Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Simulating Emotional Intelligence on Computer Systems and analysing Emotional States of South Asian Female Politicians

Prabhjot Kaur

August 9, 2021

A dissertation submitted in partial fulfilment
of the requirements for the degree of
MAI (Computer Engineering)

Abstract

Facial emotion Recognition is receiving substantial attention over the last three decades due to it's application in commercial and non commercial areas. Facial Action Coding system developed by Ekman and Friesen is still the most successful technique used in affective computing nowadays, to extract emotions from human face. FACS has enabled discovery of new relationships among emotional states and facial moments.It encodes the facial structure into sets on Action Units, whose change in intensities predicts the emotion present on the face.It was shown that each emotion is a combination of several Action Units (AU) present in FACS. And thus, concluded that facial muscular movement maps to an emotion There are few more techniques developed in this field, which are currently under research are mentioned in this dissertation.

Development in the field of behavioral sciences states making sense of decisions that are made by political leaders entails being able to tap into their emotions. To address this concern, study of politicians under emotional intelligent lens is very crucial in order to analyse behaviour aspects of leading people who govern the county. Politicians are the decision making bodies of the nation but we still know very little about the emotions of these people. We argue that politicians represent the preferences of their constituents and react differently to various levels of critical situations.According to international stats, the percentage of women sharing the political leadership in different states/ countries across the world is considerably less than men in this field. It has been stated in psychology that men and women express and perceive emotions differently. There is a subjectivity to how emotions are displayed and perceived among females and males and different cultures.There are variations in how emotions present themselves as Action Units. In the current basic emotion perspective of Automated Facial Expression analysis using FACS, contextual and cultural aspects are generally ignored.

Our main goal in this research is to adapt affective facial emotion encoding system to the study of South Asian female politicians discourse. The work present in this paper underlines the pipeline/ architecture built on different types technologies to analyse the emotional states of South Asian Female Politicians. The results prove that South Asian female leader represents themselves differently with respect to the levels of critical situation they are subjected to.

Acknowledgements

I would like to express my sincere thanks to Dr. Khurshid Ahmad who gave me the opportunity to work on this project. His advice, ideas, and supervision have made this research possible.

I would like to thank Shirui Wang, who was extremely helpful with her feedbacks, ideas and literature material.

I would like to thank Patrick and Oscar with whom I discussed my ideas for this project.

I would like to thank my family and friends for support reassurance.

Contents

List of Figures

1 Introduction

According to Oxford Dictionary Emotion is recognized as "any strong mental or instinctive feeling, as pleasure, grief, hope, fear, etc., deriving esp. from one's circumstances, mood, or relationship with others" Intelligence is defined as "The action or fact of mentally apprehending something; understanding, knowledge, comprehension (of something)."

"Emotional intelligence (EI) or emotional quotient (EQ) is the ability to recognize an individual's emotions, to discriminate between different feelings and label them appropriately, and to use emotional information to guide thinking and behavior"

An impressive breadth of interdisciplinary research suggests that emotions have an influence on human behavior. Automatic analysis of human emotional states has attracted increasing attention from developers and researchers in computer science, neuroscience, psychology and other fields. The field of emotional Intelligence is widely used in a variety of fields, such as criminology, psychology, leadership and management and other commercial usages. There have been claims that accurate information about psychopathology, personality traits and intelligence can be extracted from facial behaviour. There are various modes of expressing emotions among human beings . Facial expression is considered to be one of the most natural, powerful ways of communicating their intentions and emotions. Human face expresses an emotion earlier than they verbalize or realize their inner feelings.There has been a significant rise in facial recognition technologies to extract emotions from the human face. This study focuses on the working and analysis of existing technologies for emotional state analysis through facial expression.

Facial expression can be defined as a set of ways different facial muscles under the skin change their orientation, when they encounter any situation. Development in Artificial intelligence, Machine learning , Deep learning , Computer Vision etc, the cutting edge technologies have opened the area to develop the system that can perform emotional analysis. In previous years lots of algorithms and methods were developed to extract the presence of emotions on the face. On the basis of those algorithms, a lot of technologies are built to perform emotional analysis. One of the most successful algorithms is Facial Action Coding System FACS. FACS has enabled discovery of new relationships among emotional state and facial movements. We also discuss some new techniques in this paper that are under research and performs really well. This research focuses on the working and analysis of such technologies in order to extract emotions from the human face.

In the area of psychology, the gender difference in perception and emotion expression is an ongoing research area. It has been claimed that there is a measurable disparity between male and females in expressing their emotions in various situations. Studies have shown that women show higher levels of expression accuracy and judgement of nonverbal emotional cues than men. But these patterns cannot be 100 percent consistent. Along with the gender difference, cultural and geographical boundaries do have an impact on emotional behaviour of individuals. This is because cultural scripts dictate to an individual the norms and types of emotions that should be experienced and displayed in public and personal spaces. Cultural context acts as cues when we are trying to interpret an emotional expression.

Emotion state analysis is widely used in many areas such as criminology, leadership and management principles, psychology, and commercial areas. In leadership and management, politics is a major field of research. Politicians of any country are very important people who undertake the responsibility to run a country. Thus, it requires great leadership principles to be a successful politician.

According to Daniel Goleman

"It's worth looking at successful politicians through the emotional intelligence lens to get a sense of what a politician requires in order to be successful "

Studies show that women possess all the traits to be an influential leader. On an average, women hold only 7 percent of ministerial positions and 15 percent in national parliaments in all the 9 countries of the South Asia region. As a minority group in politics, women are often under rated for their performance and are not considered as strong politicians, when compared with male. Due to traditional cultural inequality around the world, it is still a long journey for women to mark their contributions in the political world. This research will contribute towards emotional analysis of South Asian female politicians at different occasions to further study leadership qualities among them.

1.1 Problem Statement

In the era of rapid advances in emotion detection software and technologies, there are varieties of automatic classifiers that are built using Facial Action Coding System (FACS). It is important to study the basic algorithms and techniques that act as a core for developing these systems. This research involves study of effective computing systems to recognize emotional expressions of humans and build a system pipeline that analyzes the output from the classifier to generate simplified results.

Emotions are subjective and variable, so when it comes to accuracy in emotion recognition, the matters are not that self-evident. While using machine learning algorithm's accuracy can be determined through test datasets. The comparison among all the underlying algorithms used in different emotional intelligent systems is very crucial. This helps to set the objective flagship for the upcoming new technologies and stimulate the development of better algorithms that could eliminate the shortcomings of previous algorithms and could perform better. Thus we need to perform analysis on different software kits/ API's to study the accuracy of existing techniques.

Studies prove that biological factors and cultural factors do play a significant role in determining the emotional state of human beings. Thus developing emotional intelligence requires special emphasis on these variations. Most of the Artificial intelligence systems are trained on samples that consist of Caucasian. There is a low scope of Asian faces to be subjected to emotion recognition simulation and to determine facial expression at a very high accuracy. In the study of leadership principles through emotional behaviour, women do have a very small contribution in politics. This project focuses on the study of South Asian female emotions determined by emotional recognition systems on a variety of situations. We analyse how South Asian females project their emotions in different types of critical and non-critical situations . We use these results to examine changes in aggregate levels of emotional polarities and test those emotional responses with change in critical levels of the situations.

1.2 Summary of datasets

The data chosen for performing emotion analysis comprises South Asian Female political leaders. Subjects are strong female politicians, of asian ethnicity under different situations. Politicians are categorized as semi trained actors, and the emotions they display during their interactions are semi-posed. Considering they have developed required speaking skills, the emotional analysis system will work with greater efficiency as posed emotions are easier to detect. But under critical situations subjects are likely to have emotion leakage, which can be easily detected by the system.

1.3 Structure of Report

The first section of this report provides the brief introduction and overview of study, background and core ideology that motivated this study.

The second chapter is dedicated to the research and background studies in facial expression analysis. This section provides a detailed explanation of the algorithms and techniques that develop over the span of years to achieve accuracy in emotional intelligent systems.

The third chapter focuses on the methods that are deployed to perform emotional analysis on the politicians. It also provides insight into software used for analysis and data analytic functions used to derive results and conclusions.

The fourth chapter discusses the evaluations and outcome of the simulations performed on the dataset.

The fifth chapter presents the conclusion and scope of the study conducted.

The last chapter explores the scope of future work that can be continued in this field along with some recommendations that can be beneficial for people conducting new studies related to this field.

2 Motivation

Our decision to carry out this research was motivated by the rise in importance of four influential current researches and discussions.

2.1 Emotions and Human Psychology

One concerns the psychology of human behaviour. Development in the field of behavioral sciences argue the psychological aspect of leadership and decision making are governed by emotions. The concern behind these studies is that making sense of decisions that are made by political leaders entails being able to tap into their emotions. To address this concern, we have to get over the hindrance of the apparent difficulty of monitoring these leaders to measure their state of mind. This study is focused on filling the gap by focusing on systems to extract their emotions during different situations.

2.2 Affective Computing Systems to built emotional intelligent computer systems

Second stream of research is related to computer science and engineering. It concerns the detection of human emotional states from the face. Affective Computing deals with an enormous range of computational tools for measuring emotional states communicated by humans through their verbal and non verbal cues. The most successful work in this field was published by Ekman and Friesen, known as FACS. Most of the successful technologies used in the market are built on the variants of FACS. They proposed a system which can be built using the existing technologies, such as computer vision, machine learning, deep learning etc. This study is motivated to perform analysis and understand how computing deals with facial expression encoding systems.

2.3 Politics and emotions

Politicians are semi-trained actors whose daily decisions have a lasting effect on our society. Nonetheless, we still know very little about the emotional states of those people. This research illustrates the potential of the facial emotion encoding system by tackling questions about the emotional states of our politicians. We argue that politicians represent the preferences of their constituents and react differently to various levels of critical situations.

2.4 Women in politics

Fourth concerns the women in politics.Women's equal participation and leadership in political and public life are essential to achieving the sustainable development goals by 2030. However the current scenario of politics shows that women in politics are underrepresented to very depressing figures. Achieving equal participation seems to be a very far off thing. Thus to contribute in the field of Women leadership and political participation, this study focused on analysing the emotions of Female leaders subjected to different levels of critical situations.

3 Background and Related Work

3.1 Emotions surrounding human lives

Humans are socially active creatures on the earth, they act and react to their surroundings. One of the major functions of the human brain is to evaluate the situation and reflect the inner feeling in a variety of ways. But across the borders people across the globe express their feelings very similarly to each other. We are able to differentiate the anger vs happiness emotional state of an Asian or American. Although diving deeper into this subject there might be the existence of some regional or cultural characteristics among human beings that makes it difficult to categorize them in a single set. Happiness, sadness, surprise, disgust, fear, and anger/aggressiveness are stated as six basic emotions. All other set non basic emotions can be recognized inside cultural boundaries. They are termed as culture specific emotions.

There are various ways of representing emotions discussed below Speech is a basic mode of communication among human beings. In the study of emotion recognition using speech, there was a high correlation of emotional states with acoustic variables. Text is also a basic mode of representing emotion generally used as a professional way of communication.Facial Expressions are basic and important means of communicating emotions among human beings. Gestures are the ways through which human beings emit signals through their body parts such as hands, head etc. during social interactions. Among all the non verbal behaviors of human beings - body movement, voice, posture, gaze , face is considered as the most commanding and complicated. Other non verbal communications do not occur when someone is alone but facial expressions are always present on the human face. Thus to solve the most complicated form of human communication we need more developments and research in this field.

3.2 Basic Science of Facial Expressions of Emotions

Reading facial expressions of humans can contribute to the development of trust, collegiality and rapport. They can be utilized to make credible assessments, detecting deception and evaluating truthfulness. It also provides information about emotional states which provides a better stage for cooperation and influence. People get emotional when they lie, and when they are subjected to very critical conditions. Emotional leakage can occur because of fear, guilt or shame during these conditions, especially those who are present in limelight. When people are under confident in a critical situation, they try to suppress their fear by showing another emotion such a joy or anger. Overconfidence can give rise to high anger and joy values. A lot behaviour analysis can be performed by behaviour science after knowing the emotional states of human being.

3.3 Applications of analysing emotion expression

Media Measurement:Facial coding and expression study is now a very common tool used by large companies to test their new video and content. Video data is collected as recorded responses of viewers watching TV shows, ads, trailers, etc through their webcam. Facial emotion recognition is performed on the data. The results measure the effectiveness of the videos on viewers, allowing directors of the videos to get live feedback on their content.

Non verbal leakage and clues to deception: In the research conducted in the field of psychology by Paul Ekman and Friesen, provides evidence of how non verbal behaviour can provide information which differs from the information provided by words. Non verbal signals decrease probability to deceive.

Criminology and psychiatrists : Studies prove that emotional intelligent systems are able to differentiate between suicidal and non suicidally depressed people, between lying versus telling the truth.

Leadership :Studies show a positive link between leader's and follower's emotions. In the studies published by Bono and Ilies 2006, results indicate that follower's emotional expressions are influenced from leader's expressions.

Emotions are primary drivers of human cognition, motivation, behaviour and influence in public interactions. Emotional state analysis of political leaders helps to study leadership principle and effectiveness. This resolves various issues and discussions of making an impact and influencing people to pursue collective goals. It gives deeper insight to the leader-follower relationship. Goleman in his work says

The most effective leaders are all alike in one crucial way: they all have a high degree of what has come to be known as emotional intelligence.. . . My research, along with other recent studies, clearly shows that emotional intelligence is the sine qua non of leadership. Without it, a person can have the best training in the world, an incisive, analytical mind, and an endless supply of smart ideas, but he still won't make a great leader.

3.4 Gender variations in emotional states

According to international stats(Hall et al.,2000)., the percentage of women sharing the political leadership in different states/ countries across the world is considerably less than men in this field. There exists a difference in leadership characteristics among male and female political leaders. Studies prove that men tend to be more influential leaders than women.

In the study of politicians by Rosenberg (1991), it was found that women were presumed to be more trustworthy with an almond shaped face and upper curved shaped eyes. A study by Dunbar and Burgoon (2005) stated that a person with pleasant facial expressions with a calm and relaxed voice is dominant in a relationship. As a leadership skill, people who show a relaxed and poised facial expression are implicitly powerful and credible. Too much relaxation can be assumed to be apathetic whereas medium levels of relaxation reflect a confident and composed image. Smiling too much can be presumed as submissive whereas smiling occasionally may be perceived as more likable.

3.5 Facial Signal System

There are four types of facial signal systems: static, slow, artificial and rapid. Static : Size, shape, relative locations of the features and contours that exist due to bony structure. These signals reveal information about identity and beauty.

Slow: Pouches, wrinkles and bags. These signals reveal information about age.

Artificial: Makeup, Plastic surgery. These signals concentrate on artificial additives.

Rapid : These signals reveal the actions produced by muscles and changes produced by blood velocity, body temperature, color and muscle tonus. They are the most relevant sources for all the information about attitudes, traits, personality and so on. For studying emotional intelligence we concentrate more of these signals along with knowledge of the above mentioned

3.6 Facial Muscles

When people express their emotions there are a set of facial muscle movements that are involved to produce any particular expression. Contraction of the facial muscles changes the appearance of the face. The anatomical basis of movements involved in facial muscles are the basis for deriving units of emotion detection. Friesen and Ekman stimulated the individual's muscles electrically by controlling them voluntarily. They observed that expression of the individual is associated with one or more movements of facial muscles.

Superficial facial muscles interior view is represented in 3.1. This shows how face is divided into set of different muscles.

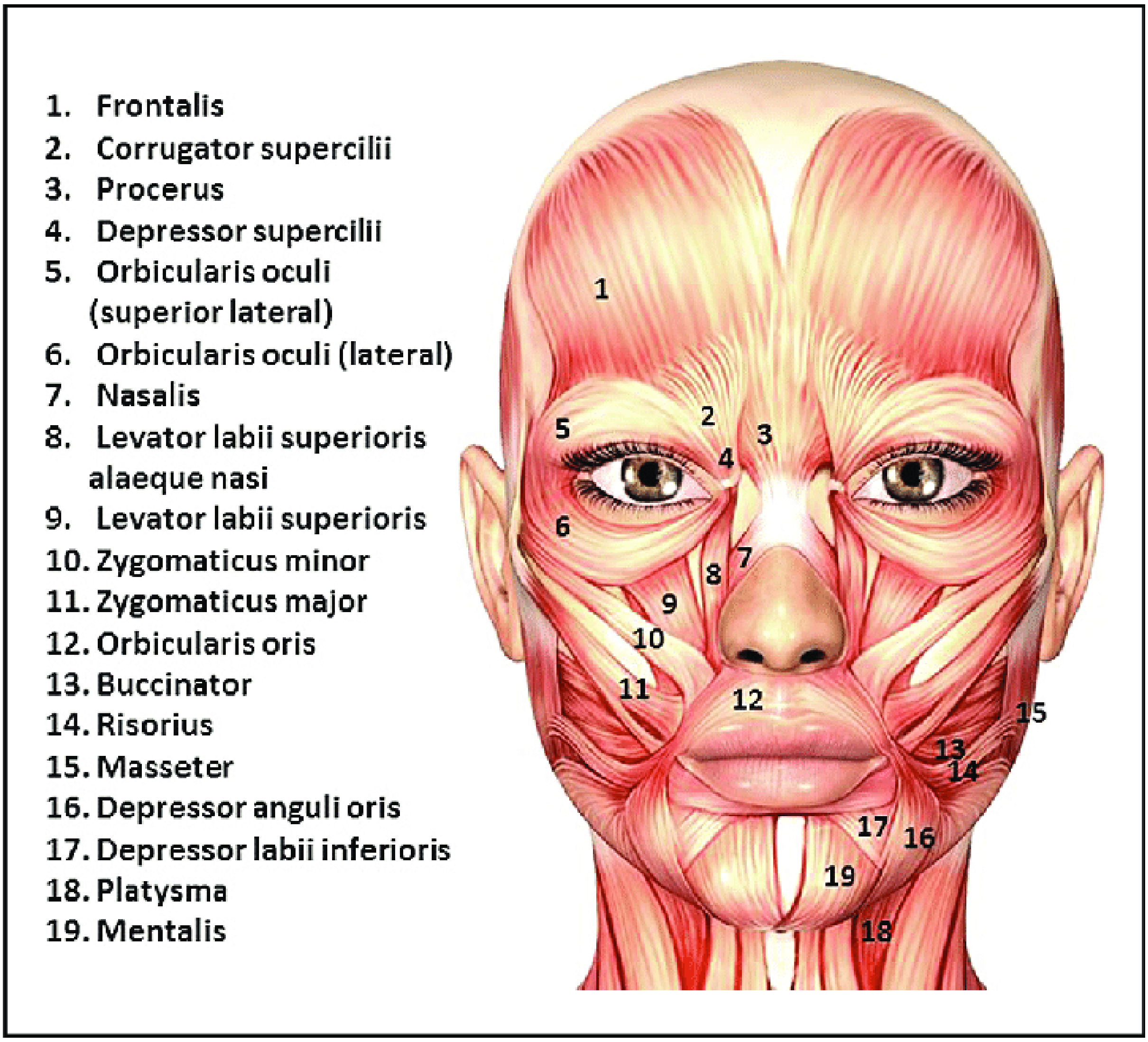


Figure 3.1: Superficial Facial Muscles- Anterior View

As a result of their studies Ekman and Friesen state an observation that in the process of detecting emotion through facial expression, they were capable of distinguishing all the visual changes that are resultant of facial muscular movements. In the case of multiple muscles participating in a single expression, the combined result produced by those muscles is evaluated.

3.7 Perceptual and affective mechanisms in facial expression recognition

According to the studies, facial expression of internal feelings or emotions is related to the physical component of morphological modulation in facial muscles. Set of movements involved in subdual facial muscles is an affective component in conveying information regarding the subject's emotions. In the research carried by (M. G. Calvo and Nummenmaa) provides the contribution of visual facial expression for analysing emotions in human beings. The research produced three conclusions:

1. "Behavioral, neurophysiological, and computational measures indicate that basic expressions are reliably recognized and discriminated from one another, albeit the effect may be inflated by the use of prototypical expression stimuli and forced-choice responses"

2. "Affective content along the dimensions of valence and arousal is extracted early from facial expressions, although this course affective representation contributes minimally to categorical recognition of specific expressions."

3. "The physical configuration and visual saliency of facial features contribute significantly to expression recognition, with "emotionless" computational models being able to reproduce some of the basic phenomena demonstrated in human observers."

Thus it is safe to presume facial muscle movements are a major factor contributing in emotional state analysis of human beings.

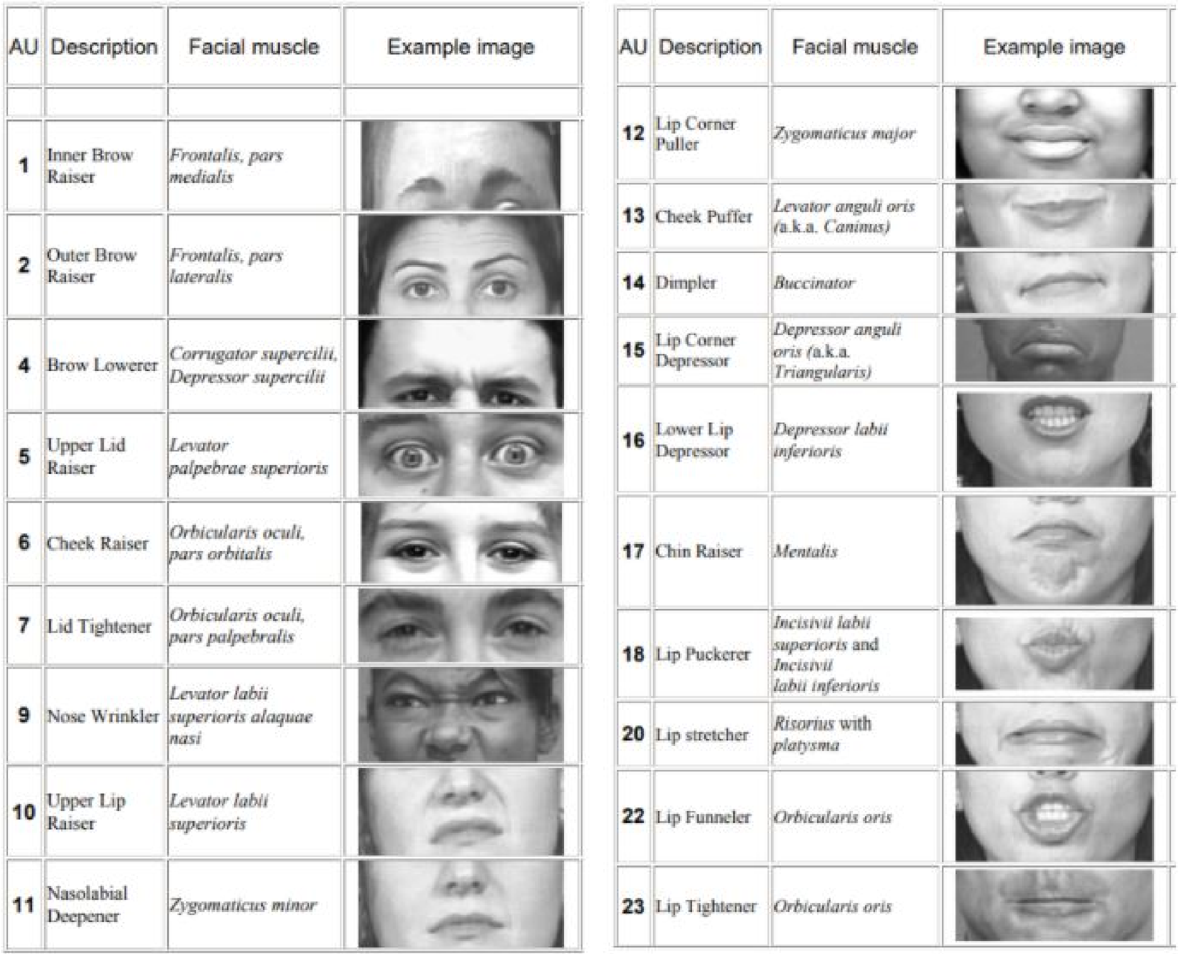
3.8 Action Units

Action Units are described as "Smallest visually discriminable facial movements". Every facial expression produced by an individual can be decomposed into Action Units (AUs). Ekman and Friesen defined forty-six AUs. All the defined AUs correspond to every independent movement/ motion of an individual' s face.

Study of AUs was a breakthrough in the field of Emotional Intelligence as other systems failed to make distinction between anatomically distinct movements. Systems usually state one to one mapping of facial expression to emotion. In the practical world emotions are not that simple, to be differentiable by few muscle movements. Individual is capable of showing a variety of mixed emotional states at a single frame of time. Thus, one to one mapping was a naïve approach to study emotional analysis. One or more AUs corresponds to the movement produced by a single muscle.

Deriving AUs:

All human beings have a similar set of facial muscles. Defined AUs are developed on the basis of "what muscles allow the face to do". To cover all the independent movements that can be produced on the human face, each Action unit is associated with one or more facial muscles. AUs are organized by the region of the face in which they reside. In 2002, Ekman and Friesen defined 18 AUs in the lower face, 9 AUs in the Upper face, 11 head movements, 4 miscellaneous AUs, 9 Action Descriptors, 4 visibility and supplementary codes, 9 Eye Positions and movements.



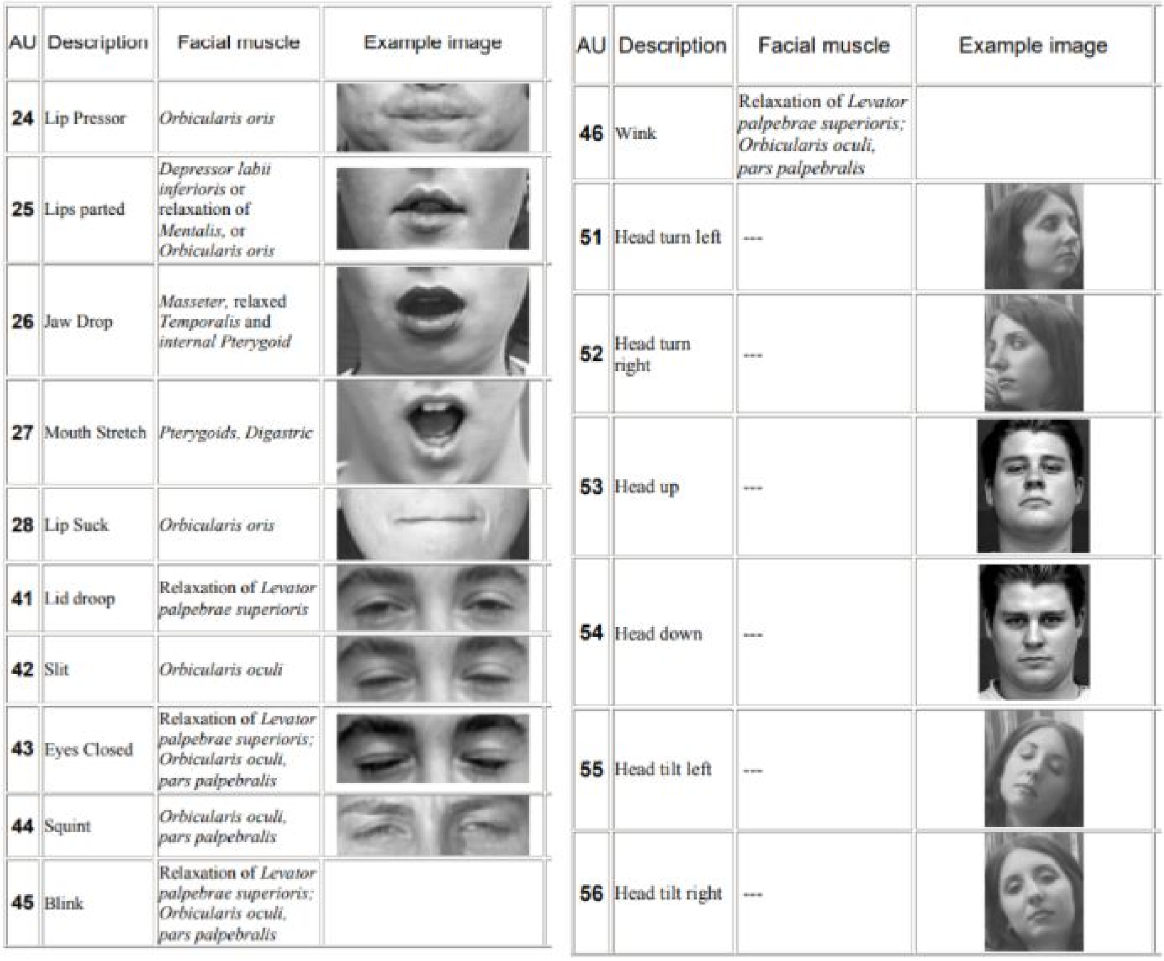
AU	Description	Facial muscle	Example image
1	Inner Brow Raiser	*Frontalis, pars medialis*	
2	Outer Brow Raiser	*Frontalis, pars lateralis*	
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*	
5	Upper Lid Raiser	*Levator palpebrae superioris*	
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*	
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*	
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*	
10	Upper Lip Raiser	*Levator labii superioris*	
11	Nasolabial Deepener	*Zygomaticus minor*	
12	Lip Corner Puller	*Zygomaticus major*	
13	Cheek Puffer	*Levator anguli oris* (a.k.a. *Caninus*)	
14	Dimpler	*Buccinator*	
15	Lip Corner Depressor	*Depressor anguli oris* (a.k.a. *Triangularis*)	
16	Lower Lip Depressor	*Depressor labii inferioris*	
17	Chin Raiser	*Mentalis*	
18	Lip Puckerer	*Incisivii labii superioris* and *Incisivii labii inferioris*	
20	Lip stretcher	*Risorius* with *platysma*	
22	Lip Funneler	*Orbicularis oris*	
23	Lip Tightener	*Orbicularis oris*	

Figure 3.2: Facial Action Units of Coding System-Table 1

3.9 Compound facial expressions of emotions

Study of different kinds of emotion expressions is a very essential factor in the development of automatic facial action coding systems(FACS). Past research in this field categorizes human emotions into six basic categories- surprise, happiness, anger, fear, sadness, disgust. But in the real world there is an existence of more emotions. There are groups of defined emotions called compound emotions. Compound emotions refer to the set of emotions that are constructed by combining basic emotions to define a new state of emotion. In FACS there exist 22 sets of emotions (shown in 3.4) which are visually discriminable from each other.

The AUs of the basic emotions are combined as shown in 3.5 to produce the compound category. The AUs of the basic expressions kept to produce the compound emotion are marked with a bounding box. These relationships define the subordinate classes of each



AU	Description	Facial muscle	Example image
24	Lip Pressor	*Orbicularis oris*	
25	Lips parted	*Depressor labii inferioris* or relaxation of *Mentalis*, or *Orbicularis oris*	
26	Jaw Drop	*Masseter*, relaxed *Temporalis* and *internal Pterygoid*	
27	Mouth Stretch	*Pterygoids, Digastric*	
28	Lip Suck	*Orbicularis oris*	
41	Lid droop	Relaxation of *Levator palpebrae superioris*	
42	Slit	*Orbicularis oculi*	
43	Eyes Closed	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	
44	Squint	*Orbicularis oculi, pars palpebralis*	
45	Blink	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	

AU	Description	Facial muscle	Example image
46	Wink	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	
51	Head turn left	---	
52	Head turn right	---	
53	Head up	---	
54	Head down	---	
55	Head tilt left	---	
56	Head tilt right	---	

Figure 3.3: Facial Action Units of Coding System-Table 2

category and their interrelatedness. In turn, these results define possible confusion of the compound emotion categories by their subordinates and vice versa.

The prototypical AU observed in each basic and compound emotion are shown in 3.7

3.10 Facial Action Coding System (FACS)

There have been a lot of studies involved in simulating emotional intelligence in computer systems. Many systems were able to detect single basic emotions such as fear, sadness, happiness, fear, disgust, etc for real time data. But analyzing emotions is a more complex system, as there are a mixture of emotions that humans can express through their facial movements. (Ekman and Friesen, 1976)FACS takes into account all the dependent and independent facial muscles movements and contractions in order to return the set of emotional states reflected in every frame. This system has proved to be the most successful system so far and has been able to differentiate between simulated and genuine pain, between suicidal and non suicidally depressed people, between lying versus telling the

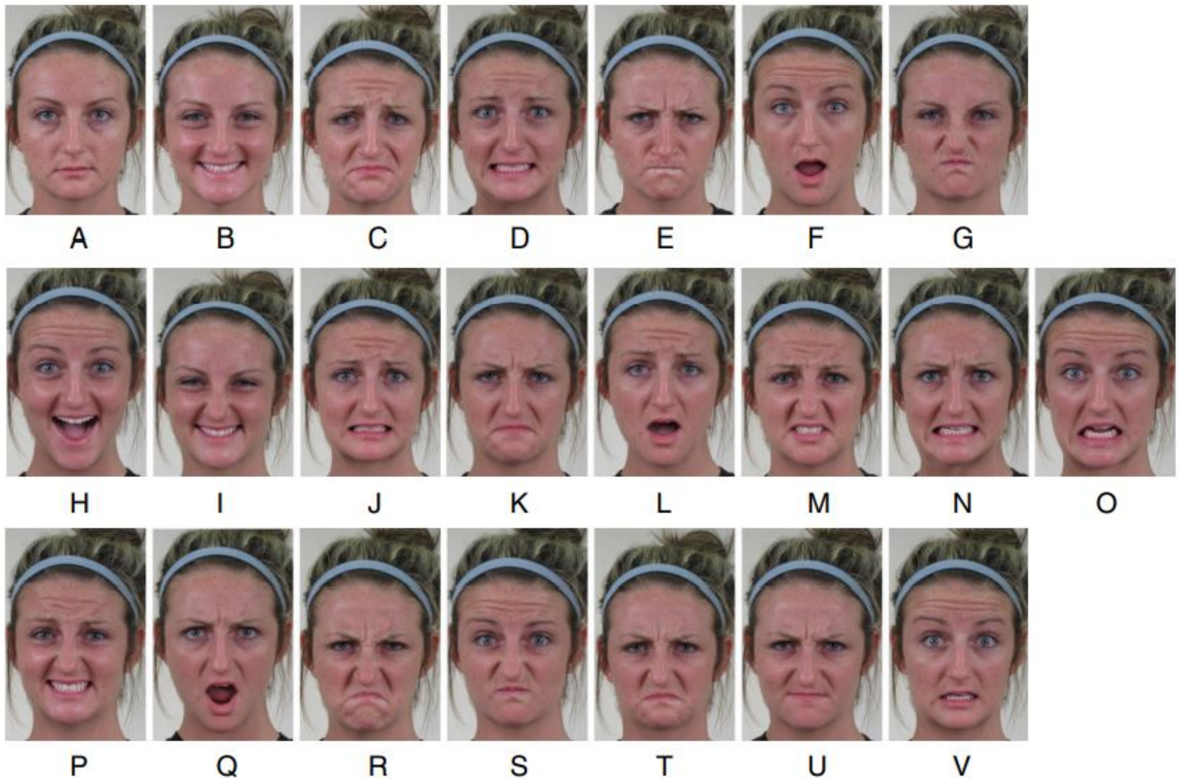


Figure 3.4: (Cohn et al., 2007).Sample images of the 22 categories in the database: (A) neutral, (B) happy, (C) sad, (D) fearful, (E) angry, (F) surprised, (G) disgusted, (H) happily surprised, (I) happily disgusted, (J) sadly fearful, (K) sadly angry, (L) sadly surprised, (M) sadly disgusted, (N) fearfully angry, (O) fearfully surprised, (P) fearfully disgusted, (Q) angrily surprised, (R) angrily disgusted, (S) disgustedly surprised, (T) appalled, (U) hatred, and (V) awed

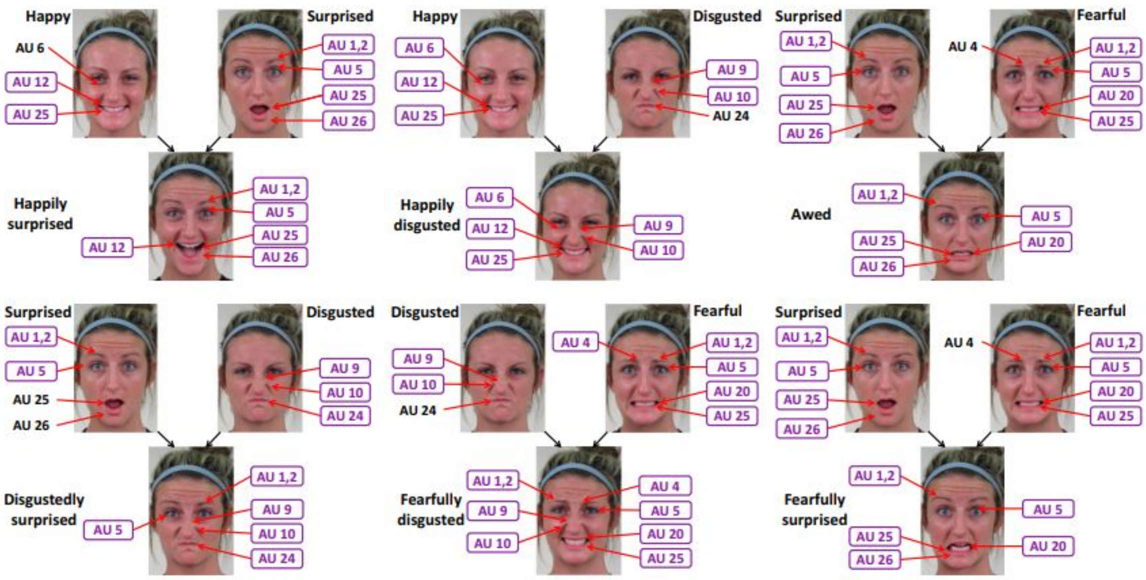


Figure 3.5: AUs of six compound facial expressions of emotion(Cohn et al., 2007).

Category	Prototypical (and variant AUs)
Happy	12, 25 [6 (51%)]
Sad	4, 15 [1 (60%), 6 (50%), 11 (26%), 17 (67%)]
Fearful	1, 4, 20, 25 [2 (57%), 5 (63%), 26 (33%)]
Angry	4, 7, 24 [10 (26%), 17 (52%), 23 (29%)]
Surprised	1, 2, 25, 26 [5 (66%)]
Disgusted	9, 10, 17 [4 (31%), 24 (26%)]
Happily surprised	1, 2, 12, 25 [5 (64%), 26 (67%)]
Happily disgusted	10, 12, 25 [4 (32%), 6 (61%), 9 (59%)]
Sadly fearful	1, 4, 20, 25 [2 (46%), 5 (24%), 6 (34%), 15 (30%)]
Sadly angry	4, 15 [6 (26%), 7 (48%), 11 (20%), 17 (50%)]
Sadly surprised	1, 4, 25, 26 [2 (27%), 6 (31%)]
Sadly disgusted	4, 10 [1 (49%), 6 (61%), 9 (20%), 11 (35%), 15 (54%), 17 (47%), 25 (43%)*]
Fearfully angry	4, 20, 25 [5 (40%), 7 (39%), 10 (30%), 11 (33%)*]
Fearfully surprised	1, 2, 5, 20, 25 [4 (47%), 10 (35%)*, 11 (22%)*, 26 (51%)]
Fearfully disgusted	1, 4, 10, 20, 25 [2 (64%), 5 (50%), 6 (26%)*, 9 (28%), 15 (33%)*]
Angrily surprised	4, 25, 26 [5 (35%), 7 (50%), 10 (34%)]
Angrily disgusted	4, 10, 17 [7 (60%), 9 (57%), 24 (36%)]
Disgustedly surprised	1, 2, 5, 10 [4 (45%), 9 (37%), 17 (66%), 24 (33%)]
Appalled	4, 10, [6 (25%)*, 9 (56%), 17 (67%), 24 (36%)]
Hatred	4, 10, [7 (57%), 9 (27%), 17 (63%), 24 (37%)]
Awed	1, 2, 5, 25, [4 (21%), 20 (62%), 26 (56%)]

AUs used by a subset of the subjects are shown in brackets with the percentage of the subjects using this less common AU in parentheses. The underlined AUs listed in the compound emotions are present in both their basic categories. An asterisk (*) indicates the AU does not appear in either of the two subordinate categories.

Figure 3.6: AUs of six compound facial expressions of emotion

truth.

The Facial Action Coding System developed by Ekman and Friesen, is the most psychometrically rigorous and comprehensive system in the field of Emotion Analysis through facial expression. FACS provides an "objective description of facial signals in terms of component motions, or facial actions". Ekman defined Action Units as the basic measurement units for his model. This model studies the correlation of a set of action units present on the face and the emotional state corresponding to it. Trained FACS coding system decomposes facial expression into action units which correspond to subdual facial muscles.

Study of facial expressions in human beings gives a very important behavioral measure in emotional analysis. The facial Action Coding System (FACS) developed by Ekman and Friesen quantifies facial movement in terms of Action Units. Computer Vision techniques are used to detect facial action units in the sequence of frames/images. Three techniques were used: holistic spatial analysis, estimation of motion (movement in facial muscles) flow fields and explicit measurement of features such as wrinkles. All three techniques are concatenated in a hybrid system which was able to classify 6 upper facial action units with an accuracy of 91 percent. The produced hybrid system shows better performance than human experts and non experts. On the basis of FACS, an automated system is developed to simulate emotional analysis using software solutions. This makes facial expression analysis accessible as an observational research tool for people studying Emotional Intelligence in Computer systems and behavioral sciences

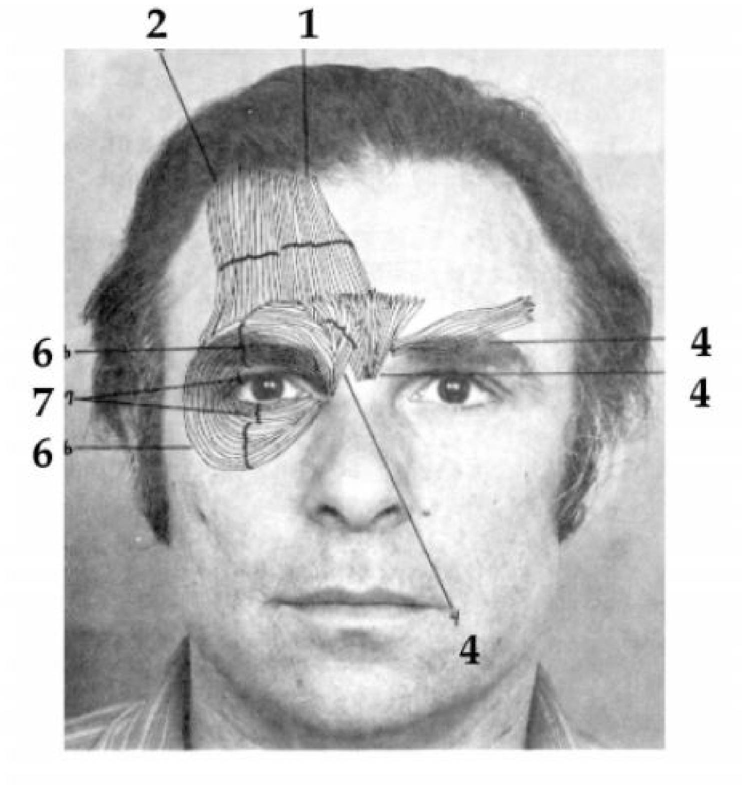


Figure 3.7: Decomposition of facial motion into component actions. Facial muscles corresponding to action units 1,2,4,6,7 are represented

3.11 FACS Design

The visual pipeline of FACS design is represented in the 3.8 and detailed implementation of the design is mentioned below.

Multiple image windows at a variety of locations and scales: Multiple windows on the face are defined to detect and differentiate multiple features on the face, such as eye, chin, nose, etc. Haar Cascades in Computer Vision recognize and define windows for different features.

Image Filters: On each window defined in the previous step multiple image filters are applied. There are a variety of image filters present in Computer Vision libraries. One or multiple filters are applied on the windows to minimize the noise in the image and spot all the useful extractions minimizing external image variables such as light, colors etc. Pixels and color intensities present in the image are distributed and quantified in every window.

Machine Learning techniques for feature selection, trained on spontaneous expression: In this step only the useful parameters and characteristics of the image required for extraction

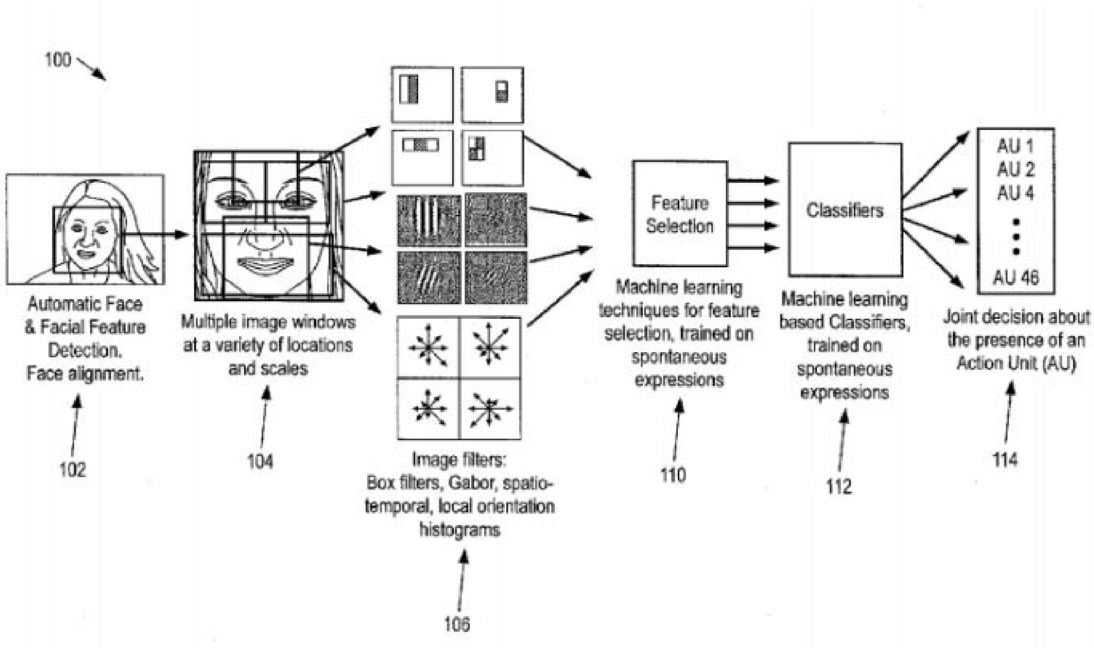


Figure 3.8: FACS architecture(Bartlett et al., 2014)

of AUs presence are selected and passed to the classification step. Spontaneous facial expressions are very much different from posed expressions in the dynamics of movements produced by muscles. On image frames, machine learning algorithms such as AdaBoost, SVM(Support Vector Machine) are applied for feature selection. Intensity evaluation in a series of frames enables the system to investigate facial features.

Machine Learning based classifiers trained on spontaneous expressions: Within the features and characteristics from the previous step, the relative likelihood of the Action Unit presence is estimated.

Joint Decision about the presence of an Action Unit(AU): Resultant of the system is calculated as a set of scores that represents the probability of a predefined Action Unit present in the frame.

3.12 FACS Scoring

FACS result is a set of scores, implying the presence of a particular action unit in a frame. Some of the AUs are omitted in scoring, such as head and eye positions. FACS score is based on 19 Action Units along with yaw degrees, pitch degrees, roll degrees, etc. Scoring in FACS can be done in two ways- Selective Coding and Comprehensive Coding- Selective Coding: In this scoring system, only predetermined AUs are coded; any other AU that might be present in the frame is ignored. It is more time efficient, but bit less reliable than comprehensive coding Comprehensive Coding: In this scoring system each and every AU, which is present in the frame, is coded. It allows the system to have a better analysis in a variety of ways, but performs very poorly in the terms of time complexity. It can take 100 mins to process a 1 min video. Most of the studies are inclined towards Selective Coding as it is more time efficient and less complex in terms of algorithmic complexity.

Intensity of the action unit plays a major role in scoring. FACS allows 5 levels of intensity scales in coding. Higher and lower levels of intensities are more reliable with central intensities being the least reliable with high error values corresponding to it.

3.13 FACS success

A study conducted with 26 participants who were videotaped under 3 conditions for carrying out research: Posed pain, real pain, baseline. Posed pain conditions were carried out using cold pressure induced by ice on the participant's arm. The objective was to compare the results of FACS and human expressions experts. And develop a classifier to automatically underline the difference in real and fake pain. This experiment produces the result using two classification stages. First one uses FACS to carry out AU scores and then classify the real vs fake pain. Humans were able to differentiate among the two scenarios with 49 percent accuracy whereas an automated FACS system was able to get 88 percent accuracy in its results.

3.14 The Computer Expression Recognition Toolbox (FACET/CERT)

CERT: "Software tool for fully automatic real-time facial expression"

This toolkit is based on the Facial Action Coding System (FACS). This toolkit is designed to code the intensity of nineteen Action Units, and six prototypical facial expressions. 3 D orientations in video are taken into account such as (yaw, roll, pitch) of head. CERT's performance in analyzing facial actions units is evaluated as 90.1

Using the FACS, CERT /FACET estimates the presence of 6 universal emotions. FACS performs the elementary task of decomposing facial muscles into an action unit, which corresponds to facial muscular movements. The result of FACS provides the activation score of Action units present on the face.

1. Face Detection: In the CERT toolbox, face detection is done using WaldBoost (automatic cascade threshold selection) and GentleBoost (boosting Algorithm). CERT's face detection achieved an accuracy of 80.6 percent. Among the group videos, CERT detected the largest face present on the screen.

2. Facial Feature Detection: CERT uses feature specific detectors for detecting 10 facial

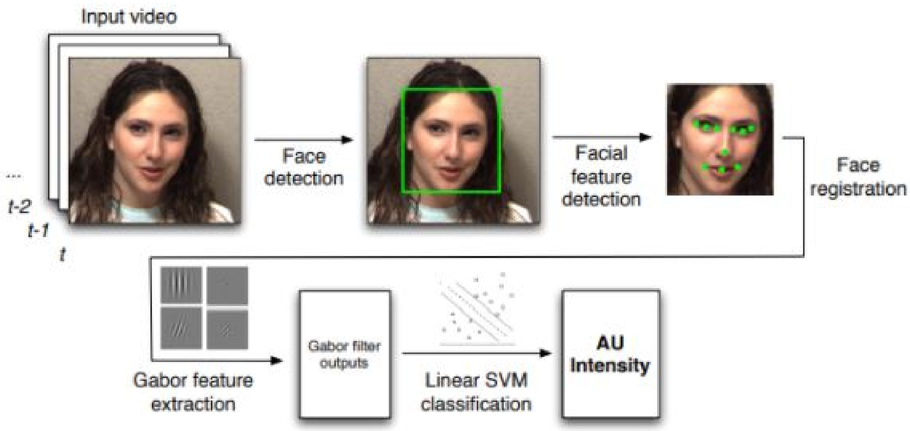


Figure 3.9: Pipeline of CERT from video to expression intensity estimates

features consisting of eye centers, mouth corners (inner and outer), nose tip, and mouth center. Feature detectors give the probability of feature being present to feature not being present at a specific location l. At location, prior and posterior probabilities of features are detected as well. The estimations at locations are recalculated by implementing linear regression

3. Face Registration: With the positions of facial features estimated in the previous step, the face patch is estimated again using an affine wrap. In fig 1 the green box is shown after re-estimation. Pixels are structured as two-dimensional arrays. This array is used in the next steps of the CERT toolkit. 4. Feature Extraction: Canonical sized picture with face patch is convolved (Filter bank of 72, Gabor filters) using Fast Fourier Transform. The values of filter outputs are combined to form a single feature vector. 5. Action Unit Recognition: The result of the previous step feature vector is fed to the support vector machine (SVM) for each Action Unit. Output of the SVM gives the estimation of Action Unit intensities.

CERT accepts real time data as an input(video). Then the simulation video is broken into frames and the emotion state of the subject is evaluated. There are many technologies that are used in CERT for running the simulation. Computer Vision is used for recognition, feature extraction, filtering etc. Machine Learning models and Convolution Neural Network Models are used by CERT to evaluate probabilities of AU intensities for a series of frames.

3.15 Facial detection Techniques and Principal Component Analysis

First step involved in the system is to process the image in order to detect faces. Face detection is performed using Computer Vision techniques. There are a number of face recognition methods and algorithms that use an unsupervised statistical framework. Face is represented as a linear combination of the set of basis images found by these algorithms. Principal Component Analysis is one of these methods. Basis images found by PCA are dependent on the pairwise relationship among the pixels in an image dataset. Image recognition process requires a higher order relationship function among pixels, we require more complex techniques to find image basis. ICA (independent component analysis) is a method derived from the Principle of Optimal Information transfer. ICA was performed on 2 different architectures. First, architecture considers pixels as random variables and images as outcome. This produces a factorial face code. Other architectures considered images as variables and pixels as an outcome. In fact, this architecture found spatially local basis images. A Classifier was used to combine these two architectures which gave better performance than PCA.

In the mechanism of face detection, there are many problems pertaining to light, facial expression, pose, orientation and picture quality. There is a requirement of image pre-processing. As we have discussed in previous sections, facial expression detection is divided into roughly four parts. Face Detection, Normalization, Facial extraction and classification.

There are two techniques used in the facial expression recognition process. Feature based approach : This approach targets local features such as nose, eyes, lips etc. Features are segmented and then used as input data, which is then passed to the classifier. Holistic approach : This approach uses statistical methods to extract statistical characterization from training samples such as support vector machine(SVM), nearest feature lines (NFL) eigen faces, probabilistic eigen faces , fisher faces etc . Hybrid approach: This approach is a combination of the above two mentioned approaches. For eg. Modular eigen face, component-based methods, hybrid local features, normalized shape etc.

3.16 Image Pre-processing techniques

Image pre processing techniques involve reading the image, detecting the image, and identifying facial feature points. The first step is to extract the shape of eyes, nose, mouth and chin. This helps to distinguish the face by scale and distance of organs. Image detection identifies the region of the face, disregarding the orientation and lightning conditions present in the frame. The last step highlights the features of the face. Some noise removing algorithms can also be used to eliminate errors caused due to noise present in the frame. Edges of the frames are detected and marked in the frame.

3.17 Emotional Facial expression study with 3D graphics

The work published in International Journal of Electrical and Computer Engineering (IJECE) 2016 aims to demonstrate the combination of Facial Action Units to depict an emotion on a virtual human face(3D graphic face). Facial Action Coding system has become a standard to control human's facial expressions. This technique has revealed a full classification of human face regions. This helps a computer graphic animator to depict the emotional facial expression of a virtual human. Thus, an avatar's expression is created by a combination of activated Action Units on facial muscles. This approach is very adaptable and attractive and can be used in the graphic animation field.

3.18 Reliability of FACS

FACS is trained on posed facial expressions by actors. We used this system for emotional state analysis of spontaneous facial behaviors in the real time video of our subjects. In spontaneous expressions face size is usually smaller, camera orientation is not at perfect straight angles, motion of the face is variable, etc. Thus, spontaneous expressions are less reliable than posed expressions. Our score values of FACS are accompanied with error values corresponding to each action unit present in the frame. Least frequently occurring AUs have the lowest levels of reliability. For each AUs there are 4 types of reliability criteria stated by FACS. Non-occurrence / Occurrence of an AU, Temporal Precision, Intensity,
Aggregates

Temporal precision deals with the timing of AU. Intensity of the AU present in the frame becomes important when hypothesis testing focuses on the probability of intensity being related to individual differences or subjective experience. By accessing all the aggregates and combining them together, Ultimate reliability of the Action Unit can be estimated. So, AU scores produced by FACS contain the error window corresponding to it.

Accuracy of the AU is represented in 3.10. This shows that the accuracy of emotion extracted by FACS is different for different emotions.

3.19 Facial Expression Recognition using Facial Landmarks

FACS is a successful method of extracting facial emotions from the human face. Last few years have recorded enormous development in GPU- based parallelism to boost performance. Convolutional Neural Networks come to the picture when image processing is involved. This

AU	Name	N	P	Hit	FA	A'
1	Inner brow raise	81	99	100	1	100
2	Outer brow raise	196	95	93	5	98
4	Brow Lower	403	88	89	12	95
5	Upper lid raise	286	92	88	7	96
6	Cheek raise	68	100	88	0	100
7	Lower lid tight	203	89	86	10	93
9	Nose Wrinkle	409	92	86	7	95
10	Lip Raise	278	93	86	6	96
11	Nasolabial	100	85	85	14	91
12	Lip Corner pull	315	88	85	12	92
14	Dimpler	376	89	83	9	93
15	Lip Corner depr.	412	89	76	9	91
16	Lower Lip depr.	49	92	64	7	88
17	Chin raise	86	93	58	5	85
29	Lipstretch	99	92	57	6	84
23	Lip tighten	57	91	42	8	70
24	Lip press	39	94	33	4	74
25	Lips part	50	97	29	2	90
26	Jaw drop	47	98	29	1	92
27	Mouth stretch	32	99	20	0	85
	Mean		92.75	68.85	6.25	90.4

Figure 3.10: Accuracy of prediction of AU

gave rise to new techniques of developing Facial Expression Recognition tools. It involves training CNNs to perform extraction. An 64*64 images are fed as an input to the CNN system which includes one input layer, five convolution layers, three pooled layers, one fully connected layer and one output layer. The fully connected layer is combined as the input of the softmax layer to obtain the output class after the convolutional pooling operation. They have used JAFFE and CK+ database for validation.

Latest study also shows a novel approach for developing facial recognition systems using facial landmarks which surprisingly gave promising results.

This proposed methodology for a novel approach is explained below:

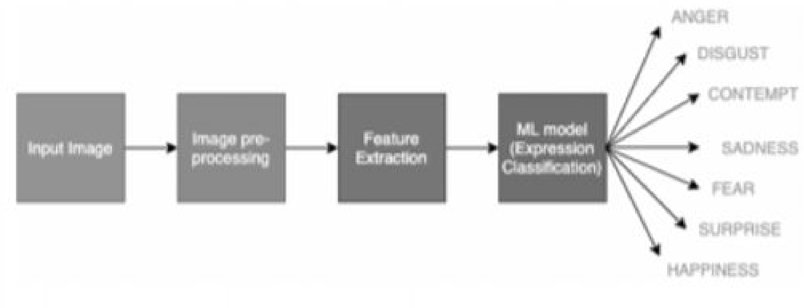


Figure 3.11: Block Diagram of proposed methodology

1. Facial Expression Image Database: The Support Vector Classifier is trained with a basic set of facial expression databases. This consists of a total 7 emotions which are expressed by 210 human beings.The basic expressions include anger, contempt, disgust, fear, happiness, sadness, and surprize. The size of image is 48*48 pixels.

2. Image pre processing: Image pre-processing is performed by inbuilt tools provided by OpenCv. This step eradicates disturbances caused due to uneven lightning by performing Contrast Limited Histogram Equalization. The image is also converted to grayscale to minimize the contrast complexities in an image.

3. Facial Landmark Prediction: In order to predict facial landmarks , face detection algorithm must run on the image. After identifying the face in the image, a classic HOG, the rectangular object detector, is developed on the face region. After face detection, the Dlib function is used to predict the 68 facial landmark points. The model is trained by iBUG 300-W face landmark dataset.

Figure 3.12: Predicted Facial Landmarks

4. Feature Extraction: For feature extraction it is very important to know the positions of the facial landmarks relative to one another. It involves calculation of the mean(both axes) which results in one center point. After we get the center point, lines are drawn from the

center point to each facial landmark .Each landmark can be now represented as a vector containing magnitude and direction.The vectors are concatenated to form feature vector that is used for training and classification stage.

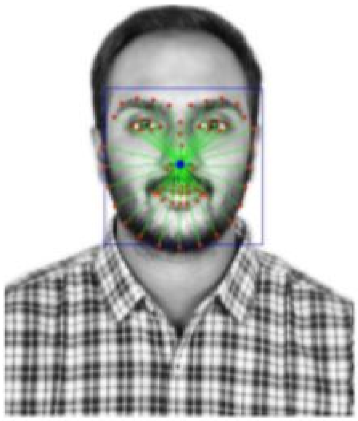


Figure 3.13: Representing feature vector with centre point

5. Training and Classification using SVC: After identifying the feature vector, facial expressions can be identified. The machine learning algorithm used for classification is Support Vector Machine. SVM provides very high accuracy with high dimensional space. The training and testing set was distributed at 80:20 ratio to provide versatility.

3.20 Reliability of facial Landmark Methodology

This method works surprisingly better than the CNN approach discussed earlier. The confusion matrix of the above experiment is displayed in n3.14:

The overall accuracy of the method is 89 percent, which is a very high accuracy on the test set. Thus it is evident that potential ability for recognizing facial expressions based on facial landmarks is quite promising. The facial landmark methodology works successfully on image sets, and further work in this field is to add support for real time recognition as done by FACS.

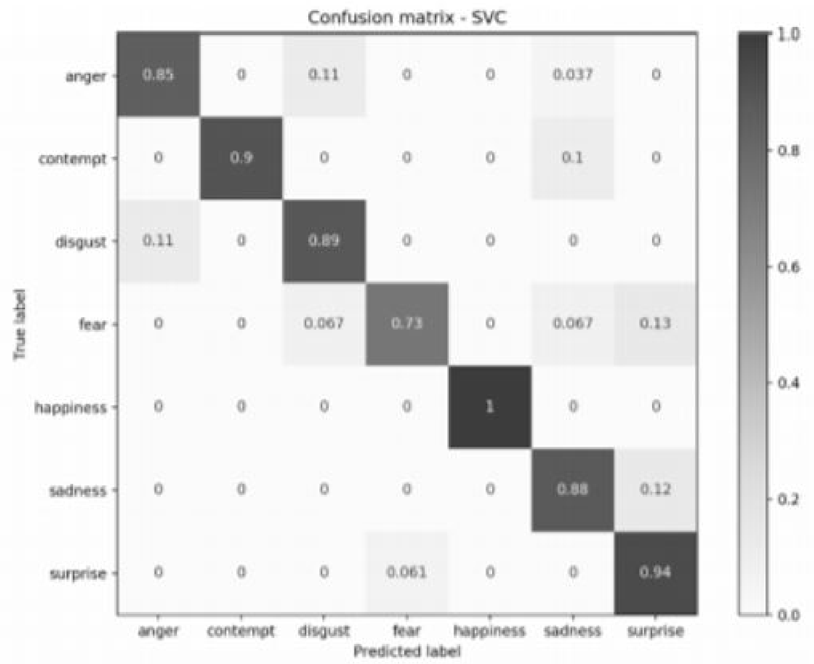


Figure 3.14: Confusion matrix on test dataset

4 Design and Methodology

This chapter contains a description of methodology used in the study and implementation of emotional intelligent systems.

4.1 Architecture

The pipeline of the design flow is represented in 4.1.

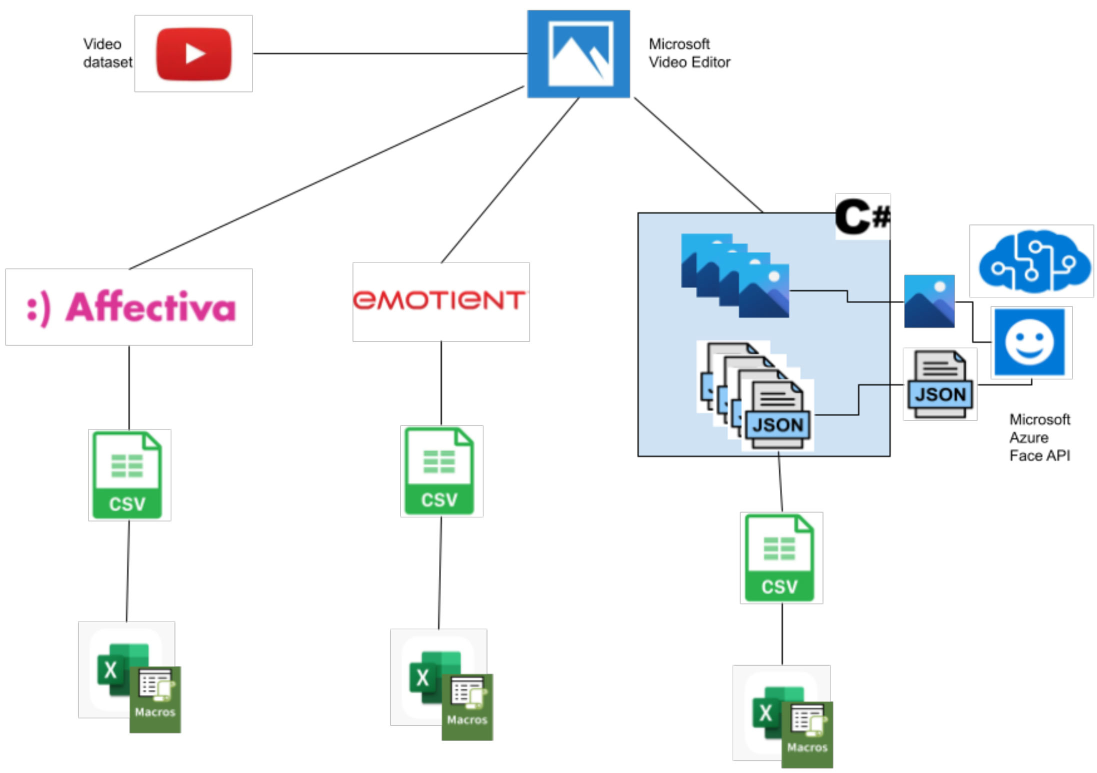


Figure 4.1: Pipeline of architecture used in this research

4.2 Collection of data

Collection of data is the first step to proceed with our studies. Our studies focus on female south Asian female politicians/leaders. The individuals chosen for this study are famous and highly trained politicians working in their fields from quite a long time. Our dataset contains the videos of politicians available on open sources such as youtube. We tried to gather different types of events for each politician. Events consist of 1:1 interviews, parliament sessions and public speeches on different topics.

1:1 Interview sessions

These sessions provide the two interaction units for a politician, Host and a Camera. This situation gives a very nice opportunity to showcase their empathy and build lasting relationships with viewers by managing their emotions in a healthy way. Personal Interviews allow politicians to create a personal connection with the masses, without actual physical interaction. Studies prove that politicians who can personally connect with the audience can be more influential than the ones with better knowledge and experience related to the field. Emotional Intelligent candidates can be more influential and have a better chance to mark a victory. 1:1 interviews are prepared situations for politicians, with minimum physical interaction. This situation lies in a lower position on critical situation analysis. We collected sessions of 1:1 interviews and fed it to Emotional Intelligent software to extract the emotions represented by the subject in the session.

Public Speaking

Public Speaking situations provide a much greater level of interaction with a large number of audience present physically at the situation. It becomes very unimportant to manage the self-emotions and emotions of the audience in order to get the message across well.

“It’s worth looking at public speaking through the emotional intelligence lens to get a sense of what a public speaker requires in order to be successful ” Influence of speech to the audience is the essence of democracy since the 90s. Thus to analyse the emotional state of subjects, we collected the public speaking videos of the subjects. The videos are general public addressing sessions, media press conferences and event gatherings. . This situation lies in a medium position on critical situation analysis.

Parliament Sessions

Parliament sessions are very important sessions to create a road-map for functioning of government, where politicians discuss various laws and policies for the country. Speeches in parliament sessions determine the career of a politician in the party. It is crucial to inspect the emotional states of politicians in parliament sessions with an emotional intelligent lens. Politicians are subjected to a very judgemental state, where there are hundreds of active

listeners ready to attack with a number of objecting to the facts of one's speech. It has been recorded in the previous sessions, discussions turning to heated debates, politicians losing their patience and attacking each other personally. Thus this situation can be classified under high position on critical situation analysis.
Dataset of South Asian Female Leaders is present in 4.2

Politician	Ethnicity	Age	Topic	Type of Speech	Length
Mayavati	Indian	65	General Politics discussion	1:1 Interview	02:53
Mayavati	Indian	65	After losing elections	Public speaking	03:59
Mayavati	Indian	65	BR Ambedkar Religious Intolerance	Parliament session	04:33
Harsimrat Kaur Badal	Indian	55	New farmer's bills 2020	Public speaking	04:10
Harsimrat Kaur Badal	Indian	55	Pulwama attack	1:1 Interview	03:05
Harsimrat Kaur Badal	Indian	55	Agriculture policies of the country	1:1 Interview	03:08
Harsimrat Kaur Badal	Indian	55	Farmer's Protest	Parliament Session	04:44
Smriti Irani	Indian	45	Farmer's Protest	Public speaking	03:49
Smiti Irani	Indian	45	New farmer's bills 2020	1:1 Interview	03:59
Smriti Irani	Indian	45	Party's Internal conflicts discussion, BJP conference	Public speaking	03:40
Smriti Irani	Indian	45	Condemn Opposition bills	Parliament session	03:27
Mamta Banerjee	Indian	65	General Politics discussion,	1:1 Interview	03:37
Mamta Banerjee	Indian	65	Nabana Press Conference	Public speaking	03:12
Mamta Banerjee	Indian	65	Reasons for Mamta Banerjee Leaving BJP in past	1:1 Interview	02:31
Mamta Banerjee	Indian	65	Railway Budget	Parliament Session	01:58
Sushma Swaraj	Indian	65	Delhi Nirbhaya gang rape case	1:1 Interview	03:38
Sushma Swaraj	Indian	65	External Affairs Policies	Public Speaking	03:12
Sushma Swaraj	Indian	65	Forign Affairs	Parliament session	03:28
Agatha Sangma	Indian	41	Parliament Speech	Public Speaking	04:03
Agatha Sangma	Indian	41	Mass exodus of the North-Eastern population	1:1 Interview	03:09
Elian Chao	Taiwanese	68	Elian Chao on returning to public services	1:1 Interview	01:49
Elian Chao	Taiwanese	68	Public conference, Personal career talks	Public speaking	05:22
Elian Chao	Taiwanese	68	Department of Transportation discussion	1:1 Interview	05:06
Elain Chao	Taiwanese	68	New York Metro Gateway Project	Parliament Sessions	02:37
Margaret Chin	Hong Kong	68	Committee on Aging Speech	Public Speaking	04:22
Margaret Chin	Hong Kong	68	City's Borough-Based Jail Plan	City Council Forum	02:33
Margaret Chin	Hong Kong	68	Planned Parenthood	1:1 Interview	00:46
Margaret Chin	Hong Kong	68	Non-partisan information about candidates in the city's elections	Vote appeal campaign	01:44
Margaret Chin	Hong Kong	68	Debate on district laws	District City Council Candidate Forum	02:23
Karim Begum	Nepalese		Karim Begum slapped CDO of Paras district	1:1 Interview	04:14
Dil Kumari Bhandari	Indian	72	Exclusive interview	1:1 Interview	04:09
Dil Kumari Bhandari	Indian	72	Constitutional recognition of Nepali language in India	Public Speaking	04:51
Azru Rana Deuba	Nepalese	59	School inauguration	Public speaking	00:45
Azru Rana Deuba	Nepalese	59	Movie release	1:1 Interview	03:59
Azru Rana Deuba	Nepalese	59	Parliament speech	Parliament Session	00:52

Figure 4.2: Dataset used for emotional analysis

4.3 Pre processing of data

Pre-processing of videos is required in order to run the simulation on the face. The videos available on open source comprise various faces in a single frame. Camera doesn't target the politician's face throughout the video, thus every other face present in the video needs to be eliminated to run the simulation on the target face. This involves working with various professional video editors.

4.4 Emotional Analysis

This is the core step of this whole study which involves running simulations on the dataset to analyse emotional states of target faces at a particular time stamp. We used three different software which takes the mp4 file as an input and returns the emotional analysis of the videos.

4.4.1 Microsoft Azure Face API

Microsoft Azure Face API performs emotional analysis to detect anger, sadness, disgust, fear, surprise, contempt, neutral and happiness. Emotional recognition by Microsoft is performed via the client library and REST API.Microsoft cognitive services allow users to access the Face API via key. Face API - v1.0 is able to return the mentioned results.

Face Detection: It detects the face and assigns an unique FaceId to it.

Recognition: It is able to recognize the face if the system already contains the faceid mapped to a particular face.

Face Rectangle: The facial area is enclosed in a rectangular box

Landmarks: Facial Landmarks are used to locate the salient features of the face such as eyes, nose, lips, jawline, ears, etc. Identification of facial landmarks are very crucial for higher order computer vision tasks such as face recognition and facial emotional analysis.

Face Attributes : These are the attributes that are derived by using multiple machine learning and deep learning models on the frames. The list of attributes is age, gender, smile(intensity),facial hair, head pose(3D roll, yaw, pitch angles), glasses, emotion, hair, makeup, accessories, blur, exposure, occlusion, mask From all the existing valuable attributes, we are interested in emotion.

The emotion recognition is performed on a local machine setup using visual studio and.NET Desktop application development workload. Face service NuGet packages are the client libraries that provide access to the Microsoft cognitive services required for FaceAPI. The NuGet Packages required for FaceAPI are mentioned in 4.3.
The services used fro FACE API are mentioned in 4.4
The data processed by the Azure Cognitive Services is an image stream. The data collected in the previous step is an mp4 file. The conversion of mp4 to image stream is required.

Video processing

This step involves conversion of mp4 to image stream in order to perform emotional analysis using Face API. This conversion is implemented by extracting screenshots from the video clip. The parameter required is the time slot after which the frame is extracted. The rate of

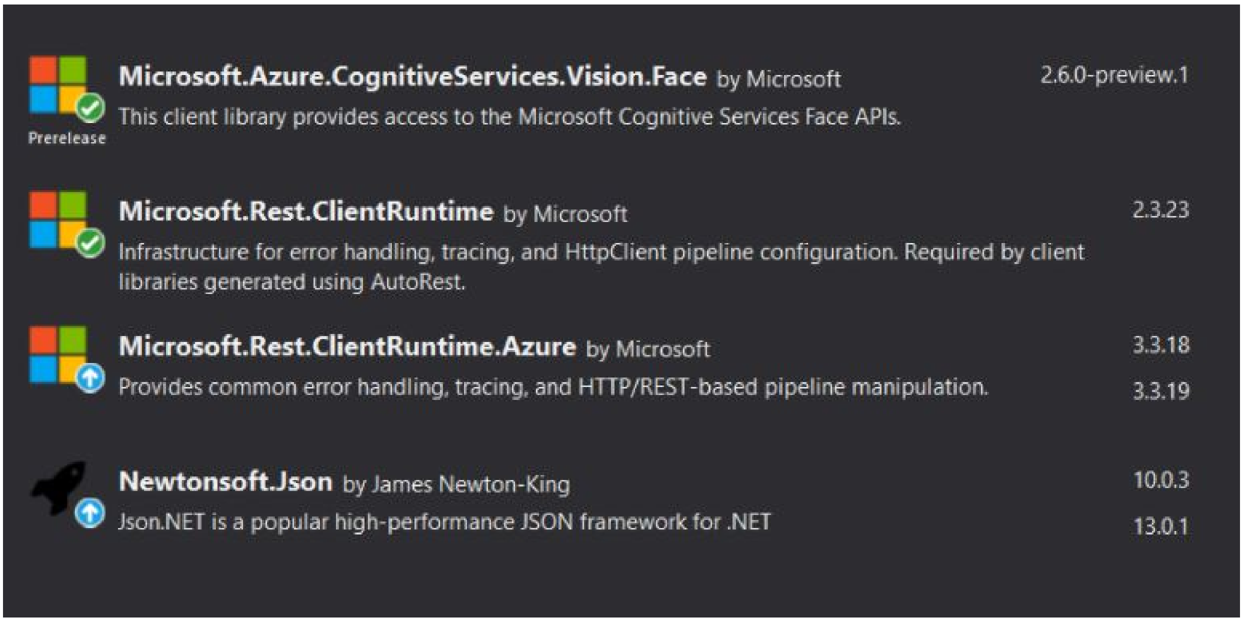


Figure 4.3: NuGet Packages provided by Microsoft used to run Face API

```
using Microsoft.Azure.CognitiveServices.Vision.Face;
using Microsoft.Azure.CognitiveServices.Vision.Face.Models;
```

Figure 4.4: Cognitive Services provided by Microsoft used to run Face API


```
[
  {
    "faceId":"8at67t5e-45654-432f-3456-ty76f002151",
    "faceRectangle":{
      "top":231,
      "left":123,
      "width":333,
      "height":333
    },
    "faceAttributes":{
      "emotion":{
        "anger":0.5,
        "contempt":0.2,
        "disgust":0.3,
        "fear":0.0,
        "happiness":0.0,
        "neutral":0.0,
        "sadness":0.0,
        "surprise":0.0
      }
    }
  }
]
```


Figure 4.5: Face Object returned by API

extraction can be customized. Smaller time slot will build a bigger image stream with greater number of frames present in it. Extracted image stream is passed to FaceAPI for emotion analysis. The c Xabe.FFmpeg library allows you to convert video files into required formats. It allows flexibility to manually set parameters to any input or output stream.

At timestamp ‘t’, there is an emotion estimation for the single frame. The results are normalised values providing the mutually exclusive probability of that emotional state being present. Graphical representation of the results is shown in 4.6

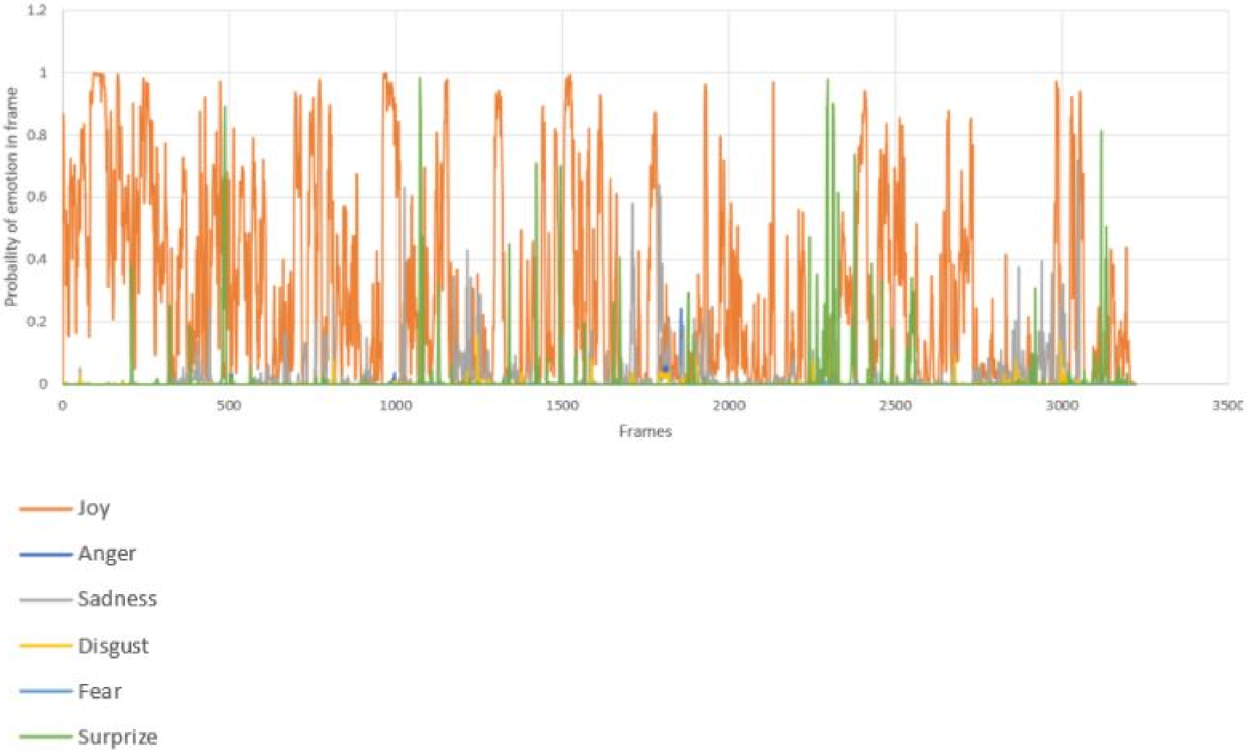


Figure 4.6: Graphical Representation of output probabilities of data

The output of the data after processing it in excel is shown in 4.7.
The probabilities are mutually exclusive, thus the summation of the values is equal to one.

Probability vector <anger, contempt, disgust, fear, happiness, neutral, sadness, surprise>
With the frame rate of 40 ms, the total numbers of frames analysed by Face API are 3221. Lower frame rate contributes to attain greater precision. Time involved to run a single simulation is considerably high. This results in a trade off between precision and time required to complete one simulation.

Critical Analysis

The emotion analyses obtained from the simulations are mutually exclusive probabilities of emotional states present in the frame. There are some strong probabilities and some weak probabilities present in the dataset. Paper “Compound facial expressions of emotion by Shichuan Du, Yong Tao, and Aleix M. Martinez ” discussed the compound categories of emotional states such as Happily surprised, Happily disgusted etc, discussed in chapter 2. The Face API is unable to detect compound categories of emotions.

Frame #	anger	contempt	disgust	fear	happiness	neutral	sadness	surprise
0	0	0.001	0	0	0.791	0.208	0	0
1	0	0	0	0	0.868	0.131	0.001	0
2	0	0.001	0	0	0.836	0.163	0	0
3	0	0.001	0	0	0.697	0.302	0.001	0
4	0	0.001	0	0	0.519	0.479	0	0
5	0	0.001	0	0	0.54	0.459	0	0
6	0	0.001	0	0	0.532	0.466	0	0
7	0	0.001	0	0	0.543	0.456	0	0
8	0	0.003	0	0	0.558	0.437	0.001	0
9	0	0.003	0	0	0.317	0.675	0.005	0
10	0	0.005	0.001	0	0.387	0.602	0.005	0
11	0	0.004	0.001	0	0.344	0.647	0.004	0
12	0	0.002	0	0	0.505	0.491	0.001	0
13	0	0.008	0	0	0.511	0.48	0.001	0
14	0	0.015	0	0	0.157	0.826	0.001	0
15	0	0.02	0	0	0.157	0.822	0.001	0

Figure 4.7: Data processed in excel

The time involved to run a single simulation is very high as discussed earlier. Running a simulation on 4 mins video costs upto 5-6 hours. Changing parameters such as frame rate is a very time-consuming job, thus experimenting with different parameters might take 2-3 days. This causes timing constraints to run simulation on a number of videos.

4.4.2 Emotient FACET

The second technology used for emotion analysis is Emotient. Emotient is a facial expression analysis engine built on the principles of a facial action coding system. The method includes processing an image to identify a face in an image. It can identify one or more windows in an image. Single or multiple pixels and color intensities can be quantified in the windows in order to derive one or more 2- dimensional intensity distributions of colors within the windows. The intensity distributions can be processed to select image features appearing in the one or more windows and to classify predefined facial actions on the face in an image. A facial action code score includes a value that indicates a relative amount of predefined facial action occurring on the face in an image.

There are various steps involved in facial expression analysis with this technology. The first step involves Face Detection, by applying Viola Jones' object detection algorithm. It detects the position of face in input video. This algorithm uses Harr-like features that look at adjacent rectangular regions in the given frame in order to compare the intensities of pixels in those regions. Human facial features shows variations of intensities throughout the face structure, such as there is a prominent difference between the intensity of pixels between cheek and the eye region. The difference of intensities of pixels is the key to detect face using harr classifier. Classifier runs throughout the frame and detects the human face. Emotient detects the face in each frame and the facial region is enclosed in a green box.

The next step involves facial feature detection, such as lips, nose eyes etc by segmentation of the face window for each of the features. As stated by Littlewort, the facial feature detector gives the log likelihood ratio of that feature being present at the location in the face window to not being present at that location. The probability of features present at the given pixels is estimated, and the resultant features are filtered out and concatenated into feature vectors.

After extracting feature vectors from a frame, they are fed as input to machine learning algorithms to identify different Action Units. The machine learning algorithm generally used in this process is Support Vector Machine. The classification derives the result as intensities of the action units prominent on the face.

Training dataset:

The dataset used in the backend for training the system is high quality supervised databases of facial expressions. The computation of feature vectors uses a support vector machine (SVM) for each AU. The action units SVM were trained from a compilation of various databases. Some of the known databases are Cohn-Kanade , Ekman-Hager, M3 , Man-Machine Interaction (MMI), and two nonpublic datasets collected by the United States government which are similar in nature to M3. These datasets are the mixture of posed and spontaneous facial expressions. Posed expressions were facial expressions of well trained actors, who were videotaped with a script to perform transformations from happy to surprised etc. Estimation of AU is succeeded by supervised learning to derive emotion status and emotion valence.

The values of AU intensities in a frame are concatenated to build an Action unit vector. Emotient encodes 19 facial action units and 6 different prototypical emotion states. Compound facial expressions such as happily surprised, happily disgusted etc can be derived from the AU intensities or derived emotion intensities.

Emotient's AU vector comprises intensities of the following vector units.
AU:(1,2,4,5,6,9,10,12,14,15,17,18,20,23,24,25,26,28,43)

For each Action Unit, Emotient outputs a continuous value corresponding to each frame.The frame-by frame intensities of AU provide information on the dynamics of the facial emotional state at temporal resolution. The output of Emotient is the intensity of Action Units along with other modules such as basic emotion recognizer, pitch degree, roll degree and yaw degree.We are mainly interested in emotion and AU evidence values.Other modules are helpful for the study of gestures, body postures and other multimodal analysis.

Emotient analysis categorizes the following six basic emotional states along with three emotion valence states. EmotionState: Joy, Anger, Surprise, Fear, Disgust, Contempt

$$EmotState\,(i,t) = \alpha_1 AU(1,t) + \alpha_2 AU(2,t) + \cdots \ldots \quad + \alpha_n AU(n,t) + \varepsilon\,(t),$$
where, i=1 → Joy, i=2 → Anger.........; n=1, N; N is the total number of action units, obtain α by multivariate analysis and statistical learning. Note estimated error, $\varepsilon(t)$.

Figure 4.8: Equation of Emotion States

Evaluation of emotion state is the second layer recognition system in FACS. An emotion state is directly derived from the multivariate equation of Action Units shown in 4.8. This equation is the result of machine learning algorithms applied to the training dataset. Weights (1, 2) are calculated by machine learning algorithms in order to minimize the error term in the equation. The machine algorithm used for this second layer recognition system is multivariate logistic regression (MLR classifier). This classifier is trained on AU intensities, on the Cohn-Kanade dataset and corresponding ground truth emotion labels. MLR outputs the posterior probability of every emotional state given the AU intensities as input. EmotValence: Positive, Negative, Neutral

$$EmotValence\,(j,t) = \beta_1 EmotState\,(1,t) + \ \ldots + \beta_3 EmotState\,(3,t) + \theta\,(t)$$, j=1 Positive Valence, j=2, Neutral Valence, j=3 → Negative valence; and $\theta(t)$ is the error.

Figure 4.9: Equation of Emotion Valance

Similarly, emotion valance is derived from the multivariate equation of Emotion states shown in 4.9. The equation EmotValance(j,t) is formed by a multivariate logistic regression (MLR classifier). Weights (ß1, ß2) are assigned the values which minimizes the error term (t). MLR outputs the emotion valence given the emotional states as input.

Simulation:

Emotient, in collaboration with iMotions, provides a Graphical User Interface for processing videos to perform facial expression recognition. In this study the latest version of Emotient is used (FACET 2.1 SDK). This technology includes robustness to extraneous factors, such as image resolution, ethnicity, head pose, change of illumination. Head pose estimates are also provided by Emotient such as yaw, pitch, and roll. the system is highly extensible and runs on Windows and Linux applications with a small memory footprint. Pre pre-processed mp4 file is fed to the system along with some customized parameters such as age, gender, etc. Emotient runs the simulation on the video and returns the csv file as output.

In 4.11The evidence is presented in the form of 10 base log likelihood of that emotion being present in the frame.

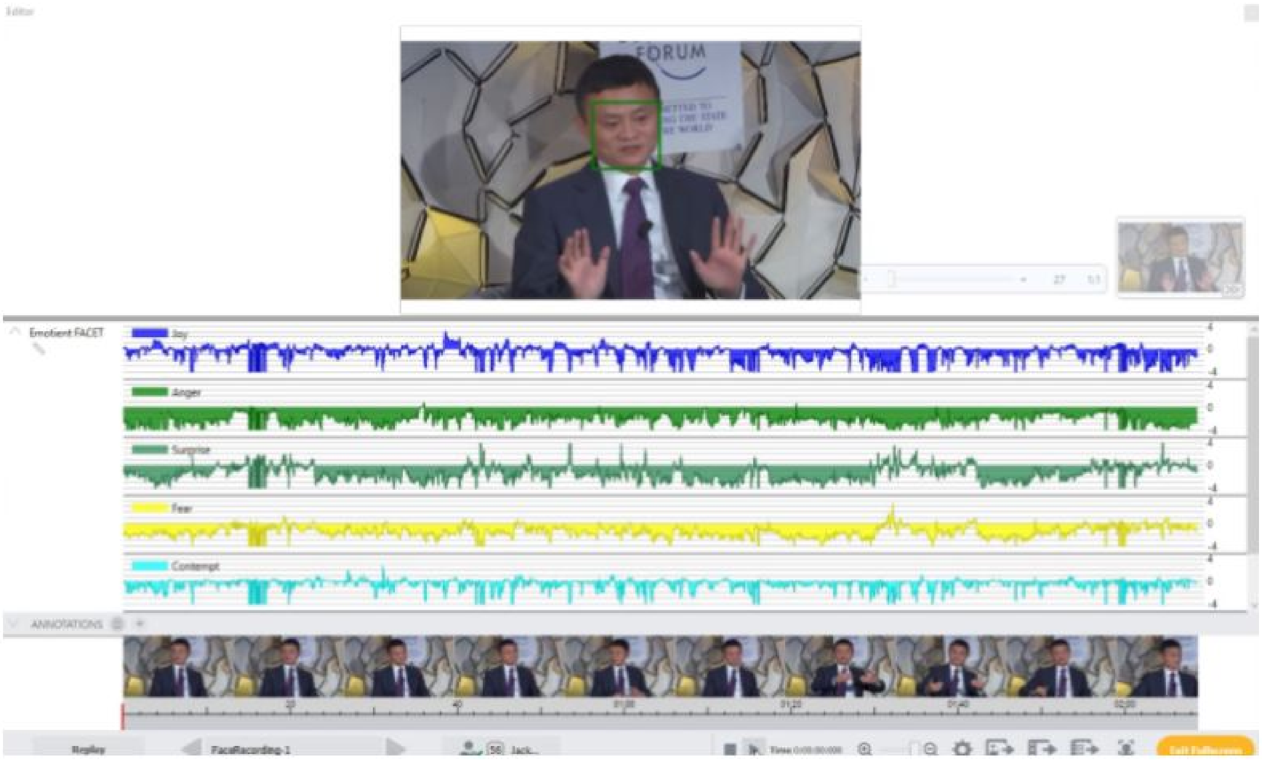


Figure 4.10: Simulation on Emotient

FrameNo	FrameTime	Joy Evidence	Anger Evidence	Surprise Evidence	Fear Evidence	Contempt Evidence	Disgust Evidence	Sadness Evidence
0	0	0.2476785	-1.075323	-0.6716394	-2.017493	0.09800221	-0.1927556	-0.9155064
1	33	0.22509	-0.9625021	-0.7799395	-2.01705	0.05986992	0.03842982	-0.7048084
2	67	0.01884632	-1.133253	-0.7936857	-1.662067	0.0846997	-0.4263833	-0.4552711
3	100	-1.057688	-1.884082	-0.6443361	-1.284373	-0.4405059	-1.1396	-0.633289
4	133	-0.5692036	-2.131799	-1.187184	-0.9201266	-0.4508249	-0.6123869	-0.4772142
5	167	0.4743144	-2.004916	-1.223915	-0.8801628	-0.4959382	-0.1234544	-0.8733523
6	200	0.489724	-2.026604	-1.18845	-0.9347656	-0.4792297	-0.09706357	-0.8951515
7	233	0.788947	-2.383477	-1.488301	-0.4335561	-0.5546108	-0.2211885	-1.31464
8	267	0.8980852	-2.742784	-1.79846	0.07271039	-0.8020142	-0.8278192	-1.835962
9	300	1.022964	-2.980765	-1.956487	0.2919402	-1.13247	-1.431405	-1.949659
10	333	1.083553	-3.317501	-2.025826	0.2368882	-1.005904	-1.749278	-1.931811

Figure 4.11: Evidence values for emotion states in a continues frames

4.4.3 Affectiva/ AFFDEX

The third software used for the analysis is Affectiva. This is an emotion recognition tool which analyse emotions based on affect. This method is disclosed for affect-based recommendations which comprises analysis of action units, gestures and mental states, along with analysis of physiological data and eye tracking. The raw video is fed to the system and is processed to obtain information about facial data, actio units,gestures, mental states. The information in facial data comprises smiles, head gestures, raised eyebrows, lowered eyebrows etc. The patent published by Affectiva claims that physiological data can be analysed along with an eye tracker. The mental state information processed by the complicated technology gives the evidence values for basic six emotions.

Anger	Sadness	Disgust	Joy	Surprise	Fear	Contempt
0.001923581	0.01190979	0.3599295	0.005373423	0.2498134	0.003120215	0.1296756
0.000830227	0.006183683	0.1639708	0.03201551	0.2326228	0.001481867	0.05561689
0.000855451	0.007549683	0.1752018	0.02306228	0.2180138	0.001665919	0.06309728
0.000910543	0.00832122	0.1876462	0.01877732	0.2152994	0.001802937	0.06899495
0.000772776	0.006771928	0.1583057	0.02991001	0.2186019	0.001499763	0.05599025
0.000812031	0.008020355	0.1705709	0.02213394	0.2090264	0.001673352	0.06329637
0.000770113	0.00761281	0.1619588	0.02517622	0.2088882	0.001588579	0.05970468
0.000932123	0.009484477	0.1965035	0.01509049	0.206845	0.001947679	0.07522814
0.001425331	0.01567885	0.302969	0.004733172	0.2012533	0.003088079	0.126629
0.00155316	0.01667725	0.3278489	0.003975329	0.2041725	0.003295557	0.1374629
0.001687896	0.01755411	0.353221	0.003400584	0.2087609	0.003464892	0.1481488
0.001776786	0.01818568	0.370129	0.003070952	0.2141074	0.003503428	0.1554906
0.001870076	0.01673204	0.3779288	0.003314181	0.2304189	0.003316922	0.1526674
0.001910471	0.01588413	0.3798575	0.003511084	0.2452117	0.003081901	0.1500438
0.001955313	0.01537696	0.3841962	0.003606257	0.2574303	0.002927854	0.1492357
0.002042326	0.01491294	0.4052801	0.003623224	0.276583	0.002736963	0.1502758
0.002060103	0.01573477	0.4653528	0.003316545	0.2873792	0.002573687	0.1555878
0.002171352	0.01558434	0.8631797	0.003199011	0.3221689	0.002255977	0.159372

Figure 4.12: Evidence values for emotion states in a continues frames by Affectiva

4.5 Data Analysis

The retrieved outputs from the software should be processed in order to retrieve results from the data. The stats and analysis is performed on the data so that conclusions and evaluations can be constructed. CSV data is imported to Excel data. Excel provides all the tools and functionalities that are required to perform data analytics on the output values. The data analysis is performed using Macros in Microsoft Excel.

The inherited data from the software simulations are accompanied with unknown error values corresponding to every value. In the previous section we discussed that emotions are derived from the resultant vector of action units. The second order derivation of action units gives rise to emotion evidence, increasing the effect of error value on the evidence of emotion present in the frame. In facial actions encoding systems, action units are classified through machine learning algorithms. Machine learning algorithms by nature are designed to improve their accuracy by reducing the error value. Minimising errors does not mean eliminating error, and studies show that FACS can estimate action units upto the accuracy of 95

The dataset used for the training comprises all the gender, ethnicity and age variations. Cohn Kanade dataset ethnicity variations comprise 80 perecent of Euro-American, 13

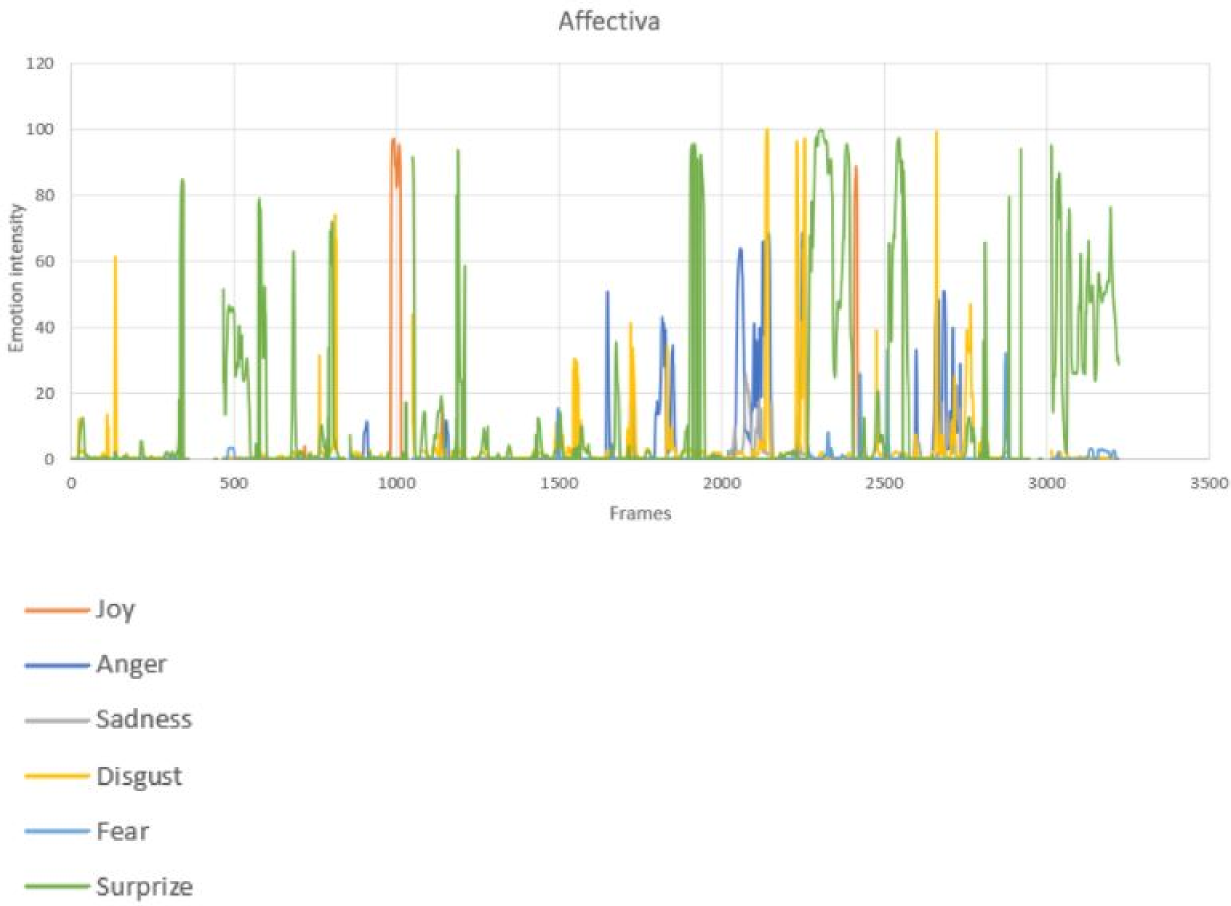


Figure 4.13: Graphical view of data processed by Affectiva

percent Afro-American and 6 percent other groups. The dataset does not have equal membership for all ethnic groups. It implies that syeten would work best for the Euro-Americans. Other ethnic groups will have lower accuracy in the result. When all the sources of errors are taken into account, this has a significant effect on the accuracy of the emotion analysis performed by these simulations. In order to normalise the data to minimise the effect of errors, fuzzy logic comes into the picture.

4.5.1 Fuzzy Logic

Fuzzy logic is defined as logic of vague concepts, and fuzzy control provides a formal methodology for representing, manipulating and implementing a human's heuristic knowledge about how to control a system. Heuristic in the study is the information based on mathematical analysis of the output results of the simulation. Fuzzy sets are the sets whose elements have degrees of membership. Fuzzy set theory permits the gradual assessment of membership of an element in a set. This is described with the aid of a membership function valued in the real unit interval [0,1]. The emotional state in the output results are in the form of emotional evidence present in the frame. As we discussed in the previous section the evidence of emotional state is the ten base log likelihood value, i.e. if the evidence value for a certain emotion is 1, then we say that there is 10 times greater likelihood of that emotion being present. Similarly, if the evidence is -1, then we can say that there is 10 times greater likelihood of the emotion not being present. Thus the range of intensities of the emotions is very large. For an emotional state of joy, the minimum evidence recorded is -7.522848, thus the intensity of the emotion would be 10 to the power of 7.522848, which is 3.00021E-08 and the maximum evidence is 2.1000, with the intensity 125.89. Thus the range of the Joy intensity in the dataset is (3.00021E-08,125.89). Fuzzy membership function assigns the new membership values to the existing values within the range of [0,1]. Linear membership function is used to evaluate membership values for an emotion state. There is a requirement of two heuristics to evaluate the function. Lower threshold value determines the lowest value of the emotion intensity to be considered for membership function similarly, upper threshold determines the largest value to be considered for membership function. All other values below the lower threshold would be considered zero and all the values greater than the upper threshold would be one. Rest of the membership values will lie between [0,1].

Linear equation:

$$\mathrm{y} = mx + c \tag{1}$$

$$\mathrm{u}(Joy) = u(Joy) + c \tag{2}$$

If 0.01 is chosen as lower threshold and 2 is chosen as upper threshold (because most of the values lie in between).

$$1 = 5m + c||0 = 0.01m + c \tag{3}$$

Solving the two equations we get the resultant equation which is the membership function of the emotion state. Overall membership function:

$$\mathrm{u}(x) = x < 0.01 : 0, x >= 2 : 1, (x * 0.50263) - 0.00506 \tag{4}$$

This membership function yields membership values for the emotion states. This minimizes the effects of error values that exist along with emotional intensity. We neglect the very small evidence(No membership) and compress the evidence with very high values(Full membership).

Fuzzy Average/ Defuzzification :

After processing the data to assign membership values to the emotion intensities, the fuzzy value produced at the composition stage needs to be converted to a single value or a crisp value. This computation takes into account the effect of each emotion intensity in a

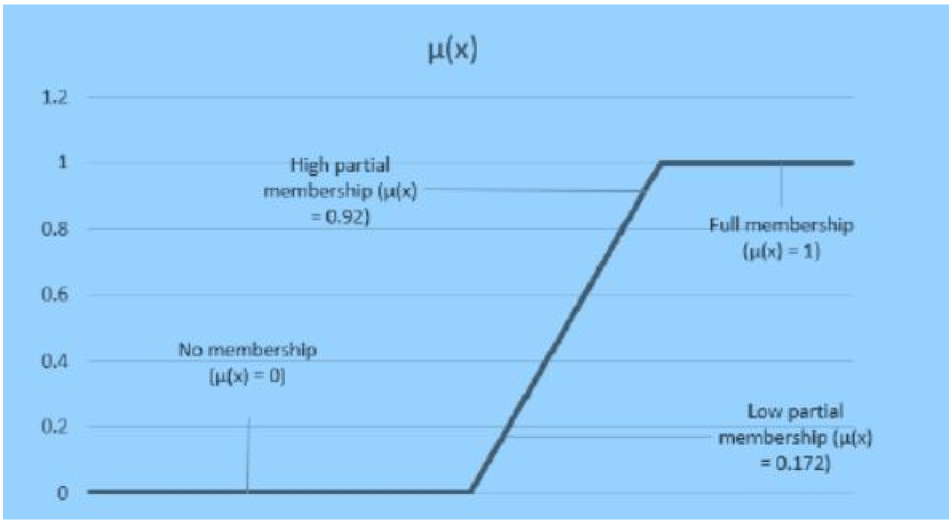


Figure 4.14: Membership function

proportionate manner. This is necessary because it is important to take only into account the effect of those datasets that have maximum effect. It almost eliminates the effect of error values in emotion intensities present in a set of frames.The crisp value is assigned to all the frames present in the set as shown in 4.15.

$$CrispValue = \frac{sum(u(emotion) * emotion)}{sum(u(emotion))} \tag{5}$$

u(surprize)	u(surprize)*surpriz	Σμ(surprize)*surprize	Σμ(surprize)	Crisp Value
0	0			0.205081231
0	0			0.205081231
0	0			0.205081231
0.010810273	0.000341329			0.205081231
0.106843715	0.023787296			0.205081231
0.011296011	0.000367582	0.117653989	0.024128625	0.205081231
0	0			0.032540858
0	0			0.032540858
0	0			0.032540858
0	0			0.032540858
0	0	0.011296011	0.000367582	0.032540858
0	0			0
0	0			0
0	0			0
0	0			0
0	0	0	0	0
0	0			0.013106116
0	0			0.013106116
0.001527527	2.00199E-05			0.013106116
0	0			0.013106116
0	0	0.001527527	2.00199E-05	0.013106116

Figure 4.15: Crisp value assigned to set of frames

5 Results and Evaluation

5.1 A Comparative Analysis of three software

In the beginning of the project we had a set of software technologies/ APIs to perform emotional analysis on the human face. Three technologies that we discussed in the previous section are Emotient, Microsoft Face API, and Affectiva. In this section we are going to compare the results produced by three technologies.

The test video used for comparison is fed to three softwares, at the frame rate of 40ms to provide standardized units for comparison. The video used for simulation is high quality video with sufficient amount of lightning on the face.

5.1.1 Face Detection Rate

The first evaluation factor is the face detection rate of the technologies.

Microsoft Face API performs the best in terms of Face Detection Success Rates, while affectiva provides minimum data for analysis.

5.1.2 Performance analysis

All the softwares provides six universal emotion values. As we already discussed in the previous section, Microsoft Azure Face API's output result provides the set of probabilities of an emotion present in the frame. The probabilities are mutually exclusive, thus failing to analyse compound emotions. On the other hand Emotient produces log likelihood evidence of emotional state present in a frame and Affectiva provides the intensities of the emotion present in the frame.

Deep analysis of the video is performed to spot some extreme emotional states throughout the video.

The experiment performed to analyse the performance of the technology involves deep manual analysis of the experiment video in order to compare them against human judgement of emotions. For each emotion 5-6 random moments are chosen from the video and are

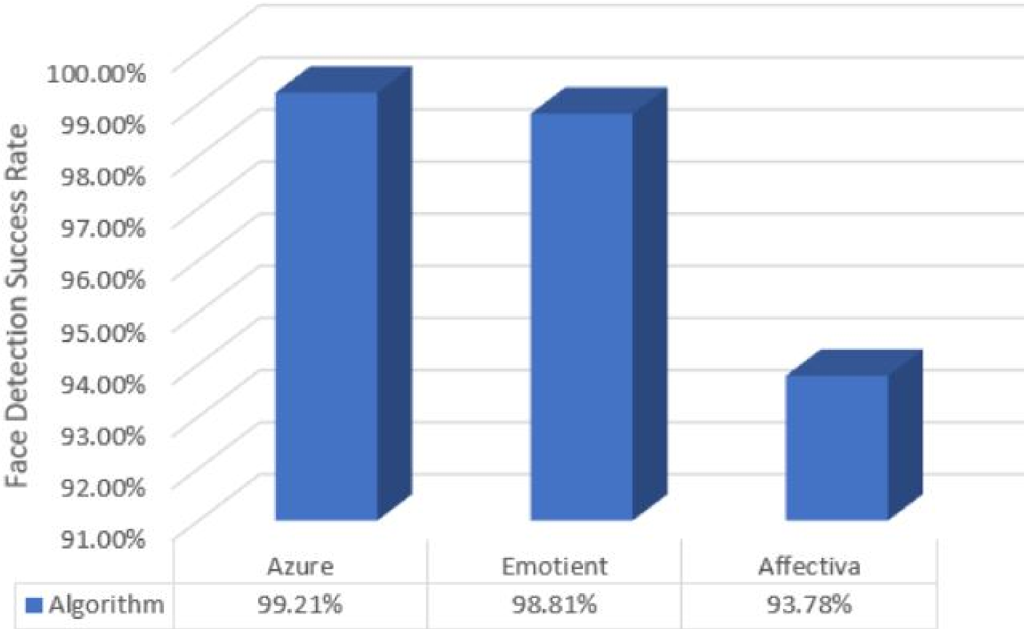


Figure 5.1: Face detection Success Rate

Figure 5.2: Some of the facial expression accessed by human judgement in the experiment video <Sad, Joy, Surprise, Disgust, Contempt and Fear>

matched to the corresponding set of frames with respect to time. One moment consists of 8-10 corresponding frames. If software is able to detect that emotion in that set of frames, it is considered a hit, else a miss.

Six random joyful moments of the subject were spotted in the video manually to run a test on the output of these softwares. The time slots of the moments were matched to the set of corresponding frames (8-10) in which they were captured.

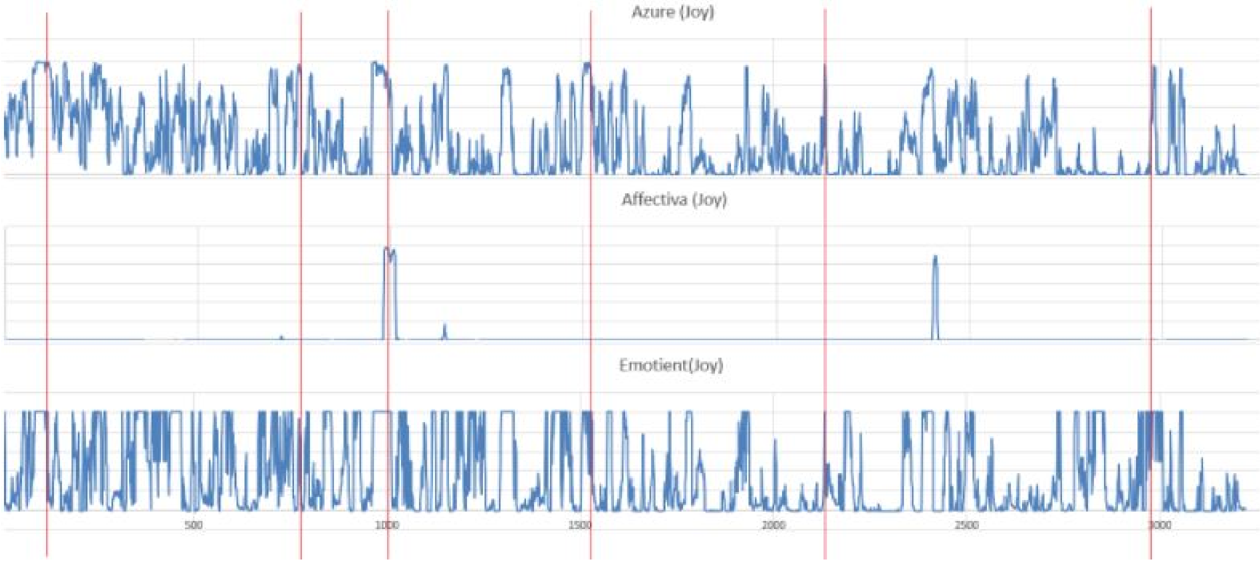


Figure 5.3: Time series plots of Joy detection for each of the software plotted against their relative scale units of emotion detection values. Frame set highlighted in red.

As the 5.4 figure depicts, Affectiva seems to miss out on maximum frame sets, whereas Microsoft Azure API and Emotient were able to detect the emotion by representing peaks/ high values at that moment.

Performing a similar test on surprise emotion.

Similarly, by performing these tests on all the emotions we analyze that accuracy of softwares differs for distinct emotions. According to the above test performed on the different technologies, Emotient performs the best whereas Affectiva's performance is not upto the mark with others. Emotient has the highest hit ratio for most of the frames subjected to analysis. This proves that FACET can perform upto the mark when compared with human judgement. It is further able to categorize compound emotion states. For eg. If a frame has high intensity values for happy along with surprise value, it can be interpreted as happily surprised as discussed in section 2. This shows that frames can predict the high evidence of action units' moments that give rise to these two emotions. From the output results, we analyse that Microsoft Azure API and Emotient results contain consistent peaks with respect to frames along the x axis, whereas Affectiva's results are not consistent with other softwares.

5.5 depicts the weighted average of the emotion values analysed by the set of technologies- Microsoft Azure Face API, FACET SDK 2.1 Emotient, AFFDEX Affectiva. Affectiva fails to determine joy, whereas over estimates the surprise emotion. Due to mutually exclusive value

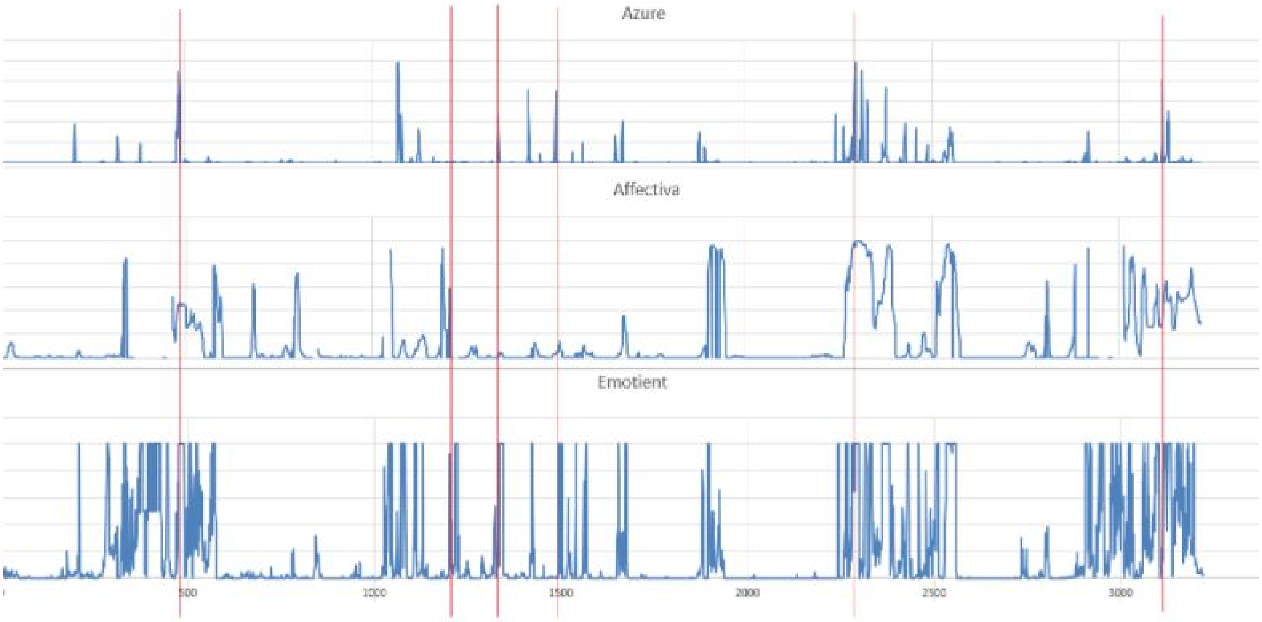


Figure 5.4: Time series plots of surprize detection for each of the software plotted against their relative scale units of emotion detection values. Frame set highlighted in red.

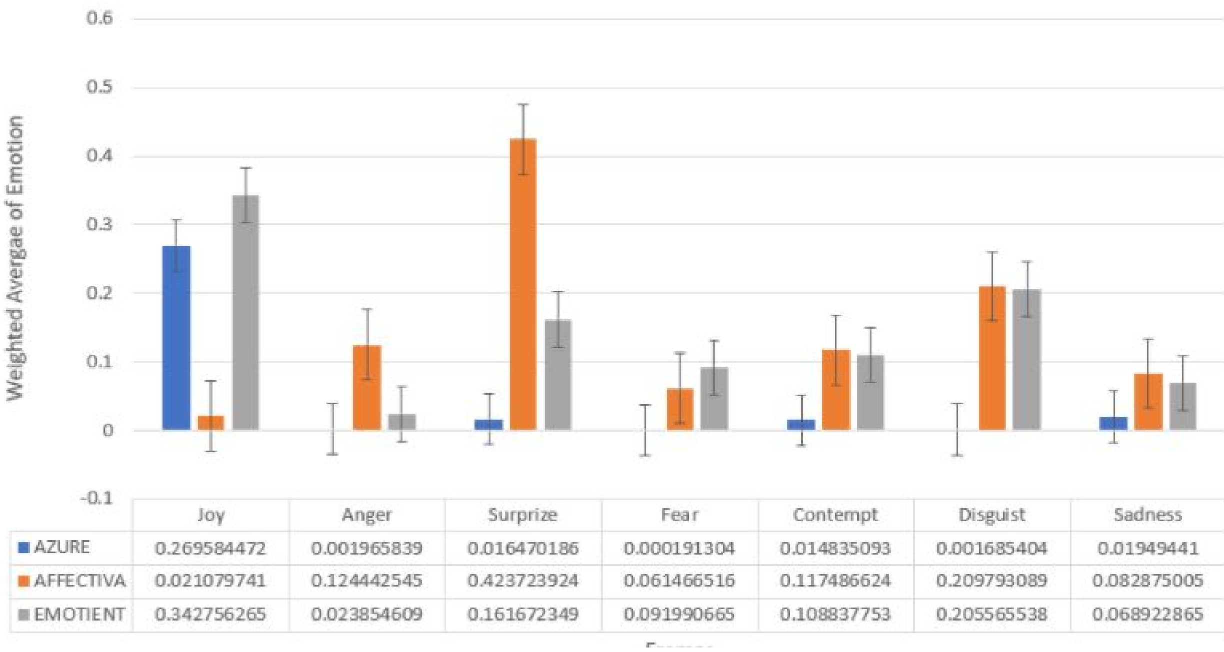


Figure 5.5: Weighted Average Emotion on relative scale

vector estimation of the Azure API, other emotions present in the frames were neglected. Emotient provides evidence of every emotion extracted from the frame, and the values are quite reasonable when compared to a human's judgement. The accuracy of the analysis is dependent on the emotion state.

	Emotient	Azure	Affectiva
Joy	6/6	6/6	1/6
Anger	3/3	1/3	3/3
Surprise	6/6	4/6	4/6
Fear	2/2	1/2	1/2
Contempt	3/3	3/3	3/3
Disgust	3/4	1/4	2/4
Sadness	4/5	4/5	3/5

Figure 5.6: Weighted Average Emotion on relative scale

This test 5.6 reveals that Emotient has the potential to measure basic and compound emotions that are expressed by human faces. FACET performs better with natural as well as prototypical facial expressions and performs with different accuracy in extracting different emotions. Thus, for conducting the emotional analysis on South Asian female politicians we will be using Emotient.

5.2 South Asian Female Emotion Analysis

The Politicians selected for emotional analysis are some of the leading ladies in their national political parties. They are supposed to be a trained group of people, representing themselves in the society to create a mass influence and win their positions in parliament. We tried to analyse their emotional states in different situations, so that more information can be extracted to study their behaviour. As we discussed in section 4.2, 1:1 Interview is considered the lowest critical situation, Public speaking the moderate one and Parliament session has the highest mark on the critical scale.

Below are the results of emotional analysis performed in South Asian Politicians.

As per the results displayed in above figures it is clear that all women represent themselves very differently, subjected to different situations. Contempt among all the women is a prominent emotion and expressed. All women do show this emotion in almost each situation they are subjected to. Joy evidence varies with respect to the situations as per the above studies. These variations in the expressions with respect to the situation highlight many personality traits to be followed up by behavioral sciences. These politicians are trained to control their emotions according to the situation in order to create maximum influence.

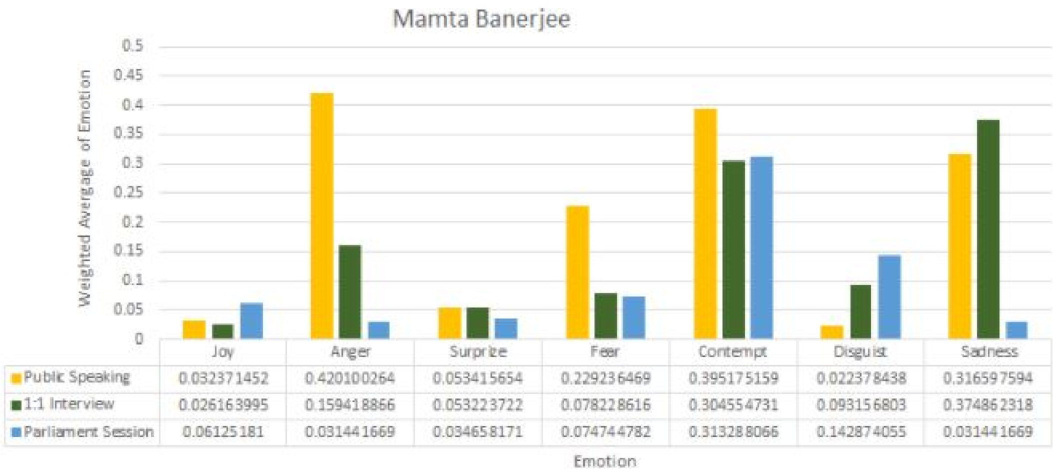


Figure 5.7: Weighted Average of Emotions at different situations

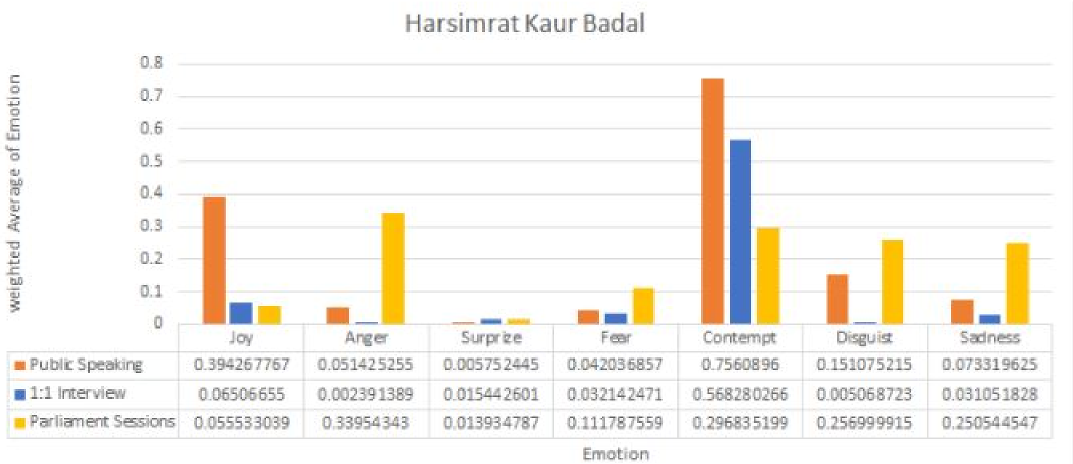


Figure 5.8: Weighted Average of Emotions at different situations

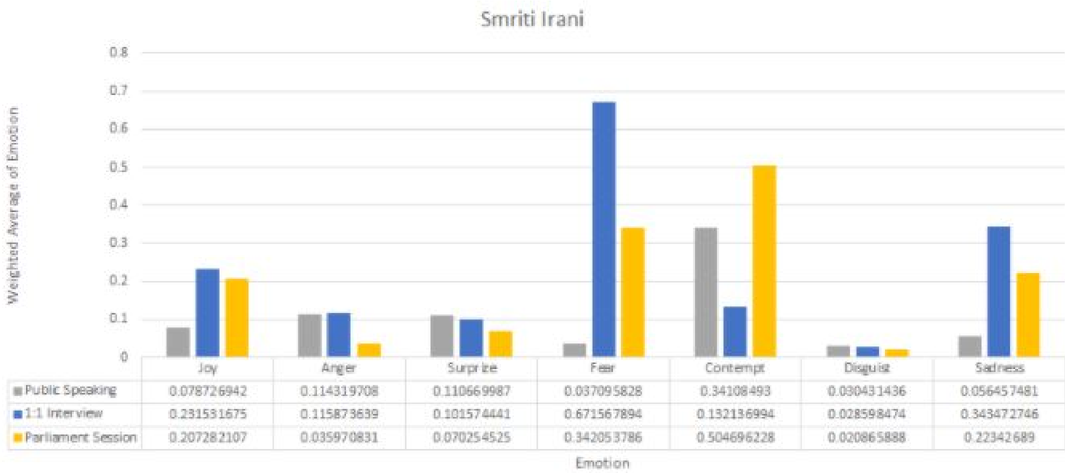


Figure 5.9: Weighted Average of Emotions at different situations

This is evident from the results that women either have strong joy emotion or strong anger emotion over their sessions. Some of the politicians show happiness in order to create a

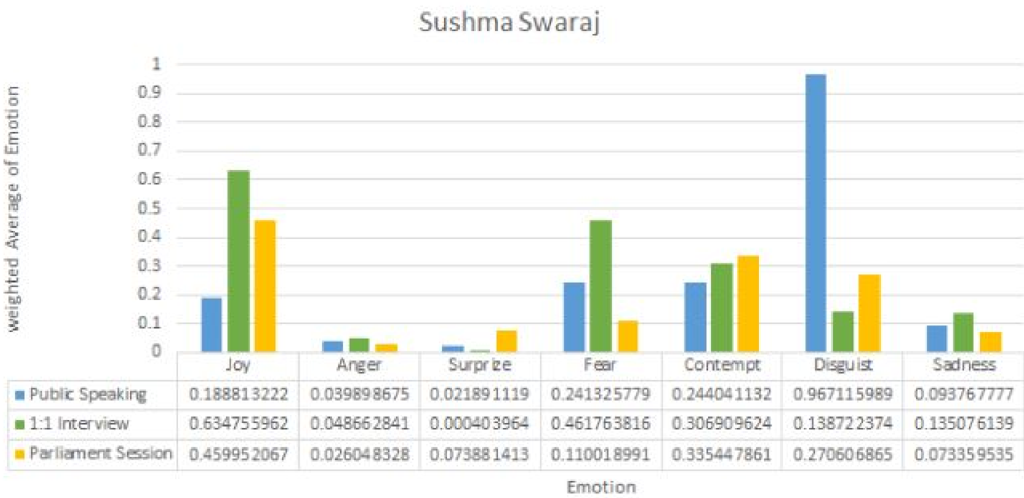


Figure 5.10: Weighted Average of Emotions at different situations

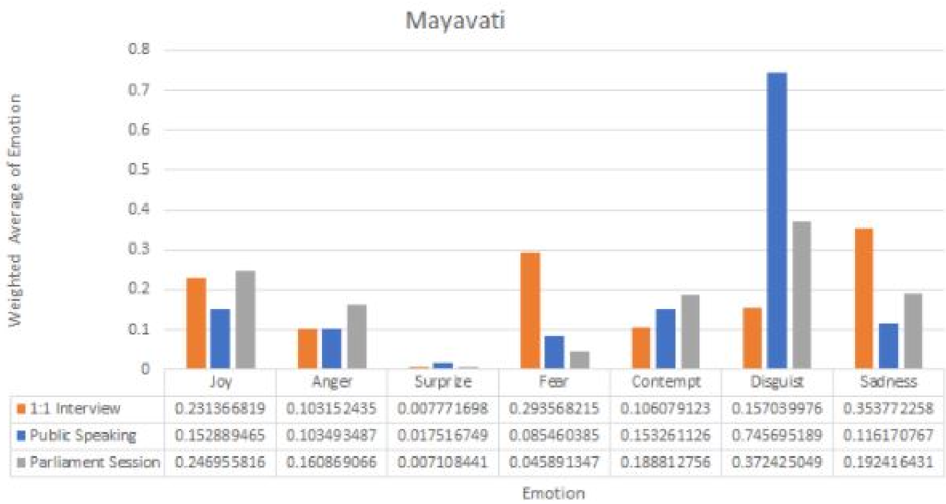
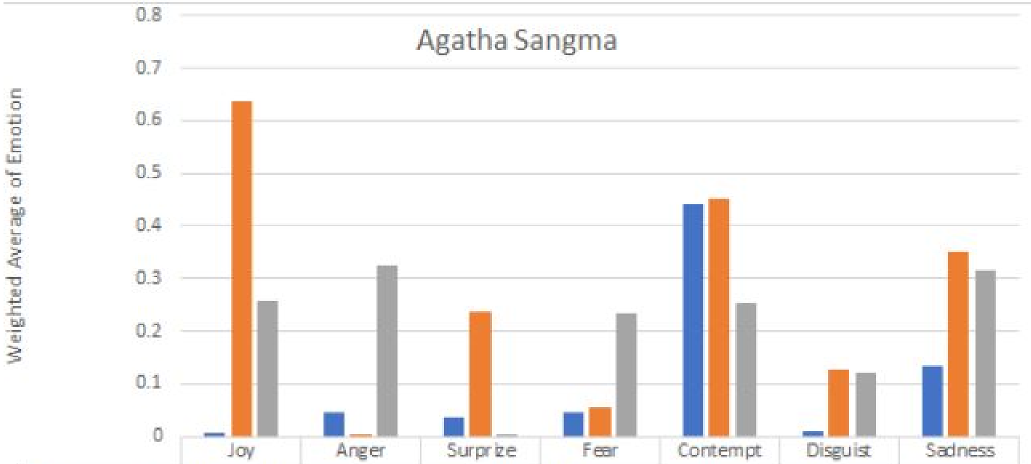


Figure 5.11: Weighted Average of Emotions at different situations

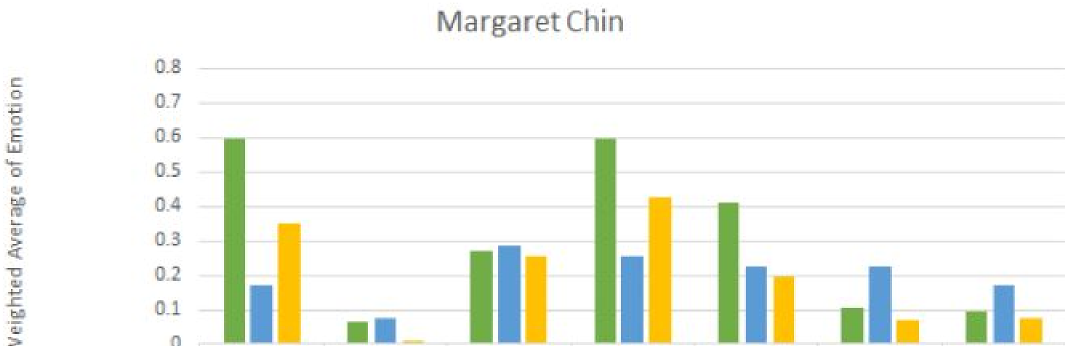
strong and confident image while there are some who show anger to show their superiority. Most of the leading female ladies, Elaino Chao, Arzu Rana, Dil Kumari Bhnadari, Margaret Chin, Mayavati, Sushma Swaraj, Smriti Irani, generally represent joy in all of their sessions to different extents. Mamta Banerjee shows anger in her public speaking and interview sessions whereas she tends to be happy in parliament sessions. On the other hand, Harsimrat Kaur Badal and Agastha Sangma show happiness in interview sessions and public speaking, but higher anger levels in parliament sessions. According to the studies, women empathise with the audience, showing a sad facial expression. Higher values of sadness can be observed in the sessions of some leading ladies. A lot more can be revealed from the emotional expression of these politicians through the study of psychology and behavioral sciences.

It can be observed from the results that women represent themselves differently in various levels of critical situations. Most women show greater happiness levels in 1:1 interviews and



	Joy	Anger	Surprize	Fear	Contempt	Disguist	Sadness
Parliament	0.008178372	0.044953686	0.035780719	0.044543074	0.440594817	0.011263296	0.133335978
1:1 Interview	0.635463	0.00125	0.236585	0.05641	0.45124	0.125874	0.35234
Public Speaking	0.25647	0.32578	0.001245	0.232561	0.25484	0.121451	0.31452

Figure 5.12: Weighted Average of Emotions at different situations

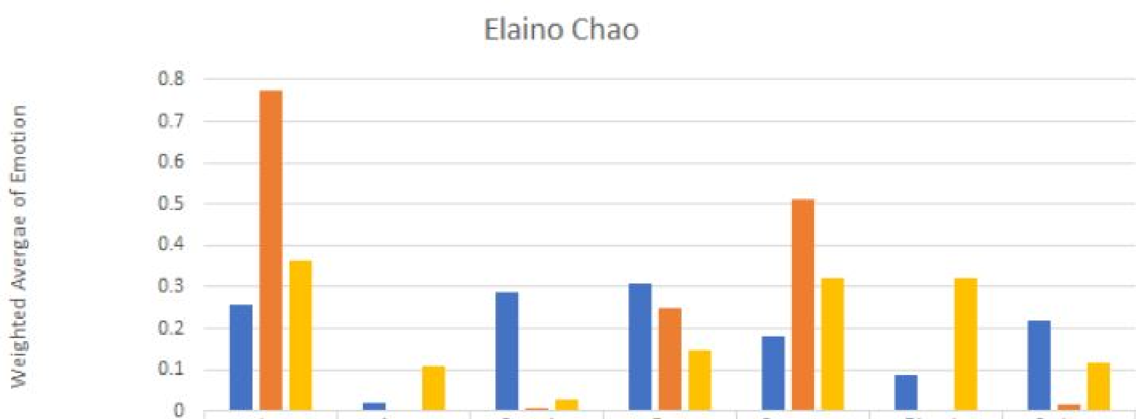


	Joy	Anger	Surprize	Fear	Contempt	Disguist	Sadness
Public Speaking	0.597732887	0.065209017	0.270922935	0.597732887	0.409970694	0.104911111	0.094035904
Parliament Session	0.169517289	0.073394804	0.283852523	0.254192901	0.223727572	0.224664599	0.170684259
1:1 Interview	0.352900836	0.011680206	0.257136456	0.425667993	0.19541634	0.068555517	0.077104193

Emotion

Figure 5.13: Weighted Average of Emotions at different situations

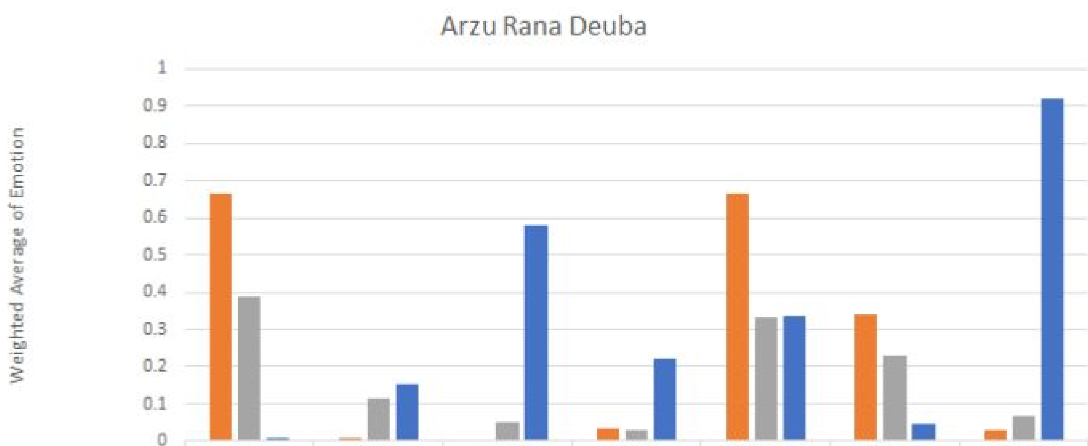
public speaking sessions than the parliament sessions. Parliament sessions are well prepared performances by the politicians hence a great balance of emotions can be seen in the results. Mamta Banerjee shows high levels of anger in her public speeches, but a very low trace of anger can be seen in her parliament sessions. Arzu Rana Deuba shows great levels of joy in 1:1 interviews, but in parliament sessions her joy disappears, showing high levels of sadness. Aagastha,Harsimrat Badal, Margreat Chin, and Elaino Chao, show similar behaviour. Whereas Sushma Swaraj recorded the highest value of joy in her parliament session, with very low levels of sadness. Sushma Swaraj is considered one of the greatest parliament speakers and has been addressed as a “daring voice of Indian pride” . Her speeches as Foreign affair minister at parliament sessions are one of the most powerful speeches recorded in Indian Parliament History. She also represented India at various International occasions.



	Joy	Anger	Surprize	Fear	Contempt	Disguist	Sadness
Parliament Session	0.258326684	0.01977928	0.28636003	0.307968043	0.179147582	0.087696497	0.219096687
1:1 Interview	0.774285306	0	0.008807661	0.249125266	0.513093324	0.000654538	0.016857819
Public Speaking	0.361895885	0.109468263	0.030193467	0.148736326	0.321262468	0.321262468	0.119527907

Emotion

Figure 5.14: Weighted Average of Emotions at different situations



	Joy	Anger	Surprize	Fear	Contempt	Disguist	Sadness
1:1 Interview	0.663997911	0.00716142	0.00467363	0.035028861	0.663997911	0.339555907	0.027614315
Public Speaking	0.385281148	0.11380476	0.051085899	0.027531565	0.33339193	0.230639398	0.066546762
Parliament Session	0.00674011	0.154188087	0.579962812	0.22120483	0.336123053	0.046888083	0.917982588

Emotion

Figure 5.15: Weighted Average of Emotions at different situations

Thus, these results can be fed to the research of behaviour sciences in order to study leadership principles among South Asian Female Politicians.

5.3 Demerits of Existing technology

Most of the existing technologies such as FACET are trained on Caucassians and Korean ethnicity. The Dravidian,Indo-Aryan, Mongoloid, along with some other Asian ethnicity variants are either minority or excluded from the training databases. On analysing emotions on Asian features, technology might perform with lower accuracy as compared to other ethnicity. When performing emotional analysis on South Asian females, we uncover some limitations of the current system.

Average Face detection success rate of South Asian Female women is 93.341 percent when fed to FACET 2.1 SDK. FACET 2.1 SDK has recorded a very high success rate (98-100) on performing simulations on Caucassion ethnicity. South Asian female leaders wear different types of traditional face marks such as bindi on their forehead (5.16). This mark on the face disrupts the intensities of other facial marks. Thus, the system is required to be trained for special facial landmarks.

Figure 5.16: Sushma Swaraj wearing Bindi on her forehead

Overall, these limitations substantiate the need to improve the existing technology (FACET) to real world applications. Affective Computing is currently addressing lots of issues, in discrimination of some very difficult expressions. Although these systems work fine with prototypical facial expressions, differentiating compound emotions is still a big milestone to reach.

6 Limitations of this research

The analysis performed on the different software- Microsoft Azure Face API, Emotient FACET, Affectiva AFFDEX was based on a single video, where we compared human judgement against the technologies. Comparison based on a single video is not enough to provide statistics for performance of these technologies. To provide reliable statistics for evaluating the performance of the softwares mentioned above, a greater number of videos should be analysed. A big amount of time is required to deep analyse the videos manually to extract results. This research might also require a number of participants who are willing to perform posed expressions in front of the camera. In the existing analysis we only considered the actual positive values for evaluation. We didn't check the presence of any other emotion at that moment. For eg, for angry facial expression we checked if the corresponding frame consists of angry value, but we didn't check what are the values of other emotions corresponding to that frame. Thus, we cannot evaluate the performance matrix for the technologies.

In emotional analysis of South Asian Women politicians, more leading ladies could be added to the list. The existing list contains very experienced politicians, who spent more than 10 years of their lives in politics. Younger politicians could be added to the list in order to compare their emotional analysis with others in the list.

7 Conclusion

Automatic recognition of facial expressions is very important to understand para linguistic communication and human emotion, to design Emotional Intelligent technologies, and to relate applications from it. Ekman and Friesen's Facial Action Coding System is currently the best and mostly accepted technique used for recognising facial expressions. Facial expression recognition is not an easy task. There can be circumstances such as illumination, face color, ethnic features, facial occlusions etc. Problems pertaining to light, pose quality of frames can be aroused for face detectors. Also, we analysed some research highlighting compound facial expressions of a single emotion, gender and ethnic differences in judgement of multiple emotions from facial expression which proves that there is a subjectivity to emotion. There is a subjectivity to how emotions are displayed and perceived among females and males. There is also an existence of situational subjectivity to how different cultures display emotion .There are differences in how emotions present themselves as Action Units. In the current basic emotion perspective of Automated Facial Expression analysis using FACS, it is generally ignored that contextual and cultural aspects can be very essential for classification of emotions with high accuracy.

Analysing the emotional states of South Asian women gave us insight into how leading ladies in politics represent themselves on different occasions. Our findings suggest that the moods of the politicians show different states of emotions when subjected to different levels of critical situations. They are also considered as semi trained actors, performing in limelight to influence the audience to their fullest potential. This data can be further used by behaviour sciences to predict more about their behaviour.

8 Bibliography

Anagnostopoulos, C.-N., Iliou, T., Giannoukos, I. (2015). Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artificial Intelligence Review, 43(2), 155-177.

Ashkanasy, N., Dorris, A. (2017). Emotions in the Workplace. Annual Review of Organizational Psychology and Organizational Behavior, 4(1). Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. Proceedings of the National Academy of Sciences, 113(4), 948-953.

Awamleh, R., Gardner, W. L. (1999). Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance. The Leadership Quarterly, 10(3), 345-373.

Bartlett, M. S., Littlewort-Ford, G., Movellan, J., Fasel, I., Frank, M. (2014). Automated facial action coding system: Google Patents.

Bartlett, M. S., Littlewort, G., Frank, M. G., Lainscsek, C., Fasel, I. R., Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. Journal of multimedia, 1(6), 22-35.

Bartlett, M. S., Movellan, J. R., Sejnowski, T. J. (2002). Face recognition by independent component analysis. IEEE Transactions on neural networks, 13(6), 1450-1464.

Bonaccio, S., O'Reilly, J., O'Sullivan, S. L., Chiocchio, F. (2016). Nonverbal Behavior and Communication in the Workplace: A Review and an Agenda for Research. Journal of Management, 42(5), 1044-1074.

Bono, J. E., Ilies, R. (2006). Charisma, positive emotions and mood contagion. The Leadership Quarterly, 17(4), 317-334.

Burgoon, J. K., Guerrero, L. K., Floyd, K. (2016). Nonverbal communication: Routledge.

Calvo, M. G., Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. Cognition and Emotion, 30(6), 1081-1106.

Calvo, R. A., D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. Affective Computing, IEEE Transactions on, 1(1), 18-37.

Calvo, R. A., D'Mello, S., Gratch, J., Kappas, A. (2014). The Oxford handbook of affective computing: Oxford University Press, USA. Cao, H., Verma, R., Nenkova, A. (2015). Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech. Computer speech language, 29(1), 186-202.

Cheng, H., Liu, Z., Zhao, Y., Ye, G., Sun, X. (2014). Real world activity summary for senior home monitoring. Multimedia Tools and Applications, 70(1), 177-197. Chu, M., Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(2), 257.

Clark, T., Greatbatch, D. (2011). Audience perceptions of charismatic and non-charismatic oratory: The case of management gurus. The Leadership Quarterly, 22(1), 22-32.

Du, S., Tao, Y., Martinez, A. M. (2014). Compound facial expressions of emotion. Proceedings of the National Academy of Sciences, 111(15), E1454-E1462. Ekman, P., Friesen, W. V. (1969). Nonverbal leakage and clues to deception. Psychiatry, 32(1), 88-106.

Ekman, P., Friesen, W. V. (1976). Measuring facial movement. Environmental psychology and nonverbal behavior, 1(1), 56-75.

Ekman, P., Friesen, W. V., Ellsworth, P. (2013). Emotion in the human face: Guidelines for research and an integration of findings: Elsevier.

Ekman, P., Oster, H. (1979). Facial expressions of emotion. Annual review of psychology, 30(1), 527-554.

El Kaliouby, R., Mahmoud, A., Turcot, P. J. (2016). Video recommendation using affect: Google Patents.

El Kaliouby, R., Picard, R., BARON-COHEN, S. (2006). Affective computing and autism. Annals of the New York Academy of Sciences, 1093(1), 228-248.

Littlewort, G., Bartlett, M. S., Lee., K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. . Image and Vision Computing,, 27(12), 1797-1803.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M. (2011).

The computer expression recognition toolbox (CERT). Paper presented at the Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. Paper presented at the Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.