ASHESI UNIVERSITY COLLEGE

DESIGN AND IMPLEMENTATION OF AN EMOTION RECOGNITION GUIDE FOR CHILDREN WITH AUTISM

APPLIED

B.Sc. Computer Science

Deborah Attuah

2018

ASHESI UNIVERSITY COLLEGE

Design and Implementation of an Emotion Recognition Guide for Children with Autism

APPLIED PROJECT

Applied Project submitted to the Department of Computer Science, Ashesi University College in partial fulfilment of the requirements for the award of Bachelor of Science degree in Computer Science

Deborah Attuah

April 2018

DECLARATION

I hereby declare that this Applied Project is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:

……………………………………………………………………………………………

Candidate's Name:

……………………………………………………………………………………………

Date:

……………………………………………………………………………………………

I hereby declare that preparation and presentation of this Applied Project were supervised in accordance with the guidelines on supervision of Applied Projects laid down by Ashesi University College.

Supervisor's Signature:

……………………………………………………………………………………………

Supervisor's Name:

……………………………………………………………………………………………

Date:

……………………………………………………………………………………

Acknowledgement

I would like to thank my supervisor Dr. Lorenzo Torresani, for his immense contribution to this project. Without his help, this project would not have been successful.

I would also like to thank my friends and family for their love and support.

Abstract

Individuals on the Autism Spectrum experience some difficult in recognizing and expressing emotions using their faces. This affects their ability to form long-term relationships and leaves them ostracized from others in our society. This project seeks to develop an application that helps children with autism recognise emotions in order to better place them in our society.

Table of Contents

Chapter 1: Introduction

1.1 Introduction

The matter of the health of citizens in any country is one of utmost priority to the government of said country. All over the world, huge portions of yearly budgets are allocated by nations towards providing inclusive and adequate healthcare to citizens. It is therefore no surprise that total expenditure on global health in 2017, as reported by the World Health Organization was US$ 6.5 trillion ("Spending on health: A global overview", 2018). In Ghana, the government in its 2017 budget indicates as one of its policy objectives, to “guarantee the right to health for all Ghanaians, through an efficient and a well-resourced health sector” (Ministry of Finance, 2017). Sadly, over the past years, contributions to the health sector in Ghana, have rarely been directed towards improving the well-being of individuals coping with neurological disorders, especially that of individuals coping with the Autism Spectrum Disorder.

1.2 Background

The term Autism Spectrum Disorder (ASD; used synonymously with pervasive developmental disorders, PDD), describes a range of neurodevelopmental disorders “diagnosed on the basis of early-emerging social and communication impairments, as well as rigid and repetitive patterns of behaviour and interests” (Frith & Happe, 2005). While the thinking and learning capabilities of such individuals can vary, ASD, which typically begins before the age of 3, creates various challenges throughout an individual’s life (Lindgren & Doobay, 2011).

The major ASDs include Autistic Disorder, Asperger’s Disorder and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS), all of which differ in terms of severity and the specific pattern of problems experienced, yet share certain behaviours.

One such behaviour exhibited by all individuals with ASD is an impairment in social interaction. This is expressed in their inability to “read” social cues which further creates a lack of social or emotional reciprocity, as well as a failure to develop peer relationships (Lindgren & Doobay, 2011). They are thus shunned by others who regard them as aloof and/or lacking empathy. According to Amaral, Cook, Leventhal & Lord (2000), individuals with autism almost never marry and only rarely form ordinary, reciprocal friendships (Amaral et al., 2000).

Considering that individuals with ASD form a part of our society, it is indeed quite sad that special attention has not been directed towards easing their burdens, as well as the burdens of others close to them. While there are a few centres available in Ghana that provide speech and language therapy for individuals on the autism spectrum, there are not enough therapists who specialise in the areas of emotion recognition and social communication ("AUTISM IN GHANA", 2018). The lack of adequate information and assistance in such areas increases the burden on families and friends, who wish to understand and be understood by their autistic relatives and peers, causing them to eventually give up. Timely interventions are therefore needed to prevent these individuals from being ostracized and enable them form close relationships.

While there are many ways of expressing emotion, facial expressions are considered by many as one of the most powerful and immediate ways of communicating one’s emotions (Tian, Kanade & Cohn, 2001). It can therefore be deduced that, developing an autistic child’s ability to identify facial expressions would improve their chances of interpreting and reciprocating emotions, thus enabling them form long-term relationships.

1.3 Project Purpose

This project therefore attempts to provide a facial expression recognition tool that can be used by children with autism as a guide to both identifying emotions expressed on the face and expressing their own emotions through their faces.

1.4 Project Objectives

The objectives of this project are therefore:

- To design and implement a system that is capable of detecting faces from video input
- To design and implement a system that is able to detect facial expressions on a given face
- To design and implement a system that is capable of relaying detected facial expressions to autistic users in an interactive and engaging manner.
- To investigate the effects of this system on the ability of its users to recognise emotions

1.5 Related Work

Many applications have been built in developed countries to aid individuals with autism in interpreting emotions from facial expressions. This section presents four of such applications, each presenting unique insight into the design and implementation of a system that utilizes facial emotion recognition algorithms in helping children with autism.

In 2011, the "LIFEisGAME" project developed a serious game[1], designed to help children with autism recognize and express emotions through facial expressions, using an avatar. The game flow of this project was influenced by Abirached et al (2011) who posited that, individuals learn to recognize emotions in a cyclical manner, involving the following stages: watch and recognize, learn by doing, recognize and mimic, and generalize or knowledge transfer to real life (Abirached et al., 2011). Thus, the "LIFEisGAME" project has users learning emotions by mimicking actions performed by an avatar in multiple game modes. While users were encouraged to play the various modes in a sequential manner, the application allowed for customizations that allowed them to begin wherever they pleased.

Two technologies enabling the game are real-time automatic facial expression analysis and virtual character synthesis which are used to map the player's facial movements onto the avatar (Abirached et al., 2011). While facial expression analysis was achieved with the help of Active Appearance Models[2], real-time virtual character synthesis was achieved through a rigging pipeline and motion capture technique. Feedback received during testing was generally favourable. However, it was noticed by the researchers that several children deliberately picked wrong answers because they preferred the wrong-answer feedback given by the application (Abirached et al., 2011).

Another application developed to help children with autism recognize emotions is "eMot-iCan". Among its many objectives is testing the theory that, "atypical attention patterns

[1] Games designed to be fun, entertaining and educational (Frutos-Pascual & Zapirain, 2017)

[2] Active Appearance Model is a computer vision algorithm for matching

are at the root of several features, such as impaired social and communicative skills, that are characteristic of individuals with ASD" (Sturm, Peppe, & Ploog, 2016). For this reason, the application follows a matching-to-sample-paradigm. That is, a sample image is given to a child, who is then prompted by the application to select a similar image from a given set of images. Depending on the images a child is selecting, the application is able to assess whether the child is paying attention to a particular part of the face, the entire face or not paying any attention at all (Sturm, Peppe, & Ploog, 2016). Unlike the "LIFEisGAME" project, users of the "eMot-iCan" application are given unique identification numbers. Players' sessions are therefore grouped using their identification numbers so that they can be accessed and analysed later. The application also provides administrative functions to guardians who wish to supervise the child's play. Guardians can also customize their child's gameplay and access logs of the child's interaction with the game.

"World of Kids" is another system designed in 2016, to help autistic children (Heni & Hamam, 2016). However, unlike the two applications presented earlier, "World of Kids" does not focus on improving the emotion recognition abilities of children with autism. Instead, the application seeks to improve their academic skills through gaming. Its implementation is however useful to this project because it uses facial emotion detection to identify the user's emotions and selects a game based on the emotion detected. All games available have the child learning one concept or another, thus improving their academic skills. The application then provides monthly analyses of the child's learning patterns to the parents or caregiver. This can be used by the guardian to optimize their ward's learning and comfort.

Other applications relating to human facial expression recognition are Affectiva (an MIT Media Lab created application for analysing smiles, which initially sought to help individuals on the autism spectrum) and EmoTrain, a mobile application that provides real-time facial expression classification using convolutional neural networks.

When playing EmoTrain, users are given images corresponding to one of seven facial expressions (happy, neutral, angry, surprise, fear, sad, disgust) and asked to mimic these expressions. The application then judges the user's performance by providing a score based on how accurate it is (Tsangouri, Li, Zhu, Abtahi, & Ro, 2016).

A similarity among the first three systems presented is a design that allows for customization. All applications however focus on recognising six basic emotions (happiness, sadness, fear, disgust, surprise and anger).

It is also worth considering research done in the development of machine learning algorithms for facial expression classification, as these will provide insight on the various algorithms behind games that use facial expression classifiers.

In 2017, Ivanovsky, Khryashchev, Lebedev, & Kosterin approached the problem of facial expression recognition in images by using a convolutional neural network.

> This neural network consisted of 4 convolutional layers, 4 layers with the ReLU activation function, 4 layers realizing the local normalization process, 3 layers describing the process of sampling using the maxpooling operation, 1 fully connected layer and 1 softmax layer. (Ivanovsky, Khryashchev, Lebedev, & Kosterin, 2017)

Due to the computational complexity of the deep convolutional neural network algorithm, acceleration of training and testing was achieved using independent streams on a GPU. A stochastic gradient descent optimization and a learning rate of 0.01 resulted in the classifier training for 60,000 iterations. Accuracy rate achieved by this algorithm in recognising six basic facial expressions (neutral, smile, surprise, squint, disgust, scream) was 84.98%.

1.6 Proposed System

Unlike the approach taken by Ivanovsky et al (2017) and Tsangouri et al (2016), this proposed system attempts to solve the problem of facial expression classification using a simple Logistic Regression model. While “EmoTrain” and “LIFEisGAME” ask users to match given expressions/faces, this system allows users the freedom to express whatever emotion they want, without scoring them. This way, the application feels more like a game to them than a task. Similar to “LIFEisGAME” and “EmoTrain” however, the proposed system provides real-time facial expression classification. It also follows the pattern of the applications identified above by predicting the six basic facial expressions ((happiness, sadness, fear, disgust, surprise and anger) with an extra expression; contempt.

1.7 Outline

This paper outlines the development of the proposed system in six chapters as follows:

Chapter 1 introduces the reader to the project being undertaken by presenting the problem being tackled and the solution this project proposes. In **Chapter 2**, a detailed requirement specification of the system to be built, including both functional and non-functional requirements is presented. **Chapter 3** presents the architecture and design of the system while **Chapter 4** details the implementation process used in developing the system. **Chapter 5** provides an outline of the testing process, as well as the results obtained. The paper concludes in **Chapter 6** with the limitations of the system, as well as recommendations for advancing the project.

Chapter 2: Requirements

This chapter presents a detailed requirement specification for the proposed system including the user and system requirements.

2.1 User Requirements

2.1.1 Users

The primary users of the proposed system are children with autism who will use the application as a guide to interpreting emotions communicated through facial expressions. Other stakeholders include the guardians of children with autism and teachers, who will be responsible for monitoring the child’s use of the application and the progress being made by their ward or student in interpreting the emotions behind facial expressions.

2.1.2 Requirements Gathering

User requirements were gathered using two approaches:

- Findings presented in existing literature.
- Observations in an autism centre.

Requirements from existing literature were gathered from research conducted either in the field of Affective Computing or Game Development for children with autism. Requirements from observation, on the other hand, were obtained after 40 hours spent in an autism centre.

2.1.3 Scenarios

The requirements gathered above were then used to frame scenarios describing various interactions the primary users and stakeholders may have with the application.

2.1.3.1 Scenario 1

Marlene, a mother of a 7-year-old son with autism has heard of the availability of an emotion recognition guide for children with autism and wishes to test its effectiveness on her son. She installs the application on her laptop and opens the application. Marlene guides Kojo in using the application by asking him to make a face in the direction of the laptop's webcam. Kojo makes a frown in the direction of the laptop's webcam and sees the word "Angry" displayed. Kojo is excited by this game and wishes to try another face. He continues playing for a while till he is tired, all the while being monitored by his mother. Marlene likes the level of engagement her son has with the application. She also likes that the application is teaching her son to express certain emotions with his face. She makes a mental note to set up more sessions between her son and the application.

2.1.3.2 Scenario 2

Angela, a teacher at an autism centre is looking for a new way to teach her students because she has noticed that they are not participating in her point-to-picture sessions. She is told by a friend that there is an application that can help children with autism recognize facial expressions. She is also told by this friend that, using computerized methods provide better results in the area of behavioural therapy than using the traditional point-to-picture methods.

She convinces the head teacher at her autism centre to give her the go-ahead to use this application with her students. After receiving approval, she installs the application and brings it to class. Instead of asking the children to point to pictures that show how they are feeling, she guides each student as they spend some time making faces to the camera. The children are intrigued by this new approach and participate fully in this session.

2.1.4 Use Cases

Below is a use case diagram illustrating derived use cases from the gathered requirements and scenarios.

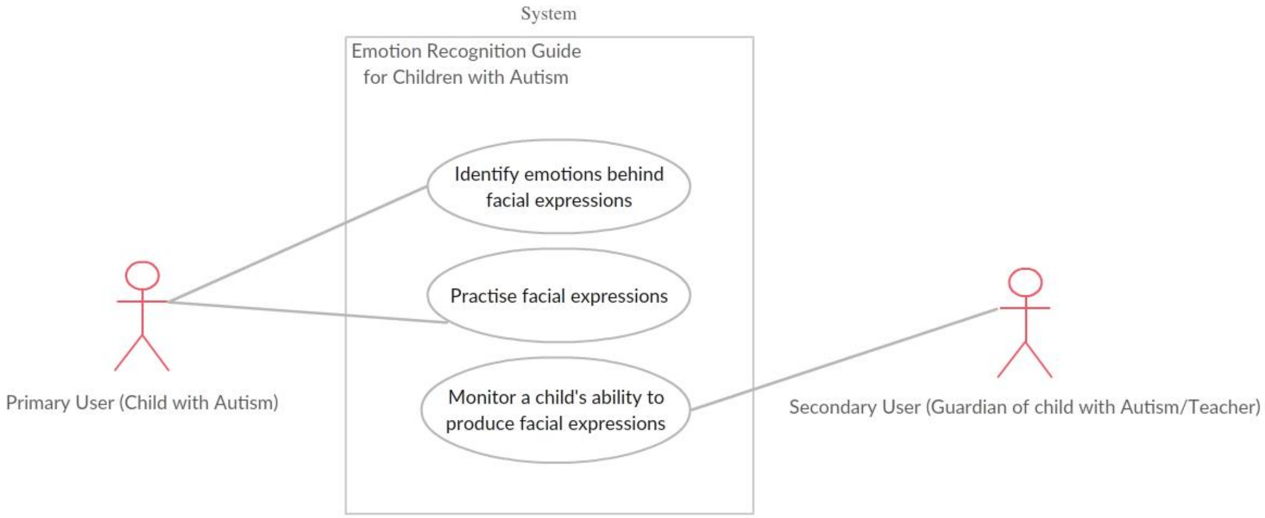

System
Emotion Recognition Guide for Children with Autism
Identify emotions behind facial expressions
Practise facial expressions
Monitor a child's ability to produce facial expressions
Primary User (Child with Autism)
Secondary User (Guardian of child with Autism/Teacher)

Figure 2.1. Use Case Diagram

2.1.5 Functional Requirements

The functional requirements in the list below describe what the system should do.

1. The application should be capable of providing feedback to the child regarding his or her current facial expression, in real time.
2. There should be a mechanism for motivating children to continue playing. Children with autism can be easily distracted and therefore, should be given sufficient motivation to continue playing the game.

2.1.6 Non-Functional Requirements

Usability

Given that children with autism prefer predictability and consistency in their environments (Heni & Hamam, 2016), the user interface design must be as consistent as possible so as not to put off the child.

2.2 System Requirements

1. The system should have access to the device’s camera.
2. The system’s Graphical User Interface should facilitate easy interaction between the application and the user, especially the child with autism. It should have as little text portions as possible and maintain a consistent layout.

Chapter 3: Architecture and Design

Based on the requirements listed in Chapter 2, the system's architecture was developed. This chapter provides a high-level overview of the system and its architecture. It also provides a detailed description of the design of the various modules that make up the system, the manner in which they communicate with each other and their necessity in the fulfilment of system and user requirements.

3.1 High Level System Overview and Architecture

3.1.1 System Overview

As stated in earlier chapters, the purpose of this system is dual-fold; to help children with autism (primary users) both understand facial expressions and express emotions through their face. This is achieved by providing primary users with real time facial expression classification. The system therefore operates by obtaining facial expressions from the user by use of a webcam and provides feedback to the user concerning the facial expression given in the image.

3.1.2 High Level Architecture

Given the mode of operation of the system described above, the application can be divided into three main modules; Facial Behaviour Analysis Module, Facial Expression Classifier Module and User Interface Module. The User Interface Module provides an environment for the user to communicate facial expressions to the system. The Facial Behaviour Analysis Module fetches this information from the User Interface Module, converts it into an appropriate format and passes this new representation to the Facial Expression Classifier Module. Upon receipt of data from the Facial Behaviour Analysis Module, the Facial Expression Class predicts the facial expression given by the user as a number. This number is then mapped to a textual facial expression, which is then presented to the user. This must be done in real-time so that the user is not distracted.

Figure 3.1 summarizes the high-level architecture of the system presented above.

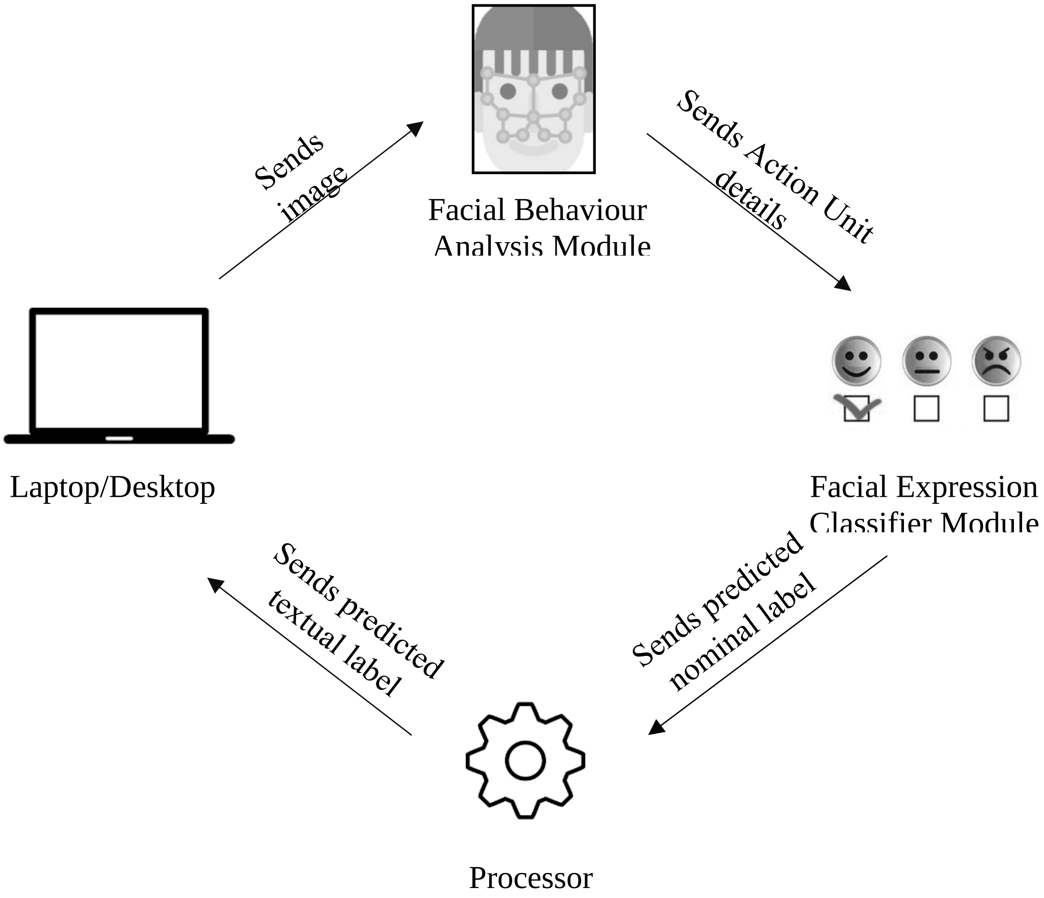

Sends image
Facial Behaviour Analysis Module
Sends Action Unit details
Laptop/Desktop
Facial Expression Classifier Module
Sends predicted textual label
Sends predicted nominal label
Processor

Figure 3.1. High-Level System Architecture

The subsequent sections of this chapter document the design of the various components that form the architecture of this system, and how these designs are informed by gathered requirements.

3.2 Hardware

This system consists of one hardware component; a desktop/laptop with a webcam. The desktop/laptop provides a platform for interaction between the autistic child and the system. It is necessary that the desktop/laptop being used has a properly functioning webcam as this will be used to capture users' facial expressions in real time.

3.3 System Modules

As stated earlier, this system consists of the following key modules communicating with one another to provide a worthwhile experience to the user:

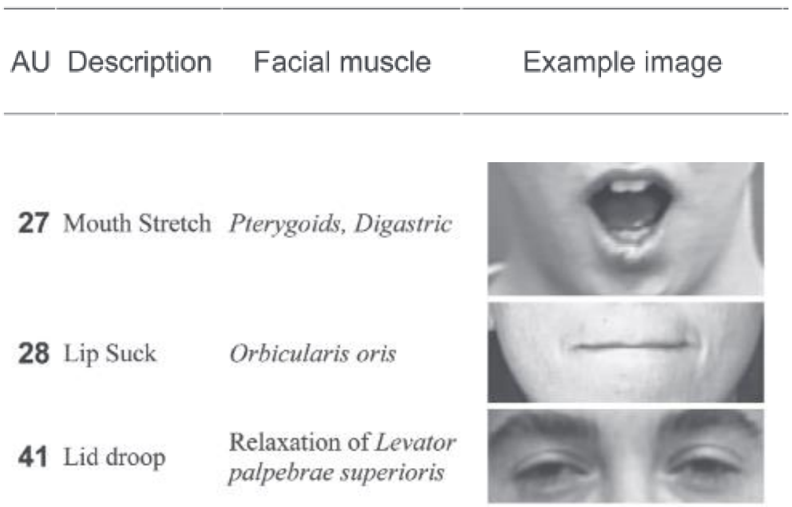
- Facial Behaviour Analysis Module
- Facial Expression Classifier Module
- User Interface Module

3.3.1 Facial Behaviour Analysis Module

The Facial Behaviour Analysis Module is responsible for identifying aspects of individuals' faces that would enable the system predict facial expressions. This is done using Computer Vision techniques to identify human faces from video input obtained using a webcam. Action Units are then retrieved from the identified faces and sent to the Facial Expression Classifier Module.

3.3.1.1 Action Units

The Facial Action Coding System (FACS), first published in 1978 by Ekman and Friesen, is an anatomical system for describing all observable facial movements ("Facial Action Coding System", 2018). “It breaks down facial expressions into individual components of muscle movement” known as Action Units ("Paul Ekman Group", 2018). The FACS 2002 specifies 9 action units in the upper face and 18 in the lower face, each of which has both a numeric and verbal label. The figure below shows some examples of Action Units, their description and the facial muscles they correspond to. AU28 for example, known as the Lip Suck, corresponds to the Orbicularis oris muscle found in the face.



AU	Description	Facial muscle	Example image
27	Mouth Stretch	*Pterygoids, Digastric*	
28	Lip Suck	*Orbicularis oris*	
41	Lid droop	Relaxation of *Levator palpebrae superioris*	

(Cohn, Ambadar, & Ekman, 2006).

Figure 3.2. Examples of Action Units and their corresponding facial muscles

These AUs, in various combinations, according to research, have shown increasing evidence of their relevance to the interpretation of facial expressions (Cohn, Ambadar, & Ekman, 2006). For example, Wegrzyn, Vogt, Kireclioglu, Schneider & Kissler (2017) report that the presence of AU12 and AU25 are highly diagnostic of a facial expression being “happiness”. This system therefore makes use of Action Units obtained from the user’s face in classifying the user’s facial expression. Again, this must be done in real-time to prevent the user from getting distracted.

3.3.2 Facial Expression Classifier Module

The Facial Expression Classifier Module has the singular responsibility of predicting the facial expression on input received by the system. It receives details concerning specific Action Units of a given image frame from the Facial Behaviour Analysis Module as input, and provides as output, a prediction of the expression on the face, based on the AU details of that image frame. Given that there exists a wide range of facial expressions available, the number of expressions capable of being classified by the Facial Expression Classifier Module has been limited to the seven basic emotions which occur in a similar manner, in all human beings, irrespective of race or gender (Heni & Hamam, 2016), for simplicity. These emotions are anger, contempt, disgust, fear, happiness, sadness and surprise.

3.3.3 User Interface Module

This module consists of all components that interact directly with the user. It provides the user with the ability to send videos of themselves making varied facial expressions and receive feedback from the application in real time.

3.3.3.1 Mock-up of the user interface

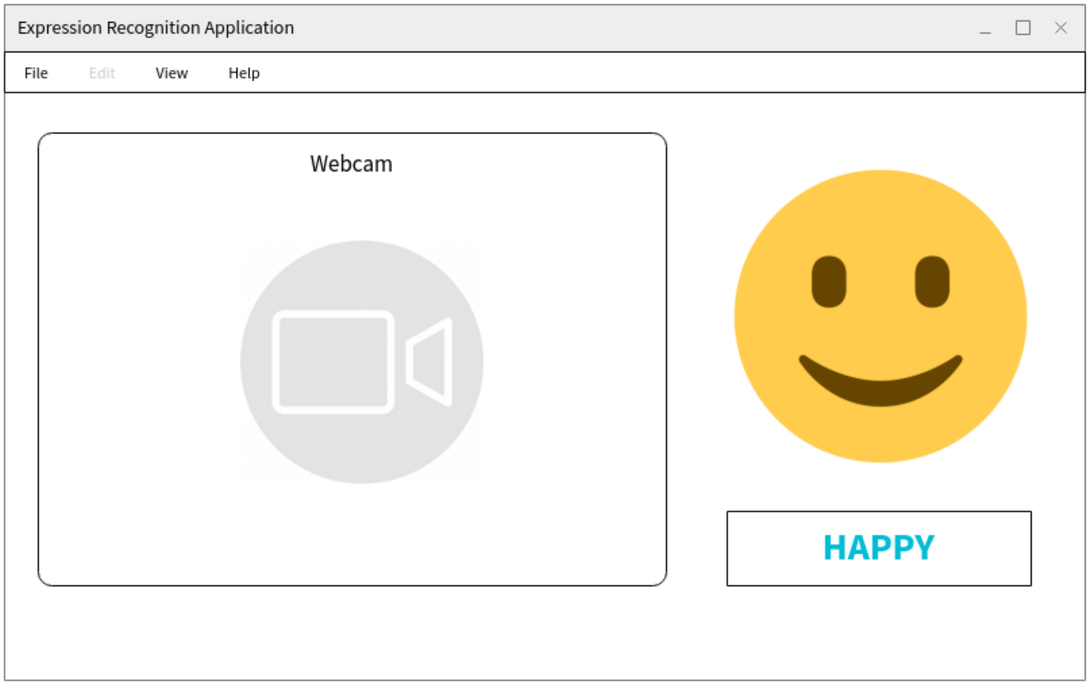


Figure 3.3. Mock-Up of User Interface

Figure 3.3 presents a Mock-Up of the User Interface through which the user interacts with the system. As seen in the image, the user will see the facial expression they are currently making in the section of the interface labelled Webcam. The right-side of the application provides feedback of the emotion being expressed with a textual label (e.g. happy, sad, angry) and an emoji that correlates to said label. The user interface is designed to be as simple and consistent as possible, while staying engaging to fulfil the requirements of preventing the user from getting distracted and maintaining consistency.

Chapter 4: Implementation

This chapter describes the various tools and techniques used in the implementation of the various modules of this system, and the manner in which they were used.

4.1 User Interface Module Implementation

The User Interface for this system was developed using the Windows Presentation Foundation (WPF) subsystem, a graphical user interface (GUI) framework, developed by Microsoft for rendering user interfaces for Windows-based applications. It enables developers create labels, textboxes, menu items and such elements for their applications without having to manually draw them, by using the Extensible Application Markup Language (XAML), an XML (Extensible Markup Language) based language, to define and link the various interface elements.

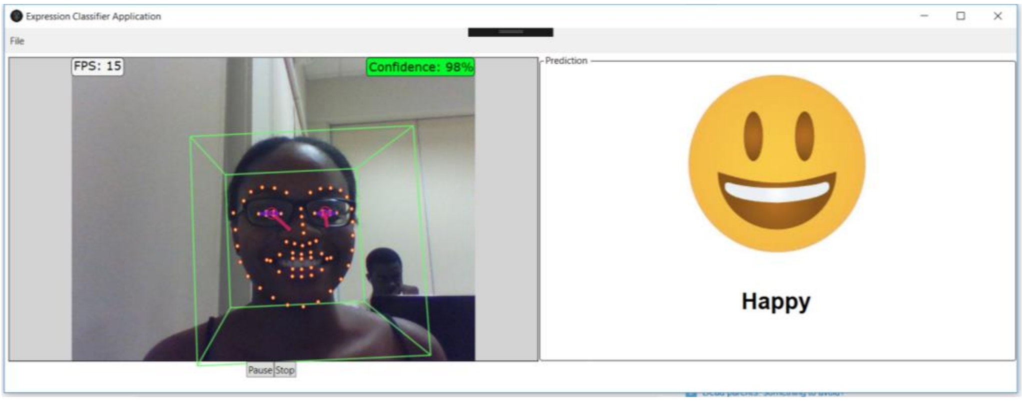


Figure 4.1. User Interface of System

The User Interface as seen in Figure 4.1, consists of two parts. The left side of the application allows the user to see the facial expression they currently have while the right provides the user with a prediction of the expression they are currently making. The interface is designed to be as simple as possible so as not to confuse the primary user.

4.2 Facial Behavioural Analysis Module Implementation

The main tool responsible for providing facial behavioural analysis is OpenFace (Baltrusaitis, Robinson, & Morency, 2016).

4.2.1 Overview of OpenFace

OpenFace is an open source tool designed to aid computer vision and machine learning researchers, as well as individuals interested in building interactive applications based on facial behaviour analysis (Baltrusaitis, Robinson, & Morency, 2016). Its capabilities include state-of-the art facial landmark detection and tracking, using Constrained Local Neural Fields (CLN[3]F), head pose estimation, facial action unit recognition and eye-gaze estimation, all done in real time, making it an invaluable tool for affective computing. With regards to the scope of this system, OpenFace serves as a tool to retrieve the presence (visibility of an AU in an image) and intensities (how intense the AU is on a 5-point scale) of AUs from the faces of individuals using the application. Of the many AUs available in the FACS, OpenFace recognizes 18 (See table B1 for a list of AUs recognized by OpenFace and their respective descriptions). OpenFace therefore returns, in terms of AUs, for each image received, an array of size 35 where the first 17 (OpenFace does not assess the intensity of one of its recognized AUs) items contain continuous values ranging between 0 and 5, representing the intensity of an AU and the

[3] The Constrained Local Neural Field is an instance of the Constrained Local Model presented by Baltrusaitis, Robinson & Morency. It is designed to handle feature detection in “complex” scenes.

remaining 18, binary values representing the presence (1) or absence (0) of a given AU. The main tools in the OpenFace toolset that provide Action Unit presence and intensity detection are the FaceLandmarkImg tool, which detects facial features, and the FeatureExtraction tool, which extracts the necessary AUs from the detected facial features.

4.3 Facial Expression Classifier Module Implementation

As stated in Section 3.3.2, the responsibility of the Facial Expression Classifier is to predict the expression on a user's face in any given video frame. Literature reviewed shows that there are many approaches to solving the problem of facial expression classification. Notable among them are the use of Convolutional Neural Networks and Active Appearance Models. For the implementation of the Facial Expression Classifier used in this system, a Logistic Regression approach was used. While research shows that neural networks perform better than Logistic Regression in terms of prediction accuracy, the Logistic Regression approach was chosen for this classifier because of its faster training ability.

4.3.1 Logistic Regression

Logistic regression is a supervised classification algorithm used to assign observations of a given problem to a discrete set of classes. That is, given an input $x^{(i)}$, the algorithm determines an output $y^{(i)}$, such that $y^{(i)} \epsilon \{ 0,1, \dots \mathrm{n}\}$, where n is a real number greater than 0. When used for addressing binary classification problems (classifying a given instance of the problem as belonging to class A or B), it is known as Binary Logistic Regression. On the other hand, when used to classify observations into three or more classes, it is known as Multiclass Logistic Regression. Given, the number of expressions under consideration for classification in this project, the classifier developed falls under Multiclass Logistic Regression.

The general approach to solving a Multiclass Logistic Regression problem is develop a hypothesis $h_\theta(x)$ that computes for a given input, the probability that $P(y = k|x)$ for each value of k = 1, … ,n. In other words, for each input, the hypothesis function $h_\theta\left(x^{(i)}\right)$ outputs an n-dimensional vector containing probabilities for each class k, all of which sum up to 1. The hypothesis function therefore takes the form,

$$h_\theta\left(x^{(i)}\right) = \begin{bmatrix} P(y = 0|z^{(i)}) \\ P(y = 1|z^{(i)}) \\ \vdots \\ P(y = n|z^{(i)}) \end{bmatrix}$$

The function $P(y = k|z^{(i)})$, known as the Softmax function,

$$P(y = j \mid z^{(i)}) = \phi_{softmax}(z^{(i)}) = \frac{e^{z^{(i)}}}{\sum_{j=0}^{k} e^{z_k^{(i)}}},$$

is used to calculate the probabilities of an event occurring given certain values. In this case, the value, $z^{(i)}$, that serves as input to the Softmax function is defined as,

$$z = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n + b = \sum_{l=0}^{n} \theta_l \, x_l \quad + b = \theta^T x + b$$

where θ = vector of weights,

x = feature vector of a training example

b = bias

The following are the stages involved in performing a Multiclass Logistic Regression:

- Input Definition
- Development of Linear Model
- Training Model
- Testing Model
- Using Model for Prediction

4.3.1.1 Input Definition

The inputs to a multiclass logistic regression, (x) are the features that are predictive of the class a given instance of a problem belongs to. In this case, the AUs of an individual's face, which are known to be predictive of a person's facial expression, serve as the inputs. Given that OpenFace recognizes the presence of 18 AUs, as well as the intensities of 17 AUs, the input to this multiclass logistic regression instance is a vector x of size 35, with the first 17 values representing the intensities of 17 out of the total 18 AUs considered by OpenFace and the remaining 18 values representing the presence of each of the 18 AUs considered by OpenFace, as illustrated in the figure below.

$$x = \begin{bmatrix} AU01\_r \\ AU02\_r \\ \vdots \\ AU45\_r \\ AU01\_c \\ AU02\_c \\ \vdots \\ AU45\_r \end{bmatrix}$$

Figure 4.2. Vector x showing the input features to the Facial Expression Classifier. Features ending in r represent AU intensities (eg: AU01_r represents the intensity of AUO1 in a specific frame) while

features ending in c represent AU occurrences (e.g.: AU01_c represents the occurrence of AUO1 in a specific frame)

4.3.1.2 Development of Linear Model

The Facial Expression Classifier was developed using the Tensorflow library.

4.3.2 TensorFlow

TensorFlow is an open source software library developed by Google that provides strong support for machine learning and deep learning computations. For the development of the Logistic Regression model used in this application, Tensorflow's LinearClassifier estimator was used, along with the library's implementation of the FtrlOptimizer (Follow the Regularized Leader Optimizer). While Tensorflow provides a number of Optimizers, the FtrlOptimizer was selected because of its adaptive learning rates. In order to address over-fitting and feature selection L1 regularization (a regularization term which adds the squared magnitude of coefficients as a penalty term to the loss function) and L2 regularization (a regularization term which adds the absolute magnitude of coefficients as a penalty term to the loss function) were used.

4.3.3 Dataset

The Extended Cohn-Kanade Dataset from the University of Pittsburgh was used in training and testing the Facial Expression Classifier. It consists of 10,633 image sequences from 123 subjects. Each sequence begins with a neutral expression and progresses to a pre-determined peak expression (Lucey et al., 2010); anger, contempt, disgust, fear, happy, sadness or surprise. All sequences are given nominal labels depending on the peak expression. For example, in the image below, the subject begins with a neutral expression and ends in a surprised expression.

(©Jeffrey Cohn) *Figure 4.3. Image showing dataset representation*

Therefore, in the dataset all three images above have the label 7, which corresponds to surprise.

Images in the dataset without labels were removed, as these would not have been able to contribute to training or testing of the facial expression classifier. This reduced the total number of images in the dataset to 6,062 images.

4.4 Training the Facial Expression Classifier

The training of the Facial Expression Classifier is the stage where the classifier is gradually optimized as it "learns" the features of the dataset.

4.4.1 Data Preparation

In order to do this, the dataset was first divided into training and test data with training data consisting of 4,985 images from 100 subjects and test data consisting of 1,077 images from a different set of 23 subjects. The purpose for division into train and test data based on subjects was to evaluate the classifier's ability to make predictions on faces it had never seen before during the testing stages. These images were then converted into AU intensity and presence values using OpenFace's FeatureExtraction class. The corresponding expression labels were then attached to the obtained AU intensity and presence values to form a training dataset of dimension 4,985 x 36 and test dataset of dimension 1,077 x 36 (36 representing the AU details and the given label).

4.4.2 First Training Session

The purpose of the first training session was to determine which group of entries in the training dataset provided better accuracy results. The first training session was therefore carried out in batches. The first batch used all entries in the training dataset to train the Logistic Regression model. The second batch on the other hand used only entries corresponding to the last frames of an image sequence, while the third batch used only entries corresponding to the last two frames. Table 4.1 provides a summary of the details of the data used for the various batches of training in the first training session

Table 4.1. Summary of Data Details for first training session

Batch	Frames Used (per sequence)	Training Data Size (# frames x #AUs + label)	Test Data Size (#frames x #AUs + label)
First Batch	All Frames	4,985 x 36	1,077 x 36
Second Batch	Last Frame	280 x 36	57 x 36
Third Batch	Last Two Frames	559 x 36	114 x 36

4.4.3 Second Training Session

The purpose of the second training session was to identify the best values of Learning Rate, L1 regularization strength and L2 regularization strength for training the logistic regression model. For this reason, different combinations of learning rate, l1 regularization and l2 regularization were applied to the Logistic Regression model.

The results of the models obtained by using the various combinations can be found in Table 5.1.

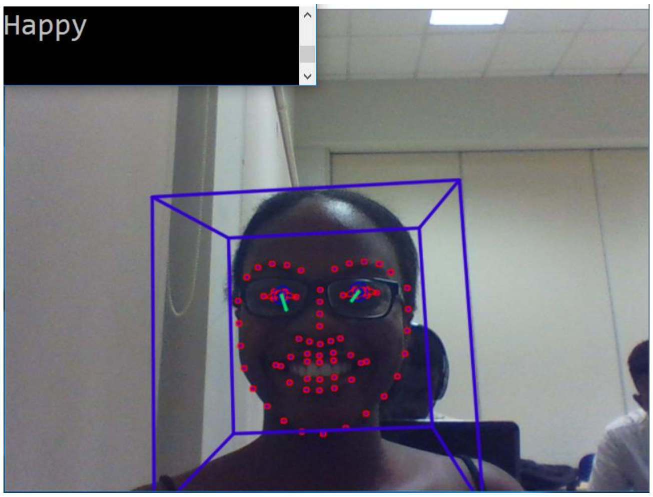
4.5 Prediction Using the Facial Expression Classifier

The Facial Expression Classifier was integrated with the Facial Behaviour Analysis Module after the second training session, to view its performance on data obtained from a webcam. This was achieved by retrieving the various weights and biases that defined the Logistic Regression Classifier and applying the hypothesis function

$$h_\theta\left(x^{(i)}\right) = \begin{bmatrix} P\left(y = 0 \middle| z^{(i)}\right) \\ P\left(y = 1 \middle| z^{(i)}\right) \\ \vdots \\ P\left(y = n \middle| z^{(i)}\right) \end{bmatrix},$$

to Action Units obtained from the Facial Behaviour Analysis Module.

As shown in Figure 4.4, the classifier was able to predict facial expressions received from a webcam.



Happy

Figure 4.4. Image showing prediction by Facial Expression Classifier given input from Facial Behaviour Analysis Module

However, it was noticed that the classifier still produced facial expressions such as happy or angry, even when a neutral expression was made. For this reason, a neutral expression needed to be added to the possible expressions that could be classified by the Logistic Regression model. A third training session was therefore conducted.

4.5.1 Third Training Session

The third training session was carried out after the first system testing session. It was necessary to perform a third training session because, it became apparent after system testing that it was necessary to include a neutral expression class to the possible classes predicted by the Facial Expression Classifier, as the system predicted one of the seven basic expressions, even when the user had a neutral facial expression. In order to rectify this, a new dataset was created composing of the dataset used in the second training session and 110 first frames of image sequences in the dataset. First frames of each image sequence were used because, as stated earlier, each image sequence began with subjects in a neutral expression. The resulting dataset was divided into 90% of subjects for training and 10% of subjects for test. The new classifier was then created with an increased number of predictable classes (8). With the exception of the number of predictable classes, this classifier resembled the old classifier.

Chapter 5: Testing and Results

This chapter describes the various processes that took place to ensure that the application fulfilled requirement specifications. Specifically, this chapter details the processes and results of component testing and system testing.

5.1 Component Testing

The Facial Expression Classifier Module was tested to assess performance of the application in facial expression prediction using Action Units.

5.1.1 Facial Expression Classifier Testing

5.1.1.1 First Test Session

The main purpose of the first test session was to determine whether or not to use all labelled entries in the dataset. As stated earlier, the dataset used for training the Facial Expression Classifier contained sets of image sequences, each given one label, although each image sequence began with a neutral expression and ended in a target expression. That is, while a sequence may have contained both neutral and angry facial expression, the sequence was given the label Anger. It therefore became important to assess the specific frames in each sequence to be included in the final dataset used for training. Figure 5.1 provides a summary of the results from the first test session.

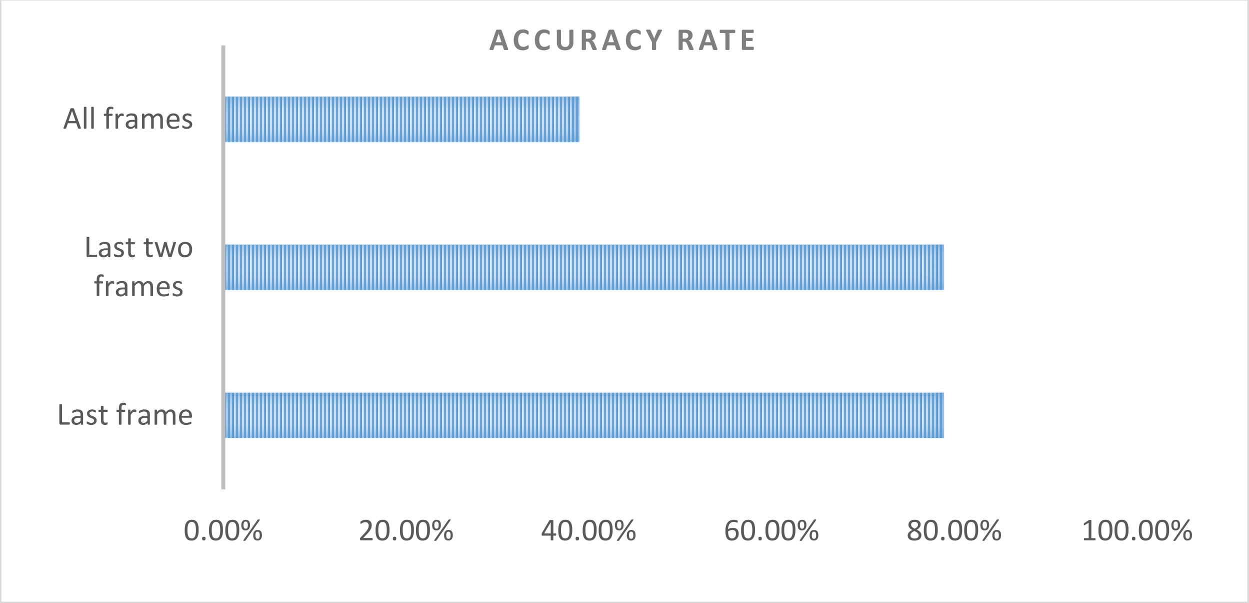


Figure 5.1. Summary of Result from First Testing Session

Testing showed that with the factors, learning rate, l1 regularization strength, l2 regularization strength, optimizer and cost function held constant, using all frames in each sequence provided a 42.3% accuracy level while using the last frames and the last two frames in each sequence both provided an equal accuracy level of 78.9%. It was therefore determined that the best approach was to use either the last frame or the last two frames of each image sequence.

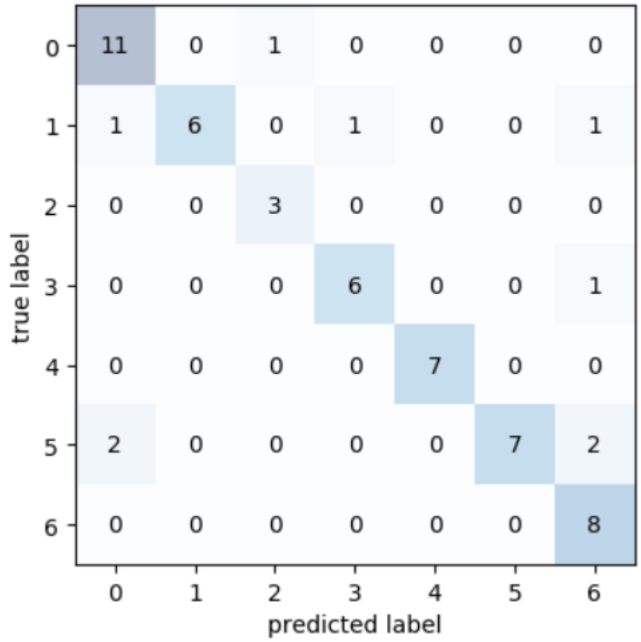
5.1.1.2 Second Test Session

After determining the best set of data to be used for training the Facial Expression Classifier, the next test session sought to determine the best learning rate, l1 regularization strength and l2 regularization strength for the classifier. Table 5.1 presents a summary of the results from the second test session

Table 5.1. Table Showing the various combinations of Learning Rate, L1 Regularization and L2 Regularization and the resulting Testing Accuracy.

Learning Rate	L1 Regularization	L2 Regularization	Testing Accuracy (after 1001 training steps)
1	1	1	73.7%
1	2	1	73.7%
1	3	1	73.7%
1	3	2	73.7%
0.1	3	2	78.9%
0.01	3	2	84.2%

The confusion matrix below shows the performance of the classifier with Learning Rate 0.01, L1 Regularization strength 3, and L2 Regularization strength 2, in predicting the various classes when given test data.



true label \ predicted label	0	1	2	3	4	5	6
0	11	0	1	0	0	0	0
1	1	6	0	1	0	0	1
2	0	0	3	0	0	0	0
3	0	0	0	6	0	0	1
4	0	0	0	0	7	0	0
5	2	0	0	0	0	7	2
6	0	0	0	0	0	0	8

Legend
0 = Anger 1 = Contempt 2 = Disgust 3 = Fear 4 = Happy 5 = Sadness 6 = Surprise

Figure 5.2. Confusion Matrix showing the performance of the Facial Expression Classifier on Test Data

Figure 5.3 below illustrates the classifier's performance on test data after 1001 training steps.

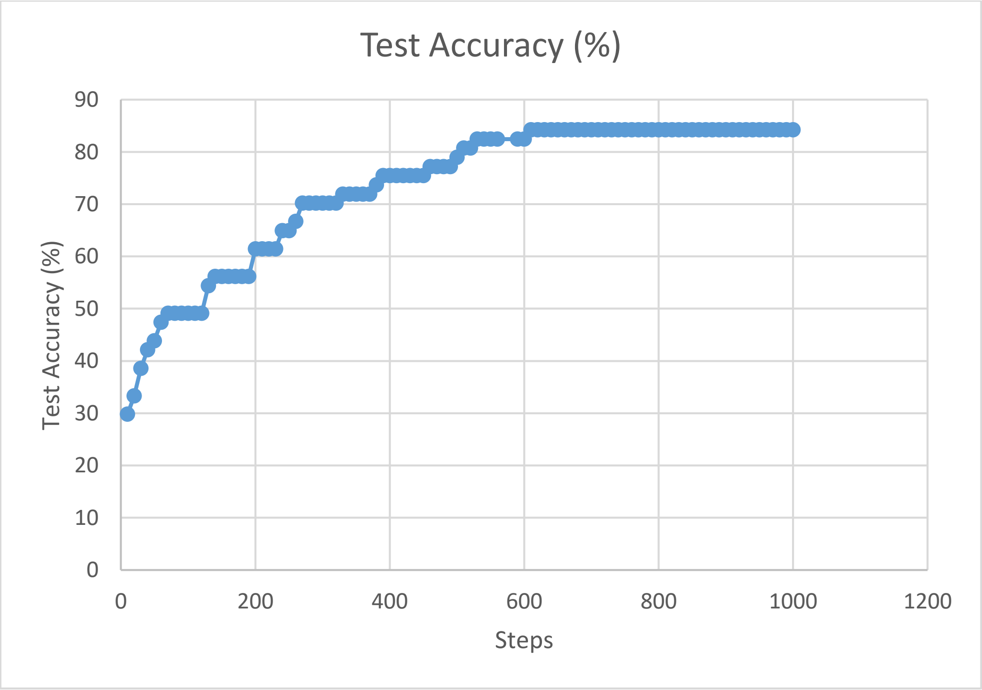


Figure 5.3. Graph illustrating classifier's performance on test data after 1001 training steps.

5.2 System Testing

System testing was first conducted after the integration of the Facial Behaviour Analysis and Facial Expression Classifier. Since a "neutral" expression was not included as one of the predictable classes, the system predicted the expressions, "Happy" and "Surprised", even when the user decided not to perform any of the predictable expressions. It therefore became necessary to include a "neutral" expression to the classifier's possible prediction. After 9001 training steps, the Facial Expression Classifier achieved an accuracy of 83.1% on the test data.

Chapter 6: Conclusions and Recommendations

6.1 Conclusions

In this paper, the design and implementation of a system capable of guiding children with autism through the process of recognizing emotions using facial expressions was presented. In fulfilment of the functional requirements specified in Chapter 2, the application identifies human faces from a webcam using computer vision techniques, obtains details of specific Action Unit presence and intensities from these identified faces and provides as feedback to the user, the facial expression on the identified faces. A prediction accuracy of 84.2% on a model that classified a given expression as either anger, happiness, fear, contempt, disgust, sadness or surprise, while a prediction accuracy of 83.1% was achieved on a model that classified a given expression as either neutral, anger, happiness, fear, contempt, disgust, sadness or surprise.

6.2 Limitations

The performance of this application, in terms of facial behaviour analysis and facial expression classification is highly dependent on the lighting conditions of the room in which it is being operated. The facial expression classifier, being trained on images from the Extended Cohn-Kanade dataset, is also limited by the interpretation of the various expressions by the subjects used in the dataset.

6.3 Future Work

The application could be improved by including temporal smoothing. That is, instead of providing feedback individually for each frame, feedback could be computed based on the previous n frames as individual’s expressions tend to be consistent over time. The application could also be improved by including game modes that allow the child to first guess the emotion before receiving positive or negative feedback from the application as is needed. The addition of a feature that allows the guardian to track the child’s progress in identifying facial expression would also contribute to the usefulness of this application.

References

Abirached, B., Zhang, Y., Aggarwal, J. K., Tamersoy, B., Fernandes, T., & Carlos, J. (2011). Improving communication skills of children with ASDs through interaction with virtual characters. In *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1–4). https://doi.org/10.1109/SeGAH.2011.6165464

Amaral, D. G., Cook, E. H., Leventhal, B. L., & Lord, C. (2000). Autism Spectrum Disorders. *Neuron*, 355-363.

AUTISM IN GHANA. (2018). *Aactgh.org*. Retrieved 17 April 2018, from http://aactgh.org/index.php/lookup

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016, March). *OpenFace: an open source facial behavior analysis toolkit.* Retrieved from University of Cambridge: https://www.cl.cam.ac.uk/~tb346/res/openface.html

Cohn, J. F., Ambadar, Z., & Ekman, P. (2006). Observer-Based Measurement of Facial Expression With the Facial Action Coding System. In *The handbook of emotion elicitation and assessment* (pp. 203-221). New York: Oxford University Press Series in Affective Science.

Facial Action Coding System. (2018). *Paul Ekman Group*. Retrieved 23 March 2018, from https://www.paulekman.com/product-category/facs/

Frith, U., & Happe, F. (2005). Autism spectrum disorder. *Current Biology*, R786-R790.

Frutos-Pascual, M., & Zapirain, B. G. (2017). Review of the Use of AI Techniques in Serious Games: Decision Making and Machine Learning. *IEEE Transactions on*

Computational Intelligence and AI in Games, *9*(2), 133–152. https://doi.org/10.1109/TCIAIG.2015.2512592

Heni, N., & Hamam, H. (2016). Design of emotional educational system mobile games for autistic children. In *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 631–637). https://doi.org/10.1109/ATSIP.2016.7523168

Ivanovsky, L., Khryashchev, V., Lebedev, A., & Kosterin, I. (2017). Facial expression recognition algorithm based on deep convolution neural network. In *2017 21st Conference of Open Innovations Association (FRUCT)* (pp. 141–147). https://doi.org/10.23919/FRUCT.2017.8250176

Lindgren, S., & Doobay, A. (2011). Evidence-Based Interventions for Autism Spectrum Disorders.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

Ministry of Finance. (2017, March 2). The Budget Statement and Economic Policy of the Government of Ghana for the 2017 Financial Year. Accra, Ghana. Retrieved from http://www.mofep.gov.gh/sites/default/files/budget/2017%20BUDGET%20STATEMENT%20AND%20ECONOMIC%20POLICY.pdf

Spending on health: A global overview. (2018). *World Health Organization*. Retrieved 17 April 2018, from http://www.who.int/mediacentre/factsheets/fs319/en/

Sturm, D., Peppe, E., & Ploog, B. (2016). eMot-iCan: Design of an assessment game for emotion recognition in players with Autism. In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1–7). https://doi.org/10.1109/SeGAH.2016.7586228

Tsangouri, C., Li, W., Zhu, Z., Abtahi, F., & Ro, T. (2016). An interactive facial-expression training platform for individuals with autism spectrum disorder. In *2016 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp. 1–3). https://doi.org/10.1109/URTC.2016.8284067

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLOS ONE*, *12*(5), e0177239. https://doi.org/10.1371/journal.pone.0177239

Appendices

Appendix A

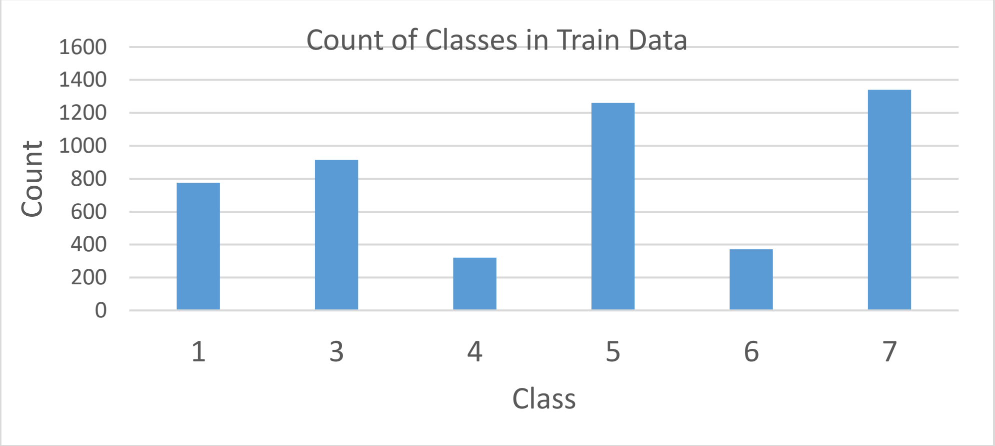


Figure A1. Graph showing distribution of the various classes in Train Data for First Test Session.

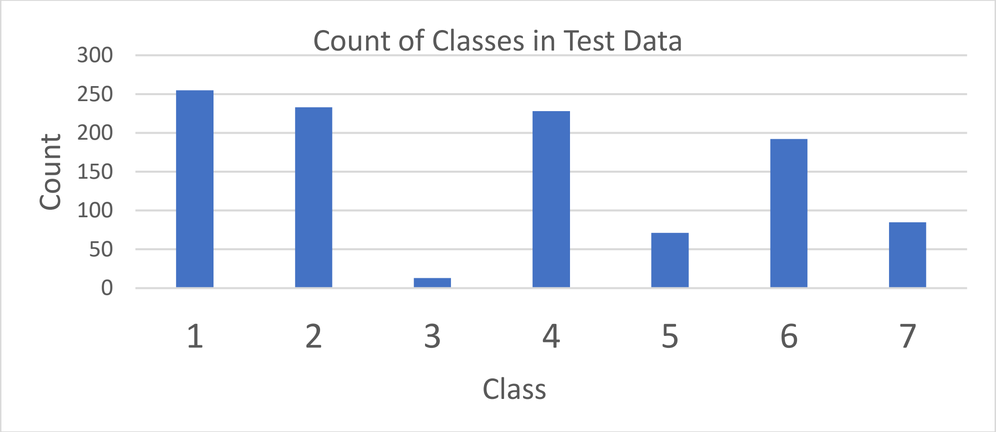


Figure A2. Graph showing distribution of the various classes in Test Data for First Test Session.

Appendix B

Action Unit	FACS name	Facial Muscle
AU1	Inner brow raiser	Frontalis (pars medialis)
AU2	Outer brow raiser	Frontalis (pars lateralis)
AU4	Brow lowerer	Depressor Glabellae, Depressor Supercilii, Corrugator Supercilii
AU5	Upper lid raiser	Levator Palpebrae Superioris, Superior Tarsal Muscle
AU6	Cheek raiser	Orbicularis Oculi (pars orbitalis)
AU7	Lid tightener	Orbicularis Oculi (pars palpebralis)
AU9	Nose wrinkler	Levator Labii Superioris Alaeque Nasi
AU10	Upper lip raiser	Levator Labii Superioris, Caput Infraorbitalis
AU12	Lip corner puller	Zygomaticus Major
AU14	Dimpler	Buccinator
AU15	Lip corner depressor	Depressor Anguli Oris
AU17	Chin raiser	Mentalis
AU20	Lip Stretcher	Risorius w/ platsyma
AU23	Lip Tightener	Orbicularis oris
AU25	Lips Part	Depressor Labii Inferioris

AU26	Jaw Drop	Masseter; Relaxed Temporalis and Internal Pterygoid
AU28	Lip Suck	Orbicularis Oris
AU45	Blink	Relaxation of Levator Palpebrae Superioris; Contraction of Orbicularis Oculi (Pars Palpebralis)

Table B1. Table of Action Units obtained by Facial Behaviour Analysis Module