An Investigation into the Performance of Ethnicity Verification Between Humans and Machine Learning Algorithms

S. K. JILANI

PhD

2020

An Investigation into the Performance of Ethnicity Verification Between Humans and Machine Learning Algorithms

A State-of-the-Art Solution for the Challenge of Face-Based Ethnicity
Verification for the Pakistani Face

Shelina Khalid JILANI

Submitted for the Degree of Doctor of Philosophy

Faculty of Engineering and Informatics
School of Media, Design and Technology
University of Bradford

2020

ABSTRACT

Shelina Khalid Jilani

An Investigation into the Performance of Ethnicity Verification Between Humans and Machine Learning Algorithms.

A State-of-the-Art Solution for the Challenge of Face-Based Ethnicity Verification for the Pakistani Face

Keywords: Artificial Intelligence, Deep Learning, Ethnicity Classification, Face Image Analysis, Human Ethnicity Discrimination Ability, Machine Learning, Own-Race Bias, Pakistani Face and Components, Pakistani Face Database.

There has been a significant increase in the interest for the task of classifying demographic profiles i.e. race and ethnicity. Ethnicity is a significant human characteristic and applying facial image data for the discrimination of ethnicity is integral to face-related biometric systems. Given the diversity in the application of ethnicity-specific information such as face recognition and iris recognition, and the availability of image datasets for more commonly available human populations, i.e. Caucasian, African-American, Asians, and South-Asian Indians. A gap has been identified for the development of a system which analyses the full-face and its individual feature-components (eyes, nose and mouth), for the Pakistani ethnic group. An efficient system is proposed for the verification of the Pakistani ethnicity, which incorporates a two-tier (computer vs human) approach. Firstly, hand-crafted features were used to ascertain the descriptive nature of a

frontal-image and facial profile, for the Pakistani ethnicity. A total of 26 facial landmarks were selected (16 frontal and 10 for the profile) and by incorporating 2 models for redundant information removal, and a linear classifier for the binary task. The experimental results concluded that the facial profile image of a Pakistani face is distinct amongst other ethnicities. However, the methodology consisted of limitations for example, low performance accuracy, the laborious nature of manual data i.e. facial landmark, annotation, and the small facial image dataset. To make the system more accurate and robust, Deep Learning models are employed for ethnicity classification. Various state-of-the-art Deep models are trained on a range of facial image conditions, i.e. full face and partial-face images, plus standalone feature components such as the nose and mouth. Since ethnicity is pertinent to the research, a novel facial image database entitled Pakistani Face Database (PFDB), was created using a criterion-specific selection process, to ensure assurance in each of the assigned class-memberships, i.e. Pakistani and Non-Pakistani. Comparative analysis between 6 Deep Learning models was carried out on augmented image datasets, and the analysis demonstrates that Deep Learning yields better performance accuracy compared to low-level features. The human phase of the ethnicity classification framework tested the discrimination ability of novice Pakistani and Non-Pakistani participants, using a computerised ethnicity task. The results suggest that humans are better at discriminating between Pakistani and Non-Pakistani full face images, relative to individual face-feature components (eyes, nose, mouth), struggling the most with the nose, when making judgements of ethnicity. To understand the effects of display conditions on ethnicity discrimination accuracy,

two conditions were tested; (i) Two-Alternative Forced Choice (2-AFC) and (ii) Single image procedure. The results concluded that participants perform significantly better in trials where the target (Pakistani) image is shown alongside a distractor (Non-Pakistani) image. To conclude the proposed framework, directions for future study are suggested to advance the current understanding of image based ethnicity verification.

DEDICATION

This work presented in this thesis is dedicated to:

Allah سُبْحَانَهُ وَتَعَالَى the Bestower, the Compassionate, the Kind,

My parents, whose unreserved love and support has helped me remain steadfast for the goals that I aspire to achieve.

My Acumé family, especially Michael, Stephen, and George.

My grandfather, Mr. Syed Abdul Quddus Jelani; who personified benevolence and courage, and was the pillar of my family. His laughter lit up any room and the loving way by which he addressed my grandmother, "Jammo" will remain with me forever.

My grandmother, Jameela Khatoon, who I miss dearly.

My paternal grandfather, Mr. Syed Abdul Bari Jelani (Consul Gen. of Cameroon) who is assiduous, compassionate and lives without fear or favour. You are a role model.

Hanai my Jaan, I love you. Your charismatic smile and endearing nature helped me through days I needed it the most.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. **Al**   Alare
2. **CAS-PEAL**   Pose, Expression, Accessory and Lighting (PEAL)
3. **Ch**   Chelion
4. **CNN**   Convolutional Neural Network
5. **En**   Endocanthion
6. **Ex**   Exocanthion
7. **G**   Glabella
8. **HURDFD**   Hajj and Umrah Face Recognition Database
9. **IC**   Identity Codes
10. **ID**   Inter-Pupillary Distance
11. **IFDB**   Iranian Face Database
12. **IMFDB**   Indian Movie Face Database
13. **JAFFE**   Japanese Female Facial Expression Database
14. **KFDB**   Korean Face Database
15. **Li**   Labiale Inferius
16. **Ls**   Labiale Superius
17. **MCP**   McCulloch Pitts Model
18. **N**   Nasion
19. **NEI**   North-East Indian Database
20. **ORB**   Own Race Bias
21. **OSH**   Optimal Separating Hyperplane
22. **PCA**   Principal Component Analysis
23. **PFDB**   Pakistani Face Database

24. PLS   Partial Least Square

25. Pog   Pogonion

26. Prn   Pronasale

27. ResNet   Residual Neural Network

28. SDE   Self Defined Ethnicity

29. SI   Sublabiale

30. Sn   Subnasale

31. Sto   Stomion

32. SVM   Support Vector Machine

33. Tr   Trichion

34. Zy   Zygion

LIST OF PUBLICATIONS

Journal Articles

1. **Jilani, S.K**., Ugail, H. and Logan, A., 2019. Inter-Ethnic and Demic-Group Variations in Craniofacial Anthropometry: A Review. *PSM Biological Research*, *4*(1), pp.6-16.
2. **Jilani, S.K**., Ugail, H., Cole, S. and Logan, A.J., 2018. Standardising the Capture and Processing of Custody Images.
3. A Comparative Analysis between Automated and Human Ethnicity Verification for Mouth Features, **Shelina Khalid Jilani**, Hassan Ugail, Irfan Mehmood and Andrew Logan. **Special Issues** for "Current Trends in Applied Artificial Intelligence" to be published in the Journal of Enterprise Information Management (*Under Review*).
4. Learnt Pakistani Ethnicity from Computer and Human-Based Classification, **Shelina Khalid Jilani**, Hassan Ugail and Andrew Logan. To be published in the Journal of Multimedia Systems (*Under Review*).

Conference Papers

1. **Jilani, S.K**., Ugail, H., Bukar, A.M., Logan, A. and Munshi, T., 2017, September. A Machine Learning Approach for Ethnic Classification: The British Pakistani Face. In *2017, International Conference on Cyberworlds (CW)* (pp. 170-173). IEEE.
2. **Jilani, S.K**., Ugail, H., Bukar, A.M. and Logan, A., 2019, October. On the Ethnic Classification of Pakistani Face using Deep Learning. In *2019, International Conference on Cyberworlds (CW)* (pp. 191-198). IEEE.
3. **Jilani, S.,** Ugail, H. and Logan, A., 2019, August. The Computer Nose Best. In *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)* (pp. 1-6). IEEE.

Book

1. Jilani., S Ugail, H., Machine Learning for Ethnicity Classification from Faces, ISBN: 9798689279930, to appear 2021.

CHAPTER 1

INTRODUCTION

1.1 Concept of Race and Ethnicity

Human ancestry dates back 150,000-200,000 years originating in Africa (Stringer and Andrews 1988). It is believed that some 60,000 years ago humans migrated from Africa in groups of colonies and eventually spread to occupy much of current day's land masses. During migration, they faced diverse climates, encountering novel pathogens and subsequently formed local communities, which were separated by geographic, linguistic and cultural barriers. Factors such as natural selection and genetic mutations functioned in parallel and introduced diversity in human populations, subsequently leading to the development of racial and ethnic diversity, which we see today (Mersha and Abebe 2015).

There are a few terms which are used to describe an individual's distinction from a certain population, subject to either physical or cultural differences. Terminology such as ancestry, ethnicity and race are widely used in literature, though the parameters for such terms are context-dependent. Race and ethnicity are two distinct concepts related to ancestry, which in broader terms relates to the connection between people i.e. a line of descent. From a genetics perspective, descendants of any ancestry will

1

display physical and/or biological characteristics of their genealogical ancestors, especially when considering that parents only transfer half of their DNA to their offspring each generation (Donnelly 1983; Mathieson and Scally 2020).

Race and ethnicity are commonly used interchangeably to describe populations. In scientific literature "race" implies biological differences that are founded on different genetic makeup, and manifest in varied physical markers for traits such as skin, eye and hair colour (Sankar and Cho 2002). On the contrary, ethnicity is a more complex construct which encompasses cultural factors, shared cultural tradition, language and religion, amongst people who may not necessarily share common biological origin. For example, the South Asian race embodies ethnicities such as Bengali, Gujarati and Pakistani, each with their own language, culture and tradition. To conclude, while features of race and ethnicity assist with the management of one's self-concept i.e. identity, there are mixed schools of thought relating to which term best represents the distinctions between human populations.

1.1.1 History of Racial Differences

The taxonomy of human populations into sub-groups is not an emerging science. There have been many contributions to the grading of humans 'Homo-Sapiens' into specific assemblies. While some scientists have proposed a categorization system based purely on skin colour (Haller 1995), others have suggested the distinction of people based on the division the

earths continents (Pickering and Hall 1854). Both views are now outdated, since skin colour is fluid and too variable for classifying human populations (Brooks and Gwinn 2010). And secondly, the mass migration of people has meant that a geographical location which once may have represented a person, is now irrelevant. Moreover, there is the factor of gene-pool mixing through interracial relations, which naturally allows for the development of facial characteristics which do not adhere to a specific group.

In the wake of the Human Genome Project (Venter et al. 2001), there has been ongoing research regarding the biological basis of race, especially since there are agreed commonalities when describing and interpreting human variance between populations. The impact of differences in DNA Methylation (i.e. the addition of Methyl groups to the DNA molecule, which modifies the function of a gene resulting in varied gene expression), have recently assisted in differentiating between the African and Caucasian population (Yuan et al. 2019). Thus, highlighting that there may be a biological basis for human variance. Nonetheless, there is still a lot of research to be conducted before a definitive answer for a biological basis of race can be concluded.

Amidst Epigenome-Wide association studies (EWAS) research, there is an abundance of anthropological research which reports that discrepancies exist between multiple groups of humans, based solely on metrics from the face and its features (Farkas et al. 2005; Jilani et al. 2019). This provides a grounding for computational methods when embarking on

ethnicity classification challenges, as it highlights that on a primary level, there are variances between human populations based also on gender. Within the field of machine learning, the terms race and ethnicity are used interchangeably. In the context of this research, ethnicity refers to a person's cultural and ancestral background. For example, a person belonging to the South Asian race can belong to either the Bangladeshi, Gujarati or Pakistani ethnicity.

1.1.2   Inter-Ethnic and Racial Variances of the Face

Anthropometric facial analysis has a direct impact for the structure of a face-feature. In research when features are extracted for ethnic classification within a machine-learning framework for example, it is those features that are used as a diagnostic of ethnic identity (Jilani et al.). Leslie Gabriel Farkas, the pioneer of anthropometry demonstrated that race could be measured in the face and classified with predefined measurements. (illustrated in Figure 1.1). The results show that the degree of agreement between measurements in three populations classed as Middle Eastern i.e. Egyptian, Turkish and Iranian, and the North American Caucasian group (the control group) is considerably greater for vertical (taken from profile), compared to horizontal (taken from the frontal view), measurements. In the three Middle Eastern male groups, all 7 vertical measurements were found to be comparable to that of the North American Caucasian group. For females, however, it was only the Turkish group who shared comparable vertical values with the North American reference group
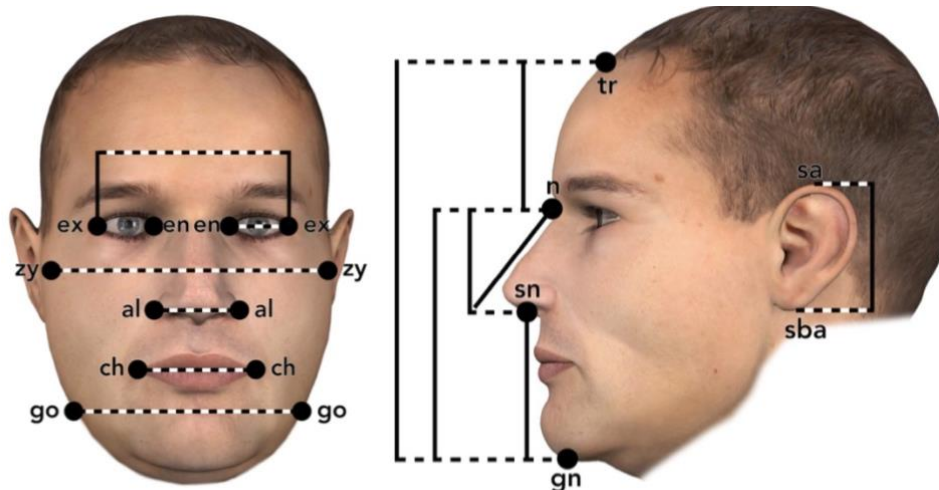
Figure 1:1: Measurements taken from fourteen anthropometric landmarks on the front and profile view of a face, for identifying race specific measurements. The image has been recreated from the original Farkas et al (2005) study. (Landmark references denote: ex:exocanthion, en:endocanthion, zy:zygion, al:alare, ch:chelion, go:gonion, tr:trichion, n:nasion and sn:subnasle.

The orbital region and nose height showed the greatest discrepancies in measurements across all the researched groups. Likely to be attributable to the finding that the nose was characteristically wide in both males and females of the Asian and African group. For the Middle Eastern group, however, the nose width was comparable to that of the North American Caucasians but differed significantly in nasal height. With regards to the concept of race and ethnicity, the study by Farkas et al., (2005) appears to outline race as the geographic location of the participants (such as the Middle East) and then group the participants by ethnicity, such as Egyptian, Iranian and Turkish. A systematic review conducted by (Fang et al. 2011) analysed data from 27 ethnic groups that were categorised into 5 principle racial groups: Europeans, Africans, East Asians, South Asians and Native Americans.

5

A total of 11 linear facial measurements were considered and inter-ethnic variability was described by 95% confidence intervals of individual measurements. The authors categorised the measurements into five degrees of variability: (1) least variable, (2) less variable intermediate, (3) intermediate, (4) more variable intermediate and (5) most variable. The results concluded that the greatest inter-ethic variation is reported in forehead height (measured tr – n) and intercanthal distance (en - en). While, measurements of the mid face width (zy - zy) and exocanthion distance (measured right ex - left ex) report the lowest degree of variability, illustrated as part of Figure 1.2. Ultimately, such findings indicate that the human face and its individual components are attributable to ethnicity.
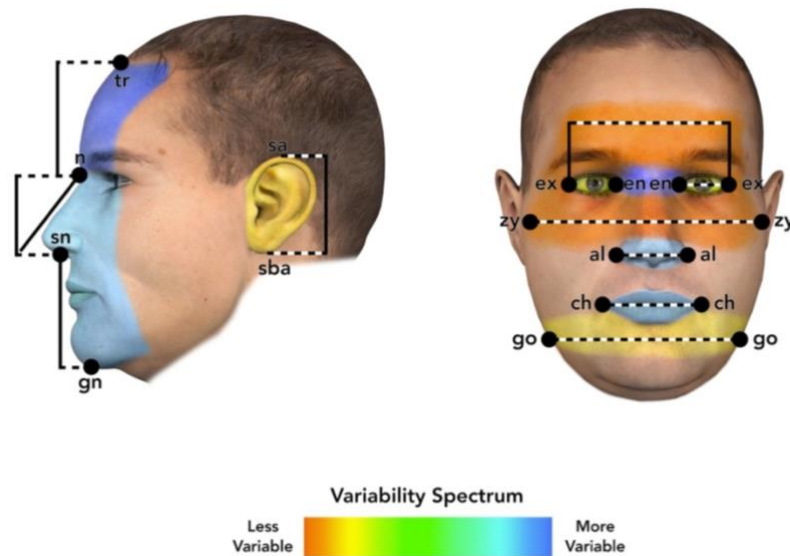
Figure 1:2: A colour variability spectrum of the front and profile view of the face, demonstrating the degree of inter-ethnic variability from different regions of the face as reported by Fang et al. (2011). (The image has been recreated from the original study).

It appears that most publications, regardless of the target population, aim to present findings that relate to an "ideal" face, defined as one with no facial deformities or previous history of medically induced facial modifications. Whilst it is understandable for a study, which aims to investigate the extent of which a race/ ethnic classification affects facial parameters, this form of selection may be problematic. A supplement to this argument is, the parents of the participants who take part in the research are presumed to be "pure" with no possibility of recessive genetic characteristics from previous generations. This is also an issue, which may grow with future data collection. Further, no robust definition is given for the participants "race" or "ethnicity", and it is used more to fit with the "race" or "ethnicity" designated for the participant, as per the researcher. Amidst the large-scale immigration, any population will inevitably see shifts to a greater or lesser extent in its ethnic or genetic make-up. Immigration not only removes an immediate relationship between racial or ethnic and geographic origin for subsequent generations, but also introduces the opportunity for more genetic or racial homogeneity, albeit over many generations with a gradual shift away from previously established norms.

There are many variables that can specifically define a population for study, geography being the most obvious but additionally language. While the use of language as a selection criterion for participants does have a very strong relationship with geographical distribution, this is not always the case. In fact, for some languages the correlation to geographical

location is far less and widely spoken international languages might give no clue to geographic origin or broad ethnic/ racial classification. To conclude, it is apparent that metric information taken from the face is a data form which describes the variability of faces across different demographic categories. Essentially, whether a human population is characterized based on their race or ethnic origin, variations in facial measurements explain the variations amongst populations.

1.1.3 Race and Ethnicity in an Applied Context

Identity Codes (IC) known also as the 6+1 system are a set of terms used to describe the ethnic identity of an unknown person by the British Police. The codes are ethnic descriptors but they are not absolute markers for recognition, since the assignment relies on an officer's subjective observation (C/I Kevin Bowsher 2007). The use of IC codes are expressed in Section 95 of the Criminal Justice Act 1991 (Association of Chief Police Officers of England and Ireland 2002; Government 2019) and by law, a Police Officer is obliged to collect information on the ethnic identity of a subject, during certain events such as, see table 1.1:

- Arrests;
- Cautions, Reprimands and Final Warnings;
- Stop and Searches.

Code	Ethnicity
IC1	White – North European
IC2	White – South European
IC3	Black
IC4	Asian – Indian Subcontinent
IC5	Chinese, Korean, Japanese or Southeast Asian
IC6	Arab or North African
IC7	Unknown

Table 1:1 Police Identity Code (IC), used to describe the ethnic identity of an unknown subject and is based on the Police Officer's subjective visual assessment. The system is also referred to as the 6+1 system.

It is apparent that the IC codes assist an officer by filtering the perceived ethnicity of a subject and will assist in producing a generic image in the mind of an officer. However, a drawback to the codes is that they do not have any scientific foundation and are mere verbal descriptors, which can be influenced by bias. What one officer may class as an Arab may be equally classed as Black, for example. Thus, with regards to the concept of race and ethnicity, IC codes prove superficial and misleading especially if officers are depending on transient features such as skin tone, to qualify physical descriptors as ethnic markers. To limit the fluidity of cases where the ethnic descriptors span across more than one code, the "16+1" Self-Defined Ethnicity (SDE) framework was presented in the 2001 United Kingdom Consensus and was implemented on 1st April 2002 (System 2011).

The 16+1 SDE allows a person, to allocate his or her own ethnicity removing any decision from the Police Officer. While the 16+1 SDE classification appears to be inclusive of broader sub-groups for a given

population some of the categories are still narrow. This is particularly apparent for descriptions such as "*Any other mixed background"* and "*Any other Asian background*" for example, since the non-specific terminology leaves the scope of ethnic assignment open to interpretation. In October 2009, the Office of a National Statistics (ONS) implemented additional codes to the 16+1 ethnic classification system (Statistics 2009). The new codes included *"Gypsy or Irish Traveler"* under the broad heading group of 'White' (with the corresponding code: W3) and 'Arab' expressed as O2. Further improving the Self-Defined Ethnicity (SDE) classification system and increasing it from 16+1 system to an 18+1. Though the addition of an 'Arab' option may be inclusive of the Middle Eastern population from Gulf countries such as Bahrain, Oman, Qatar, Saudi Arabia and the United Arab Emirates (UAE), it does not include people of Egyptian, Iranian or even Kurdish descent, leaving slight ambiguity. The mass movements of people around the world and the consequent intermingling of racial and genetic heritage will inevitably present a challenge to racial or ethnic categorisation. While diversity of all sub-ethnic groups should be represented to ensure inclusion, it may prove too difficult to implement. Therefore, broad categories within which different sub-groups can extend seems the most logical step. Moreover, the development around the concept of ethnicity is best kept isolated from national identity.

1.2 Bradford Demographics

In 2016, the population of the UK was at its largest ever at 65.6 million people, according to the Office of National Statistics (ONS). The most recent census published in 2011 shows that in England and Wales, 80% of the population were White British, and the Asian group (including Pakistani, Indian, Bangladeshi, and other) made up 6.8% of the population (Statistics 2012). While the United Kingdom is ethnically diverse, there are cities within the country which have proportionally higher numbers of specific ethnicities. Bradford is the fifth largest district in terms of population in the UK, after major cities such as Birmingham, Leeds and Sheffield. According to Figures reported by ONS an estimated 534,300 people live in Bradford (Council 2019). Although the district largely consists of people who identify themselves are White British, Bradford has the largest proportion of people of Pakistani ethnic origin in England, 20.3%. Moreover, the mixed ethnic group for populations such as Indian, Pakistani and Bangladeshi are higher for Bradford than for West Yorkshire as a county and England. Suggesting a higher rate of inter-ethnic unions which may potentially generate mixed gene pools. This is also reflected by the 10% increase of the Pakistani population within Bradford, since the 2001 consensus (Colborn 2017).

With the next census upcoming in 2021, it appears that the Pakistani ethnicity will follow the current trend by increasing in numbers. Subsequently, other changes will almost be inevitable. Therefore, more information about this ethnic group would be beneficial to both human and

computer vision related research. Especially since the current literature does not appear to have included the (British) Pakistani population as an investigated group. We consider one of the novelties of our work as being able to work with this ever-increasing ethnic population.

1.3   The Role of Face Features in Recognition

Faces are amongst the most complex objects processed by the human visual system. Within the discipline of psychology, there is a grouping system which divides the features of the face into internal (eyes, nose and mouth) and external (hair and face shape) features. The importance of internal in comparison to external features is essentially dependent on familiarity. For example, in a study conducted by Ellis et al. (1979) it was reported that the importance of the centralised internal features (eyes, nose and mouth) was greater for familiar face recognition, in contrast to external features (hair and face shape). One of the experiments consisted of showing participants a series of famous people's faces either as (i) a whole face, (ii) the hair and face contours only and (iii) the inner face only (Figure 1.3). On average, the novice participants accurately identified 80% of the familiar faces when shown as a full face, in contrast to only 50% when recognised from inner facial features only, and 30% when hair and face contours were shown (Ellis et al. 1979).

Figure 1:3: An example of a familiar face stimulus which was shown as part of experiment 1 during Ellis et al. (1979) study. The images were shown in such contrasting ways to understand which features drove face recognition. The face is of James Callaghan and was shown as one of the three forms depicted above for 9 seconds at a time, with a 6 second inter-slide interval.

An identified weakness of the study flagged by the experimenters was the use of a single image only, which may not have fully captured the face of the famous person. Additionally, the results may reflect the lack of knowledge the students have of the chosen famous people, instead of testing their recognition ability based on internal and external features. The importance of familiarity as a key contributor for the internal feature advantage has also been evidenced by Bonner et al. (2003). University students from Sterling learned the faces of 24 unfamiliar subjects from videos used as part of Bruce et al. (1999) study, over a period of 3 days.

On day 1, participants completed an unfamiliar face matching task before proceeding to watching a 30 second video clip of the target. This was conducted to generate a baseline result for unfamiliar face matching.

13

During day 2 and 3, the order was reversed and the videos were shown before the matching task was completed. The face matching task consisted of either a video image of the target face with the internal or external features visible only, or the photograph of the true target or a similar looking distractor, shown as part of Figure 1.4. The researchers concluded that with the passing of time (albeit 3 days in this study), as the face became familiar the role of internal features became more important for recognition. This notion of converting an unfamiliar face to familiar over time rather than instantaneously, has also been reported by other researchers (Clutterbuck and Johnston 2002; Clutterbuck and Johnston 2004; Clutterbuck and Johnston 2005; Osborne and Stevenage 2008).

Figure 1:4: An example of a face matching task stimuli, used as part of the Bonner et al. (2003) study. The whole face images are photographs of either the target (the true person) or a similar looking distractor. The participants were shown images of the whole face alongside external (hair, face shape and ears) and internal (eyes, nose and mouth) features, and were required to match the images.

A potential explanation for the internal feature advantage is the location of where these features are on the face. Since the features are centrally located without an element of transiency (unlike hair), it may be that human vision is naturally drawn to that region of the face. Furthermore, the expressive nature of the internal features may contribute to the successful memory recall of faces, making them more distinctive (Young 1984). While familiar face recognition is driven by internal features, there is research which suggests that external features are a contributing factor for the processing of unfamiliar faces (Ellis et al. 1979; Hancock et al. 2000).

Earlier, Pascalis et al. (1995) studied face recognition by newborn babies and found that while babies could reliably distinguish their mother's face with all the facial information (internal and external features) visible. They struggled to discriminate between the face of their mothers and a stranger when external features were covered with a scarf. Research has also indicated significant differences between the importance of the internal and external features for the identification of familiar and unfamiliar face (Sinha and Poggio 1996; Sinha 2000). Sinha and Poggio (1996) placed the internal features of Bill Clinton within the external features of Al Gore (Figure 1.5). By looking at the images side-by-side the faces depicted appear to be of the same person, with identical internal features and spatial configurations, but upon closer inspection, this is incorrect. The external features i.e. the hair and ears are different. The image of Al Gore and Bill Clinton demonstrates that in some contexts, the identification of a person

relies on the head i.e. external features, rather than the internal features. The visual system therefore is said to rely on 'head-recognition' instead of face recognition.

Figure 1:5: The Bill Clinton and Al Gore Illusion, taken from Sinha and Poggio (1996). The internal features (eyes, nose and mouth) of Al Gore (dark haired male) have been replaced with Bill Clintons, while the external features (hair and ears) of Al Gore are left unchanged.

Andrews and Thompson (2010) identified a similar pattern of results for images of David Cameron and Nick Clegg. Ultimately, both the Sinha and Poggio (1996) and Andrews & Thompson, (2010) studies demonstrate that when a face is disected into internal and external features, those aspects of the face can either be observed as similar or dissimilar but on a perceptual level, external features tend to be more informatic compared to interal features. However it is apparent that the relative importance of internal and external features for face recognition, vary across cultures.

Megreya and Bindemann (2009) reported that, while British participants demonstrated the typical external feature advantage for

unfamiliar faces, Egyptian adults demonstrated a reliance on the internal features. The results were interpreted as evidence that the covering of the external features with a headscarf (a common, religious-based practice in Egypt) had resulted in a shift in reliance from the external to internal features in Egyptian participants. In line with this result, recognition accuracies for the internal features of unfamiliar faces is significantly greater in participants from the United Arab Emirates, compared to individuals from the USA (Wang et al. 2015). The internal feature advantage was also investigated by Kemp et al. (2016), who tested the ability of unfamiliar face matching of students, by showing them face-pairs on a computer screen, and instructing them to determine whether the two face images were the same or different. The face pairs depicted in the 'same' condition consisted of either minor changes in the hairstyle of the same person (same-hard condition) or no changes (same-easy condition). Equally, in the 'different' conditions, the face pairs either represented similar looking people (same-hard) or dissimilar people (same-easy). The research reported an improvement of 2.4% in the matching of unfamiliar faces where the external features were masked and the internal features were visible. Moreover, in the conditions when hard-face pairs were shown, participants made accurate judgements of faces based on the internal features and made quicker decisions.

A suggested explanation for the external feature advantage for unfamiliar faces centres on the notion that since face shape and hair cover a large area it is possible to differentiate between such features over long

viewing distances (Logan et al. 2017). Compared to familiar faces where internal features may be inspected more closely, (i.e.: at social gatherings), unfamiliar faces seem to be observed from a distance and tend to be less affected by facial movement and expressions, making them more reliable cues for discrimination.

1.4 Own Race Bias (ORB)

Own-Race Bias (ORB) interchangeably referred to as the Cross-Race Effect (CRE), Other Race Effect (ORE) or Cross Race Bias (CRB), is the disadvantaged recognition accuracy for other race faces, relative to those of one's own race (Barkowitz and Brigham 1982; Ayuk 1990; Meissner and Brigham 2001; Sporer 2001a). An alternative definition for ORB is the "*relative inability to recognise persons of another race*" (Malpass and Kravitz 1969). For example, a South Asian person will innately be better at recognizing another South Asian in comparison to East Asian faces, however it is subject to context. Research already evidences that racial group membership influences babies as young as 3 months who preferentially look at own race faces (Kelly et al. 2005). The most commonly attributed hypothesis for such recognition is the contact hypothesis, which proposes that familiarity through increased exposure with members of one's own race is key for the onset of ORB. Research suggests that ORB is dependent on both the quantity and quality of interracial contact. To reduce the effects of ORB, one must increase both the quantity and quality of contact of other race faces (Sporer 2001b).

There are two theoretical concepts for ORB: perceptual expertise-based and social-motivation based. Perceptual expertise suggests that ORB in face recognition may be due to a difference in the initial perceptual encoding of own-relative to other-race faces (Walker and Tanaka 2003). The theory suggests that other race faces are not as efficiently represented in comparison to own race faces, impairing the memory of other race faces. This impairment results in the poor recognition of other race faces due to restricted inter-racial contact. Numerous proposals have been put forward with respect to perceptual encoding of both own and other-race faces. Tanaka and Farah (1993) proposed that holistic processing of own race faces is greater than that of other-race faces (i.e.: the facial features are processed collectively as a whole). Whereas Valentine (1991) suggested separate perceptual representation of own and other race faces, such that individual faces are represented based on the values underlying face dimensions in face space. These dimensions are then tweaked to own race faces and are not well suited for individualizing other race faces.

In brief, the Valentine model (Face Space) suggests that the number of faces we encounter during our life are represented by hypothetical points (one point per face) mapped within a multidimensional space, where each dimension is a physiognomic trait useful for the encoding of a face. To account for ORB, Valentine proposed that since the points of the face space are based on an individual's own encounters, consisting of own race faces, it would be difficult to encode another-race face, due to the lack of

19

dimensions within the required face space.

In contrast, Social motivational approaches for the onset of ORB propose that people tend to process other race faces at a low level. Generally requiring less effort and potentially less cognitive resources, when compared to individualizing other race faces, on a minute descriptive level i.e.: "the face stimuli shown are all East Asian" as supposed to "those faces are of Japanese males, those are Korean and the last three are Chinese." This in turn suggests that without any meaningful requirement, the observer generally does not apply individuation processes to other race faces to distinguish between them. Studies have shown that selective attention does strongly play a role in other race faces (Michel et al. 2006; Ho and Pezdek 2016).

Malpass and Kravitz (1969) were one of the first to study ORB during their face recognition study, testing 20 Black and 20 White undergraduate students. They proposed that ORB in face recognition may be a result of limited contact with other race faces compared to own race faces. Consequently, suggesting that people residing in racially homogenous cities/countries with no contact with other race faces would more likely exhibit a strong ORB. Since then a series of experimental studies have been set up replicating the testing strategy for ORB amongst different racial groups from different geographical locations. ORB has been demonstrated in demographic groups other than the White population, since literature has

reported the prevalence in East Asians (MacLin et al. 2001), Hispanics (Gross 2009), Black (Gross 2009), and the Middle-Eastern, Egyptian population (Megreya et al. 2011b). During perceptual discrimination studies, face matching tasks are used to assess the effects of ORB. Face matching tasks can either be sequential or simultaneous. In a sequential matching task two facial images are presented side-by-side, after which a same or different judgement is made by the participant. The extent of perceptual ORB was investigated using a same/different perceptual discrimination task between British Caucasian and British South Asian participants, by Walker and Hewstone (2006). Face stimuli were generated by morphing a South Asian and a Caucasian face along a linear scale (Figure 1.6).
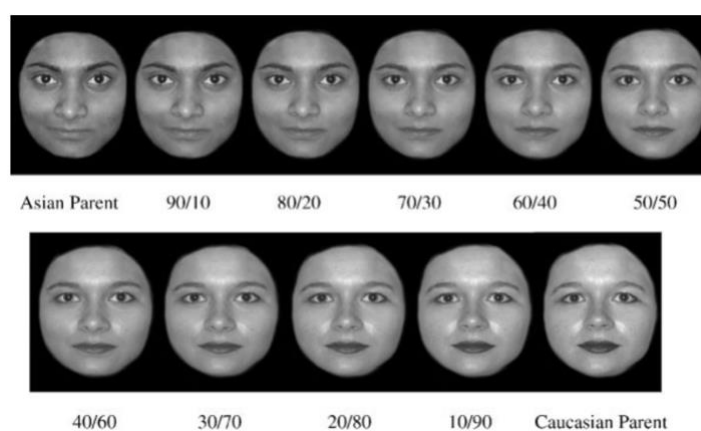
Figure 1:6: An example of face stimuli taken from the Walker and Hewstone (2006) experiment. A total of 16 face images were paired based on similarities in face shape and using the Morph 2.5 program, each face pair was averaged together on a linear continuum. (e.g.: Asian parent face was morphed with a Caucasian face and vice versa) to generate an average face at 10% intervals until a 90% contribution was reached.

The task required participants to judge whether the morphed face was physically identical to or different from, an original 'parent' face. Own Race Bias was reported in the British Caucasian results since they recognised White faces significantly better than South Asian faces; White faces: $m = 46.28$, SD = 1.90, South Asian face $m = 39.27$, SD = 1.61. Whereas, the South Asian participants recognised both sets of faces equally (White faces: $m = 36.64$, SD = 1.77 and South Asian faces $m = 38.87$ SD = 2.55), suggesting a lack of an ORB in these participants.

The difference in ORB onset between the two groups of participants was explained by the observation that the South Asian participants had comparable contact with both own and other-race faces, since birth. The heterogeneous assortment of people the South Asian participants encountered helped to build a broader cognitive schema, which may have assisted with the discrimination task. However, not all research supports the contact hypothesis as an explanation for the ORB.

Tullis et al. (2014) conducted experiments to evaluate whether participants exhibited ORB during two conditions: self-paced (participants learned a face with no time constraints) and fixed-rate (average time taken by a self-paced learner). Caucasian and Chinese participants took part in a series of tests which consisted of presenting individual face stimuli (Caucasian and Chinese) on a computer screen. Participants were asked to memorize the faces before partaking in a memory test. It was concluded that self-pacing i.e. allocating equal amounts of time for studying own and

other-race faces, has no effect on ORB. Moreover, for the Caucasian participants, interaction with other-race faces (Asians) did not reduce the extent of ORB.

In brief, the ORB is the tendency of perceivers to recognise own race faces better than other race faces. It is generally considered to develop from individual face experience and, perhaps, social motivation.

1.5 Computer Vision

Researchers in the field of vision science have worked for decades on understanding how the human visual system works, and is affected by illusions of the face. As humans, we can look at a bowl of food and accurately categorise what we see whilst interpreting factors such as colour, contrast, texture and shape. As humans, we have no difficulty in directing our attention, a somewhat underrated human ability.

The simple definition of computer vision is the computer's ability to simulate human brain behaviour. One of the most renowned application of computer vision which is full of commercial interest is facial recognition, see Zhao et al. (2003). The automated computer-based process for person verification has solutions to a host of challenges such as; (1) controlled access, (2) rapid suspect identification by law enforcement agencies (FaceFirst 2019) and, (3) biometric airport check-in points. While, research evidences the broad applications of computer vision techniques, the ability of a computer to fully operate on-par with the human ability remains a

challenge. And human performance is considered the gold-standard against which machines must compete.

Consequently, a common criticism of machine-based systems is that for real-world applications, the machine must always perform at the same level as, but ideally above human performance.
Although currently implemented face-related systems do provide numerous advantages, there is evidence of inaccuracies and misidentification (Alexander J Martin 2017; Lamb 2017).

A racial bias similar to the phenomenon evidenced in humans (ORB) is also alleged in facial recognition algorithms (Phillips et al. 2011). In a study conducted by researchers at Massachusetts Institute of Technology (MIT), inaccuracies in gender identification was reported because of skin tones. Three facial recognition algorithms developed by Microsoft, IBM and a Chinese firm, called Face++, who reported that gender was misidentified in up to 7% of cases for light-skinned female images and up to 1% of light-skinned males (Buolamwini and Gebru). In contrast, greater inaccuracies were reported for the darker-skinned subjects; 12% for males and up to 35% for females.

Such results point to the requirement of diverse datasets for algorithms. Essentially, computer algorithm performance will only continue to improve if the datasets on which they are trained and tested, are continually challenging. A novel dataset has been produced as part of this

thesis to address the challenge of ethnicity verification and is presented in Chapter 2.

1.5.1 Deep Learning

Artificial intelligence (AI) is a technique which enables a computer to mimic human behaviour and Machine Learning (ML) is a sub-category (Michalski et al. 2013), within which Deep Learning is it's subfield. Deep Learning allows computational models to learn and represent data in a manner mimicking how the human brain perceives and understands multi-modal information (Bengio et al. 2015). The first computational model of a neuron was proposed by Warren McColloch and Walter Pitts. They worked to understand how it was possible for the human brain to produce intricate patterns, by using interconnected neurons and used it as a foundation for their proposed MCP model. A mathematical concept founded on the basis of a biological neuron (Hayman; McCulloch and Pitts 1943). In brief, Neurons are interconnected nerve cells within the brain (collectively known as a Neural Network) and function by processing and transmitting information via electrical and chemical signals, see Figure 1.7.
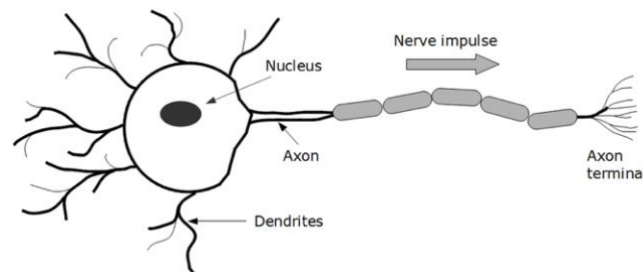
Figure 1:7: An illustration of a single Neuron (known also as a Nerve cell) taken from Zaitcev et al. (2015). A Neuron is an electrically excitable cell which is connected with other Neurons, which produce a network of cells where electric impulses can be originated, transmitted and received.

25

A biological neurone is stimulated when an action potential is generated because of a change in the ion concentration across the cell membrane. Generally, there is a higher concentration of sodium ions in the extracellular space while there is a higher concentration of potassium ions within the intracellular space. During an action potential, ions are transported back and forth across the neurones membrane causing an electrical change that transmits the nerve impulse. Similar to the functioning of a human neuron, the MCP Neural model receives a series of incoming signals $x1$, $x2$, $x3$, which would either be excitory of inhibitory. If the weighted sum of the incoming signals is at its threshold the model gives an output of 1, if not the output is 0. The model bases its output decision on the input signals i.e. $x1$, $x2$, $x3$, by performing a weighted sum, which in turn generates a binary output i.e. 0 or 1, see Figure 1.8 (Hayman).
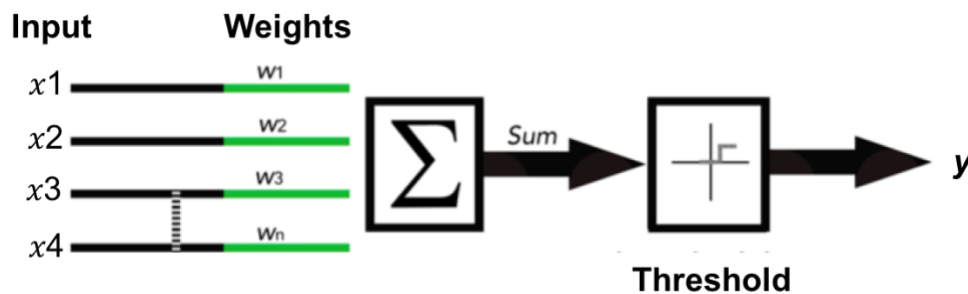
Figure 1:8: An illustration of the McCulloch and Pitts Computational Neuron Model (MCP) recreated from Kawaguchi (2017). The model is the first proposed computational neural network, founded on the mechanism of a biological Neuron. Essentially, the model consists of inputs i.e. $x1, x2, x3$ to x^n with weights (0,1 or -1, 1) which are summed together and used to provide a binary output, y.

The MCP model was the first of its kind and is agreed to be the foundation of AI and Neural Networks. Since then, there has been a host of developed models such as Neocognitrion (Fukushima et al. 1983), Perceptron (Rosenblatt 1958), LeNet (LeCun et al.; Al-Jawfi 2009) and more recently Convolutional Neural Networks (CNNs) (LeCun et al. 2015) and Deep Residual Neural Network (ResNet) (He et al.)

Keeping to the context of face image analysis for the classification of ethnicity, several frameworks have attempted to classify demographic profiles i.e. race and ethnicity. The human face is understood to be the most informative region of a person, while the extraction of race specific features has demonstrated robustness in anthropometry. The study by Bagchi et al. (2019) investigated a binary ethnicity classification problem between Asian and Non-Asian faces. A dataset of 3,105 front-facing colour images were collected from Google and Facebook profile images (53 for testing and 3,052 for training). Using a CNN algorithm implemented in MATLAB a classification accuracy of 84.91% was reported.

Saliha et al. (2019) recently proposed a two-class i.e. Asian vs Non-Asian race classification framework, called Spark (Saliha et al. 2019). Race specific features were extracted from frontal images using Local binary patterns (LBP), a texture tool, which labels the pixels of an image. Using a dataset of combined images from the colour FERET and CAS-PEAL facial image database and using a combination of classifiers such as the Linear Support Vector Machine (SVM), Decision Trees (DT) and Random Forest,

classification accuracies of >90% were reported; 97.59%, 99.26% and 98.65%, respectively.

1.5.1.1 *Convolutional Neural Network (CNN)*

Convolutional Neural Networks (CNN) are one of the distinguished Deep Learning approaches, "*where multiple layers are trained in an end-to-end manner*" (Guo et al. 2016). CNNs are inspired by the animal visual cortex (Hubel and Wiesel 1968) and grew prominence when Alex Krizhevsky utilized the computer model to win an ImageNet competition (Krizhevsky et al.). Cementing this method of Deep Learning to be central to image recognition. A CNN consists of three core neural layers, (1) Convolutional layers, (2) Pooling layers (down-sampling) and (3) Fully connected (FC) layers (Guo et al. 2016; Yamashita et al. 2018). Each of the layers perform different tasks yet work together to generate an output for a given class, see Figure 1.9.
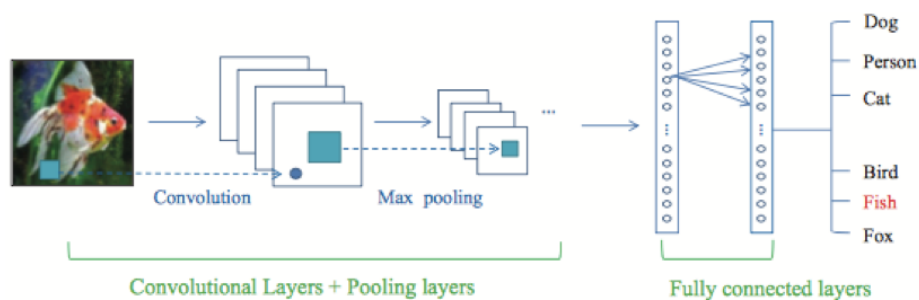
Figure 1:9: An illustrative example of the general architecture of a Convolutional Neural Network (CNN) taken from Guo et al. (2016). The process shows the classification of an input image, i.e. a fish. Convolutions of the image generates feature maps at each pixel stride which then pass the feature maps as vectors to fully-connected (FC) layers, which have the role of classification.

(1) Convolutional Layer: The CNN will utilise a Kernel (i.e. a filter) to convolve an image, section by section. Since the scanning process of the Kernal is linear, calculations are made by multiplying the generated kernel values with the original pixel value of the input data. Upon completion of the calculations, a single value is generated to represent the parameter of the initial area, prompting the kernel to move to the right by 1 stride. This step is repeated until the the entire image has been processed. By the end of the convolving several pixel arrays i.e. feature maps are generated.

(2) Pooling layer: Reduce the spatial dimensions of the input image i.e. the width and height, for the following convolutional layer. This layer is also referred to as the down-sampling stage, which results in the loss of arbitrary information whilst retaining dominant feature information. There are two reason for this, firstly, to reduce computation time for data processing and secondly to control overfitting (Guo et al. 2016). This is where the model essentially extracts noise, believing it to be representative of the model. Although data augmentation can limit the effect of overfitting (Zhong et al. 2017). For a detailed review on max pooling see Boureau et al.

(3) Fully-Connected (FC) layer: Convert two-dimensional feature maps into 1-dimensional feature vectors. FC layers function to identify high level features which best correlate to a class with specific weights, so that the output is the correct probability for a given feature (Yamashita et al. 2018). For example, given the input image of a sunflower, the FC layer will locate

high level feature maps for petals and seeds to assist with a probability given to the output being a sunflower. The process of classification can either be conducted by the FC layer or the derived vector can be processed forward for classification using Support Vector Machine (SVMs) algorithms (Krizhevsky et al.).

1.5.1.2   *Residual Neural Network (ResNet)*

In 2015 Microsoft Research Asia developed Deep Residual Network (ResNet). The critical feature of the framework is its "identity shortcut connection" which functions to skip layers during learning, without compromising accuracy. A residual block consists of 2 paths, first the 'main branch' which is a series of neural network layers, and second, a direct path from the input to the output (i.e. the shortcut branch). An "identity shortcut" enables the input to be passed directly to the output, avoiding all the intermediate layers. The advantage of the shortcut is that it helps the method of learning to become faster (He et al.; Bukar and Ugail 2017). An advantage of ResNet is that the technique does not add any extra parameters. The parameters trained are the same had there been no residual connections. This is particularly important because a greater number of parameters would lead to overfitting.

VGG-Face (VGG-F) was developed at Oxford University by researchers belonging to the Visual Geometry Group (VGG) (Parkhi et al.). It is based on the application of VGG-Very-Deep-16 CNN architecture, described in research by Simonyan and Zisserman (2014). The model is trained on a dataset of 2.6 million face images of 2622 unique identities and uses approximately 1000 images per subject. The network consists of a sequence of Convolutional, Max-Pooling and Fully-Connected (FC) layers, and is designed to work with inputs of fixed dimensions; 224 x 224.

VGG-F consists of 13 convolutional layers with filters that have an unchanging field size of 3 × 3 and a set convolutional stride of 1 pixel. Shown as part of Figure 1.10, the network also consists of 5 max-pooling layers and 3 Fully-Connected (FC) layers; FC6, FC7 and FC8. FC6 and FC7 consists of 4096 parameters, whereas FC8 has 2622 parameters.
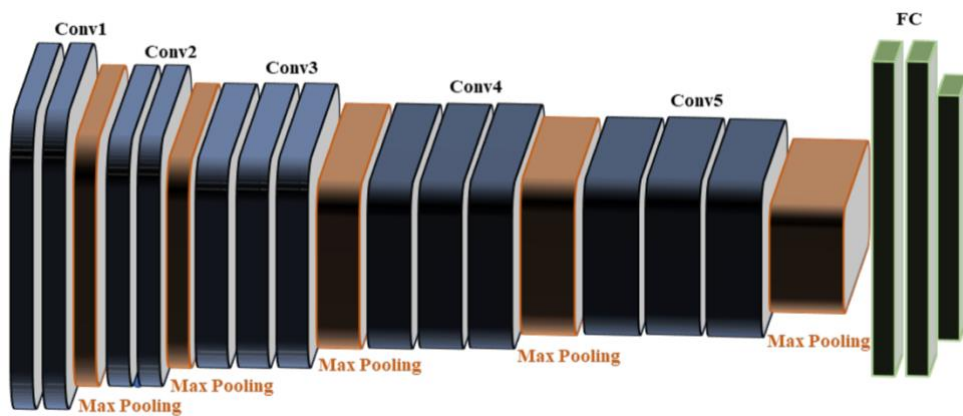
Figure 1:10: A schematic showing the VGG-Face architecture which consists of thirteen convolutional layers, five max-pooling layers and three fully-connected (FC) layers. (Recreated from El Khiyari and Wechsler (2016).

31

1.6 Problem Definition

The aim for this research is to develop an inter-disciplinary framework for the task of ethnicity classification, which encompasses Machine-Learning and Human-discrimination ability. The task of ethnicity classification is to identify to which set of sub-populations, a new observation belongs, based on a training set of data, containing observations whose category membership is known. Consider a dataset of face images F, consisting of N number of face images as given:

$$\mathcal{F} = \{F_1, F_2, ..., F_{N-1}, F_N\} \qquad (1\text{-}1)$$

This is a binary ethnicity classification E^C which aims to differentiate between a Pakistani ethnicity E^P and a Non-Pakistani ethnicity E^{NP}, as accurately as possible. Suppose, x represents multiple data points or features belonging to a face image. However, within a real-world scenario, input data for ethnicity classification can be based on either full face or partial face images. In this study, we explore both full-face x^F and partial face x^{PF} features for ethnicity classification. Both hand-crafted and Deep Learning features are incorporated in training. This ethnicity classification problem can be defined as:

$$H(x) = \theta_0 + \theta_1 x; \quad x = x^F \bigcup x^{PF} \qquad (1\text{-}2)$$

where *H(x)* is the predictor function, θ_0 and θ_1 are constants. Our goal is to find the perfect values of θ_0 and θ_1 to make our predictor work as well as possible. Optimizing the predictor *H(x)* is done using training examples. For

each training example, we have an input value x_train, for which a corresponding output, *y*, is known in advance. We find the difference between the known, correct value y, and our predicted value h (x_train). With enough training examples, these differences provide a useful way to measure the "inaccuracy" of *H(x)*. We can then tweak h(x) by tuning the values of θ_0 and θ_1 to make it "less wrong". This process is repeated over and over until the system has converged on the best values for θ_0 and θ_1.

1.7 Problem Statement

The research objective is to develop an intelligent system for image-based ethnicity classification, by probing the challenge of ethnicity from a multi-layered, computer and human approach.

- Firstly, a semi-automatic framework using low-level, manually annotated facial landmark features, is applied to an independent dataset of frontal and profile images. Supervised (Partial Least Squares) and Unsupervised (Principal Component Analysis) Machine Learning algorithms are applied in isolation of one another, to each dataset to discern the most enriched features of the face. A robust binary classifier (Support Vector Machine) is applied for reporting the rate of ethnicity classification.

- Improving on the primary testing conditions (low-level features) we advance to the implementation of a fully

automatic, Deep Learning approach. By independently applying 3 Residual-Learning Algorithms (ResNet 50/101/152) and 3 Neural-Network Algorithms (VGG-F, VGG-16 and VGG-19) to datasets of 1,000 eyes, 1,000 nose and 1,000 mouth images, we investigate the role of individual face features for ethnicity classification (High-Level features).

- To further examine the discriminatory ability of face components and the Deep Learning models (ResNet-50/101/152, VGG-F/16/19), individual datasets of 2,200 nose and 2,200 mouth images are further scrutinised. Deep Learning models are used for feature extraction and classification is carried out using a Linear Support Vector Machine. Using the same methodological approach, supplementary experiments are also conducted on the concept of feature combinations, using 1,200 eyes and nose, 1,200 nose and mouth and 1,200 mouth and eyes images.

- Finally, we investigate human ethnicity discrimination ability by conducting a series of computerised experiments to quantify the accuracy with which the Human Visual System discriminates between face and face features (eyes, nose and mouth) of different ethnicities. Participants are categorised by ethnicity (Pakistani/Non-Pakistani/Other), and gender (Male/Female), and are shown a total of 800

greyscale images, i.e. 200 full faces, 200 eyes only, 200 nose only and 200 mouth only.

1.8 Framework

Humans have increased perceptual capacity for processing the faces they encounter as either familiar or unfamiliar. Such visual judgements are based on extracting information from the feature of the face, and may also consider aspects of ORB. A primary aspect of the visual system, specific to face recognition relates to the human brain and the specialist region of the Fusiform Face Area (FFA), which reports high sensitivity for faces. By extracting information from the face a person can make attributions on gender, race, emotional state and trustworthiness. In this work, the focus is on the classification of ethnicity and the research framework is founded on three stages; (1) Low-level features, (2) High-level features and lastly, (3) Human discrimination ability.

Low-level features. Herein, a total of 26 hand-crafted features (16 on frontal images and 10 for profile) are manually annotated onto RGB images. The images are of Pakistani and other-ethnicity participants, self-assigned under strict select criterion. To remove redundant information, two different methods were used; Principal Component Analysis (PCA) and Partial Least Squares (PLS), and a Linear classifier i.e. Support Vector Machine was applied for the two-class verification. The experimentations

demonstrate that low-level features are an enriched foundation for ethnicity classification.

Progressing on from low-level features is the second stage, where high-level i.e. deep face features are implemented from deep learning frameworks. The difference between low-level and high level features is the depth of information. Low level features are essentially facial landmarks on the surface of a face image, whereas high-level features extend deeper and may be elusive differences which are not visible to the naked eye.

High-level features. A series of comparative analysis have been conducted between 6 pre-trained models (ResNet-50/101/152, VGG-F/16/19) on a range of augmented full face and isolated-feature i.e. eyes, nose and mouth, datasets. The 6 Machine Learning models were used to extract the distinguished features from the images and the binary (Pakistani vs Non-Pakistani) classification was conducted using a Linear SVM classifier. The discriminatory ability of the internal components of the face i.e. eyes, nose and mouth were also tested, to understand their contributing effect for determining ethnicity.

Human discrimination ability. To understand how humans, use the visual information contained within images of the full face and isolated feature components (eyes, nose, mouth), novice participants were recruited to partake in a custom-designed, computerised ethnicity classification task. Colour images were converted to greyscale to prevent any colour cues from assisting the participant during the computer-based ethnicity task. Using two conditions: Two-Alternative Forced Choice (2-AFC) and a single image display, ethnicity discrimination performance was investigated. See Figure 1.11 for a visual representation of the research framework.
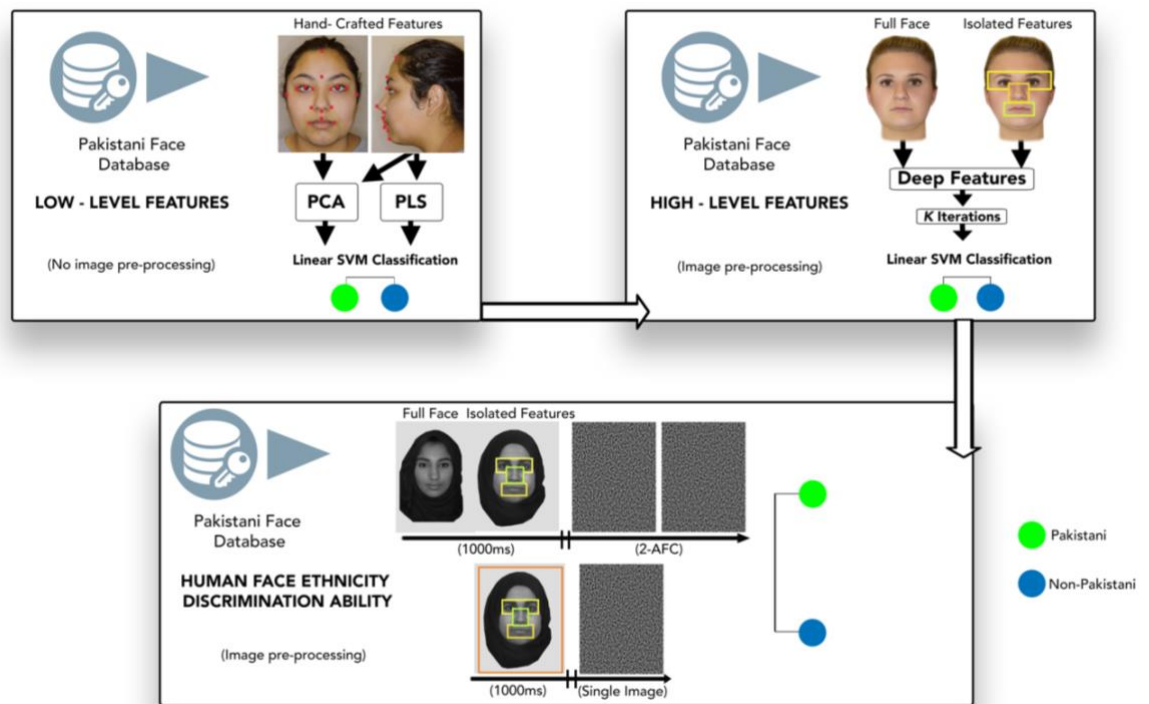
Figure 1:11: Proposed ethnicity classification framework for the Pakistani Face using a three-tier approach: (1) Low-level features, (2) High-level features and, (3) Human visual system based. The 'tick' indicates accurate classification, and the star indicates a high classification.

CHAPTER 2

PAKISTANI FACE DATABASE

2.1 Overview

The dependence of researchers on diverse datasets of human face and face-feature images, is evident within the field of computer vision. However, issues arise when researchers are dependent on restrictive datasets that are unable to encompass diversity. While, a common justification to such an error is the requirement of a universally balanced dataset, upon which models can be trained. It is an exhaustive task, trying to create a balanced dataset inclusive of all genders, race and ethnicities, which also incorporates people of dual-heritage. And, to ensure that the data is representative of all the facial characteristics seen to man. So, instead researchers are developing criterion-specific datasets which are reflective of their research proposition. This in turn, generates insular datasets which may not be complimentary to other publically available data. Although, sub-sampling from multiple datasets seems to be an effective alternative for many researchers, when trying to create multi-purpose data.

Face stimuli used within experimental settings are varied; researchers either use low-resolution photographs generated under uncontrolled illumination for various experimental aims (Phillips et al. 2000), images may also be taken from moving video files (O'Toole et al. 2005) or, some researchers use computer generated face models (Todorov et al.

2009; Logan et al. 2017). Researchers are essentially faced with a trade-off between their study requirements and the objectives they wish to achieve. This is particularly important since it is well known that for a Machine Learning algorithm to perform optimally and in an unbiased manner, the training data must be diverse. More significantly, the training data must be representative of the testing data, to prevent the Machine Learning model compromising its performance. Given the ubiquitous nature of faces and their importance in research, there will always be a demand for high quality databases which are detailed and representational.

2.2 Assessment of Facial Images for Databases

Creating a facial image database is a big responsibility which requires effort to identify and recruit criterion-specific participants, standardise the process of image capture and conduct any pre-or post-image processing. For this reason, databases of faces that are readily available offer a hassle-free alternative to researchers who may be subject to time constraints. Equally, a shortage of a specific human population may drive a researcher to combine data from a range of databases, to best fit the research question at hand. The selection criteria for facial image capture is varied and while there appears to be no gold-standard in academia, government bodies have produced standards. With the rise in biometric data for person verification, it is hoped the standards will help capture the most resourceful images.

In 2007, the National Policing Improvement Agency (NPIA) released a set of guidelines for digital image capture which considered factors such as (Islam 2007):

- pose,
- pose with and without glasses,
- facial expression,
- lighting conditions,
- lighting for glasses and
- image background.

Additional considerations relate to the minimum number of images, contact lenses, hair and head coverings. Countries such as Germany (Castillo 2006), India (India 2010) and the United States (Paul Griffin 2005) have also produced facial image standards. While the image capture standards do not transpire into the world of academia, they play an active role in assisting law enforcement and government officials. Especially since the proliferation of CCTV and facial recognition technology.

2.3 Population Specific Databases

Creating a facial image database specific for demographic information i.e. race or ethnicity, has led to a widespread interest into the "identity" of a person. Table 2.1 presents a list of published facial image databases that have not been designed exclusively for ethnic or racial classification,

however, contain facial images representative of either (1) single, (2) dual

or, (3) multiple demographic profiles i.e. race and/or ethnicity.

Name of Database	Images	Target Race/ Ethnicity	Author and Year
NimStim Database	672 images (18 female, 25 male)	African American (10), American-Asian (6) European-American (25) and Latino-American (2).	(Tottenham et al. 2009)
CAS-PEAL (Pose, Expression, Accessory, Lighting)	99,594 images of 1040 individuals (595 males and 445 females).	All Chinese subjects.	(Gao et al. 2008)
CAS-PEAL-R1	30,900 images of 1,040 individuals.	All Chinese subjects.	(Cao et al.)
Asian Face Image Database PF01	1,751 images of 103 people individuals. (53 men and 50 women)	Chinese (1), Vietnamese (1), Bangladeshi (1). The remaining subjects are Korean (100).	(Dong and Gu 2001)
Korean Face Database (KFDB)	1,000 images per subject	All Korean subjects.	(Bon-Woo et al.) (Hwang et al.)
Japanese Female Facial Expression Database (JAFFE)	213 images of 10 Japanese female models.	All Japanese subjects.	(Lyons et al. 2014)
The PUT Face Database	9,971 images of 100 individuals.	All European subjects (Polish).	(Kasinski et al. 2008)
Iranian Face Database	Over 3,600 images of 616 individuals.	All Iranian subjects.	(Bastanfard et al. 2007)

Indian Movie Face Database	34,512 images of 100 Indian actors. (67 male and 33 female actors)	All South Asian subjects(Indian).	(Setty et al.)
North-East Indian Face Database **(NEI)**	27,740 images of 292 individuals. (152 male, 140 female)	All South Asian subjects (Indian). Multi ethnic subjects from 7 states in India; Runachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura.	(Saha et al. 2012)
Hajj and Umrah Facial Recognition Dataset **(HUFRD)**	Unknown (All male)	Multi-racial and multi ethnic. (Over 25 countries)	(Aly 2012)
FEI Face Database	2,800 images, 14 images for each of 200 individuals.	All Brazilian subjects. (Latin ethnicity)	(Thomaz and Giraldi 2010)
Senthilkumar Face Database	80 images from 5 individuals. (All male).	All South subjects. Dravidian ethnic group (Telegu).	(Senthil face database 2005)
Ethnicity, Gender and Age Face Database.	2,345 images of 469 subjects.	Five ethnicities: African-American (53), Asian (111), Caucasian (162) Indian (75), and Latinos (68).	(Riccio et al.)
Facial Recognition Technology Database (FERET)	14,126 images of 1,199 individuals. (and 365 duplicates)	Heterogeneous with respect to ethnicity/ age and gender.	(Phillips et al. 1998)

The MR2: A multi-racial, mega-resolution database of facial stimuli.	Images of 74 individuals. (41 female, 33 male)	All subjects were classed as either African, European or Asian as a perceived race.	(Strohminger et al. 2016)

Table 2:1 A list of published facial image databases developed either to be used exclusively by Machine Learning algorithms for racial and/or ethnic classification or, have assisted with evaluating the performance of Facial Recognition Algorithms.

When investigating racial and ethnic classification, particularly within a Machine Learning framework, there is an obvious absence of data relating to the Pakistani ethnicity, especially data which is balanced in terms of gender. With the aim to fill the void the core of the database was to generate high resolution images of British (British in the sense that they were born in the UK, hence nationals) Pakistani Males and Females amongst other racial and ethnically defined participants. A dataset of this sort would provide accurate ethnic classification of people from a Pakistani origin and would be a bench-mark for ethnic classification. The applicability of database also lends itself to human based experiments, since the controlled set of stimuli may be used to investigate strategies of human identification or unfamiliar face matching and Own Race Bias (ORB) for example.

2.4   Advantages of Ethnicity Verification and the PFDB

Given the significance of faces, stimuli collection is important for the progression of ethnicity verification as a soft biometric. The attributes of a person such as ethnicity, has a range of applications, for example,

automated recognition systems where such demographic information can be matched against a range of pre-enrolled candidates. Essentially, the by-product of such a 'filtered search' would see an overall reduction in computational search time. Additionally, a level of an understandable interpretation from a human's perspective, especially when humans use obvious characteristics to verbally explain themselves or another person, e.g. "he has black curly hair in-style of an afro".

Thus, the benefit of studying ethnicity a soft biometric feature is its descriptive nature and semantic description. Something which is proved by the 18+1 Self-Defined Ethnicity framework, utilsied by Police forces nationwide. Further benefits of ethnicity verification from facial images is consent free acquisition, the fact that such demographic information can be attained without prior consent or physical subject involvement. An obvious associated advantage when considering ethnicity verification, is the generation of taxonomy, which will facilitate both the organisation and classification of human populations, especially fine-grain, sub-populations. Data collection shortcomings are accepted which is why the exact number of individual's belongings to a specific racial and ethnic group is not known, but also why diverse and inclusive datasets are needed.

Table 2 of Section 2.4 displays a list of currently published face image databases which are specific for either a racial or ethnic group, subject to how the researchers have labelled the images. The most closely related datasets to the Pakistani Face Database (PFDB) in terms of the

geographical aspect i.e. Southern region of the Asian continent, is the Indian Movie Face Database (Setty et al.), North-East Indian (NEI) Face Database (Saha et al. 2012) and the Senthilkumar Face Database (Senthil face database 2005).

A preliminary aim of the present research was to create a face database which is both rich in ethnic diversity and high in ecological validity. There is a need for an open-access, multi-racial and multi-ethnicity face database, which includes subjects of both genders. To maximise ecological validity and emulate naturalistic settings (e.g. an international airport), items of religious dress (e.g. headscarf) were not removed. Moreover, many of the female subjects were wearing make-up. However, objects without ethnicity-specific associations, on the other hand, were removed such as spectacles, lanyards, high collar jackets and scarfs. Presently, there is a lack in dataset which consists of Pakistani face images, either as a pure ethnic set or in addition to other ethnic and racial participants. This lack of data is the void which the Pakistani Face Database (PFDB) completes, especially since the world is diverse and progressive. A key point of consideration for the database, is that the participants are photographed in the city of Bradford, although the dataset is not representative of the city but more so, it's Pakistani population.

2.5 Pakistani Face Database

The Pakistani Face Database (PFDB) is a visual face image database, which has been predominantly generated at the University of Bradford, United Kingdom. A total of 463 students (280 male and 183 female) from the University of Bradford consented to have their photograph taken. The database is made up of 224 Pakistani and 239 Non-Pakistani participants. In total, 5 photographs were taken per subject using a capture system called Halo. The database was developed in November 2015 and included 137 subjects with a total image count of 685. In 2016, a further 79 participants were photographed (5 images per subject) while in 2017, the database again increased by another 80 participants (5 images per subject). A total of 1,480 facial images have been collected from 296 participants. Further, 167 front-facing images were captured of participants who only consented to having a single image taken. The Pakistani Face Database consists a total of 1,647 images.

Details surrounding participant recruitment and digital processing is discussed in Section 2.6.3 of this Chapter. Since an aspect of the upcoming ethnicity classification framework, requires human based experimentation, which examines novice participant's discrimination ability. To quantify the contributions of different face features to judgements of ethnicity, for the human based experiments, the full-face images were then manipulated to create images of individual components (eyes, nose and mouth). To do this, the full-face images were cropped using Adobe Photoshop (CS6 Extended,

version 13.0 x 64). A cropping template was created, within which each image was aligned and manually cropped to factor-in individual head shape and headscarf type. The greyscale full face images in addition to the internal feature crops is considered a separate database.

2.5.1   The Capture System

A high resolution multi-image capture system was used to photograph the faces of all consented participants. Halo has been designed and manufactured by Acumé Forensic. While the development is targeted for law enforcement officials within a custody environment. This is the first time it has been applied to the field of academic research. The camera to subject distance is an important determinant of the consistency of custody images. Images which are taken too close to the subject may distort individual features. While, long distances significantly reduce image resolution. Halo controls the image distance by maintaining a distance of 80cm from the center panel. This fixed distance optimizes the consistency of custody images. Each of the three panels houses a single machine vision camera. Six LED light sources (two per panel) provide a consistent and bright light source. In line with the national protocol for custody image capture, this is a balanced 3-point light source. This enables the Halo to produce images which are free from shadows, across a wide range of viewpoints, see Figure 2.1.

Figure 2:1: A set of 5 sample images taken using Halo, demonstrating the viewpoints under which images are captured. For illustration purpose the images have been processed using Adobe Photoshop and the background has been removed.

The process of capturing an image requires the participant to face the central panel. The Halo then utilizes three cameras (left, center and right panel) to simultaneously capture an image of the subject from each of the frontal, 45° left and 45° right viewpoints. To obtain profile (90°) images, the subject is guided to rotated their face towards the right and left panels, respectively.

2.5.2   Participants

Participants were recruited to partake in the image capture sessions by advertisements which were posted in the School of Engineering and Informatics. This was to ensure a higher number of participants could agree to assist, since the Halo set-up was in a room in the same building. Each participant was given an information sheet to read, prior to being given the opportunity to ask any questions before their photograph was taken. The information sheet had a brief outline of the session, in addition to points

relating to associated risks, potential benefits, protection of confidentiality, voluntary participation and contact information, should there be any complaint. Participants were requested to sign a consent form which also allowed them to provide consent to having their photograph be used as part of any published work. Table 2.2 provides a breakdown of participant's demographic information i.e. ethnicity.

Gender	Assigned Race/ Ethnicity	Number	Head Covering
Male	Pakistani	137	N/A
Female	Pakistani	87	47
Male	Bengali	11	N/A
Female	Bengali	5	5
Male	Gujarati	4	N/A
Female	Gujarati	4	4
Male	White	62	N/A
Female	White	37	N/A
Male	Nigerian	22	N/A
Female	Nigerian	23	1
Male	Kurdish	4	N/A
Female	Kurdish	4	2
Male	Arab	5	N/A
Female	Arab	1	1
Male	Egyptian	8	N/A
Female	Egyptian	11	11
Male	Baganda	6	N/A
Female	Baganda	2	N/A
Male	Indian	2	N/A
Male	Han Chinese	7	N/A
Female	Han Chinese	2	N/A
Male	Japanese	1	N/A
Male	Greek Cypriot	6	N/A
Male	Black British	2	N/A
Female	Black British	1	N/A
Male	Poles	2	N/A
Female	Poles	6	N/A
Male	South Slavic	1	N/A

Table 2:2 A Table showing the demographic information of participants in the Pakistani Face Database (PFDB). The ethnicity was self-defined by the participant, although the was a strict criterion for participants of Pakistani heritage, as both maternal and paternal side were required to be of Pakistani heritage.

2.5.3 Image Collection

Participants were assigned time slots during which they would sign the consent form and then have their photograph taken. While additional personal information was not required, participants did provide their names when signing for consent as well as their self-defined ethnicity. Participants were requested to remove any items of clothing which obstructed the view of their face including their neck. Any hooded items were either removed or lowered. Those who wore a headscarf were requested to attend their session wearing a dark, non-patterned headscarf. However, some participants did not follow the instruction. Instead of turning them away, their image was still photographed. It was requested that make-up be kept to a minimum, this again was a choice of the participant and those who did not adhere to the instruction were not turned away.

Participants who wore glasses were requested to remove them during the session. Participants were requested to not wear any jewellery/accessories from the collarbone above during their session. For participants who had either open hair or hair which swept across the face. They were requested to either sweep it away from their face to ensure the periocular region was visible, or, they were requested to put it behind the ear. During the process of photography participants were requested to keep a neutral facial expression and to look directly into the camera ahead with minimum movement. A foot marker was mapped out onto the floor to ensure the participants stood at the correct distance from the centre camera panel.

2.5.4 Image Processing

All the raw captured facial images inherently contained redundant information, which was not required such as the upper body and the background. As a result, all the facial images went through a rigorous 'clean up' procedure. All the images were cropped around the face and neck, and effort was made to reduce any stray hair and remove earrings, if they were visible. Participants who wore a headscarf also went through the process of cropping, whereby the image was cropped in line with the drape of the headscarf, see Figure 2.2. This ensured that the image did not look deliberately manipulated. The headscarves of participants who did not wear a dark, non-patterned headscarf were also digitally processed. Moreover, blemishes such as moles or acne, for example were processed to prevent them being used as identification cues during the human experiments. Finally, all the images were duplicated and converted to greyscale, so they could be used for the human experiments. Whereas the original cropped coloured images were used during the Machine Learning experiments.

To ensure a consistent set of data used in both the Machine Learning based experimentation and the human-based discrimination ethnicity classification task. A total of 200 images, split equally in gender and ethnicity (100 per ethnic class and 50 per gender) were selected from the PFDB. To ensure the database was not homogenous images of multiple Non-Pakistani participants were also captured and consisted of Caucasian

(including Irish), East/Central Asian (Chinese and Japanese) Arab, and

Black (Nigerian) participants.

Figure 2:2: **Above:** Post-processed images using Adobe Photoshop CS6. Images A-C show the process from; (A) initial image, (B) background removal and, (C) conversion to greyscale. **Below:** Images showing the removal of blemishes from the nose and cheek of the male, while the female image has been processed to remove accessories from the ears and nose, in addition to a blemish removal.

2.5.5   Image Storage

During each photography session, the initials of each participant i.e. SJ, were assigned to a single folder which stored the images. All the images contained within the folder were automatically assigned the following labels and stored as Jpg files:

- Front_Face
- Left_45_Degree
- Left_Profile
- Right_45_Degree
- Right_Profile

At the end of each image session, all the images were transferred from the laptop linked to the Halo, onto an encrypted hard-drive.

2.6 Conclusion

The Pakistani Face Database (PFDB) is the first known example of a facial image database which includes participants of the Pakistani origin amongst other ethnicities. The images captured are realistic, well illuminated and high in racial and ethnic diversity. More importantly, the images captured are ecological valid and emulate realistic settings (i.e. international airports), since images of females for example, are wearing headscarfs of varied styles. The controlled environment within which the images have been photographed ensure their usability. Especially since the images are not limited by extraneous variables such as lighting or facial expression. Ultimately, a human face is informative and the PFDB presents this point efficiently. The remainder of the thesis discusses the use of the captured images to investigate performance of ethnicity verification between humans and Machine Learning algorithms.

CHAPTER 3

LOW-LEVEL HAND-CRAFTED FEATURES FOR ETHNICITY

3.1 Introduction

The human face includes a range of cues which form part of an individual's identity. Accordingly, faces are important for recognition and identification of specific individuals. In the context of a crime, this presents the following important challenge: how can the perpetrator's face be accurately extracted and described from the memory of a witness? One approach is to rely upon an artist to replicate the face based on verbal descriptions from witnesses. Having a witness verbally describe the complex spatial configuration of the perpetrator's face, often after a period, limits the amount of face information which can be captured within a schematic representation.

In recent years, the use of demographic information i.e. ethnicity, for facial image classification, within the Machine Learning framework, has drawn increased research attention. Ethnicity is a variable visual trait of the human face, which is measurable, something which has been demonstrated in the literature presented as part of the introduction (Chapter 1). Reasons for studying facial-image based ethnicity classification, include the wide-ranging practical applications associated with such type of work. For example, (i) Policing (law enforcement), which would ensure accurate classification of unknown subjects without an element of bias, (ii) automated

facial recognition systems, which may utilise ethnicity as a pre-assessment criterion for wanted criminal searchers.

The process of ethnicity classification involves a class categorization problem which requires the identification of the ethnic group a facial image belongs to. To do so, specific hand-crafted features i.e. facial landmarks, of the face are manually selected or, in automated cases, pre-trained, Machine Learning models are used for feature extraction.

While significant advances have been made in feature extraction from front view facial images for ethnicity assignment (Hosoi et al.; Jilani et al.; Lu and Jain; Riccio et al.). Ethnicity classification based on profile images, however, is limited to profile silhouettes (Tariq et al.). Further, there has not been a lot of research considering the South Asian population, specifically those belonging to the Pakistani heritage, hence there is a paucity of evidence regarding the Pakistani face. To bridge this gap, a Machine Learning based framework is proposed for the binary, Pakistani vs Non-Pakistani, classification task. To our knowledge, this is the first attempt to incorporate profile facial images for the classification of Pakistani ethnicity.

The upcoming Chapter reports on two separate experiments which use hand-crafted face features to assist image-based ethnicity classification. A total of 26 features are manually annotated onto the facial images; 16 for the frontal image, and 10 or the profile image. The hand-crafted feature enable feature extraction, and are used to represent the

most enriched features. To test the enriched nature of the feature vectors, an unsupervised pre-processing algorithm has been used called Principal Component Analysis (PCA), in addition to a supervised model; Partial Least Squares (PLS). Both models are used independently of one-another for dimensionality reduction. A robust binary classifier, i.e. Support Vector Machine (SVM), is implemented for the binary classification challenge. The experimental objective is to determine whether low-level, hand-crafted features are robust markers of Pakistani ethnicity, when classifying the ethnic origin of front facing images, in addition to restrictive profiles.

The Chapter is organised as follows; a brief collection of literature is presented on the topic of ethnicity classification from frontal and profile images in Section 3.2. In Section 3.3 the experimental stimuli are discussed, and in Section 3.4 the experimental methodology is presented. The results are reported and discussed in Section 3.5 and in Section 3.6 conclusions of the experimentations are drawn.

3.2 Literature Review

Anthropometric studies demonstrate that there are distinctions between feature measurements and characteristics of people from different racial and ethnic backgrounds. Leslie Farkas and colleagues carried out a comprehensive anthropometric study of facial morphology and facial parameters comparing 14 normative measurements of the face across

ethnic and racial groups. Several differences between groups were reported (Farkas et al. 2005).

In the field of Machine Leaning there is a wide-ranging spectrum of research on face-based ethnicity classification, and the area of research continues to grow. Research conducted by Duan et al (2010), investigated the task of ethnicity classification using images from 3 ethnic groups residing in China; Tibetan, Uighur and Zhuang. All three of the human groups are known to have obvious distinctions in their facial make-up, and differences are also reported between hair texture as well as the shape of their eyes, face, nose and mouth (Duan et al.). Images underwent pre-processing since backgrounds were removed from the images and pose correction was conducted, to ensure the eyes were aligned. Geometric features relating to nose width, face width, right eye width, in addition to crossover measurements such as face width to the nasal alar, were extracted. The measurements highlighted obvious differences in the facial make-up of the three investigated groups, and assisted in generating facial templates for the three ethnic groups (Tibetan, Uighur and Zhuang). Ethnicity classification was carried out using facial recognition technology and the results highlight the efficiency of geometric features as well as the descriptive nature of the human face; 88.6% for Tibetan, 90.0%, for Uighur and 94.3% for the Zhuang group.

More recently Becerra-Riera et al.(2018) combined local appearance and geometric features to estimate race, using images from the

Ethnicity Gender and Age (EGA) Database. Face images were divided into 10 areas; hair, face, contour, forehead, eyebrows, eyes, nose, cheeks, mouth and chin. To gather appearance based information, 17 filters were used to extract color and texture information which could be exploited to ascertain racial origin. Per face region, a total of 34-component descriptors processed the area and produced a mean and variance value, which, when concatenated best represented that area of the face. Thus, a total of 34 descriptors were used per image. In contrast, 68 facial landmarks were used to represent geometric features; 17 points for the face contour, 12 for the eyes, 10 for the eyebrows, 9 related to the nose and 20 to the mouth, and ratios between the distances was calculated. For classification, a Support Vector Machine (SVM) and a Forest Random (FR) classifier were employed, and a combined performance accuracy (i.e. local appearance and geometric feature approach) of race estimation was as follows; 90.9% for African, 87.5% for Asian, 90.4% for Caucasian and 87.5% and 74.7% for the Indian and Latin race, respectively.

Muhammad et al. (2012) investigated the use of Local Descriptors for the racial classification of 5 distinct classes: Asian, Black, Hispanic, Middle-Eastern and White. Two types of Local Descriptors were used, (i) Local Binary Patterns (LBP) and (ii) Weber Local Descriptors (WLD). Since the researchers did not create their own database, they used images from the FERET database. A total of 1,188 images were used for training while 1,180 were used for the testing. The accuracy achieved for LBP was

98.42%, 95.56%, 93.65%, 100% and 98.18% for Asian, Black, Hispanic, Middle-Eastern and the White race respectively. And comparable results were achieved with WLD; 97.74%, 96.89%, 92.06%, 98.33% and 99.53% again for the Asian, Black, Hispanic, Middle-Eastern and the White race respectively.

Research into ethnicity verification from silhouetted facial profiles was first conducted by Tariq et al. The novelty was in the ethnic categorisation of the 441 images and consisted of (as described in the study): Black (African, African-American), East/South-East Asian (Chinese, Japanese, Korean, Vietnamese, Filipino, Singaporean) South-Asian (Indian, Pakistani, Sri-Lankan and Bengali) and White (Caucasian and Middle Eastern). This research is one of the few which considers South Asian as an investigated participant demographic, although the dataset used is not big. While the accurate ethnic assignment from the profile silhouettes averaged at 71.66%, it is important to mention that the literature did not focus specifically on the Pakistani heritage, even though a few participants were reported to be Pakistani. Tin and Sein, (2011) investigated a two-class, fine-race classification task between Myanmar and Non-Myanmar face images. In brief, a total of 250 Images were collected from the internet and the results concluded were promising (on average 94%) given the small dataset (Tin and Sein 2011).

Having reviewed some of the literature on ethnic and racial classification from face images, it is obvious that researchers typically rely

on previously published databases. This may be due to the limited demographic variability or restrictions which mean researchers are forced to used internet derived images. Further, it is evident that data surrounding the South Asian race, but specifically participants of the Pakistani heritage remains unchallenged within an ethnicity classification framework. Moreover, there is a requirement for a criterion-specific database, which consists of high resolution, multi-racial and multi-ethnic participants, especially of the Pakistani heritage.

3.3 Dataset

A total of 270 images (from 135 participants of which 63 identified as Pakistani and 72 of Non-Pakistani heritage) were selected from PFDB (see Chapter 2). Of the 270 images, 135 were of the frontal view and 135 were of the left profile view. To reduce computation time the original dimensions were too large (4000 x 6016 pixel) and therefore were cropped to 1500 x 2100 pixels. The face images portrayed neutral facial expression and full visibility of the face was ensured, during the photography session.

3.4 Methodology

3.4.1 Hand-Crafted Features

Hand-crafted facial features are manually mapped onto the boundaries of the selected face regions. A variety of landmarks was selected for experiment 1, which involves front-facing images, and for experiment 2,

which made use of profile images. For example, the inner and outer eye corners of the eyes, i.e. Endocanthion and Exocanthion, as well as the right and left Nasal Alars (nostril) were mapped on the frontal images. Whereas, for experiment 2, which used profile images, only the Pronasale i.e. the tip of the nose, and the Subnasale (the most inferior region of the Nasal Septum) were selected. Importantly, published literature has shown that there are no set criteria for annotating facial landmarks. Especially since the number of landmarks are dependent on the type of analysis being conducted. Literature on 3D facial recognition for example, suggests that anything between 11-59 facial points on the face will be sufficient (Hutton 2004), while other literature have reported the use of in excess of 100 facial landmarks (Jain and Park). See Wu and Ji (2019) for a comprehensive literature review on facial landmark detection.

Face landmarking is a common method used for localising characteristic regions of the face and have been applied to other face-related work such as facial expression and face ageing (Elmahmudi and Ugail 2020). Consisting of two categories; (i) primary/ fiducial (e.g. corner of the eyes and mouth) and (ii) secondary (e.g. nostrils), which are often guided by the primary landmarks (Çeliktutan et al. 2013), a combination of landmark types were implemented for the upcoming experiment.

A combined total of 26 facial landmarks were selected, 16 from front facing images and 10 from profile images, see Figure 3.1. The rationale behind selecting such a small number of landmarks is because

anthropometric studies already report on the informative nature of regions of the face which are robust markers for discrimination, for example; forehead, eyes, nose etc., moreover, by choosing a small set of appropriate features, there will be a reduction in computation time (Tang et al. 2014).

Figure 3:1: An illustration to the manually placed hand-crafted face features used in experiment 1 (frontal view images) and experiment 2 (profile view image). A combined total of 26 landmarks were used, 16 were mapped onto frontal images, while 10 hand-crafted features were selected for the face-profile images.

Properties of the detected features plus their relation to one another (such as distances and angles) are used as the primary descriptors of the face. This method of feature selection is a pre-processing step and removes distracting variance from a dataset, enabling the classifier to perform efficiently. Features of interest within each facial image are extracted using landmarks, which are positioned on the boundary of important components of the face. Each hand-crafted feature i.e. mapped landmark is represented

by a 2-dimensional vector, which subsequently represents both the x and y coordinates i.e. x_i, y_i (Fodor 2002).

Given an *n* set of 2-dimensional landmarks, the face can be represented by a single linear equation:

$$\textbf{\textit{Feature Points}} = (x_1, x_2,...., x_n, \ y_1, \ y_2,..., y_n)^T \qquad\qquad (3\text{-}1)$$

Where *n* represents the total number of facial landmarks ($n = 16$ or $n = 10$) the corresponding *x* and *y* values, $x1 - x_n$ and $y1 - y_n$ respectively. The superscript 'T' denotes the transpose function which in brief, switches rows to columns and columns to rows. For example, a 3 x 2 matrix transposed becomes a 2 x 3 matrix.

Each of the measurements for the frontal images are represented in a 16 x 2 matrix which represent the x *and* y co-ordinates of the 16 landmarks. Whereas, for the profile images the x *and* y co-ordinates are represented in a 10 x 2 matrix. After the landmarks are mapped a total of 16 values are outputted from each front face image and 10 values are generated from a single profile image. Ultimately this form of feature extraction will allow for the use of uncorrelated hand-crafted features, which do not represent more than one feature point on a given facial image, and will highlight features with the highest degree of variance. A brief description of the facial landmarks used as part of this study is shown in Table 3.1.

Landmark	Description	Region
Trichion *(tr)*	Mid-point of the hairline.	Cranial
Glabella *(g)*	Most prominent part between supraorbital ridge.	Cranial
Nasion *(n)*	A point in the midline of the nasal root.	Face
Exocanthion (right & left) *(ex)*	A point at the inner commissure of the eye fissure.	Face
Endocanthion (right and left) *(en)*	A point at the outer commissure of the eye fissure.	
Zygion (right & left) *(zy)*	Most lateral point of the zygomatic arch.	Face
Interpupillary Distance *(ID)*	Distance between the centre of each pupil.	Eyes
Subnasale *(sn)*	Mid-point of the pronasale (tip of the nose) where the lower border of the nasal septum and the philtrum (upper lip) meet.	Face
Chelion (right & left) *(ch)*	Located at each labial commissure (corner of the mouth).	Orolabiale
Alare (right & left) *(al)*	Most lateral point on each alar (nostril) contour.	Nose
Pronasale *(prn)*	Tip of the nose.	Nose
Labiale Inferius *(li)* **Labiale Superius** *(ls)*	Mid-point of lower vermilion line. Mid-point of upper vermilion line.	Orolabiale
Stomion *(sto)*	Mid-point of the labiale fissue.	Orolabiale
Sublabiale *(sl)*	Mid-point of the labiomental-sulcus.	Face
Pogonion (*pog*)	Anterior mid-point of the chin.	Face

Table 3:1 Facial landmarks selected for anatomical distinctiveness as hand – crafted features for the classification of ethnicity, using front and profile face images from the Pakistani Face Database (PFDB).

3.4.2 Dimensionality Reduction

To discard redundant data and retain the most significant features that capture ethnic variations amongst each facial image, two different dimensionality reduction techniques were used. Feature compression was achieved by two methods: Principal Component Analysis (PCA) (Fodor 2002) and Partial Least Square Regression (PLS) (Dormann et al. 2013). For experiment 1 which considered frontal images only, Principal Component Analysis (PCA) was used. In experiment 2, which considered profile images both Principal Component Analysis (PCA) and Partial Least Square regression (PLS) were used independently of one another to remove redundant data.

Principal Component Analysis is an unsupervised algorithm which seeks to reduce the dimensionality of the data by locating the Principal Components (PCs) of the original variables, along which the variation of the data is maximal. By identifying a reduced number of variables, the data can be analysed efficiently (Abdi and Williams 2010). Each data point can be divided into eigenvectors (Principal Components, k) which have a corresponding eigenvalue. The eigenvector determines the direction of the new feature space, while the eigenvalue highlights the amount of variance i.e. the magnitude and how spread out the data is on an x and y axis. Ultimately, the number of eigenvectors and eigenvalues corresponds with the number of dimensions in the data. By implementing the PCA model, the number of Principal Components, k, were chosen by selecting the

eigenvectors that accounted for 95% of the variance. Similarly, 10 Principal Components, k, described 96% variance for experiment 2.

Partial Least Square regression (PLS) is a supervised dimensionality reduction algorithm that captures the variation that exists between the dependent and independent variables (Chierchia et al.). PLS is more powerful in its regression abilities because it searches for components that capture the highest degree of variance in x, in addition to the direction which best describes x and y. Whereas, PCA works to find the direction of highest variance only in x (Fodor 2002; Bukar et al. 2016).

3.4.3 Ethnicity Classification

Having used hand-crafted features to highlight the pertinent areas of the face, followed by the subsequent reduction in the dimensionality, a Linear Support Vector Machine (SVMs) is employed for binary data classification. Support Vector Machine (SVMs) are supervised Machine Learning models that function to identify a hyperplane, which best classifies data points within a given data space. Previous studies have demonstrated that SVMs is a powerful binary classifier and operates by defining an Optimum Separating Hyperplane (OSH) between two classes of data (Chierchia et al.; Nalavade and Meshram).

Given a training set, the input to the SVM algorithm is a set of trained labelled data $\{(x_i, y_i)\}$ where, x_i is the data and y_i is the class membership i.e. the label; $+1$ or -1. In the context of the experiments

described within this chapter, the label $+1$ denotes a Pakistani image and the label -1 denotes a Non-Pakistani image.

3.4.4 Model Evaluation

A k-fold cross validation technique was employed to evaluate the performance of the learning system, where 90% of the data was used for training and the remaining 10% for testing. During the process of k-fold cross validation, the data is randomly separated into 10 partitions of equal size. Nine of the partitions are used to train the model, and the remaining partition is used for testing. This is repeated until each partition has been used to test the model. The results are then averaged and analysed to measure performance. We have used k-fold cross validation where k=10: which means the value for k is fixed to 10. The value is identified through experimentation to generally result in a model skill estimate with low bias a modest variance. Importantly k-fold cross-validation was not manually implemented, instead the scikit-learn library was used, which provides an application that will split a given data sample up. We created an instance that splits the dataset into 3 folds, shuffles prior to the split, and uses a value of 1 for the pseudorandom number generator. This is where the k-fold cross-validation procedure is repeated n times, where importantly, the data sample is shuffled prior to each repetition, which results in a different split of the sample.

3.5 Experimental Results

Table 3.2 reports the ethnicity classification accuracies for the use of PCA and a Linear SVM for frontal and profile images. The classification rates are computed as percentages of the total number of test images that are accurately classified as Pakistani from within the dataset.

Experimental Stimuli	SVM Classification Rate with PCA (%)
Frontal Images	65.71%
Profile Images	71.42%

Table 3:2: Linear Support Vector Machine (SVM) binary classification (%) of Pakistani ethnicity using Principal Component Analysis (PCA) for **(i)** front face images and **(ii)** profile images.

It is evident from Table 3.2 that the ethnicity classification results attained using PCA (experiment 1) were above chance level (>50%) for front view, whole-face images; 65.71 %. In contrast, a higher ethnicity classification accuracy is achieved when using profile images (71.42%), which is surprising considering the limited facial data. As, the profile face image reported a higher ethnicity classification using PCA (unsupervised model), it was considered a logical step to test whether, by implementing a supervised model, for redundant data removal, a higher performance accuracy would be achieved. Therefore, experiment 2 used Partial Least Square (PLS) reduction algorithm especially since published reports

highlight the reliability of PLS in classifying facial images by demographic information i.e. race/ethnicity (Guo and Mu; Phillips et al. 2011). Table 3.3 shows the results achieved for using both the reduction techniques on the dataset of 135 profile images.

	SVM Classification Rate (%) with PCA	SVM Classification Rate (%) with PLS
Profile Images	71.42%	**76.03%**

Table 3:3: Linear Support Vector Machine (SVM) binary classification (%) of Pakistani ethnicity from profile face images using an unsupervised model: Principal Component Analysis (PCA), and a supervised model: Partial Least Square (PLS).

To evaluate the accuracy of the ethnicity classification algorithm, Receiver Operating Characteristics (ROC) Curve was generated. The curve provides a graphical comparison of the true positive rates (TPR) and the false positive rates (FPR) at different thresholds. An ROC is produced to demonstrate the performance of the binary classifier and to evaluate its accuracy (Lever et al. 2016). The straight line which starts at the co-ordinates of (0,0) and ends at the co-ordinates of (1,1) indicates a random classifier, which acts as an invisible baseline. A key advantage of using ROC curve analysis is that the optimal point of interest i.e. the best trade-off between sensitivity and specificity, can be chosen for the investigated variable. According to the ROC curve (Figure 3.2) the experiment on binary ethnicity classification from profile images, which used PLS, outperformed the experiment with

PCA, demonstrating its strength as an algorithm. This is visually illustrated by the position of the blue PLS curve, which is shifted to the left of the red curve for PCA.
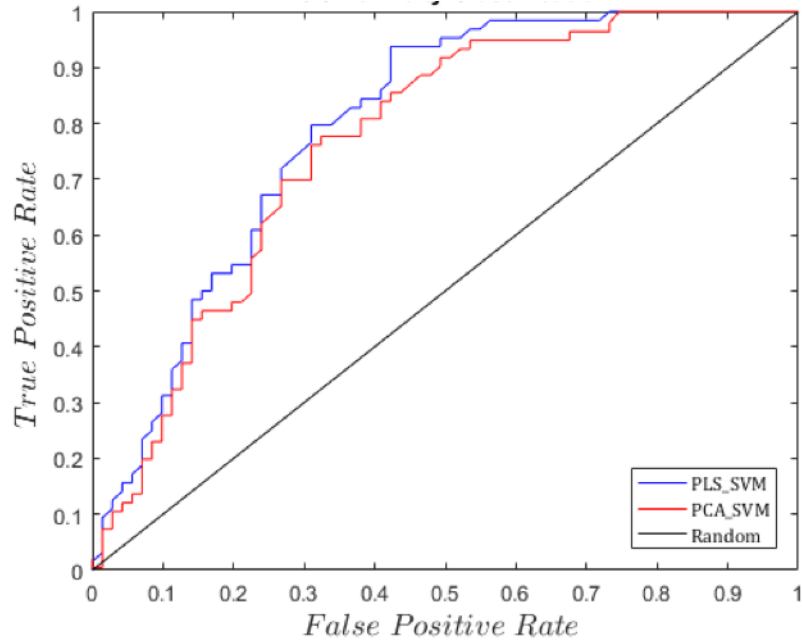
Figure 3:2: Receiver Operating Characteristics Curve (ROC) for the binary classification of the Pakistani profile using PLS and PCA features.

The results show that the Pakistani facial profile is distinguishable from other ethnic facial profiles within the database. It is presumed that the differences are inherent in the measurements between each hand-crafted feature. The reason for this belief is because all the images within the database were annotated using the same landmarks and therefore any variation is presumed to be a direct result of the measurements. This is particularly important as published literature on facial anthropometry,

already reports distinctions between sub-groups of human populations. While the results presented are based on experiments using low-level features, they are the first of its kind considering the research question posed, therefore it is not possible to conduct comparative analyses with other published results. While the results are not significantly high, they are considered promising and do illustrate that the ethnic classification of the Pakistani face is achievable, even from a profile view face. Further, the results demonstrate that low-level face features do hold sufficient, discriminative information to allow the accurate categorization of a Pakistani face. There are however, evident drawbacks to the applied framework which relate to the repeatability of the results and the small dataset, which affect the accuracy of performance. Dealing with the issue of repeatability, each of the hand-crafted features were selected by a single user, and although the landmarks have predefined locations on the face, it is possible that when annotating the face images there may be slight discrepancies in the placement of the hand-crafted feature. This suggests that there is a chance for error and possibly lower results, since the process is manual and not automated. Further, the individual mapping of the low-level features is time consuming task which further makes the task prone to error, especially due to fatigue. With regards to the dataset, it is relatively small (270 images only) and may be considered restrictive when considering the varying characteristics of the face. Hence, this is a known limitation of the current experimental framework.

3.6 Conclusion

A binary ethnicity classification challenge has been attempted using images from the PFDB of front and profile images of 135 participants. To retrieve the most useful ethnic markers, low-level, hand-crafted features were used to extract discriminatory features. While a combined total of 26 features were used, i.e.16 features on frontal images and 10 features for the profile face image. The low-level features were selected for their anatomical position on the face as well as their ecological validity, especially when considering that the region of the frontal face and view of the profile face, are typically examined during forensic facial image analysis.

Following on from feature extraction, the removal of redundant information was carried out using an unsupervised (PCA) and a supervised (PLS) algorithm. Since the ethnicity classification task is binary, the obvious choice of classifier was a Linear Support Vector Machine (SVMs). The highest reported ethnicity classification within the current Machine Learning framework, was using PLS for dimensionality reduction on profile view face images. A reported classification of 76.03% was achieved. The result demonstrates that low-level features do hold informative ethnic data. Given the inherent limitations, the present study is considered promising and is used as a benchmark for upcoming experiments. With an aim to attain better accuracy signature for image-based ethnicity classification, a natural progression to the use of Deep Learning (a sub-set of Machine Learning) is warranted. Given the numerous publications which report on the state-of-

the-art-classification Deep Learning achieves, it is hypothesized that by implementing Deep Learning, the performance accuracy for Pakistani ethnicity will grow significantly. The upcoming chapter aims to explore the role of high-level face features within a Deep Learning framework, which incorporates, augmented, criterion specific datasets. Chapter 4 presents a Deep Learning specific framework for the task of binary ethnicity classification, based on the full face in addition to testing the role of individual face components, i.e. eyes, nose and the mouth. It is hypothesized to be both robust, efficient and generate higher accuracy.

CHAPTER 4

DEEP-FACE FEATURES FOR ETHNICITY VERIFICATION

4.1   Introduction

The classification of ethnicity from facial images has gained significant popularity within Machine Learning. The motivation to categorize ethnicity from images of the face and its components i.e. eyes, nose and mouth stems from the role ethnicity plays in technology with applications in face-related biometric systems such as iris recognition (Gangwar and Joshi; Dua et al. 2019). Ethnicity is a stand-alone, variable trait (Fu et al. 2014), and within the Machine Learning arena demographic (ethnicity), classification has been reported for a spectrum of face datasets. For example, whole-face images (Hosoi et al.; Jilani et al.; Lu and Jain; Riccio et al.), face profiles (Jilani et al.) including silhouetted face profiles (Tariq et al.), and more recently with the use of 3D Facial Landmark data in Kendall Shape Space (Lv et al. 2020). In contrast, researchers have also used isolated face components namely the eyes (Mohammad and Al-Ani; Qiu et al.) and nose (Chang et al. 2006; Song et al. 2009), to address the problem of ethnicity classification.

Low-level, hand-crafted features have proven reliable as ethnic markers for Pakistani face images, as reported in Chapter 3. As low as 10 hand-crafted from coloured, profile images, are sufficient for image-based, ethnicity classification, of the Pakistani origin. Importantly, the results are

higher compared to previously published literature by Tariq et.al, who used profile silhouettes. However, manually annotating images is time-consuming, subjective (with regards to placement) and maybe subject to varied results based on the testing criteria as previously demonstrated, i.e. profile images give a higher accuracy rate compared to frontal images. With these drawbacks in mind, the decision to adopt an automated Deep Learning approach for feature extraction is justified.

Pre-trained Deep Learning models are known to achieve high classification rates; thus, this Chapter discusses a series of experiments, which implement 6 pre-trained Deep Learning algorithms (VGG-Face, VGG-16, VGG-19, ResNet-50, ResNet-101 and ResNet-152) on high-level full and partial-face data. The proposed approach for the experimental framework comprises of 2 components; (i) feature extraction using the weights of 3 Residual Learning models (ResNet-50, ResNet-101, ResNet-152), and 3 Convolutional Neural Networks namely, VGG-Face, VGG-16 and VGG-19, to learn information (ii) demographic (ethnicity) classification using Linear Support Vector Machine (SVM) algorithm as a binary, two-class technique. The goal is to identify the experimental conditions within which the model's parameters are fine-tuned to yield a high ethnicity classification rate. To achieve this goal, a purpose-built dataset of images has been created.

The Chapter is organised as follows; literature related to the ethnic classification of facial images is discussed as part of Section 4.2. In Section

4.3, the experimental stimuli are discussed. Section 4.4 presents the proposed experimental framework. In Section 4.5 the experimental results are presented and Section 4.6 consists of a discussion. Finally, the Chapter is summarized in form of a conclusion in Section 4.7.

4.2 Literature Review

Ethnicity classification has been reported for the Chinese (Gao et al. 2008), Japanese (Bastanfard et al. 2007) and Korean race (Hwang et al.), however the South Asian, Pakistani ethnic group remains untried, with a single exception (Jilani et al.). Ou et al., carried out a binary classification task (Asian Vs Non-Asian class) using frontal face images. Real-time analysis was achieved using images from uncontrolled environments. Principal Component Analysis (PCA) was used to obtain the most variant features and a novel "321" algorithm was combined with a Support Vector Machine (SVM) to boost classification. The researchers reported an 82.5% classification accuracy for a database of 750 face images taken from The Facial Recognition Technology Database (FERET). In contrast, Hosoi et al. investigated a three-class (Asian Vs African Vs European) ethnicity classification task. From 1,991 facial images, key features were extracted by a combination of Gabor Wavelets Transformation (GWT) and retina sampling techniques. Since the classification was not binary, the Support Vector Machine used was expanded to a multi-clustering classifier. Based on the applied technique the estimation accuracy achieved was as follows: Asian: 96.3%, European: 93.1% and African: 94.3%. More recently, Han et

al. (2015) proposed an automatic multi-demographic (age, gender and race) framework. Biologically-Inspired Features (BIFs) were used to extract demographic informative features from facial images. And a Hierarchical classifier was applied to categorize face race. Lakshmiprabha (2016) produced a multi-modal framework for gender, ethnicity, age and expression classification. Images were collected from publically available databases such as the Japanese Female Facial Expression (JAFFE) and Facial Recognition Technology (FERET). A total of 357 images were collected and categorized into the following ethnic groups; White, Black, Indian and "Other" (consisted of Asian and Hispanic). Of the 357 images, 187 were used for testing and it was reported that the Active Appearance Model (AAM) generated the highest ethnic classification rate of 93.83%.

More recent methods of Deep Learning, ethnicity classification has been attempted by Anwar and Islam, 2017, who proposed a method of race classification using Convolutional Neural Networks (CNN) for deep feature extraction and a supervised learning algorithm (SVM) for classification. A total of 3 racial groups were tested; Asian, African-American and Caucasian and using *k*-fold cross validation the accuracies reported were above 90% at 98.28%, 99.66% and 99.05% for the Asian, African-American and Caucasian group respectively (Anwar and Islam 2017).

The literature on the use of facial images for ethnic and racial classification features a few consistent themes. Firstly, experiments have been conducted on previously published databases for example FERET.

This may be due to the limited demographic variability or participant restrictions which meant researchers resorted to images harvested from the internet. The issue here is that one is presented with a spectrum of results for various image-based classification problems, on homogenous datasets. Nonetheless it seems that the use of the FERET database remains the most common choice for researchers, possibly due to the size and variability of the dataset. However, the problem with using such an old database is that it is outdated in terms of the equipment used to capture the images, and may not reflect the most recent demographic trends. Another observation is the complexity surrounding the definitions of the terms "race" and "ethnicity". The use of the words within the Machine Learning literature remains interchangeable but, where possible, researchers have drawn an association between facial appearance and geographical origin, i.e.: Middle-Eastern. However, drawing an association between geographical origin and appearance can prove difficult, especially since migration is associated with a geographical shift, and the transition introduces the opportunity for heterogeneity through intermarriage. Nonetheless, it is a given that facial appearances vary across the world. It is apparent that data for the South Asian race is limited. Further, ethnicity verification for facial images of the Pakistani ethnicity, remains under-investigated with a single exception (Jilani et.al).

4.3 Dataset

There are relatively few databases of facial images which are specific for demographic information such as race or ethnicity. Further, due to the lack of a standardised image capture process, there is considerable variability between published databases of facial images. Images derived from the PFDB (Chapter 2) are used as experimental stimuli for the upcoming experiments, but they have been manipulated to increase the data size. Students from the University of Bradford consented to have their face photograph taken using the HALO system: a multi-view, custody image capture system (Jilani et al. 2018). All raw (pre-processed) images were labelled with the ethnicity of the participant, which was self-defined. However, for participants of Pakistani origin, eligibility was dependant on whether both the maternal and paternal parents were of Pakistani ethnicity.

4.3.1 Image Collection

Participants were requested to remove any items of clothing which obstructed the view of their face including their neck. Any hooded item was either removed or lowered. Those who wore a headscarf were requested to attend their photography session wearing a dark, non-patterned headscarf. Further, it was requested that make-up be kept to a minimum and participants who wore glasses were requested to remove them during image capture. Participants were requested to not wear any jewellery/accessories from the collarbone above during their photography

session. Also, participants with long hair were requested to sweep it away from their faces, to ensure the periocular region was visible with clear visibility of the face. During photography, participants were requested to keep a neutral facial expression and to look directly into the camera lens. A foot marker was plotted out onto the floor to ensure the participants stood at the correct distance from the central camera panel.

4.3.2   Image Pre-Processing

All captured images went through a rigorous 'clean up' process. All the images were cropped around the face and neck (with background removal), and an effort was made to reduce any stray hair and jewellery, if they were visible. Participants who wore a headscarf also went through the process of cropping, whereby the image was cropped in line with the drape of the headscarf, see Figure 4.1. The reason for this was to remove redundant image information and secondly, to ensure the proportions of all the images are uniform and closely matched as possible. Image backgrounds have been removed to ensure maximum face data is used during the process of training and testing. For the upcoming experiments a total of 4 different augmented datasets have been created from original images that are derived from PFDB and include: 1,000 full-faces, 1,000 eyes, 1,000 nose and 1,000 mouth images, see Figure 4.2 for an example of the mouth stimuli. Augmented images are artificially expanded to increase the training

dataset, and is achieved by generating modified versions of existing data i.e. rotations, flips etc.

Figure 4:1: An example of multi-ethnic and multi-racial face images from the Pakistani Face Database, that have been processed using Adobe Photoshop CS6. Images have been cropped around the face and neck and have undergone processing to remove any visible jewelry.
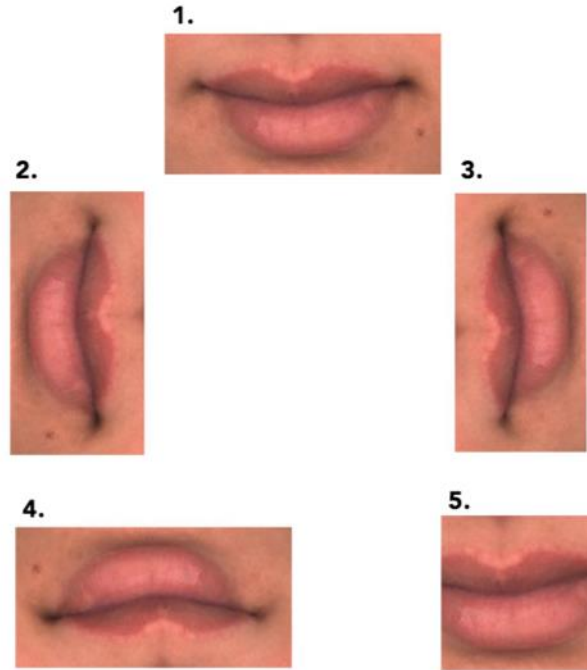
Figure 4:2: A schematic of the isolated face-feature (mouth) augmentations. The original image of the mouth has undergone 4 additional manipulations, to increase the sample size, creating 5 images per face feature. The additional augmentations include (i) 90⁰ clockwise rotation, (i) (ii) 90⁰ counter-clockwise rotation, (iii) 180⁰ rotations and, (iv) a section crop. The remaining face features (eyes and nose) were also augmented using Photoshop CS6 with the same degree of rotations.

4.4   Methodology

The experimental procedure entails extracting deep features using the 6 pre-trained Deep Learning models; ResNet models; 50/101/152 (Residual Learning based), and 3 Neural Network algorithms; VGG-F, VGG-16 and VGG-19. After feature extraction supervised learning is used to perform a binary classification using a Linear Support Vector Machine. Each Machine Learning model is used in isolation of one another for each experiment. A total of 6 experiments were conducted per feature condition i.e. 1,000 full-

face images, 1,000 isolated eye-crops, 1,000 isolated nose-crops and 1,000 isolated mouths-crops.

4.4.1   Data Pre-Processing

All images were resized to 224 × 224 pixels to ensure they conform to the input criteria of the pre-trained models.

4.4.2   Pre-Trained Model: VGG-F

The VGG-F model is a 37-layer deep Convolutional Neural Network developed by researchers from the Visual Geometry Group (VGG) at the University of Oxford (Parkhi et al.) The Machine Learning model is trained on a dataset of 2.6 million face images and achieved state-of-the-art results on benchmark datasets such as Labelled Faces in the Wild dataset (LFW) and the YouTube Faces (YTF).

4.4.3   Pre-Trained Model: VGG-16

The VGG-16 model is developed by Simonyan and Zisserman from the Visual Geometry Group (VGG) at the University of Oxford. The model consists of 16 layers (Simonyan and Zisserman 2014) and has an input criteria of RGB images at 224 × 224 image dimension. There are 13 convolutional layers with 3 × 3 size filters, 5 max-pooling layers with a 2 × 2 stride and 3 Fully-Connected (FC) layers; FC-6, (4096 channels), FC-7, (4096 channels) and FC-8, (1000 channels representative of the 1000

classes of the tested subset of the ImageNet dataset). The final layer is the soft-max layer, which functions to ascertain the probability of an outcome.

4.4.4   Pre-Trained Model: VGG-19

VGG-19 is an improved version of the VGG-16 model but with an increased depth of 43 layers, thus a deeper model. Similar to the other VGG-based models, the input data criteria is a RGB image of 224 × 224 dimension. There are 16 convolutional layers which use a 3 × 3 filter, and the convolution stride is fixed to 1-pixel. There are 3 Fully-connected layers, FC-6 and FC-7 that have 4096 channels each and FC-8 which contains 1000 channels. (Simonyan and Zisserman 2014). The final layer is the soft-max layer.

4.4.5   Residual Network

In 2015 Microsoft Research Asia developed Deep Residual Network (ResNet) (He et al.) which are deep compared to Convolutional Neural Networks, such as VGG-F. The ResNet model achieved the highest state-of-the-art performance on ILSVRC- 2015 (The ImageNet Large Scale Visual Recognition Challenge). The critical feature of the residual networks are the "skip connections" known also as "shortcuts", which function to avoid parts within the network, to get straight to the output, quicker without compromising accuracy.

4.4.6 ResNet-50

The ResNet-50 model (50-layer deep), is a deep Convolutional Neural Network (dCNN) architecture that was developed in 2015 and utilises residual learning (He et al. 2015). The model is trained on the ImageNet Dataset and can classify up to 1000 separate classes of objects such as mouse, pencil and keyboard and has achieved an accuracy of 92.02% on the ILSVRC 2012 dataset. Each residual block is 3 layers deep and like the CNN models, ResNet-50 has a fixed input size of 224 × 224 but instead consists of two paths; (1) main branch which are a series of neural network layers and (2) shortcut branch which is a direct path from the input to the output i.e.; identity shortcut.

4.4.7 ResNet-101

The ResNet-101 model is 101-layer deep and has been trained on over a million images of a 1000 different classes of objects originating from the ImageNet database. The model has a fixed input size of 224 x 224 dimension and each residual block is 3 layers deep.

4.4.8 ResNet-152

The main idea that went into building the ResNet-152 model was to stack residual blocks to form a very deep, 152-layer network. The input size for the model is 224 x 224 and like the other ResNet models, each residual block contains 3 convolutional layers. Residual networks have shown to improve the task of image classification over CNN (VGG-based) models.

When comparing CNN with ResNet models' researchers have demonstrated that as the residual network gets deeper the accuracy increases, because of the identity-shortcut connection. The shortcuts can learn residual information without compromising on performance and without exhibiting signs of overfitting.

Given the known advantages of using residual learning models for classification of non-face stimuli, it is an obvious choice to investigate its efficiency for face data, especially ethnicity. Hence all three of the ResNet models have been implemented in our framework for the classification of binary ethnicity from face data including isolated face features. Alternatively, the VGG-Face dataset contains over 2.6 million face images of 2,622 different identities and was used to train the VGG-F model. Since the model has been previously applied on face data, it was a natural choice to select it for the proposed ethnicity classification framework. In contrast, VGG-16 and VGG-19 have been trained on a dataset of 1000 varied classes of objects instead of human faces, so it will be interesting to observe how the models behave when given face data.

4.4.9   Feature Extraction

Feature extraction is the process of parameterizing the input i.e. the face image, with a view to defining the most discriminating features. Since published literature has shown that Deep Learning models are capable of learning generic image features (LeCun et al. 2015), it is possible to use such features directly with a classifier in order to address an image-based

task. A total of three sets of features are retrieved using ResNet-50, RestNet-101, and ResNet-152. For the VGG-based experiments, layer 34 of the VGG-Face and VGG-16 model is used to extract discriminatory features, while layer 41 is used for the VGG-19 model. The initial layers of ConvNet represent local visual features of image and the deeper layers represent more global features. It is proven from recent studies (Mehmood et al. 2019; Ul Haq et al. 2019) that the convolutional layers are more suitable for localization and saliency extraction kinds of tasks and fully connected layers are more suitable for recognition and classification tasks. Therefore, we utilized features of FC7 layer and the FC6 layers is not utilized because it is the first deeper layer and features are not as mature as FC7. The FC8 layer is discarded because the VGG model is trained on ImageNet dataset which has 1000 classes the FC8 layer has 1000 activation which is used for the classification of the ImageNet dataset categories.

Given an input image, the process of feature extraction will entail learning parameters such as image height, image width, colour channels, skin colour image edge information, in addition to face-shape data from localized face-features such as the eyes, nose and mouth. Essentially, each input image is filtered through different layers of each pre-trained model independently. The pre-trained models consist of learnable parameters i.e. layered ConvNet, '**C**' which in turn generate outputs '**X**' for each layer in the network,

$$\mathbf{C} = c^1, \; c^2, \; c^3 \dots c^n \text{ with the outputs such as } \mathbf{X_1, X_2, X_3 \dots X_n} \qquad (4\text{-}1)$$

4.4.10 Ethnicity Classification

A Linear classifier was employed for binary ethnicity classification. Support Vector Machine (SVMs) are supervised Machine Learning models that function to identify a hyperplane, which best classifies data points within a given data space. Published studies have demonstrated that SVMs is a powerful binary classifier and operates by defining an Optimum Separating Hyperplane (OSH) between two classes of data (Chierchia et al.; Nalavade and Meshram). For ethnicity classification, the two classes presented within the model are Pakistani $(+1)$ and Non-Pakistani (-1).

4.4.11 Model Evaluation

A K-fold cross validation technique is used to evaluate the performance of all 6 of the pre-trained models (ResNet-50, ResNet-101, ResNet-152, VGG-F, VGG-16 and VGG-19). The face data 'D' is split into 10 mutually-exclusive parts of the same size, D_1, D_2, D_3..., D_k. The SVM algorithm is trained 10 times and during each rotation, 1 set is selected for testing and the remaining 9 sets are combined to create a training set. A K-fold cross-validation is conducted by 10 similar executions by using each of the ten partitions as a testing set in each iteration. During each iteration, relevant features are extracted from the training set and the classification is applied. The outcome of the classification is used to identify face images of Pakistani ethnicity.

We have used k-fold cross validation where k=10: which means the value for k is fixed to 10. The value is identified through

89

experimentation to generally result in a model skill estimate with low bias a modest variance. Importantly k-fold cross-validation was not manually implemented, instead the scikit-learn library was used, which provides an application that will split a given data sample up. We created an instance that splits the dataset into 3 folds, shuffles prior to the split, and uses a value of 1 for the pseudorandom number generator. This is where the k-fold cross-validation procedure is repeated n times, where importantly, the data sample is shuffled prior to each repetition, which results in a different split of the sample.

4.5 Experimental Results

The previous Section details the framework and methodology applied for image based ethnicity classification of Pakistani face images. In this Section, observations from the multiple experiments performed to evaluate the methodology, is presented. Experiments have been conducted on a dataset of 1,000 (500 Pakistani and 500 Non-Pakistani) face images, 1,000 isolated eyes, nose and mouth crops. By extracting features from images using the pre-trained ResNet-50, ResNet-101, and ResNet-152 models, close to near-perfect results were achieved for the binary classification of a Pakistani face using a dataset of 1,000 frontal images, Table 4.1.

Feature Extraction Model	Classifier
	Linear Support Vector Machine (SVM)
ResNet-50	98.8%
ResNet-101	**99.2%**
ResNet-152	99.0%

Table 4:1: The performance accuracy (%) of Residual Learning models (ResNet-50/101/152) for the binary classification of the Pakistani Face, using a Linear Support Vector Machine Algorithm (SVM).

Since the reported results for each of the 3 ResNet-models are very similar with little margin between each of the three models. For ease of data visualization, Figure 4.3 displays the results in a graph format
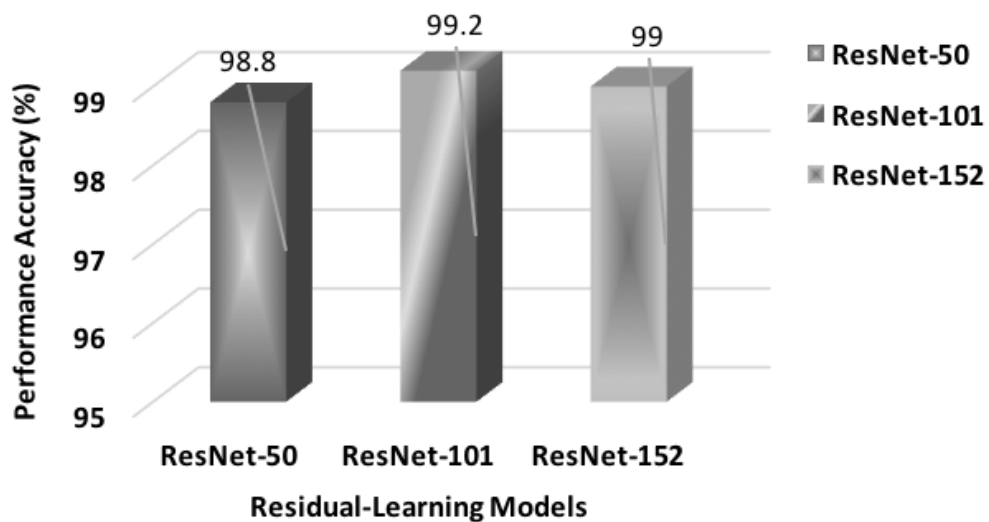
Figure 4:3: Efficiency of Residual Learning Models (ResNet-50/101/152) for the binary classification of the Pakistani full face, using a dataset of 1,000 images.

It is clear from visually examining the data that firstly, Deep Residual Networks are efficient in extracting feature detail from the face to ascertain

ethnicity, especially for the Pakistani population. The strength of residual learning is the ability to skip layers within deeper network structures and yield higher precision for image-based classification tasks, namely ethnicity in this context. Thus, the results shown in Table 4.1 and Figure 4.3 suggest that ResNet-50, ResNet-101 and ResNet-152 are efficient in extracting high-level face information to address the ethnicity classification problem. In contrast, the performance accuracy for the VGG-based pre-trained models (VGG-F, VGG-16 and VGG-19) is shown in Table 4.2 and as part of Figure 4.4.

Feature Extraction Model	Classifier
	Linear Support Vector Machine (SVM)
VGG-Face	87.1%
VGG-16	**93.2%**
VGG-19	87.6%

Table 4:2: The performance accuracy (%) of Convolutional Neural Network models (VGG-F/VGG-16/VGG-19) for the binary classification of the Pakistani Face, using a Linear Support Vector Machine Algorithm (SVM).
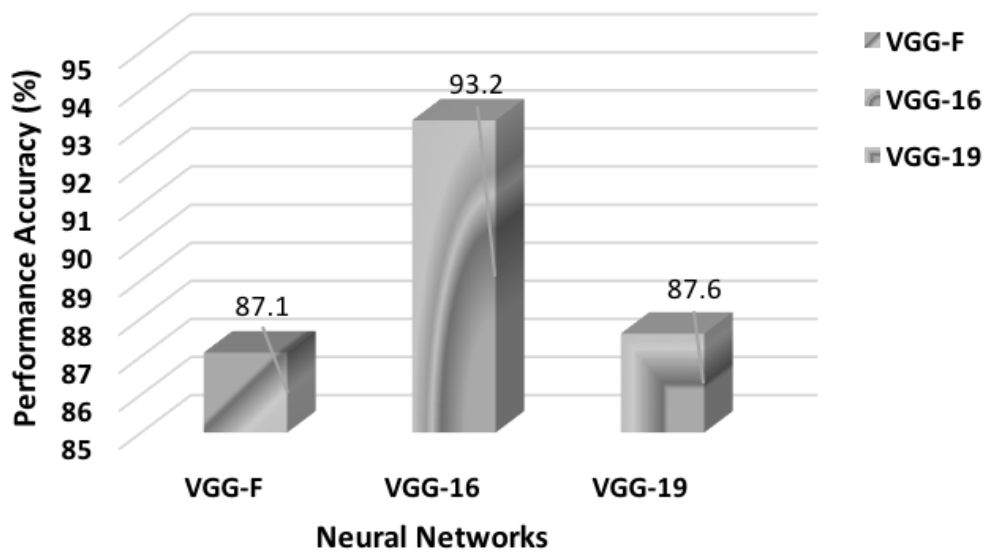
Figure 4:4: Efficiency of Convolutional Neural Network Models (VGG-F/VGG-16/VGG-19) for the binary classification of the Pakistani full face, using a dataset of 1,000 images.

It is evident from looking at Figure 4.4 that while the VGG-based models perform competitively, the VGG-16 model outperforms both VGG-F and VGG-19, with an accuracy of 93.2%. On the contrary, VGG-F achieved the lowest classification rate, albeit above 80% with a hit rate of 87.1%. This is comparable to the accuracy performance achieved by VGG-19, with a hit rate of 87.6%. While the results differ for each of the 6 evaluated Deep Learning models, it is important to acknowledge that the dataset for each of the models was the same. Thus, the results report the adequacy and robustness of each of the models without any external factors, influencing the results. It is fair to draw the conclusion that Deep Learning models are far more accurate and competent in extracting high-level data from face images, to classify the Pakistani ethnicity. Further, the results demonstrate

that all 6 of the Deep Learning models performed above chance (> 50%) suggesting the repeatability of the results and validity of the applied methodology. Having obtained the classification performance of the 6 pre-trained models on a dataset of 1,000 face images, experiments were conducted to ascertain whether isolated face components i.e. eyes, nose and mouth are also as informative and discriminatory for binary ethnicity classification. Essentially, it is deemed important to see whether the features are enriched when tested alone, or whether they all have equal importance when presented collectively. Additional experiments were carried out using the same 6 pre-trained models (ResNet-50/101/152, VGG-F/VGG-16/VGG-19) on datasets of 1,000 eyes, nose and mouth (Table 4.3).

Feature Extraction Model	Linear Support Vector Machine (SVM)	Face Feature
ResNet-50	85.9%	Eyes
ResNet-101	85.6%	
ResNet-152	**87.4%**	
ResNet-50	91.5%	Nose
ResNet-101	**91.8%**	
ResNet-152	91.7%	
ResNet-50	94.8%	Mouth
ResNet-101	94.3%	
ResNet-152	**95.7%**	

Table 4:3: Performance accuracy (%) for the binary ethnicity classification of isolated face-components: (1) Eyes, (2) Nose and (3) Mouth using ResNet-50, ResNet-101 and ResNet-152, and a Linear Support Vector Machine (SVM) Classifier.

The results achieved from the isolated face-components demonstrate that the eyes, nose and mouth are reliable features, which can be used in isolation of one another to separate between the Pakistani and Non-Pakistani class. Firstly, by visually assessing the Table of results, the mouth is a feature of the lower face, which in isolation reports the highest classification, followed by the nose and then lastly, the eyes. ResNet-152 looks to outperform both ResNet-50 and ResNet-101 on mouth data, achieving 95.7%. The performance accuracy between ResNet-101 and ResNet-152 is very close at 91.8% and 91.7% respectively on the dataset

of 1,000 nose images. Concluding that ResNet-101 generates the highest accuracy for the nose. The eyes are the only face-feature which reports a performance accuracy of below 90% with 85.9%, 85.6% and 87.4% for ResNet-50, ResNet-101 and ResNet-152, respectively. Experiments implementing the VGG-based models were also conducted to the dataset of isolated face-components for comparative analyses, Table 4.4. Importantly, the experimental set-up and testing dataset remained the same.

Feature Extraction Model	Linear Support Vector Machine (SVM)	Face Feature
VGG-Face	71.2%	Eyes
VGG-16	**78.0%**	
VGG-19	74.2 %	
VGG-Face	74.3%	Nose
VGG-16	81.5%	
VGG-19	**82.9%**	
VGG-Face	76.9%	Mouth
VGG-16	**86.7%**	
VGG-19	81.0%	

Table 4:4: Performance accuracy for binary ethnicity classification of isolated face-components: (1) Eyes, (2) Nose and (3) Mouth using VGG-F, VGG-16 and VGG-19, and a Linear Support Vector Machine (SVM) Classifier.

Similarly, the results reported by the VGG-based models confirm that the tested face-components i.e. eyes, nose and mouth, are enriched and can be applied to the task of ethnicity classification. Further demonstrating that there is no requirement to train Machine Learning models with full-face images. While VGG-F, VGG-16 and VGG-19 perform competitively, VGG-16 outclassed VGG-19 and VGG-F on the eye and mouth dataset, achieving 78% and 86.7% respectively. In contrast, VGG-19 marginally outclassed VGG-16 on the nose dataset by a margin of 1.4%. When comparing the results on the dataset of the 1000 eyes, nose and mouth images, between the 6 investigated models, the Residual-Learning models (ResNet-50, 101 and 152) exceed the performance accuracy achieved by VGGF, VGG-16 and VGG-19, see Figure 4.6.
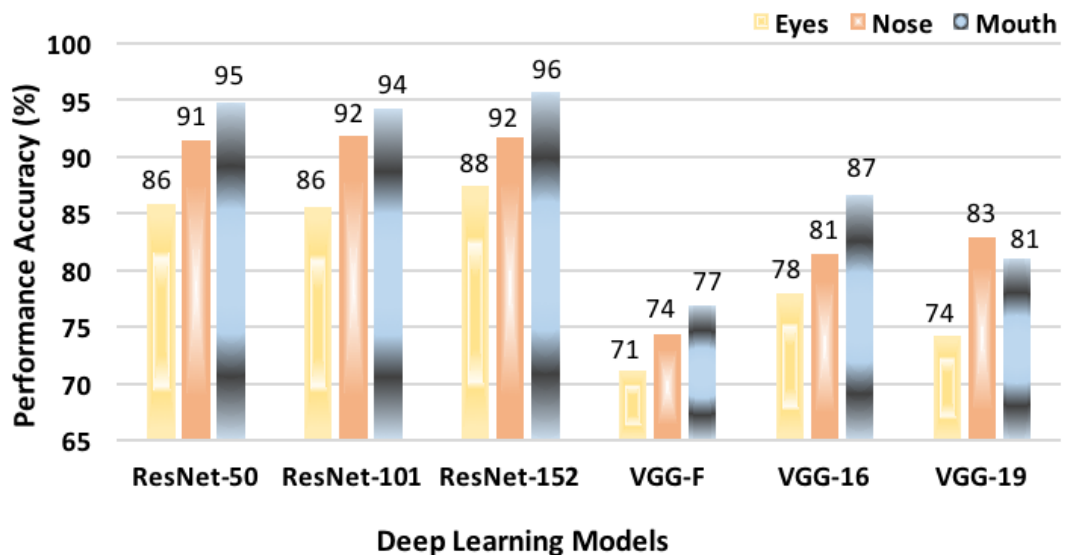
Figure 4:5: A combined column graph to show the performance accuracy (%) between different Deep Learning models (ResNet-50/101/152, VGG-F, VGG-16 and VGG-19) for the classification of isolated Pakistani face-components; Eyes, Nose and Mouth, using a Linear Support Vector Machine (SVM) classifier.

When analyzing Figure 4.5, it is apparent that the features with the greatest potential are those which constitute the lower portion of the face, i.e. the nose and the mouth. The graph clearly illustrates the robustness of individual face-components when using Deep Learning models for image based ethnicity classification. By visual observation it is apparent that the nose and the mouth in isolation yield a high classification rate and are strong ethnic markers for the Pakistani face. To determine the effectiveness of the results achieved, performance metrics such as Sensitivity and Specificity are calculated,

$$\text{Specificity} = \frac{\textbf{TN}}{\textbf{TN+ FP}} \tag{4-2}$$

$$\text{Sensitivity} = \frac{\textbf{TP}}{\textbf{TP+ FN}} \tag{4-3}$$

Sensitivity is the value of the positive class that has been correctly identified (True Positive Ratio), while Specificity is the value of the negative class which has been correctly identified (True Negative Ratio) (Fawcett 2006; Powers 2011). The true positive ratio (TPR) and the false positive ratio (FPR) are defined as,

$$\textbf{TPR} = \frac{\textbf{True Positive}}{\textbf{True Positive+False Positive}} \tag{4-4}$$

$$\textbf{\textit{FPR}} = \frac{\textbf{True Negative}}{\textbf{True Negative+False Negative}} \tag{4-5}$$

"True Positive" is the total number of instances the Machine Learning model makes a correct prediction; "False Positive" is the total number of cases for which the model yields an incorrect prediction; "True Negative" is the total of cases for where the model has correctly predicted a non-match; "False Negative" is the total number of cases where the model has incorrectly predicted a non-match (Fawcett 2006). By combining Sensitivity and Specificity the overall performance accuracy of the classification algorithm can be shown in form of a confusion matrix.

A confusion matrix highlights that for any given experiment there are 1 of 4 outcomes (Figure 4.6). Given a set of data points ($\boldsymbol{x}^{(i)}$, …, $\boldsymbol{x}^{(n)}$), where each $\boldsymbol{x}^{(i)}$ has n features associated to a set of outputs ($\boldsymbol{y}^{(i)}$, …, $\boldsymbol{y}^{(n)}$). The confusion matrix represents how a given classifier learns to predict $\boldsymbol{y}$ from $\boldsymbol{x}$.

TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
TRUE NEGATIVE (TN)	FALSE POSTIVE (FP)
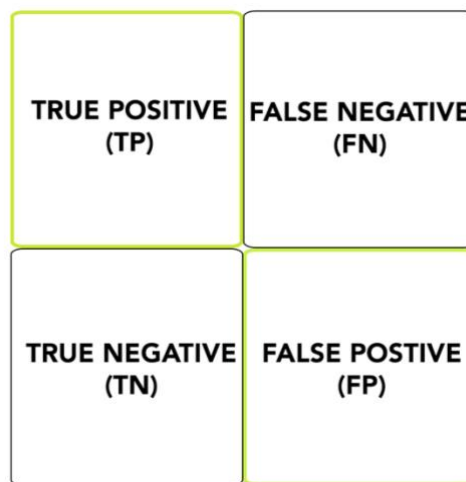
Figure 4:6: A schematic of a Confusion Matrix which is a layout that allows visualization of the performance of an algorithm.

A visual representation of the values attained from the confusion matrix is shown in form of a Receiver Operating Characteristics (ROC) Graph. Figure 4.6 represents the generated ROC for the full-face experiments conducted using ResNet-features, while Figure 4.7 shows the ROC for the VGG-features.
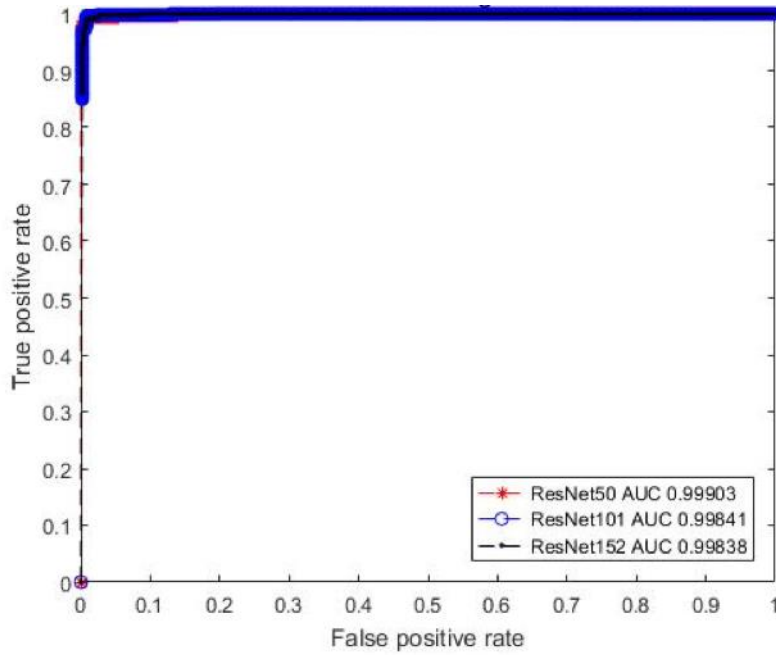
Figure 4:6: Receiver Operating Characteristics (ROC) Curve using ResNet features for the binary classification of full-faces on a dataset of 1,000 images.

Figure 4.6 shows the ROC for the full-face experiment using three Deep Learning algorithms; ResNet-50, ResNet-101 and ResNet-152. All three models performed marginally close-to-perfect, which is depicted by the position of the performance lines, i.e. positioned to the left of the axis. Despite having seemingly lower accuracy (98.8%), RestNet50, having an area under the curve (AUC) of 0.9990 demonstrates competitive performance. This proves that in the presence of sufficient face detail, the

50-layered architecture can capture strong ethnicity traits.
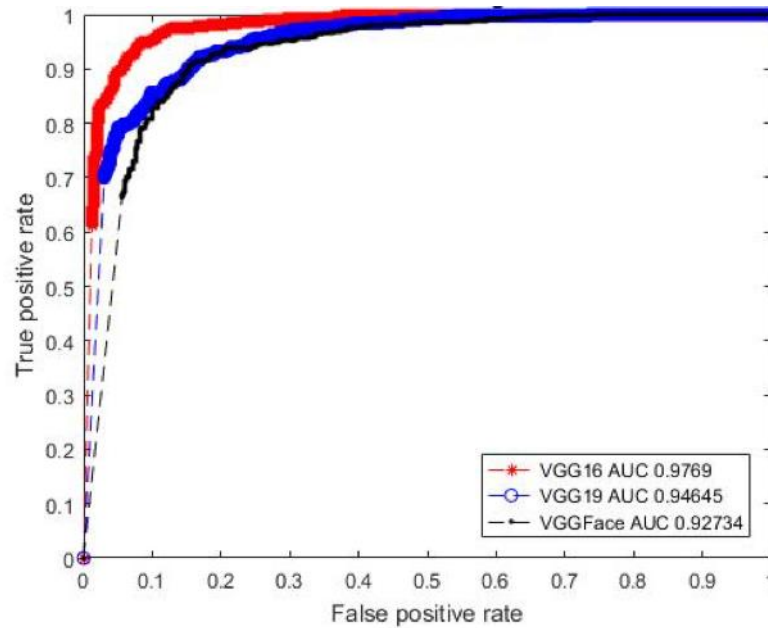
Figure 4:7: Receiver Operating Characteristics (ROC) Curve using VGG features for the binary classification of full-faces on a dataset of 1,000 images.

In comparison, Figure 4.7 shows the ROC for the full-face experiment using the three VGG-based Deep Learning algorithms; VGG-F, VGG-16 and VGG-19. By visually analyzing the placement of the performance lines on the ROC, it appears that the VGG-based experiments do not perform equally and with reduced performance, to the Residual learning models. This is depicted by the position of the performance lines on the graph since the lines are closer to the center. Further, a low accuracy is reported in the values for the area under the curve (AUC). AUC values which fall between .90 – 1 are known to be excellent indicators of accuracy. When comparing the strength of performance, it is evident that all 6 of the

Deep Learning models report excellent classification, and the results are a clear indication of the discriminatory abilities of Deep Learning models for binary ethnicity classification from full-face and face component data.

4.6 Discussion

The classification of full faces and its internal-feature components i.e. eyes, nose and mouth, is an interesting challenge within the field of Machine Learning. The results presented in this Chapter are novel because the application of Deep Learning algorithms for the ethnic classification of Pakistani face and face-features particularly, remains unchallenged. The results achieved for the full-face ethnicity experiment using Residual Networks (ResNet), outperforms Ou et. al, study on the binary (Asian Vs. Non-Asian) classification and Hosoi et. al, study of a three class (Asian Vs. African Vs. European) ethnicity task. However, the results are also in line with those reported by Muhammad et. al. (2012) who achieved 98.42%, 95.56%, 93.65%, 100% and 98.18% for the Asian, Black, Hispanic, Middle-Eastern and the White racial class respectively, albeit implementing a different methodology with Local Binary Patterns (LBP).

Islam (2017) implemented VGG-F for feature extraction using 68 points from the face to investigate ethnicity classification of the White, Black and the Asian class. A total of 10 varied face image databases including the Chicago Face Database, Facial Recognition Technology Database (FERET), Japanese Female Face Expression (JAFFE) were used for

training and testing. By combining a range of the databases, Islam and colleagues created a pool of 8291 Asian face images, 1035 Black face images and 4068 White face images. K-fold cross validation was conducted and a linear classifier was applied. The reported results were above 90% even with the inter-database differences such as lighting, pose and emotion.

However, while the results are higher in comparison to those achieved with the full-face experiment presented within this Chapter, a key disadvantage is the lack of sufficient high quality face data. The researchers do not provide a definition of what ethnic group constitutes 'Asian' or 'Black', what is known, is that the classification does not consider the South Asian race, nor the Pakistani ethnic group. A potential reason for this being the lack of face data available for the group. Thus, while the reported results from the Islam 2017 study are higher compared to those presented in this Chapter, the novelty of the research remains. A key challenge which is of great importance is computational time, as it is important to ensure a framework is computationally efficient. From the moment of data input to the point where the images have been classified accurately, required on average 20 minutes. While, a comparison between computational time from previously published literature is not feasible (due to the unavailability of such knowledge), it is expected that other methods of ethnicity classification may also be time consuming. Especially when considering the human perspective for the classification of ethnicity based on face images (discussed in the upcoming Chapter 6). Ultimately, by working with Deep

Learning the efficiency of the work is not compromised when considering the high-performance accuracies that are achieved.

4.7   Conclusion

Multiple, binary ethnicity classification experiments have been conducted using multi-ethnic frontal face images, as well as face components. Deep Learning models namely ResNet-50, ResNet-101, ResNet-152, VGG-F, VGG-16 and VGG-19 were used to extract face information, which subsequently was forwarded to a Linear classifier, for a two-class (Pakistani or Non-Pakistani) grouping.

The results achieved show the strength of performance of Deep Learning algorithms since the implemented framework concludes 99.2% accuracy on a dataset of 1,000 full face images for the ethnic classification of the Pakistani face using the model ResNet-101. In contrast, a performance accuracy of 93.2% was achieved on a dataset of 1,000 full face images using the VGG-16 model. Since such promising results were achieved using full face information, it was important to ascertain whether the Machine Learning models would perform comparably, with augmented face-components, i.e. eyes, nose and the mouth. Further experiments were carried out using isolated face-features i.e. the eyes, nose and mouth. Again, high performance accuracies are reported for the 3 different datasets, which concluded 91.8% on nose images using ResNet-101 and 95.7% on mouth images, using ResNet-152. The reported results highlight

the vigor of Deep Learning models in addition to the descriptive nature of the nose and mouth. The lowest performance accuracy reported by all 6 of the pre-trained models was on the data of 1,000 eye images, suggesting that it is the least informative feature when attempting to classify ethnicity. Although the classification rates are above chance (50%), since ResNet-152 attained 87.4% and VGG-16 concluded 78% accuracy. The margin of error is far greater, when comparing to the other features.

To further study the discriminatory power of Deep Learning models, and the enriched nature of the nose and mouth, further experiments are required. Moreover, in a test to investigate whether discrimination ability can be increased, face features are combined for the upcoming experimental Chapter. It is hypothesized that performance accuracy will increase as the data sample is further augmented, although it will not be surprising if there isn't a significant increase, since a challenging dataset will be probed.

CHAPTER 5

PARTIAL FACE DATA FOR ETHNICITY VERIFICATION

5.1 Introduction

To further explore the role of face-components this Chapter presents a series of experiments on a set of challenging image datasets. The expressive nature of the nose and mouth features are further tried, using a novel catalogue of 2,200 images, which contains partial data, potentially making the task of ethnicity classification increasingly demanding. The motivation behind investigating the nose and mouth features is driven by the high-performance accuracy reported in the previous experiments, thus it is natural to devise supplementary experiments, to determine their true robustness.

In line with this, additional experiments are also proposed based on the concept of feature combinations. The feature combination datasets consist of two features such as: (i) eyes and nose, (ii) nose and mouth and (iii) mouth and eyes. The theory behind combining face features is to explore whether by combining the internal components of the face, a higher ethnic recognition rate is achieved. This notion doesn't appear to have been previously attempted within the discipline of Machine Learning, nor for ethnicity classification.

The human nose is the most centralized and protrusive feature on the face (Harkema et al. 2006). By using images of the nose as a primary feature for image-based classification, there are several advantages. Firstly, the shape of the nose is robust to changes caused by variations in facial expression, and is unaffected by facial hair. Furthermore, anthropometric studies report both gender and ethnicity differences in the shape and size of the nose (Ohki et al. 1991; Ozdemir and Uzun 2015; Elsamny et al. 2018). The stability of the nose appearance within an individual, and the invariance of nasal shape over time, highlights the significance of this surface feature for image-based ethnicity classification.

The mouth is positioned directly beneath the nose and is the oral cavity through which food and the air enters the body. The impact of ethnicity on the mouth region has been reported by researchers who have demonstrated that the mouth width (measured chelion to chelion) is identical between males and females from the North American White population, Egyptian, Iranian and Turkish. However, significantly smaller for those of Vietnamese heritage (Farkas et al. 2005). While the Farkas study reports on variances of the face, which includes the mouth area, there is no published literature with a Machine Learning framework, which investigates the role of mouth features for ethnicity discrimination. Therefore, it is envisaged that the upcoming studies will be a first, and serve as a benchmark for future studies.

The experiments discussed within this Chapter, implement pre-trained Machine Learning algorithms for nose and mouth feature based ethnicity classification. A sub-set of these primary experiments, is the feature combination experiments. The proposed method for ethnicity classification consists of 2 components; (i) Feature extraction using weights of VGG-Face, VGG-16, VGG-19, ResNet-50, ResNet-101 and ResNet-152 and (ii) ethnicity classification using a Linear Support Vector Machine (SVM) algorithm as a binary procedure.

The Chapter is organised as follows; a literature review on ethnicity classification, related to the role of face features specifically the nose and mouth, is presented as part of Section 5.2. In Section 5.3, the experimental stimuli are discussed, while Section 5.4 reports the methodology. In Section 5.5 the experimental results are presented then the discussion Section is reported in Section 5.6. To conclude the Chapter, Section 5.7 presents a synopsis of the findings.

5.2   Literature Review

Ethnic variations in facial appearance are significant and the physiological makeup is dependent upon a person's ethnicity (Greenwell et al.; Islam et al.). It has been reported that asymmetric dimensions in anthropometry can be an indicator of gender and ethnicity. Images taken from the MORPH and FERET database were used to demonstrate male faces to be more asymmetric compared to females. African participants were reported to have asymmetric face dimensions compared to Europeans, and

comparable face measurements were noted amongst Hispanic and European participants (Sajid et al. 2018).

While the mouth was not used as a stand-alone feature, in the Srinivas et. al, (2017) study. A CNN-based framework was used for predicting fine-grain ethnicity (defined by the authors as a refined category of a human group) for 5 classes: Chinese, Japanese, Korean, Vietnamese and Filipino (Srinivas et al. 2017). Face and face regions from the Wild East Asian Face Dataset (WEAFD), were located and annotated using Dlib. Regions included the eyes, mouth, lower face (inclusive of all three face features as well as the chin), center of the face (nose only) and the left and right side of the face.

The methodology employed by the Srinivas et. al, (2017), made use of two different CNN models, one for the whole face condition (Network A), and another for the sub-region condition (Network B). A key difference between the two models were the feature parameters; Network A had 3,573,028, while Network B had 1,688,550. While the results for fine grain ethnicity were low, 24.0% for experiments 1 and 33.33% for experiment 2. The authors acknowledged the limitations of their training and testing data, was the predominant reason for the low accuracy. Although, the results did highlight that the features learned in the first convolutional layers and ultimately used to discern ethnicity include the shape of the nose, eyebrows and the lips. This provides a rational foundation to suggest that the region of the mouth can be an enriched feature for demographic classification.

The classification of ethnicity from distances between the internal components of the face (i.e. eyes, nose and mouth) was carried out by Masood et. al, (2018). Using a total of 357 images for training and 90 for testing, three ethnic groups were investigated; Mongolian, Caucasian and Negro (described as per the researcher), and a comparative analysis was conducted between 2 methods of analysis. Artificial Neural Network (ANN) and Convolutional Neural Network (CNN). It was concluded that CNN reported higher results with 98.6% accuracy compared to 82.4% for ANN (Masood et al. 2018).

The nose is an useful biometric (Chang et al. 2006), although the availability of nose-related literature is limited. Lv et. al (2018) introduced a novel 3D Nose Shape Net classifier for gender and ethnicity classification. To construct a 3D Nose Shape Net, a nose measurement method is required to ascertain the distances between the different noses within a given dataset and then the images are clustered into groups. By using the nose clustering results, the 3D nose is constructed. Using the Bosphorus3D face image database and the FRGC2.0, the authors reported an ethnicity classification of 89.2% for the Asian ethnicity and 87.4% for the White. A similar study on the use of 3D nose shape information has also been reported by Drira et. al (2019).

Such literature demonstrates the strength of the internal components of the face, inclusive of the nose and mouth, as information-rich features. The literature review demonstrates that the use of nose and

mouth images for demographic (ethnicity) identification has proven successful. The nose and mouth is a robust are discriminative features and the literature review proves their reliability.

5.3   Dataset

A total of 100 Pakistani and 100 Non-Pakistani images were selected from the Pakistani Face Database. Images which constitute the Non-Pakistani class, are varied and include Black, Caucasian, East Central Asian, Kurds etc. as well as the South Asian, Gujarati (Indian) ethnicity. The assigned ethnicity of each image was determined by asking the participants to self-identify their ethnicity. For participants of Pakistani origin, eligibility was dependant on whether both the maternal and paternal parents were of Pakistani ethnicity. All the facial images were cropped around the region of the nose using Photoshop CS6 to generate a sub-dataset of the varied nose regions, see Figure 5.1. To keep all the information, it was ensured that the entire nose was cropped, showing clearly the nasal bridge, the pronasale (nasal tip) as well as the nasal alars (nostrils).

Figure 5:1: An example of feature-component (nose) stimuli that have been processed using Adobe Photoshop CS6. The nose images are cropped from full face raw images.

A total of 11 images have been created (from the baseline 200 participants) per subject by the means of augmentation, consequently building a database of 2,200 nose images, see Figure 5.2. The dataset is equally split between the Pakistani and Non-Pakistani class (1,100 images per class). A further 4 datasets have been generated for ethnicity classification; (1) database of 2,200 mouth images (Figure 5.3), (2a) feature combination consisting of eyes and nose, (b) nose and mouth and (c) mouth and eyes. The feature-combination datasets consist of 1,200 images in total (600 Pakistani images and 600 Non-Pakistani), there are 6 images per participant; 3 images per feature at rotations of 90° clockwise, 90° counter-clockwise and 180° rotation (see Figure 5.4).
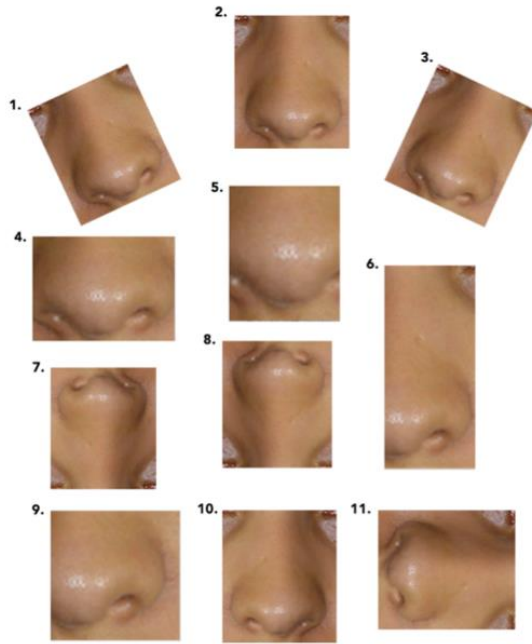
Figure 5:2: An example of the augmented Nose dataset consisting of 11 images per participant. The image shown is of a South Asian, Pakistani participant. (Augmented positions: 1. 25° counter-clockwise, 2. Full nose, 3. 25° clockwise, 4. Pronasale and alars, 5. Pronasale, 6. Partial nose, 7. 180° rotation, 8. Vertical flip, 9. Single alar, 10. Horizontal flip and 11. 90° rotation.)
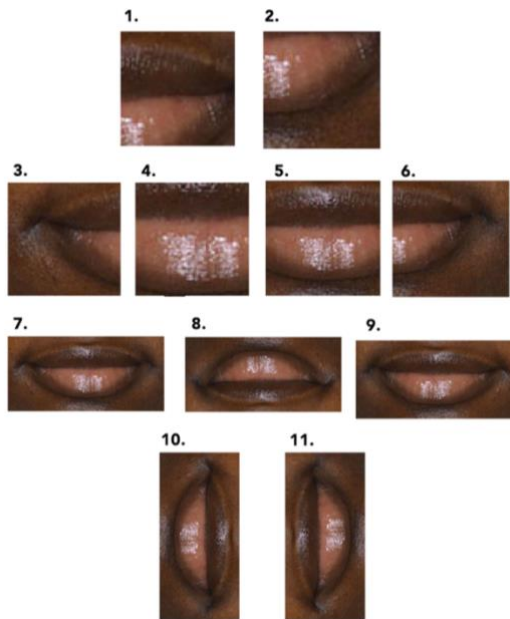
Figure 5:3: An example of the augmented mouth image dataset, showing 11 images for a single participant. (Augmented positions: 1. Upper vermillion, 2. Lower vermillion, 3. Left chelion, 4. Lips zoom, 5. Stomion, 6. Right chelion, 7. Normal mouth, 8. 180° rotation, 9. Horizontal flip, 10. 90° clockwise 11. 90° counter-clockwise.
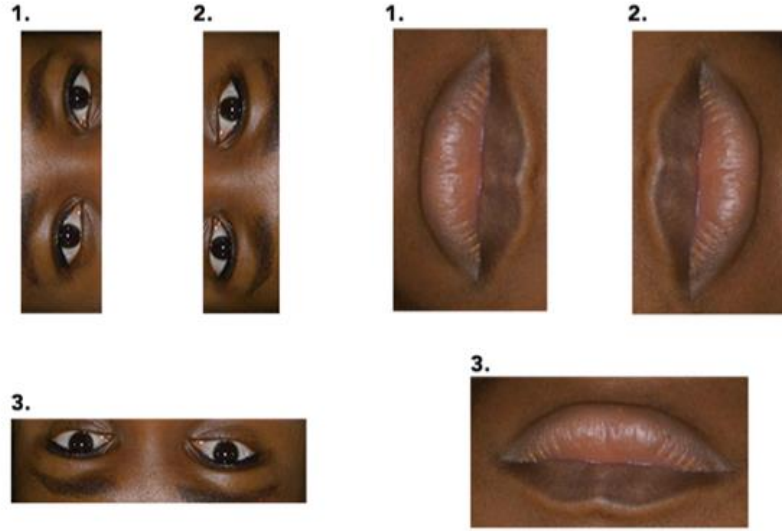
Figure 5:4: Face-feature combination dataset which consists of 6 images per participant. (**above:** eyes and mouth combination). The same rotations are used for the nose and mouth combination and the mouth and eyes feature combination.

5.4 Methodology

The approach to feature extraction and classification is the same as previously described in Chapter 4. Deep features are extracted using the 6 pre-trained Machine Learning models; ResNet-50, ResNet-101, ResNet-152, VGG-Face, VGG-16 and VGG-19. All images used as part of the experiments were resized to a dimension of 224 × 224 to ensure they conformed to the input criteria of the pre-trained models.

5.4.1 Feature Extraction

The activation of the last pooling layer (2048 dimension) of each of the ResNet models was used, for feature representation. Whereas, for VGG-feature representation the Fully-Connected layer 7 (FC-7) with 4096

dimension was used, discounting FC6 and FC8. In total, three sets of features were retrieved using ResNet-50, RestNet-101, and ResNet-152 Neural Networks. Similarly, 3 sets of features were extracted using each of the three VGG-based models.

5.4.2 Classification

Having used Machine Learning-based feature extraction with ResNet-50, ResNet-101, ResNet-152, VGG-Face, VGG-16 and VGG-19, models independently. A Linear classifier was employed for binary classification (Pakistani vs Non-Pakistani ethnicity). For ethnicity classification for the different datasets, the two categories are Pakistani $(+1)$ and Non-Pakistani (-1).

5.4.3 Model Evaluation

A k-fold cross validation technique was employed to evaluate the performance of the pre-trained models. The data '**D**' is divided into 10 equal sections and the SVM algorithm is trained 10 times, whereby during each cycle, 9 sets of data are used for training and 1 is used for testing. By repeating the process 10 times it is ensured that each mutually exclusive dataset has been used for both training and testing. This is important or else the model could fail to recognise trends in data. We have used k-fold cross validation where k=10: which means the value for k is fixed to 10. The value is identified through experimentation to generally result in a model skill estimate with low bias a modest variance. Importantly k-fold cross-

validation was not manually implemented, instead the scikit-learn library was used, which provides an application that will split a given data sample up. We created an instance that splits the dataset into 3 folds, shuffles prior to the split, and uses a value of 1 for the pseudorandom number generator. This is where the k-fold cross-validation procedure is repeated n times, where importantly, the data sample is shuffled prior to each repetition, which results in a different split of the sample.

5.5 Experimental Results

A total of 5 experimental conditions (i.e. nose image only, mouth images only, eyes and nose feature-set, nose and mouth feature-set and mouth and eyes feature-set) have been investigated. Using 6 pre-trained Machine Learning models, experiments have first been conducted on a dataset of 2,200 nose and 2,200 mouth images (1,100 Pakistani and 1,100 Non-Pakistani). After which, experiments were conducted using face-feature combinations. By extracting features from images using three pre-trained Deep Learning models; ResNet-50, ResNet-101 and ResNet-152, a classification result of above 90% is reported, for the binary classification of a Pakistani nose on a dataset of 2,200 nose images, see Table 5.1.

Feature Extraction Model	Classifier
	Linear Support Vector Machine (SVM)
ResNet-50	94.1%
ResNet-101	93.4%
ResNet-152	93.6%

Table 5:1: Performance accuracy (%) for the binary ethnicity classification of 2,200 isolated face-feature (nose) images, using Residual-Networks, for the Pakistani ethnicity.

The ResNet-50 model outclassed ResNet-101 and ResNet-152, achieving 94.1% whereas ResNet-152 marginally outperformed ResNet-101 attaining 93.6% and 93.4% respectively. However, when assessing the results visually, the Residual Learning models have performed comparably, and this is reflected in the closeness in their reported accuracy, demonstrating competitiveness. In contrast, the experimental results for the three VGG-based models (VGG-Face/VGG-16/VGG-19) are varied and lower compared to the ResNet models. The performance accuracies are not comparable to the ResNet models, and may be attributable to the finding that the ResNet models are classed as deeper models, which are architecturally different and report higher accuracies. However, VGG-16 does outperform VGG-F and VGG-19 in the binary classification task, see Table 5.2.

Feature Extraction Model	Classifier
	Linear Support Vector Machine (SVM)
VGG-16	90.8%
VGG-19	87.5%
VGG-Face	79.9%

Table 5:2: Performance accuracy (%) for the binary ethnicity classification of 2,200 isolated face-feature (nose) images, using Convolutional Neural Networks, for the Pakistani ethnicity.

In addition to the nose being tested as a feature of ethnic relevance for the Pakistani face, a series of experiments were conducted using a challenging dataset of 2,200 mouth images. The experimental results in Chapter 4 has demonstrated the robustness of the lower-face feature, as an ethnic marker, hence it has been further interrogated. All 6 Machine Learning models were used to discern whether the mouth is an informative feature and Table 5.3 shows the ethnicity classification of the region. It is apparent that while the Deep Learning models are effective at attaining information sufficient to categorise ethnicity from the mouth, the classification of the ethnicity from the nose outperforms the region of the mouth. A possible explanation as to why, may be the challenging nature of the mouth data which consisted part mouth images, thus the restrictive nature may have impacted performance accuracy.

Feature Extraction Model	Linear Support Vector Machine (SVM)	Face Feature
ResNet-50	94.1%	Nose
ResNet-101	93.4%	
ResNet-152	93.6%	
VGG-16	90.8%	
VGG-19	87.5%	
VGG-Face	79.9%	
ResNet-50	91%	Mouth
ResNet-101	91.5%	
ResNet-152	90.6%	
VGG-16	86.2%	
VGG-19	84.4%	
VGG-Face	79.5%	

Table 5:3: Performance accuracy (%) for the binary ethnicity classification of Pakistani face-feature components i.e. nose and mouth, using 6 pre-trained Machine Learning Algorithms on individual datasets of 2,200 Nose images and 2,200 Mouth images.

Upon visual observation of Table 5.3, ResNet-101 (91.5%) marginally outperforms both ResNet-50 and ResNet-152, when classifying the mouth as a hallmark of ethnicity. Whereas, VGG-16 (86.2%) outclasses VGG-19 and VGG-F when classifying Pakistani ethnicity from mouth images only. Ultimately, the classification of both the lower-face, isolated face regions (mouth and nose) is above 90% when residual learning is applied, in

119

comparison to the VGG-based models, which concludes a lower classification.

Supplementary experiments were also conducted which focused on the use of feature combinations i.e. eyes and nose, nose and mouth and mouth and eyes, which successfully revealed that the features most central to the face are the most discerning for ethnicity, i.e. the nose and mouth combination (Table 5.4). The results are in-line with those reported in Chapter 4 earlier, which highlighted that the features of the lower face are the most enriched.

Feature Extraction Models	Linear Support Vector Machine (SVM)		
	Eyes & Nose	Nose & Mouth	Mouth & Eyes
ResNet-50	85.8%	**91.8%**	87.8%
ResNet-101	85.5%	**90.5%**	88.8%
ResNet-152	84.4%	**90.2%**	84.8%
VGG-16	81.5%	**85.2%**	82.8%
VGG-19	79.5%	**80.3%**	77.6%
VGG-Face	72.6%	**75.8%**	73.6%

Table 5:4: Performance accuracy (%) for the binary ethnicity classification of the Pakistani face using face-feature combinations: (i) eyes and nose, (ii) nose and mouth and, (iii) mouth and eyes.

The results achieved from the isolated feature combinations also demonstrate that, the nose and mouth features yield the highest results consistently. The results are closely followed by the mouth and eyes and the eyes with nose combination. The results show that the internal features

i.e. eyes, nose and mouth do provide reliable information to separate between the Pakistani and Non-Pakistani features. However, the nose and mouth are features which yield the highest accuracy and therefore are the most robust for image-based ethnicity classification (Figure 5.4).
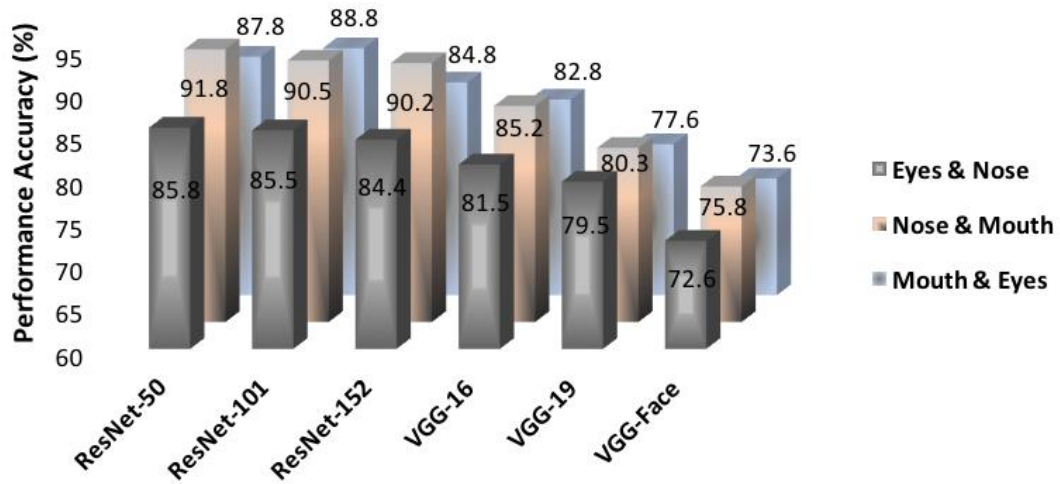
Figure 5:5: Comparative analysis between Residual Learning and Convolutional Neural Network models, for the classification of Pakistani ethnicity using isolated face-feature combinations; (i) eyes and nose only, (ii) nose and mouth and (iii) mouth and eyes.

Importantly, the reported results are in line with literature which proposes that isolated face regions, specifically the nose and mouth are important in determining ethnicity in humans (Hosoi et al.; Azzakhnini et al. 2018). To determine the effectiveness of the results achieved, performance metrics such as sensitivity and specificity were calculated. A visual representation of the values is shown in form of a Receiver Operating Characteristics (ROC) Graph. Figure 5.5 shows the ROC generated for the nose experiment using ResNet-features (experiment 1). While, Figure 5.6 shows

the ROC for VGG-features also using nose data (experiment 2). By looking at Figure 5.5 it is evident that all three ResNet models performed similarly (i.e. within the region of >90%), which is depicted by the position of the performance lines, that are positioned to the left of the axis.
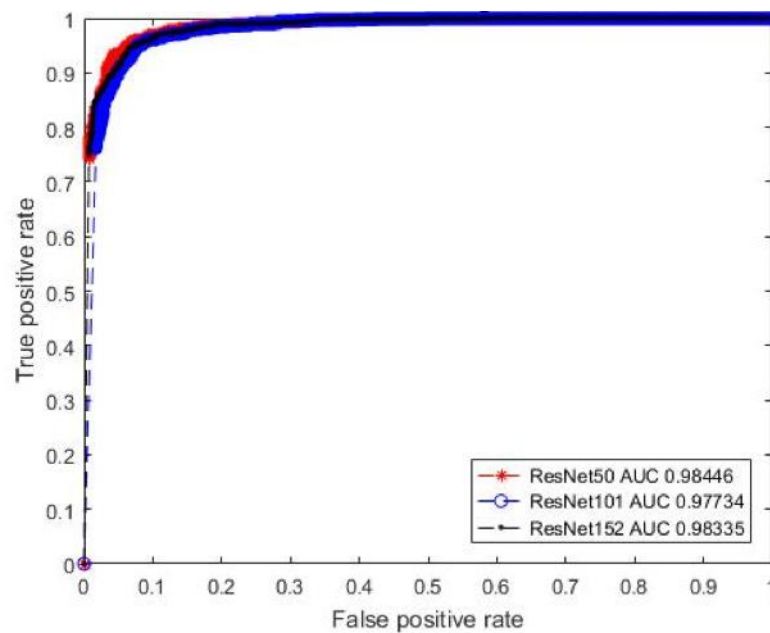
Figure 5:6: Receiver Operating Characteristics (ROC) Curve for the binary classification of the Pakistani nose, using ResNet-Features. The True Positive Rate (TPR) is the value of cases where the data has been correctly identified. False Positive Rate (FPR) is the value of cases where data has been incorrectly identified as a positive.

Despite having seemingly lower accuracy, RestNet-101, having an area under the curve (AUC) of 0.9773 demonstrates competitive performance. This shows that even with limited facial detail i.e. nose images only, the deep-layered architecture is sufficient at capturing ethnicity traits. Figure 5.6 shows the ROC for the isolated nose images using VGG-features. Unlike the ResNet models which are trained on none-face images and performed competitively. The VGG-based models, which are trained specifically on a

dataset of faces, did not perform comparably to the ResNet models. Such results are surprising especially since they are trained on aspects of the face. However, it is apparent from the ROC curve that VGG-16 outperforms the other two models, and VGG-19 outclassed VGG-Face.
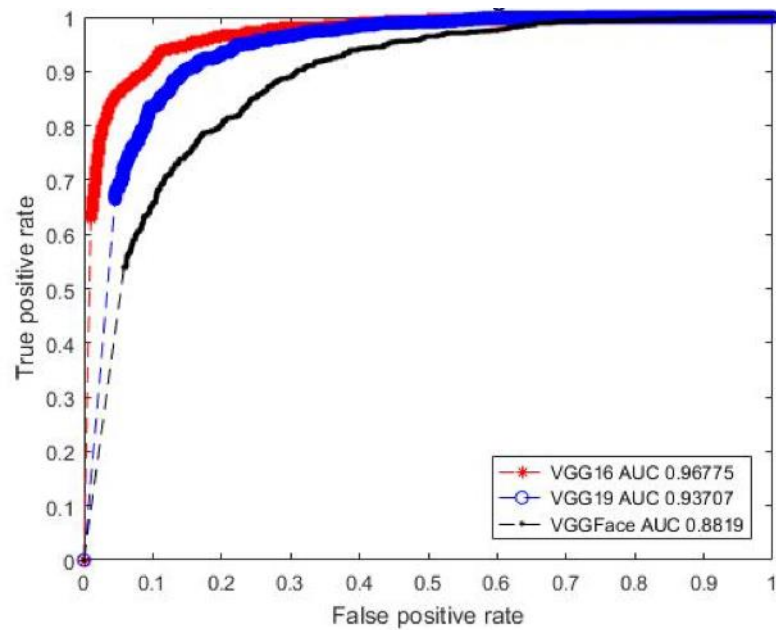
Figure 5:6: Receiver Operating Characteristics (ROC) Curve for the binary classification of the Pakistani nose, using VGG-Features. The True Positive Rate (TPR) is the value of cases where the data has been correctly identified. False Positive Rate (FPR) is the value of cases where data has been incorrectly identified as a positive.

Comparing both the ROC curves, Deep Neural Networks extract relevant information, especially when there is very little facial detail, to yield a high-performance accuracy, in comparison to the VGG-based models.

In contrast, the ROC shown in Figure 5.7 is for the ResNet mouth features, whereas Figure 5.8 shows the ROC for the VGG-features based on the models; VGG-Face, VGG-16 and VGG-19 models.
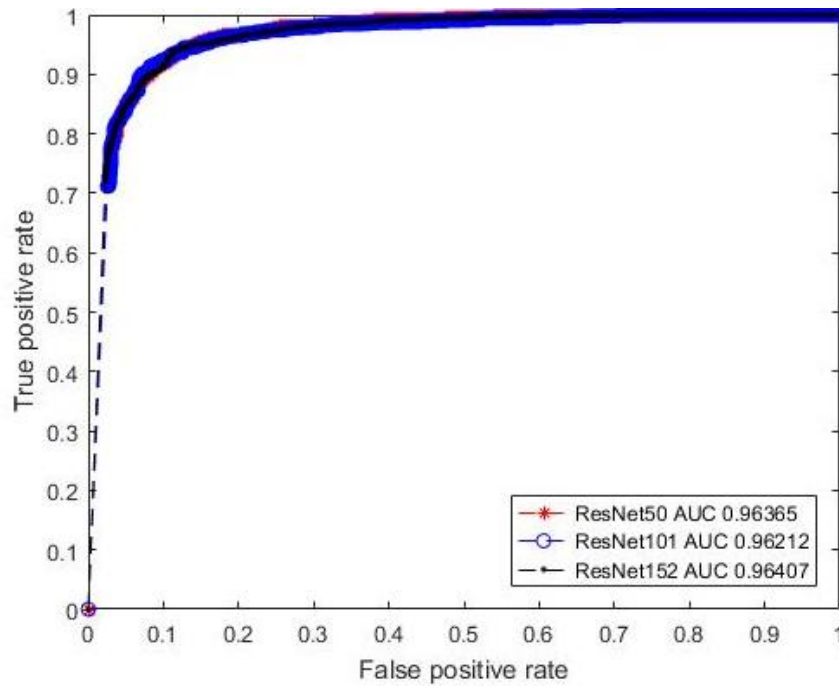
Figure 5:7: Receiver Operating Characteristics (ROC) Curve for the binary classification of the Pakistani mouth, using ResNet-Features. The True Positive Rate (TPR) is the value of cases where the data has been correctly identified. False Positive Rate (FPR) is the value of cases where data has been incorrectly identified as a positive.

By looking at Figure 5.7 it is obvious that all three ResNet-models performed similarly (i.e. within the region of >90%), which is depicted by the position of the performance lines, that are positioned to the left of the axis. ResNet-101 marginally outclassed both ResNet-50 and ResNet-152. The competitiveness of all 3 Residual learning models is reflected in the values for the Area Under the Curve (AUC); 0.963, 0.962 and 0.964 for ResNet-50, 101 and 152 respectively.
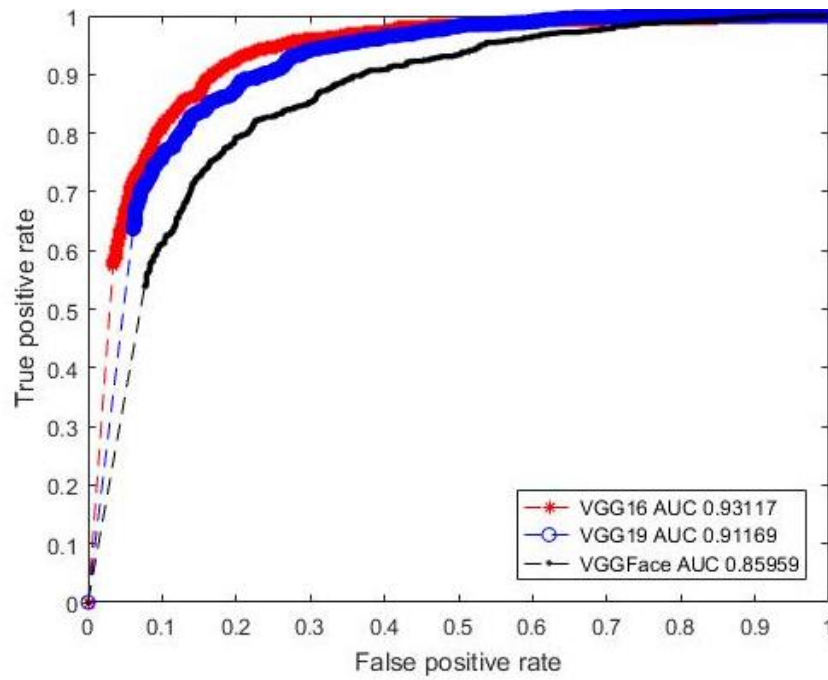
Figure 5:8: Receiver Operating Characteristics (ROC) Curve for the binary classification of the Pakistani mouth, using VGG-Features. The True Positive Rate (TPR) is the value of cases where the data has been correctly identified. False Positive Rate (FPR) is the value of cases where data has been incorrectly identified as a positive.

In comparison, Figure 5.8 shows the ROC for the mouth data images using VGG-features. Unlike the ResNet models which are not trained on face images and performed competitively. The VGG-based models, did not perform comparably to the ResNet models, even though they are trained on a dataset of faces. By comparing both the ROC curves, it is reasonable to suggest that deep neural networks (i.e. residual learning) can extract relevant information, from limited mouth-feature detail, to yield a high-performance accuracy, in comparison to the VGG-based models.

5.6 Discussion

The aim of the presented research was to investigate the distinctiveness of isolated face-components, specifically the nose and the mouth, as physical markers of ethnicity for the Pakistani origin. Supplementary experiments were also conducted to ascertain whether by combining face-features, there is a richness to the regions of the face, for accurate ethnicity classification.

The concluded results show that Deep Learning, ResNet algorithms outperform VGG-based models, on the classification of isolated face-components i.e. eyes, nose and mouth. The results advance current understanding of ethnicity determination because the application of Deep Learning algorithms for the ethnic classification of localised Pakistani features, have not previously been attempted. With regards to the classification of ethnicity from the nose, Azzakhnini et. al (2018) utilized 2 classifiers and reported 80.18% and 91.15% accuracy using Boosting classifier and SVM, respectively. In contrast, the ethnic classification of the nose using the experimental framework discussed as part of this Chapter has achieved state-of-the-art results 94.1%, and outclasses the reported values by Azzakhnini.

ResNet-50 achieved a performance accuracy of 94.1% for the classification of ethnicity from the nose, followed closely by ResNet-152 and ResNet-101 with a classification at 93.6% and 93.4%, respectively. The closeness in performance between the ResNet models is consistent with results presented by He et. al (2016). The 50/101/152-layer ResNet models yield

high and accurate results given the considerably increased depth of the architectures.

Due to the novelty of the experimental stimuli (i.e. isolated nose and mouth data) and the specificity of the approach (the determination of ethnicity based on nose and mouth information only), the opportunity to compare the results with previous literature is limited. However, the presently attained results for the nose have greater accuracy than those presented by Lv et. al (2018) who utilized 3D nose shape classifiers and reported an ethnicity classification of 89.2% for the Asian and 87.4% for the White ethnic group. The results achieved support the suggestion that the nose is a beneficial feature for determining Pakistani ethnicity. Importantly, the results demonstrate that to classify ethnicity, it is feasible to focus on localized face features only, instead of training with images of the full face. This is in agreement with findings by Zehngut et. al (2015) who reported a Machine Learning based experiment on the distinctiveness of the nose as a feature. However, while the authors did not investigate ethnicity classification, the research highlights the value of the nose. It is a viable approach to investigating the relevance of the localized feature in determining the overall classification accuracy. Moreover, the discriminative nature of the mouth as a marker of ethnicity for the Pakistani population, has also been investigated. From the results, it is apparent that Deep Learning (ResNet) algorithms outperform VGG-based models, on the classification of isolated mouth images.

In conclusion, the results reported as part of this Chapter are state-of-the art for the nose and mouth feature condition. Moreover, experiments centering feature combinations, also report on the expressive nature of the nose and mouth as combined indicators of ethnicity. However, when considering the data, an acknowledged limitation is apparent. The Pakistani Face Database is a visual image dataset which is predominantly collected in the city of Bradford, United Kingdom and the Pakistani population it houses may be homogenous in terms of the geographical area of Pakistan from which they originate. This is an inherent limitation and uncontrollable unless a larger dataset is created, constituting different regions of Pakistan.

Secondly, given the diversity associated with the nose and the mouth, it is hypothesized that in situations where humans may struggle to classify the ethnicity of an isolated feature, and discriminate between Pakistani and Non-Pakistani. Deep Learning methods verify competence in providing information, which is not available from the human visual system. Thus, computational methods exploit deep features which reside far deeper within a feature. The results achieved for the mouth condition, is the most exciting and now a proven feature for ethnicity classification, irrespective of the variables which may affect its visibility, e.g.: facial hair and cosmetic makeup. Ultimately, the mouth is a feature which may require fine-tuning, although it is apparent that ethnicity judgements can be made based on the mouth alone.

5.7   Conclusion

In this Chapter, the binary classification of Pakistani face-features i.e.: the nose and mouth have been attempted. In addition, isolated face-feature combinations have also been tested, using a novel, criterion specific dataset of images. ResNet and VGG-features were extracted and fed to a Linear Support Vector Machine (SVMS) classifier. A performance accuracy of 94.1% was concluded for nose-based ethnicity classification using ResNet-50, whereas VGG-16 classified nose images with 90.8% accuracy. The reported results are state-of-the-art and higher than those reported by researchers previously. When considering the discriminatory ability of computer models for mouth features, a performance accuracy of 91.5% was concluded for ResNet-101, while VGG-16 achieved 86.2% accuracy. Ultimately, the experimental conditions i.e. limiting ethnicity classification to nose and mouth features only, are innovative considering the ethnic classification of the Pakistani heritage has not been previously attempted.

Essentially, all the conducted experiments have reinforced a known fact that pre-trained Deep Learning algorithms can extract enriched-features from images of isolated face-components, to determine ethnicity. The nose and mouth show potential as hallmarks of determining ethnicity. Literature relating to the ethnic classification of the nose and the mouth is limited, specifically for 2-Dimensional images. However, there is an acknowledged limitation to the data.

There is a possibility that the data relating to the Pakistani ethnic group, may have their own characteristics manifesting only in the specified geographical region, i.e. the Pakistani participants from the North of the country (Bradford) who may exhibit certain physicality's distinguishable, when compared to Pakistani participants from London or Scotland. The PFDB did not consider geographical aspects and it may be a future study, to compare images of Pakistani participants from different regions of the United Kingdom, to ascertain whether features are consistent or not.

As part of future work, it will be interesting to explore the ethnicity discrimination ability of novice human participants from full faces images, in addition to isolated face-components (i.e. eyes, nose, mouth). To ensure a fairer comparison, the novice tested participants will have no dealings with forensic human face analysis, as it may be that due to work practices, they may show superior ability. By testing human observers, a comparison can be drawn between the performance of both machines and humans, which may provide an insight into which experimental condition yields the best performance accuracy, and within which context. Experiments into the strategies of ethnicity classification from a human perspective are discussed in Chapter 6.

CHAPTER 6

QUANTIFYING THE FACE ETHNICITY DISCRIMINATION ABILITY OF THE HUMAN VISUAL SYSTEM

6.1 Introduction

The face is a complex visual object and a rich source of information. The face enables humans to extract socially-relevant information pertaining to an individual, such as details of their age, gender, identity, emotional state, direction of attention, as well as demographic information relating to ethnicity. The earlier Chapters have reported on the robustness of Machine Learning, particularly the branch of Deep Learning, which reports a natural affinity for image-based ethnicity classification tasks. Face-related analysis is a widely-researched branch within Computer Vision and Machine Learning environments, and the significance is due to its universal application within biometrics and human authentication.

The enriched nature of low-level, hand-crafted features were investigated as part of Chapter 3 whereby it was concluded that a total of 16 hand-crafted face-features are sufficient for front-face ethnicity classification for a Pakistani image, within a Machine Learning framework. The working framework reported a performance accuracy of 65.71%. A significant limitation of this work was this low classification rate, especially considering that other geometric-feature based systems have reported

ethnicity classification at an overall rate of 82. 4% (Masood et al. 2018). As a result, there was a need for further development. This included the use of profile face images, coupled with a powerful statistical technique for the removal of redundant face data. Following this development, 10 enriched, hand-crafted features increased the classification rate to 76.03%. Evidently, low-level face features exemplify the prospect of organizing demographic, ethnic information into binary categories i.e. Pakistani and Non-Pakistani.

Like the Masood et al., 2018 study, which reported an increase to 98.6% when applying Deep Learning. A logical progression to Deep Learning was decided, especially since Deep Learning methods are known to outperform state-of-the-art techniques for a range of Computer Vision tasks, see for a review (Voulodimos et al. 2018). Experimental conditions relating to high-level face data evidences that the Residual-learning model, ResNet-101, achieves the highest performance accuracy of 99.2% for full-face ethnicity classification, from an augmented dataset of 1000 images.

To further test the robustness of high-level features, experiments relating to isolated internal features, especially the nose revealed that ResNet-50 achieves 94.1% accuracy on a dataset of 2,200 isolated nose images. Similarly, experiments relating to the mouth features disclosed that ResNet-101 achieves 91.5%. Results from the Machine Learning experiments demonstrate the strength of computer-based algorithms. The

next logical phase of this work is to compare this performance to that of humans.

In the present study, we quantified the accuracy with which the human visual system discriminates between faces of different ethnicities. Further, this work aimed to understand how humans use the visual information contained within faces to make decisions about ethnicity. We focused specifically upon the information used to identify Pakistani faces from those of different ethnicities.

The present study focused on this demographic because we identified a gap in current understanding of processing Pakistani ethnicity. Specifically, there is, currently, a lack of literature that explores strategies of ethnicity classification from a South Asian perspective, even though longitudinal studies into face perception and recognition have been carried out, from a Caucasian (United Kingdom) and an Egyptian perspective (Megreya and Bindemann 2009; Megreya et al. 2012), as well as from an Arab (United Arab Emirates) versus White (United States of America) viewpoint (Wang et al. 2015). Such research brings to the forefront the advantage of culture-based experiences; especially how culture, traditions and experience fine-tune a perceiver's processing skills.

The Pakistani observer's own-race bias (ORB) has been under-investigated. As discussed in the introduction to this thesis, the ORB is the superior ability of a perceiver to recognize people belonging to his/her

same-race, relative to that of other races (see Section 1.4 of Chapter 1 for further detail). The lack of understanding of the ORB with Pakistani people is surprising, given that people of Pakistani ethnicity represent a significant proportion of the population. The United Kingdom houses the largest proportion of Pakistani people, when considering the entire population within Europe and exceeds 1.17 million (*United Kingdom: Minorities and Indeginous People* 2019). Statistics published by the Metropolitan District Council refer to Bradford as having the greatest proportion of Pakistani people in the UK (Council 2019).

While evidence is currently lacking, it seems reasonable to hypothesise that because of the homogeneity of their local community, it is possible that Pakistani people may have developed a more substantial advantage for own-race faces and may exhibit reduced perceptual expertise for other-race faces (Walker and Hewstone 2006). We expect that, since the Pakistani people are possibly within a dense own-race environment, they may demonstrate a strong ORB: a significantly improved ability to recognize fellow Pakistani faces and face-features, relative to those of other races.

The primary aim of this study is to quantify the magnitude of the own race bias for individuals of Pakistani ethnicity living in the UK. Before collecting data, the following points were hypothesised;

(1) Pakistani participants will demonstrate superior ethnicity discrimination accuracy, relative to Non-Pakistani participants. This hypothesis is formulated on the effect of own-race bias in face recognition (Malpass and Kravitz 1969; Meissner and Brigham 2001; Wright et al. 2001). We expect superior own-race face recognition accuracy to extend to ethnicity discrimination accuracy.

(2) Full face images, in which all features will be available, will be more accurately classified, based on ethnicity, than individual face parts (e.g. eyes, nose, mouth) presented in isolation. This is based on the premise that full faces, relative to face parts, hold more information, which may be used to inform the discrimination judgment (Ellis et al. 1979).

6.2   Literature Review

Own race bias (ORB) is the ability to recognise own-race faces more accurately than faces from another race (Malpass and Kravitz 1969). A study by Wright and Boyd tested ORB effects by asking either a Black or White subject to approach either a Black or White member of public, seeking assistance on locating lost items of jewelry. A couple of minutes later, the member of public was approached and asked questions relating to the unknown subject. They were also asked to identify the subject from a sequential line-up and a forced-choice recognition test. In all tasks, the unknown subject was better recognized by the same-race member of public

(i.e. higher accuracy for Black faces by Black participants, compared to Black faces by White participants). A corresponding effect was found for the White participants (Wright et al. 2001).

Several competing theories have attempted to explain the ORB. Firstly, the contact hypothesis which argues that through abundant and accessible contact with own-race faces, an individual naturally conditions his or herself to become an expert at recognition of such faces (Brigham and Malpass 1985; Slone et al. 2000). According to the contact hypothesis, the more experience an individual gains with various racial groups, the more accurate the individual becomes at identifying members of that particular group (Brigham et al. 2007). A related effect, however, is that the individual becomes *less* accurate at recognized faces of other races.

For example, Sangrigoli and colleagues investigated whether the ORB was modifiable in the case of adopted children. The study showed that children born in Korea, to Korean parents, who were subsequently adopted by Caucasian people, and moved to France, later demonstrated superior recognition of Caucasian, relative to Korean, faces. This illustrates that an individual's environment, rather than hard-wired genetics, influences the ORB (Sangrigoli et al. 2005). Relatedly, limited contact with other-race faces also leads an individual to be inexperienced at differentiating between other race faces (Hugenberg et al. 2007).

Early research suggested the effect of ORB as being specific to only White participants since they demonstrated a higher degree of accuracy in

recognising White faces, relative to Black (Malpass and Kravitz 1969). However, this initial view has now been discredited. For example, research by (Malpass et al. 1973; Slone et al. 2000) demonstrated a similar manifestation of the ORB for both White and Black observers. The ORB is reported in the meta-analysis conducted by (Meissner and Brigham 2001) and has also been described to manifest with blurred and scrambled faces (Hayward et al. 2008).

A paradigm of unfamiliar Caucasian and Chinese male faces within two conditions (scrambled and blurred) were shown to participants from two different universities; University of Western Australia (predominantly Caucasian participants) and Chinese university of Hong Kong (predominantly Chinese participants), to test the own-and-other-race- face recognition. A total of 4 conditions were tested: study condition (full face stimuli), scrambled condition (segmented features), blurred condition (full face stimuli blurred with Gaussian filter) and scrambled-blurred conditions, which consisted of the segmented features shown as blurred (Figure 6.1).
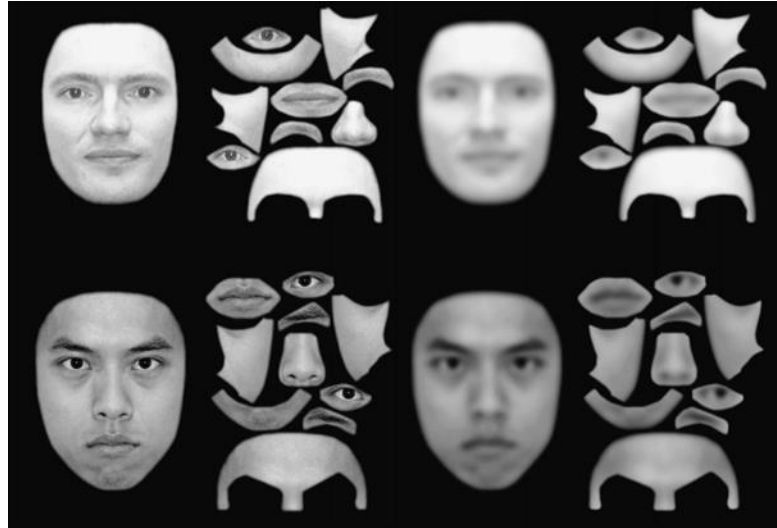
Figure 6:1: A Caucasian and a Chinese face shown as part of 4 investigated conditions for own race bias (ORB): Left: study condition with the full face followed by scrambled condition of 10 segmented face features, followed by the blurred condition for the same full face shown as part of the 'study condition' and the scrambled-blurred condition which shows the segmented features blurred. A Gaussian filter was applied to the images to reduce image opacity (Hayward et al. 2008).

The researchers investigated the effect of own race bias from 2 aspects: full face (configural) information and isolated/ scrambled features. As hypothesised, participants showed strong ORB effects for the blurred full face condition compared to the scrambled, isolated feature condition. This was because the blurred image conditions still allowed visibility of the full face, which subsequently allowed featural and configural processing.

The presence of ORB for the full face (configural) condition was expected, as literature reports configural processing, (the analysis of the face, based on the relative positions of face features), plays an active role in face recognition (Maurer et al. 2002). Further, participants of both groups performed better in face conditions depicting own-race faces, relative to other i.e. demonstrating ORB (Hayward et al. 2008). Moreover, the

conclusions highlight the role of holistic processing for own race bias, whereby a familiar race-face is individualized by the perceiver, in contrast to other 'unfamiliar' race-face, which tend to be assessed on a feature-by-feature basis. With regards to the upcoming study on ethnicity discrimination, the research by Hayward et al. (2008) has implications for the comparison of ethnicity discrimination accuracy for full faces versus individual features.

6.3   Experimental Stimuli

New, photographic stimuli were specifically created for this work. Firstly, a series of high resolution photographs were taken of 200 participants from the Pakistani Face Database. Of the 200 individuals, 100 were of South Asian, Pakistani origin and 100 were of Non-Pakistani origin. The latter category consisted of Caucasian (including Irish), East/Central Asian (Chinese) Middle Eastern (Jordanian, Omani, Syrian) and Black (Nigerian and Ugandan) participants. Those labelled as of Pakistani origin were required to be of Pakistani heritage on both the maternal and paternal side. The gender split for the stimuli was 100 females and 100 males aged between the range of 19 – 56 years old, split equally with 50 males and 50 females per ethnic group. Subjects for the photographs were recruited from among the students of the University of Bradford. Photographed subjects gave informed consent and the procedure was approved by the Chair of Humanities, Social and Health Sciences Research Ethics Panel at the

University of Bradford on 9th September 2015. Images were captured in a darkened room, with a plain white background which was illuminated with a custom-designed lighting set-up (see Figure 6.2, further information can be found in Section 2.6 of Chapter 2). Photographed subjects were asked to look straight ahead and maintain a neutral facial expression.

Figure 6:2: HALO set-up. Left: The subject stands facing the centre tower which comprises one camera flanked by two LED panels. The left and right towers are identical but orientated at 45° left and right. Viewing distance is maintained by asking the subject to stand on a specific mark. To capture right and left profile (i.e. 90°) images, the subject is asked to turn themselves to face the left and right towers respectively. Middle: view of HALO from the subject's perspective. Right: view of HALO from operator's perspective.

A preliminary aim of the present study was to create a face database which is both rich in ethnic diversity and high in ecological validity. There is a need for an open-access, multi-racial and multi-ethnicity face database, which includes subjects of both genders. To maximise ecological validity and emulate naturalistic settings (e.g. an international airport), items of religious dress (e.g. hijab or headscarf) were not removed. Moreover, many of the female subjects were wearing make-up. Objects without ethnicity-specific

associations, on the other hand, were removed such as spectacles, lanyards, high collar jackets, scarfs. In all experiments, the face images were displayed in greyscale and to remove recognition cues from clothing, the faces were also cropped below the neck line, see Figure 6.3.

Figure 6:3: Greyscale full face stimuli (all features present). The left face is a Caucasian male, the middle face is a Pakistani male and the right face is a Caucasian female

To quantify the contributions of different face features to judgements of ethnicity, the full-face images were then manipulated to create images of individual components (eyes, nose and mouth) (see Figure 6.4). To do this, the full-face images (Figure 6.3) were cropped using Adobe Photoshop (CS6 Extended, version 13.0 x64). To ensure consistency across images, a cropping template was created, within which each image was aligned and manually cropped to factor-in individual head shape and headscarf type. Importantly, this approach meant that the individual feature conditions were directly comparable to the full-face condition. For example, the 'nose' individual feature condition contained the same visual information as the nose when it was part of a full face. As a result, ethnicity judgements for full

faces and their individual features can be directly compared.

Figure 6:4: Individual feature conditions. The features have been extracted from the full-face images shown in Figure 2. Specifically, the eyes (left) are extracted from the full face shown in the left of Figure 2. The mouth (middle) is extracted from the full face shown in the middle of Figure 2. The nose (right) is extracted from the full face shown to the right of Figure 2. This set-up ensures that the full face and individual feature conditions are directly comparable.

6.4   Methodology

We aimed to use our new stimuli to quantify the ability of human participants to make ethnicity discrimination judgements based solely on information available from the face. We further aimed to determine if there is any evidence of an own race bias and own gender bias when making this type of face ethnicity judgement.

6.4.1   Participants

To investigate the existence of an own race bias (ORB) for ethnicity classification between faces of Pakistani and Non-Pakistani origin, participants from both ethnic groups were recruited. A total of 74 students from the University of Bradford took part in this experiment, of which 38 were Male and 36 were Females. From the male participants, 15 were of British Pakistani (British in the sense that they were born in the UK, hence nationals) ethnicity while 23 were Non-Pakistani, in comparison the ethnic

split for the female participants was as follows: 18 British Pakistani females and 18 Non-Pakistani females. Collating all this information into the 2-class ethnic groups, a total of 33 Pakistani participants and 41 Non-Pakistani participants were recruited. All the participants reported normal or corrected-to-normal vision. Observers gave informed consent and the study was approved by the Chair of the Biomedical, Natural, Physical and Health Sciences Research Ethics Panel at the University of Bradford on 23rd May 2017.

6.4.2   Design

Participants were categorised by ethnicity (Pakistani/ Non-Pakistani/other ethnicity) and gender (Male/Female). The stimuli on which participants were tested were also organised into categories based on Gender (Male and Female) and Ethnicity (Pakistani and Non-Pakistani).

6.4.3   Tested Stimuli

A total of 800 greyscale images were used for this experiment. This included full face images and images of isolated features (eyes, nose and mouth) extracted from the full faces. Of the 800 images, 200 depicted full faces (with ethnicity and gender balanced). A further 200 depicted the eyes, another 200 depicted the nose, and the final 200 depicted the mouth.

6.4.4   Apparatus

Images were presented, using routines from the Psychtoolbox (Brainard 1997) on a Sony Trinitron CRT monitor (1024 X 768 at 85Hz) of 65 cd/m^2 mean luminance which was controlled by a Mac mini computer. 150 equally spaced grey levels were used to maximize contrast linearity. Participants were seated 1.2m from the monitor. Accurate viewing distance was maintained with a chin and forehead rest. At the test distance, the computer monitor subtended 13.4° by 10.1° of visual angle; one pixel was 0.018°.

6.4.5   Full-Face Procedure

Ethnicity discrimination accuracy was measured with a custom-designed computerised test. On each trial, participants were presented with two simultaneously-visible faces. One of the faces (target) depicted an individual of Pakistani origin, the other face (distracter) belonged to a different ethnicity. The pairings of target and distracter were kept consistent for all participants. The participant was asked to indicate the location (left or right) of the Pakistani face via computer key press. The position (left or right) of the target face (i.e. Pakistani origin) was randomly determined prior to each trial. To minimise eye movements and preclude scrutiny of individual featural cues (e.g. eyebrow width, lip thickness), faces were presented for a maximum of 1 second. After this time, the faces were replaced by a low-level, greyscale luminance noise mask until a decision was made. Early responses (<1s) were accepted and resulted in immediate progression to the next trial.

Ethnicity discrimination accuracy was tested for each gender in separate blocks. Each block comprised 25 trials (i.e. 50 faces in total). At the end of each block, ethnicity discrimination accuracy was calculated as the total number of correct responses as a proportion of the maximum possible score (i.e. 25).

6.4.6   Individual-Feature Procedure

The individual feature procedure was identical to that outlined above for full faces, other than the participants were now shown a single feature (eyes, nose or mouth) only. Within a single block, discrimination accuracy was measured for 25 sets of eyes, noses and mouths (i.e. 75 trial in total). Individual features were presented in a random order using an interleaved design. Accordingly, participants could not predict which feature would be tested on the next trial. As above, ethnicity discrimination accuracy for individual features was tested in gender-specific blocks. Each block comprised 75 trials (25 nose, 25 eyes and 25 mouth). All participants undertook the four experimental blocks in the same order: full face (male), individual feature (male), full face (female) and individual features (female), see Figure 6.5. The photographs used for this experiment were the same for each participant, but the order of their presentation was pseudo-randomized to minimize order effects.
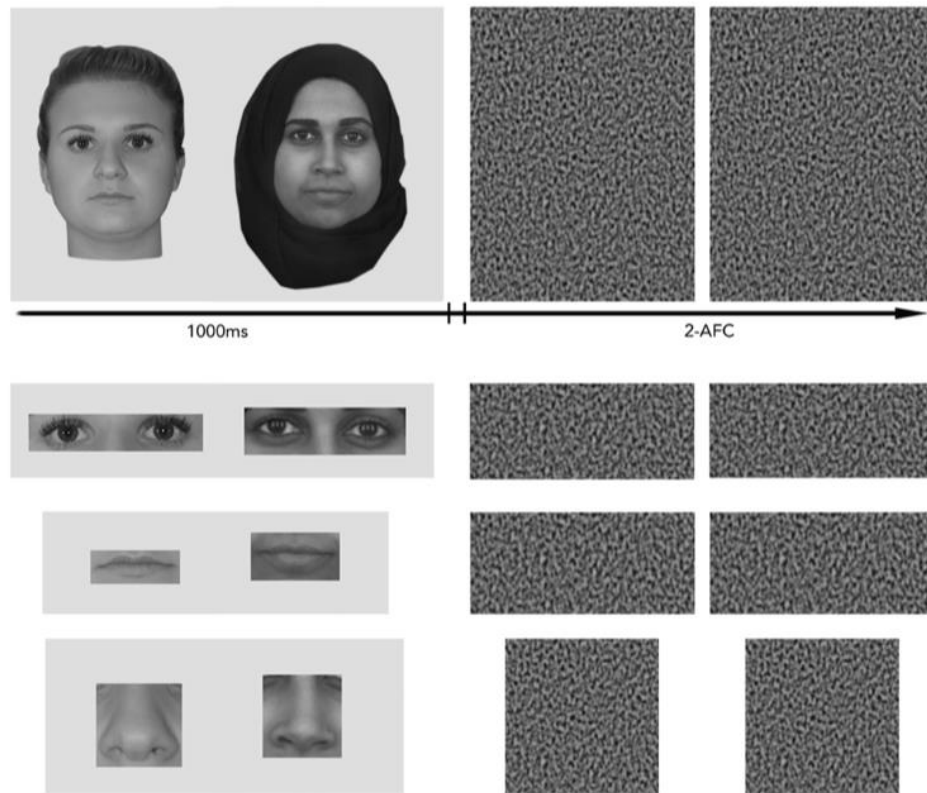
Figure 6:5: Procedure schematic. Top: A single trial for the full-face condition where gender is tested in separate blocks. A pair of target faces are shown for 1000ms, followed by a luminance noise mask, presented until the participant responds. Participants must select the face which they believe is of Pakistani origin. The remainder of the diagram depicts isolated feature (i.e. eyes, nose and mouth) trials. These trials progressed in the same way as for full faces, although participant were now only presented with a single feature on which to base their decision of ethnicity. As before, isolated features were shown in pairs and the participant was required to select the feature which corresponds to an individual of Pakistani origin.

6.5 Experimental Results

The present study measured face ethnicity discrimination ability in participants of both Pakistani and Non-Pakistani heritage.

Since our paradigm included a mixture of male and female participants, we initially sought to determine if performance depended upon participant gender. The mean identification thresholds for participants of each gender per group are presented in Figure 6.6.
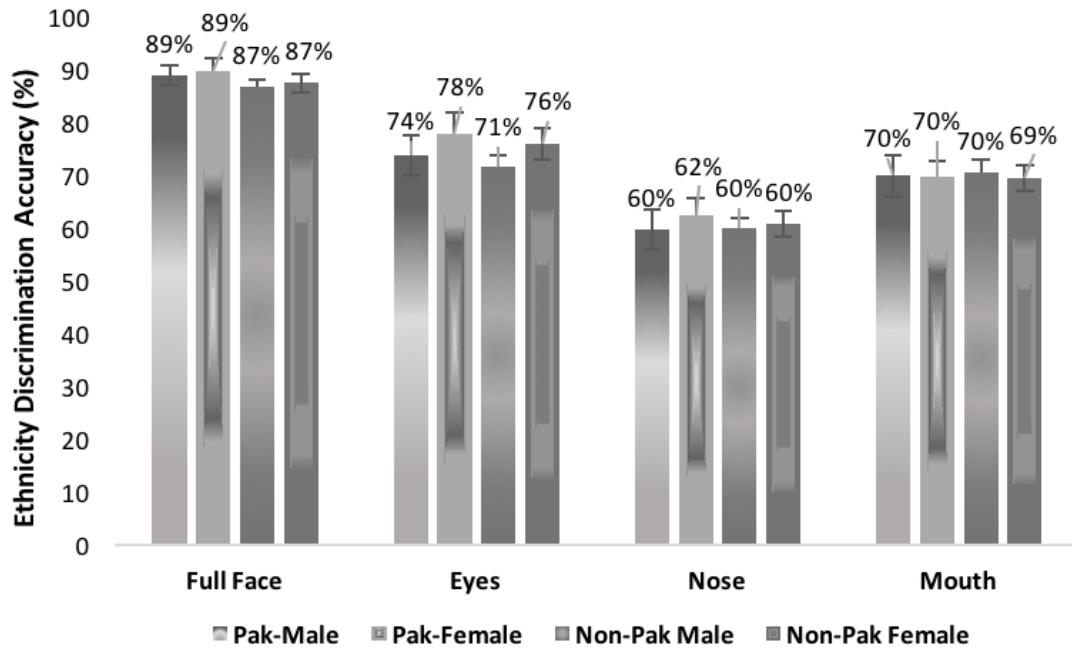
Figure 6:6: Mean participant-gender to feature-gender discrimination thresholds. The dark grey bars represent Pakistani male responses and the light grey bars represent Pakistani female responses. The gradient filled, dark grey columns represent Non-Pakistani male responses and the while the lighter gradient filled columns represent the Non-Pakistani female responses. The results are split into specific isolated gender features, across all four viewing conditions i.e. full face, eyes only, nose only and mouth only. The error bars denote 95% confidence Intervals.

A two-factor (feature and participant gender) repeated-measures, mixed ANOVA found no significant effect of participant gender on ethnicity

147

classification ability ($F_{3,216}$ =3.08; p = 0.29). This lack of an effect of gender is, perhaps, unsurprising since gender was counter-balanced in the stimuli. This leaves no scope for the own gender bias to influence the overall accuracy levels.

Visual inspection of Figure 6.7 also suggests that full faces are classified more accurately than component features (e.g. eyes, nose, mouth). The effect of face feature on performance will be analysed in detail later in this section.
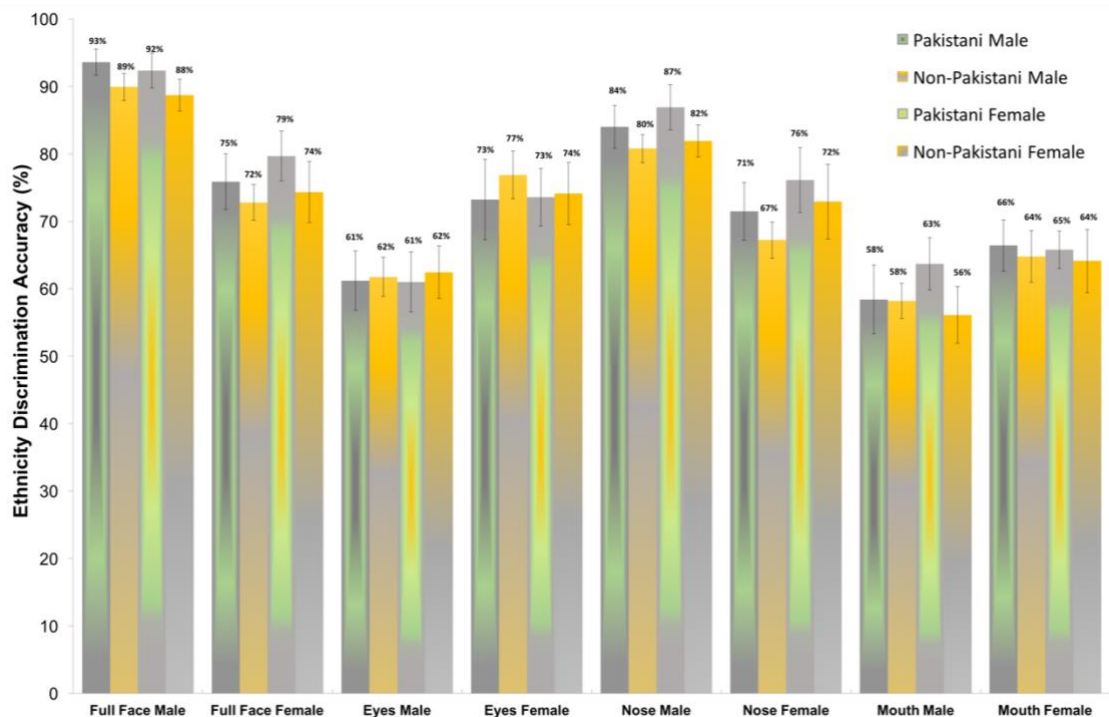
Figure 6:7: Averaged ethnicity classification results for the effect of participant-gender bias. Dark grey denotes Pakistani male responses, light grey denotes Pakistani female responses. While, Dark gradient filled grey denotes Non-Pakistani male responses and light gradient filled column represents the Non-Pakistani female responses. The error bars represent 95% Confidence Interval.

Further, upon inspecting Figure 6.7, there is no evidence that male participants performed better with male faces and, similarly, no evidence that female participants performed better with female faces. This result suggests that there is no own gender effect for this type of face ethnicity discrimination task.

The data from Figure 6.8 indicates that performance was poorest for the nose. This suggests that the nose was the feature which participants found it most difficult to base judgements of ethnicity, even though the nose is the most central and protrusive feature of the face (Harkema et al. 2006).
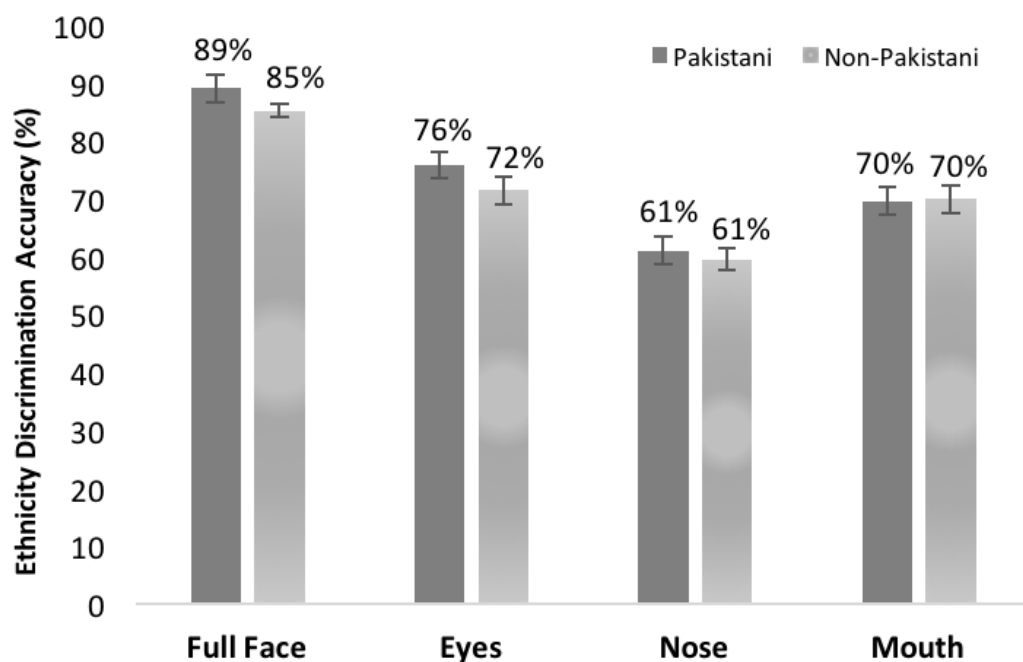
Figure 6:8: Performance Accuracy between Pakistani (dark grey column) and Non-Pakistani (lighter toned column) Observers for the Classification of Ethnicity from Face Images and Face Features).

Figure 6.8 presents ethnicity classification accuracy, measured for Pakistani (dark bars) and Non-Pakistani (light bars) participants, for each of the different face features that were tested. A two-factor, mixed repeated measures ANOVA (participant's ethnicity; Pakistani and Non-Pakistani); Face features [full face, eyes, nose and mouth]) identified a significant main effect of face features ($F_{3,216}= 304.255$; $p =0.028$). Full faces are the most informative sources for ethnicity classification. While, all tested conditions are above the chance boundary i.e. > 50%, the feature least representative of ethnicity, is the nose, see Figure 6.8. Both Pakistani and Non-Pakistani participants demonstrated the same pattern of reliance on individual features: accuracy was highest for full faces and poorest for the nose. The eyes and mouth led to intermediate levels of performance.

The two-factor, mixed repeated measures ANOVA (participant's ethnicity [Pakistani and Non-Pakistani]; Face features [full face, eyes, nose and mouth]) also revealed a main effect of participant ethnicity, (repeated measures ANOVA; $p<0.05$). Pakistani participants outperformed participants of other-ethnic backgrounds (pairwise comparisons with Bonferroni correction; $p = 0.039$), with this effect appearing to be mainly driven by the full-face condition (Pakistani participant accuracy = 89%, Non-Pakistani participant accuracy = 85%). The mean thresholds are shown in Figure 6.8. Features alike to one's own-race are typically easier to recall and this is also evident in Figure 6.8; however, the nose as a stand-alone feature still poses difficulty.

The experiments conducted so far directly compare ethnicity discrimination thresholds for 2 ethnically distinct cohorts (Pakistani and non-Pakistani), for a range of visual conditions, i.e. full face, eyes, nose and mouth. In sum, the results of the present study thus far support the initial hypotheses; (1) Judgements of ethnicity are significantly more accurate when based upon full-face images, compared to individual features presented in isolation (2) there is no significant effect of participant gender on the ability to make judgements of ethnicity and finally, (3) participants of Pakistani heritage performed significantly more accurately than those of other ethnicities in an ethnicity categorisation task which was based on Pakistani faces. This latter finding is suggestive of an own race effect and extends the premise of this phenomenon from identification of individual faces to judgements of ethnicity.

To investigate if differences in display (i.e. 2-Alternative-Forced Choice (2-AFC) or single image option), affect the magnitude of the own race bias; 11 students were recalled to undertake a modified version of the ethnicity discrimination task, 12 months after the first session. The reason for this additional work was to investigate the possibility that participants made their ethnicity discrimination judgement by the process of elimination. This may have been in the case in some trials where a hijab wearing face was shown beside a Caucasian face. However, equally there have been more difficult trials where 2 faces of hijab wearing female faces were shown. Thus, it was considered important to understand whether a difference in

display format affected performance. Especially, the modified (single option) paradigm challenges the participant solely on the face and feature presented, without any comparison. More importantly, the single option task allows for a fairer comparison with the Machine Learning experiments. This is since the Machine Learning framework processes the input data i.e. face images, one at a time. The 11 re-called participants consisted of 8 Pakistani, 3 non-Pakistani, 7 females and 4 males in total.

One possible outcome was that a single image display would reduce cognitive load by allowing the observer to remain fixated on a single image, instead of looking between two. By doing so, the task could be relatively simpler and thus there is the possibility of a higher accuracy rate. However, the observed effects are not in accordance with this hypothesis: observers performed less accurately in the single, compared to the 2-AFC condition. A repeated measures ANOVA confirmed that this difference was significant (p<0.05).
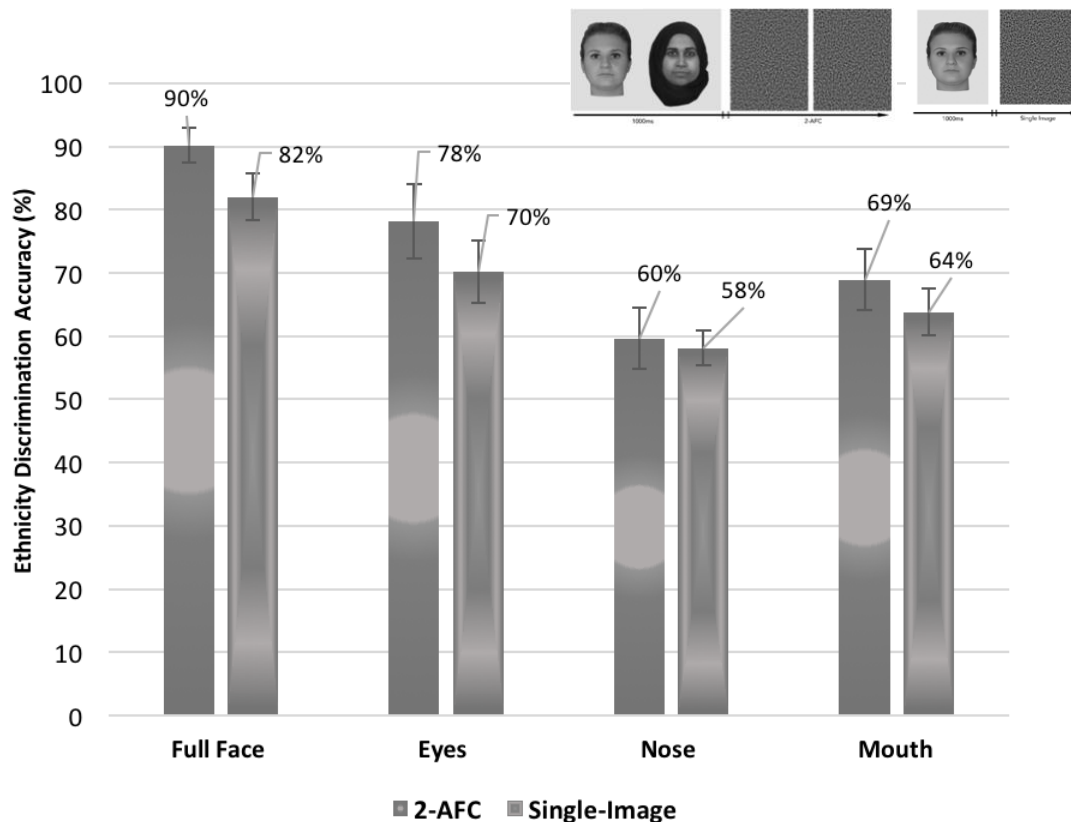
Figure 6:9: Classification of the Pakistani ethnicity from images shown as 2-AFC (Two-Alternative Forced Choice; dark column) and as a single image display (lighter, designed column).

Figure 6.9 clearly shows that given 2 options of images, the full face is the most informative ethnic marker with an accuracy performance of above 90%, followed by the eyes, mouth and nose at 78%, 71% and 59%, respectively. Yet, the most difficult feature to make assertions of ethnicity from is the most centralised i.e. the nose. Firstly, it is important to recognise that participants demonstrated the same pattern of results when tested by the 2-AFC and single image paradigms. Specifically, performance was highest for the full-face condition and lowest in the nose condition in both experiments. When presented single images, rather than a binary choice, however, graph 6.9 suggests that there is an overall reduction in

discrimination accuracy. This global reduction in sensitivity indicate that in trials where a target face is presented alongside a distractor, the participant is more likely to select the correct choice, compared to when a single image is shown. This may be explained by the process of elimination, whereby an observer may make comparisons between the two images and select the one which they consider to appear most similar to a face of Pakistani heritage; something which is not possible when a single image is shown.

6.6 Discussion

The human ethnicity discrimination data reveals that participants could perform ethnicity discrimination at a level beyond that which would have been expected based on chance performance. Importantly, the results are novel because they show that own race bias extends to an ethnicity classification task, i.e. participants demonstrated improved performance when judging the ethnicity of own-race, relative to other race faces and component features. These results are in-line with literature which examined the robust phenomenon of Own-Race Bias (ORB), where an observer recognizes faces from one's own-racial group with an increased accuracy, relative to others (Malpass and Kravitz 1969; Brigham and Malpass 1985; Bothwell et al. 1989; O'toole et al. 1994; Meissner and Brigham 2001).

In addition to extending the premise of an own-race bias to an ethnicity classification task, a major finding of the present study relates to

the relative importance of individual face features. Firstly, the data clearly show that ethnicity classification judgements were most accurate when observers viewed full faces, as opposed to individual features. This is in line with the premise of holistic processing- the whole face is greater than individual parts (Hancock et al. 2000; Megreya and Burton 2006; Johnston et al. 2013; Richler and Gauthier 2014). The results presented here are consistent with the seminal work of Tanaka and Farah (1993) who investigated the difference in recognition between whole faces and isolated face-features. These authors showed that recognition was significantly more accurate when participants were presented with whole faces, compared to isolated features (Tanaka and Farah 1993).

The present study also found that the nose was the least informative features (compared to eyes and mouth) for ethnicity classification judgements. It may be that while the nose is the most central, invariant feature of the face, it is not a descriptive ethnic marker, when shown in isolation without any face context i.e. in the presence of surrounding face features. This finding may seem surprising, given that the nose if the most central feature of the face, and the one which remains constant over time (unlike the mouth, for example, which undergoes significant deformations in shape and size to portray facial expressions). Importantly, however, our results also reveal that face processing depends strongly on ethnicity. In line with this premise, Wang et al. (2015) demonstrated that Emirati (United Arab Emirates Native) participants significantly outperformed Caucasian

participants on an identity recognition task in which only information from the nose region was visible. The authors proposed that this result could be explained by the fact that Emirati males commonly practice the custom of greeting friends and family with a nose touch. As a result, Emirati people may encode visual information about the nose in a different way to that of Caucasians. Other evidence points to the conclusion that there are significant differences in the strategies employed for face processing for people from different racial backgrounds. Specifically, Blais et al. (2008) reported a striking cultural contrast demonstrating that while Western Caucasian participants preferentially fixate the eye region, East Asian participants were drawn to the centre of the face (i.e. nose region), during face processing.

6.7   Conclusion

The present study investigated the ability of the human visual system to make judgements about the ethnicity of individuals based on face information alone. Our data clearly demonstrate that- despite the lack of colour information (all images were presented in grayscale)- observers could perform ethnicity categorization at a level significantly better than would be expected based on chance performance. A particularly novel aspect of our results is that they provide evidence of the own race bias extending from recognition of identity to the classification of ethnicity. Specifically, our results highlight an own-race bias for an ethnicity classification task- participants of a Pakistani background demonstrated

significantly improved accuracy, relative to participants of other ethnicities. A significant effect of face features is also reported: we found that full faces are the most revealing sources of information for ethnicity classification. While all other tested conditions (eyes, nose and mouth) are above the chance boundary (i.e. > 50%), the feature least informative of ethnicity is the nose, possibly due to a lack of contextual information from surrounding face features or because of a reduction in race specific information, which is visually attainable from nose images.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1   Overview

The classification of ethnicity from face images has the potential to be a particularly informative biometric trait for the field of Forensic Science, Policing and intelligence gathering. A significant advantage of ethnicity classification, is its non-invasive nature. In the field of Machine Learning, the challenge of heritage classification (as an over-arching concept incorporating race and ethnicity, which is based on superficial differences in face morphology, and metric differences between face features) has been addressed. One of the considered novelties of the research presented as part of this thesis is the use of a human population that has not been researched previously; the South Asian, more specifically, the Pakistani ethnicity.

Discussions surrounding the definition of ethnicity are unresolved, although the consensus agrees on the complexity surrounding the definition. Factors combining biological and social traits may assist in defining variations of human population. However, Epigenome-Wide association studies (EWAS) have reported that variations do exist in DNA methylation (DNAme) between different human populations (Fraser et al. 2012; Banda et al. 2015; Yuan et al. 2019). Thus, providing evidence for

158

some biological basis of ethnicity in humans.

To address the challenge of ethnicity determination, a new framework is proposed for the Pakistani ethnicity, using a two-tier, human and machine perspective. Fundamental to the processing of demographic information, is the requirement of abundant training data. Therefore, a novel facial image database, entitled Pakistani Face Database (PFDB) has been created. The principal measure of ethnicity classification was investigated from three distinct routes; (1) Low-level features, (2) High-level features and, (3) the Human visual system.

In Chapter 1, a detailed stratification of distinct human populations into sub-groups, based on differences in facial metrics, i.e. anthropometry, is presented. Moreover, the significance of race and ethnicity labels was discussed with a real-world application i.e. Policing. This was followed by a Section on the defined research problem and proposed solution, for this thesis. In Chapter 2, a new face-image dataset, the Pakistani Face Database (PFDB) was discussed and detailed information relating to the hardware i.e. Halo, used to capture the images. The use of images is a less-intrusive form of testing any hypotheses. A first of its type, the PFDB is the only database to offer images of South Asian, Pakistani participants and includes female participants with headscarves. Such images may be useful to further investigate the headscarf-effect, for example (Megreya and Bindemann 2009; Megreya et al. 2012; Toseeb et al. 2014). Notably, the

database bridges an obvious gap between multi-view, high-quality images and ethnicity specific data exclusive to the South Asian, Pakistani population. Although the database does contain participants of other racial and ethnic backgrounds.

Formal work is presented as part of Chapter 3, which investigated the enriched nature of hand-crafted features to determine ethnicity, using 2 separate experimental frameworks. Firstly, using front-view face images, 16 individually marked feature points were manually selected. Principal Component Analysis (PCA), algorithm was used to remove redundant information and a linear classifier (Support Vector Machine) used for binary classification. The results reported were not exceptional, but promising since a performance accuracy of 65.71% was concluded. In contrast, the second analyses, exhibited a stronger correlation between face and ethnicity. By using a more robust dimensionality reduction algorithm, Partial Least Square (PLS), and a lower number of hand-crafted features i.e. 10, which outlined the facial profile within an image. The classification of Pakistani ethnicity was increased; 71.42%. While the initial experimentations are successful in demonstrating the face as being a robust indicator of ethnicity, there are limitations to using hand-crafted (manually-mapped) features. Firstly, the low classification accuracy and secondly the time-consuming nature of individually marking images with facial points. Nonetheless, the experiments are considered successful as they captured essential baseline information, which demonstrates that the classification of

the Pakistani ethnicity is achievable.

To further the investigation into the strength of face information for ethnicity classification, a Deep Learning based framework is presented in Section 4.4 of Chapter 4. Deep Learning based experimentations have achieved state-of-the-art classifications on several image-based tasks and therefore is considered a rational approach. The Police, for example classify perceived race/ethnicity as physical features such as hair and skin colour, making it a subjective assessment. However, the classification of ethnicity from automated methods will remove bias, error (to an extent) and the efficiency of functioning will directly reduce issues surrounding time constraints.

In the proposed framework, deep face features are extracted using pre-trained Deep Learning models, that are transferred to a robust binary classifier to allow for ethnicity classification of the Pakistani face. In an approach to test the informative nature of individual face segments, internal features from the "T-Zone" region of the face, namely the eyes, nose and mouth have also been individually scrutinized. In isolation, each of the feature components concluded high accuracy for ethnicity classification, but the nose and mouth marginally outperformed the eyes. The experimental results demonstrate ResNet-101 achieves the highest performance accuracy of 99.2% for full-face ethnicity classification, followed closely by 91.7% and 95.7% for the nose and mouth respectively. In comparison, by

using VGG-based models, a performance accuracy of 93.2% was achieved for the binary classification of a Pakistani face using VGG-16. Whereas, when the VGG-based algorithms were applied to internal face components, i.e. eyes, nose and mouth, accuracy did not reach above 90% although the results were above chance level (>50%). Interestingly though, the results did follow the same trend as seen with the ResNet algorithms because the nose and mouth features were classified with more accuracy; 82.9% (VGG-19) and 86.7% (VGG-16) respectively. The indications from the experimental findings of Chapter 4 provide a good base for continuing the pursuit of further scrutinizing, the informative threshold of nose and mouth features, for the problem of ethnicity classification.

An adaptive framework of high-level face data is proposed in Chapter 5, in which the same 6 pre-trained Deep Learning models were now applied to challenging datasets of nose and mouth images (2,200 images per feature), in addition to datasets of feature combinations. ResNet-50 achieved results above 90% with an accuracy of 94.1%, while VGG-16 outperformed VGG-F and VGG-19 reporting 90.8% accuracy. The experimental results also reported that the nose and the mouth as a combined feature-set, consecutively reported the highest performance accuracy, relative to- (1) eyes and nose and (2) mouth and eyes. By directing an extensive range of computational experimentation which implement Deep Learning models, it is concluded that ethnicity verification is feasible from the face (1) holistically and (2) feature-specifically. The

frameworks applied to the image classification tasks are essentially time-efficient and repeatable. In addition, the performance achieved for each experimentation block i.e. hand-crafted feature, low-level features and high level features, are an improvement on previously published data, in terms of the implemented processes only because they have not investigated the Pakistani ethnicity. In some respects, the data used and the performance accuracy achieved are novel and thus, they can be used as a benchmark for further work.

In summary, all the Machine Learning frameworks presented as part of the computer-assisted experiments report an unbiased and efficient classification of face and face components for the Pakistani ethnicity. The experimental conditions which have tested low-level features (i.e. chapter 3) demonstrate the robustness of the semi-automatic approach to ethnicity classification, even though it is based on the manual annotation of face features, which presents a challenge within itself, especially for issues relating to experimental repeatability. Even though the experimental framework was predominantly devised as a baseline experiment, testing the informative nature of face features for ethnicity. By progressing to a fully automatic framework (as seen in experimental chapters 4 and 5), it is apparent that the employment of Deep Learning methodologies not only increased the recognition rates significantly. But, when comparing the experimental frameworks from Chapter 4 and 5 to Chapter 3, technical challenges such as facial occlusion i.e. individual face-components (eyes,

nose, mouth), readily assigned ethnicity.

The final experimental Chapter of this thesis investigated human ethnicity discrimination. Participants undertook a computerised ethnicity discrimination test, which showed a series of face photographs on the screen and participants were prompted to make decisions about them. Specifically, participants were asked to indicate which of two presented faces appeared to belong to an individual of Pakistani origin. On some trials, full face images were shown; in others, specific features (e.g. eyes, nose, mouth) were presented in isolation. We found that participants of Pakistani origin correctly identified Pakistani face and face-features relatively higher to those of another ethnicity. This extends the premise of an own-race bias from face recognition to an ethnicity discrimination task. While ethnicity verification from full-face images was considered the easiest, the nose proved to be the most challenging feature for both participant groups, suggesting that the nose carried limited information about ethnicity

In conclusion, the experimental findings suggest that while humans are good at recognizing full-faces due to holistic processing (Hancock et al. 2000; Megreya and Burton 2006; Johnston et al. 2013). Humans struggle to discern ethnicity from images of the nose in isolation, irrespective of self-reported ethnicity. There is a range of published literature which supports the notion that there are both gender and ethnic-specific differences in the shape and size of a human nose (Ozdemir and Uzun 2015; Zaidi et al. 2017;

Elsamny et al. 2018; Mohammed et al. 2018). Similarly, there have been a handful of studies that have investigated the nose as a standalone marker for ethnicity (Zehngut et al.; Song et al. 2009). Lv et al. (2018) reported ethnicity classification rates of 89.2% and 97.4% for the White and Asian population, respectively. The results are in line with those reported for high-level face features (Chapter 5), 94.1% for ResNet-101 and 90.6% for VGG-16 and is successful in highlighting the richness of the feature.

7.2 Limitations and Future Work

While the results for the image-based ethnicity classification task are encouraging, there are several possibilities from unexploited perspectives, which can be considered to further enhance the results. In the following, recommendations are made as to what future directions should be explored to assist in advancing the current accomplishments relating to image-based ethnicity classification. For ease, the recommendations are grouped into separate points and include: (1) database improvements, (2) reduced features (3) international perspective to ethnicity discrimination.

7.2.1 Database Improvements

The Pakistani Face Database (PFDB) was created using a hardware designed for Police Custody suites, allowing the capture of 5 images, from different viewpoints, for each subject. As previously described in Section 2.6 of Chapter 2, the process of image capture was conducted in 2 phases.

There are ways in which the database can be enhanced, which would subsequently increase consistency in the recorded information.

First, the participants of Pakistani heritage were all recruited at the University of Bradford, with a large proportion originating from Bradford. The limitation this poses, is that of regional homogeneity whereby the participants of Pakistani heritage only represent an insular region (i.e. Bradford) of a wider country (i.e. United Kingdom). Thus, it can be argued that the facial aesthetics are not universal nor representative of every 'Pakistani' face. This is somewhat true however, it would be virtually impossible to collate face images of the entire Pakistani population, thus a decision to limit data collection to the University of Bradford, was considered suitable. To ensure a wider inclusivity of facial morphology for participants belonging to the Pakistani heritage, further image capture session can be coordinated.

Age is another factor for consideration when determining the wider applications of the dataset. The human face undergoes a range of sequential changes relating to skin texture, facial features, facial hair and weight (Coleman and Grover 2006). Since participants were actively recruited at the university, the consensus may be that such an approach is restrictive. However, the decision to select participants from the older student population (i.e. university students) was deliberate, especially when considering the physical stability of adult (post-pubescent) face features, relative to children. The Facial Identification Scientific Working Group

166

(FISWG) document an extensive list of facial features that are categorised as either high, moderate or low in physical stability, subject to changes based on short-term (period of 5 years or less) and long-term (excess of 5 years) changes, including considerations relating of weight, health and intentional facial modifications (FISWG 2019). The FISWG research indicates that the position of internal features (eyes, nose and mouth) remains stable under most conditions.

Whilst female participants were requested to keep make-up to a minimum for their photography session, it was difficult to maintain. There were a few participants who wore a lot of makeup, because of which their images were not used for any experiments. A way of improving this issue would be to make clear from the initial point of recruitment that heavy make-up is a barrier to inclusion. Additionally, during each photography session, participants were instructed to stand on a foot placement marked onto the floor. While this placement functions to maintain a known distance between the participant and the camera panel, it does not assist with posture/body position. If the body is not positioned correctly, then participants tended to lean back, which positions their face further away from the centre of the screen.

As a point of note for future data collection, a seat could be used to allow the participants to sit upright during image capture session. There are other commercially known custody image capture systems, which also require the user to be seated instead of standing, such as the Viper ID Booth

(Bureau 2017). By unifying the approach to participant body posture, during image capture, the face would align accurately. Such normalisation may prove beneficial to the process and assist in efficiency of image capture. The Pakistani Face Database (PFDB) consists of 224 participants of Pakistani origin. Although this is a considerable number, it cannot represent the facial traits of all Pakistanis. Hence there is a need to create a larger dataset of Pakistani face images, which is balanced in terms of gender. The associated advantage is a greater, in-depth understanding of the differences in facial traits between males and females.

7.2.2 Feature Reduction

The Machine Learning experiments described as part of this thesis have relied on pre-trained Convolutional Neural Networks and Deep Learning models. The process of ethnicity classification has almost been sequential with feature extraction first, followed by binary classification. VGG-F for example, extracts 4096 features per input, some of which may contain redundant information. While, the model is trained on an extensive dataset of 2.6 million images of 2,600 people, the dataset is not pure in terms of ethnicity, since the model is evaluated on a few databases such as; Labelled Faces in the wild (LFW), and YouTube Faces in the Wild (YFW). The datasets are not designed to discern demographic attributes but in fact, are used for largescale face recognition and verification tasks.

Further work can be conducted to reduce the number of features which are extracted by creating a purpose-built ethnicity verification model,

which is trained on images of full faces, partial faces and isolated face features of South Asian identities, including Pakistani. The model would be trained and tested on Pakistani images with an in-built classifier to allow efficient classification. More importantly, the images with which the model would be tested would include ladies wearing the headscarf of different variations i.e. hijab, turban, as this would allow diversity and ecological validity to the data. In addition, it would ensure robustness of the model as it will be tested on a range on input data.

7.2.3   International Perspective to Ethnicity Classification

Presently, the participants who contributed in taking the computerised ethnicity test, were all recruited at the University of Bradford. A future consideration is to request students belonging to 2 distinctly separate regions of different demographic profiles, to take part. The test could be conducted with students in a predominately Caucasian region such as South West England.

According to the most recent consensus, there is very limited ethnic diversity in Plymouth and it is 92.9% White British (Public Health 2014). It would be interesting to see how the discrimination ability performance varies between students in Plymouth and Bradford. It is hypothesised that participants belonging to a predominantly Caucasian region would perform poorly on the face ethnicity discrimination task and may struggle equally between the full face and isolated features i.e. eyes, nose and mouth, condition.

7.2.4 Masked-Face Ethnicity

Given the global spread of COVID-19 (Coronavirus) there is an increased awareness internationally for the use of face masks in communal areas including university campuses and shopping centers. The World Health Organization (WHO) recommends the use of a mask as (i) a form of protection and (ii) for a source of control i.e. onward transmission (WHO 2020). When considering the current environmental climate and the necessity for face covering, there have been increased efforts towards understanding the 'face mask effect.' Authentication applications such as restricted access buildings and face-related payment systems, (which rely on facial recognition) have met issues when presented mask-obscured face images. Subsequently, there has been a surge in the development of operative technology and masked face-image datasets (Wang et al. 2020). Especially when considering that pre-pandemic facial recognition technologies were not routinely applied to masked face images. When considering the face mask effect, the experiments conducted on the individual face features, (i.e. eyes, nose and mouth) are the most prominent and in-line with the regions of the face which contribute to the face-mask effect. Additional work can be conducted to investigate the implications of the face mask effect for the challenge of ethnicity classification. Moreover, since there are numerous styles of face-coverings which disguise mixed regions of a person's face, it would be a novel investigation to see how

ethnicity is discerned. Especially when considering that currently, the domain of masked face recognition remains an immature field of research.

BIBLIOGRAPHY

Abrams, R. A. and Christ, S. E. (2003) Motion onset captures attention. *Psychological Science* 14 (5), 427-432.

Abdi, H. and Williams, L. J. (2010) Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2 (4), 433-459.

Ahn, C. H., Choi, J.-W., Beaucage, G., Nevin, J. H., Lee, J.-B., Puntambekar, A. and Lee, J. Y. (2004) Disposable smart lab on a chip for point-of-care clinical diagnostics. *Proceedings of the IEEE* 92 (1), 154-173.

Akyildiz, I. F., Lo, B. F. and Balakrishnan, R. (2011) Cooperative spectrum sensing in cognitive radio networks: A survey. *Physical Communication* 4 (1), 40-62.

Akyildiz, I. F., Melodia, T. and Chowdhury, K. R. (2007a) A survey on wireless multimedia sensor networks. *Computer networks* 51 (4), 921-960.

Akyildiz, I. F., Melodia, T. and Chowdury, K. R. (2007b) Wireless multimedia sensor networks: A survey. *Wireless Communications, IEEE* 14 (6), 32-39.

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. (2002) A survey on sensor networks. *Communications magazine, IEEE* 40 (8), 102-114.

Al-Jawfi, R. (2009) Handwriting Arabic character recognition LeNet using neural network. *Int. Arab J. Inf. Technol.* 6 (3), 304-309.

Alexander J Martin (2017) *Police facial recognition trial led to 'erroneous arrest'.* Sky News. https://news.sky.com/story/police-facial-recognition-trial-led-to-erroneous-arrest-11013418 Accessed 18th May.

Aly, S. A. (2012) Pilgrims Face Recognition Dataset--HUFRD. *arXiv preprint arXiv:1205.4463.*

American Society for Gastorintestinal Endoscopy. http://www.asge.org/patients/patients.aspx?id=390

Andrews, T. J. and Thompson, P. (2010) Face-to-face coalition. *i-Perception* 1 (1), 28-30.

Angier, N. (2000) Do races differ? Not really, genes show. *New York Times* 22, F1.

Anwar, I. and Islam, N. U. (2017) Learned features are better for ethnicity classification. *Cybernetics and Information Technologies* 17 (3), 152-164

Aspinall, P. J., Song, M. and Hashem, F. (2006) Mixed race in Britain: a survey of the preferences of mixed race people for terminology and classifications.

Association of Chief Police Officers of England and Ireland, W. a. N. (2002) *Guide to Self-Defined Ethnicity and Descriptive Monitoring.* Business, A. R. A. D. and Area. https://www.npcc.police.uk/2017 FOI/IM/029 17 16 plus 1 Policy.pdf

Attneave, F. (1954) Some informational aspects of visual perception. *Psychological review* 61 (3), 183.

Ayuk, R. E. (1990) Cross-racial identification of transformed, untransformed, and mixed-race faces. *International Journal of Psychology* 25 (2), 509-527.

Azzakhnini, S., Ballihi, L. and Aboutajdine, D. (2018) Combining Facial Parts For Learning Gender, Ethnicity, and Emotional State Based on RGB-D Information. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14 (1s), 19.

Baig, M. M. and Gholamhosseini, H. (2013) Smart health monitoring systems: an overview of design and modeling. *Journal of medical systems* 37 (2), 1-14.

Bagchi, C., Amali, D. G. B. and Dinakaran, M. (2019) Accurate Facial Ethnicity Classification Using Artificial Neural Networks Trained with Galactic Swarm Optimization Algorithm. *Information Systems Design and Intelligent Applications.* Springer. 123-132.

Barkowitz, P. and Brigham, J. C. (1982) Recognition of Faces: Own-Race Bias, Incentive, and Time Delay 1. *Journal of Applied Social Psychology* 12 (4), 255-268.

Basar, M. R., Malek, F., Juni, K. M., Idris, M. S. and Saleh, M. I. M. (2012) Ingestible wireless capsule technology: A review of development and future indication. *International Journal of Antennas and Propagation* 2012.

Bastanfard, A., Nik, M. A. and Dehshibi, M. M. (2007) Iranian face database with age, pose and expression. *Machine Vision*, 50-55.

BBCResearch BBC Research. http://www.bccresearch.com/report/minimally-invasive-medical-devices-hlc051e.html

Becerra-Riera, F., Llanes, N. M., Morales-González, A., Méndez-Vázquez, H. and Tistarelli, M. On Combining Face Local Appearance and Geometrical Features for Race Classification. 2018. Springer.

Bengio, Y., Goodfellow, I. J. and Courville, A. (2015) Deep learning. *Nature* 521 (7553), 436-444.

Bergmann, J. H., Chandaria, V. and McGregor, A. (2012) Wearable and Implantable Sensors: The Patient's Perspective. *Sensors* 12 (12), 16695-16709.

Bhopal, R. (2004) Glossary of terms relating to ethnicity and race: for reflection and debate. *Journal of Epidemiology & Community Health* 58 (6), 441-445.

Bianconi, F. and Fernández, A. (2014) An appendix to "Texture databases–A comprehensive survey". *Pattern Recognition Letters*.

Bianconi, F., Harvey, R., Southam, P. and Fernández, A. (2011) Theoretical and experimental comparison of different approaches for color texture classification. *Journal of Electronic Imaging* 20 (4), 043006-043006-17.

Bicen, A. O., Gungor, V. C. and Akan, O. B. (2012) Delay-sensitive and multimedia communication in cognitive radio sensor networks. *Ad Hoc Networks* 10 (5), 816-830.

Bilstrup, K. (2008) A preliminary study of wireless body area networks.

Blakemore, C. and Over, R. (1974) Curvature detectors in human vision. *Perception* 3 (1), 3-7.

Blumenbach, J. F. and Bendyshe, T. (1865) *The Anthropological Treatises of Johann Friedrich Blumenbach.* Anthropological Society.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D. and Caldara, R. (2008) Culture shapes how we look at faces. *PloS one* 3 (8), e3022

Boesen, K., Rehm, K., Schaper, K., Stoltzner, S., Woods, R., Lüders, E. and Rottenberg, D. (2004) Quantitative comparison of four brain extraction algorithms. *NeuroImage* 22 (3), 1255-1261.

Bolnick, D. A. (2008) Individual ancestry inference and the reification of race as a biological phenomenon. *Revisiting race in a genomic age*, 70-85.

Bon-Woo, H., Byun, H., Myoung-Cheol, R. and Seong-Whan, L. Performance evaluation of face recognition algorithms on the Asian face database, KFDB. 2003. Springer.

Bonner, L., Burton, A. M. and Bruce, V. (2003) Getting to know you: How we learn new faces. *Visual cognition* 10 (5), 527-536.

Bothwell, R. K., Brigham, J. C. and Malpass, R. S. (1989) Cross-racial identification. *Personality and Social Psychology Bulletin* 15 (1), 19-25.

Boureau, Y. L., Ponce, J. and LeCun, Y. A theoretical analysis of feature pooling in visual recognition.   2010.

Brainweb *BrainWeb: Simulated Brain Database.* http://brainweb.bic.mni.mcgill.ca/brainweb/

Brigham, J. C. and Malpass, R. S. (1985) The role of experience and contact in the recognition of faces of own-and other-race persons. *Journal of social issues* 41 (3), 139-155.

Brigham, J. C., Bennett, L. B., Meissner, C. A. and Mitchell, T. L. (2007) The influence of race on eyewitness memory. *The Selected Works of Christian A. Meissner, Ph. D.*, 28

Brooks, K. R. and Gwinn, O. S. (2010) No role for lightness in the perception of black and white? Simultaneous contrast affects perceived skin tone, but not perceived race. *Perception* 39 (8), 1142-1145.

Brown, A. G. (2012) *Nerve cells and nervous systems: an introduction to neuroscience.* Springer Science & Business Media.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M. and Miller, P. (1999) Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied* 5 (4), 339.

Bukar, A. M. and Ugail, H. (2017) Automatic age estimation from facial profile view. *IET Computer Vision* 11 (8), 650-655.

Bukhari, A. A. (2011) The distinguishing anthropometric features of the Saudi Arabian eyes. *Saudi Journal of Ophthalmology* 25 (4), 417-420.

Buolamwini, J. and Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification.   2018.

C/I Kevin Bowsher, D. a. C. F. D. M. (2007) *The code systems used within the Metropolitan Police Service (MPS) to formally record ethnicity.* http://policeauthority.org/metropolitan/publications/briefings/2007/0703/index.html

CancerSociety *Canadian cancer society.*   https://www.cancer.ca/Canada-wide/About cancer/Cancer statistics.aspx?sc_lang=EN

Cao, B., Shan, S., Zhang, X. and Gao, W. Baseline evaluations on the CAS-PEAL-R1 face database.   2004. Springer.

Casares, M., Vuran, M. C. and Velipasalar, S. (2008) Design of a wireless vision sensor for object tracking in wireless vision sensor networks. *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on.* IEEE.

Castano, R., Anderson, R. C., Estlin, T., DeCoste, D., Fisher, F., Gaines, D., Mazzoni, D. and Judd, M. (2003) Rover traverse science for increased mission science return.

Castano, R., Wagstaff, K., Song, L. and Anderson, R. (2005) Validating rover image prioritizations. *The Interplanetary Network Progress Report* 42, 160.

Castillo, O. Y. G. (2006) *Survey About Facial Image Quality.* Fraunhofer Institute for Computer Graphics Research. http://www.face-rec.org/interesting-papers/Standards/2dfiqsurveyreport_OYGC.pdf

Çeliktutan, O., Ulukaya, S. and Sankur, B. (2013) A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing* 2013 (1), 13.

Chakraborty, C., Gupta, B. and Ghosh, S. K. (2013) A Review on Telemedicine-Based WBAN Framework for Patient Monitoring. *Telemedicine and e-Health* 19 (8), 619-626.

Chapelle, O., Vapnik, V., Bousquet, O. and Mukherjee, S. (2002) Choosing multiple parameters for support vector machines. *Machine Learning* 46 (1-3), 131-159.

Chang, K. I., Bowyer, K. W. and Flynn, P. J. (2006) Multiple nose region matching for 3D face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (10), 1695-1700.

Chen, D., Meng, M.-H., Wang, H., Hu, C. and Liu, Z. (2011) A novel strategy to label abnormalities for wireless capsule endoscopy frames sequence. *Information and Automation (ICIA), 2011 IEEE International Conference on.* IEEE.

Chen, Y.-T. and Chen, C.-S. (2008) Fast human detection using a novel boosted cascading structure with meta stages. *Image Processing, IEEE Transactions on* 17 (8), 1452-1464.

Chiroro, P. and Valentine, T. (1995) An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology* 48 (4), 879-894.

Claesen, M., De Smet, F., Suykens, J. and De Moor, B. (2013) EnsembleSVM: A Library for Ensemble Learning Using Support Vector Machines. *Journal of Machine Learning Research* (accepted).

Clutterbuck, R. and Johnston, R. A. (2002) Exploring levels of face familiarity by using an indirect face-matching measure. *Perception* 31 (8), 985-994.

Clutterbuck, R. and Johnston, R. A. (2004) Matching as an index of face familiarity. *Visual Cognition* 11 (7), 857-869.

Clutterbuck, R. and Johnston, R. A. (2005) Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology* 17 (1), 97-116.

Colborn, C. (2017) *Understanding Bradford District.* City of Bradford Metropolitan District Council: https://ubd.bradford.gov.uk/media/1348/ethnicity-in-bradford.pdf Accessed 22nd January 2019.

Collins, A.-P. (July 2014) *2011 Census analysis: What does the 2011 Census tell us about Inter-ethnic Relationships?* Office of National Statistics:

Coleman, S. R. and Grover, R. (2006) The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesthetic surgery journal* 26 (1_Supplement), S4-S9.

Coren, S., Porac, C. and Theodor, L. H. (1986) The effects of perceptual set on the shape and apparent depth of subjective contours. *Attention, Perception, & Psychophysics* 39 (5), 327-333.

Cosmides, L., Tooby, J. and Kurzban, R. (2003) Perceptions of race. *Trends in cognitive sciences* 7 (4), 173-179.

Council, C. o. B. M. D. (2019) *Population.* https://www.bradford.gov.uk/open-data/our-datasets/population/ Accessed 18th January 2019.

Crow, F. C. (1984) Summed-area tables for texture mapping. *ACM SIGGRAPH Computer Graphics.* Vol. 18. ACM.

Darwin, C. (1859) On the origin of species by means of natural selection.

Darwin, C. (1968) *On the Origin of Species: Or the Preservation of Favoured Races in the Struggle for Life.* Penguin.

Davis, J. W., Morison, A. M. and Woods, D. D. (2007) An adaptive focus-of-attention model for video surveillance and monitoring. *Machine Vision and Applications* 18 (1), 41-64.

Delvaux, M. and Gay, G. (2008) Capsule endoscopy: technique and indications. *Best Practice & Research Clinical Gastroenterology* 22 (5), 813-837.

Dobbins, A., Zucker, S. W. and Cynader, M. S. (1987) Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature* 329 (6138), 438-441.

Dobra, A., Garofalakis, M., Gehrke, J. and Rastogi, R. (2002) Processing complex aggregate queries over data streams. *Proceedings of the 2002 ACM SIGMOD international conference on Management of data.* ACM.

Dong, H. and Gu, N. (2001) *Asian face image database PF01*.

Donnelly, K. P. (1983) The probability that related individuals share some section of genome identical by descent. *Theoretical population biology* 23 (1), 34-63.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B. and Leitão, P. J. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27-46.

Drira, H., Amor, B. B., Daoudi, M. and Srivastava, A. Nasal region contribution in 3D face biometrics using shape analysis framework. 2009. Springer.

Dua, M., Gupta, R., Khari, M. and Crespo, R. G. (2019) Biometric iris recognition using radial basis function neural network. *Soft Computing*, 1-15.

Duan, L., Wu, C., Miao, J. and Bovik, A. C. (2011) Visual Conspicuity Index: Spatial Dissimilarity, Distance, and Central Bias. *Signal Processing Letters, IEEE* 18 (11), 690-693.

Duan, X.-d., Wang, C.-r., Liu, X.-d., Li, Z.-j., Wu, J. and Zhang, H.-l. Ethnic features extraction and recognition of human faces. 2010. Vol. 2. IEEE.

Durtschi, R. B., Chung, D., Gentry, L. R., Chung, M. K. and Vorperian, H. K. (2009) Developmental craniofacial anthropometry: Assessment of race effects. *Clinical Anatomy* 22 (7), 800-808.

Ebner, N. C., Riediger, M. and Lindenberger, U. (2010) FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods* 42 (1), 351-362.

Ehrenstein, W. (2003) Basics of seeing motion. *Arquivos brasileiros de oftalmologia* 66 (5), 44-52.

Ejaz, N., Mehmood, I. and Baik, S. W. (2013a) MRT letter: Visual attention driven framework for hysteroscopy video abstraction. *Microscopy research and technique* 76 (6), 559-563.

Ejaz, N., Mehmood, I. and Baik, S. W. (2013b) MRT letter: Visual attention driven framework for hysteroscopy video abstraction. *Microscopy research and technique*.

Ejaz, N., Mehmood, I. and Wook Baik, S. (2012) Efficient visual attention based framework for extracting key frames from videos. *Signal Processing: Image Communication*.

Elmahmudi, A. and Ugail, H. (2020) A framework for facial age progression and regression using exemplar face templates. *The Visual Computer*, 1-16.

El Khiyari, H. and Wechsler, H. (2016) Face recognition across time lapse using convolutional neural networks. *Journal of Information Security* 7 (03), 141.

Elsamny, T. A., Rabie, A. N., Abdelhamid, A. N. and Sobhi, E. A. (2018) Anthropometric Analysis of the External Nose of the Egyptian Males. *Aesthetic plastic surgery* 42 (5), 1343-1356.

Ellis, H. D., Shepherd, J. W. and Davies, G. M. (1979) Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception* 8 (4), 431-439.

Engel, S., Zhang, X. and Wandell, B. (1997) Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature* 388 (6637), 68-71.

Eriksson, A., Olsson, C. and Kahl, F. (2011) Normalized cuts revisited: a reformulation for segmentation with linear grouping constraints. *Journal of Mathematical Imaging and Vision* 39 (1), 45-61.

FaceFirst (2019) *FaceFirst Software.* https://www.facefirst.com/ Accessed 29th January 2020.

Fang, F., Clapham, P. J. and Chung, K. C. (2011) A systematic review of inter-ethnic variability in facial dimensions. *Plastic and reconstructive surgery* 127 (2), 874.

Farkas, L. G., Katic, M. J. and Forrest, C. R. (2005) International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery* 16 (4), 615-646.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D. and Csibra, G. (2005) Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences* 102 (47), 17245-17250.

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern recognition letters* 27 (8), 861-874.

Feingold, G. A. (1914) Influence of environment on identification of persons and things. *J. Am. Inst. Crim. L. & Criminology* 5, 39.

Fischl, B., Sereno, M. I., Tootell, R. B. and Dale, A. M. (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping* 8 (4), 272-284.

FISWG (2019) *Physical Stability of Facial Features of Adults*. www.fiswg.org: https://fiswg.org/documents.html Accessed 02nd October 2020

Fletcher, A. (2011) Nerve cell function and synaptic mechanisms. *Anaesthesia & Intensive Care Medicine* 12 (6), 253-257.

Flinn, J. and Satyanarayanan, M. (1999) Powerscope: A tool for profiling the energy usage of mobile applications. *Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA'99. Second IEEE Workshop on.* IEEE.

Fodor, I. K. (2002) *A survey of dimension reduction techniques.* Lawrence Livermore National Lab., CA (US).

Fu, S., He, H. and Hou, Z.-G. (2014) Learning race from face: A survey. *IEEE transactions on pattern analysis and machine intelligence* 36 (12), 2483-2509.

Fukushima, K., Miyake, S. and Ito, T. (1983) Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE transactions on systems, man, and cybernetics* (5), 826-834.

Gangwar, A. and Joshi, A. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition.   2016. IEEE.

Greenwell, C., Spurlock, S., Souvenir, R. and Jacobs, N. GeoFaceExplorer: Exploring the geo-dependence of facial attributes.   2014. ACM.

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X. and Zhao, D. (2008) The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38 (1), 149-161.

Gao, Y., Zheng, Y., Diao, S., Toh, W.-D., Ang, C.-W., Je, M. and Heng, C.-H. (2011) Low-power ultrawideband wireless telemetry transceiver for medical sensor applications. *Biomedical Engineering, IEEE Transactions on* 58 (3), 768-772.

GastroLab http://www.gastrolab.net/.

Gianluigi, C. and Raimondo, S. (2006) An innovative algorithm for key frame extraction in video summarization. *Journal of Real-Time Image Processing* 1 (1), 69-88.

Glukhovsky, A. (2003) Wireless capsule endoscopy. *Sensor Review* 23 (2), 128-133.

Goldstein, A. G. and Chance, J. E. (1985) Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society* 23 (3), 211-214.

Goncalves, W. N. and Bruno, O. M. (2013) Combining fractal and deterministic walkers for texture analysis and classification. *Pattern Recognition* 46 (11), 2953-2968.

Google App Engine.   https://developers.google.com/appengine/

Goren, C. C., Sarty, M. and Wu, P. Y. K. (1975) Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56 (4), 544-549.

GOV.UK.   *List   of   Ethnic   Groups.*   https://www.ethnicity-facts Figures.service.gov.uk/ethnic-groups. Accessed 29th January 2020.

Criminal Justice Act 1991 2019 https://www.legislation.gov.uk/ukpga/1991/53/section/95. Accessed 16th March 2019.

Goyette, N., Jodoin, P.-M., Porikli, F., Konrad, J. and Ishwar, P. (2012) Changedetection. net: A new change detection benchmark dataset. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.* IEEE.

Grgic, M., Delac, K. and Grgic, S. (2011) SCface–surveillance cameras face database. *Multimedia tools and applications* 51 (3), 863-879.

Gross, R., Matthews, I., Cohn, J., Kanade, T. and Baker, S. (2010) Multi-pie. *Image and Vision Computing* 28 (5), 807-813.

Gross, T. F. (2009) Own-ethnicity bias in the recognition of Black, East Asian, Hispanic, and White faces. *Basic and Applied Social Psychology* 31 (2), 128-135.

Grother, P. J., Quinn, G. W. and Phillips, P. J. (2010) Report on the evaluation of 2D still-image face recognition algorithms. *NIST interagency report* 7709, 106.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. and Lew, M. S. (2016) Deep Learningfor visual understanding: A review. *Neurocomputing* 187, 27-48.

Guo, G. and Mu, G. Joint estimation of age, gender and ethnicity: CCA vs. PLS. 2013. IEEE.

Gurney, K. (2014) *An introduction to neural networks.* CRC press.

Haller, J. S. (1995) *Outcasts from evolution: Scientific attitudes of racial inferiority, 1859-1900.* SIU Press.

Halpern, M. and Jacob, H. (2002) *Atlas of capsule endoscopy.* Given Imaging, Incorporated.

Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. and Shraiman, B. I. (1986) Fractal measures and their singularities: The characterization of strange sets. *Physical Review A* 33 (2), 1141.

Harel, J., Koch, C. and Perona, P. (2007) Graph-based visual saliency.

Harnett, C. (2010) Open wireless sensor network telemetry platform for mobile phones. *Sensors Journal, IEEE* 10 (6), 1083-1084.

Han, H., Otto, C., Liu, X. and Jain, A. K. (2015) Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (6), 1148-1161.

Harkema, J. R., Carey, S. A. and Wagner, J. G. (2006) The nose revisited: a brief review of the comparative structure, function, and toxicologic pathology of the nasal epithelium. *Toxicologic pathology* 34 (3), 252-269.

Hayman, S. The mcculloch-pitts model. 1999. Vol. 6. IEEE.

Hayward, W. G., Rhodes, G. and Schwaninger, A. (2008) An own-race advantage for components as well as configurations in face recognition. *Cognition* 106 (2), 1017-1027.

He, K., Zhang, X., Ren, S. and Sun, J. Deep residual learning for image recognition. 2016.

Ho, M. R. and Pezdek, K. (2016) Postencoding cognitive processes in the cross-race effect: Categorization and individuation during face recognition. *Psychonomic bulletin & review* 23 (3), 771-780.

Hoffman, D. D. and Singh, M. (1997) Salience of visual parts. *Cognition* 63 (1), 29-78.

Hong, Y.-W., Huang, W.-J., Chiu, F.-H. and Kuo, C.-C. (2007) Cooperative communications in resource-constrained wireless networks. *Signal Processing Magazine, IEEE* 24 (3), 47-57.

Hosoi, S., Takikawa, E. and Kawade, M. Ethnicity estimation with facial images. 2004. IEEE.

Hu, M.-K. (1962) Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on* 8 (2), 179-187.

Hu, Z., Cai, Y., Li, Y. and Xu, X. (2005) Support vector machine based ensemble classifier. *American Control Conference, 2005. Proceedings of the 2005.* IEEE.

Huang, L. and Pashler, H. (2007) A Boolean map theory of visual attention. *Psychological review* 114 (3), 599.

Hubel, D. H. and Wiesel, T. N. (1968) Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* 195 (1), 215-243.

Hudson, N. (1996) From" Nation to" Race": The Origin of Racial Classification in Eighteenth-Century Thought. *Eighteenth-century studies* 29 (3), 247-264.

Hugenberg, K., Miller, J. and Claypool, H. M. (2007) Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology* 43 (2), 334-340.

Hussain, O., Hussain, F., Parvin, S., Han, S., Tian, B. and Chang, E. (2012) Cognitive radio network security: A survey.

Hwang, B.-W., Roh, M.-C. and Lee, S.-W. Performance evaluation of face recognition algorithms on asian face database. 2004. IEEE.

Iakovidis, D., Tsevas, S. and Polydorou, A. (2010) Reduction of capsule endoscopy reading times by unsupervised image mining. *Computerised Medical Imaging and Graphics* 34 (6), 471-478.

IDS (c. 2018a) *Imaging Development Systems.* https://en.ids-imaging.com/home.html

IDS (c. 2018b) *Lens, Tamron.* https://en.ids-imaging.com/store/products/accessory/lenses/lens-tamron-m23fm08-8-mm-2-3.html

India, G. o. (2010) *Face Image Data Standard for e-Goverance Applications in India.* Ministry of Communications and Information Technology: http://egovstandards.gov.in/sites/default/files/Face_Image_Data_Standard_Ver1.0.pdf

Irfan Mehmood, M. S., S. Rho, and Sung Wook Baik (January 26-28, 2015) Audio-visual and EEG-based Attention Modeling for Extraction of Affective Video Content. In *(PlatCon 2015.* Jeju island, Korea.

Irgan, K., Ünsalan, C. and Baydere, S. (2013) Low-cost prioritization of image blocks in wireless sensor networks for border surveillance. *Journal of Network and Computer Applications*.

Islam, R. (2007) *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark & Tattoo Images.* National Policing Improvement Agency. https://www.whatdotheyknow.com/request/383407/response/939050/attac Islam, M. T., Workman, S., Wu, H., Jacobs, N. and Souvenir, R. Exploring the geo-dependence of human face appearance. 2014. IEEE.

Islam, M. T., Workman, S., Wu, H., Jacobs, N. and Souvenir, R. Exploring the geo-dependence of human face appearance. 2014. IEEE.

h/6/017 16 NPCC Response attachment 03 of 05.pdf

Islam, M. T., Workman, S., Wu, H., Jacobs, N. and Souvenir, R. Exploring the geo-dependence of human face appearance. 2014. IEEE.

Itti, L., Koch, C. and Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (11), 1254-1259.

Jaffar, M. A., Hayder, Z., Hussain, A. and Mirza, A. M. (2009) An Intelligent Ensemble Based Systems for Breast Cancer Diagnosis. *Proceedings of 2009 International Conference on Computer Engineering and Applications, Philippine.*

Jahanshahi, M., Golalipour, M. J. and Heidari, K. (2008) The effect of ethnicity on facial anthropometry in Northern Iran. *Singapore medical journal* 49 (11), 940.

Jain, A. K. and Park, U. Facial marks: Soft biometric for face recognition. 2009. IEEE.

Jilani, S. K., Ugail, H., Bukar, A. M., Logan, A. and Munshi, T. A Machine LearningApproach for Ethnic Classification: The British Pakistani Face. 2017. IEEE.

Jilani, S. K., Ugail, H. and Logan, A. (2019) Inter-Ethnic and Demic-Group Variations in Craniofacial Anthropometry: A Review. *PSM Biological Research* 4 (1), 6-16.

Johnson, M. H., Dziurawiec, S., Ellis, H. and Morton, J. (1991) Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40 (1), 1-19.

Jovanov, E. and Milenkovic, A. (2011) Body area networks for ubiquitous healthcare applications: opportunities and challenges. *Journal of medical systems* 35 (5), 1245-1254.

Judd, T., Durand, F. and Torralba, A. (2012) A Benchmark of Computational Models of Saliency to Predict Human Fixations.

Kanwisher, N., McDermott, J. and Chun, M. M. (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience* 17 (11), 4302-4311.

Kanwisher, N., Stanley, D. and Harris, A. (1999) The fusiform face area is selective for faces not animals. *Neuroreport* 10 (1), 183-187.

Kanwisher, N. and Yovel, G. (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1476), 2109-2128.

Karargyris, A. and Bourbakis, N. (2010) Wireless capsule endoscopy and endoscopic imaging: A survey on various methodologies presented. *Engineering in Medicine and Biology Magazine, IEEE* 29 (1), 72-83.

Karimaa, A. (2011) Efficient video surveillance: performance evaluation in distributed video surveillance systems. *Video Surveillance. InTech*.

Kasinski, A., Florek, A. and Schmidt, A. (2008) The PUT face database. *Image Processing and Communications* 13 (3-4), 59-64.

Kawaguchi, K. (2017) *The McCulloch-Pitts Model of Neuron.* http://wwwold.ece.utep.edu/research/webfuzzy/docs/kk-thesis/kk-thesis-html/node12.html Accessed 01st February.

Keith A. Johnson, M. D., J. Alex Becker. *The Whole Brain Atlas.* http://med.harvard.edu/ *AANLIB/*

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L. and Pascalis, O. (2005) Three-month-olds, but not newborns, prefer own-race faces. *Developmental science* 8 (6).

Kemp, R. I., Caon, A., Howard, M. and Brooks, K. R. (2016) Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology* 30 (4), 622-627.

Keogh, E., Chakrabarti, K., Pazzani, M. and Mehrotra, S. (2001) Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems* 3 (3), 263-286.

Khan, T. H. and Wahid, K. A. (2012) An advanced physiological data logger for medical imaging applications. *EURASIP Journal on Embedded Systems* 2012 (1), 10.

Kim, H., Kim, W., Lee, M., Song, E. and Loh, J. J. (2007) Tumor volume and uterine body invasion assessed by MRI for prediction of outcome in cervical carcinoma treated with concurrent chemotherapy and radiotherapy. *Japanese journal of clinical oncology* 37 (11), 858-866.

Kim, H.-C., Pang, S., Je, H.-M., Kim, D. and Yang Bang, S. (2003) Constructing support vector machine ensemble. *Pattern recognition* 36 (12), 2757-2767.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V. and Jain, A. K. (2012) Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security* 7 (6), 1789-1801.

Kotsiantis, S. B., Zaharakis, I. and Pintelas, P. (2007) Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160, 3-24.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. Imagenet classification with deep convolutional neural networks. 2012.

Krupinski, E. A. (2000) The importance of perception research in medical imaging. *Radiation Medicine* 18 (6), 329-334.

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., Graham, A. R., Descour, M. R., Davis, J. R. and Weinstein, R. S. (2006) Eye-movement study and human performance using telepathology virtual slides: implications for medical education and differences with experience. *Human pathology* 37 (12), 1543.

Kumar, K. and Lu, Y.-H. (2010) Cloud computing for mobile users: Can offloading computation save energy? *Computer* 43 (4), 51-56.

Kundel, H. and Nodine, C. (1983) A visual concept shapes image perception. *Radiology* 146 (2), 363-368.

Kundel, H. L. (2006) History of research in medical image perception. *Journal of the American College of Radiology* 3 (6), 402-408.

Lakshmiprabha, N. S. (2016) Face Image Analysis using AAM, Gabor, LBP and WD features for Gender, Age, Expression and Ethnicity Classification. *arXiv preprint arXiv:1604.01684*.

Lamb, H. (2017) *Facial recognition technology 'fails' at Notting Hill Carnival.* https://eandt.theiet.org/content/articles/2017/08/facial-recognition-technology-fails-at-notting-hill-carnival/

Latré, B., Braem, B., Moerman, I., Blondia, C. and Demeester, P. (2011) A survey on wireless body area networks. *Wireless Networks* 17 (1), 1-18.

Lawson, J. A., Vogrin, S., Bleasel, A. F., Cook, M. J. and Bye, A. M. (2005) Cerebral and cerebellar volume reduction in children with intractable epilepsy. *Epilepsia* 41 (11), 1456-1462.

Lazarevic-McManus, N., Renno, J. and Jones, G. (2006) Performance evaluation in visual surveillance using the F-measure. *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks.* ACM.

Lazarevic-McManus, N., Renno, J., Makris, D. and Jones, G. A. (2008) An object-based comparative methodology for motion detection based on the F-Measure. *Computer Vision and Image Understanding* 111 (1), 74-85.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *nature* 521 (7553), 436.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E. and Jackel, L. D. Handwritten digit recognition with a back-propagation network. 1990.

Lee, H. and Byun, H. (2010) Temporal gradient pattern for the near-duplicate video clustering. *Machine Learningand Cybernetics (ICMLC), 2010 International Conference on.* Vol. 6. IEEE.

Lee, H.-G., Choi, M.-K., Shin, B.-S. and Lee, S.-C. (2013) Reducing redundancy in wireless capsule endoscopy videos. *Computers in biology and medicine*.

Lemieux, L., Hagemann, G., Krakow, K. and Woermann, F. G. (1999) Fast, accurate, and reproducible automatic segmentation of the brain in T1-

weighted volume MRI data. *Magnetic Resonance in Medicine* 42 (1), 127-135.

Levina, E. and Bickel, P. (2001) The earth mover's distance is the Mallows distance: some insights from statistics. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on.* Vol. 2. IEEE.

Lever, J., Krzywinski, M. and Altman, N. (2016) *Points of significance: classification evaluation*. Nature Publishing Group.

Li, B. and Meng, M.-H. (2012) Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. *Information Technology in Biomedicine, IEEE Transactions on* 16 (3), 323-329.

Lim, M., Chan, C., Monekosso, D. and Remagnino, P. (2014) Detection of salient regions in crowded scenes. *Electronics Letters* 50 (5), 363-365.

Lison, P. (2015) *"An introduction to machine learning.*

Logan, A. J., Gordon, G. E. and Loffler, G. (2017) Contributions of individual face features to face discrimination. *Vision Research* 137, 29-39.

Lucas, B. D. and Kanade, T. (1981) An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th international joint conference on Artificial intelligence.*

Lu, X. and Jain, A. K. Ethnicity identification from face images. 2004. International Society for Optics and Photonics.

Lv, C., Wu, Z., Wang, X., Dan, Z. and Zhou, M. (2020) Ethnicity classification by the 3D Discrete Landmarks Model measure in Kendall shape space. *Pattern Recognition Letters* 129, 26-32.

Lv, C., Wu, Z., Zhang, D., Wang, X. and Zhou, M. (2018) 3D Nose shape net for human gender and ethnicity classification. *Pattern Recognition Letters*.

Lyons, M. J., Kamachi, M. and Gyoba, J. (2014) *Japanese Female Facial Expressions (JAFFE), Database of digital images. 2007*.

MacLin, O. H., MacLin, M. K. and Malpass, R. S. (2001) Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law* 7 (1), 134.

Makhoul, J., Kubala, F., Schwartz, R. and Weischedel, R. (1999) Performance measures for information extraction. *Proceedings of DARPA Broadcast News Workshop.*

Masood, S., Gupta, S., Wajid, A., Gupta, S. and Ahmed, M. (2018) Prediction of human ethnicity from facial images using neural networks. *Data Engineering and Intelligent Computing.* Springer. 217-226.

Mathieson, I. and Scally, A. (2020) What is ancestry? *PLOS Genetics* 16 (3), e1008624.

Malpass, R. S. and Kravitz, J. (1969) Recognition for faces of own and other race. *Journal of personality and social psychology* 13 (4), 330.

Manning, D., Gale, A. and Krupinski, E. (2005) Perception research in medical imaging. *British journal of radiology* 78 (932), 683-685.

Manzoor, U. and Nefti, S. (2012) iDetect: Content Based Monitoring of Complex Networks using Mobile Agents. *Applied Soft Computing* 12 (5), 1607-1619.

Manzoor, U., Nefti, S. and Rezgui, Y. (2012) Categorization of Malicious Behaviors using Ontology-based Cognitive Agents. *Data & Knowledge Engineering*.

Margolin, R., Zelnik-Manor, L. and Tal, A. (2012) Saliency for image manipulation. *The Visual Computer*, 1-12.

Massanes, F. and Brankov, J. G. (2011) Motion perception in medical imaging. *SPIE Medical Imaging.* International Society for Optics and Photonics.

Maurer, D., Le Grand, R. and Mondloch, C. J. (2002) The many faces of configural processing. *Trends in cognitive sciences* 6 (6), 255-260.

McCulloch, W. S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5 (4), 115-133.

Megreya, A. M. and Bindemann, M. (2009) Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception* 38 (12), 1831-1848.

Megreya, A. M., White, D. and Burton, A. M. (2011) The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology* 64 (8), 1473-1483.

Mehmood, I., Baik, R. and Baik, S. W. (2013a) Automatic Segmentation of Region of Interests in MR Images Using Saliency Information and Active Contours. *IT Convergence and Security 2012.* Springer. 537-544.

Mehmood, I., Ejaz, N., Sajjad, M. and Baik, S. W. (2013b) Prioritization of brain MRI volumes using medical image perception model and tumor region segmentation. *Computers in biology and medicine* 43 (10), 1471-1483.

Mehmood, I., Sajjad, M. and Baik, S. W. (2014a) Mobile-Cloud Assisted Video Summarization Framework for Efficient Management of Remote Sensing Data Generated by Wireless Capsule Sensors. *Sensors* 14 (9), 17112-17145.

Mehmood, I., Sajjad, M. and Baik, S. W. (2014b) Video summarization based tele-endoscopy: A service to efficiently manage visual data generated during wireless capsule endoscopy procedure. *Journal of medical systems* 38 (9), 1-9.

Mehmood, I., Sajjad, M., Ejaz, W. and Baik, S. W. (2014c) Saliency-directed prioritization of visual data in wireless surveillance networks. *Information Fusion.*

Mehmood, I., Ullah, A., Muhammad, K., Deng, D.-J., Meng, W., Al-Turjman, F., Sajjad, M. and De Albuquerque, V. H. C. (2019) Efficient image recognition and retrieval on IoT-assisted energy-constrained platforms from big data repositories. *IEEE Internet of Things Journal* 6 (6), 9246-9255.

Meissner, C. A. and Brigham, J. C. (2001) *Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review.* American Psychological Association.

Mersha, T. B. and Abebe, T. (2015) Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. *Human Genomics* 9 (1), 1.

Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (2013) *Machine learning: An artificial intelligence approach.* Springer Science & Business Media.

Michel, C., Corneille, O. and Rossion, B. (2006) Categorization of face race modulates holistic face processing. *Journal of Vision* 6 (6), 435-435.

Microchip (2014) http://www.microchip.com/.

Milgrim, L. M., Lawson, W. and Cohen, A. F. (1996) Anthropometric analysis of the female Latino nose: revised aesthetic concepts and their surgical implications. *Archives of Otolaryngology–Head & Neck Surgery* 122 (10), 1079-1086.

Money, A. G. and Agius, H. (2010) Elvis: entertainment-led video summaries. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 6 (3), 17.

Morton, J. and Johnson, M. H. (1991) CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological review* 98 (2), 164.

Mohammad, A. S. and Al-Ani, J. A. Convolutional Neural Network for Ethnicity Classification using Ocular Region in Mobile Environment. 2018. IEEE.

Muhammad, G., Hussain, M., Alenezy, F., Bebis, G., Mirza, A. M. and Aboalsamh, H. (2012) Race classification from face images using local descriptors. *International Journal on Artificial Intelligence Tools* 21 (05), 1250019.

Murphy, T., Matlin, M. and Finkel, L. H. (2003) Curvature covariation as a factor in perceptual salience. *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on.* IEEE.

Nesbitt, T. S., Cole, S. L., Pellegrino, L. and Keast, P. (2006) Rural outreach in home telehealth: assessing challenges and reviewing successes. *Telemedicine Journal & E-Health* 12 (2), 107-113.

Ng, H.-W. and Winkler, S. A data-driven approach to cleaning large face datasets. 2014. IEEE.

Nicholls, J. G., Martin, A. R., Wallace, B. G. and Fuchs, P. A. (2001) *From neuron to brain.* Vol. 271. Sinauer Associates Sunderland, MA.

Niedermeyer, E. and da Silva, F. L. (2005) *Electroencephalography: basic principles, clinical applications, and related fields.* Lippincott Williams & Wilkins.

O'Toole, A. J., Harms, J., Snow, S. L., Hurst, D. R., Pappas, M. R., Ayyad, J. H. and Abdi, H. (2005) A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (5), 812-816.

O'Toole, A. J. and Natu, V. (2013) Computational perspectives on the other-race effect. *Visual Cognition* 21 (9-10), 1121-1137.

O'Toole, A. J., Phillips, P. J. and Narvekar, A. Humans versus algorithms: Comparisons from the face recognition vendor test 2006. 2008. IEEE.

Ofodile, F. A. and Bokhari, F. (1995) The African-American Nose: Part I I. *Annals of plastic surgery* 34 (2), 123-129.

Ofodile, F. A., Bokhari, F. J. and Ellis, C. (1993) The black American nose. *Annals of plastic surgery* 31 (3), 209-219.

Ohki, M., Naito, K. and Cole, P. (1991) Dimensions and resistances of the human nose: racial differences. *The Laryngoscope* 101 (3), 276-278.

Ojala, T., Pietikäinen, M. and Harwood, D. (1996) A comparative study of texture measures with classification based on featured distributions. *Pattern recognition* 29 (1), 51-59.

Okun, O. and Priisalu, H. (2007) Unsupervised data reduction. *Signal processing* 87 (9), 2260-2267.

Ou, Y., Wu, X., Qian, H. and Xu, Y. A real time race classification system. 2005. IEEE.

Osborne, C. D. and Stevenage, S. V. (2008) Internal feature saliency as a marker of familiarity and configural processing. *Visual Cognition* 16 (1), 23-43.

Ozdemir, F. and Uzun, A. (2015) Anthropometric analysis of the nose in young Turkish men and women. *Journal of Cranio-Maxillofacial Surgery* 43 (7), 1244-1247.

Pal, S. K., Petrosino, A. and Maddalena, L. (2012) *Handbook on Soft Computing for Video Surveillance.* CRC Press.

Palermo, R. and Rhodes, G. (2007) Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia* 45 (1), 75-92.

Pan, G., Yan, G., Qiu, X. and Cui, J. (2011) Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. *Journal of medical systems* 35 (6), 1477-1484.

Parkhi, O. M., Vedaldi, A. and Zisserman, A. Deep Face Recognition. 2015. Vol. 1.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. and Fabre-Grenet, M. (1995) Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development* 18 (1), 79-85.

Pashler, H. E. (1999) *The psychology of attention.* The MIT Press.

Paul Griffin, P. D. (2005) *Understanding The Face Image Format Standards.* NIST: https://www.nist.gov/sites/default/files/documents/2016/12/12/griffin-face-std-m1.pdf

Peter Wade, A. S. (2017) *Race.* Encyclopaedia Britannica inc. https://www.britannica.com/topic/race-human Accessed 18th January 2019.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J. and Worek, W. Overview of the face recognition grand challenge. 2005. Vol. 1. IEEE.

Phillips, P. J., Grother, P., Micheals, R., Blackburn, D. M., Tabassi, E. and Bone, M. Face recognition vendor test 2002. 2003. IEEE.

Phillips, P. J., Jiang, F., Ayyad, J. and Pénard, N. (2007) Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (9), 1642-1646.

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J. and O'Toole, A. J. (2011) An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception (TAP)* 8 (2), 14.

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.

Phillips, P. J., Wechsler, H., Huang, J. and Rauss, P. J. (1998) The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing* 16 (5), 295-306.

Pickering, C. and Hall, J. C. (1854) *The races of man: and their geographical distribution.* George Bell.

Pizer, S. M. and ter Haar Romeny, B. M. (1991) Fundamental properties of medical image perception. *Journal of Digital Imaging* 4 (1), 1-20.

Porter, K. R., McCarthy, B. J., Freels, S., Kim, Y. and Davis, F. G. (2010) Prevalence estimates for primary brain tumors in the United States by age, gender, behavior, and histology. *Neuro-oncology* 12 (6), 520-527.

Precup, O. V. and Iori, G. (2007) Cross-correlation measures in the high-frequency domain. *European Journal of Finance* 13 (4), 319-331.

Puzicha, J., Buhmann, J. M., Rubner, Y. and Tomasi, C. (1999) Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.* Vol. 2. IEEE.

Puzicha, J., Hofmann, T. and Buhmann, J. M. (1997) Non-parametric similarity measures for unsupervised texture segmentation and image retrieval.

Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on. IEEE.

Qiu, X., Sun, Z. and Tan, T. Learning appearance primitives of iris images for ethnic classification. 2007. Vol. 2. IEEE.

Rakotomamonjy, A. and Guigue, V. (2008) BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. *Biomedical Engineering, IEEE Transactions on* 55 (3), 1147-1154.

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T. and Reissland, N. (2017) The human fetus preferentially engages with face-like visual stimuli. *Current Biology*.

Reljin, I. S. and Reljin, B. D. (2002) Fractal geometry and multifractals in analyzing and processing medical data and images. *Archive of Oncology* 10 (4), 283-293.

Riaz, F., Silva, F. B., Ribeiro, M. D. and Coimbra, M. T. (2012) Invariant gabor texture descriptors for classification of gastroenterology images. *Biomedical Engineering, IEEE Transactions on* 59 (10), 2893-2904.

Ricanek, K. and Tesafaye, T. Morph: A longitudinal image database of normal adult age-progression. 2006. IEEE.

Riccio, D., Tortora, G., De Marsico, M. and Wechsler, H. EGA—Ethnicity, gender and age, a pre-annotated face database. 2012. IEEE.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. and Feldman, M. W. (2002) Genetic structure of human populations. *science* 298 (5602), 2381-2385.

Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review* 65 (6), 386.

Rubner, Y., Puzicha, J., Tomasi, C. and Buhmann, J. M. (2001) Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding* 84 (1), 25-43.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A. and Bernstein, M. (2015) Imagenet large scale visual recognition challenge. *International journal of computer vision* 115 (3), 211-252.

Saha, K., Debnath, R., Bhowmik, M. K., Bhattacharjee, D. and Nasipuri, M. (2012) North-east Indian face database: its design and aspects.

Sahoo, B., Rath, S. and Puthal, D. (2012) Energy Efficient Protocols for Wireless Sensor Networks: A Survey and Approach. *International Journal of Computer Applications (0975–8887)* 44 (18).

Sainju, S., Bui, F. M. and Wahid, K. A. (2014) Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. *Journal of medical systems* 38 (4), 1-11.

Sajjad, M., Mehmood, I. and Baik, S. W. (2014) Sparse Representations-Based Super-Resolution of Key-Frames Extracted from Frames-Sequences Generated by a Visual Sensor Network. *Sensors* 14 (2), 3652-3674.

Sajid, M., Shafique, T., Riaz, I., Imran, M., Jabbar Aziz Baig, M., Baig, S. and Manzoor, S. (2018) Facial asymmetry-based anthropometric differences between gender and ethnicity. *Symmetry* 10 (7), 232.

Saliha, M., Ali, B. and Rachid, S. (2019) Towards large-scale face-based race classification on spark framework. *Multimedia Tools and Applications* 78 (18), 26729-26746.

Sankar, P. and Cho, M. K. (2002) *Toward a new vocabulary of human genetic variation.* American Association for the Advancement of Science.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G. and de Schonen, S. (2005) Reversibility of the other-race effect in face recognition during childhood. *Psychological Science* 16 (6), 440-444.

Scharcanski, J. and Gavidao, W. (2006) Hierarchical summarization of diagnostic hysteroscopy videos. *Image Processing, 2006 IEEE International Conference on.* IEEE.

Sciences, P. I. o. M. (2010) http://www.pims.gov.pk/radiology.htm. Pakistan: http://www.pims.gov.pk/radiology.htm Accessed 2010.

Senthil face database. (2005) 1 edition.

Setty, S., Husain, M., Beham, P., Gudavalli, J., Kandasamy, M., Vaddi, R., Hemadri, V., Karure, J. C., Raju, R. and Rajan, B. Indian movie face database: a benchmark for face recognition under wide variations. 2013. IEEE.

Šerić, L., Stipaničev, D. and Štula, M. (2011) Observer network and forest fire detection. *Information Fusion* 12 (3), 160-175.

Shan, Z. Y., Yue, G. H. and Liu, J. Z. (2002) Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. *NeuroImage* 17 (3), 1587.

Shapley, R. and Hawken, M. J. (2011) Color in the cortex: single-and double-opponent cells. *Vision research* 51 (7), 701-717.

Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A. and Leahy, R. M. (2001) Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13 (5), 856-876.

Shi, J. and Malik, J. (2000) Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 (8), 888-905.

Shu, P., Liu, F., Jin, H., Chen, M., Wen, F., Qu, Y. and Li, B. (2013) eTime: energy-efficient transmission between cloud and mobile devices. *Proc. of IEEE INFOCOM.*

Sim, T., Baker, S. and Bsat, M. The CMU pose, illumination, and expression (PIE) database. 2002. IEEE.

Simonyan, K. and Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sinha, P. (2000) Last but not least. *Perception* 29 (8), 1005-1008.

Sinha, P. and Poggio, T. (1996) I think I know that face.

Slone, A. E., Brigham, J. C. and Meissner, C. A. (2000) Social and cognitive factors affecting the own-race bias in Whites. *Basic and applied social Psychology* 22 (2), 71-84.

Song, S., Ohnuma, K., Liu, Z., Mei, L., Kawada, A. and Monma, T. (2009) Novel biometrics based on nose pore recognition. *Optical Engineering* 48 (5), 057204.

Spillner, J., Müller, J. and Schill, A. (2013) Creating optimal cloud storage systems. *Future Generation Computer Systems* 29 (4), 1062-1072.

Sporer, S. L. (2001a) Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law* 7 (1), 36.

Sporer, S. L. (2001b) The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law* 7 (1), 170.

Srinivas, N., Atwal, H., Rose, D. C., Mahalingam, G., Ricanek, K. and Bolme, D. S. Age, gender, and fine-grained ethnicity prediction using convolutional neural networks for the East Asian face dataset. 2017. IEEE.

Statistics, O. o. N. (2009) *Final recommended questions for the 2011 Census in England and Wales- Ethnic Group.* gov.uk: https://www.ethnicity-facts-Figures.service.gov.uk/ethnicity-in-the-uk/ethnic-groups-and-data-collected

Statistics, O. o. N. (2012) *Ethnicity and National Identity in England and Wales: 2011.* Vol. 2. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11 Accessed 18th January 2019.

Statistics, O. o. N. (2015) *2011 Census analysis: Ethnicity and religion of the non-UK born population in England and Wales: 2011.* https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/2011censusanalysisisethnicityandreligionofthenonukbornpopulationinenglandandwales/2015-06-18 Accessed 18th January 2019.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C. and Heagins, T. B. (2016) The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior research methods* 48 (3), 1197-1204.

Stringer, C. B. and Andrews, P. (1988) Genetic and fossil evidence for the origin of modern humans. *Science* 239 (4845), 1263-1268.

Stuurman, S. François Bernier and the invention of racial classification. 2000. Vol. 2000. Oxford University Press.

Su, Y., Zhao, Q., Zhao, L. and Gu, D. (2013) Abrupt motion tracking using a visual saliency embedded particle filter. *Pattern Recognition*.

Sze, G., Milano, E., Johnson, C. and Heier, L. (1990) Detection of brain metastases: comparison of contrast-enhanced MR with unenhanced MR and enhanced CT. *American Journal of Neuroradiology* 11 (4), 785-791.

Sztojánov, I., Crisan, D., Mina, C. P., Voinea, V. and Chen, Y. (2009) Image processing in biology based on the fractal analysis. *Image Proceesing. InTech*, 323-344.

Takahashi, T., Murata, T., Narita, K., Hamada, T., Kosaka, H., Omori, M., Takahashi, K., Kimura, H., Yoshida, H. and Wada, Y. (2006) Multifractal analysis of deep white matter microstructural changes on MRI in relation to early-stage atherosclerosis. *NeuroImage* 32 (3), 1158-1166.

Tanaka, J. W. and Farah, M. J. (1993) Parts and wholes in face recognition. *The Quarterly journal of experimental psychology* 46 (2), 225-245.

Tanaka, J. W. and Sengco, J. A. (1997) Features and their configuration in face recognition. *Memory & cognition* 25 (5), 583-592.

Tang, J., Alelyani, S. and Liu, H. (2014) Feature selection for classification: A review. *Data classification: algorithms and applications*, 37.

Tariq, U., Hu, Y. and Huang, T. S. Gender and ethnicity identification from silhouetted face profiles. 2009. IEEE.

Tatler, B. W. (2007) The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision* 7 (14).

Thomaz, C. E. and Giraldi, G. A. (2010) A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing* 28 (6), 902-913.

Thompson, P. (1980) Margaret Thatcher: a new illusion. *Perception*.

Tian, Y.-L. and Hampapur, A. (2005) Robust salient motion detection with complex background for real-time video surveillance. *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on.* Vol. 2. IEEE.

Tin, H. H. K. and Sein, M. M. (2011) Race identification for face images. *ACEEE Int. J. Inform. Tech* 1 (02), 35-37.

Todorov, A., Pakrashi, M. and Oosterhof, N. N. (2009) Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition* 27 (6), 813-833.

Tome, P., Fierrez, J., Vera-Rodriguez, R. and Ramos, D. (2013) Identification using face regions: Application and assessment in forensic scenarios. *Forensic science international* 233 (1), 75-83.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J. and Nelson, C. (2009) The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research* 168 (3), 242-249.

Touati, F. and Tabish, R. (2013) U-healthcare system: State-of-the-art review and challenges. *Journal of medical systems* 37 (3), 1-20.

Tsang, I. W., Kocsor, A. and Kwok, J. T. (2006) Diversified svm ensembles for large data sets. *Machine Learning: ECML 2006.* Springer. 792-800.

Tullis, J. G., Benjamin, A. S. and Liu, X. (2014) Self-pacing study of faces of different races: Metacognitive control over study does not eliminate the cross-race recognition effect. *Memory & cognition* 42 (6), 863-875.

Ugochukwu, E. G., Ijomone, O. M., Ude, R. A. and Nandi, E. M. (2014) Anthropometric analysis of the nose of the Ejagham ethnic group in Cross River State, Nigeria. *Annals of Bioanthropology* 2 (1), 13.

Ul Haq, I., Ullah, A., Muhammad, K., Lee, M. Y. and Baik, S. W. (2019) Personalized movie summarization using deep cnn-assisted facial expression recognition. *Complexity* 2019

Ullah, S., Higgin, H., Siddiqui, M. A. and Kwak, K. S. (2008) A study of implanted and wearable body sensor networks. *Agent and Multi-Agent Systems: Technologies and Applications.* Springer. 464-473.

Umar, R. and Sheikh, A. U. (2012) A comparative study of spectrum awareness techniques for cognitive radio oriented wireless networks. *Physical Communication.*

Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology* 43 (2), 161-204.

Valentini, G., Muselli, M. and Ruffino, F. (2004) Cancer recognition with bagged ensembles of support vector machines. *Neurocomputing* 56, 461-466.

Valenza, E., Simion, F., Cassia, V. M. and Umiltà, C. (1996) Face preference at birth. *Journal of experimental psychology: Human Perception and Performance* 22 (4), 892.

Vapnik, V. (1999) *The nature of statistical learning theory.* springer.

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

Von der Heydt, R., Peterhans, E. and Baumgartner, G. (1982) Illusory contours and cortical neuron responses. *Ann. Inst. Oceanogr* 58, 297.

Walker, P. M. and Hewstone, M. (2006) A perceptual discrimination investigation of the own-race effect and intergroup experience. *Applied Cognitive Psychology* 20 (4), 461-475.

Walker, P. M. and Tanaka, J. W. (2003) An encoding advantage for own-race versus other-race faces. *Perception* 32 (9), 1117-1125.

Wang, L. and Markham, R. (1999) The development of a series of photographs of Chinese facial expressions of emotion. *Journal of Cross-Cultural Psychology* 30 (4), 397-410.

Wang, X.-W., Nie, D. and Lu, B.-L. (2014) Emotional state classification from EEG data using Machine Learningapproach. *Neurocomputing* 129, 94-106.

Wang, Y., Thomas, J., Weissgerber, S. C., Kazemini, S., Ul-Haq, I. and Quadflieg, S. (2015) *The headscarf effect revisited: Further evidence for a culture-based internal face processing advantage*. SAGE Publications Sage UK: London, England.

Wang, Z., Lim, E. G., Tillo, T. and Yu, F. (2012) Review of the wireless capsule transmitting and receiving antennas. *Book title: Wireless Communications and Networks-Recent Advances. Intech Open. ISBN*, 978-953.

Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N. and Pei, Y. (2020) Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*.

WCEVideoAtlas http://www.wceatlas.org/index.php.

WHO (2020) *Advice on the use of masks in the context of COVID-19*. Coronavirus disease (COVID-19) advice for the public: When and how to use masks: World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks Accessed 18th August 2020.

Wilson, H. R., Wilkinson, F. and Asaad, W. (1997) Concentric orientation summation in human form vision. *Vision research* 37 (17), 2325-2330.

Wolf, H. and Deng, D. (2005) How interesting is this? Finding interest hotspots and ranking images using an MPEG-7 visual attention model.

Wolfe, J. M., Yee, A. and Friedman-Hill, S. R. (1992) Curvature is a basic feature for visual search tasks. *PERCEPTION-LONDON-* 21, 465-465.

Wright, D. B., Boyd, C. E. and Tredoux, C. G. (2003) Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology* 17 (3), 365-373.

Wu, Y. and Ji, Q. (2019) Facial landmark detection: A literature survey. *International Journal of Computer Vision* 127 (2), 115-142.

Xu, L., Li, W. and Schuurmans, D. (2009a) Fast normalized cut with linear constraints. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE.

Xu, Y., Ji, H. and Fermüller, C. (2009b) Viewpoint invariant texture description using fractal analysis. *International Journal of Computer Vision* 83 (1), 85-100.

Yamashita, R., Nishio, M., Do, R. K. G. and Togashi, K. (2018) Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 1-19.

Yoon, U.-C., Kim, J.-S., Kim, J.-S., Kim, I.-Y. and Kim, S. I. (2001) Adaptable fuzzy c-means for improved classification as a preprocessing procedure of brain parcellation. *Journal of Digital Imaging* 14, 238-240.

Young, A. W. (1984) Right cerebral hemisphere superiority for recognizing the internal and external features of famous faces. *British Journal of Psychology* 75 (2), 161-169.

Yu, S. X. and Shi, J. (2004) Segmentation given partial grouping constraints. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26 (2), 173-183.

Yuce, M. R., Ng, S. W., Myo, N. L., Khan, J. Y. and Liu, W. (2007) Wireless body sensor network using medical implant band. *Journal of Medical Systems* 31 (6), 467-474.

Zaitcev, A., Cook, G., Liu, W., Paley, M. and Milne, E. (2015) Source localization for brain-computer interfaces. *Brain-Computer Interfaces.* Springer. 125-153.

Zehngut, N., Juefei-Xu, F., Bardia, R., Pal, D. K., Bhagavatula, C. and Savvides, M. Investigating the feasibility of image-based nose biometrics. 2015. IEEE.

Zhang, J. and Tan, T. (2002) Brief review of invariant texture analysis methods. *Pattern recognition* 35 (3), 735-747.

Zhang, T., Yue, D., Gu, Y., Wang, Y. and Yu, G. (2009a) Adaptive correlation analysis in stream time series with sliding windows. *Computers & Mathematics with Applications* 57 (6), 937-948.

Zhang, W., Mallik, R. and Letaief, K. (2009b) Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. *Wireless Communications, IEEE Transactions on* 8 (12), 5761-5766.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A. and Liu, P. A high-resolution spontaneous 3d dynamic facial expression database. 2013. IEEE.

Zhao, S., Yao, H. and Sun, X. (2013) Video classification and recommendation based on affective analysis of viewers. *Neurocomputing* 119, 101-110.

Zhao, W., Chellappa, R., Phillips, P. J. and Rosenfeld, A. (2003) Face recognition: A literature survey. *ACM computing surveys (CSUR)* 35 (4), 399-458.

Zheng, J. and You, H. (2013) A new model-independent method for change detection in multitemporal SAR images based on Radon transform and Jeffrey divergence. *Geoscience and Remote Sensing Letters, IEEE* 10 (1), 91-95.

Zhong, Z., Zheng, L., Kang, G., Li, S. and Yang, Y. (2017) Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*.