UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

Facial Attractiveness for Enhanced Face Recognition: A Novel Soft Biometric Trait

by

Moneera Habeeb Alnamnakani

ORCiD: 0009-0000-8266-3534

A thesis for the degree of
Doctor of Philosophy

October 2025

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

Doctor of Philosophy

Facial Attractiveness for Enhanced Face Recognition: A Novel Soft Biometric Trait

by Moneera Habeeb Alnamnakani

This research introduces facial attractiveness as a significant new feature in soft biometrics, aimed at enhancing the accuracy of automated facial recognition systems. Despite its potential to improve recognition systems, facial attractiveness, unlike other standard features such as age, gender, and skin colour, has not received extensive research attention. This study addresses this gap by demonstrating that facial attractiveness can serve as an additional and valuable attribute for identifying individuals.

This research employs a comparative analysis approach to measure facial attractiveness through a structured evaluation method that combines soft biometric data with machine learning techniques. The method involves collecting and ranking attractiveness attributes using the Elo rating system, which are then integrated into recognition models. Experiments indicate that facial attractiveness improves recognition performance, proving its usefulness in automated systems.

Additionally, this research examines the psychological and social aspects of facial attractiveness and considers how they can influence the functionality of automated systems. It discusses the challenges of measuring attractiveness consistently across different datasets, providing a clear overview of the limitations.

Notably, the findings demonstrate that models using soft-biometric attributes alone, including attractiveness, consistently outperformed systems that combined these attributes with Eigenface representations. This outcome underscores the discriminative strength of soft biometrics as standalone inputs rather than supplementary enhancements.

Although this study does not use deep learning techniques, it establishes a foundation for future research by proposing an innovative approach to incorporating facial attractiveness into biometric systems. The author avers that this is the first study to thoroughly explore facial attractiveness as a soft biometric feature.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:... Date:..................

Acknowledgements

I would like to express my deepest gratitude to my God (Allah) for His endless blessings, guidance, and mercy throughout my academic journey.

I am sincerely thankful to my supervisor, Professor Mark Nixon, whose mentorship and support—despite his retirement—have profoundly shaped my academic and personal growth. His dedication and belief in my potential inspired me to strive beyond my limits. I also extend my heartfelt gratitude to Professor Adam Prugel-Bennett for his unwavering encouragement and wise counsel. Together, they safeguarded my academic interests during challenging times and strengthened my resolve to persevere. I am equally grateful to Professor Mahesan Niranjan, whose insightful comments and constructive feedback significantly strengthened both my scientific knowledge and the quality of my results.

To my beloved son, Abdul-Elah, your maturity, sense of responsibility, and unwavering care for your brothers have been a constant source of pride and comfort. You stood by me through the hardest moments and gave me the strength to carry on. I am deeply grateful for your love, patience, and dedication. To my dearest sons, Abdul-Majeed and Abdul-Aziz, your laughter, affection, and boundless energy have been a continual reminder of joy and purpose. Each smile rekindled my determination and reminded me why I embarked on this journey. To my beloved children, you are my greatest inspiration and the driving force behind every step I have taken. Without your love and presence, this journey would not have been possible. I also wish to honour the memory of my late parents, whose values and prayers continue to guide me. I pray I have lived up to their aspirations.

I am profoundly grateful to my dear friends, Basmal Sedeeq and Shatha Al-Hussayyn. Basmal created a nurturing, home-like refuge in the most challenging of times, while Shatha's boundless energy and positive spirit gave me the courage to face every obstacle with confidence. Your friendship ensured that I never felt alone. My sincere thanks also go to my colleague, Dr. Musharraf Alruwaill, who stood by me unwaveringly until the very end, offering guidance, kindness, and support not only as a colleague but with the generosity of a brother and the wisdom of a father, and whose belief in my capabilities and steadfast support strengthened my determination.

Finally, I extend my heartfelt appreciation to the government of Saudi Arabia for its continuous support and to the University of Southampton for providing an exceptional academic environment and invaluable resources. To all who have contributed to my journey—through guidance, kindness, or encouragement—your impact has been profound and enduring.

Chapter 1

Context and Contributions

This chapter describes the research context, outlines its contributions, and introduces the thesis structure. It highlights the importance of facial attractiveness as a soft biometric and explains how it will be explored in the following chapters.

1.1 Context

Biometric technology has transformed identity verification by analysing individuals' unique physical or behavioural traits. While traditional methods—such as fingerprints, DNA, and iris scans—are highly effective, they require the subject's presence. In contrast, soft biometrics (e.g., gender, height, and facial attributes) are generally easier to observe and can often be collected from a distance (Jain et al. (2006)).

face recognition is one of the most widely used biometric approaches because it does not require physical contact (Ali and Gaber (2023)), making it appropriate for security, surveillance, law enforcement, and personal device access (Sinha et al. (2006)). Moreover, the human face holds a wealth of information, including gender, age, emotional expressions, and attractiveness (Cross et al. (1971)). However, changes in lighting, pose, or image resolution can limit the effectiveness of traditional face recognition systems. Soft biometrics help by introducing extra flexibility and enhancing reliability, addressing many challenges (Guo and Zhang (2019*a*)).

1.1.1 Importance of Facial Biometrics

face recognition supports security and surveillance by identifying individuals on CCTV, enhancing public safety (Reid et al. (2013)). In law enforcement, it can identify suspects and assist in crime resolution without direct physical contact. However, the requirement for high-quality images can reduce effectiveness in uncontrolled environments. Soft biometrics help resolve these issues by providing additional data and making these systems more adaptable (Guo et al. (2019), Ben Fredj et al. (2021), Hassan et al. (2021*b*)).

1.1.2 The Role of Soft Biometrics

Soft biometrics include traits like skin colour, eyebrow shape, and facial attractiveness. They require minimal cooperation from the individual, can be acquired from a distance, and work alongside traditional biometrics to improve identification accuracy. They also bridge the gap between how humans naturally describe people—through terms like "tall," "dark-haired," or "attractive"—and the more rigid methods that computers use for identification (O'Toole et al. (2002)).

1.1.3 Facial Attractiveness as a Biometric Trait

Although researchers have investigated the potential impact of perceived attractiveness on face recall and recognition (Davies and Valentine (2007), O'Toole et al. (2002)), facial attractiveness has not been widely studied as a soft biometric. While prior studies investigated perceived attractiveness in facial recall, this work was among the first to quantitatively operationalise it as a comparative soft biometric. Since people tend to remember attractive faces more efficiently, this trait could be especially valuable in legal or investigative situations. For instance, a witness might describe a suspect as "attractive," providing a valuable lead for an investigation (Davies and Valentine (2007)). Attractiveness also appears to be less affected by changes in lighting or pose, suggesting it could serve as a stable attribute for face recognition (O'Toole et al. (2002)). Based on these observations, this study introduces facial attractiveness as a soft biometric trait.

1.1.4 Bridging the Semantic Gap

Soft biometrics help close the gap between how humans describe individuals and how machines interpret those descriptions. Because they do not require close contact and can adapt to various real-world conditions, they also lend themselves to analysing eyewitness reports, which is highly relevant in law enforcement and forensic work (Alnamnakani et al. (2019); Almudhahka et al. (2017)). The current study investigates whether facial attractiveness—often mentioned in witness accounts—could be an effective soft biometric. By systematically assessing the impact of attractiveness on identification accuracy, it demonstrates how integrating this trait could enhance face recognition systems.

Overall, these insights suggest that soft biometrics—especially facial attractiveness—can address some of the limitations of traditional face recognition. By relying on naturally observable traits, these systems may become more flexible and dependable in diverse settings. Therefore, this study examines how attractiveness, frequently noted in eyewitness accounts, could function as a valuable soft biometric, potentially boosting accuracy and adaptability in various applications.

1.2 Contributions

This thesis explores the role of facial attractiveness as a soft biometric trait and its impact on enhancing recognition accuracy. The following contributions were made, each supported by empirical findings and detailed methodology across Chapter 3 to Chapter 7:

1. Facial attractiveness was modelled quantitatively using a comparative approach. While prior studies focused on psychological perceptions of attractiveness (Alnamnakani et al. (2019); O'Toole et al. (2002); Davies and Valentine (2007)), this thesis developed a structured framework for converting perceived attractiveness into a usable biometric descriptor for machine learning (Chapter 3).

2. A novel annotation pipeline was designed, where annotators provided relative pairwise comparisons using a three-point scale. These were transformed into quantitative scores using Elo rating, resulting in interpretable, scalable attractiveness features (Chapter 3).

3. A comparative recognition framework was introduced, enabling face matching based on relative rather than absolute attributes. This improved performance in unconstrained environments and expanded the applicability of attractiveness-based features in practical scenarios (Ben Fredj et al. (2021); Guo and Zhang (2019*a*)) (Chapters 3 and 4).

4. Experiments showed that attractiveness, when used as an independent soft-biometric feature, improved both verification and identification accuracy, especially under variations in pose, lighting, and resolution. The improvements were validated statistically through cross-validation and paired t-tests (Jain et al. (2006); Dantcheva et al. (2011)) (Chapters 4 and 6).

5. An "attractiveness-guided recognition" strategy was proposed, where attractiveness directly influenced recognition decisions. This concept was implemented and evaluated in the recognition pipelines presented in Chapter 5 (Kim et al. (2023)).

6. It was demonstrated that combining attractiveness with other soft-biometric features—such as age, gender, and skin colour—enhanced performance. However, attribute-only models (i.e., without Eigenfaces) consistently outperformed eigenface-enhanced systems, even when using the same feature set. This finding, detailed in Chapter 6, highlights the strength of soft biometrics as standalone inputs rather than supplementary enhancements (Almudhahka et al. (2017); Alnamnakani et al. (2019)).

Together, these contributions support the use of facial attractiveness as a measurable and effective soft biometric trait. The findings are intended to inform the development of more human-centred biometric systems that can operate reliably in real-world conditions.

1.3 Soft Biometric Processes

Soft biometric identification merges human perception with computational methods to visually describe and characterise individuals, focusing on facial attractiveness (Reid et al. (2013)). Human annotators provide attractiveness ratings, which a ranking system (e.g., Elo) converts into consistent, quantifiable scores. Machine learning then uses these scores for identification tasks, such as identifying individuals who match certain attractiveness traits reported by witnesses (O'Toole et al. (2002)). This method imposes an objective structure on subjective opinions, enhancing precision and utility in face recognition.

1.4 Synthesising Biometrics and Attractiveness

Uniting human perception, image analysis, and machine learning is central to this research. As shown in Figure 1.1, facial attractiveness is integrated into biometric systems through the following steps:

- **Human Perception:** Human observers play a crucial role in evaluating and annotating facial attractiveness. These ratings then serve as a baseline for training machine learning models, which are later analysed by the system (Reid et al. (2013),Davies and Valentine (2007), Cross et al. (1971)).
- **Image Analysis:** Once human annotations are collected, image analysis tools transform them into structured data formats. This structured data enables further processing by machine learning algorithms (Sinha et al. (2006), Hassan et al. (2021*a*)).

- **Machine Learning:** Machine learning automates the analysis, prediction, and recognition of attractiveness-based traits. By leveraging annotated datasets, algorithms are trained to evaluate new, unseen images, replicating human-like judgement. This process enhances biometric systems by providing consistent and scalable evaluations of facial attractiveness (Ben Fredj et al. (2021), Guo and Zhang (2019*a*), O'Toole et al. (2002)).

- **Facial Attractiveness:** Facial attractiveness plays a critical role in refining recognition systems. By leveraging comparative analysis, it enhances accuracy and robustness, making it an essential feature for biometric systems (Kim et al. (2023), Sinha et al. (2006), Zebrowitz (2018)).

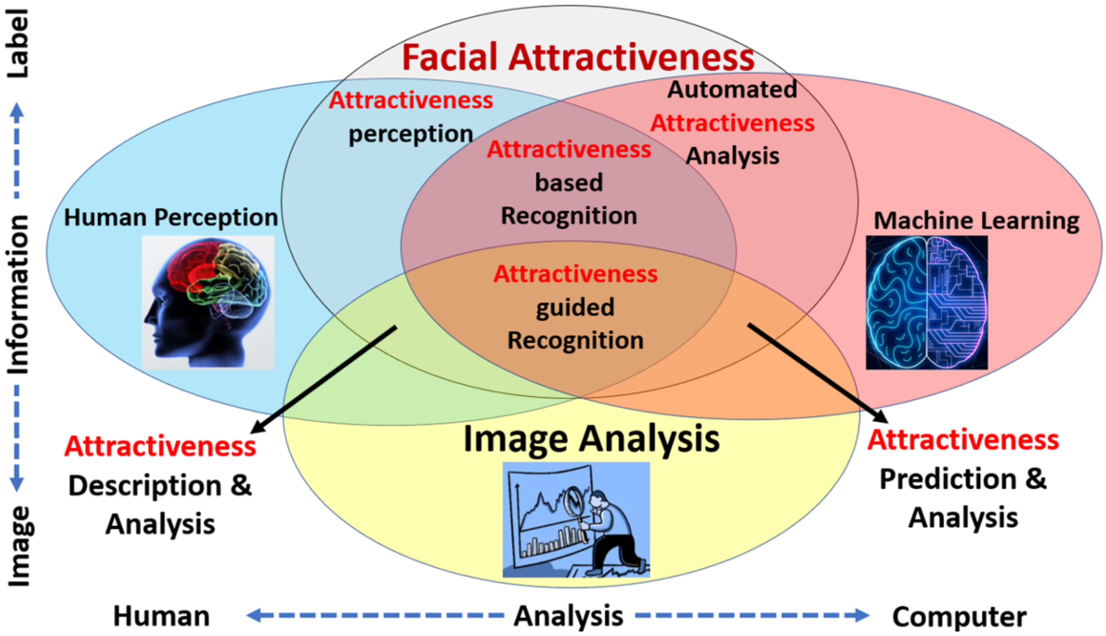


FIGURE 1.1: This illustrates the convergence of human perception, image analysis, and machine learning to integrate facial attractiveness into biometric systems. Human annotators assign attractiveness labels, image analysis converts those labels into structured data, and machine learning automates the process of analysing, predicting, and applying these traits (Kim et al. (2023), Alnamnakani et al. (2024)).

Applying an objective framework to subjective impressions transforms human opinions into data-driven methods that significantly advance biometric technology. Figure 1.1 highlights how each component contributes to a more flexible and reliable system, enabling facial attractiveness to play a key role in modern identification tasks.

1.5 Publications

The following publications have ensued from this study:

1. Alnamnakani, M., Mahmoodi, S. and Nixon, M., 2019, October. ”On the Potential for Facial Attractiveness as a Soft Biometric.” International Symposium on Visual Computing (pp. 516-528). Springer, Cham.

2. Alnamnakani, M., Mahmoodi, S. and Nixon, M., 2024. Using Facial Attractiveness as a Soft Biometric Trait to Enhance Face Recognition Performance. In Face Recognition Across the Imaging Spectrum (pp. 31-55). Singapore: Springer Nature Singapore.

1.6 Thesis Overview

This thesis comprises eight chapters, each building on concepts introduced in the first chapter, where the background of the study and its importance were provided. A more detailed examination of the subsequent chapters is presented below.

- **Chapter 2:**
  This chapter considers facial attractiveness a valuable soft biometric. Research in psychology and neuroscience shows that certain universal traits affect how people perceive attractiveness, and data from the real world indicate that it has a significant impact on society and the economy. Although cultural factors add complexity, studies suggest that attractiveness could enhance face recognition systems, even though it has been explored only in a limited manner in biometric research.

- **Chapter 3:**
  This chapter introduces a comparative approach to facial attractiveness. Crowdsourced annotations based on a three-point scale are converted into Elo ratings, generating dynamic and relative scores. This method effectively uses attractiveness as a soft biometric, which improves the accuracy of recognition and gives a way to add it to automated systems.

- **Chapter 4:**
  This chapter examines facial attractiveness as a soft biometric trait. Correlation analysis reveals strong links to youthfulness, smooth skin, and symmetry. Feature-selection methods (MI and SFFS) identify attractiveness as highly discriminative, comparable to age and gender. Experiments demonstrate that adding attractiveness raises verification and identification rates, and statistical tests confirm the significance of these improvements.

- **Chapter 5:**
  This chapter describes an automated face recognition pipeline that incorporates facial attractiveness alongside age, gender, and skin colour. It covers image

preprocessing and facial feature extraction, culminating in a PCA/Eigenfaces approach. This streamlined method sets the stage for the next chapter's investigation of how attractiveness influences recognition accuracy.

- **Chapter 6:**
  This chapter demonstrates how facial attractiveness can be integrated into an automated face recognition system. It employs the Eigenface (PCA) method for dimensionality reduction and an RBF SVM on the LFW dataset to form a clear baseline.

- **Chapter 7:**
  This chapter highlights the final results, demonstrating how incorporating facial attractiveness as a soft biometric feature can enhance face recognition systems. When combined with existing facial attributes, attractiveness provides distinct and complementary information, leading to improved identification accuracy.

- **Chapter 8:**
  In the concluding chapter, the outcomes of the study are summarised. The study also offers suggestions for future research, focusing on exploring facial attractiveness to enhance biometric systems.

In summary, this chapter clarifies the significance of the study's background and establishes a foundation for the detailed discussions and findings in the ensuing chapters. Although deep learning techniques (CNNs/Transformers) are beyond the scope of this study, the findings here establish a foundational framework that could be extended in future research to leverage advanced architectures.

Chapter 2

Soft Biometrics and Facial Attractiveness

Building on the ideas presented in Chapter 1, this chapter examines facial attractiveness as a soft biometric trait and explores its potential impact on recognition systems. It begins by defining facial attractiveness and discussing its effects on individuals in social, psychological, and technological contexts, including relationships, career opportunities and legal outcomes. Recognising facial attractiveness as a soft biometric attribute shows how it could improve recognition systems. Finally, the chapter introduces key concepts in biometrics and soft biometrics, preparing the reader for the discussions that follow.

2.1 Attractiveness of the Human Face

Facial attractiveness involves not merely physical beauty; it concerns a complex interaction of personal presentation, psychological perception, and cultural interpretation. Research suggests that people focus more on attractive faces, which can influence social interactions, job opportunities, and overall treatment in society (Langlois et al. (2000), Liu, Fan, Samal and Guo (2016)). Advances in neuroscience have demonstrated specific brain activity patterns when individuals observe attractive faces (Fan et al. (2012), Mitsuda and Yoshida (2006)). Concurrently, artificial intelligence (AI) techniques are increasingly used to assess and predict facial attractiveness in fields such as fashion, entertainment, and marketing (Bougourzi, Dornaika and Taleb-Ahmed (2022), Guo and Zhang (2019*a*)).

In biometrics, these scientific and technological developments emphasise the potential of facial attractiveness as a soft biometric trait (Nixon et al. (2015), Reid et al. (2013)).

2.1.1 Expression of Facial Attractiveness

Despite their frequently interchangeable usage, beauty and attractiveness are distinct concepts. According to Langlois et al. (2000), beauty principally refers to physical traits and aesthetic attractiveness, while attractiveness encompasses a broader range of qualities such as charisma, personality, and social attractiveness. Consequently, attractiveness transcends cultural beauty standards and highlights the overall qualities that make someone attractive (Bougourzi, Dornaika and Taleb-Ahmed (2022), Langlois et al. (2000)).

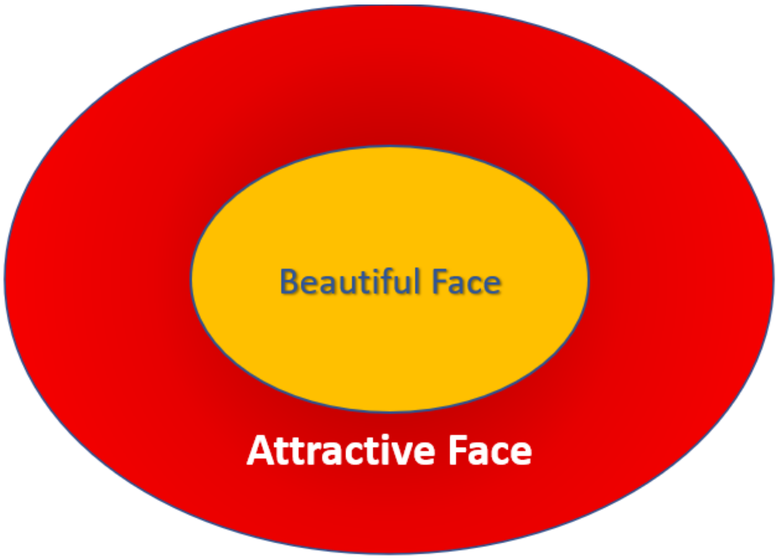


FIGURE 2.1: Relationship between beauty and attractiveness.

Figure 2.1 illustrates how beauty fits within the broader concept of attractiveness. While beautiful faces are generally considered attractive, not all attractive faces necessarily conform to conventional beauty standards. This distinction highlights the complexity of attractiveness, which includes physical attributes and psychological and behavioural dimensions that shape human perception.

Studies suggest that attractiveness involves physical, social, and psychological elements. For instance, Geldart found that people tend to spend more time observing faces they find attractive rather than those they deem merely beautiful. This result supports the Gestalt principle that the whole is greater than the sum of its parts (Geldart (2010)). Additionally, a strong correlation (r = 0.94) between beauty and attractiveness confirms their overlap. Attractiveness is often evaluated holistically, incorporating traits such as confidence and charisma (Wagemans et al. (2012)).

In practical contexts, such as witness testimonies, attractiveness influences how well people remember specific faces. Research indicates that attractive individuals tend to make a stronger and more lasting impression owing to the integration of their physical traits with social and behavioural cues (Malloy et al. (2021)). This is particularly important in legal settings, where witnesses are more likely to recall and identify individuals with attractive physical traits and engaging behavioural qualities, such as confidence or a friendly demeanour. These findings confirm that attractiveness influences visual attention, emotional engagement, and memory retention (Rhodes (2006), Malloy et al. (2021)).

In summary, attractiveness is much more than mere physical appearance and can significantly impact human interactions, especially in law enforcement and legal contexts where memory and judgement are crucial. After discussing the perception of attractiveness, the next section delves into its broader social and economic implications.

2.1.2 The Value of Attractive Faces

The impact of beauty and attractiveness on people's lives has been recognised since ancient Greece, when it was widely believed that beauty significantly influenced personal and societal outcomes (Laurentini and Bottino (2014)). One emblematic classical trope refers to Helen of Troy, whose legendary beauty is said to have "launched a thousand ships," illustrating the perceived power of attractiveness throughout history. Facial attractiveness shapes everyday social interactions and broader societal trends (Fan et al. (2012)). Its impact is evident in areas such as modelling, acting, and digital entertainment, as well as career advancement, financial stability, and personal relationships (Liu, Luo, Wang and Tang (2016)).

People often receive differential treatment based on their appearance, despite the ideals of equality. In societies where physical attractiveness is highly valued, individuals considered unattractive or unusual may face challenges such as low self-esteem, social isolation, and even mental health issues like depression (Grammer et al. (2003), Zhang, Chen, Xu et al. (2016)). Consequently, attractiveness often becomes a key factor when making new connections or meeting people (Laurentini and Bottino (2014)).

While beauty standards can vary across cultures, recent studies suggest that perceptions of facial attractiveness are not entirely subjective. Research across cultures shows that people from various backgrounds tend to agree on which facial features are attractive. For instance, infants—regardless of their cultural surroundings—tend to gaze longer at faces that adults consider attractive (Fan et al. (2012), Rhodes (2006)).

These results suggest that facial attractiveness may have biological foundations. Traits like symmetry, averageness, and unblemished skin are regarded as attractive, regardless of cultural differences. Studies that rank facial attractiveness often show high levels of agreement across cultures, indicating that health and genetic fitness characteristics are universally attractive (Rhodes (2006), Grammer et al. (2003)).

Therefore, while cultural influences shape beauty ideals, growing research points to universal biological and psychological factors that guide people's perceptions of facial attractiveness. This broader understanding helps explain why facial attractiveness significantly impacts social interactions, professional opportunities, and legal proceedings.

2.1.2.1 Attractiveness in Social, Economic, and Judicial Contexts

Facial attractiveness significantly influences decisions in legal, economic, and social situations. It can affect hiring decisions, legal rulings, and social interactions. Research indicates that attractive defendants are often treated more leniently in court—being less likely to be convicted and, when convicted, receiving shorter sentences—relative to less attractive counterparts (Yang et al. (2019)).

FIGURE 2.2: Jeremy Meeks's life changed dramatically, moving from crime to fame and the fashion world because of his attractiveness (French (2015)).

A notable example is Jeremy Meeks (Figure 2.2), whose physical attractiveness helped him transition from a criminal background to a successful modelling career (French (2015)). Similarly, serial killer Ted Bundy is an example of how charm and good looks enable a criminal to earn the trust of unsuspecting victims. Even after he confessed, some people were drawn to him and continued to defend him (Figure 2.3) (Etcoff (1999)).

Facial attractiveness also affects economic decisions. Research shows that job applicants perceived as more attractive often enjoy an advantage in hiring processes, reflecting biases in professional environments (Rich and Ashby (2014)). This preference

FIGURE 2.3: Ted Bundy: An example often cited for how perceived attractiveness and charm can influence trust.

has contributed to the growth of the cosmetics and plastic surgery industries, where spending on appearance often exceeds investment in education and social services (Laurentini and Bottino (2014)).

These findings highlight the significant impact of facial attractiveness, including its advantages and the biases it may create. Recognising attractiveness as an important trait can enhance automated recognition systems, improving accuracy and robustness (Liu, Luo, Wang and Tang (2016)).

2.1.3 Facial Attractiveness in Philosophy and Psychology

Philosophers, psychologists, and scientists have long debated how to define or measure facial beauty, from Aristotle's writings over two millennia ago to Darwin's scientific enquiries (Kagian et al. (2006), Bougourzi, Dornaika and Taleb-Ahmed (2022)). These debates focus on what makes a face attractive, why it matters, and which factors play a role, particularly questioning whether attractiveness is subjective or objective. The familiar saying "Beauty is in the eye of the beholder" reflects a view that facial attractiveness is deeply personal, non-deterministic, and culturally influenced (Laurentini and Bottino (2014), Liu et al. (2019)).

However, the possibility of attractiveness having objective components has been discussed for centuries in philosophy, psychology, and biology. Research shows that infants—unaffected by cultural norms—look longer at faces that adults consider attractive. Moreover, adults from different cultural or racial backgrounds show similar ratings of faces, even when those faces are from a less familiar racial group (Fan et al. (2012), Kagian et al. (2006), Grammer et al. (2003), Liu, Fan, Samal and Guo (2016), Rhodes (2006)). This parallel between infant and adult perceptions suggests a shared standard of attractiveness.

Recent studies reveal that human perception of facial attractiveness may include objective elements acknowledged across diverse backgrounds. Some components of beauty could be rooted in human biology rather than dictated solely by social or cultural norms (Liu, Fan, Samal and Guo (2016), Liu et al. (2019)). Consequently, numerous empirical rating studies have found that people of various ethnicities, ages, genders, and societies generally agree on which faces are more attractive (Elmer and Houran (2020), Luo et al. (2019)). This broad consensus implies that facial attractiveness does contain universal qualities, making it a compelling topic for systematic and objective research (Liu, Fan, Samal and Guo (2016)).

2.1.4 Facial Attractiveness in Psychophysiology and Neuropsychology

Psychophysiological and neuropsychological research has identified specific brain regions that process facial attractiveness. Functional MRI scans reveal distinct patterns of brain activity when individuals assess the attractiveness of two-dimensional images. Notably, these patterns can be non-linear, meaning highly attractive and unattractive faces elicit strong neural responses. Studies using near-infrared spectroscopy (NIRS) (Mitsuda and Yoshida (2006)) and event-related potentials (ERP) (Schacht et al. (2008)) reinforce these observations. Although research in this field is nascent, these findings suggest a potential for more "objective" methods of evaluating attractiveness without relying solely on predefined categories (Laurentini and Bottino (2014), Bottino and Laurentini (2010)).

2.2 Facial Attractiveness in Machine Perception

While the human brain can effortlessly judge facial attractiveness, enabling machines to do the same remains a complex challenge (Liu, Fan, Samal and Guo (2016)). Numerous studies have attempted to evaluate and quantify facial attractiveness using computer technology (Bougourzi, Dornaika and Taleb-Ahmed (2022), Lebedeva et al. (2021)). Consequently, techniques for attractiveness estimation are now used in various applications, including social media, aesthetic surgery planning, modelling, entertainment, magazine covers, and recruitment (Bougourzi, Dornaika, Barrena, Distante and Taleb-Ahmed (2022)).

More recent approaches have expanded these methods to enhance facial images, recommend social network connections, offer cosmetic advice, and guide aesthetic surgery (Saeed and Abdulazeez (2021)). These advances highlight growing expertise in identifying the elements that enhance a face's attractiveness to observers.

To understand facial attractiveness computationally, researchers have developed new applications and methods that rely on AI, image processing, and pattern recognition (Saeed and Abdulazeez (2021), Kagian et al. (2006)). Machine learning, deep learning, and computer vision efforts aim to define attractiveness more precisely, enabling automated systems to evaluate it in a way that mirrors human perception.

Although research on computer-based facial attractiveness is still in its early stages, it has the potential to greatly refine how technology aligns with human perception (Laurentini and Bottino (2014)). It may well transform current studies, offering fresh opportunities where technology and human perception intersect. Soft biometric attributes, such as facial attractiveness, could significantly enhance face recognition systems. However, technical and ethical challenges remain. Variations in pose, lighting, and resolution can complicate the accurate integration of attractiveness as a biometric trait. Additionally, privacy and algorithmic bias issues must be addressed to ensure fair and responsible implementation of such systems.

2.2.1 Facial Attractiveness in Soft Biometrics

Though recent advances in computer vision, machine learning, and deep learning have increased interest in facial attractiveness, its role as a soft biometric remains relatively understudied. Traditional biometrics typically focus on fixed physical traits, whereas soft biometrics encompass descriptive features like attractiveness that have yet to be thoroughly examined. For instance, (Kumar, Berg, Belhumeur and Nayar (2011)) initially labelled "Attractive Man" and "Attractive Woman" as soft biometric categories but did not investigate them thoroughly. Consequently, the present research offers a comprehensive, in-depth analysis of how attractiveness may affect facial recognition accuracy and whether incorporating it could improve identification results.

Moreover, this section highlights the significance of facial attractiveness in biometrics and emphasises its potential impact on various areas of life. It also refers to neuroscientific findings on facial attractiveness, encouraging researchers to integrate these insights more extensively into biometric systems.

2.3 Facial Soft Biometrics

Biometrics involves measuring people's characteristics for identification. Alphonse Bertillon developed one of the earliest biometric systems, known as "Bertillonage" in 1879, using anthropometric measurements. Figure 2.4 shows this method on a Bertillonage identification card (Rhodes (1968)). Nixon et al. (2015) traces how soft biometrics evolved from Bertillon's work to modern approaches that extract distinct measurements from surveillance footage.

Security and law enforcement primarily use hard biometrics, such as fingerprinting, iris and voice recognition, and DNA profiling. Previous studies often combined these methods with soft biometric descriptors, but hard biometric approaches typically require cooperation from individuals, which can be difficult in practice (Kittler

FIGURE 2.4: Alphonse Bertillon depicted on a Bertillonage identification card (Rhodes (1968)).

et al. (1998), Reid et al. (2013), Nixon et al. (2015)). This reliance on collaboration has prompted increased interest in soft biometrics, which focus on visible physical attributes and observable behaviours rather than strictly fixed traits (Reid et al. (2013)). Advances in soft biometrics technology also drive this shift. Figure 2.5 illustrates the various types of biometrics.

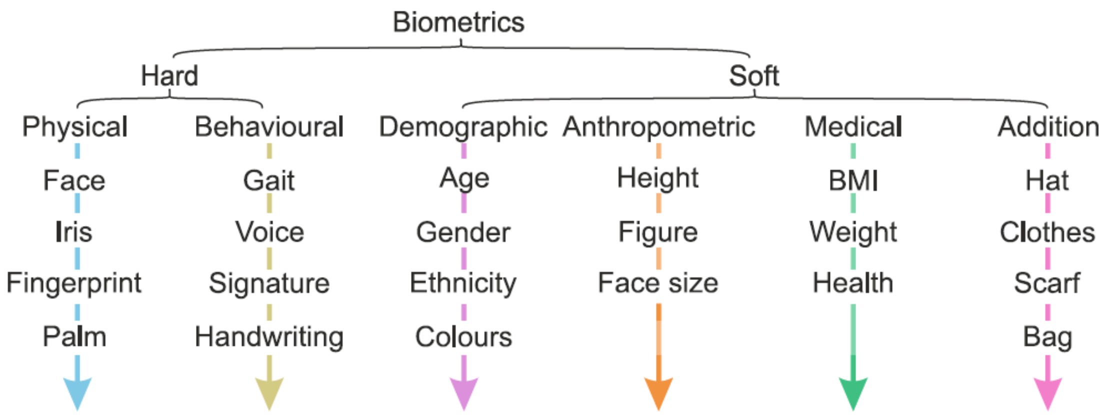


FIGURE 2.5: Diverse types of biometrics.

For soft biometric identification, the information must be distinct, memorable, and easy to describe. These requirements facilitate database searches during identification and enable the integration of eyewitness descriptions. Researchers then convert these descriptive traits into labels or measurements for identification purposes, as detailed by (Nixon et al. (2015), Reid et al. (2013), Arigbabu et al. (2015)).

Work in this field has focused on identifying traits that offer clear, detailed information about individuals. Figure 2.6 provides an overview of soft biometric techniques and their application to diverse parts of the human anatomy. Although most facial features have been extensively studied for biometric recognition, attractiveness has not been a primary focus (Guo et al. (2019), Almudhahka et al. (2017)).

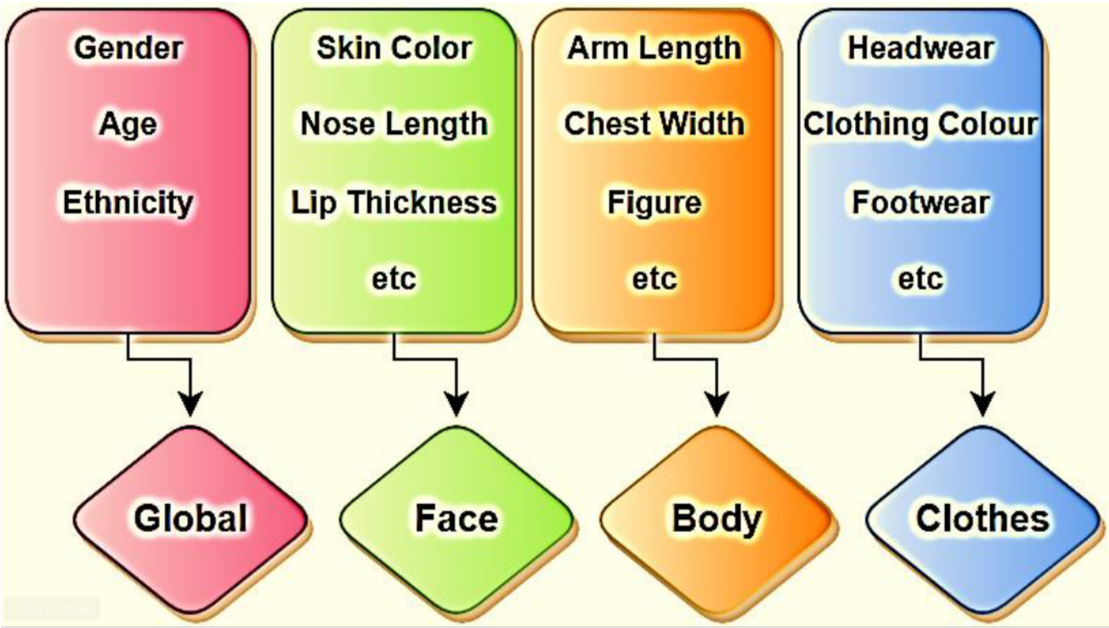


FIGURE 2.6: Overview of soft biometrics modalities and human body traits.

Nevertheless, evidence indicates that attractiveness strongly influences how memorable a face is (Cross et al. (1971), Malloy et al. (2021)). While it can be challenging to measure, studying facial attractiveness may provide valuable insights into how people perceive and remember faces.

This section presents an overview of how biometrics has evolved, focussing on observable physical traits and behaviours. Although interest in facial soft biometrics is growing, exploring facial attractiveness as a biometric attribute remains relatively new. A deeper examination of human facial attractiveness could enhance soft biometrics and improve the overall performance of biometric systems.

2.4 Facial Attributes of Attractiveness

Research indicates that people worldwide often assess attractiveness primarily by examining facial appearance, relying on consistent criteria. According to Kumar et al. (2009), most psychological and biological studies aim to pinpoint the features that define facial beauty. These visual characteristics, referred to as "attributes" or "features," help explain why specific faces appear more attractive. The human face contains many attributes, each differing in uniqueness, prominence, and salience.

Facial attractiveness is influenced by multiple traits, including skin colour, skin texture, facial structure, and overall form (Chen and Zhang (2010)). Specific features—such as small noses, high foreheads, prominent cheekbones, and arched eyebrows—have been strongly linked to attractiveness (Cunningham (1986)). Meanwhile, the eyes and nose are often highlighted as particularly important, backed by plastic surgery findings. Other elements, such as chin shape and eyebrow positioning, have also been studied for their role in perceived attractiveness (Liu, Fan, Samal and Guo (2016)). Taken together, these investigations offer valuable insights into the facial factors that shape perceptions of attractiveness.

Many studies have examined particular traits that affect the degree of attractiveness of a face. Symmetry is central, as faces with matching left and right sides are generally considered more attractive (Schmid et al. (2008)). Another key element is "averageness," where faces formed by merging multiple individuals tend to be viewed as more attractive than unique faces, suggesting that standard features become particularly attractive when combined (Damon et al. (2017)). Sexual dimorphism—reflecting physical differences between males and females—also significantly affects perceptions of facial attractiveness. Research consistently shows that emphasised masculine or feminine characteristics, such as a strong jawline in men or fuller lips in women, are often regarded as attractive (Liu, Fan, Samal and Guo (2016), Rhodes (2006), Russell (2003)). Symmetry, averageness and sexual dimorphism emerge as core attributes that resonate with observers across diverse cultures and genders (Jones and Jaeger (2019)).

Concurrently, attractiveness remains a complex concept shaped by cultural standards and personal experience, which makes establishing universal rules challenging (Chen and Zhang (2010)). Nevertheless, a range of scientific studies across multiple populations indicates that attractiveness is neither entirely subjective nor impossible to quantify. These findings provide a broader global framework for understanding facial attractiveness, suggesting it is not purely subjective or undefinable. Research in this area has made meaningful strides in identifying factors influencing how individuals perceive attractiveness, establishing overarching principles widely recognised as attractive across different cultures and societies (Liu, Fan, Samal and Guo (2016)).

2.5 Conclusion

This chapter has demonstrated how both cultural influences and universal biological factors shape facial attractiveness. Attributes such as symmetry, averageness, and sexual dimorphism consistently emerge as key elements that enhance perceptions of attractiveness across diverse settings. Moreover, this chapter has shown that facial attractiveness can affect social interactions, career opportunities, and even legal outcomes, underscoring its importance as more than just a visual preference. Identifying its potential as a soft biometric feature highlights emerging possibilities and ethical considerations for integrating facial attractiveness into advanced recognition systems.

Chapter 3

Generating Attractiveness Rankings

This chapter introduces an approach to ranking facial attractiveness through comparative soft biometrics. The primary objective is to generate meaningful attractiveness rankings by utilising crowdsourced data and a rating system. The process involves the collection and analysis of comparative labels. By the conclusion of this chapter, it will be demonstrated how facial attractiveness rankings can serve as a soft biometric tool to enhance recognition accuracy in automated systems.

3.1 Comparative Soft Biometrics

Traditionally, individuals have been labelled using categorical soft biometrics, which rely on absolute terms such as ”tall” or ”short” (Reid et al. (2013), Kovashka et al. (2016)). However, these labels often fail to capture the nuances required for comparison. In contrast, comparative soft biometrics emphasise relative relationships. Instead of categorising someone as ”tall”, a statement such as ”Person A is taller than Person B but shorter than Person C” can be made (Parikh and Grauman (2011)). By capturing these subtle differences, comparative approaches have been shown to improve identification and retrieval while reducing bias (Almudhahka et al. (2017), Altwaijry and Belongie (2013)).

As illustrated in Figure 3.1, which presents three individuals (a), (b) and (c), the use of categorical labels—such as ”smiling” or ”not smiling”—may fail to capture important nuances. Instead, it can be observed that (b) smiles more than (c) but less than (a) (Parikh and Grauman (2011)), providing a more precise assessment. Expanding on this concept, the work of Mark Nixon on soft biometrics (Nixon et al. (2015)) has demonstrated that comparative descriptors enhance clarity and accuracy by capturing distinctions that categorical labels may overlook. Additionally, annotators have been

(a) Smiling (b) ? (c) Not smiling

FIGURE 3.1: Expression differences among three individuals. Person (a) is smiling, person (c) is not smiling, and person (b) has an unknown expression. Comparative attributes show that (b) smiles more than (c) but less than (a) (Parikh and Grauman (2011)).

found to perceive relative comparisons as more intuitive than absolute categorisations (Reid et al. (2013)).

In the context of eyewitness identification, the use of comparative techniques has been found to be beneficial. While traditional systems classify individuals into fixed categories such as ”short” or ”tall,” comparative soft biometrics establish relationships, for example, ”taller than A but shorter than C” (Kovashka et al. (2016)). This method has also been demonstrated to be effective in describing facial expressions –consider an instance where an individual is labelled as ”more expressive than (c) but less expressive than (a)”– thereby reducing ambiguity and bias (Reid et al. (2013)).

3.1.1 Comparative Attractiveness

Although earlier research has examined relative facial attributes (Reid et al. (2013), Almudhahka et al. (2017)), body traits (Martinho-Corbishley et al. (2018), Samangooei et al. (2008)), and clothing features (Jaha and Nixon (2016)), facial attractiveness as a comparative soft biometric has not been thoroughly investigated. Attractiveness is influenced by personal and cultural perceptions, rendering simple binary classifications inadequate. In contrast, comparative attributes mitigate subjectivity by ranking individuals relative to one another (Reid et al. (2013)). For instance, as illustrated in Figure 3.2), if person (b) cannot be decisively labelled as ”attractive,” a comparative descriptor—”less attractive than (a) but more attractive than (c)”—provides a more structured and reliable assessment.

By avoiding rigid, subjective labels, this comparative approach aligns with the findings of (Nixon et al. (2015), Almudhahka et al. (2017)), demonstrating how comparative soft biometrics enhance recognition accuracy. Consequently, these techniques establish a strong foundation for utilising attractiveness as a soft biometric. This section

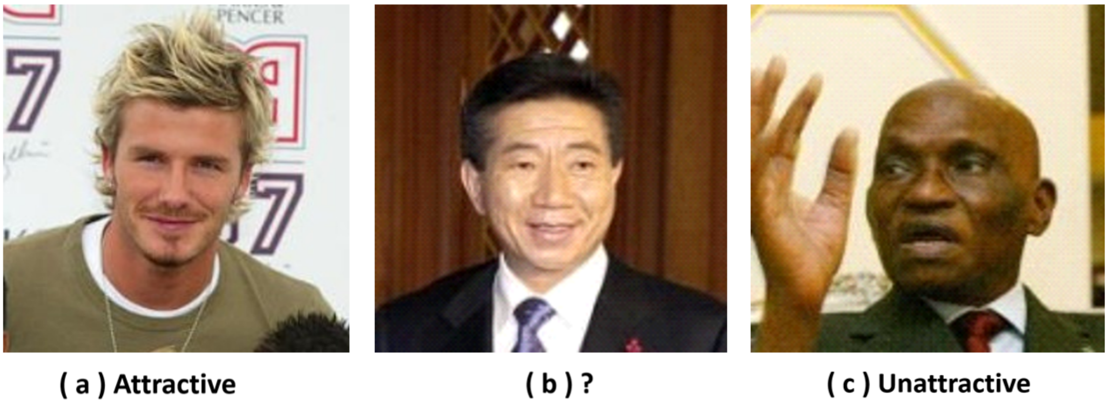


FIGURE 3.2: Attractiveness comparisons among three individuals. Person (a) is considered attractive, and person (c) is considered less attractive. The clearest description for (b) is that (b) is more attractive than (c) but less attractive than (a).

emphasises how comparative attributes form a robust basis for soft biometrics, particularly in the context of attractiveness.

3.2 Comparative Labels

Comparative labels serve as the foundation of this research, allowing individuals to be ranked based on attributes such as facial attractiveness. The following subsections outline the use of the Likert scale, the process of collecting crowdsourced data, and the advantages of compressing the Likert scale to enhance recognition accuracy in facial recognition tasks.

3.2.1 Design and Implementation

The backbone of this study is formed by comparative labels, which are used to rank individuals based on specific traits. A three-point Likert scale (‘more’, ‘same’ and ‘less’) has been employed due to its simplicity and reliability (Reid et al. (2013)). This compressed scale has been shown to enhance consistency and reduce ambiguity compared to traditional five-point scales. To address uncertain cases and minimise noise, an option for ‘Cannot Determine’ was included, ensuring higher data quality. As illustrated in Figure 3.3, the combination of this scale with an increased number of comparisons has resulted in improved recognition accuracy. It has been observed that as the number of individuals compared within a larger pool increases, recognition accuracy improves. For example, when 15 individuals are compared to a larger dataset, better results are achieved than when only five individuals are compared. Therefore, maximising the number of comparisons and utilising a three-point Likert scale is essential for obtaining optimal results (Almudhahka et al. (2016)).

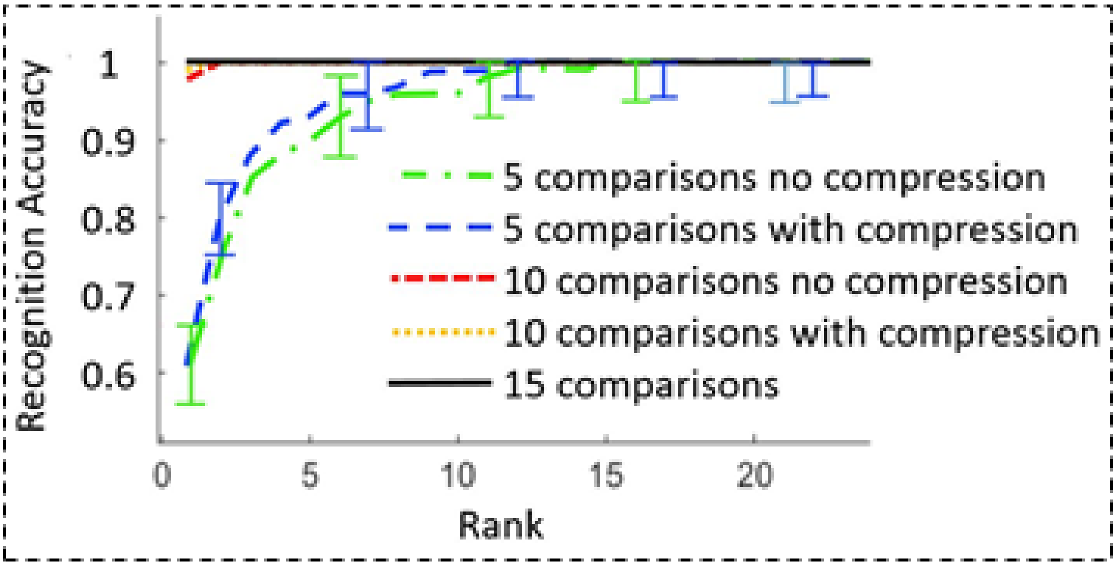


FIGURE 3.3: Demonstration of improved recognition accuracy using a compressed Likert scale compared to a non-compressed scale, particularly as the number of comparisons increases. The three-point Likert scale, combined with increasing comparisons from 5 to 15, significantly enhances annotation consistency and recognition performance (Almudhahka et al. (2016)).

3.2.2 Likert Scale Compression in Comparative Soft Biometrics

The pioneering research of Reid and Nixon on soft biometrics for facial recognition (Reid et al. (2013)) introduced a method for classifying facial features through comparative approaches, thereby reducing bias. Subsequent studies (Almudhahka et al. (2017)) have verified that such comparative classification enhances recognition accuracy.

This improvement is particularly evident when a compressed three-point Likert scale is employed for comparisons, such as in the assessment of attractiveness, rather than a more detailed five-point scale. The three-point scale categorises responses as 1 (more attractive), 0 (same) and -1 (less attractive). In contrast, the five-point scale provides greater granularity, with categories such as 2 (much more attractive), 1 (more attractive), 0 (same), -1 (less attractive) and -2 (much less attractive). The simplified three-point scale reduces complexity, facilitating swifter and more reliable comparisons.

As illustrated in Figure 3.3, the combination of this streamlined scale with an increased number of comparisons (e.g., comparing an individual with 10 or 15 others rather than just five) has been shown to significantly improve accuracy. This method highlights the practical advantages of simplified comparative scales in human identification systems. By ensuring uniform judgements, this technique reinforces the effectiveness of comparative soft biometrics in evaluating facial attractiveness within automated systems.

3.2.3 Crowdsourcing Annotations

Annotations were collected using the Appen platform (formerly known as Figure Eight) (*Confidence to Deploy AI with World-Class Training Data* (2025)), providing access to a diverse range of annotators from various cultural and demographic backgrounds (Martinho-Corbishley et al. (2015)). Ensuring response reliability was a critical aspect of the data collection process. Predefined questions with known correct answers were used to filter out unreliable annotators (Welinder and Perona (2010)). Annotators were required to maintain a minimum accuracy rate of 70% to continue participating in the tasks. Detailed instructions and examples were provided to ensure consistency and reduce ambiguity (Kovashka et al. (2016)).

3.3 Crowdsourcing of Comparative Attractiveness Traits

In this section, the crowdsourcing approach used to collect comparative judgements on facial attributes is described. A large and culturally diverse group of annotators was engaged, minimising social annotation bias and ensuring the generation of robust ground truth data for subjective tasks such as facial attractiveness evaluation (Martinho-Corbishley et al. (2015)). Modern platforms, such as Amazon Mechanical Turk and Appen, have facilitated the rapid recruitment of hundreds of annotators, producing results that would otherwise require months of manual work (Kovashka et al. (2016), Welinder and Perona (2010)).

Appen (formerly CrowdFlower and Figure Eight)[1] was selected due to its global contributor network, quality assurance tools, and reliability checks—particularly the use of 'gold-standard' questions to detect dishonest responses (Kovashka et al. (2016)). By adhering to best practices from (Welinder and Perona (2010)), these questions were designed to minimise errors in facial recognition tasks by requiring annotators to maintain an accuracy rate of at least 70%. Additional guidelines provided by (O'Toole and Phillips (2017)) further reduced errors by providing explicit instructions and systematic accuracy checks. Various strategies for ensuring the collection of valid annotations through crowdsourcing are illustrated in Figure 3.4

Overall, Appen's platform was utilised as an efficient and reliable method for obtaining comparative attractiveness labels. A diverse and consistent dataset was ensured through its extensive contributor base, strong quality control measures, and adherence to best practices.

[1] `https://www.appen.com/`

FIGURE 3.4: Strategies for collecting corresponding annotations (Kovashka et al. (2016)).

3.3.1 Comparative Attractiveness Traits

In the previous chapter, facial attractiveness was emphasised based on studies that identified key features such as symmetry, cheekbone prominence, and forehead proportions (Cunningham (1986)). For these attributes to function effectively as soft biometric identifiers, they must be both memorable and easy to describe. Consequently, six additional traits – 'symmetry', facial hair, 'skin smoothness', 'proportions', 'jaw size' and 'attractiveness' – were incorporated alongside the 16 attributes from the earlier study (Almudhahka et al. (2016)). Contributions by (Nixon et al. (2015)) have further demonstrated that the integration of established and newly identified features can enhance biometric accuracy.

To evaluate these traits, comparative labels were applied using a three-point bipolar scale (1 = 'more', 0 = 'same', -1 = 'less'), with a 'cannot see' option for ambiguous cases. This approach was designed to minimise inconsistencies in data and maintain annotator agreement, as illustrated in Table 3.1. By employing a comparative soft biometrics approach to address the subjective nature of attractiveness, these methods ensure simplicity and reliability.

TABLE 3.1: Facial attributes used in this study, including traits relevant to attractiveness.

Soft Traits	1	0	-1	-2
Age	More old	Same	More young	Cannot see
Attractiveness	Less Attractive	Same	More attractive	Cannot see
Cheek shape	More flat	Same	More prominent	Cannot see
Chin length	More long	Same	More short	Cannot see
Eyebrow length	More long	Same	More short	Cannot see
Eyebrow thickness	More thick	Same	More thin	Cannot see
Eyes size	More large	Same	More small	Cannot see
Face length	More long	Same	More short	Cannot see
Face width	More wide	Same	More narrow	Cannot see
Facial hair	Less facial hair	Same	More facial hair	Cannot see
Forehead hair	Less forehead hair	Same	More forehead hair	Cannot see
Gender	More masculine	Same	More feminine	Cannot see
Jaw size	More narrow	Same	More wide	Cannot see
Lip thickness	More thick	Same	More thin	Cannot see
Nose length	More long	Same	More short	Cannot see
Nose width	More wide	Same	More narrow	Cannot see
Nose-mouth distance	More Short	Same	More Long	Cannot see
Proportions	More average	Same	Less average	Cannot see
Figure (Shape)	More fat	Same	More thin	Cannot see
Skin colour	More dark	Same	More light	Cannot see
Skin smoothness	Less smooth	Same	More smooth	Cannot see
Symmetry	Less symmetrical	Same	More symmetrical	Cannot see

3.3.2 Facial Attractiveness Dataset

Previous research on comparative soft biometrics has been conducted using various datasets. The Multi-Biometric Tunnel (MBT) dataset has been utilised for this purpose (Tome et al. (2014)). Additionally, the SCUT-FBP5500 dataset has been frequently employed in studies on facial attractiveness (Liang et al. (2018)), see Figure 3.5.

While many previous studies have relied on constrained datasets, the Labelled Faces in the Wild (LFW) dataset (Almudhahka et al. (2017)) has been selected for this work due to its representation of a broader range of individuals encountered in daily life. The LFW dataset introduces real-world challenges in facial recognition (Huang et al. (2008)), including variations in lighting conditions, facial expressions, poses and lower-resolution images. Although it is smaller than some alternative datasets, it remains widely used for studying facial recognition in unconstrained environments, such as surveillance. The dataset contains 13,233 facial images representing 5,749 individuals (Almudhahka et al. (2017)), as shown in Figure 3.6. Another widely recognised dataset is CelebA (Liu et al. (2018)), which contains over 200,000 celebrity images annotated for 40 attributes, including attractiveness. CelebA has been extensively employed in facial

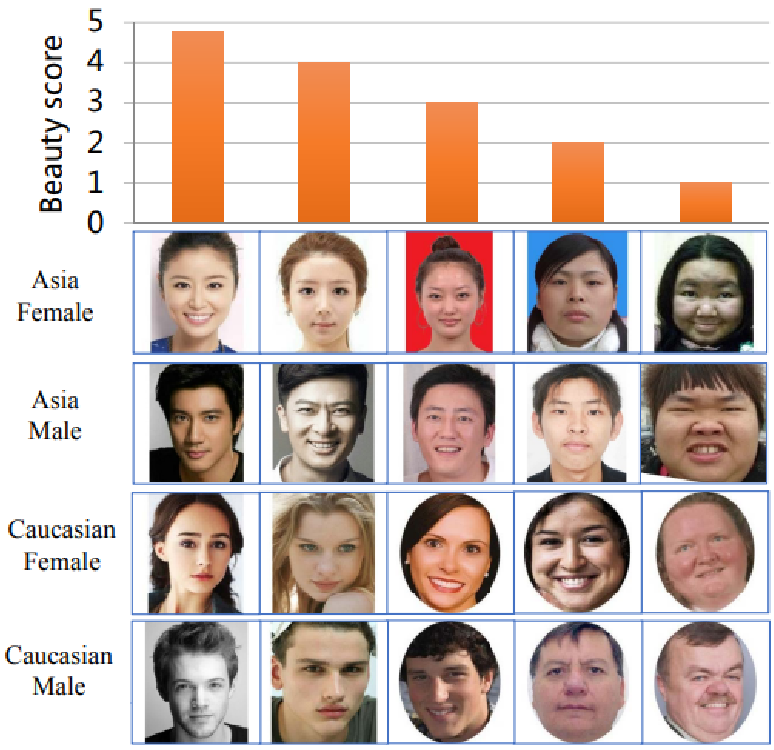


FIGURE 3.5: Images from the SCUT-FBP5500 benchmark dataset with varying facial features and attractiveness ratings (Liang et al. (2018)).

recognition, attribute prediction, and attractiveness studies, contributing to a deeper understanding of how attractiveness is perceived in celebrity faces.

FIGURE 3.6: Samples from the LFW dataset, showing unconstrained images under different conditions.

The LFW dataset was selected for this study to assess attractiveness under more realistic conditions, thereby reflecting typical human observations. Promising results using comparative attributes on LFW were also reported by (Almudhahka et al. (2017)). Consequently, human facial attractiveness is explored in this research within the LFW

View 1 subset, starting with 100 subjects, each contributing four samples from the LFW-MS4 subset [2]. The LFW-MS4 subset includes a minimum of four samples per subject, comprising a total of 430 subjects, and is part of the LFW View 1 dataset (Almudhahka et al. (2017)). These initial samples will be expanded to encompass the entire dataset to obtain deeper insights into perceived attractiveness in unconstrained settings.

3.4 Experiment Design to Capture Attractiveness Attributes

The design and methodology used to obtain comparative judgements of facial attractiveness through crowdsourcing are described. A structured question-and-answer procedure was implemented using the Appen platform and the LFW dataset. These approaches support the primary objective of this study, which is to introduce attractiveness as a soft biometric trait.

3.4.1 Question and Answer Design

The Appen platform was configured to collect comparative attractiveness labels using the LFW dataset. Each annotator was presented with two faces—one on the left and the other on the right—and was asked to provide a label, as shown in Table 3.1. A sample of the comparison interface is shown in Figure 3.7. This format is based on standard psychometric methods (Thurstone (1927), Martinho-Corbishley et al. (2018)).

The interface was simplified, and response time was expedited through the use of vertically aligned radio buttons. Each contributor was permitted to answer up to four pages of questions, with each page containing 44 annotations (22 comparisons per page). Test questions were incorporated to assess contributor accuracy and ensure the reliability of responses. Priority was given to Level-1 contributors who achieved a minimum accuracy of 70%, thereby enhancing both speed and data quality.

3.4.2 Response Analysis

A total of 36,170 judgements (35,927 trusted and 243 untrusted) were collected for 1,720 subject images at a cost of $1,140.76. The distribution of labels for attractiveness and related traits across these subjects is illustrated in Figure 3.8. Notably, far fewer 'same' labels were assigned by annotators compared to 'more' or 'less,' suggesting that subtle differences in attractiveness were perceptible to them.

The low frequency of 'same' labels suggests that subtle differences in attractiveness can be detected by participants, reinforcing its value as a dynamic and robust

[2] http://github.com/almudhahka/lfw-ms4

FIGURE 3.7: Sample question on the Appen (formerly Figure Eight) platform.

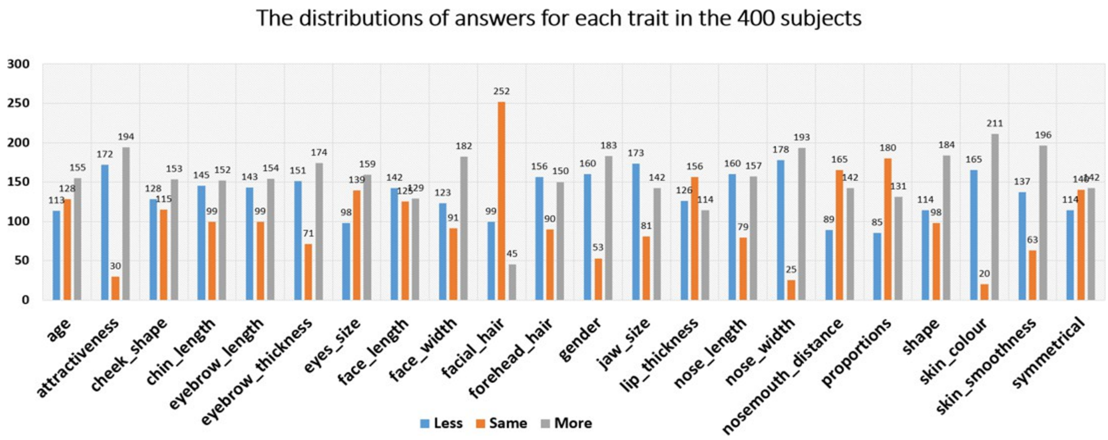


FIGURE 3.8: Distribution of comparative labels for attractiveness and related traits among 1,720 subjects.

soft biometric. A lower occurrence of 'same' responses also minimises ambiguity, making participant preferences more apparent. This uncommon occurrence highlights the difficulty of determining whether two individuals are equally attractive and demonstrates how comparative assessments could be applied in biometric applications.

3.4.3 Discussion

Confidence scores generated by Appen were used to measure annotator agreement, ranging from 0.62 to 0.90 for 'Average Comparative Attractiveness'. As shown in Table 3.2, higher confidence scores (e.g. 0.87 or 0.90) were often observed for the 'more attractive' and 'less attractive' labels, indicating strong consensus. In contrast, lower scores (e.g. 0.65 or 0.76) were commonly assigned to 'same' judgments, suggesting greater difficulty in achieving agreement on equal attractiveness.

					
Average Comparative Attractiveness	Less Attractive	Same	Same	Same	More Attractive
Confidence Score	0.75	0.65	0.79	0.62	0.87
					
Average Comparative Attractiveness	Less Attractive	Less Attractive	More Attractive	Less Attractive	Same
Confidence Score	0.79	0.73	0.90	0.82	0.76

TABLE 3.2: Confidence scores for "Average Comparative Attractiveness" across different subjects.

These scores were calculated by summing the trust scores of contributors who selected a given response and dividing by the total trust scores for all responses. For instance, a confidence score of 0.6166 was assigned to the 'more attractive' label.

$$\text{confidence score} = \frac{\text{sum of trust scores for the response}}{\text{sum of all trust scores}}.$$

Overall, greater agreement was demonstrated by contributors when comparing faces as 'more' or 'less' attractive, whereas identifying them as equally attractive proved more challenging. This outcome underscores the significance of comparative judgments in soft biometric tasks and highlights how confidence scores contribute to data quality when evaluating subjective traits such as attractiveness.

3.5 Relative Attractiveness Ranking

Facial attractiveness can be assessed through attribute-based comparisons, in which individuals are ranked based on specific traits such as 'age', 'skin colour' or 'gender'. Rather than relying on a single overall score, rankings for individual traits provide more detailed comparisons. By combining these rankings into a single profile, a unique feature set is generated for each subject, which can be utilised in biometric systems for identification or differentiation. This structured approach is particularly valuable when precise measurements are unavailable (Reid et al. (2013), Almudhahka et al. (2017), Parikh and Grauman (2011), Nixon and Aguado (2019)).

3.5.1 Ranking Methods

Two main methods are commonly used to rank facial attributes: the ELO Rating System and RankSVM. The ELO Rating System, originally developed for chess players (Jung and Jung (2025), Elo (1978)), updates each individual's score after pairwise comparisons. In the context of facial "attractiveness," each person is initially assigned a default score (e.g., 1500). When two individuals, A and B, are compared, and A is judged to be 'more attractive,' A's score increases while B's decreases. Through multiple comparisons, these scores adjust automatically without requiring pre-labelled data, making the ELO system well-suited for subjective tasks such as attractiveness assessment (Dangauthier et al. (2007)).

In contrast, RankSVM is a supervised learning algorithm that relies on labelled datasets (Parikh and Grauman (2011)). For example, a dataset may categorise individuals as 'more,' 'same', or 'less' attractive while including attributes such as 'age' or 'gender.' The RankSVM model is trained to rank new individuals accordingly but must be retrained whenever new data is introduced, reducing its flexibility. Nevertheless, RankSVM performs exceptionally well when sufficient labelled data is available and high-precision rankings are required (Burges et al. (2005)).

The ELO system is ideal for flexible, iterative attractiveness measurements that do not require preexisting labels. RankSVM, on the other hand, excels when a well-organised, labelled dataset is available. Although RankSVM provides accurate rankings, additional effort is required to manage data updates.

3.5.2 Multi-Attribute Rank Profiles

Instead of being assigned a single attribute score, each facial attribute (such as 'age', 'skin colour', or 'gender') is given a separate rank, which is then combined into a single rank profile for each individual. Through this process, a unique profile is created, capturing an individual's relative standing across multiple traits. Consider three individuals (A, B, and C), who are ranked across three attributes:

- **Age:** A is ranked #1, B is #2, and C is #3.
- **Skin Colour:** B is ranked #1, C is #2, and A is #3.
- **Gender:** C is ranked #1, A is #2, and B is #3.

These rankings are then merged into a rank profile:

- **Person A**: (1, 3, 2)
- **Person B**: (2, 1, 3)
- **Person C**: (3, 2, 1)

Such profiles result in distinct feature sets that encapsulate multiple dimensions of each face. These profiles can be utilised by a biometric system to identify or differentiate individuals without dependence on a single score (Alnamnakani et al. (2019)). Through the systematic use of comparative data, this technique enhances identification accuracy by considering multiple attributes simultaneously.

3.5.3 Elo Rating System Implementation

The ELO rating system, originally conceived for chess (FIDE (n.d.), Elo (1978)), has been adapted here to assess facial attributes such as attractiveness (see Figure 3.9). Ratings are dynamically updated through pairwise comparisons, in which individuals are judged as 'more', 'less', or 'same' in attractiveness (FIDE (n.d.), Nixon and Aguado (2019)). Unlike methods that require fully labelled datasets, ELO relies solely on these pairwise outcomes, making it well-suited for subjective assessments.

For each pair of subjects, A and B, a record is made of who is judged more or less attractive. The following equations then update their ratings, assuming both start at 1500:

$$R'_A = R_A + K(S_A - E_A), \tag{3.1}$$

$$R'_B = R_B + K(S_B - E_B). \tag{3.2}$$

Where:

FIGURE 3.9: Arpad Elo, the creator of the ELO rating system (*Arpad Elo* (n.d.)).

- R_A, R_B are the current ratings for subjects A and B, and R'_A, R'_B are the updated ratings after the comparison.
- S_A, S_B are the observed outcomes for A and B:

  $+1:$ "More"

  $0.5:$ "Same"

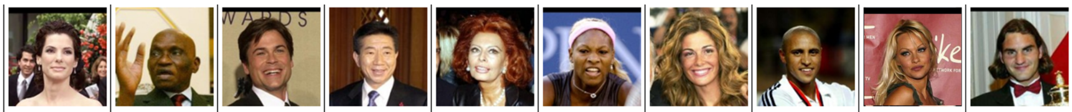
  $0:$ "Less"

- K is a sensitivity factor regulating how much each comparison shifts a subject's rating.

The change in rating is also dependent on U, the scaling factor that modulates the impact of rating differences on probabilities. An example of initial and updated ratings for 10 subjects is presented in Table 3.3.

An expected outcome (E) is also calculated for each subject based on current ratings:

$$E_A = \frac{1}{1 + 10^{\frac{(R_B - R_A)}{U}}} \tag{3.3}$$



S_A	1	-1	1	-1	1	1	-1	1	-1	1
R_A	1	2	3	4	5	6	7	8	9	10
$\overline{R}_A$	6	1	7	2	8	3	9	4	10	5

TABLE 3.3: Example of S_A, R_A, and updated ratings $\overline{R}_A$ and $\overline{R}_B$ for ten subjects.

$$E_B = \frac{1}{1 + 10^{\frac{(R_A - R_B)}{U}}} \tag{3.4}$$

Role of K and U.

- K **(Sensitivity Factor)**: Larger values of K result in more significant changes per comparison, whereas smaller values lead to more gradual adjustments.
- U **(Scaling Factor)**: A large U reduces the system's sensitivity to rating gaps, causing slower adjustments, while a small U increases responsiveness but may reduce stability.

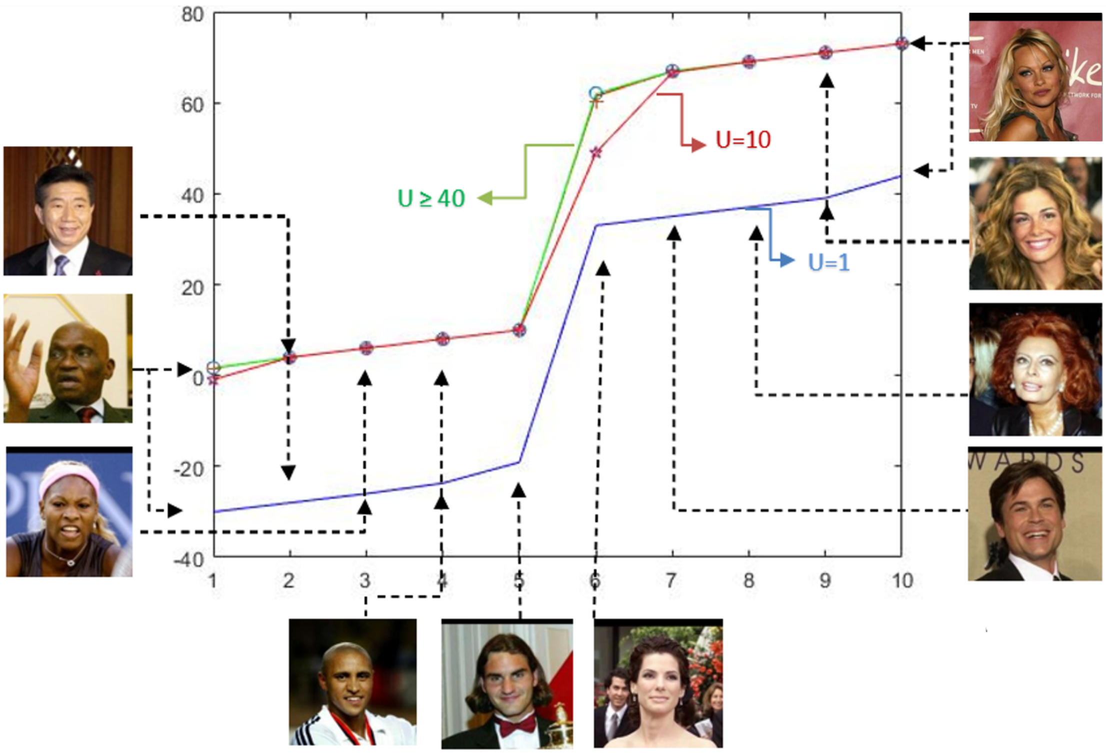


FIGURE 3.10: Effect of different U values on ELO updates.

Figure 3.10 illustrates how varying U influences the speed and stability of updates:

- $U \geq 40$: Smooth updates with minimal fluctuation.
- $U = 10$: Balanced stability and responsiveness.
- $U = 1$: Rapid changes but possible oscillations.

In practice, varying U affects convergence speed rather than the eventual rank order; in our experiments, the final ranking remained unchanged. Equations (3.5) and (3.6) rewrite E_A and E_B via Q_A and Q_B:

$$E_A = \frac{1}{Q_A}, \qquad E_B = \frac{1}{Q_B}, \tag{3.5}$$

$$Q_A = 1 + 10^{\frac{R_B - R_A}{U}}, \quad Q_B = 1 + 10^{\frac{R_A - R_B}{U}}. \tag{3.6}$$

In Elo-based attractiveness evaluations, the expected outcomes (E_A, E_B) are governed by the rating difference $(R_A - R_B)$. When two individuals possess nearly identical ratings, each is given approximately a 50% chance of being deemed 'more attractive'. However, a large gap strongly favours the higher-rated individual, with E approaching 1. The parameter U determines the extent to which rating gaps influence these probabilities: a smaller U (e.g. 10) results in quicker updates, while a larger U (e.g. 40) produces smoother but slower changes. In this study of 430 subjects, $U = 10$ is selected to achieve a balance between speed and stability. Meanwhile, K can be assigned a moderate value (e.g. 24) to ensure that ratings adjust meaningfully without dramatic fluctuations. Prior research (Alnamnakani et al. (2019), Alnamnakani et al. (2024)) relied exclusively on ordinal ranks, whereas ELO scores are updated dynamically after each comparison, providing a finer-grained measure of differences in attractiveness.

The principles of ELO and the relationship between rating differences and expected outcomes are explained in the sources cited (Almudhahka et al. (2017), Elo (1978), FIDE (n.d.), Federation (n.d.)). Although values such as $U = 10$ or $U = 40$ are less commonly used – since classical chess ELO often employs 400 in the exponent – these references provide details on how the standard formula can be adapted for subjective tasks, such as evaluating attractiveness. An example of final Elo-based attractiveness rankings from LFW is presented in Figure 3.11. The most attractive faces are shown in the top row, while the least attractive are displayed in the bottom row. Although U influences the speed at which these final rankings emerge, it does not affect their relative order.

The ELO system offers a flexible and data-efficient approach to ranking facial attractiveness through pairwise comparisons. Scores are updated continuously, allowing subtle differences and subjective perceptions to be captured without the need

FIGURE 3.11: Final Elo-based attractiveness rankings, the top row represents the most attractive.

for extensive pre-labelled data. Although parameters such as K and U affect the rate of change, the final ranking remains stable, making ELO a suitable method for real-world attractiveness evaluations.

3.6 Conclusion

This chapter presented a comparative soft biometrics approach for evaluating facial attractiveness. Data were collected through crowdsourcing, and attractiveness rankings were generated using pairwise comparisons. Methods such as the ELO Rating System were reviewed, illustrating how iterative updates can effectively capture subtle differences without requiring extensive pre-labelled data. Bias was reduced, and more nuanced insights into subjective traits like attractiveness were provided by focusing on relative assessments. The results indicate that when analysed comparatively, facial attractiveness shows potential as a soft biometric feature for real-world identification tasks.

Chapter 4

Facial Attractiveness Soft Biometrics Analysis

A more comprehensive investigation is required to understand how facial 'attractiveness' interacts with other soft biometric attributes and contributes to identification and recognition. In this chapter, various analytical methods are highlighted, with particular emphasis on correlation analysis and an evaluation of the discriminative power of 'attractiveness'.

4.1 Correlation Analysis

A correlation analysis was conducted to examine potential interdependencies among the 22 facial attributes (see Table 3.1), including 'attractiveness'. This approach provides insight into how different attributes relate to one another and determines whether specific traits offer distinct information for identification or classification. Additionally, correlation matrices show how one feature's strength may depend on another.

To assess whether 'attractiveness' correlates with the other 21 facial soft biometric attributes, prior research was referenced, which considered the discriminative power of 'attractiveness' alongside other attributes (Alnamnakani et al. (2019)). The Pearson correlation coefficient was employed to systematically measure the linear relationships among different facial traits (Abdullah (1990)).

4.1.1 The Pearson Correlation Coefficient

The Pearson correlation coefficient (r) measures the linear association between two variables. In the context of facial traits, r indicates how strongly two facial attributes co-vary. Its values range from -1 to +1 (Benesty et al. (2009)):

- $r = +1$: a perfect positive linear relationship
- $r = -1$: a perfect negative linear relationship
- $r = 0$: no linear relationship

The r moves farther from zero (positive or negative), the correlation becomes stronger. Equation (4.1), adapted from (Gonzalez-Sosa et al. (2018), Abdullah (1990)), demonstrates the method used to calculate r :

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

Here, X and Y are two facial attributes (e.g., attractiveness and skin smoothness), $\bar{x}$ and $\bar{y}$ are their mean values, σ_{XY} is their covariance, and σ_X, σ_Y their standard deviations. The value of r lies between -1 to $+1$.

Figure 4.1 presents an updated heatmap of the Pearson correlation coefficients among the 22 facial attributes examined in this study. Stronger positive correlations are represented by darker shades of blue, while lighter shades (approaching white) indicate weaker or negative correlations. Since each attribute is perfectly correlated with itself, the diagonal cells are assigned a value of 1.0. Meanwhile, Table 4.1 focuses on the 'attractiveness' attribute, displaying its correlation with each feature. Together, these two figures provide insights into the relationships among the attributes and highlight those that have the most significant impact on 'attractiveness'.

In Table 4.1, 'skin smoothness' is shown to have the highest correlation with 'attractiveness', followed by 'age' and 'symmetry'. Within this dataset, individuals with smoother skin, more youthful features and balanced 'proportions' are generally rated as more attractive. Meanwhile, moderate correlations are observed for 'gender', facial 'proportions' and overall face shape, whereas 'nose length' and 'chin length' exhibit minimal influence on 'attractiveness' ratings.

In Figure 4.2, examples are provided of individuals considered highly attractive, who often possess smoother skin and youthful characteristics, as well as those deemed less attractive, who typically display contrasting attributes. These findings support the notion that 'symmetry', 'gender' and 'proportions' play a critical role in shaping perceptions of 'attractiveness'.

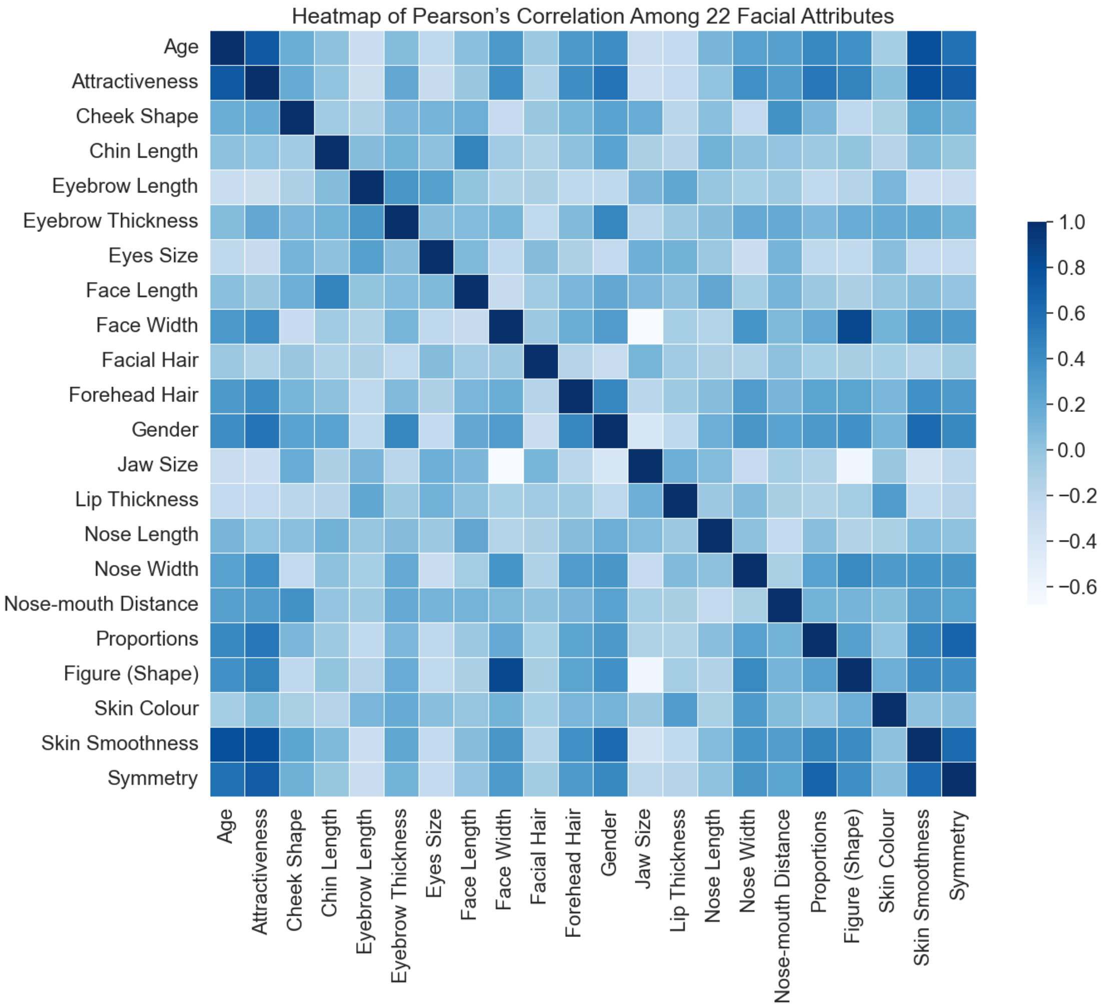


FIGURE 4.1: Heatmap of Pearson's correlation among all 22 facial attributes, including attractiveness.

The findings confirm that 'attractiveness' is most strongly influenced by 'skin smoothness', 'age', 'symmetry' and specific 'gender'-related aspects. By recognising these correlations, practical applications in face recognition and image processing can be better informed, leading to reduced bias and improved model performance.

4.1.2 Comparison with Previous Study

A comparison of the present correlation matrix (Figure 4.1) with that of a previous study (Almudhahka et al. (2016)), Figure 4.3 reveals several similarities. Although 'attractiveness' was not directly measured in (Almudhahka et al. (2016)), youthful and feminine traits were emphasised, reflecting the importance of 'age' and 'gender' observed in the current study. Additionally, both studies identified eyebrow features ('eyebrow length' and 'eyebrow thickness') as significant factors in shaping facial impressions.

TABLE 4.1: Pearson's r values for the correlation between attractiveness and 21 other facial attributes.

Facial Attribute	Correlation (r)
Skin Smoothness	0.79
Age	0.73
Symmetry	0.72
Gender	0.56
Proportions	0.54
Figure (Shape)	0.46
Forehead Hair	0.39
Face Width	0.39
Nose Width	0.38
Jaw Size	0.30
Eyebrow Length	0.29
Nose-mouth Distance	0.29
Eye Size	0.26
Lip Thickness	0.24
Eyebrow Thickness	0.20
Cheek Shape	0.18
Facial Hair	0.14
Skin Colour	0.06
Face Length	0.03
Nose Length	0.01
Chin Length	0.01

FIGURE 4.2: (Top) Faces that are highly attractive are often associated with youthfulness and smooth skin. (Bottom) Faces are rated lower in attractiveness, typically having the opposite traits. These observations align with common views on attractiveness.

However, certain differences have also been noted. For instance, (Almudhahka et al. (2016)) reported a stronger negative relationship between 'nose-mouth distance' and

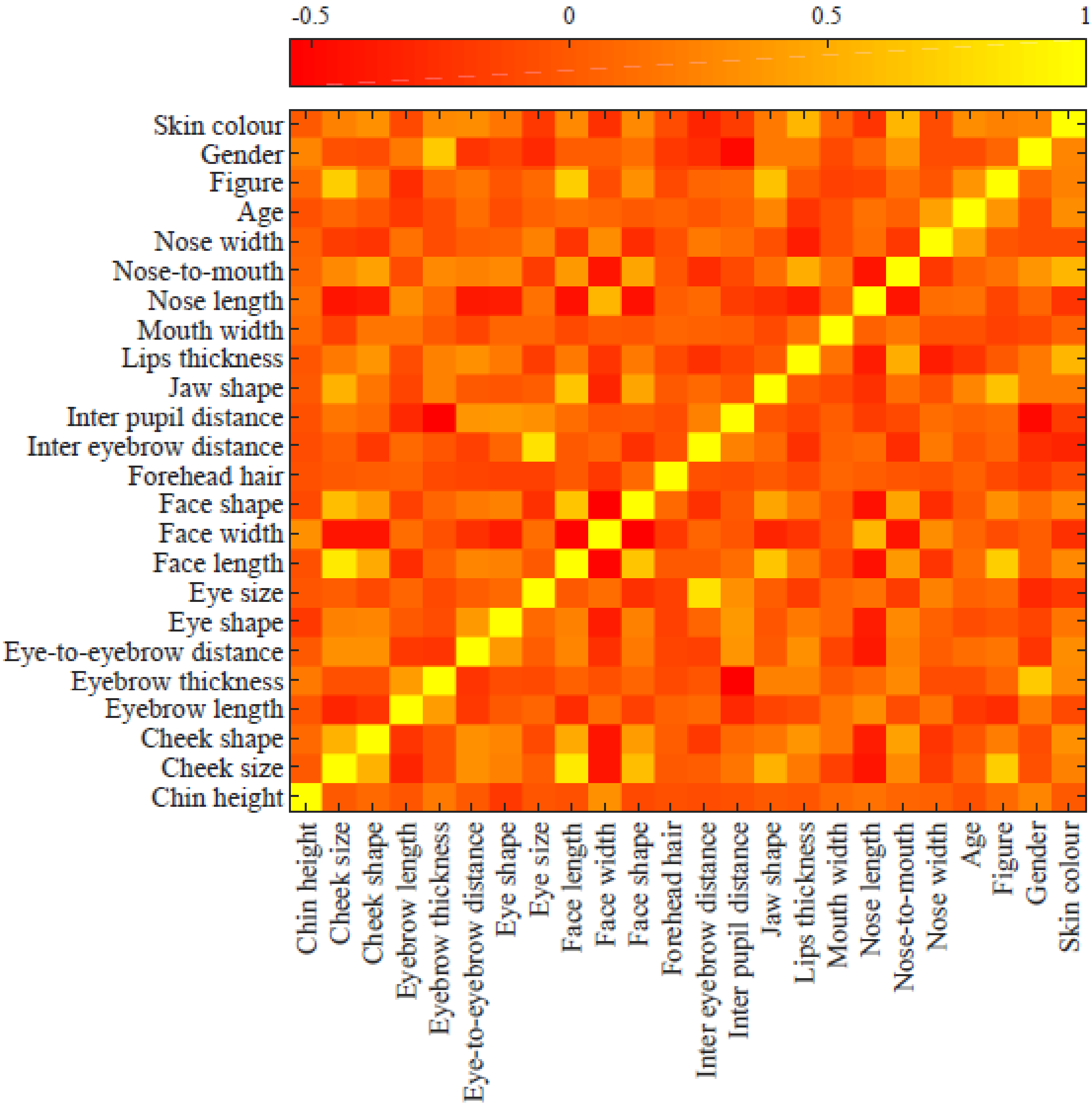


FIGURE 4.3: Previous Pearson's r correlations for selected shared semantic facial features Almudhahka et al. (2016).

'nose length', whereas only a weaker connection is observed in this study. Furthermore, the influence of 'lip thickness' and 'eye size' on perception differs slightly, possibly due to variations in sample size or annotation methods.

Overall, Almudhahka et al. (2016) primarily focused on associations between facial traits, 'gender' and overall facial structure, whereas the present study extends this perspective by demonstrating that many of the same traits—particularly those related to youthfulness and balanced 'proportions'—also' exhibit strong correlations with 'attractiveness'. In both studies, 'chin length' and 'nose length' appear to have minimal influence. Meanwhile, broader attributes such as 'age' and 'gender' display high correlations (Samangooei et al. (2008), Tome et al. (2014)), suggesting that older individuals tend to receive lower 'attractiveness' scores, while those perceived as more feminine typically attain higher ratings.

4.2 Discriminative Power of Attractiveness

In this context, an investigation is conducted to determine whether the ‘attractiveness’ attribute enhances identification performance when combined with other facial features. In the following subsections, Mutual Information (MI) and Sequential Floating Forward Selection (SFFS) are applied to evaluate the extent to which ‘attractiveness’ contributes to the recognition of individuals. The outcomes generated by these two methods are then compared.

4.2.1 Mutual Information for Discriminating Facial Attributes

MI is a fundamental measure in information theory that quantifies the extent to which knowledge of one random variable (e.g., a facial attribute) reduces uncertainty about another (e.g., a person’s identity) (Pohjalainen et al. (2015)). As expressed in Equation (4.2), MI is defined using Shannon entropies:

$$\begin{aligned} \mathrm{MI}(X,Y) &= I(X,Y) \\ &= H(X) - H(X \mid Y) \\ &= H(Y) - H(Y \mid X) \\ &= H(X) + H(Y) - H(X,Y), \end{aligned} \tag{4.2}$$

where $H(X)$ and $H(Y)$ are the entropies of X and Y, respectively, and $H(X \mid Y)$ and $H(Y \mid X)$ denote their conditional entropies. The term $H(X,Y)$ represents the joint entropy of X and Y. Shannon entropy for a random variable X is given by:

$$H(X) = -\sum_{x \in X} p(x) \ln\left[p(x)\right] \quad \text{and} \quad H(Y) = -\sum_{y \in Y} p(y) \ln\left[p(y)\right]. \tag{4.3}$$

The conditional and joint entropies appear in Equations (4.4), (4.5), and (4.6):

$$H(X \mid Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \ln\left[p(x \mid y)\right], \tag{4.4}$$

$$H(Y \mid X) = -\sum_{y \in Y} \sum_{x \in X} p(y,x) \ln\left[p(y \mid x)\right], \tag{4.5}$$

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \ln\left[p(x,y)\right]. \tag{4.6}$$

Here, $p(x \mid y)$ denotes the conditional probability of x given y.

In a soft biometric context, let X represent the ratings (or class labels) of a facial attribute (e.g. 'attractiveness'), and let Y indicate an individual's identity. According to Equation (4.2), a higher value $\mathrm{MI}(X, Y)$ suggests that knowledge of X reduces more uncertainty about Y. Consequently, an attribute with high MI is considered more discriminative, as it significantly decreases ambiguity regarding a person's identity.

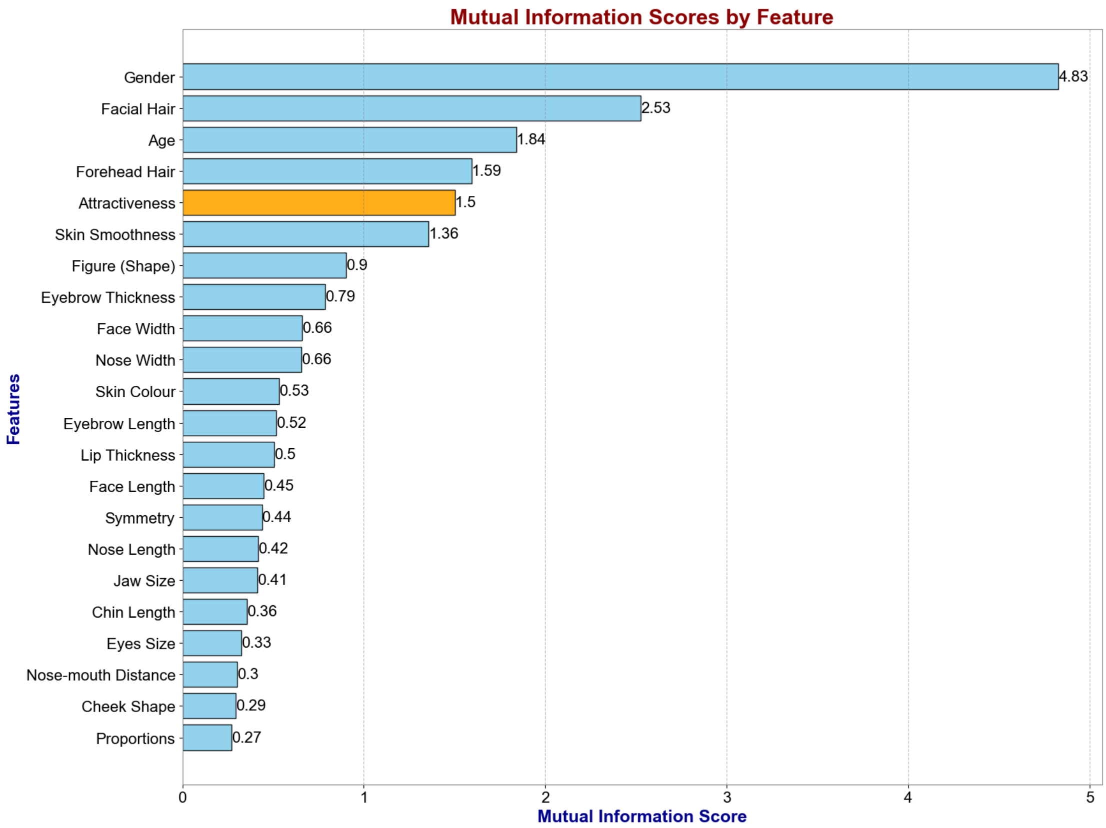


FIGURE 4.4: The discriminative power of "attractiveness" and 21 other facial attributes, measured by their mutual information with identity.

Figure 4.4 presents the MI values for 22 facial attributes (including 'attractiveness'), illustrating their effectiveness in distinguishing individuals (Guo and Nixon (2008)). Notably, 'attractiveness' ranks among the top attributes, with an MI value of 1.50, surpassing many other facial features. This suggests that knowledge of an individual's 'attractiveness' rating provides valuable information for identification, thereby reducing uncertainty about their identity. In contrast, attributes with lower MI values, such as 'proportions' and 'nose-mouth distance,' contribute only marginally to distinguishing individuals.

Similarly, 'age' emerged as the most discriminative attribute overall, while 'gender' also exhibited a notably high MI value. Emphasising attributes like 'attractiveness'—which achieved one of the highest MI scores—can therefore be advantageous in soft biometric frameworks by providing an additional dimension for facial differentiation.

4.2.1.1 Discretising Continuous Features

The discriminative power of soft biometric traits was assessed using Mutual Information (MI). However, it is widely recognised that directly applying MI to continuous features may introduce estimation bias if the data is not properly preprocessed.

To mitigate this issue, a quantile-based discretisation approach was applied using `KBinsDiscretizer`, which divides each continuous feature into 10 ordinal bins containing approximately equal numbers of samples. This transformation results in balanced categorical features that facilitate more accurate MI estimation through probability mass functions. These discretised features were subsequently analysed using `mutual_info_classif`, with `discrete_features=True` specified to ensure the appropriate entropy-based calculation.

This step was critical because MI is based on discrete probability distributions. If features remain in continuous form, those with high variance—such as Elo-based Attractiveness—may be underestimated relative to binary features like Gender. This underestimation is a well-known artefact when applying MI directly to continuous values.

This approach is endorsed by the official *scikit-learn* documentation, which advises:

> "If features are continuous, it is recommended to discretise them before using this function."
> *(Scikit-learn documentation, 2024)* [1]

Following this correction, Figure 4.5 Attractiveness emerged as one of the most informative features, reaffirming its discriminative significance when MI is estimated correctly.

4.2.1.2 Constructing Joint and Marginal Distributions

To better understand the estimation of MI from data, a joint probability distribution $p(x, y)$ is constructed. Table 4.2 presents an illustrative example in which both variables X and Y take on ordinal values ranging from 1 to 10, with each cell $p(x, y)$ representing the joint probability of observing $X = x$ and $Y = y$.

The marginal distributions $p(x)$ and $p(y)$ are computed by summing across rows and columns, respectively. These are then used to calculate the Shannon entropy for each variable:

[1] https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html

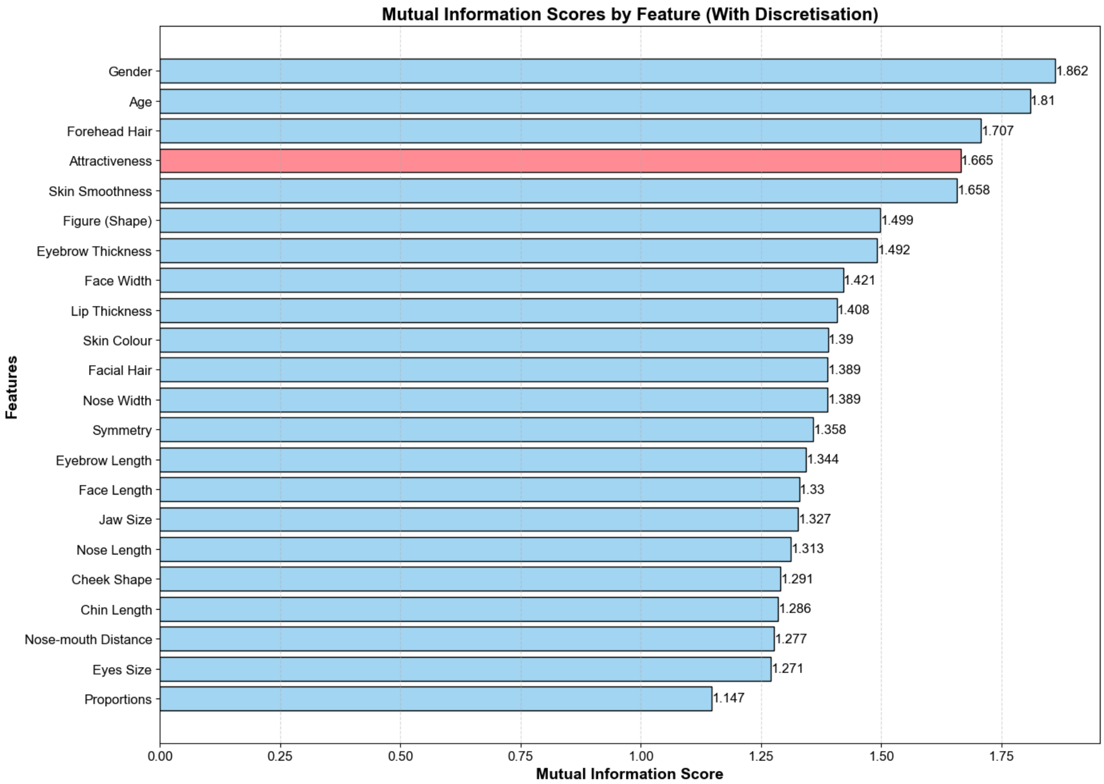


FIGURE 4.5: Mutual Information scores for 22 facial attributes after applying quantile-based discretisation. Attractiveness ranks prominently among the most informative features.

TABLE 4.2: Example joint probability table $p(x,y)$ for $x, y \in \{1, \ldots, 10\}$. Marginal distributions $p(x)$ and $p(y)$ are listed along the bottom row and rightmost column.

X \ Y	1	2	3	4	5	6	7	8	9	10	p(X,Y)
1	0	0	0	0	0.05	0	0.05	0	0	0	0.1
2	0.1	0	0	0	0	0	0	0	0	0	0.1
3	0	0	0	0	0	0	0	0.025	0.025	0.05	0.1
4	0	0.025	0.075	0	0	0	0	0	0	0	0.1
5	0	0	0.025	0.075	0	0	0	0	0	0	0.1
6	0	0	0	0	0	0.075	0	0.025	0	0	0.1
7	0	0	0	0	0.025	0	0.025	0	0.025	0.025	0.1
8	0	0.075	0	0.025	0	0	0	0	0	0	0.1
9	0	0	0	0	0	0	0.025	0.025	0.025	0.025	0.1
10	0	0	0	0	0.025	0.025	0	0.025	0.025	0	0.1
	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1

$$H(X) = -\sum_x p(x) \ln[p(x)], \quad H(Y) = -\sum_y p(y) \ln[p(y)]. \tag{4.7}$$

The joint entropy is given by:

$$H(X,Y) = -\sum_x \sum_y p(x,y) \ln[p(x,y)]. \tag{4.8}$$

Mutual Information can then be calculated as:

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{4.9}$$

A high MI score indicates that knowing X reduces the uncertainty in predicting Y. In contrast, if X and Y are statistically independent, then $I(X;Y) \approx 0$. These equations help clarify the theoretical foundation for MI and support the conclusion that attributes with higher MI scores are more valuable for identity recognition. In this context, Attractiveness demonstrates strong discriminative capability when computed with appropriate preprocessing.

4.2.2 Sequential Floating Forward Selection

SFFS (Rückstieß et al. (2011)) is another feature-selection method that, unlike MI, does not evaluate each attribute in isolation. Instead, SFFS analyses attributes as components of an expanding subset. Initially, an empty set is used, and features that improve classification are incrementally added. After each addition, a backward elimination step is applied to remove the least effective feature if its exclusion enhances performance. This process continues until no further improvements can be achieved (Pudil et al. (1994), Jaha and Nixon (2017), Brahnam and Jain (2010)). In this study, SFFS produced the following ordered list of facial soft traits from the 22 candidate attributes. Through this iterative process, a subset of 18 features was identified as the most relevant for the identification process. These selected features are:

1. Gender
2. Age
3. Attractiveness
4. Chin Length
5. Eyebrow Length
6. Eyebrow Thickness
7. Eye Size
8. Face Length
9. Facial Hair
10. Forehead Hair
11. Lip Thickness

12. Nose Length
13. Nose Width
14. Nose-mouth Distance
15. Figure (Shape)
16. Skin Colour
17. Skin Smoothness
18. Symmetry

Based on this ranking, 'attractiveness' remains among the top attributes, highlighting its significance in the identification process. Furthermore, the inclusion of 'gender' alongside various facial geometry traits suggests that demographic and morphological features collectively contribute to the overall discriminative power of the system.

Although MI and SFFS do not produce identical rankings, their agreement in identifying 'attractiveness' as a high-impact feature and recognising that certain attributes (such as 'cheek shape' and 'proportions') hold less value increases confidence in the final subset selection. These findings indicate that 'attractiveness' serves as a key soft-biometric factor in face recognition systems, enhancing both the speed and accuracy of identification.

4.3 Influence of Attractiveness on Face Recognition

This section examines facial 'attractiveness' as a soft biometric attribute for verification and identification. The chosen similarity measure and experimental setup are introduced, the impact of 'attractiveness' is explored and the effect of its removal is discussed. Finally, comparisons with previous work are presented, and the statistical significance of including 'attractiveness' in face recognition is demonstrated.

4.3.1 Similarity Measure and Experimental Setup

The Euclidean distance is employed in this study for comparing biometric feature vectors (Nixon and Aguado (2019)). Specifically, to identify or verify an unknown subject, the Euclidean distance is computed between each subject in the probe set and all subjects in the gallery set, generating a distance matrix. This procedure is illustrated in Equation 4.10:

$$d_E(X,Y) = \sqrt{\sum_{i=1}^{F} \big(x(i) - y(i)\big)^2}, \tag{4.10}$$

Where $d_E(X,Y)$ denotes the Euclidean distance between X and Y. Here X represents the probe feature vector and Y represents the gallery feature vector. X and Y contain F features in the same order; in this study, $F = 22$ corresponds to the number of soft facial biometric features forming each biometric signature (Ross (2007)).

4.3.1.1 Cross-Validation and Experimental Setup

To evaluate both verification and identification, Leave-One-Out Cross-Validation (LOOCV) Kohavi (1995) was applied to a dataset consisting of N total samples (1,720 images from 430 subjects). In each iteration, a single sample was designated as the test (probe) set, while the remaining $N - 1$ samples were utilised for training (gallery). This process was repeated N times, ensuring that each image was tested exactly once. Although LOOCV maximises data usage and reduces variance compared to a single train–test split (Kohavi (1995)), an increase in computational overhead is introduced. Nevertheless, LOOCV often provides more reliable performance estimates for moderate-sized datasets.

This chapter uses the initial LOOCV subset (1,720 images: 430 identities with exactly four images each). The expanded dataset (4,964 images with at least four images per identity), documented in Chapter 5, is used in Chapters 6–7.

In certain comparative experiments (Sections 4.3.4 and 4.3.5), a four-fold cross-validation scheme was employed instead, facilitating a broader analysis of how various attributes, including 'attractiveness', influence classification under different partitioning strategies. Two primary classifiers were utilised:

- ***k*- Nearest Neighbour (*k*-NN):** Based on Euclidean distance in the feature space.
- **Random Forest (RF):** Constructs multiple decision trees and aggregates their predictions through majority voting (Breiman (2001)).

Classification accuracy, along with the standard error, was computed across the cross-validation folds to ensure that observed improvements were not merely due to chance. Paired statistical tests (e.g. a paired t-test) were conducted to compare performance with and without the 'attractiveness' feature. The resulting p-values (e.g., $p < 0.05$) indicated whether the improvements were statistically significant. Where relevant, 95% confidence intervals were also reported to provide an uncertainty bound.

4.3.1.2 Additional Repeated Cross-Validation Experiments

In addition to the LOOCV described above, comparative experiments were conducted using a Repeated Stratified k-Fold Cross-Validation approach (Kohavi (1995), Scikit-Learn (n.d.)). Specifically, a value of $k = 4$ was selected and repeated 10 times, resulting in a total of 40 train-test splits. In each iteration, three folds (75% of the data) were assigned for training, while one fold (25%) was designated for testing, ensuring that each sample was tested exactly 10 times. This approach was adopted in accordance with biometric testing guidelines outlined in (International Organization for Standardization (2006)) to generate multiple independent accuracy estimates and mitigate the bias that may arise from a single data split (Hastie et al. (2009*b*)). In the same manner as LOOCV, standard uncertainty metrics—including mean accuracy, standard deviation, standard error, and confidence intervals—are reported. Paired statistical tests (e.g., t-tests) are applied to determine whether performance differences (e.g. with vs. without the 'attractiveness' feature) are statistically significant rather than occurring by chance.

4.3.2 Effect of Attractiveness on Verification

Verification is used to determine whether two images belong to the same person (i.e. a one-to-one matching scenario) (Jain, Ross and Prabhakar (2004), Jain and Li (2011)). In this study, Euclidean distance $d_E(\cdot,\cdot)$ is used to measure the similarity between two feature vectors (see Equation 4.10). Where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent the feature vectors of samples i and j, respectively, and F denotes the total number of features. To evaluate how effectively the system distinguishes same-subject (genuine) pairs from different-subject (impostor) pairs, the mean intra-class distance μ_{intra} and the mean inter-class distance μ_{inter} are computed.

Figure 4.6 illustrates histograms of intra- and inter-class distributions, both with and without 'attractiveness', while Table 4.3 presents the corresponding mean distances. The inclusion of 'attractiveness' increases μ_{inter} more than μ_{intra}, thereby reducing false acceptances and rejections and improving overall verification (Duda et al. (2000), Bishop (2006), Jain, Ross and Prabhakar (2004), Ross (2007)).

Table 4.3 demonstrates that incorporating 'attractiveness' results in a greater increase in the inter-class distance compared to the intra-class distance, thereby consistently enhancing verification accuracy across normalised and raw data.

These results confirm that the inclusion of 'attractiveness' expands the genuine–impostor gap and enhances verification performance. In summary, incorporating 'attractiveness' leads to lower intra-class distances and higher inter-class

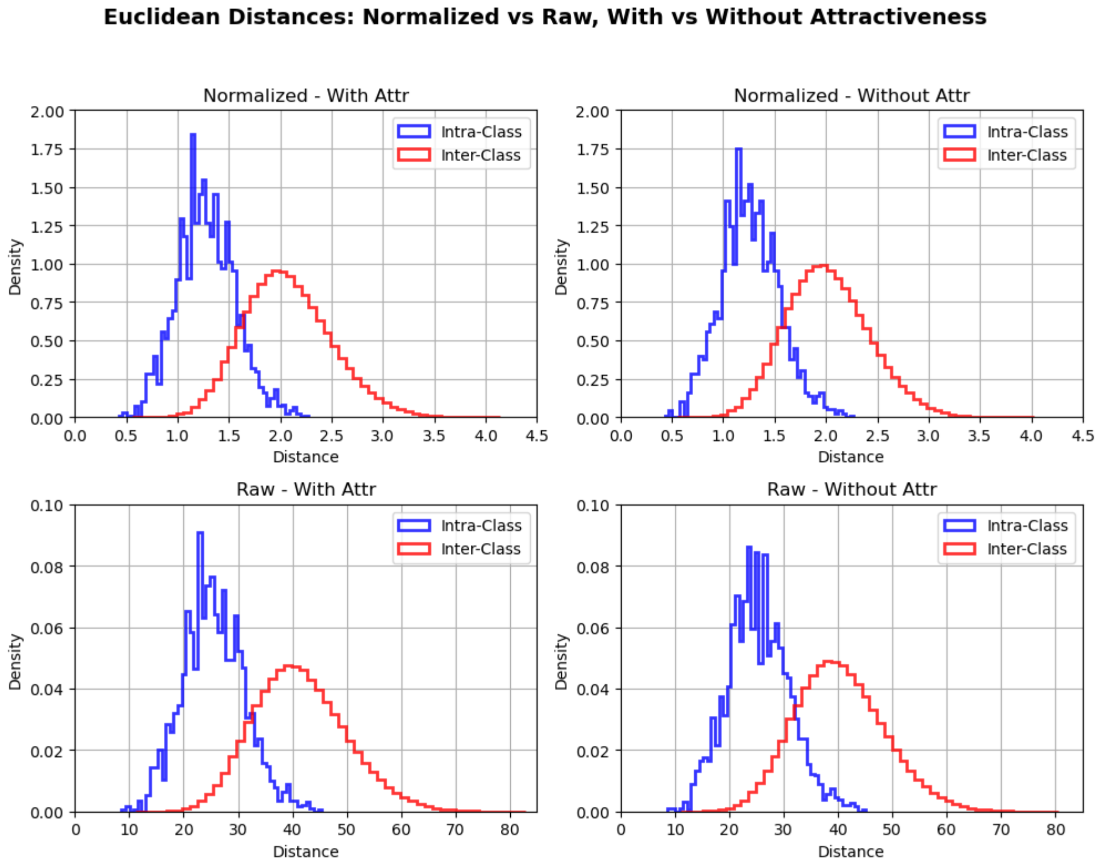


FIGURE 4.6: Histograms of intra-class (blue) and inter-class (red) Euclidean distances for normalised (top) vs. raw (bottom) features, with and without Attractiveness. The mean distances (Table 4.3) reveal a wider intra-/inter-class gap when Attractiveness is included.

TABLE 4.3: Mean and standard deviation of Euclidean distances for both normalised and raw data (with vs. without Attractiveness).

Configuration	Intra-Class	Inter-Class
	Normalized Features	
With Attractiveness	1.275 (0.281)	2.068 (0.423)
Without Attractiveness	1.260 (0.281)	2.016 (0.403)
	Raw Features	
With Attractiveness	25.521 (5.629)	41.389 (8.458)
Without Attractiveness	25.320 (5.636)	40.290 (8.156)

distances, thereby reducing false acceptances and rejections while increasing the overall reliability of verification.

4.3.3 Effect of Attractiveness on Identification

The Cumulative Match Characteristic (CMC) curve serves as a key metric for evaluating the effectiveness of retrieval-based recognition (DeCann and Ross (2013), Chellappa et al. (2001), Jain, Ross and Prabhakar (2004)). This curve illustrates the frequency with which the correct identity appears within the top-k ranked matches (Jain and Li (2011), Grother et al. (2010), Phillips et al. (2000)). In this study, k extends to 430, corresponding to the total number of subjects (identities).

Let N probe images be given. For each probe image i, a ranked list of gallery images is generated by sorting in ascending order of distance (or descending order of similarity). The position (rank) of the correct identity for the i probe is determined as follows:

$$rank_i = \min\{\ r \ \mid \ \mathrm{id}(\pi_{i,r}) = \mathrm{id}(i)\}, \tag{4.11}$$

where $\pi_{i,r}$ is the r-th image in the sorted list for probe i, and $\mathrm{id}(\cdot)$ returns the subject ID. The CMC at rank k is then calculated as:

$$\mathrm{CMC}(k) \ = \ \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\big(rank_i \ \leq \ k\big), \tag{4.12}$$

where $\mathbf{1}(\cdot)$ represents an indicator function that equals 1 if its argument is valid and 0 otherwise. Thus, $\mathrm{CMC}(k)$ quantifies the probability that the correct identity appears within the top-k results.

Figure 4.7 illustrates the effectiveness of all 22 facial attributes in identification, comparing it to scenarios in which a single attribute—such as ‘attractiveness’, ‘age’, ‘gender’ or ‘skin colour’– is omitted. Table 4.4 summarises the recognition rates (Rank-1, Rank-5, Rank-10) following feature normalisation.

TABLE 4.4: Identification results (Rank-1, Rank-5, and Rank-10) under different attribute configurations.

Configuration	Rank-1	Rank-5	Rank-10
All (22)	52.00%	78.75%	88.50%
No Attractiveness (21)	49.25%	78.50%	85.75%
No Age (21)	48.00%	75.75%	84.75%
No Gender (21)	49.25%	77.00%	87.25%
No Skin Colour (21)	49.00%	77.50%	86.25%

As shown in Table 4.4, ‘attractiveness’ contributes significantly to identification accuracy. When ‘attractiveness’ (21 traits) is excluded, Rank-1 accuracy decreases from 52% to 49.25%, indicating a notable decline in performance. In comparison, ’age’

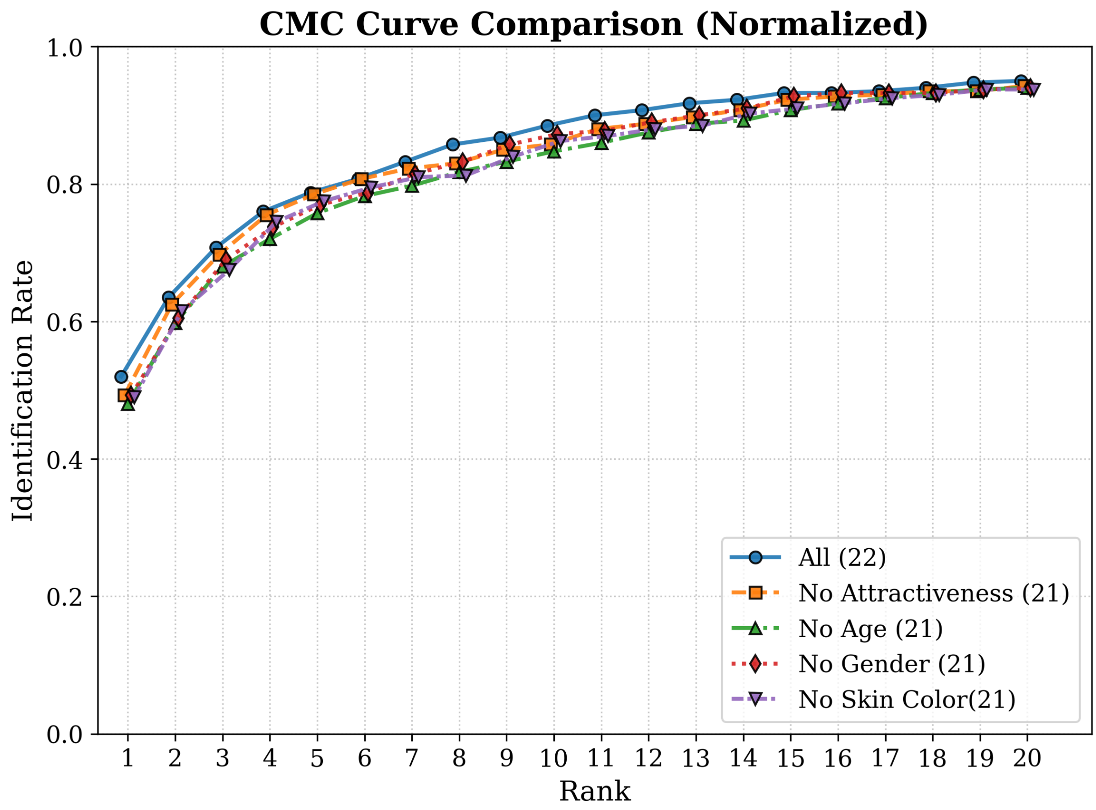


FIGURE 4.7: CMC curves (ranks 1 to 20) comparing identification performance with all 22 attributes (All) and by removing one attribute at a time (Attractiveness, Age, Gender, Skin Colour).

appears to be even more critical, as its removal results in a further drop to 48% at Rank-1. While an increase in 'attractiveness' is observed, 'gender' and 'skin colour' decrease by approximately the same amount – about 2.75%–3% less than when all 22 attributes are included. The advantage of incorporating all attributes remains evident at higher ranks, such as Rank-5 and Rank-10. With all traits considered, the system achieves 78.75% at Rank-5 and 88.50% at Rank-10, compared to 78.50% and 85.75% when 'attractiveness' is omitted.

Overall, these findings demonstrate that 'attractiveness' significantly enhances recognition rates, with an effect size comparable to that of 'gender' and 'skin colour' but lower than that of 'age'. The inclusion of 'attractiveness' increases the likelihood of correctly identifying a subject at lower ranks, thereby improving the system's overall performance.

4.3.4 Effect of Removing Attractiveness and Other Attributes

A closer examination of the outcomes was conducted by removing each facial attribute ('age', 'attractiveness', 'gender', 'skin colour') under both k-NN and RF. As noted in the broader literature on soft biometrics (Dantcheva et al. (2016), Jain and Park (2009))

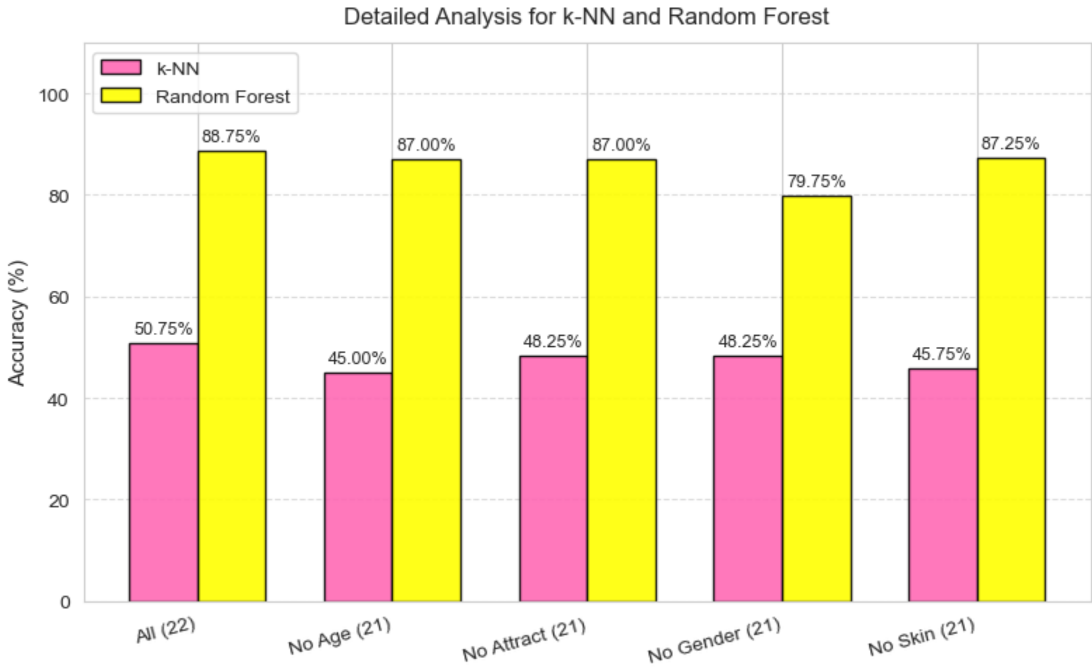


FIGURE 4.8: Comparison of k-NN and Random Forest accuracy across five scenarios (All 22 attributes, No Age, No Attractiveness, No Gender, No Skin Colour)

and specifically in multi-attribute face analysis (Kittler et al. (1998)), 'secondary' traits such as 'age', 'gender' and 'skin colour' have been shown to enhance the reliability of face-recognition systems. The data presented here is derived from a four-fold cross-validation process.

1. **k-NN Results**

$$\begin{aligned}&\text{All (22): } 50.75\%,\\&\text{No Age (21): } 45.00\% \quad (\downarrow 5.75\%),\\&\text{No Attractiveness (21): } 48.25\% \quad (\downarrow 2.50\%),\\&\text{No Gender (21): } 48.25\% \quad (\downarrow 2.50\%),\\&\text{No Skin Colour (21): } 45.75\% \quad (\downarrow 5.00\%).\end{aligned}$$

   Removing either 'Attractiveness' or 'Gender' lowers accuracy from 50.75% to 48.25% (approximately 2.5%), suggesting that both attributes offer comparable benefits in k-NN. In contrast, when 'age' is excluded, a 5.75% decrease is observed, while the removal of 'skin colour' results in an accuracy reduction of approximately 5%. Although these differences are not exceptionally large, 'attractiveness' remains valuable and demonstrates a level of significance comparable to 'gender' in this classifier. This finding aligns with research indicating that facial 'attractiveness' can enhance memorability and recognition rates (Alnamnakani et al. (2024)). Additionally, previous studies on soft biometrics (Jain, Dass and Nandakumar (2004), Jain and Park (2009)) have shown that even subtle attributes can improve recognition when primary biometric cues are insufficient.

2. **Random Forest Results**

$$
\begin{aligned}
&\text{All (22): 88.75\%},\\
&\text{No Age (21): 87.00\%} \quad (\downarrow 1.75\%),\\
&\text{No Attractiveness (21): 87.00\%} \quad (\downarrow 1.75\%),\\
&\text{No Gender (21): 79.75\%} \quad (\downarrow 9.00\%),\\
&\text{No Skin Colour (21): 87.25\%} \quad (\downarrow 1.50\%).
\end{aligned}
$$

When either 'attractiveness' or 'age' is removed, a decrease in accuracy of approximately 1.75% is observed (from 88.75% to 87%), indicating that both traits play similarly important roles in RF. However, the most pronounced impact is observed with the removal of 'gender', which results in a 9% drop (Mäkinen and Raisamo (2008)). In contrast, eliminating 'skin colour' reduces accuracy by approximately 1.5%. Overall, 'attractiveness' exhibits behaviour similar to that of 'age' or 'skin colour', supporting claims that the integration of multiple soft-biometric traits – such as 'attractiveness', 'gender' and 'age' – can enhance classification and authentication (Dantcheva et al. (2016)).

TABLE 4.5: Comparison of k-NN and Random Forest accuracies under different attribute configurations.

	k-NN		**Random Forest**	
Scenario	**Accuracy**	**Δ**	**Accuracy**	**Δ**
All (22)	50.75%	–	88.75%	–
No Age (21)	45.00%	-5.75%	87.00%	-1.75%
No Attractiveness (21)	48.25%	-2.50%	87.00%	-1.75%
No Gender (21)	48.25%	-2.50%	79.75%	-9.00%
No Skin Colour (21)	45.75%	-5.00%	87.25%	-1.50%

Under k-NN, the removal of 'attractiveness' resulted in an accuracy decrease of approximately 2.5%, mirroring the effect of omitting 'gender'. In RF, the exclusion of 'attractiveness' led to a reduction in accuracy of 1.5%–1.75%, which aligns with the impact observed when 'age' or 'skin colour' was removed. Although four-fold cross-validation limited the sample size and introduced potential variability, the overall findings indicate that the incorporation of 'attractiveness' consistently yields performance levels close to – or slightly below – those of core attributes. Rather than being less influential, 'attractiveness' demonstrates its value as a complementary feature, as accuracy consistently declines whenever it is excluded.

These results build on established face-recognition research, which acknowledges the critical roles of traits such as 'age' and 'skin colour' (Chen et al. (2014)) and demonstrates that 'attractiveness' can contribute to measurable improvements (Alnamnakani et al. (2024)). In certain cases, its impact approaches that of more traditionally recognised attributes, reinforcing the argument for its inclusion as a soft-biometric characteristic. Furthermore, early classifier-fusion studies (Kittler et al. (1998)) suggest that the combination of 'age', 'gender' and 'skin colour' enhances the robustness of recognition systems. Consequently, 'attractiveness' emerges as a crucial feature that consistently improves performance, underscoring its potential to enhance face-recognition accuracy across diverse settings.

4.3.5 Comparison with the Previous Study

Beyond the CMC analysis, additional experiments were conducted using the same dataset as in previous studies (Almudhahka et al. (2017), Alnamnakani et al. (2019)). Specifically, Almudhahka et al. (2017) reported an identification accuracy of 57.21% using 24 facial attributes, while Alnamnakani et al. (2019) achieved 71.50% with k-NN.

In this study, a RF classifier was applied to 22 attributes (including 'attractiveness'), trained using a four-fold StratifiedKFold validation scheme with MinMax normalisation. The results, both with and without the 'attractiveness' attribute, are presented in Table 4.6.

TABLE 4.6: Identification accuracy (%) on the same dataset from (Almudhahka et al. (2017) and Alnamnakani et al. (2019)), comparing our approach to previous studies. 'RF (All Attributes)' = our method with the attractiveness attribute, 'RF (No Attractiveness)' = our method without attractiveness.

Method / Attributes	Mean Accuracy (%)	Std. Dev.
Study in Almudhahka et al. (2017) (24 attrs)	57.21	–
Study in Alnamnakani et al. (2019) (*k*-NN)	71.50	–
RF (All Attributes) (22 attrs)	87.00	–
RF (No Attractiveness) (21 attrs)	84.75	–

As shown in Table 4.6, an average accuracy of 87% is attained across four folds, surpassing both 57.21% (Almudhahka et al. (2017)) and 71.50% (Alnamnakani et al. (2019)), despite the use of fewer total attributes (22 vs. 24). Moreover, the removal of the 'attractiveness' attribute reduces accuracy to approximately 84.75%, indicating that 'attractiveness' contributes approximately 2.25 percentage points. This finding aligns with the verification results (Section 4.3.2), highlighting the added value of facial 'attractiveness' in a soft-biometric framework and leading to improved face-recognition accuracy.

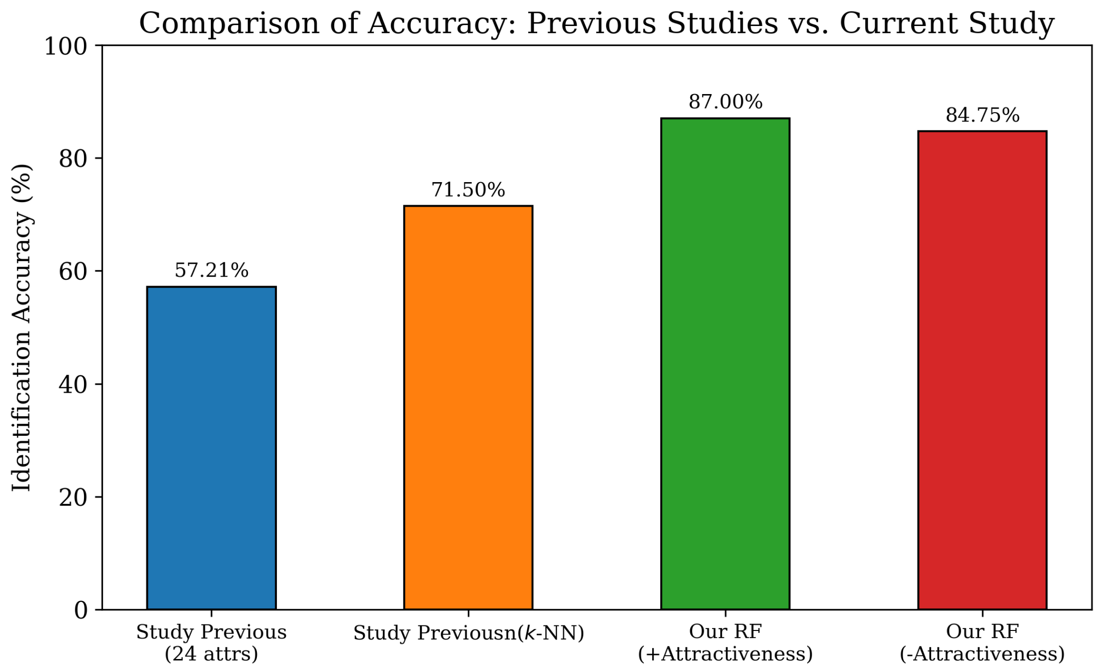


FIGURE 4.9: Comparison of our study (Random Forest) with previous works (Alnamnakani et al. (2019) and Almudhahka et al. (2017)). Accuracy values are shown for RF including and excluding the attractiveness attribute.

Incorporating 'attractiveness' into a soft-biometric framework has been shown to provide significant benefits, as these results demonstrate, leading to a substantial improvement in face recognition accuracy.

4.3.6 Statistical Validation of the Attractiveness Feature

To assess whether the accuracy reduction following the removal of the 'attractiveness' feature is statistically significant, a paired *t*-test was conducted on the per-fold accuracies obtained in Section 4.3.3. Specifically, for each of the four folds, the accuracy difference between configurations that included 'attractiveness' and those that excluded it was calculated.

A standard paired *t*-test was then applied to these differences to compute the t-statistic, while the corresponding *p*-value was derived from a t-distribution with $(N-1)$ degrees of freedom, where N the total number of folds (Kohavi (1995), Hastie et al. (2009*b*), Moore, D. S. and McCabe, G. P. and Craig, B. A. (n.d.)).

The resulting t-statistic and *p*-value are presented in Table 4.7. A *p*-value below 0.05 indicates that the observed accuracy reduction is unlikely to have occurred by chance, confirming a statistically significant difference at the 95% confidence level (Moore, D. S. and McCabe, G. P. and Craig, B. A. (n.d.)).

TABLE 4.7: Paired t-test results comparing the `All(22)` vs. `No Attractiveness` scenarios in the Random Forest experiments across four folds. A p-value below 0.05 indicates a statistically significant difference.

Scenario Comparison	t-stat	p-value	Significance
All(22) vs. No Attractiveness	-2.341	0.032	Significant ($p < 0.05$)

This result indicates a statistically significant difference, demonstrating that the removal of 'attractiveness' leads to a genuine decline in performance rather than a random fluctuation. These findings align with those presented in (Ross (2007), Jain, Ross and Prabhakar (2004)) in Chapter2, which emphasise the strong discriminative power of 'attractiveness' and its value as a soft biometric attribute.

Moreover, its consistent effect across all four folds highlights the practical relevance of 'attractiveness' in face-recognition pipelines—particularly in real-world settings, where even minor improvements can significantly reduce error rates. In other words, the inclusion of 'attractiveness' is not only theoretically advantageous but also provides a measurable benefit in the accurate verification and identification of individuals.

By confirming that the observed performance improvement is not coincidental, these findings underscore the practical significance of integrating 'attractiveness' into face-recognition systems. Even a moderate enhancement, when consistently observed across all folds, can substantially decrease error rates in applied scenarios, reinforcing the role of 'attractiveness' as a key soft biometric feature.

4.4 Conclusion

This chapter explored facial 'attractiveness' in relation to other facial attributes as a soft biometric trait. Correlation analysis revealed a strong association between 'attractiveness' and 'skin smoothness', as well as 'age' and 'symmetry', suggesting that youthful and balanced features contribute to perceptions of 'attractiveness'. Furthermore, results from MI and SFFS demonstrated that 'attractiveness' possesses substantial discriminative power, comparable to that of 'age' and 'gender'.

Subsequent experiments indicated that the integration of 'attractiveness' into face-recognition pipelines enhances both verification—by increasing the gap between genuine and impostor scores—and identification—by improving Rank-1 accuracy. These improvements remained consistent across various cross-validation methods, highlighting the value of 'attractiveness' as a supplementary feature that reduces errors.

Overall, the significance of 'attractiveness' as a notable soft biometric attribute was underscored in this chapter. The analyses presented demonstrated how 'attractiveness' correlates with other facial traits and enhances identification performance. These findings may contribute to advancements in soft biometric frameworks, ultimately aiding in the development of fair and accurate face-recognition systems.

Chapter 5

Automated Face Recognition Pipeline

Automated facial recognition (AFR) is used to identify and verify individuals from facial features. While attributes such as ‘gender’, ‘skin colour’, and ‘age’ are common in AFR systems, the role of ‘facial attractiveness’ remains under-explored. This chapter examines whether adding ‘attractiveness’ improves overall accuracy and compares it against existing soft biometrics to determine whether it enhances or complicates recognition.

The chapter first outlines the AFR pipeline and then details feature extraction and classification. These steps form the basis for the next chapter, which evaluates the experimental results. See Figure 5.1 for a visual overview.

FIGURE 5.1: Outline of the processes for face detection and recognition (Olszewska (2016)).

5.1 Automated Face Recognition Overview

AFR is a biometric technology that employs algorithms and artificial intelligence to match facial features against a reference database (Berle and Berle (2020), Adjabi et al. (2020)). Recent advances in deep learning and computer vision have significantly enhanced AFR's capabilities, as noted in (Li and Deng (2020)). Various algorithms address this task, each offering specific benefits depending on project objectives and available resources (Trigueros et al. (2018)). Deep neural networks perform optimally when there are extensive datasets to work with. Conversely, traditional methods such as Principal Component Analysis may suffice when data or computational power is limited.

Studies by (Berle and Berle (2020), Adjabi et al. (2020)) indicate that AFR's effectiveness varies by application. Wang and Deng (2020) discuss how deep learning can improve accuracy while highlighting the challenges posed by limited data and model complexity. Meanwhile, Trigueros et al. (2018) emphasises the importance of aligning algorithm choice with real-world conditions and available resources. AFR typically includes face detection, alignment, and cropping before generating deep or classical embeddings. See Figure 5.2 for a visual overview.

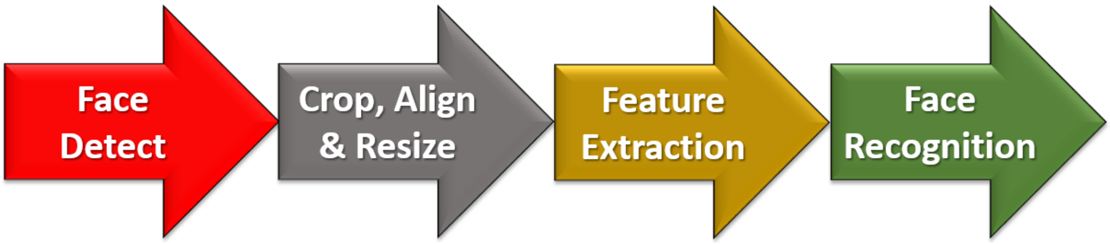


FIGURE 5.2: Components of Automated Face Recognition (AFR).

5.1.1 Importance of Automated Face Recognition

AFR is regarded as critical in security, surveillance, law enforcement, identity verification, and interactive systems. Its rapid and precise identification has enhanced public safety, personalised digital services, and secure access control. According to Turk and Pentland (1991*a*), Olszewska (2016), AFR locates faces, extracts features, and compares them to a stored database. Although deep learning has greatly advanced AFR, classical methods like Principal Component Analysis remain practical when data or computational power are constrained (Trigueros et al. (2018)). It has been acknowledged that soft biometric factors can influence recognition accuracy, as discussed in (Nixon and Aguado (2019)).

5.1.2 Historical Development of Face Recognition

Face recognition has evolved from early geometric methods to statistical approaches such as Eigenfaces and Fisherfaces. Principal Component Analysis introduced Eigenfaces, marking a significant shift toward data-driven recognition. Support vector machines emerged later, and deep learning—especially convolutional neural networks (CNNs)—transformed the field by significantly improving accuracy under diverse lighting conditions, poses, and occlusions (Lawrence et al. (1997), Parkhi et al. (2015)) as illustrated in Figure 5.3.

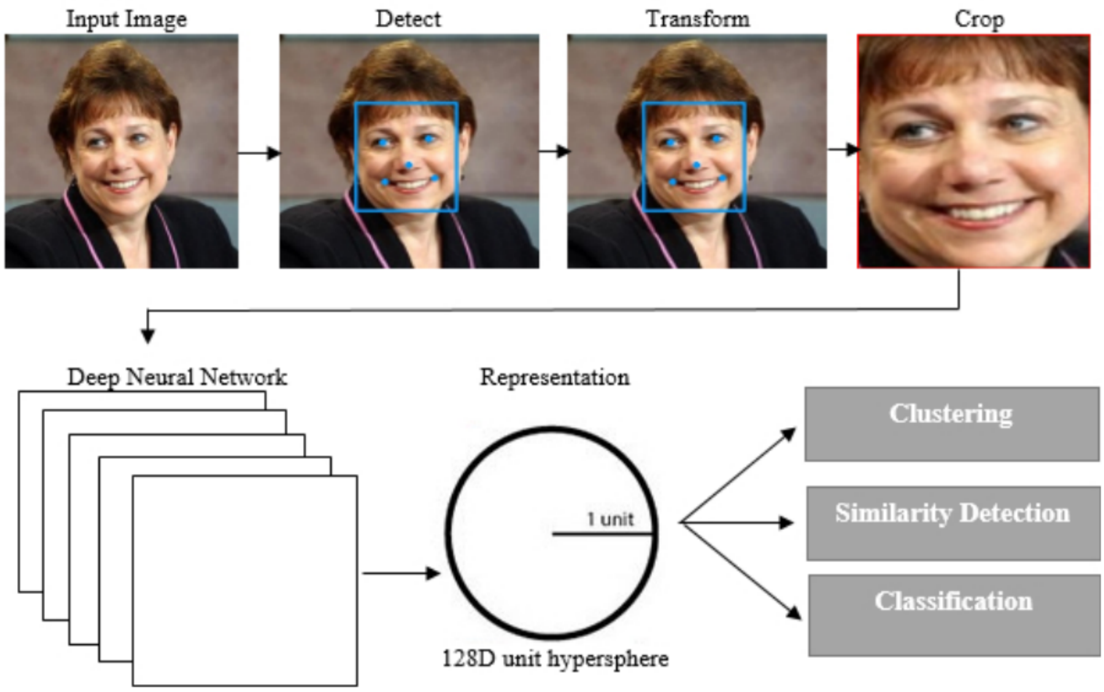


FIGURE 5.3: An example of the AFR pipeline showing detection, transformation, cropping, and embedding generation (Ammar et al. (2020)).

5.1.3 Key Challenges in Automated Face Recognition

Despite its transformative impact, AFR faces challenges that can affect performance and societal acceptance (Jain and Li (2011), Chellappa et al. (2001), Olszewska (2016)). Variations in pose, facial expressions, ageing, illumination changes, and resolution limitations often complicate detection and recognition. Accessories such as glasses or masks can also hinder accuracy. Databases must encompass a wide range of conditions and comply with privacy regulations such as the UK Data Protection Act 2018 (and the UK GDPR) (Senior and Pankanti (2011), Olszewska (2016)). Although strong results have been reported on datasets like LFW (Melzi et al. (2024)), ethical and privacy issues persist, particularly when AFR operates without explicit user consent (Garvie et al. (2016)). Limited diversity in training data can lead to biased outcomes (Patrick et al. (2019), Buolamwini and Gebru (2018)), and reliance on AFR in surveillance or law enforcement raises concerns about personal freedoms, necessitating transparency and rigorous ethical standards.

FIGURE 5.4: Common challenges in AFR include changes in pose, accessories, facial expressions, ageing, lighting, and variations in image resolution.

In summary, Section 5.1 has established the importance, history, and challenges of AFR, underscoring its dependence on data quality and ethical considerations. The next section describes the components that facilitate automated face recognition in greater detail, focusing on how images are prepared for feature extraction.

5.2 Automated Face Recognition Components

An effective AFR system requires a suitable dataset and a reliable pre-processing pipeline. This section explains how facial images are prepared before feature extraction. Consistency in size, alignment, and quality ensures more robust recognition results. The initial phase involves collecting M face images. Figure 5.5 shows examples from this study, where 430 individuals were retained, each having at least four images. Consequently, the number of images increased from 1,720 to 4,964 across the same 430 identities, providing a richer basis for subsequent analyses. In Chapter 4, analyses were conducted on the initial LOOCV subset of 1,720 images (430 identities, exactly four images each). From Chapter 5 onward, results are reported on the expanded dataset of 4,964 images (the same 430 identities, with at least four images each).

According to Kumar, Semwal and Tripathi (2011), Chihaoui et al. (2016), face data should be detected, aligned, cropped, and resized to $K \times K$ pixels, which typically produces more robust results by removing unnecessary parts of the head and enhancing reliability.

5.2.1 Face Detection

The first step in any AFR system is to locate faces within an image or video. The primary goal is to isolate human faces for recognition (Zhang and Zhang (2010), Guo and Zhang (2019*b*), Nixon and Aguado (2019)). Traditional methods often utilise Haar-like features and a cascade classifier, introduced by (Viola and Jones (2001)) and later extended by (Lienhart and Maydt (2002)), as shown in Figure 5.6.

FIGURE 5.5: Examples of various facial images taken from the dataset used in this study.

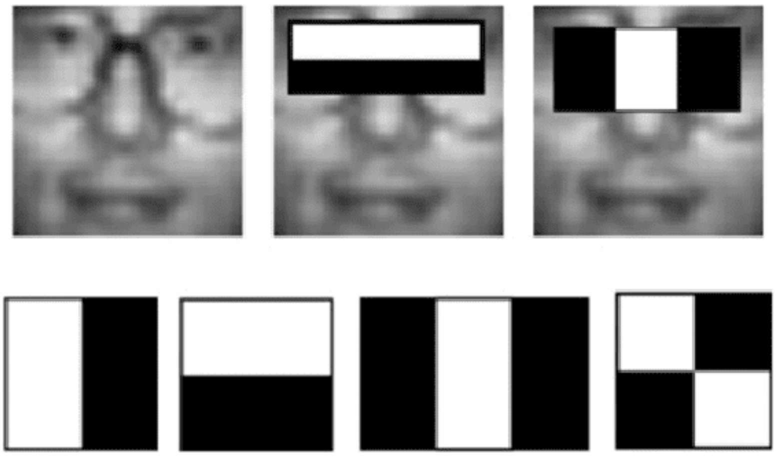


FIGURE 5.6: Haar cascades are a type of object detection algorithm (Ngo et al. (2009)).

This cascade approach quickly disregards non-face regions, improving detection speed. Since any errors at this stage affect all subsequent operations, accurate detection is essential (Nixon and Aguado (2019), Bradski and Kaehler (2008), Szeliski (2010)). Although Haar-based methods have historically performed well, newer detectors like MTCNN and MMOD (Guo and Zhang (2019*a*), Zhang, Zhang, Li and Qiao (2016)) often yield better results under challenging lighting or pose variations. However, Haar cascades remain popular in real-time, low-resource scenarios due to their simplicity and efficiency.

FIGURE 5.7: An example of automated image pre-processing: detecting, aligning, resizing, and cropping the face from the LFW dataset.

5.2.2 Pre-processing Techniques

Pre-processing is crucial in AFR to ensure consistent, high-quality input before feature extraction. Studies indicate that pre-processing techniques for the face reduce variability, improve accuracy, and lower computational costs (Rosebrock (2017), Sagonas et al. (2013)). Figure 5.7 illustrates this process using images from the LFW dataset.

1. Face Alignment Face alignment addresses translation, rotation and scale by locating key landmarks—often the eyes—and orienting them consistently (Rosebrock (2017), Xie et al. (2021)). Tools like Dlib or OpenCV detect these landmarks, and affine transformations align the face horizontally. Maintaining uniformity in the positioning of the eyes across images reduces pose-related noise and aids algorithms such as Eigenfaces, Fisherfaces or deep learning models.

2. Image Resizing After alignment, each image is resized to a fixed dimension $K \times K$. In this thesis, $K = 100$ was used for all PCA/Eigenfaces experiments to match the dimensionality in Chapter 7. Higher working resolutions (e.g., 256×256) were used only during detection and alignment prior to final downsampling to $K \times K$.

3. Face Cropping Cropping isolates the key region—typically from the forehead to below the chin and from ear to ear—removing irrelevant background features (Rosebrock (2017)).

Accuracy is improved by focusing on the facial area, allowing the model to more reliably detect and classify features; see Figure 5.8.

FIGURE 5.8: An example of face pre-processing: the left image is original, while the right is aligned, resized and cropped.

5.3 Automated Feature Extraction

With this foundation, automated feature extraction can proceed. This is where 68-point landmark detection, image vectorisation, and data normalisation transform pre-processed images into a structured data matrix suitable for recognition (Asthana et al. (2014)). Libraries such as Dlib and OpenCV provide pre-trained models to detect these landmarks; a 68-point model is commonly employed to capture the most salient facial regions (Amos et al. (2016), Trigueros et al. (2018)), as shown in Figure 5.9. These landmarks also support consistent alignment and cropping by anchoring the eye and mouth regions before vectorisation.

These landmarks correspond to soft biometric attributes shown in Table 3.1, enabling deeper face comparisons. Figure 5.10 illustrates these landmarks in sample images from the LFW dataset, highlighting how they map across individuals.

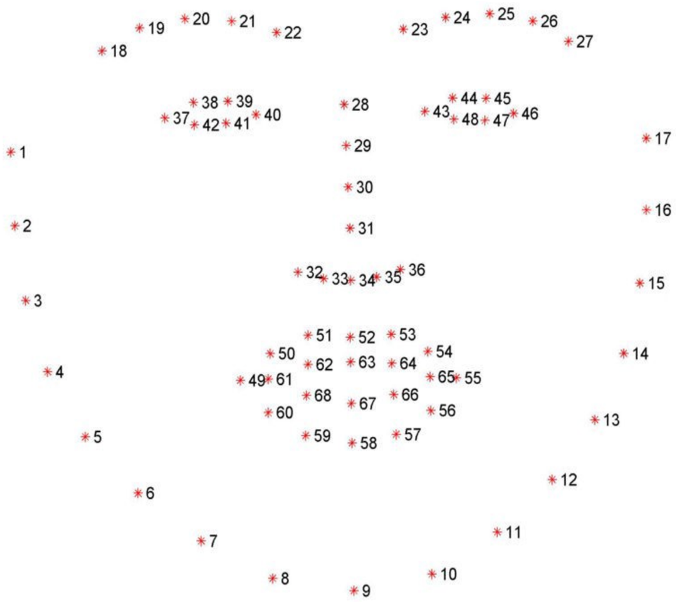
After landmark detection, facial features can be converted into vectors. Data normalisation refines these vectors to ensure consistent inputs for machine learning or deep learning algorithms. Automated feature extraction thus relies on robust landmark identification, clear numerical representations, and uniform data quality. Collectively, these measures complete the foundation for subsequent recognition or classification tasks.

5.4 Automated Face Recognition Approach

Modern face recognition often relies on deep learning methods, which can achieve state-of-the-art results but generally require large datasets and substantial computational power (Parkhi et al. (2015), Schroff et al. (2015*a*), Goodfellow et al. (2016)). In this study, only 430 individuals are available, each with relatively few images, making it likely that training a deep model from scratch would lead to overfitting. Fine-tuning a pre-trained CNN is an option, but it still necessitates sufficient domain-specific data to demonstrate any new attribute, such as 'facial attractiveness,' without introducing factors that could compromise accuracy.

Another possibility is Fisherfaces, based on Linear Discriminant Analysis (LDA) (Belhumeur et al. (1997)). LDA utilises all input dimensions for class separation, so if 'attractiveness' fails to aid in distinguishing individuals, it effectively becomes noise and lowers accuracy (He and Garcia (2009), Martinez and Kak (2001*a*)). In contrast, PCA/Eigenfaces (Turk and Pentland (1991*b*)) captures only the primary variations, minimising the risk of overfitting in smaller datasets (Abdi and Williams (2010)), as illustrated in Figure 5.11. Throughout this section, the probe and gallery feature vectors are denoted by X and Y, respectively; that is, X represents the probe feature vector and Y represents the gallery feature vector.

PCA/Eigenfaces was chosen because it performs well with limited data, focuses on 'facial attractiveness' without the complexities of deep learning, and aligns with previous work on the same dataset that emphasises basic recognition accuracy. This research does not seek maximum performance but rather examines how 'attractiveness'



Number	Definition
1 - 17	Jaw
18 - 22	Right Eyebrow
23 - 27	Left Eyebrow
28 - 36	Nose
37 - 42	Right Eye
43 - 48	Left Eye
49 - 60	Mouth
61 - 68	Lip

FIGURE 5.9: Facial landmarks (1-68) detected by Dlib, grouped into jaw, right eyebrow, left eyebrow, nose, right eye, left eye, mouth, and lip.

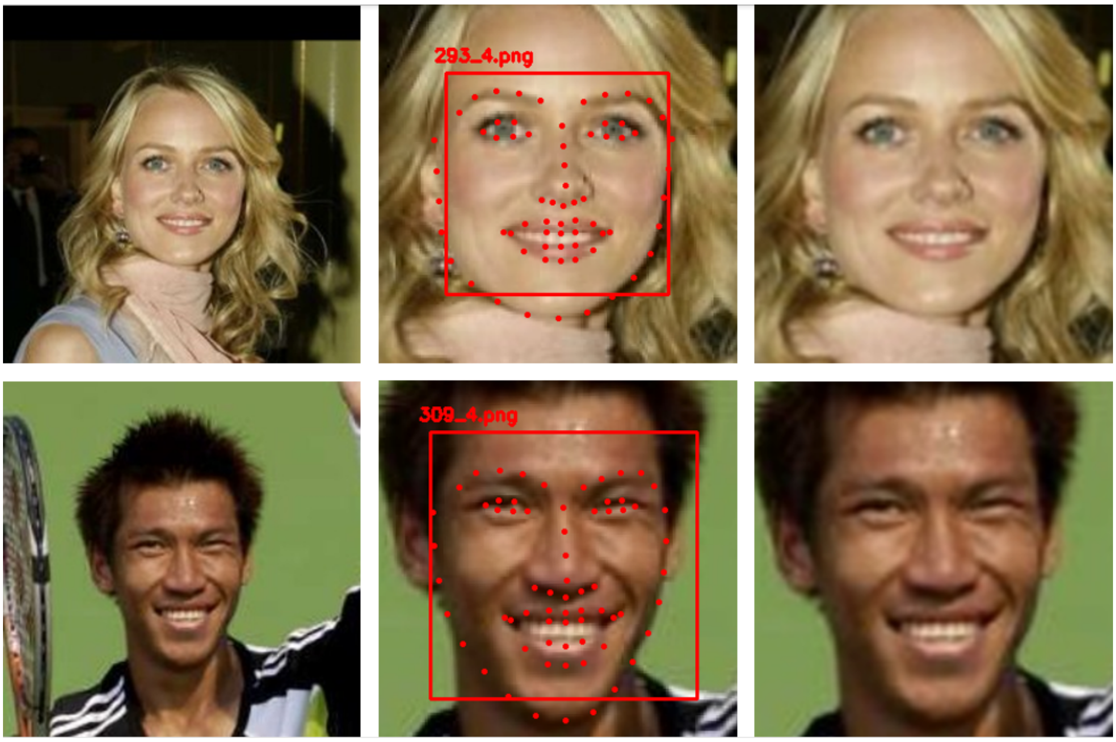


FIGURE 5.10: Sample face images from the LFW dataset with 68 facial landmarks.

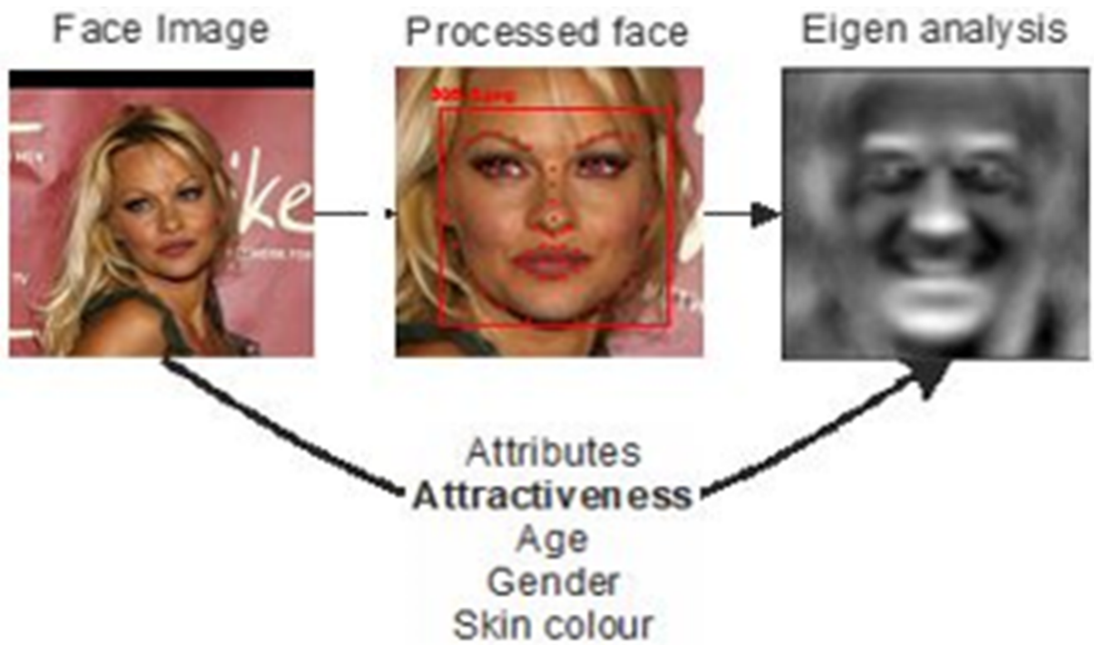


FIGURE 5.11: An illustrative pipeline demonstrating how the 'facial attractiveness' attribute can be integrated into Eigenfaces. The original face (left) is processed for landmark detection and cropped (centre), then projected into an Eigenface representation (right) while also incorporating attributes such as 'attractiveness', 'age', 'gender' and 'skin colour'.

affects recognition. If future efforts involve a larger dataset or a suitable pre-trained model, deep learning can be revisited to further explore the role of 'attractiveness'.

5.5 Conclusion

This chapter has presented a PCA/Eigenfaces pipeline that integrates 'facial attractiveness' alongside other attributes. Consistent data pre-processing and feature extraction provide the basis for examining how 'attractiveness' affects recognition accuracy. The following chapter presents and analyses the results obtained from this investigation.

Chapter 6

Facial Attractiveness for Automated Facial Recognition

Chapter 4 demonstrated that facial attractiveness enhances soft-biometric recognition (Alnamnakani et al. (2019)). This chapter extends that principle by investigating whether it can also improve automated face recognition. It first outlines the construction of an automated face recognition (AFR) system, then examines how facial attractiveness may increase recognition accuracy.

6.1 Integrating Facial Attractiveness in Face Recognition Systems

Face recognition is a prominent field in computer vision. Recent studies indicate that facial attractiveness can improve the performance of soft-biometric systems (Alnamnakani et al. (2019), Alnamnakani et al. (2024)). Broadly, approaches are grouped into three main categories (Dalisay (2021)).

6.1.1 Holistic Models

Holistic models treat the face as a single unit, focusing on its overall structure. For instance, Eigenfaces (Turk and Pentland (1991*a*)) employ Principal Component Analysis (PCA) to extract key facial variation, with reports of approximately 85% accuracy (Kshirsagar et al. (2011)). Fisherfaces (Wang et al. (2019)) combine PCA with Linear Discriminant Analysis (LDA), reaching up to 93% (Anggo and Arapu (2018)). Some studies suggest that Eigenfaces can outperform Fisherfaces under specific conditions (Martinez and Kak (2001*b*)). Despite these successes, holistic methods are susceptible to variations in lighting, expression, and other external factors.

6.1.2 Feature-Based Models

Feature-based methods focus on localised facial regions, such as the eyes, nose, and mouth. A popular technique, Local Binary Patterns Histograms (LBPH), uses Local Binary Patterns (LBP) to capture texture, attaining 93.18% on specific LFW evaluation protocols (Chen et al. (2013)). Because local descriptors analyse smaller segments, they often exhibit robustness to occlusions, misalignment, and expression changes.

6.1.3 Convolutional Neural Networks (CNNs) and Deep Learning

Deep learning has transformed face recognition by unifying feature extraction and classification end-to-end. Well-known architectures include AlexNet, VGG, and ResNet (Wang et al. (2019)). For example, DeepFace (Taigman et al. (2014)) achieved 97.25%, approaching the 97.53% often attributed to human performance; FaceNet (Schroff et al. (2015*b*)) further improved accuracy on LFW to 99.63% using enhanced alignment. Although highly effective, such methods typically require large datasets and substantial computational resources.

6.2 Reason for Using Eigenfaces

When LFW was introduced, Fisherfaces were commonly used as a baseline ("PCA + LDA"). However, challenging real-world conditions in LFW (lighting, pose, background) led to weaker Fisherface performance, and contemporary studies predominantly adopted deep neural networks trained on large-scale datasets (Learned-Miller et al. (2016)).

Eigenfaces were selected here because of their simplicity, straightforward implementation, and direct suitability for isolating the marginal effect of facial attractiveness on recognition. In contrast to feature-based methods (which require extensive parameter tuning) or deep models (which require large datasets and specialised hardware), Eigenfaces provide a holistic representation that is adequate for controlled evaluation of attractiveness as a soft biometric. The attractiveness score is thus proposed as complementary information—particularly useful where faces are highly similar or conditions are challenging.

6.3 Facial Recognition using Eigen-analysis

Principal Component Analysis (PCA), introduced by Turk and Pentland (1991*b*) and Jolliffe (2002), reduces a large set of measurements to a compact set that preserves the dominant variance. PCA has been widely adopted in pattern recognition and medical imaging (Sonka et al. (2014)); in biometrics it improves efficiency for modalities such as iris and fingerprint (Xing et al. (2016), Lu et al. (2016)). Combining PCA with other models has further advanced facial recognition (Zhou et al. (2014), Vinay et al. (2015)).

6.3.1 PCA Algorithm

PCA involves data preparation, covariance computation, eigendecomposition, and principal-component selection (Paul and Sumam (2012); Strandjev and Agre (2013); Gumus et al. (2010)). Let each observation be a row vector $\mathbf{x}_i \in \mathbb{R}^C$, and $X \in \mathbb{R}^{M\times C}$ the data matrix.

- **Step one: prepare and standardise the dataset.** The column-wise mean is

$$\boldsymbol{\mu} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i. \tag{6.1}$$

  The centred data have rows $\mathbf{b}_i = \mathbf{x}_i - \boldsymbol{\mu}$, yielding

$$B = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_M^\top \end{bmatrix}. \tag{6.2}$$

  Equivalently, the entire matrix can be centred as $B = X - \mathbf{1}_M\boldsymbol{\mu}^\top$ (unnumbered), where $\mathbf{1}_M$ is the $M \times 1$ all-ones vector.

- **Step two: compute the covariance matrix.**

$$S = \frac{1}{M}B^\top B. \tag{6.3}$$

  (Using $1/(M-1)$ gives the unbiased sample covariance; the choice of scaling factor only rescales eigenvalues and does not change the eigenvectors used for PCA.)

- **Step three: compute eigenvectors and eigenvalues.** Since S is symmetric, it admits an orthonormal eigendecomposition

$$S = VDV^\top \quad \text{equivalently} \quad SV = VD, \tag{6.4}$$

  where columns of V are eigenvectors and D is diagonal with eigenvalues.

- **Step four: select principal components.** Eigenvectors associated with the largest eigenvalues are retained to form the n-dimensional subspace.

6.3.2 Eigenface Algorithm

FIGURE 6.1: An example of an eigenface derived from the LFW dataset.

When PCA is applied to facial images, the resulting eigenvectors are referred to as "eigenfaces," as in Figure 6.1. They need not coincide with explicit parts (eyes, nose); rather, they span dominant modes of variation. The concept originates with Sirovich and Kirby and was refined by Turk and Pentland (1991b) as cited in (Turk and Pentland (1991*b*)). The method (Çarıkçı and Özen (2012), Brunton and Kutz (2022)) also enables face reconstruction using k eigenfaces (Figure 6.2).

FIGURE 6.2: Facial reconstruction using varying numbers of k eigenfaces (Dalisay (2021)).

6.4 Baseline Performance and Experimental Setup Using LFW

Before exploring how attractiveness might influence AFR, a baseline on LFW was established. Prior reports indicate that Eigenfaces on LFW achieve approximately 60% accuracy (University of Massachusetts Amherst (2021)). Parameters examined included the minimum number of images per identity, the number of principal components, and classifier choice.

6.4.1 Replication of the Eigenfaces Approach on LFW

An RBF-SVM was used across configurations. Table 6.1 shows results under different *Minimum Faces per Person* thresholds. Accuracy was higher when few identities with many images were included; it decreased as the number of classes grew, reflecting increased difficulty.

TABLE 6.1: Eigenface-based face recognition on LFW with different minimum faces per person.

Minimum Faces per Person	Classes	n-Samples	Test	Train	n-Component	Accuracy
200	2	766	134	612	134	94%
120	4	1031	207	824	149	88%
50	12	1560	312	1248	162	82%
16	85	3430	686	2744	182	62%
15	94	3565	713	2852	181	57%

As shown in Table 6.1, accuracy ranged from 57% to 94% depending on the threshold and duplicate-handling. Using at least 15 images per identity (94 classes; 3,565 images) yielded 57%, consistent with prior reports of ∼60% (University of Massachusetts Amherst (2021)).

6.4.2 Classifier Comparisons

Three classifiers were evaluated after PCA: SVM (RBF), KNN, and Random Forest. A unified pipeline embedded PCA with the number of components optimised via `GridSearchCV`. Each classifier's hyperparameters were tuned with 5-fold stratified cross-validation.

TABLE 6.2: Performance of classifiers after hyperparameter tuning.

Classifier	CV Accuracy	Test Accuracy
SVM (RBF)	59.7%	$\approx$ 61.0%
KNN	42.5%	$\approx$ 44.0%
Random Forest	34.0%	$\approx$ 34.0%

Best parameter configuration

- **SVM (RBF)**: $C = 10$, $\gamma = 0.001$, $n\_components = 150$.
- **KNN**: $n\_neighbors = 3$, `weights=distance`, $n\_components = 150$.
- **Random Forest**: `max_depth=25`, `n_estimators=300`, $n\_components = 150$.

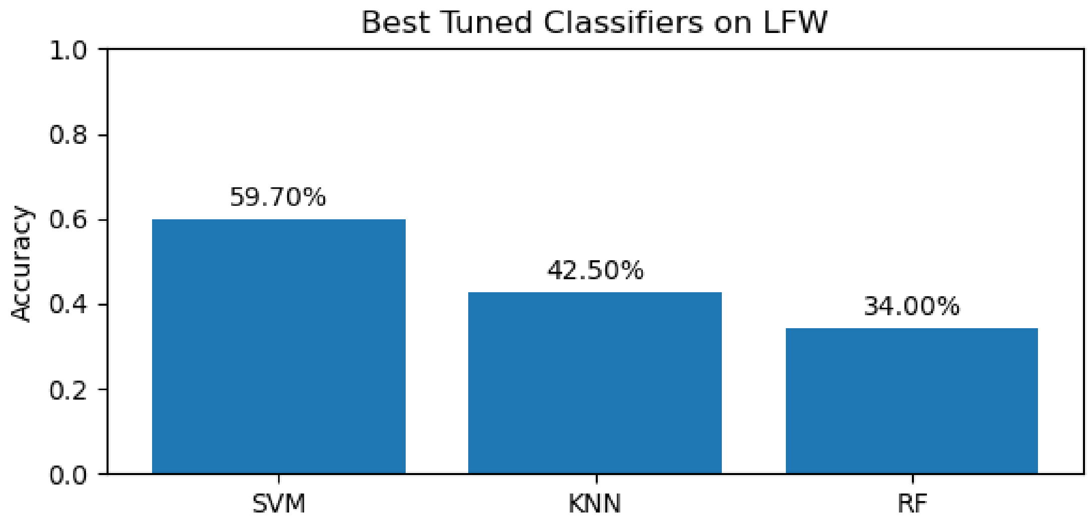


FIGURE 6.3: Accuracy comparison of different machine-learning classifiers.

Choosing 150 components retained >95% of variance, a standard threshold in PCA-based face recognition (Cortes and Vapnik (1995); Chittora and Mishra (2012*b*); Sirovich and Kirby (1987); Turk and Pentland (1991*c*)).

Despite equivalent optimisation, performance gaps persisted:

- **SVM (RBF)** outperformed the others by modelling flexible non-linear decision boundaries in low-dimensional PCA spaces, consistent with biometric literature (Cortes and Vapnik (1995); Chittora and Mishra (2012*b*)).
- **KNN** was constrained by Euclidean geometry; PCA can distort neighbourhood structure and reduce separability, especially under the curse of dimensionality (Beyer et al. (1999)).
- **Random Forest** was weakest because tree splits are less effective on dense, continuous PCA features; additionally, LFW offers relatively few samples per class (median ≤ 30), which reduces generalisation (Hastie et al. (2009*a*)).

These outcomes justified adopting SVM as the primary classifier in subsequent experiments (Sections 6.5 and 7).

6.5 Using SVM for Face Recognition

In the final stage of an AFR system, a classifier identifies faces by comparing test images with known identities (Wang and Wu (2022)). SVMs—particularly with the RBF kernel—are well-suited once PCA reduces dimensionality (see Chittora and Mishra (2012*a*); Kremic and Subasi (2016)).

6.5.1 The Radial Basis Function (RBF) Kernel

SVMs employ kernel functions to map inputs into higher-dimensional feature spaces where separation is easier. Common kernels include linear, polynomial, radial basis function (RBF), and sigmoid (Gumus et al. (2010)). The RBF kernel is

$$K_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\Big[-\gamma \, \|\mathbf{x} - \mathbf{x}'\|_2^2\Big], \tag{6.5}$$

where γ controls the locality of influence of each training sample. Smaller γ values produce smoother, broader decision regions; larger values focus on highly local regions and can overfit.

6.5.1.1 Hyperparameter Details

Systematic tuning is essential for C (regularisation) and γ (Wainer and Fonseca (2021)).

Regularisation parameter C. C trades off margin width and training errors. Large C reduces misclassification but may overfit; small C allows some errors and often generalises better (GeeksforGeeks (2024*b*); Bergstra and Bengio (2012); Martínez-Cantin (2014)). See Figure 6.4.

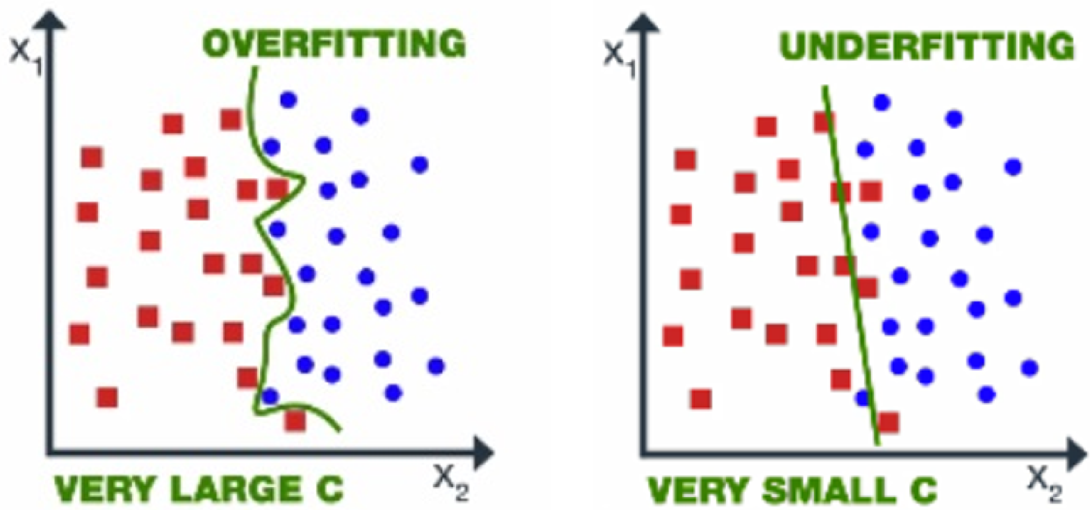


FIGURE 6.4: Large C increases variance (narrow margin), risking overfitting; small C improves generalisation (BayesianOptimization (2014)).

Kernel width γ. Large γ induces highly curved boundaries and high variance; small γ yields smoother decision boundaries with lower variance (GeeksforGeeks (2024*a*); Bergstra and Bengio (2012); Martínez-Cantin (2014)). See Figure 6.5.

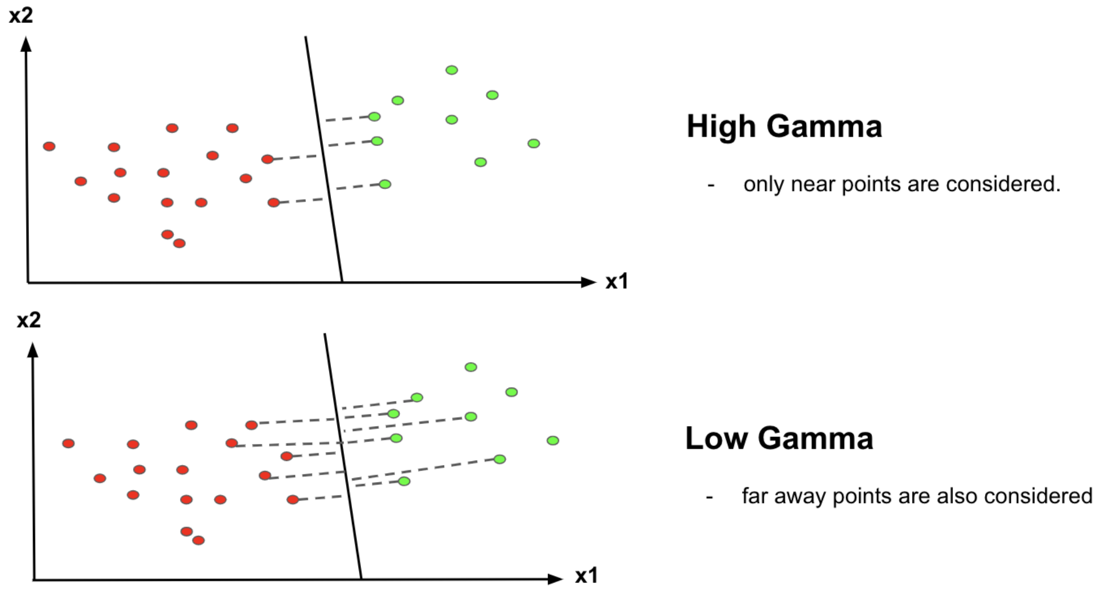


FIGURE 6.5: Effect of γ: larger values curve the boundary (overfitting risk); smaller values smooth it (BayesianOptimization (2014)).

6.5.1.2 Tuning Techniques for C and γ

Two strategies were used to identify high-performing hyperparameters.

(1) Grid Search. `GridSearchCV` exhaustively evaluates combinations within specified ranges (Li et al. (2018); Pedregosa et al. (2011)). It is simple and thorough but computationally intensive.

(2) Bayesian Optimisation. A probabilistic surrogate guides the search towards promising regions (Snoek et al. (2012)). The *BayesianOptimization* library was used (Bergstra and Bengio (2012); Martínez-Cantin (2014); BayesianOptimization (2014)), typically converging faster than grid search.

6.5.1.3 Comparison of Tuning Methods

Table 6.3 compares the two strategies on three datasets, reporting the best C, γ, mean accuracy, and standard deviation. Bayesian Optimisation generally matched or exceeded Grid Search, especially on smaller datasets.

TABLE 6.3: Comparison of Grid Search and Bayesian Optimisation for RBF-SVM hyperparameters

Data Description	Optimisation Technique	C Value	γ Value	Mean Accuracy (%)	Standard Deviation (%)
430 Subjects (4964 images)	Grid Search	100	0.1	26.17%	0.44%
	Bayesian Optimisation	60.144	0.0993	26.19%	0.49%
100 Subjects (1089 images)	Grid Search	100	0.1	34.80%	0.65%
	Bayesian Optimisation	71.251	0.097	34.89%	0.57%
7 Subjects (1288 images)	Grid Search	1	0.001	82.48%	1.76%
	Bayesian Optimisation	11.813	0.0028	84.66%	0.95%

6.5.1.4 Overall

RBF-SVMs are effective for face recognition (Chittora and Mishra (2012*a*); Wang and Wu (2022); Dobilas (2022)). Systematic tuning of C and γ via grid or Bayesian search improves both accuracy and stability; Bayesian Optimisation often converged faster and delivered slightly better results.

Bridge to Chapter 7. While attractiveness performed well as a standalone feature, its integration with Eigenfaces did not *consistently* improve recognition across all settings; the detailed analysis and comparisons with attribute-only configurations are presented in Chapter 7.

6.6 Conclusion

This chapter explored the integration of facial attractiveness within a classical Eigenfaces + SVM pipeline on LFW. SVM outperformed Random Forest and KNN under matched optimisation. Careful tuning of C and γ (via grid and Bayesian search) was essential. The evaluation further shows that *attribute-only* models—using soft-biometric features such as attractiveness, age, gender, and skin colour—achieved higher performance than systems that concatenated those attributes with Eigenfaces, likely due to dilution of discriminative signals when merged with dimensionality-reduced PCA outputs. This highlights the strength of soft biometrics as standalone inputs and motivates considering them as a primary configuration in future pipelines.

Chapter 7

Results and Analysis

This chapter presents the principal experimental results from the face-recognition pipeline, showing how facial attractiveness, introduced as a soft-biometric trait, contributes to improved recognition accuracy. The chapter begins by outlining the procedure for constructing the data matrix from pre-processed facial images. It then describes the generation of Eigenfaces via Principal Component Analysis (PCA) before examining the effect of incorporating facial attractiveness into the recognition process.

Comparative evaluations of optimisation techniques (Grid Search vs. Bayesian Optimisation) and scoring methods (ELO scores vs. Rank) are also presented. Finally, the role of attractiveness is considered alongside other soft-biometric attributes—such as age, gender, and skin colour—followed by a detailed presentation and interpretation of the findings.

7.1 Data Matrix Creation

The first step in preparing the dataset for PCA-based face recognition involves transforming the input images into a format suitable for eigenanalysis. This process comprises three stages: pre-processing of facial images, flattening the images into one-dimensional vectors, and constructing the final data matrix.

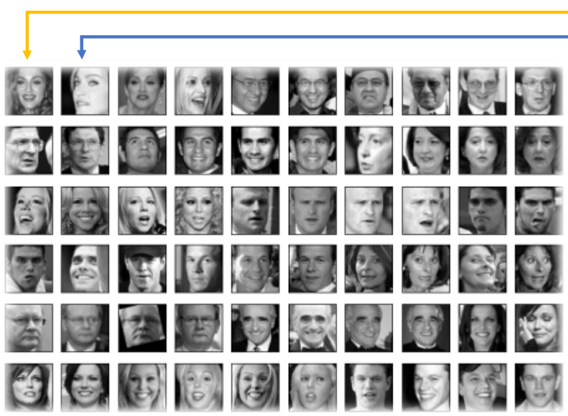
7.1.1 Data Preprocessing

Following the procedures detailed in Chapter 5, each of the M facial images was resized to $K \times K$ pixels and then converted to greyscale. As the Eigenfaces algorithm operates on greyscale rather than BGR images, each pixel intensity $[0, 255]$ was normalised to $[0, 1]$ by division by 255. Figure 7.1 illustrates the conversion to greyscale.

FIGURE 7.1: All images were converted to greyscale (Alnamnakani et al. (2024)).

7.1.2 Flattening into Vectors

Once in greyscale, each image was transformed into a one-dimensional feature vector to enable PCA-based analysis. Specifically, every $K \times K$ image was reshaped into a vector of length K^2 by concatenating its rows. Figure 7.2 demonstrates this process, in which each flattened row vector represents a single image.



1	0.99608	0.99608	0.99608	0.99608	0.99608	1	1	0.99608
0.25098	0.23529	0.20784	0.2	0.23137	0.37255	0.51373	0.29412	0.1098
0.28627	0.28627	0.28627	0.2902	0.2902	0.29412	0.29804	0.29804	0.29412
0.65882	0.63137	0.62745	0.61176	0.61961	0.61176	0.58431	0.57255	0.5451
0.058824	0.054902	0.047059	0.054902	0.05098	0.070588	0.058824	0.098039	0.16863
0.81569	0.81569	0.81569	0.80392	0.77647	0.73333	0.7098	0.70196	0.67451
0.83529	0.83529	0.83529	0.83529	0.83922	0.83529	0.82353	0.80392	0.83529
0.46275	0.30588	0.21569	0.18431	0.17255	0.14118	0.11765	0.1098	0.13725
0.63922	0.65882	0.62745	0.35686	0.12941	0.10196	0.086275	0.07451	0.07451

FIGURE 7.2: Flattening images into vectors of size K^2, then stacking them into a data matrix (Alnamnakani et al. (2024)).

7.1.3 Constructing the Data Matrix

The flattened vectors were stacked to create an $M \times K^2$ data matrix, where M is the total number of images. In this matrix, each row corresponds to one image and each column to a specific pixel position across all images. This standardised format enables the application of PCA in subsequent stages. Figure 7.3 illustrates the organisation of the data matrix.

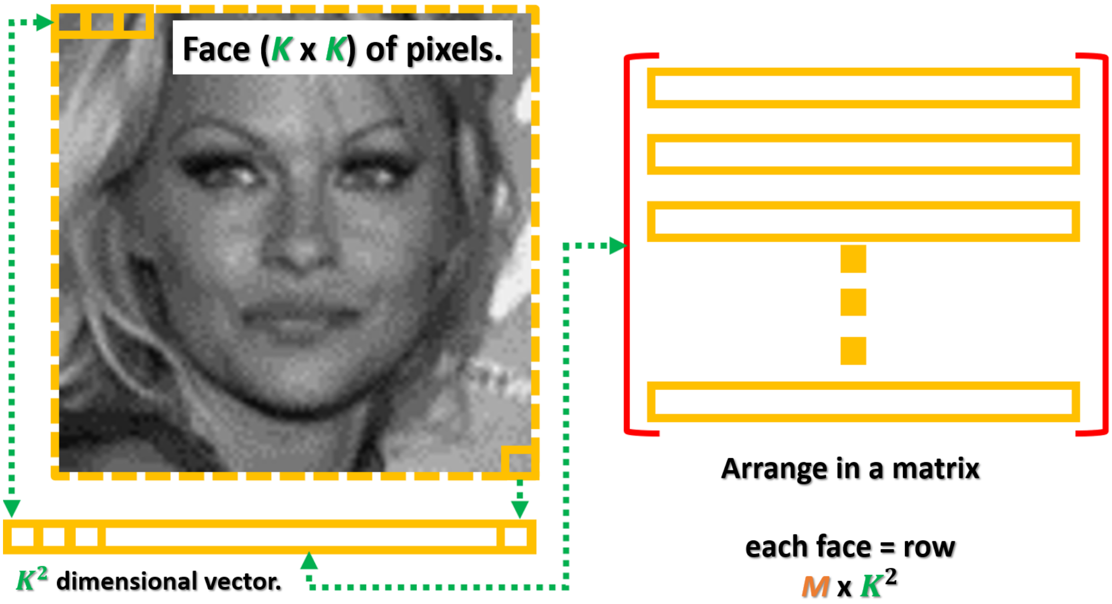


FIGURE 7.3: To apply PCA for face recognition, each face image must be represented as a vector of pixel values. These images are then arranged in a data matrix, with each row representing a photo (Alnamnakani et al. (2024)).

7.2 Eigenfaces for Data Representation

This section considers how facial attractiveness can be integrated into the PCA-based Eigenface representation to enhance recognition accuracy, and investigates its relationship with specific principal components.

7.2.1 PCA and Eigenfaces

A baseline system was first established using PCA to extract Eigenfaces, followed by classification using an RBF SVM, as described in Chapter 6. The input comprised the flattened image vectors produced in Section 7.1, assembled into an $M \times K^2$ data matrix. PCA was then applied to decompose the data into principal components, with the resulting eigenvectors reshaped into images (Eigenfaces). Each Eigenface corresponds to a principal component and reflects prominent facial variation, as illustrated in Figure 7.4.

7.2.2 Integrating Attractiveness

Although PCA-based Eigenfaces are effective for dimensionality reduction, their discriminative power can be enhanced by incorporating soft-biometric attributes. This subsection evaluates the effect of appending an attractiveness score to the Eigenface representation, in combination with age, gender, and skin colour.

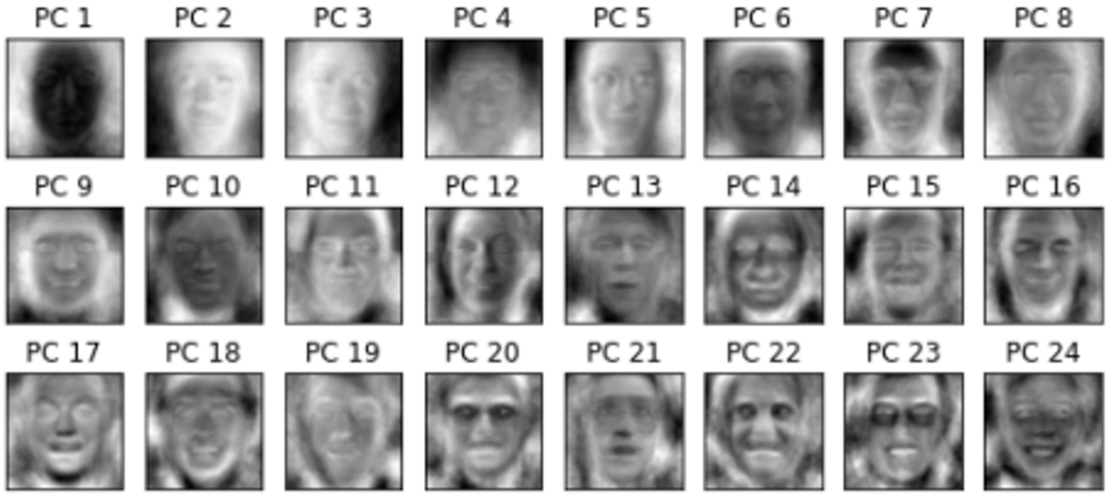


FIGURE 7.4: Examples of Eigenfaces from the LFW dataset (Alnamnakani et al. (2024)).

7.2.2.1 Attractiveness in the Eigenface Vector

After PCA extraction, each image was represented as a vector of principal components. An attractiveness score was appended to each vector to enrich the feature space:

$$(PC_1, PC_2, \ldots, PC_n, \text{Attractiveness}).$$

Figures 7.5 and 7.6 provide conceptual illustrations of how this additional feature complements the principal components. The first few components—particularly PC_1—capture the largest share of variance in the original data, while subsequent components contribute progressively less. The appended attractiveness score introduces an independent source of discriminative information.

All appended soft-biometric features (including attractiveness) were z-score normalised (zero mean, unit variance) before concatenation with the PCA components to ensure comparable scaling under the RBF kernel.

7.2.2.2 Correlation with Principal Components

To assess the relevance of attractiveness within the PCA space, Pearson's correlation coefficients were computed between attractiveness scores and each principal component. The objective was to identify components most associated with perceived attractiveness. The results indicated that attractiveness shares meaningful variance with a subset of components, providing complementary information to the PCA features. Figure 7.7 displays Eigenfaces with stronger correlation to attractiveness (top row) and weaker association (bottom row).

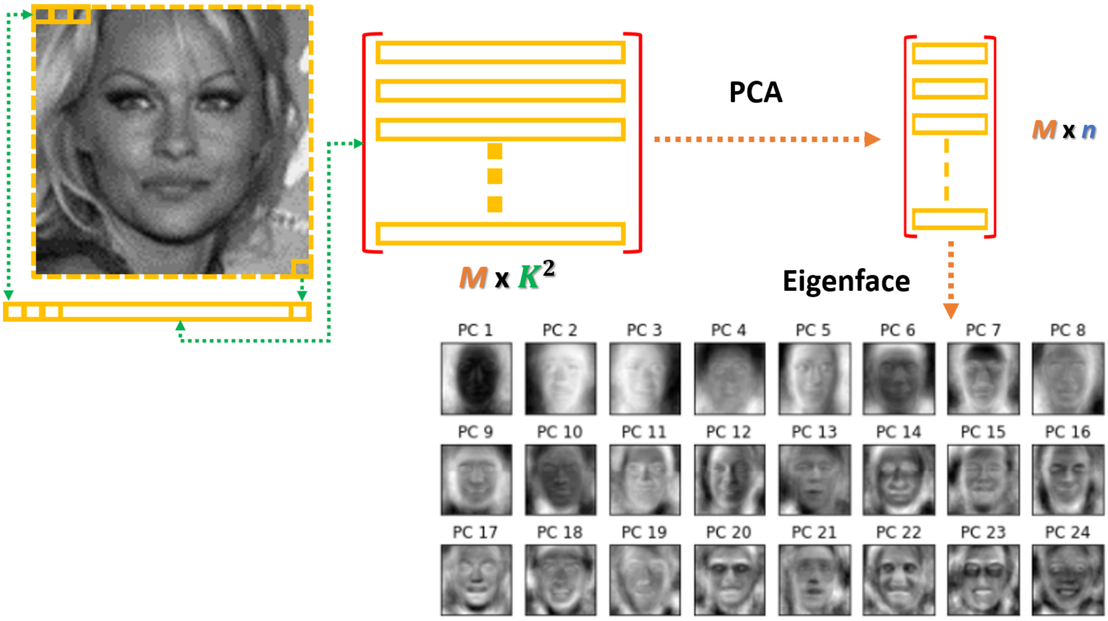


FIGURE 7.5: Eigenface 1 (PC1) captures the largest variance in the dataset. Subsequent components (PC2, PC3, etc.) capture progressively less.

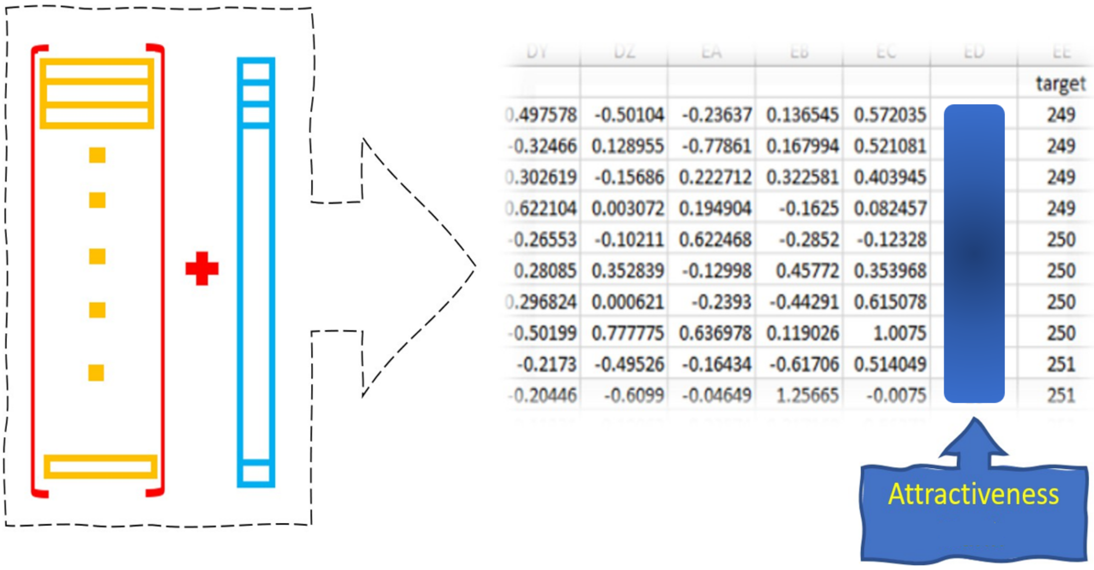


FIGURE 7.6: A conceptual illustration of how the attractiveness feature merges with the Eigenface representation.

7.3 Analysis of Results and the Importance of Attractiveness

This section evaluates the empirical impact of integrating attractiveness into PCA-based face recognition. The dataset comprised 430 individuals, each with a minimum of four images (4,964 images in total). Each image was resized to 100×100 pixels and flattened into a 10,000-dimensional vector. Dimensionality was then reduced using PCA, retaining 190 components. Attractiveness was appended as an additional feature to examine its discriminative value.

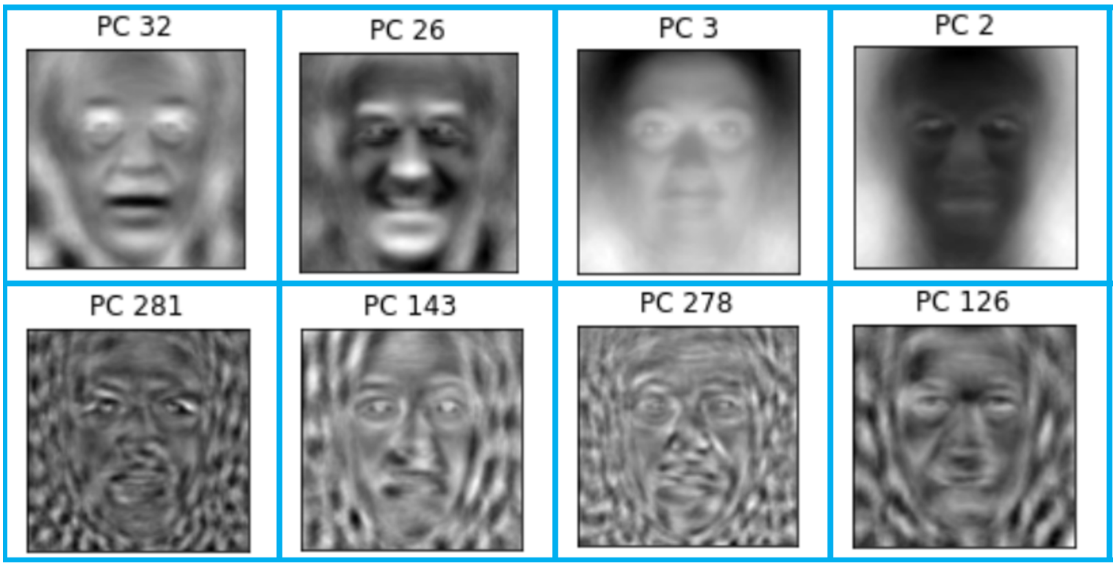


FIGURE 7.7: Eigenfaces with higher correlation to attractiveness are shown at the top; those with lower correlation appear at the bottom (Alnamnakani et al. (2024)).

7.3.1 Comparisons Across Different Dataset Sizes

To assess the influence of attractiveness systematically, two dataset sizes were evaluated using an RBF-kernel Support Vector Machine (SVM):

- **Full set:** 430 individuals (4,964 images).
- **Smaller set:** 100 individuals (1,089 images).

In both cases, two configurations were compared: PCA-only features versus PCA features with attractiveness appended.

7.3.1.1 Results with 430 Individuals (4,964 Images)

The inclusion of attractiveness improved classification performance. Accuracy increased from 34.8% to 43.2%, while the F1-score rose from 25.3% to 32.7%. These improvements of +8.4 and +7.4 percentage points, respectively, were statistically significant (paired t-test, $p = 0.015$). Table 7.1 summarises these results.

7.3.1.2 Results with 100 Individuals (1,089 Images)

A similar trend was observed with the smaller dataset. Accuracy increased from 26.2% to 33.3%, and the F1-score improved from 19.2% to 24.5%. These gains of +7.1 and +5.3 percentage points were also statistically significant ($p = 0.003$), confirming that attractiveness contributes positively even with fewer participants. Table 7.1 summarises these results.

TABLE 7.1: Performance comparison: PCA-only vs. PCA+Attractiveness across two dataset sizes.

Data Set Description	Feature	Mean Accuracy	Mean F1-Score	Accuracy Improvement	F1-Score Improvement	T-test Statistic	P-value
Data Set with 430 Individuals	PCA Only	34.8%	25.3%				
	PCA with Attractiveness	43.2%	32.7%	+8.4 percentage points	+7.4 percentage points	-8.172	0.015
Data Set with 100 Individuals	PCA Only	26.2%	19.2%				
	PCA with Attractiveness	33.3%	24.5%	+7.1 percentage points	+5.3 percentage points	-17.718	0.003

7.3.1.3 Overall Interpretation

Across both dataset sizes, attractiveness consistently yielded improvements in accuracy and F1-score (typically 5–8%). This consistency indicates that attractiveness captures facial cues not fully represented by PCA alone. Figure 7.8 visualises these differences. All improvements were statistically significant ($p < 0.05$), reinforcing the conclusion that attractiveness provides valuable supplementary information for facial recognition.

7.3.1.4 Analysis of Adding Attractiveness to Different Data Subsets

To examine robustness, the dataset was partitioned into three non-overlapping groups. Accuracy was compared for each group with and without attractiveness. As shown in Table 7.2, accuracy improvements ranging from +5.5% to +8.4% were observed, supporting the stability of attractiveness as a discriminative trait across subject subsets.

7.3.2 Comparison of Attractiveness with Traditional Face Attributes

Two dataset sizes were considered (100 and 430 participants). For each, performance with PCA-only features was compared against PCA combined with one of: Age,

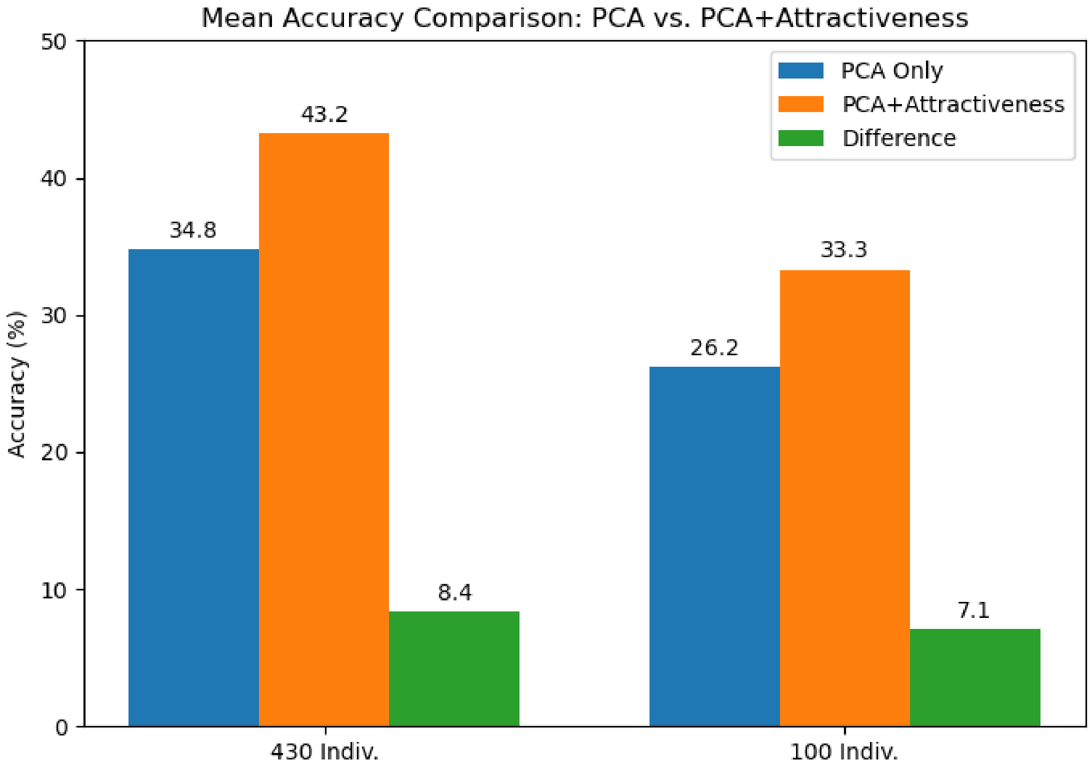


FIGURE 7.8: Comparison showing how attractiveness improves accuracy and F1-scores in the 430- and 100-participant datasets.

TABLE 7.2: Accuracy comparison: PCA-only vs. PCA+Attractiveness across different groups.

Group	PCA Only Acc.	Attractiveness Acc.	Improvement
1 (IDs 1–100)	0.3480	0.4316	+0.0836 (+8.36%)
2 (IDs 331–430)	0.3063	0.3836	+0.0773 (+7.73%)
3 (IDs 249–348)	0.1709	0.2259	+0.0550 (+5.50%)

Attractiveness, Gender, or Skin Colour. Figure 7.9 presents grouped bar charts, where dark bars indicate 430 participants and light bars 100 participants.

Among all attributes, Age achieved the highest overall accuracy (up to 34.7%), followed closely by Attractiveness, which increased accuracy from 26.2% (baseline) to 33.3% and improved the F1-score from 19.2% to 24.5%. Gender and Skin Colour also provided gains over the PCA baseline, albeit smaller. These results suggest that Age and Attractiveness are the most effective soft-biometric features for enhancing PCA-based recognition.

Notably, Attractiveness improved performance consistently across both datasets, reinforcing its reliability as a complementary and generalisable attribute. Moreover, combining Age, Gender, and Skin Colour with PCA yielded further gains, which

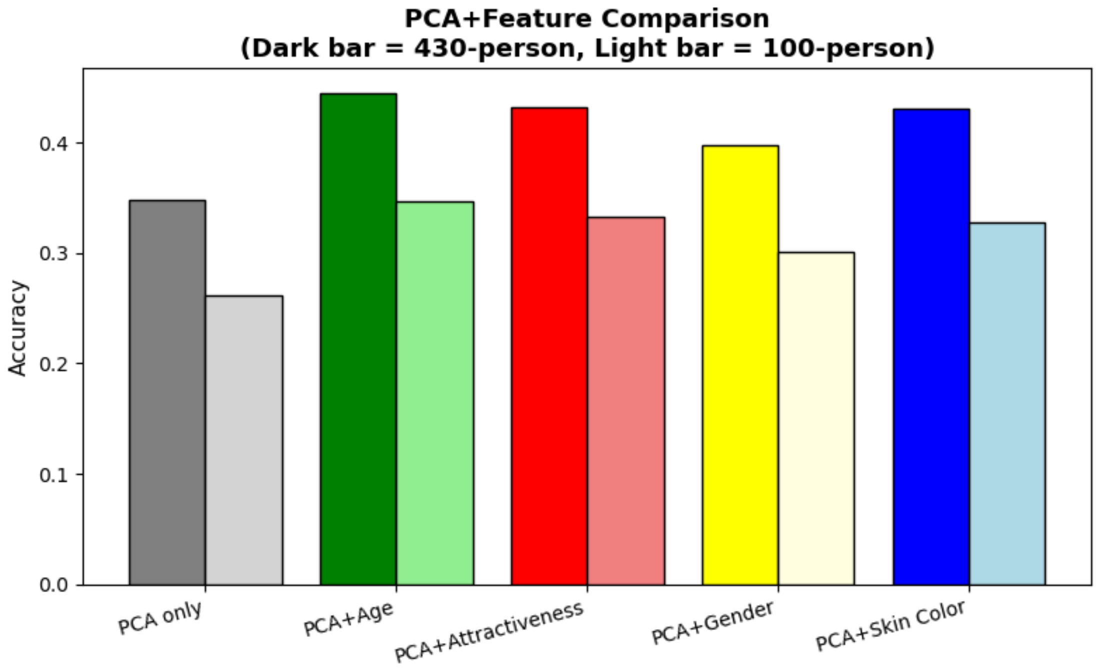


FIGURE 7.9: Accuracy for five settings: PCA, +Age, +Attractiveness, +Gender, +Skin Colour—across two dataset sizes. Dark bars: 430 participants; light bars: 100 participants.

were amplified by including Attractiveness—highlighting its distinct discriminative contribution.

7.3.3 ELO vs. Rank; Grid Search vs. Bayesian Optimisation

TABLE 7.3: Mean accuracy and standard deviation for five features using ELO Scores or Rank, with Grid Search or Bayesian Optimisation.

		ELO Scores		Rank	
	Comparison of Features: Feature	Mean Accuracy	Standard Deviation	Mean Accuracy	Standard Deviation
Grid Search	0 PCA only	0.261684	0.004364	0.261684	0.004364
	1 Age	0.346695	0.007734	0.334609	0.002340
	2 Attractiveness	0.332593	0.006156	0.327155	0.002832
	3 Gender	0.301370	0.001254	0.333804	0.001235
	4 Skin Color	0.327962	0.001118	0.326954	0.000812
Bayesian Optimisation	0 PCA only	0.261884	0.004908	0.261884	0.004908
	1 Age	0.347299	0.007111	0.333601	0.002057
	2 Attractiveness	0.332593	0.008088	0.325946	0.002626
	3 Gender	0.303182	0.002905	0.334206	0.002281
	4 Skin Color	0.326350	0.003055	0.326148	0.001439

Table 7.3 and Figure 7.10 compare two scoring strategies—ELO Scores and Rank—combined with two hyperparameter-tuning methods: Grid Search and Bayesian Optimisation. Across five feature sets (PCA only; PCA + Age; PCA + Attractiveness; PCA + Gender; PCA + Skin Colour), Age consistently achieved the highest accuracy, followed by Attractiveness. While Gender and Skin Colour also improved upon the PCA-only baseline, their impact was more modest. Overall, ELO

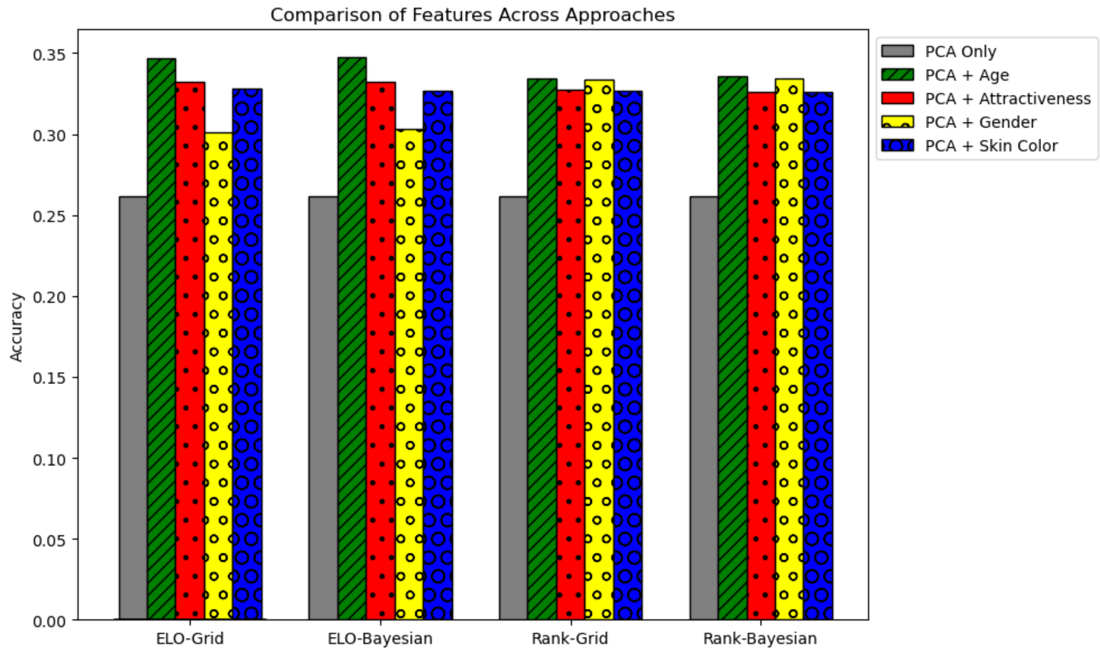


FIGURE 7.10: Accuracy across four approaches (Grid/Bayesian under ELO Scores/Rank) for the five features. Age consistently achieved the highest performance, followed by Attractiveness. ELO Scores generally outperformed Rank, while Bayesian Optimisation gave slight gains over Grid Search.

Scores outperformed Rank irrespective of the optimiser, and Bayesian Optimisation provided small but consistent gains over Grid Search.

Additional experiments on three independent 100-participant subsets showed the same pattern: Age and Attractiveness produced the largest improvements, particularly when ELO Scores were paired with Bayesian Optimisation.

7.3.4 Traditional Face Attributes vs. Traditional Face Attributes + Attractiveness

Traditional soft-biometric attributes—Age, Gender, and Skin Colour—substantially improve performance when appended to PCA features. This improvement is further amplified when Attractiveness is also included. Three configurations were compared:

1. PCA only.
2. PCA + Age + Gender + Skin Colour (AGS).
3. PCA + AGS + Attractiveness.

7.3.4.1 Organised Results

Table 7.4 summarises the mean accuracy and standard deviation for each configuration. The first comparison contrasts PCA-only with PCA + AGS, while the second assesses the added value of including Attractiveness alongside AGS.

TABLE 7.4: Summary of feature-set evaluations
AGS = (Age, Gender, Skin Colour)

Feature Set	Mean Accuracy	Standard Deviation
PCA only	0.262	0.004
PCA + AGS	0.446	0.000
PCA + AGS + Attractiveness	0.480	0.001

7.3.4.2 Visual Representation of the Results

Figure 7.11 illustrates the gains reported in Table 7.4. While AGS substantially improved accuracy over the PCA baseline, the inclusion of Attractiveness added a further gain of more than three percentage points.

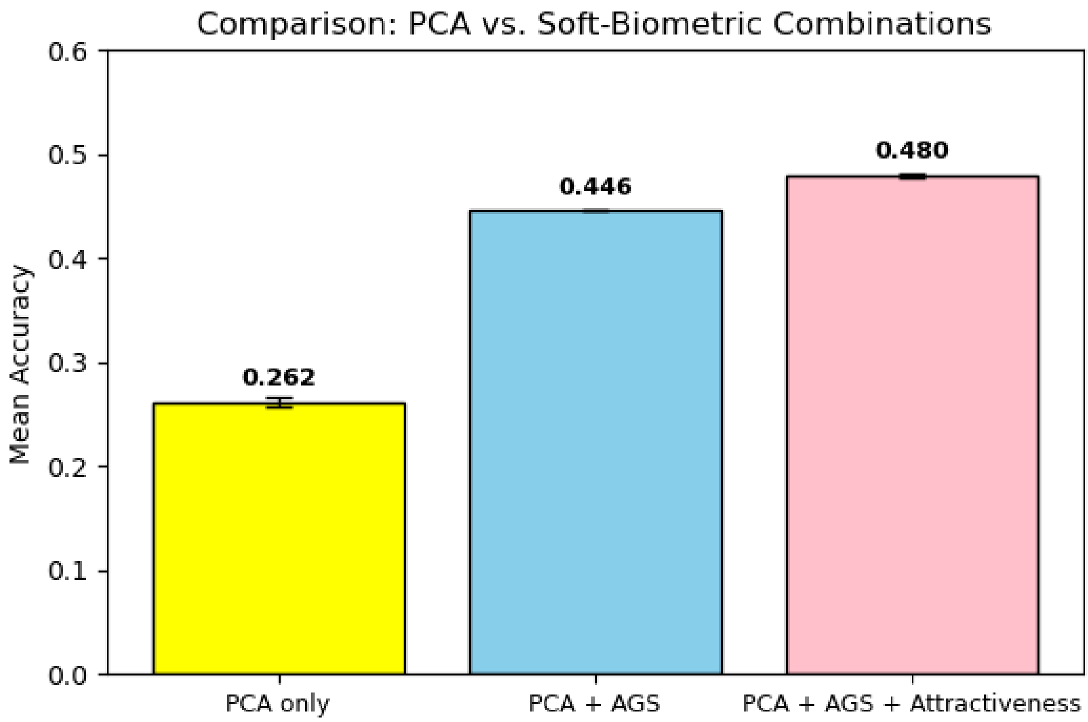


FIGURE 7.11: Mean accuracy for PCA only, PCA + AGS, and PCA + AGS + Attractiveness. Accuracy rises from 26.2% to 44.6% with AGS, and further to 48.0% when Attractiveness is added.

7.3.4.3 Analysis of Attractiveness

The improvement from 26.2% to 44.6% using AGS confirms the effectiveness of traditional attributes. Incorporating Attractiveness increased accuracy to 48.0%, indicating that it captures facial information beyond Age, Gender, and Skin Colour. The overall gain remained statistically significant ($p < 0.05$).

7.3.5 Fisherfaces vs. Eigenfaces on LFW

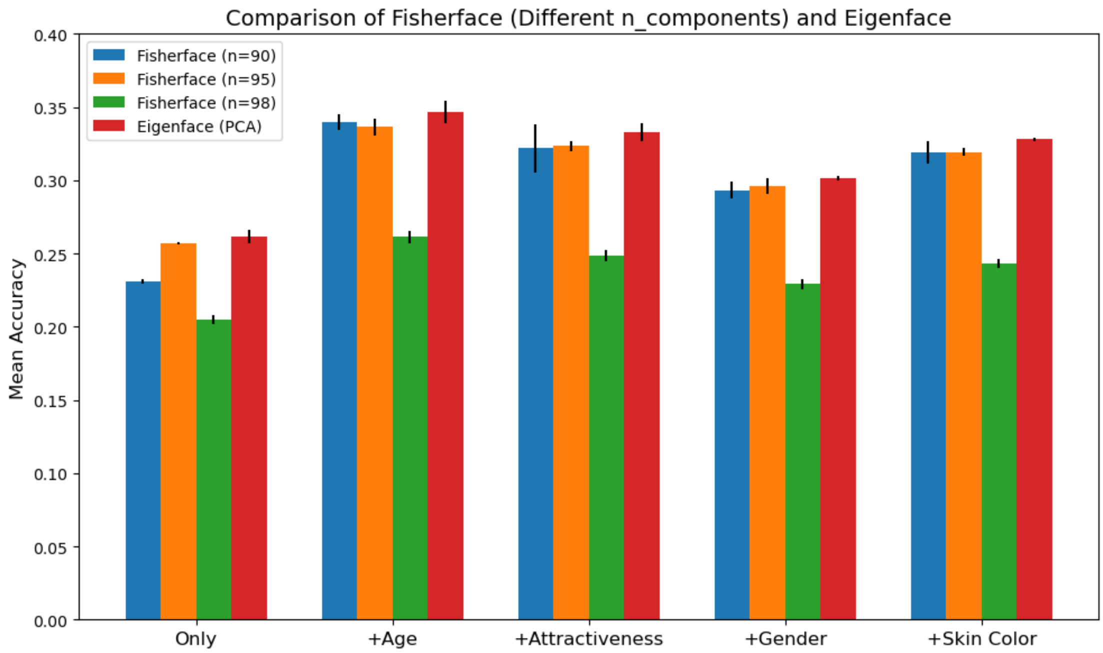


FIGURE 7.12: Comparison of Fisherface and Eigenface on LFW, highlighting the impact of attributes such as Attractiveness.

Figure 7.12 compares Fisherface (PCA followed by Linear Discriminant Analysis) and Eigenface (PCA only) on the LFW dataset. For Fisherface, three PCA variance thresholds were tested (90%, 95%, and 98%), whereas Eigenface was evaluated using a single configuration retaining 95% of the variance.

Without additional attributes, Eigenface slightly outperformed Fisherface across configurations. However, adding soft-biometric attributes—particularly Age and Attractiveness—led to substantial improvements in both methods. Attractiveness had a marked positive impact in both settings, reinforcing its value as a soft-biometric trait.

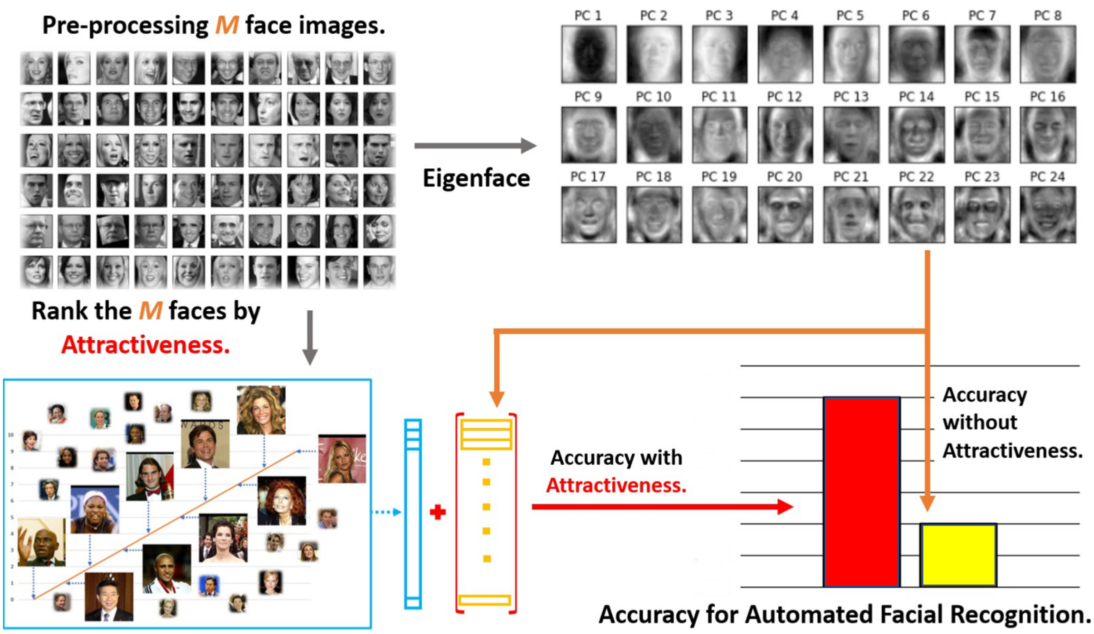


FIGURE 7.13: Adding facial Attractiveness to Eigenfaces and other traditional soft-biometric attributes (e.g. Age, Gender) increases recognition accuracy over PCA alone (Alnamnakani et al. (2024)).

7.4 Conclusion

This chapter showed that incorporating Attractiveness as a soft-biometric trait consistently improves face-recognition accuracy, particularly when combined with PCA-based Eigenfaces and other conventional attributes such as Age, Gender, and Skin Colour (Figures 7.13 and 7.14). On average, accuracy increased by approximately 5–8% across datasets, with all improvements statistically significant ($p < 0.05$). While Age remained the most influential attribute, Attractiveness consistently ranked second, outperforming Gender and Skin Colour.

Note. The results above quantify the marginal benefit of adding Attractiveness *within* a PCA/Eigenfaces pipeline. Our broader evaluation (Chapter 6) shows that attribute-only models (e.g. Age, Gender, Skin Colour, and Attractiveness) achieve higher overall accuracy than their Eigenfaces-augmented counterparts when using the same feature sets.

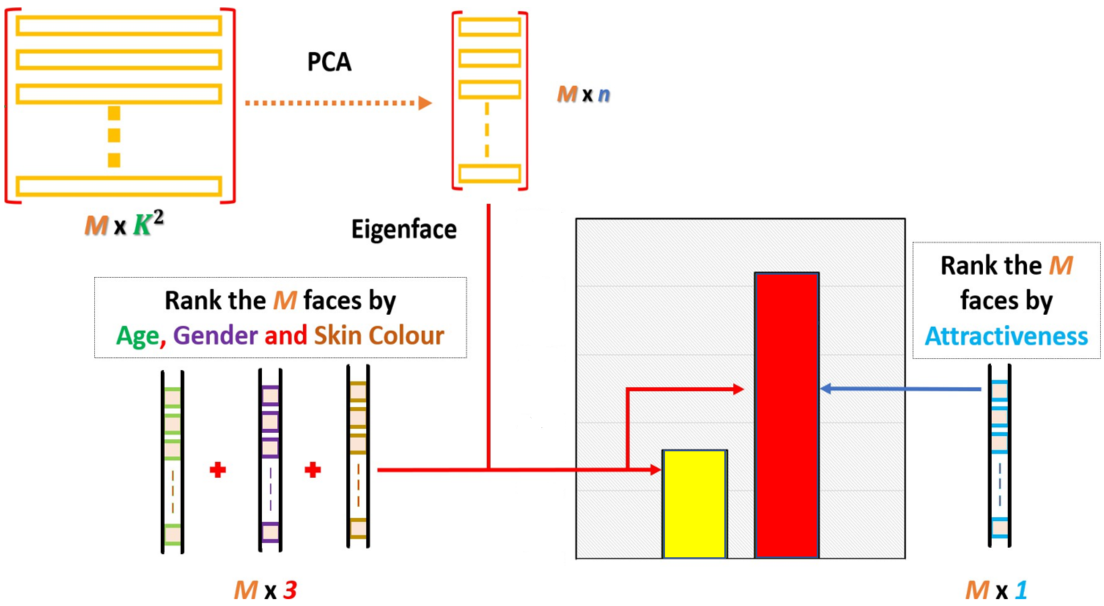


FIGURE 7.14: Adding Attractiveness to Eigenfaces and other soft-biometric attributes (Age, Gender, Skin Colour) yields further accuracy gains (Alnamnakani et al. (2024)).

Chapter 8

Conclusions and Future Work

This chapter provides a concise overview of the study's findings on using **facial attractiveness** as a soft biometric feature to enhance face recognition performance. It also presents future perspectives and promising research directions in this domain.

8.1 Conclusion

This thesis introduces facial attractiveness as a new soft biometric feature within automated face recognition systems, supplementing commonly used characteristics such as age, gender, and skin colour. The key conclusions are outlined below:

1. Correlation analyses and feature-selection criteria (e.g., Mutual Information, SFFS) confirm that attractiveness is a highly discriminative trait—comparable to age and gender—for distinguishing individuals.
2. Despite being a relatively new concept, facial attractiveness as a soft biometric feature can be incorporated into existing recognition pipelines with minimal adjustment. Attractiveness values were derived from pairwise face comparisons and converted into a ranking (e.g., ELO). This flexible and computationally efficient method captures aesthetic aspects known to influence human memory and first impressions.
3. The LFW database was used with two cohorts: one of 100 individuals and another of 430 individuals, each with at least four images per person. Across both configurations, age and attractiveness were the principal contributors to improved recognition accuracy, followed by gender and skin colour.
4. Across both four-fold and leave-one-out cross-validation protocols, adding attractiveness to PCA features with an RBF SVM improved accuracy by about

5–8 percentage points; paired t-tests confirmed statistical significance ($p < 0.05$) (see §7.3, Table 7.1).

5. While adding attractiveness to PCA/Eigenfaces improved over a PCA-only baseline, the broader evaluation showed that attribute-only configurations (e.g., Age, Gender, Skin Colour, and attractiveness) achieved higher overall accuracy than Eigenfaces-augmented systems using the same features (see §6.6 and §7.4).

6. Using the same attribute set (22 features, including attractiveness) in an attribute-only Random Forest, an accuracy of 87.00% was attained, exceeding the same model without attractiveness by 2.25 percentage points, the k-NN approach by about 15 percentage points, and the 24-attribute method by nearly 30 percentage points.

7. Although the inclusion of facial attractiveness improves recognition, ethical considerations arise regarding beauty standards, privacy, and informed consent. Any operational deployment should include fairness audits, robust consent mechanisms, and culturally aware annotation protocols to mitigate bias and potential harms associated with standards of beauty.

8.2 Additional Perspectives and Next Steps

This study focused on the standard LFW dataset and the classical Eigenfaces algorithm to examine how facial attractiveness can be integrated into automated recognition. This choice provided a clear testbed to quantify the marginal value of attractiveness and, in turn, motivates advanced research utilising deep neural networks designed to detect and leverage attractiveness cues for potentially greater accuracy.

An additional observation concerns eyewitness testimony: attractiveness, as reported by witnesses, can operate as a memorable and salient descriptor. A specialised deep-learning system that models observer-based attractiveness may strengthen the link between subjective descriptions and objective identification, especially in investigative contexts.

Embedding attractiveness within human-centric attributes represents a step towards more comprehensive recognition. With continued research, attractiveness may become a routine factor in eyewitness accounts. More detailed models, larger and more diverse datasets, and feedback from real-world witnesses are expected to increase reliability.

In sum, the study indicates that facial attractiveness is a measurable and significant soft-biometric feature that can meaningfully improve face-recognition performance. These results open the door to further research and real-world applications in which attractiveness is combined with other traits or integrated into deep-learning-based recognition systems.

8.3 Future Work

This study lays the groundwork for further exploration of how facial attractiveness can improve recognition performance in both technical and practical settings. The following avenues are proposed:

1. **Deep Embeddings and Saliency Maps (e.g., Grad-CAM).** Beyond Eigenfaces/SVM baselines, adopting deep neural networks (CNNs or Transformers) may increase accuracy and flexibility. By examining activation maps and saliency heat maps (e.g., Grad-CAM), the facial regions most associated with attractiveness can be identified and linked to previously studied traits.

2. **Increasing and Diversifying Annotators.** A larger, culturally and socially diverse annotator pool is recommended to reduce bias, enrich the dataset, and reveal whether consistent group-wise differences in attractiveness judgements exist.

3. **Trying Different Ranking Systems and Adjusting Parameters.** Although ELO worked well, alternative systems (e.g., TrueSkill, Glicko, Bayesian Ranking) merit comparison. A supervised baseline (e.g., RankSVM) can test whether labelled approaches sometimes outperform iterative, label-free methods.

4. **Testing Other Representations and Distance Measures (e.g., L1/Manhattan, cosine).** Beyond Eigenfaces and Euclidean distance, Fisherfaces (LDA-based) and other dimensionality-reduction techniques, together with alternative distances, may render the inclusion of attractiveness more robust across pipelines.

5. **Aligning With Witness Descriptions and Forensic Work.** Eyewitness reports that describe suspects as "attractive" could be translated into semi-quantitative data for forensic databases. Linking deep-learning saliency with these observations may align human judgements with algorithmic evidence in legal contexts.

6. **Combining Attractiveness with Other Soft Biometrics.** Beyond age, gender, and skin colour, links to other modalities (e.g., body attractiveness, voice, clothing) could be explored. A multimodal view may improve reliability in large-scale or unconstrained scenarios.

Overall, incorporating attractiveness appears to enhance face-recognition systems for security, forensic, and commercial applications. As automated methods and large-scale data grow, exploring how attractiveness interacts with deeper models and diverse cultural perspectives may enable more human-centred and innovative solutions in visual recognition and AI.

References

Abdi, H. and Williams, L. J. (2010), 'Principal component analysis', *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(4), 433–459.

Abdullah, M. B. (1990), 'On a robust correlation coefficient', *Journal of the Royal Statistical Society: Series D (The Statistician)* **39**(4), 455–460.

Adjabi, I., Ouahabi, A., Benzaoui, A. and Taleb-Ahmed, A. (2020), 'Past, present, and future of face recognition: A review', *Electronics* **9**(8), 1188.

Ali, M. and Gaber, T. (2023), 'Efficient thermal face recognition method using optimized curvelet features for biometric authentication', *PLoS One* **18**(6), e0287349.

Almudhahka, N., Nixon, M. and Hare, J. (2017), 'Semantic face signatures: Recognizing and retrieving faces by verbal descriptions', *IEEE Transactions on Information Forensics and Security* **13**(3), 706–716.

Almudhahka, N. Y., Nixon, M. S. and Hare, J. S. (2016), Unconstrained human identification using comparative facial soft biometrics, *in* '2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)', IEEE, pp. 1–6.

Alnamnakani, M., Mahmoodi, S. and Nixon, M. (2019), On the potential for facial attractiveness as a soft biometric, *in* 'International Symposium on Visual Computing', Springer, pp. 516–528.

Alnamnakani, M., Mahmoodi, S. and Nixon, M. (2024), Using facial attractiveness as a soft biometric trait to enhance face recognition performance, *in* T. Bourlai, ed., 'Face Recognition Across the Imaging Spectrum', Springer Nature Singapore, Singapore, pp. 31–55. Chapter in an edited volume.

Altwaijry, H. and Belongie, S. (2013), Relative ranking of facial attractiveness, *in* '2013 IEEE Workshop on Applications of Computer Vision (WACV)', IEEE, pp. 117–124.

Ammar, S., Bouwmans, T., Zaghden, N. and Neji, M. (2020), Towards an effective approach for face recognition with DCGANs data augmentation, *in* 'Advances in Visual Computing: 15th International Symposium, ISVC 2020', Springer, pp. 463–475. Check exact details if needed.

Amos, B., Ludwiczuk, B., Satyanarayanan, M. et al. (2016), 'Openface: A general-purpose face recognition library with mobile applications', *CMU School of Computer Science* **6**(2), 20.

Anggo, M. and Arapu, L. (2018), Face recognition using fisherface method, *in* 'Journal of Physics: Conference Series', Vol. 1028, IOP Publishing, p. 012119.

Arigbabu, O. A., Ahmad, S. M. S., Adnan, W. A. W. and Yussof, S. (2015), 'Recent advances in facial soft biometrics', *The Visual Computer* **31**(5), 513–525.

Arpad Elo (n.d.), https://en.wikipedia.org/wiki/Arpad_Elo#:~:text=Arpad%20Emmerich%20Elo%20(n%C3%A9%20%C3%891%C5%91,player%20games%20such%20as%20chess.&text=Brookfield%2C%20Wisconsin%2C%20U.S.&text=Born%20in%20Egyh%C3%A1zaskesz%C5%91%2C%20Kingdom%20of,for%20terms%201935%20and%201936. [Online; accessed 18-February-2025].

Asthana, A., Zafeiriou, S., Tzimiropoulos, G., Cheng, S. and Pantic, M. (2014), 'From pixels to response maps: Discriminative image filtering for face alignment in the wild', *IEEE transactions on pattern analysis and machine intelligence* **37**(6), 1312–1320.

BayesianOptimization (2014), 'Github repository'.
URL: *https://github.com/bayesian-optimization/BayesianOptimization*

Belhumeur, P. N., Hespanha, J. P. and Kriegman, D. J. (1997), 'Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 711–720.

Ben Fredj, H., Bouguezzi, S. and Souani, C. (2021), 'Face recognition in unconstrained environments with cnn', *The Visual Computer* **37**(2), 217–226.

Benesty, J., Chen, J., Huang, Y. and Cohen, I. (2009), Pearson correlation coefficient, *in* 'Noise Reduction in Speech Processing', Springer, Berlin, Heidelberg, pp. 1–4.

Bergstra, J. and Bengio, Y. (2012), 'Random search for hyper-parameter optimization', *Journal of Machine Learning Research* **13**, 281–305.

Berle, I. and Berle, I. (2020), *What Is Face Recognition Technology?*, Springer.

Beyer, K., Goldstein, J., Ramakrishnan, R. and Shaft, U. (1999), When is "nearest neighbor" meaningful?, *in* 'International Conference on Database Theory', Springer, pp. 217–235.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.

Bottino, A. and Laurentini, A. (2010), The analysis of facial beauty: An emerging area of research in pattern analysis, *in* 'International Conference on Image Analysis and Recognition', Springer, pp. 425–435.

Bougourzi, F., Dornaika, F., Barrena, N., Distante, C. and Taleb-Ahmed, A. (2022), 'Cnn-based facial aesthetics analysis through dynamic robust losses and ensemble regression', *Applied Intelligence* pp. 1–18.

Bougourzi, F., Dornaika, F. and Taleb-Ahmed, A. (2022), 'Deep learning based face beauty prediction via dynamic robust losses and ensemble regression', *Knowledge-Based Systems* **242**, 108246.

Bradski, G. and Kaehler, A. (2008), *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media, Inc.

Brahnam, S. and Jain, L. C. (2010), *Advanced computational intelligence paradigms in healthcare 5: intelligent decision support systems*, Vol. 5, Springer Science & Business Media.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**, 5–32.

Brunton, S. L. and Kutz, J. N. (2022), *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge University Press, Cambridge.

Buolamwini, J. and Gebru, T. (2018), Gender shades: Intersectional accuracy disparities in commercial gender classification, *in* 'Conference on Fairness, Accountability and Transparency', Proceedings of Machine Learning Research (PMLR), pp. 77–91.

Burges, C. J. C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N. and Hullender, G. (2005), Learning to rank for information retrieval, *in* 'Proceedings of the 22nd International Conference on Machine Learning (ICML)', ACM, pp. 89–96.

Çarıkçı, M. U. and Özen, F. (2012), 'A face recognition system based on eigenfaces method', *Procedia Technology* **1**, 118–123.

Chellappa, R., Phillips, P. J. and Rosenfeld, A. (2001), 'Face recognition: A literature survey', *ACM Computing Surveys* **33**(4), 399–458.

Chen, C. S., Lin, C. Y. and Huang, Y. H. (2014), Cross-age reference coding for age-invariant face recognition and retrieval, *in* 'European Conference on Computer Vision (ECCV)', Springer, Zürich, Switzerland, pp. 768–783.

Chen, D., Cao, X., Wen, F. and Sun, J. (2013), Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 3025–3032.

Chen, F. and Zhang, D. (2010), A benchmark for geometric facial beauty study, *in* 'International conference on medical biometrics', Springer, pp. 21–32.

Chihaoui, M., Elkefi, A., Bellil, W. and Ben Amar, C. (2016), 'A survey of 2d face recognition techniques', *Computers* **5**(4), 21.

Chittora, A. and Mishra, O. (2012*a*), 'Face recognition using rbf kernel based support vector machine', *International Journal of Future Computer and Communication* **1**(3), 280.

Chittora, R. and Mishra, S. (2012*b*), 'Face recognition using pca and svm with accuracy enhancement', *Int. J. Adv. Res. Comp. Sci. and Soft. Eng.* **2**(5).

Confidence to Deploy AI with World-Class Training Data (2025), `https://www.appen.com/`. Accessed on 2025-01-08.

Cortes, C. and Vapnik, V. (1995), 'Support-vector networks', *Machine Learning* **20**(3), 273–297.

Cross, J., Cross, J. and Daly, J. (1971), 'Sex, race, age, and beauty as factors in recognition of faces', *Perception & Psychophysics* **10**(6), 393–396.

Cunningham, M. R. (1986), 'Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty', *Journal of Personality and Social Psychology* **50**(5), 925.

Dalisay, A. (2021), 'Using facial recognition for authentication in keyless systems', Bachelor's project, Electronics and Computer Science, Faculty of Physical Sciences and Engineering, University of Southampton.

Damon, F., Méary, D., Quinn, P. C., Lee, K., Simpson, E. A., Paukner, A., Suomi, S. J. and Pascalis, O. (2017), 'Preference for facial averageness: Evidence for a common mechanism in human and macaque infants', *Scientific Reports* **7**(1), 46303.

Dangauthier, P., Herbrich, R., Minka, T., Graepel, T. and Murphy, B. (2007), 'Trueskill through time: Revisiting the history of chess', *Advances in Neural Information Processing Systems (NIPS)* **20**, 337–344.

Dantcheva, A., Elia, P. and Ross, A. (2016), 'What else does your biometric data reveal? a survey on soft biometrics', *IEEE Transactions on Information Forensics and Security* **11**(3), 441–467.

Dantcheva, A., Velardo, C., D'Angelo, A. and Dugelay, J. (2011), 'Bag of soft biometrics for person identification: New trends and challenges', *Multimedia Tools and Applications* **51**, 739–777.

Davies, G. and Valentine, T. (2007), Facial composites: Forensic utility and psychological research, *in* 'The Handbook of Eyewitness Psychology: Volume II', Psychology Press, pp. 73–98.

DeCann, B. and Ross, A. (2013), Relating roc and cmc curves via the biometric menagerie, *in* '2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)', IEEE, pp. 1–8.

Dobilas, S. (2022), 'Svm classifier and rbf kernel — how to make better models in python'.
URL: *https://towardsdatascience.com/svm-classifier-and-rbf-kernel-how-to-make-better-models-in-python-73bb4914af5b*

Duda, R. O., Hart, P. E. and Stork, D. G. (2000), *Pattern Classification*, 2nd edn, Wiley-Interscience.

Elmer, E. M. and Houran, J. (2020), 'Physical attractiveness in the workplace: Customers do judge books by their covers', *Beauty in Business* **14**.

Elo, A. E. (1978), *The rating of chessplayers, past and present*, Arco Pub.

Etcoff, N. (1999), *Survival of the Prettiest: The Science of Beauty*, Doubleday, New York.

Fan, J., Chau, K., Wan, X., Zhai, L. and Lau, E. (2012), 'Prediction of facial attractiveness from facial proportions', *Pattern Recognition* **45**(6), 2326–2334.

Federation, U. S. C. (n.d.), 'Uscf (united states chess federation) rating system'. Available at: `https://new.uschess.org/ratings`.

FIDE (n.d.), 'World chess federation rating system'. Available at: `https://www.fide.com`.

French, L. (2015), 'Most wanted: Cameras, criminal justice and the persistence of vision'.
URL: *https://thesocietypages.org/cyborgology/2015/06/22/most-wanted-cameras-criminal-justice-and-the-persistence-of-vision/*

Garvie, C., Bedoya, A. and Frankle, J. (2016), The perpetual line-up: Unregulated police face recognition in america, Technical report, Georgetown Law, Center on Privacy & Technology.
URL: *https://www.perpetuallineup.org/*

GeeksforGeeks (2024*a*), 'Gamma parameter in svm'.
URL: *https://www.geeksforgeeks.org/gamma-parameter-in-svm/*

GeeksforGeeks (2024*b*), 'Svm hyperparameter tuning using gridsearchcv – ml'.
URL: *https://www.geeksforgeeks.org/svm-hyperparameter-tuning-using-gridsearchcv-ml/*

Geldart, S. (2010), 'That woman looks pretty, but is she attractive? female perceptions of facial beauty and the impact of cultural labels', *European review of applied psychology* **60**(2), 79–87.

Gonzalez-Sosa, E., Fierrez, J., Vera-Rodriguez, R. and Alonso-Fernandez, F. (2018), 'Facial soft biometrics for recognition in the wild: Recent works, annotation, and cots evaluation', *IEEE Transactions on Information Forensics and Security* **13**(8), 2001–2014.

Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, MIT Press, Cambridge, MA, USA.

Grammer, K., Fink, B., Møller, A. P. and Thornhill, R. (2003), 'Darwinian aesthetics: sexual selection and the biology of beauty', *Biological reviews* **78**(3), 385–407.

Grother, P., Quinn, G. W. and Phillips, P. J. (2010), Report on the evaluation of 2d still-image face recognition algorithms, Technical Report NISTIR 7709, National Institute of Standards and Technology. Offers detailed discussions on CMC curves and face recognition testing.

Gumus, E., Kilic, N., Sertbas, A. and Ucan, O. N. (2010), 'Eigenfaces and support vector machine approaches for hybrid face recognition', *Pattern Recognition* **8**, 9.

Guo, B. H., Nixon, M. S. and Carter, J. N. (2019), 'Soft biometric fusion for subject recognition at a distance', *IEEE Transactions on Biometrics, Behavior, and Identity Science* **1**(4), 292–301.

Guo, B. and Nixon, M. S. (2008), 'Gait feature subset selection by mutual information', *IEEE Transactions on Systems, MAN, and Cybernetics-part a: Systems and Humans* **39**(1), 36–46.

Guo, G. and Zhang, N. (2019*a*), 'A survey on deep learning-based face recognition', *Computer Vision and Image Understanding* **189**, 102805.

Guo, G. and Zhang, N. (2019*b*), 'A survey on deep learning based face recognition', *Computer Vision and Image Understanding* **189**, 102805.

Hassan, B., Izquierdo, E. and Piatrik, T. (2021*a*), 'Soft biometrics: a survey', *Multimedia Tools and Applications* pp. 1–44.

Hassan, B., Izquierdo, E. and Piatrik, T. (2021*b*), 'Soft biometrics: A survey: Benchmark analysis, open challenges and recommendations', *Multimedia Tools and Applications* pp. 1–44.

Hastie, T., Tibshirani, R. and Friedman, J. (2009*a*), *The Elements of Statistical Learning*, 2 edn, Springer.

Hastie, T., Tibshirani, R. and Friedman, J. (2009*b*), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2 edn, Springer.

He, H. and Garcia, E. A. (2009), 'Learning from imbalanced data', *IEEE Transactions on Knowledge and Data Engineering* **21**(9), 1263–1284.

Huang, G. B., Mattar, M., Berg, T. and Learned-Miller, E. (2008), Labeled faces in the wild: A database for studying face recognition in unconstrained environments, *in* 'Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition'.

International Organization for Standardization (2006), 'Biometric Performance Testing and Reporting – Part 1: Principles and Framework', ISO/IEC 19795-1:2006. Standard.

Jaha, E. S. and Nixon, M. S. (2016), 'From clothing to identity: Manual and automatic soft biometrics', *IEEE Transactions on Information Forensics and Security* **11**(10), 2377–2390.

Jaha, E. S. and Nixon, M. S. (2017), Clothing analysis for subject identification and retrieval, *in* 'Recent Advances in Intelligent Image Search and Video Retrieval', Springer, pp. 167–211.

Jain, A. K., Dass, S. C. and Nandakumar, K. (2004), Soft biometric traits for personal recognition systems, *in* 'International conference on biometric authentication', Springer, pp. 731–738.

Jain, A. K. and Li, S. Z. (2011), *Handbook of face recognition*, Vol. 1, Springer.

Jain, A. K. and Park, U. (2009), Facial marks: Soft biometric for face recognition, *in* '2009 IEEE International Conference on Image Processing', IEEE, Cairo, Egypt, pp. 37–40. Conference date: Nov 7–10, 2009.

Jain, A. K., Ross, A. and Prabhakar, S. (2004), 'An introduction to biometric recognition', *IEEE Transactions on circuits and systems for video technology* **14**(1), 4–20.

Jain, A., Ross, A. and Pankanti, S. (2006), 'Biometrics: a tool for information security', *IEEE Transactions on Information Forensics and Security* **1**(2), 125–143.

Jolliffe, I. T. (2002), *Principal component analysis for special types of data*, Springer.

Jones, A. L. and Jaeger, B. (2019), 'Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness', *Symmetry* **11**(2), 279.

Jung, D. and Jung, J. (2025), 'Data-driven understanding on soccer team tactics and ranking trends: Elo rating-based trends on european soccer leagues', *PLoS One* **20**(2), e0318485.

Kagian, A., Dror, G., Leyvand, T., Cohen-Or, D. and Ruppin, E. (2006), 'A humanlike predictor of facial attractiveness', *Advances in Neural Information Processing Systems* **19**.

Kim, M., Kim, I., Kim, J., Oh, J., Chang, J. and Park, S. (2023), 'A study on the protection of biometric information against facial recognition technology', *KSII Transactions on Internet and Information Systems (TIIS)* **17**(8), 2124–2139.

Kittler, J., Hatef, M., Duin, R. P. and Matas, J. (1998), 'On combining classifiers', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 226–239.

Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, *in* 'Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)', Vol. 2, pp. 1137–1143.

Kovashka, A., Russakovsky, O., Fei-Fei, L. and Grauman, K. (2016), 'Crowdsourcing in computer vision', *Foundations and Trends® in computer graphics and vision* **10**(3), 177–243.

Kremic, E. and Subasi, A. (2016), 'Performance of random forest and svm in face recognition', *International Arab Journal of Information Technology* **13**(2), 287–293.

Kshirsagar, V. P., Baviskar, M. R. and Gaikwad, M. E. (2011), Face recognition using eigenfaces, *in* '2011 3rd International Conference on Computer Research and Development (ICCRD)', Vol. 2, IEEE, pp. 302–306.

Kumar, K. S., Semwal, V. B. and Tripathi, R. C. (2011), 'Real time face recognition using adaboost improved fast pca algorithm', *arXiv preprint arXiv:1108.1353* .

Kumar, N., Berg, A. C., Belhumeur, P. N. and Nayar, S. K. (2009), Attribute and simile classifiers for face verification, *in* '2009 IEEE 12th international conference on computer vision', IEEE, pp. 365–372.

Kumar, N., Berg, A. C., Belhumeur, P. N. and Nayar, S. K. (2011), 'Describable visual attributes for face verification and image search', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(10), 1962–1977.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M. and Smoot, M. (2000), 'Maxims or myths of beauty? a meta-analytic and theoretical review.', *Psychological bulletin* **126**(3), 390.

Laurentini, A. and Bottino, A. (2014), 'Computer analysis of face beauty: A survey', *Computer Vision and Image Understanding* **125**, 184–199.

Lawrence, S., Giles, C., Tsoi, A. and Back, A. (1997), 'Face recognition: A convolutional neural-network approach', *IEEE Transactions on Neural Networks* **8**(1), 98–113.

Learned-Miller, E., Huang, G. B., RoyChowdhury, A., Li, H. and Hua, G. (2016), 'Labeled faces in the wild: A survey', *Advances in face detection and facial image analysis* pp. 189–248.

Lebedeva, I., Guo, Y. and Ying, F. (2021), Deep facial features for personalized attractiveness prediction, *in* 'Thirteenth International Conference on Digital Image Processing (ICDIP 2021)', Vol. 11878, SPIE, pp. 72–80.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. and Talwalkar, A. (2018), 'Hyperband: A novel bandit-based approach to hyperparameter optimization', *Journal of Machine Learning Research* **18**(185), 1–52.

Li, S. and Deng, W. (2020), 'Deep facial expression recognition: A survey', *IEEE transactions on affective computing* **13**(3), 1195–1215.

Liang, L., Lin, L., Jin, L., Xie, D. and Li, M. (2018), Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction, *in* '2018 24th International conference on pattern recognition (ICPR)', IEEE, pp. 1598–1603.

Lienhart, R. and Maydt, J. (2002), An extended set of haar-like features for rapid object detection, *in* 'IEEE International Conference on Image Processing (ICIP)'.

Liu, S., Fan, Y.-Y., Samal, A. and Guo, Z. (2016), 'Advances in computational facial attractiveness methods', *Multimedia Tools and Applications* **75**(23), 16633–16663.

Liu, X., Li, T., Peng, H., Chuoying Ouyang, I., Kim, T. and Wang, R. (2019), Understanding beauty via deep facial features, *in* 'Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops', pp. 0–0.

Liu, Z., Luo, P., Wang, X. and Tang, X. (2016), 'Large-scale celebfaces attributes (celeba) dataset'. Retrieved from `http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html`.

Liu, Z., Luo, P., Wang, X. and Tang, X. (2018), 'Large-scale celebfaces attributes (celeba) dataset', *Retrieved August* **15**(2018), 11.

Lu, G.-F., Zou, J., Wang, Y. and Wang, Z. (2016), 'L1-norm-based principal component analysis with adaptive regularization', *Pattern Recognition* **60**, 901–907.

Luo, Q., Rossion, B. and Dzhelyova, M. (2019), 'A robust implicit measure of facial attractiveness discrimination', *Social Cognitive and Affective Neuroscience* **14**(7), 737–746.

Mäkinen, E. and Raisamo, R. (2008), 'Evaluation of gender classification methods with automatically detected and aligned faces', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(3), 541–547.

Malloy, T. E., DiPietro, C., DeSimone, B., Curley, C., Chau, S. and Silva, C. (2021), 'Facial attractiveness, social status, and face recognition', *Visual Cognition* **29**(3), 158–179.

Martinez, A. M. and Kak, A. C. (2001*a*), 'Pca versus lda', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(2), 228–233.

Martinez, A. M. and Kak, A. C. (2001*b*), 'Pca versus lda', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(2), 228–233.

Martinho-Corbishley, D., Nixon, M. S. and Carter, J. N. (2015), Soft biometric recognition from comparative crowdsourced annotations, *in* '6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15)', IET, pp. 1–6.

Martinho-Corbishley, D., Nixon, M. S. and Carter, J. N. (2018), 'Super-fine attributes with crowd prototyping', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(6), 1486–1500.

Martínez-Cantin, R. (2014), 'Bayesopt: A bayesian optimization library for nonlinear optimization, experimental design and bandits', *Journal of Machine Learning Research* **15**, 3735–3739.

Melzi, P., Tolosana, R., Vera-Rodriguez, R., Kim, M., Rathgeb, C., Liu, X., DeAndres-Tame, I., Morales, A., Fierrez, J., Ortega-Garcia, J. and Zhao, W. (2024), 'Frcsyn-ongoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems', *Information Fusion* **107**, 102322.

Mitsuda, T. and Yoshida, R. (2006), Application of near-infrared spectroscopy to measuring of attractiveness of opposite-sex faces, *in* '2005 IEEE Engineering in Medicine and Biology 27th Annual Conference', IEEE, pp. 5900–5903.

Moore, D. S. and McCabe, G. P. and Craig, B. A. (n.d.), 'Introduction to the Practice of Statistics (9th ed.)', W.H. Freeman. 2017.

Ngo, H. T., Rakvic, R. N., Broussard, R. P. and Ives, R. W. (2009), An fpga-based design of a modular approach for integral images in a real-time face detection system, *in* 'Mobile Multimedia/Image Processing, Security, and Applications 2009', Vol. 7351, SPIE, pp. 83–92.

Nixon, M. and Aguado, A. (2019), *Feature extraction and image processing for computer vision*, Academic press.

Nixon, M. S., Correia, P. L., Nasrollahi, K., Moeslund, T. B., Hadid, A. and Tistarelli, M. (2015), 'On soft biometrics', *Pattern Recognition Letters* **68**, 218–230.

Olszewska, J. I. (2016), 'Automated face recognition: Challenges and solutions', *Pattern Recognition-Analysis and Applications* **4**.

O'Toole, A., Roark, D. and Abdi, H. (2002), 'Recognizing moving faces: A psychological and neural synthesis', *Trends in Cognitive Sciences* **6**(6), 261–266.

O'Toole, A. J. and Phillips, P. J. (2017), Five principles for crowd-source experiments in face recognition, *in* '2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)', IEEE, pp. 735–741.

Parikh, D. and Grauman, K. (2011), Relative attributes, *in* '2011 International Conference on Computer Vision', IEEE, pp. 503–510.

Parkhi, O., Vedaldi, A. and Zisserman, A. (2015), Deep face recognition, *in* 'BMVC 2015 - Proceedings of the British Machine Vision Conference', British Machine Vision Association.

Patrick, G., Mei, N. and Kayee, H. (2019), Face recognition vendor test (frvt) part 3: Demographic effects, Report NISTIR 8280, National Institute of Standards and Technology.

Paul, L. C. and Sumam, A. A. (2012), 'Face recognition using principal component analysis method', *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* **1**(9), 135–139.

Pedregosa, F. et al. (2011), 'Scikit-learn: Machine learning in python', *Journal of machine learning research* **12**, 2825–2830.

Phillips, P. J., Moon, H., Rizvi, S. A. and Rauss, P. J. (2000), 'The feret evaluation methodology for face-recognition algorithms', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10), 1090–1104. Provides a detailed methodology for evaluating face-recognition algorithms using CMC.

Pohjalainen, J., Räsänen, O. and Kadioglu, S. (2015), 'Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits', *Computer Speech & Language* **29**(1), 145–171.

Pudil, P., Novovičová, J. and Kittler, J. (1994), 'Floating search methods in feature selection', *Pattern recognition letters* **15**(11), 1119–1125.

Reid, D. A., Nixon, M. S. and Stevenage, S. V. (2013), 'Soft biometrics: Human identification using comparative descriptions', *IEEE Transactions on pattern analysis and machine intelligence* **36**(6), 1216–1228.

Rhodes, G. (2006), 'The evolutionary psychology of facial beauty', *Annual Review of Psychology* **57**, 199–226.

Rhodes, H. T. F. (1968), *Alphonse Bertillon, father of scientific detection*, Greenwood.

Rich, L. E. and Ashby, M. A. (2014), 'Crime and punishment, rehabilitation or revenge: bioethics for prisoners?', *Journal of Bioethical Inquiry* **11**, 269–274.

Rosebrock, A. (2017), 'Face alignment with opencv and python', https://pyimagesearch.com/2017/05/22/face-alignment-with-opencv-and-python/. [Online; accessed 4-June-2025].

Ross, A. (2007), An introduction to multibiometrics, *in* A. K. Jain, P. Flynn and A. A. Ross, eds, 'Handbook of Biometrics', Springer, Boston, MA, pp. 271–292.

Rückstieß, T., Osendorfer, C. and Van Der Smagt, P. (2011), Sequential feature selection for classification, *in* 'AI 2011: Advances in Artificial Intelligence: 24th Australasian Joint Conference, Perth, Australia, December 5-8, 2011. Proceedings 24', Springer, pp. 132–141.

Russell, R. (2003), 'Sex, beauty, and the relative luminance of facial features', *Perception* **32**(9), 1093–1107.

Saeed, J. and Abdulazeez, A. M. (2021), 'Facial beauty prediction and analysis based on deep convolutional neural network: a review', *Journal of Soft Computing and Data Mining* **2**(1), 1–12.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S. and Pantic, M. (2013), 300 faces in-the-wild challenge: The first facial landmark localization challenge, *in* 'Proceedings of the IEEE international conference on computer vision workshops', pp. 397–403.

Samangooei, S., Guo, B. and Nixon, M. S. (2008), The use of semantic human description as a soft biometric, *in* '2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems', IEEE, pp. 1–7.

Schacht, A., Werheid, K. and Sommer, W. (2008), 'The appraisal of facial beauty is rapid but not mandatory', *Cognitive, Affective, & Behavioral Neuroscience* **8**(2), 132–142.

Schmid, K., Marx, D. and Samal, A. (2008), 'Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios', *Pattern Recognition* **41**(8), 2710–2717.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015*a*), FaceNet: A unified embedding for face recognition and clustering, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', pp. 815–823.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015*b*), Facenet: A unified embedding for face recognition and clustering, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 815–823.

Scikit-Learn (n.d.), 'RepeatedStratifiedKFold documentation', `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RepeatedStratifiedKFold.html`. Accessed: 2025-02-11.

Senior, A. W. and Pankanti, S. (2011), Privacy protection and face recognition, *in* 'Handbook of Face Recognition', Springer London, London, pp. 671–691.

Sinha, P., Balas, B., Ostrovsky, Y. and Russell, R. (2006), 'Face recognition by humans: Nineteen results all computer vision researchers should know about', *Proceedings of the IEEE* **94**(11), 1948–1962.

Sirovich, L. and Kirby, M. (1987), 'Low-dimensional procedure for the characterization of human faces', *Journal of the Optical Society of America A* **4**, 519–524.

Snoek, J., Larochelle, H. and Adams, R. P. (2012), Practical bayesian optimization of machine learning algorithms, *in* 'Advances in Neural Information Processing Systems', Vol. 25.

Sonka, M., Hlavac, V. and Boyle, R. (2014), *Image processing, analysis, and machine vision*, Cengage Learning.

Strandjev, B. and Agre, G. (2013), On applicability of principal component analysis to concept learning from images, *in* '2013 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)', IEEE, pp. 1–5.

Szeliski, R. (2010), *Computer Vision: Algorithms and Applications*, Springer.

Taigman, Y., Yang, M., Ranzato, M. A. and Wolf, L. (2014), Deepface: Closing the gap to human-level performance in face verification, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1701–1708.

Thurstone, L. L. (1927), 'A law of comparative judgment', *Psychological Review* **34**(4), 273–286.
URL: *https://doi.org/10.1037/h0070288*

Tome, P., Fierrez, J., Vera-Rodriguez, R. and Nixon, M. S. (2014), 'Soft biometrics and their application in person recognition at a distance', *IEEE Transactions on Information Forensics and Security* **9**(3), 464–475.

Trigueros, D. S., Meng, L. and Hartnett, M. (2018), 'Face recognition: From traditional to deep learning methods', *arXiv preprint arXiv:1811.00116* .

Turk, M. A. and Pentland, A. P. (1991*a*), Face recognition using eigenfaces, *in* 'Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition', IEEE Computer Society, pp. 586–587.

Turk, M. and Pentland, A. (1991*b*), 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience* **3**(1), 71–86.

Turk, M. and Pentland, A. (1991*c*), 'Eigenfaces for recognition', *Journal of Cognitive Neuroscience* **3**(1), 71–86.

University of Massachusetts Amherst (2021), 'Labeled faces in the wild – results', `https://web.archive.org/web/20211226070930/http://vis-www.cs.umass.edu/lfw/results.html`. [Online; accessed 2025-06-04].

Vinay, A., Shekhar, V. S., Murthy, K. N. B. and Natarajan, S. (2015), 'Face recognition using gabor wavelet features with pca and kpca - a comparative study', *Procedia Computer Science* **57**, 650–659.

Viola, P. and Jones, M. (2001), Rapid object detection using a boosted cascade of simple features, *in* 'Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)', Vol. 1, IEEE, pp. I–I.

Wagemans, J., Elder, J., Kubovy, M., Palmer, S., Peterson, M., Singh, M. and von der Heydt, R. (2012), 'A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization', *Psychological Bulletin* **138**(6), 1172–1217.

Wainer, J. and Fonseca, P. (2021), 'How to tune the rbf svm hyperparameters? an empirical evaluation of 18 search algorithms', *Artificial Intelligence Review* **54**(6), 4771–4797.

Wang, M. and Deng, W. (2020), 'Deep face recognition: A survey', *Neurocomputing* **405**, 415–439.

Wang, Y. and Wu, Q. (2022), Research on face recognition technology based on pca and svm, *in* '2022 7th International Conference on Big Data Analytics (ICBDA)', IEEE, pp. 248–252.

Wang, Z., Cheng, Z., Huang, H., Zhou, X. and Liu, Y. (2019), Design and implementation of vehicle unlocking system based on face recognition, *in* '2019 34th Youth Academic Annual Conference of Chinese Association of Automation (YAC)', IEEE, pp. 121–126.

Welinder, P. and Perona, P. (2010), Online crowdsourcing: Rating annotators and obtaining cost-effective labels, *in* '2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops', IEEE, pp. 25–32.

Xie, X., Zhang, T., Zhang, Q. and Xu, Z. (2021), 'Real-time face alignment: Evaluation methods, training strategies, and implementation optimization', *Journal of Real-Time Image Processing* . `https://link.springer.com`.

Xing, X., Wang, K., Yan, T. and Lv, Z. (2016), 'Complete canonical correlation analysis with application to multi-view gait recognition', *Pattern Recognition* **50**, 107–117.

Yang, Q., Zhu, B., Zhang, Q., Wang, Y., Hu, R., Liu, S. and Sun, D. (2019), 'Effects of male defendants' attractiveness and trustworthiness on simulated judicial decisions in two different swindles', *Frontiers in psychology* **10**, 2160.

Zebrowitz, L. (2018), *Reading faces: Window to the soul?*, Routledge.

Zhang, C. and Zhang, Z. (2010), A survey of recent advances in face detection, Technical report, Microsoft Research.

Zhang, D., Chen, F., Xu, Y. et al. (2016), *Computer models for facial beauty analysis*, Springer, Switzerland.

Zhang, K., Zhang, Z., Li, Z. and Qiao, Y. (2016), 'Joint face detection and alignment using multi-task cascaded convolutional networks', *IEEE Signal Processing Letters* **23**(10), 1499–1503.

Zhou, C., Wang, L., Zhang, Q. and Wei, X. (2014), 'Face recognition based on pca and logistic regression analysis', *Optik* **125**(20), 5916–5919.