convolutional Neural Network," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1242-1251, 2020.
- [25] S. Wang, A. Liew, W. Lau and S. & Leung, "Lip region segmentation with complex background," *Visual Speech Recognition: Lip Segmentation and Mapping: Lip Segmentation and Mapping*, p. 150, 2009.
- [26] A. W. C. Liew, S. H. Leung and W. H. Lau, "Lip contour extraction from color images using a deformable model," *Pattern Recognition*, vol. 35, p. 2949 – 2962, 2002.
- [27] S. Sangve and N. Mule, "Lip Recognition for Authentication and Security," *IOSR Journal of Computer Engineering*, vol. 16, no. 3, pp. 18-23, 2014.
- [28] Y. Lu and Q. Liu, "Lip segmentation using automatic selected initial contours based on localized active contour model," *EURASIP Journal on Image and Video Processing*, 2018.
- [29] Y. M. Cheung, M. Li, X. C. Cao and X. G. You, "Lip segmentation under MAP-MRF framework with automatic selection of local observation scale and number of segments," *IEEE Transactions on Image Processing*, vol. 23, no. 8, p. 3397–3411, 2014.
- [30] S. Werda, W. Mahdi and A. B. Hamadou, "Colour and geometric based model for lip localisation: Application for lip-reading system," *14th International Conference on Image Analysis and Processing*, pp. 9-14, 2007.
- [31] L. L. Mok, W. H. Lau, S. H. Leung, S. L. Wang and H. Yan, "Person authentication using ASM based lip shape and intensity information," *2004 International Conference on Image Processing*, vol. 1, pp. 561-564, 2001.
- [32] Y. Tian, T. Kanade and C. J., "Robust lip tracking by combining shape, color and motion," *4th Asian Conference on Computer Vision*, pp. 1040-1045, 2000.

- [33] B. Johnston and P. D. Chazal, "A review of image-based automatic facial landmark identification techniques," *EURASIP Journal on Image and Video Processing*, 2018.
- [34] D. Ma and B Wittenbrink, "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data," *Behavior Research Methods*, vol. 47, pp. 1122-1135, 2015.
- [35] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016.
- [36] M. Buric, M. Pobar and M. Ivašić-Kos, "Ball Detection Using Yolo and Mask R-CNN," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018.
- [37] C.-Y. Wang, I.-H. Yeh and H.-Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks," 2021.
- [38] F. Fuentes-Hurtado, V. Naranjo, J. A. Diego-Mas and M. Alcañiz, "A hybrid method for accurate iris segmentation on at-a-distance visible-wavelength images," *EURASIP Journal on Image and Video Processing*, 2019.

