Emotion recognition from facial expressions using SVM algorithm

Ioannis Lemonis

Thesis Committee:

Prof. Michalis Zervakis
Prof. Kostas Kalaitzakis
Dr. Eleftheria Sergaki

SCHOOL of ELECTRICAL & COMPUTER ENGINEERING
Chania January 2019

ABSTRACT

The recognition of facial expressions and the corresponding emotion that they convey is something every human understands and it is universal without requiring training. It is instinctively “coded” in humans DNA. Human – Computer interaction can greatly benefit from recognizing various human emotions just by looking at us (no need for same language or any spoken language at all). The way we can achieve this is by taking frames (or live video) that depict human faces showing different emotional expressions and convert them to greyscale which makes it easier and faster to locate the facial area. We then apply a pre-trained mask with 68 flags-points (each one with unique coordination) based on a method that tries to apply these flags around the main facial areas that show details of emotions, around the orifices (mouth, eyes, nose). We proceed by taking all possible combinations of these 68 points and calculate the Euclidean distance between each set and then by using an SVM Machine learning method we train the system to recognize four main emotions (Anger, Joy, Tranquility, Sadness) experimented on two Data Sets (with depictions of expression and corresponding feeling). The data sets being used for the training and validation of the SVM are: Patras A.I.nD.M. data set of 84 directed facial poses (portrait angle) and Fer2013 data set of 960 Not directed facial poses (random angle). The Fer2013 data set was created for a Facial Recognition competition to test the recognition algorithms to their limits, with very small resolution 48x48 pixels and full with out of focus, obscured (either by hair, hands, sunglasses, hats, bad angle etc) pictures, some of them are not even from real photographs rather portraits or drawings of faces, with wide variety of different people from around the world. After the training we apply a validation test to find out how accurate it is. Of the four examined emotions *Joy* is the overall *best* distinguished with 100% (in all three statistical measures: Sensitivity, Specificity and Accuracy) in both Patras data sets tests 72%, 75% and 85% (Sen, Spe, Acc) in the Fer2013 data set accordingly. *Anger* comes second with 100%, 83%, 87.5% in Patras data set and 56%, 85%, 78% in Fer2013 data set. *Tranquility* is third with 50%, 100%, 87.5% (Sen, Spe, Acc) in Patras data set and 45%, 67% and 62% in Fer2013 data set. Finally *Sadness* with 87.5%, 50% and 83% in Patras data set and 21%, 80% and 62% in Fer2013 data set.

Keywords: facial expressions, emotion, computer vision, SVM, Dlib, OpenCV, image processing, classifier, Fer2013, IBUG 300-W

ΠΕΡΙΛΗΨΗ

Η αναγνώριση των εκφράσεων του προσώπου και των αντίστοιχων συναισθημάτων που εμφανίζει είναι οικείο γνώρισμα για κάθε άνθρωπο και αποτελεί παγκόσμιο φαινόμενο, χωρίς να χρειάζεται κάποιου είδους εκμάθηση, είναι ενστικτωδώς "προγραμματισμένο" στους ανθρώπους. Μία διεπαφή Ανθρώπου – Υπολογιστή μπορεί να ενισχυθεί κατά πολύ από την αναγνώριση των ανθρώπινων συναισθημάτων απλά κοιτώντας μας (δεν υπάρχει λόγος χρήσης της ίδιας γλώσσας ή γενικότερα λόγου). Αυτό μπορούμε να το πετύχουμε παίρνοντας φωτογραφίες (ή ζωντανό βίντεο) που δείχνουν ανθρώπινα πρόσωπα τα οποία εκφράζουν διαφορετικά συναισθήματα και μετατρέποντας τες σε τόνους γκρίζου (για λόγους ευκολίας και ταχύτητας). Εφαρμόζουμε μια ήδη εκπαιδευμένη "μάσκα" με 68 σημεία (κάθε ένα από αυτά με συγκεκριμένες συντεταγμένες) βασισμένη σε μια μέθοδο που προσπαθεί να προσαρμόσει τα σημεία αυτά γύρω απ'τις βασικές περιοχές του προσώπου που απεικονίζουν λεπτομέρειες συναισθημάτων, δηλαδή γύρω απο τις οπές του (στόμα, μάτια, μύτη). Κατόπιν παίρνουμε όλους τους πιθανούς συνδυασμούς μεταξύ όλων αυτών των σημείων ανά δύο και υπολογίζουμε την Ευκλείδεια απόσταση μεταξύ τους, ενώ χρησιμοποιώντας έναν SVM αλγόριθμο μηχανικής μάθησης εκπαιδεύουμε το σύστημα μας να αναγνωρίζει τέσσερα διαφορετικά συναισθήματα (θυμός, χαρά, ηρεμία, λύπη) βασιζόμενοι σε δύο σετ αρχείων (με παρουσίαση εκφράσεων και αντίστοιχου συναισθήματος). Τα δύο σετ δεδομένων με τα οποία εκπαιδεύουμε και επικυρώνουμε το SVM είναι: το Patras A.I.nD.M. με 84 σκηνοθετημένες φωτογραφίες (πορτραίτο) και το Fer2013 με 960 αυθόρμητες φωτογραφίες (τυχαία γωνία). Το Fer2013 σετ δημιουργήθηκε για έναν διαγωνισμό Αναγνώρισης εκφράσεων με σκοπό να τεστάρει τους αλγόριθμους αναγνώρισης στα όρια τους, με πολύ μικρή ανάλυση 48x48 εικονοστοιχεία και γεμάτο με θολές εικόνες, εμπόδια μπροστά απο το πρόσωπο (είτε απο μαλλιά, καπέλα, γυαλιά, κακή γωνία λήψης κτλπ). Κάποιες απο αυτές δεν είναι καν πραγματικές φωτογραφίες αλλά σκίτσα η ζωγραφίες προσώπων με πληθώρα ανθρώπων απο όλο τον κόσμο. Τέλος μετά την εκπαίδευση, διενεργούμε τεστ απόδοσης για να δούμε πόσο ακριβής είναι. Από τα τέσσερα συναισθήματα η Ευτυχία αναγνωρίστηκε με μεγαλύτερη ακρίβεια με 100% (και στις τρείς στατιστικές μετρήσεις: Sensitivity, Specificity και Accuracy) στο σετ δεδομένων της Πάτρας, και 72%, 75% and 85% (Sen, Spe, Acc) στο Fer2013 σετ δεδομένων αντίστοιχα. Ο *Θυμός* έρχεται δεύτερος με 100%, 83%, 87.5% στο Patras σετ δεδομένων και 56%, 85%, 78% στο Fer2013 σετ δεδομένων. Η *Ηρεμία* είναι τρίτη σε στατιστικά με 50%, 100%, 87.5% στο Patras σετ δεδομένων και 45%, 67% και 62% στο Fer2013 σετ. Τέλος η *Λύπη* με 87.5%, 50% και 83% στο Patras σετ δεδομένων και 21%, 80% και 62% στο Fer2013 σετ δεδομένων.

Λέξεις κλειδιά: εκφράσεις προσώπου, συναίσθημα, ρομποτική όραση, SVM, Dlib, OpenCV, επεξεργασία εικόνας, ταξινόμητής, Fer2013, IBUG 300-W

Table of Contents

CHAPTER 1....9
Introduction – related work....9
1.1) Introduction - Thesis goals....9
1.2) Thesis structure....9
1.3) Related work....10
CHAPTER 2....17
Facial anatomy and emotions....17
2.1) Facial expressions....17
2.2) Physiology of facial expressions....18
2.3) Facial Action Coding System (FACS)....20
2.4) Psychology of emotions - basic categories....21
CHAPTER 3....27
Prerequisite mathematical knowledge....27
3.1) Support Vector Machine (SVM)....27
3.2) The “Kernel trick”....29
3.3) Radial basis function Kernel....30
3.4) Sensitivity, Specificity and Accuracy (evaluation stats)....31
3.5) Norm (Mathematics)....33
3.6) OpenCV library....34
3.7) Dlib library....35
3.8) Face Landmark Detection (I-Bug 300-W)....36
CHAPTER 4....37
Proposed method....37
4.1) Input data (image pre-proccessing)....37
4.2) Analyzing experimental procedure....38
4.3) Input data analysis....41
4.4) Understanding Dlib’s library facial landmark detector....42
4.5) Support vector machine (SVM) training....44
4.6) Support vector machine (SVM) prediction....44
4.7) Support vector machine (SVM) execution results....45
CHAPTER 5....49
Datasets and results analysis....49
5.1) Data sets....49
5.2) Testing and use of data sets....51
5.3) Analyzing results....59
CHAPTER 6....61
Future work and proposed method improvements....61
ACRONYMS....63
LINKS – REFERENCES....65

CHAPTER 1

Introduction – related work

1.1) Introduction - Thesis goals

This thesis studies the detection of human facial expressions using computer vision and tries to predict it's corresponding emotional state (Joy, Anger, Tranquility, Sadness) using a *Support Vector Machine* algorithm. The motivation behind this thesis is that it would be greatly beneficial to human-computer interaction if computers could identify human emotions especially since robotics made major advancements over the last decades with different kinds of robotic smart assistants. It could be used in health robotic assistants checking patients and saving time for the doctors for more serious cases in Hospitals, or domestic assistants helping people with disabilities faster and more accurate and it's very practical way of "training" computers to best help us in our needs based on our reactions (feedback) just by "looking" at us. The proposed method is to train a Support Vector Machine program written in C++ with data sets of Images of facial expression of different people and their corresponding emotion by analyzing them , locating the facial area, applying 68 landmarks around the three orifices of the human face, calculating the distances of each and every two flag/point combination and making one clasifier for each different emotion based on these stats with the help of a Gausian Kernel. We use two data sets that I managed to acquire for the training and verification of the Support Vector Machine (SVM). These data sets were created for Facial Recognition competitions in order to test different recognition algorithms to their limit.

1.2) Thesis structure

Chapter 1 Introduction of the Thesis and similar projects.

Chapter 2 Introduces the facial expressions from a physiological and psychological points of view.

Chapter 3 Explains the basic concepts of the tools we are using and some Insight in the Math needed.

Chapter 4 Image pre-proccessing and how the proposed method works in steps.

Chapter 5 Datasets and proposed method's statistical results analysis.

Chapter 6 Future work and improvements/upgrades over proposed method.

1.3) Related work

1.3.1) <u>Facial Expression Recognition using SVM classification in Perceptual Color Space</u> *by Ms. Aswathy.R, Department of Computer Science & Engineering, Nehru College of Engineering & Research Center, Thrissur, Kerala, India.*

Facial expression analysis is an important area of Human Robot Interaction (HRI) because facial expressions represent human emotions. Here, a new facial expression recognition system is introduced which uses tensor concept. Here perceptual color space is used instead of RGB color space since it cannot work well with illumination and pose variations. Also for classification purpose SVM classifier is used. The experimental results are compared using accuracy and the proposed method shows significant improvement in terms of these factors.

1.3.2) <u>Emotion classification using Adaptive SVMs</u> *by P. Visutsak, International Journal of Computer and Communication Engineering, Vol. 1, No. 3.*

The study of the interaction between humans and computers has been emerging during the last few years. This interaction will be more powerful if computers are able to perceive and respond to human nonverbal communication such as emotions. In this study, we present the image-based approach to emotion classification through lower facial expression. We employ a set of feature points in the lower face image according to the particular face model used and consider their motion across each emotive expression of images. The vector of displacements of all feature points input to the Adaptive Support Vector Machines (A-SVMs) classifier that classify it into seven basic emotions scheme, namely neutral, angry, disgust, fear, happy, sad and surprise. The system was tested on the Japanese Female Facial Expression (JAFFE) dataset of frontal view facial expressions. Our experiments on emotion classification through lower facial expressions demonstrate the robustness of Adaptive SVM classifier and verify the high efficiency of our approach.

1.3.3) ***Real-time Facial Expression Recognition from image sequences using Support Vector Machines*** *by I. Kotsia and I. Pitas, Aristotle University of Thessaloniki, Department of Informatics.*

In this paper, a real-time method is proposed as a solution to the problem of facial expression classification in video sequences. The user manually places some of the Candide grid nodes to the face depicted at the first frame. The grid adaptation system, based on deformable models, tracks the entire Candide grid as the facial expression evolves through time, thus producing a grid that corresponds to the greatest intensity of the facial expression, as shown at the last frame. Certain points that are involved into creating the Facial Action Units movements are selected. Their geometrical displacement information, defined as the coordinates' difference between the last and the first frame, is extracted to be the input to a six class Support Vector Machine system. The output of the system is the facial expression recognized. The proposed real-time system, recognizes the 6 basic facial expressions with an approximately 98% accuracy.

1.3.4) ***A real-time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor Classifier*** *by Hend Ab. ELLaban, A. A. Ewees, Elsaeed E. AbdElrazek, Computer Inst. Prep. Dep. Damietta University Egypt.*

Faces are a unique feature of human being that can detect a great deal of information about age, health, personalities and feelings. Facial Expressions are the main sources in determining the internal impressions of the individual. RealTime system for facial expression recognition is able to detect and locate human faces in image sequences obtained in real environments then extracts expression features from these images finally recognize facial expressions. In this paper, the proposed system presents a real-time system for facial expression recognition that aims to recognize 8 basic facial expressions of students: anger, disgust, fear, happy, nervous, sad, surprise and natural inside E-learning environment. The primary objective is to use k-NN and SVM classifiers to test the efficiency of the proposed system and compared the results of them. There are some techniques has been used in this study for facial expression recognition such as Viola-Jones approaches to detect a face from images, Gabor Feature approach to extract features, and Principal Component Analysis (PCA) to select features and k-NN, SVM classifiers to recognize expressions from facial image.. The result showed that the SVM classifier has the best recognition rate in general thank-NN classifier. From these results, it can say that SVM classifier is more suitable for recognition of facial expression in a real-time system.

1.3.5) <u>Facial Expression Recognition through machine learning</u> *by Nazia Perveen, Nazir Ahmad, M. Abdul Qadoos Bilal Khan, Rizwan Khalid, Salman Qadri, International Journal of Scientific & Technology Research Volume 5, Issue 03.*

Facial expressions communicate non-verbal cues, which play an important role in interpersonal relations. Automatic recognition of facial expressions can be an important element of normal human-machine interfaces; it might likewise be utilized as a part of behavioral science and in clinical practice. In spite of the fact that people perceive facial expressions for all intents and purposes immediately, solid expression recognition by machine is still a challenge. From the point of view of automatic recognition, a facial expression can be considered to comprise of disfigurements of the facial parts and their spatial relations, or changes in the face's pigmentation. Research into automatic recognition of the facial expressions addresses the issues encompassing the representation and arrangement of static or dynamic qualities of these distortions or face pigmentation. We get results by utilizing the CVIPtools. We have taken train data set of six facial expressions of three persons and for train data set purpose we have total border mask sample 90 and 30% border mask sample for test data set purpose and we use RST- Invariant features and texture features for feature analysis and then classified them by using k-Nearest Neighbor classification algorithm. The maximum accuracy is 90%.

1.3.6) <u>Emotion recognition through facial expression analysis based on a Neurofuzzy Network</u> *by Spiros V. Ioannou, Amaryllis T. Raouzaiou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, Stefanos D. Kollias, Article:Literature review in Neural Networks.*

Extracting and validating emotional cues through analysis of users' facial expressions is of high importance for improving the level of interaction in man machine communication systems. Extraction of appropriate facial features and consequent recognition of the user's emotional state that can be robust to facial expression variations among different users is the topic of this paper. Facial animation parameters (FAPs) defined according to the ISO MPEG-4 standard are extracted by a robust facial analysis system, accompanied by appropriate confidence measures of the estimation accuracy. A novel neurofuzzy system is then created, based on rules that have been defined through analysis of FAP variations both at the discrete emotional space, as well as in the 2D continuous activation-evaluation one. The neurofuzzy system allows for further learning and adaptation to specific users' facial expression characteristics, measured though FAP estimation in real life application of the system, using analysis by clustering of the obtained FAP values. Experimental studies with emotionally expressive datasets, generated in the EC IST ERMIS project indicate the good performance and potential of the developed technologies.

1.3.7) <u>Emotion recognition from facial expression based on Benzier curve</u> *by Shruti Bansal, Pravin Nagar, International Journal of Advanced Information Technology (IJAIT) Vol. 5, No. 3/4/5/6.*

Human emotions are conveyed by different medium such as behaviours, actions, poses, facial expressions and speech. Multitudinous researches have been carried out to find out the relation between these mediums and emotions. This paper proposes a system which automatically recognizes the emotion represented on a face. Thus, a Bezier curve based solution together with image processing is used in classifying the emotions. Coloured face images are given as input to the system. Then, Image processing based feature point extraction method is applied to extract a set of selected feature points. Finally, extracted features like eyes and mouth, obtained after processing is given as input to the curve algorithm to recognize the emotion contained. Experimental results show average 60% of success to analyze and recognize emotion detection.

1.3.8) <u>Emotion recognition from facial expression based on fiducial points detection and using Neural Network</u> *by Fatima Zahra Salmam, Abdellah Madani, Mohamed Kissi, Article in International Journal of Electrical and Computer Engineering 2018.*

The importance of emotion recognition lies in the role that emotions play in our everyday lives. Emotions have a strong relationship with our behavior. Thence, automatic emotion recognition, is to equip the machine of this human ability to analyze, and to understand the human emotional state, in order to anticipate his intentions from facial expression. In this paper, a new approach is proposed to enhance accuracy of emotion recognition from facial expression, which is based on input features deducted only from fiducial points. The proposed approach consists firstly on extracting 1176 dynamic features from image sequences that represent the proportions of euclidean distances between facial fiducial points in the first frame, and faicial fiducial points in the last frame. Secondly, a feature selection method is used to select only the most relevant features from them. Finally, the selected features are presented to a Neural Network (NN) classifier to classify facial expression input into emotion.

1.3.9) <u>Automatic Facial Expression Analysis and Emotional Classification</u> *by Robert Fischer, Submitted to the Department of Math and Natural Sciences MIT – Thesis.*

In this thesis, a system for automatic facial expression analysis was designed and implemented. This system includes a monocular 3d head pose tracker to handle rigid movements, feature extraction based on Gabor wavelets and gradient orientation histograms, and a SVM classifier. Further more, a database with video sequences of acted and natural facial expression was compiled and tests were done, including comparisons to other systems.

1.3.10) <u>Emotional Facial Expression Recognition & Classification</u> *by Galateia Iatraki, Master of Science in Media and Knowledge Engineering TUDelft University – Thesis.*

This thesis is focused on the recognition of emotional facial expressions, finding interrelation between the facial expressions and labels and finally the classification of the expressions. One of the parts of this project is an emotional database which will contain images of faces, their corresponding Action Units and their labels. The contribution of this database to the problem stated above is that it can be used by systems in order to recognize emotional facial expressions given one of the database data i.e. action units' combination. The other part of the project, which is an expert system for emotional classification, will enable to classify emotional expressions, the ones included in the database and all the possible combinations of them.

1.3.11) <u>Emotion Recognition Using Facial Expression</u> *by Santosh Kumar, Shubam Jaiswal, Rahul Kumar, Sanjay Kumar Singh, Innovative Research in Attention Modeling and Computer Vision.*

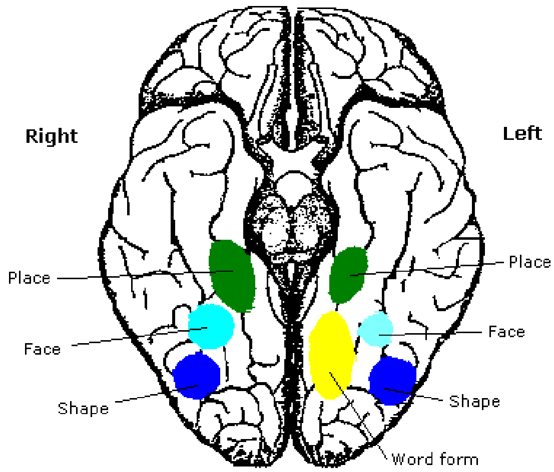
Recognition of facial expression is a challenging problem for machine in comparison to human and it has encouraged numerous advanced machine learning algorithms. It is one of the methods for emotion recognition as the emotion of a particular person can be found out by studying his or her facial expressions. In this paper, we proposes a generic algorithms for recognition of emotions and illustrates a fundamental steps of the four algorithms such as Eigenfaces (Principal Component Analysis [PCA]), Fisherfaces, Local Binary Pattern Histogram (LBP) and SURF with FLANN over two databases Cohn-kanade database and IIT BHU student face images as benchmark database.The objective of this book chapter is to recognize the emotions from facial images of individuals and compare the performances of holistic algorithms like Eigenfaces, Fisherfaces, and texture based recognition algorithms LBPH, hybrid algorithm SURF and FLANN. Matching efficiency of individual emotions from facial expression databases are labeled for training and testing phases. The set of features is extracted from labeled dataset for training purpose and test images are matched with discriminative set of feature points. Based on that comparison, we conclude that Eigenfaces and Fisherfaces yields good recognition accuracy on the benchmark database than others and the efficiency of SURF with FLANN algorithm can be enhanced significantly by changing the parameters.

CHAPTER 2

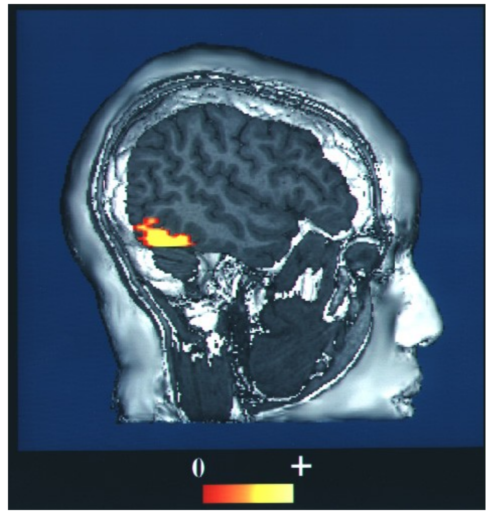
Facial anatomy and emotions

2.1) Facial expressions

The face is the feature which best distinguishes a person. Specialized regions of the human brain, such as the fusiform face area (FFA), enable facial recognition; when these are damaged, it may be impossible to recognize faces even of intimate family members. The pattern/shape of specific organs, such as the eyes, is used in biometric identification(a process the brain does instinctively) to uniquely identify individuals. The face is the only area on the human body that is designed to show emotions which are very important for our non-verbal social communication skills. The human face is a remarkable and complex structure of groups of muscles which work in specific distinct combinations and formations for each emotional state. There are a lot of research studies about the universality of facial expressions and the emotion they represent across the globe. Dr. Paul Ekman has one of the most thorough studies on the matter based on the research among others of Charles Darwin's "The Universality Hypothesis" and his own observations of an isolated tribe in Nea Guinea. All evidence points that the way humans show emotions with their facial expressions is the same across the world irrelevant of knowledge, race, age, sex, geography and location, it's instinctive.



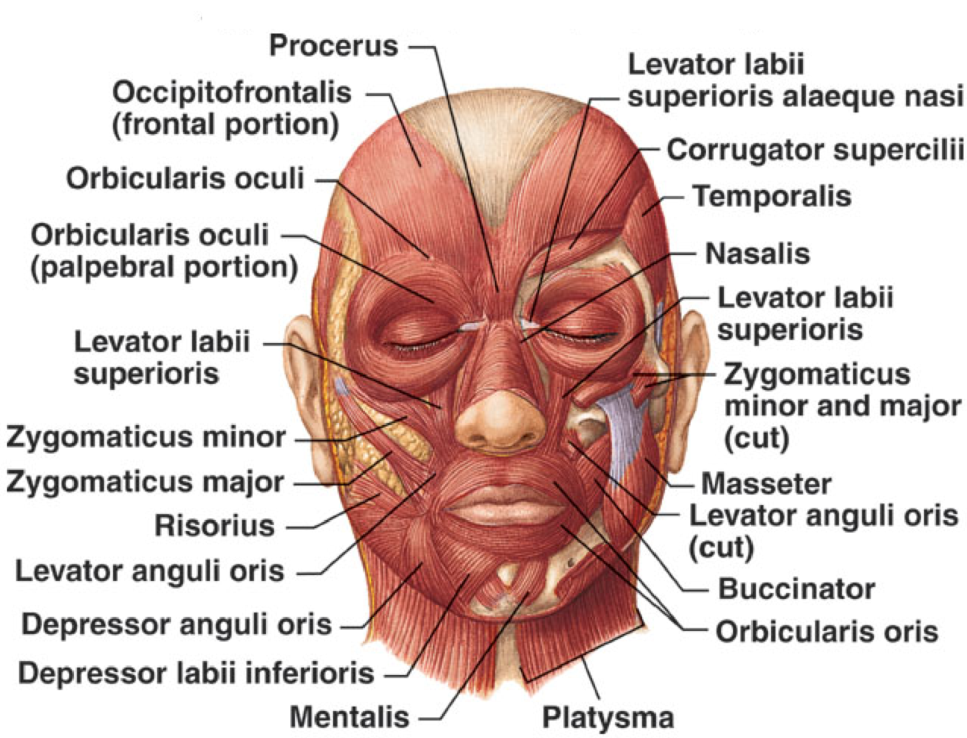
Picture 2.1 - human brain, bottom view.
Fusiform face area shown in bright blue.



Picture 2.2 – *c*omputer-enhanced fMRI scan of a person who has been asked to look at faces. The image shows increased blood flow in cerebral cortex that recognizes faces (FFA).

2.2) Physiology of facial expressions

These complex facial expressions are made possible by a unique group of muscles called *facial expression muscles* or *mimetic muscles* and are part of our head muscles that also contain muscles of the scalp, mastication (responsible for the movement of our jaw) etc. These muscles are attached to the front cranial bone structure and the skin behind our face and are controlled and connected to the Facial Nerve responsible for their contractions in order to show expressions on our face. They are located around the orifices on our face and divided in groups depending on their area of effect (mouth, eyes, nose). The muscles in each group work together to control the corresponding orifice and also take their name from it: Oral, Nasal, Orbital. In more detail, the human face consists of about 20 flat skeletal muscles. The muscles are located under the skin, attached to the skull bone and inserting the facial skin, but not the bones or joints as other muscles responsible for body movements do. However, unlike other Cranial muscles, they do not move with joints and bones, but mainly with the skin. Consequently, they cause facial surface deformations, which result in a variable facial expression representing emotions. The oral muscles alter the shape of the oral orifice. This group is responsible for complex mouth motions and allows sophisticated shaping of the mouth, e.g. encircle the mouth, control angle of the mouth, elevate or depress the lower and upper lip separately or lift and depress the left and right corner or move with cheeks. The nasal group is responsible for the compression and opening of the nostrils. One of the muscles, critical for facial expressions, is located between the eyebrows and pulls the eyebrows downwards and causes wrinkles over the nose.



Picture 2.3 - muscles of facial expressions

The orbital group of three muscles is primarily responsible for the motion of the eyelid and protecting eyes. All the muscles are inserting the skin around the eyebrows and form vertical wrinkles between the eyebrows.

2.3) Facial Action Coding System (FACS)

The Facial Action Coding System (FACS), developed by Dr. Paul Ekman, is a tool describing every facial muscle activity motion by a set of Action Units (AUs). A particular AU represents a certain component of facial muscles movement. Each emotion in the face can be described by a set of AUs, see Table 2.1, 2.2.

AU	Description	Facial muscle
1	Inner Brow Raiser	Frontal Belly (Epicranius)
2	Outer Brow Raiser	Frontal Belly (Epicranius)
4	Brow Lowerer	Frontal Belly (Epicranius)
5	Upper Lid Raiser	Frontal Belly (Epicranius)
6	Cheek Raiser	Orbicularis Oculi
7	Lid Tightener	Orbicularis Oculi
9	Nose Wrinkler	Levator Labii Superioris
10	Upper Lip Raiser	Levator Labii Superioris
11	Nasolabial Deepener	Zygomaticus Minor
12	Lip Corner Puller	Zygomaticus Major
13	Cheek Puffer	Levator Anguli Oris
14	Dimpler	Buccinator
15	Lip Corner Depressor	Depressor Anguli Oris
16	Lower Lip Depressor	Depressor Labii Inferioris
17	Chin Raiser	Mentalis
20	Lip stretcher	Risorius w/ Platysma

Table 2.1 - description of several AUs together with the muscle name.

Emotion	AU
Anger	4, 5, 7
Disgust	9, 15, 16
Fear	1, 2, 7, 20
Joy	6, 12
Sadness	1, 15
Surprise	1, 2

Table 2.2 **- b**asic AUs connected with emotions.

2.4) Psychology of emotions - basic categories

For more than 40 years, Paul Ekman has supported the view that emotions are discrete, measurable, and physiologically distinct. Ekman's most influential work revolved around the finding that certain emotions appeared to be universally recognized, even in cultures that were preliterate and could not have learned associations for facial expressions through media. Another classic study found that when participants contorted their facial muscles into distinct facial expressions (for example, disgust), they reported subjective and physiological experiences that matched the distinct facial expressions. His research findings led him to classify six basic emotions as:
anger, disgust, fear, joy, sadness and surprise.

2.4.1 Anger

Anger or **wrath** is an intense emotional response usually involving agitation, malice, or retribution. It is an emotion that involves a strong uncomfortable and hostile response to a perceived provocation, hurt or threat. Anger can occur when a person feels their personal boundaries are being or are going to be violated. Some have a learned tendency to react to anger through retaliation as a way of coping. Anger is an emotional reaction that impacts the body. A person experiencing anger will also experience physical conditions, such as increased heart rate, elevated blood pressure, and increased levels of adrenaline and nor-adrenaline. Some view anger as an emotion which triggers part of the fight or flight brain response. Anger is used as a protective mechanism to cover up fear, hurt or sadness. Anger becomes the predominant feeling behaviorally, cognitively, and physiologically when a person makes the conscious choice to take action to immediately stop the threatening behavior of another outside force.

Anger has a substantial impact on the whole body.

- The increase of the blood pressure, red face, and tension in the muscles are usually reflected.
- The eyebrows are lowered and squeezing together. There are vertical wrinkles between the eyebrows.
- The eyelids are tight and straight.
- The eyes are tight, focusing on the source of anger. Pupils are narrowed, focusing on the source of anger.
- The lips are either closed tight or gently opened (preparing for yelling).

2.4.2 Disgust

Disgust is an emotional response of revulsion to something considered offensive, distasteful, or unpleasant. In *The Expression of the Emotions in Man and Animals*, Charles Darwin wrote that disgust is a sensation that refers to something revolting. Disgust is experienced primarily in relation to the sense of taste (either perceived or imagined), and secondarily to anything which causes a similar feeling by sense of smell, touch, or vision. Musically sensitive people may even be disgusted by the cacophony of inharmonious sounds. Research continually has proven a relationship between disgust and anxiety disorders such as arachnophobia, blood-injection-injury type phobias, and contamination fear related obsessive–compulsive disorder (also known as OCD). Disgust is one of the basic emotions of Robert Plutchik's theory of emotions and has been studied extensively by Paul Rozin.

Unlike the emotions of fear, anger, and sadness; disgust is associated with a decrease in heart rate. Effects on the body:

- The upper lip is lifted.
- There are wrinkles on the nose.
- The cheeks are lifted.
- The eyelids are lifted but are not tight. There are wrinkles under the eyes.
- The eyebrows are pulled down.

2.4.3 Fear

Fear is a feeling induced by perceived danger or threat that occurs in certain types of organisms, which causes a change in metabolic and organ functions and ultimately a change in behavior, such as fleeing, hiding, or freezing from perceived traumatic events. Fear in human beings may occur in response to a specific stimulus occurring in the present, or in anticipation or expectation of a future threat perceived as a *risk* to body or life. The fear response arises from the perception of danger leading to confrontation with or escape from avoiding the threat (also known as the fight-or-flight response), which in extreme cases of fear (horror and terror) can be a freeze response or paralysis.

Effects on the body:

- The eyebrows are lifted and pulled inward.
- There are wrinkles on the forehead.
- The upper eyelids are lifted and eyes widen up.

- The mouth is open, and the lips are tight according to the intensity of the emotion.
- Adrenaline rushes through the body heightening all senses (fight-or-flight response).

2.4.4 Joy

In philosophy, **Joy - Happiness** is translated from the Greek concept of *eudaimonia*, and refers to the good life, or flourishing, as opposed to an emotion. In psychology, happiness is a mental or emotional state of well-being which can be defined by, among others, positive or pleasant emotions ranging from contentment to intense joy. Happy mental states may reflect judgements by a person about their overall well-being.

Effects on the body:

- The lips corners are pulled back and up.
- The mouth can be open, and the teeth might be visible.
- The cheeks can be raised.
- The wrinkles under the lower eyelid might appear.
- The wrinkles appear outside the eye corners.

2.4.5 Sadness

Sadness is an emotional pain associated with, or characterized by, feelings of disadvantage, loss, despair, grief, helplessness, disappointment and sorrow. An individual experiencing sadness may become quiet or lethargic, and withdraw themselves from others. An example of severe sadness is depression. Crying can be an indication of sadness. During the emotion, the facial muscles lose the tension which may result in typical physiological features.

Effects on the body:

- The inner parts of the eyebrows are pulled down.
- Lips corners pull down, and lips shake.

2.4.6 Surprise

Surprise is a brief mental and physiological state, a startle response experienced by animals and humans as the result of an unexpected event. Surprise can have any valence; that is, it can be neutral/moderate, pleasant, unpleasant, positive, or negative. Surprise can occur in varying levels of intensity ranging from very-surprised, which may induce the fight-or-flight response, or little-surprise that elicits a less intense response to the stimuli.

- Eyebrows that are raised so they become curved and high.
- Horizontal wrinkles across the forehead.
- Open eyelids: the upper lid is raised and the lower lid is drawn down, often exposing the white sclera above and below the iris.
- Pupil dilation mydriasis or pupil constriction miosis.
- Dropped jaw so that the lips and teeth are parted, with no tension around the mouth.

CHAPTER 3

Prerequisite mathematical knowledge (tools used in this Thesis)

3.1) Support Vector Machine (SVM)

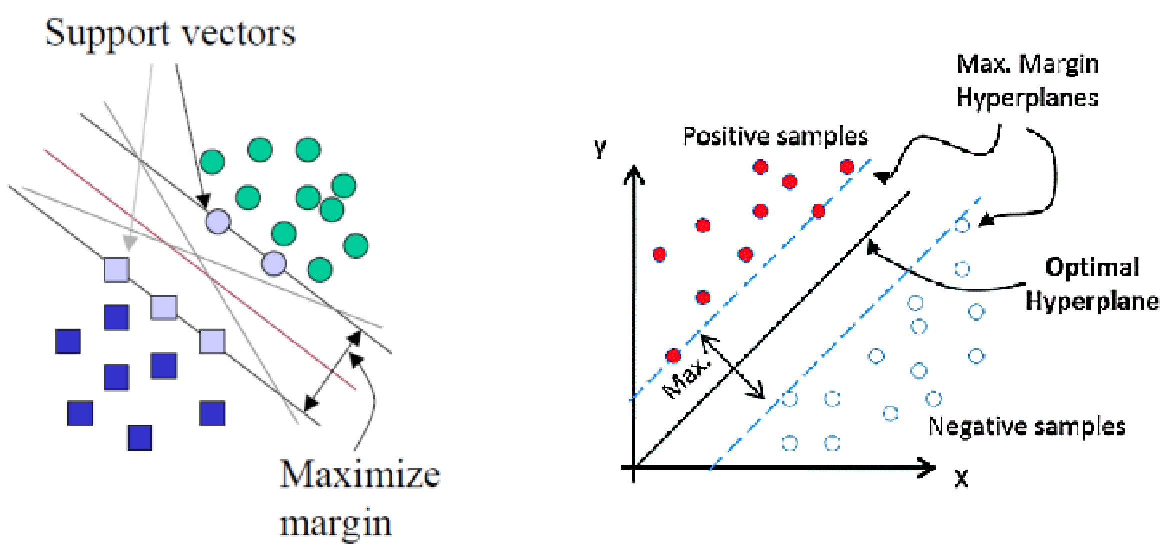
Machine learning is a computer science field that gives computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Unsupervised Machine learning can be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

In machine learning, **Support Vector Machines** (**SVMs**, also **support vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick (which is what we chose), implicitly mapping their inputs into high-dimensional feature spaces.

Unlike other classifiers, the support vector machine is ***explicitly*** told to find the best separating line. How? The support vector machine searches for the closest points (Picture 3.1), which it calls the "support vectors" (the name "support vector machine" is due to the fact that points are like vectors and that the best line "depends on" or is "supported by" the closest points).

Once it has found the closest points, the SVM draws a line connecting them. It draws this connecting line by doing vector subtraction (point A - point B). The support vector machine then declares the best separating line to be the line that bisects -- and is perpendicular to -- the connecting line.

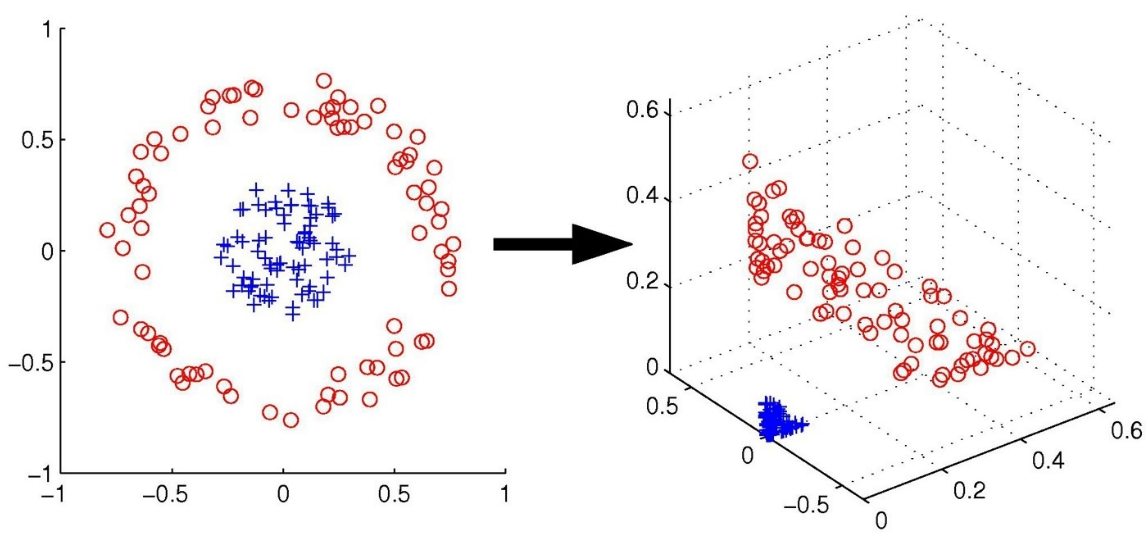
The support vector machine is better because when you get a new sample (new points), you will have already made a line that keeps B and A as far away from each other as possible, and so it is less likely that one will spillover across the line into the other's territory.



Picture 3.1 – *v*isual examples of the core idea of how SVM works.

3.2) The “Kernel trick”

In machine learning, **kernel methods** are a class of algorithms for pattern analysis, whose best known member is the support vector machine (SVM). The general task of pattern analysis is to find and study general types of relations (for example clusters, rankings, principal components, correlations, classifications) in datasets. In its simplest form, the kernel trick means transforming data into another dimension that has a clear dividing margin between classes of data. For many algorithms that solve these tasks, the data in raw representation have to be explicitly transformed into feature vector representations via a user-specified *feature map*: in contrast, kernel methods require only a user-specified *kernel*, i.e., a similarity function over pairs of data points in raw representation.
Kernel methods owe their name to the use of kernel functions, which enable them to operate in a high-dimensional, *implicit* feature space without ever computing the coordinates of the data in that space, but rather by simply computing the inner products between the images of all pairs of data in the feature space. This operation is often computationally cheaper than the explicit computation of the coordinates. This approach is called the "**kernel trick**". Kernel functions have been introduced for sequence data, graphs, text, images, as well as vectors.



Picture 3.2 - example of a labeled data inseparable in 2-Dimension is separable in 3-Dimension.

3.3) Radial basis function Kernel

In machine learning, the **radial basis function kernel**, or **RBF kernel**, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

The RBF kernel on two samples **x** and **x'**, represented as feature vectors in some *input space*, is defined as

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \qquad \textbf{\textit{(3.1)}}$$

$\|\mathbf{x} - \mathbf{x}'\|^2$ may be recognized as the squared Euclidean distance between the two feature vectors. “σ” is a free parameter. An equivalent, but simpler, definition involves a parameter:

$$\gamma = \tfrac{1}{2\sigma^2}: \quad K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2) \qquad \textbf{\textit{(3.2)}}$$

Since the value of the RBF kernel decreases with distance and ranges between zero (in the limit) and one (when **x** = **x'**), it has a ready interpretation as a similarity measure. The feature space of the kernel has an infinite number of dimensions; for σ = 1 , its expansion is:

$$\begin{aligned}\exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|^2\right) &= \sum_{j=0}^{\infty} \frac{(\mathbf{x}^\top \mathbf{x}')^j}{j!} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \exp\left(-\frac{1}{2}\|\mathbf{x}'\|^2\right) \\ &= \sum_{j=0}^{\infty} \sum_{\sum n_i = j} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \frac{x_1^{n_1} \cdots x_k^{n_k}}{\sqrt{n_1! \cdots n_k!}} \exp\left(-\frac{1}{2}\|\mathbf{x}'\|^2\right) \frac{{x'}_1^{n_1} \cdots {x'}_k^{n_k}}{\sqrt{n_1! \cdots n_k!}} \qquad \textbf{\textit{(3.3)}}\end{aligned}$$

3.4) Sensitivity, Specificity and Accuracy (evaluation stats)

Sensitivity and **specificity** are statistical measures of the performance of a binary classification test, also known in statistics as classification function:

Sensitivity (also called the **true positive rate**, the **recall**, or **probability of detection** in some fields) measures the proportion of positives that are correctly identified as such (e.g. the percentage of facial expressions which are correctly identified with the correct emotion).
Specificity (also called the **true negative rate**) measures the proportion of negatives that are correctly identified as such (e.g. the percentage of facial expressions which are correctly dismissed for not having the emotion we want).

condition positive (P) = the number of real positive cases in the data
condition negative (N) = the number of real positive cases in the data

Match: positive for the specific emotion

Missed: negative for the specific emotion

True positive (TP) = the number of cases *correctly* identified as ***Match***

False positive (FP) = the number of cases *incorrectly* identified as ***Match***

True negative (TN) = the number of cases *correctly* identified as ***Missed***

False negative (FN) = the number of cases *incorrectly* identified as ***Missed***

Sensitivity or True Positive Rate (TPR): The sensitivity of a test is its ability to determine the patient cases correctly. To estimate it, we should calculate the proportion of true positive in patient cases. Mathematically, this can be stated as:

$$Sensitivity = \quad TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$$ *(When it's actually yes, how often does it predict yes?)* ***(3.4)***

Specificity or True Negative Rate (TNR): The specificity of a test is its ability to determine the healthy cases correctly. To estimate it, we should calculate the proportion of true negative in healthy cases. Mathematically, this can be stated as:

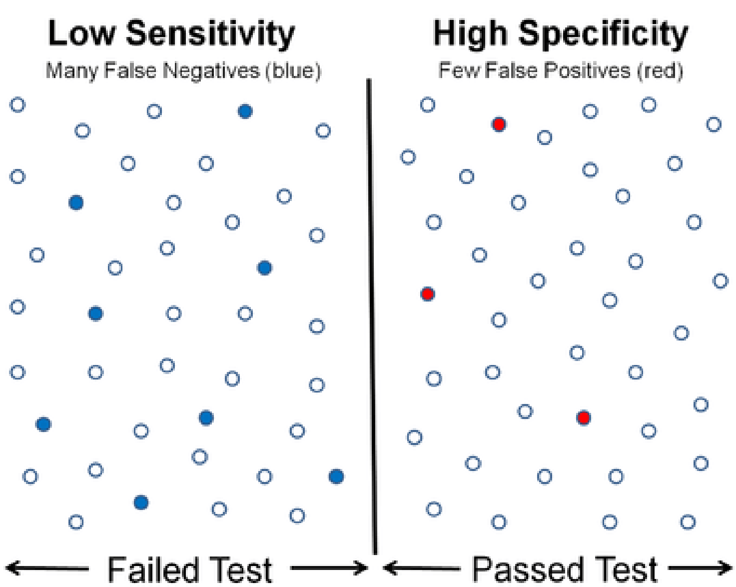
$$Specificity = \quad TNR = \frac{TN}{N} = \frac{TN}{TN+FP}$$ *(When it's actually no, how often does it predict no?)* ***(3.5)***

Accuracy (ACC): The accuracy of a test is its ability to differentiate the Match and Correctly Missed cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \quad ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$$ *(Overall, how often is the classifier correct?)*

(3.6)

example:



Picture 3.4 – graphic example.

3.5) Norm (Mathematics)

In linear algebra, functional analysis, and related areas of mathematics, a **norm** is a function that assigns a strictly positive *length* or *size* to each vector in a vector space—save for the zero vector, which is assigned a length of zero. A **seminorm**, on the other hand, is allowed to assign zero length to some non-zero vectors (in addition to the zero vector).

A norm must also satisfy certain properties pertaining to scalability and additivity which are given in the formal definition below.

A simple example is two dimensional Euclidean space **R**2 equipped with the "Euclidean norm" (the same norm that I am using for the SVM classifier to work) Elements in this vector space are usually drawn as arrows in a 2-dimensional cartesian coordinate system starting at the origin (0, 0). The Euclidean norm assigns to each vector the length of its arrow. Because of this, the Euclidean norm is often known as the magnitude.

- **Euclidean norm:**

On an *n*-dimensional Euclidean space **R***n*, the intuitive notion of length of the vector ***x*** = (*x*1, *x*2, ..., *xn*) is captured by the formula:

$$\|\boldsymbol{x}\| := \sqrt{x_1^2 + \cdots + x_n^2} \qquad \textbf{(3.7)}$$

This gives the ordinary distance from the origin to the point ***x***, a consequence of the Pythagorean theorem. The Euclidean norm is by far the most commonly used norm on **R***n*, but there are other norms on this vector space as will be shown below. However all these norms are equivalent in the sense that they all define the same topology.
On an *n*-dimensional complex space **C***n*
the most common norm is:

$$\|\boldsymbol{z}\| := \sqrt{|z_1|^2 + \cdots + |z_n|^2} = \sqrt{z_1 \bar{z}_1 + \cdots + z_n \bar{z}_n} \qquad \textbf{(3.8)}$$

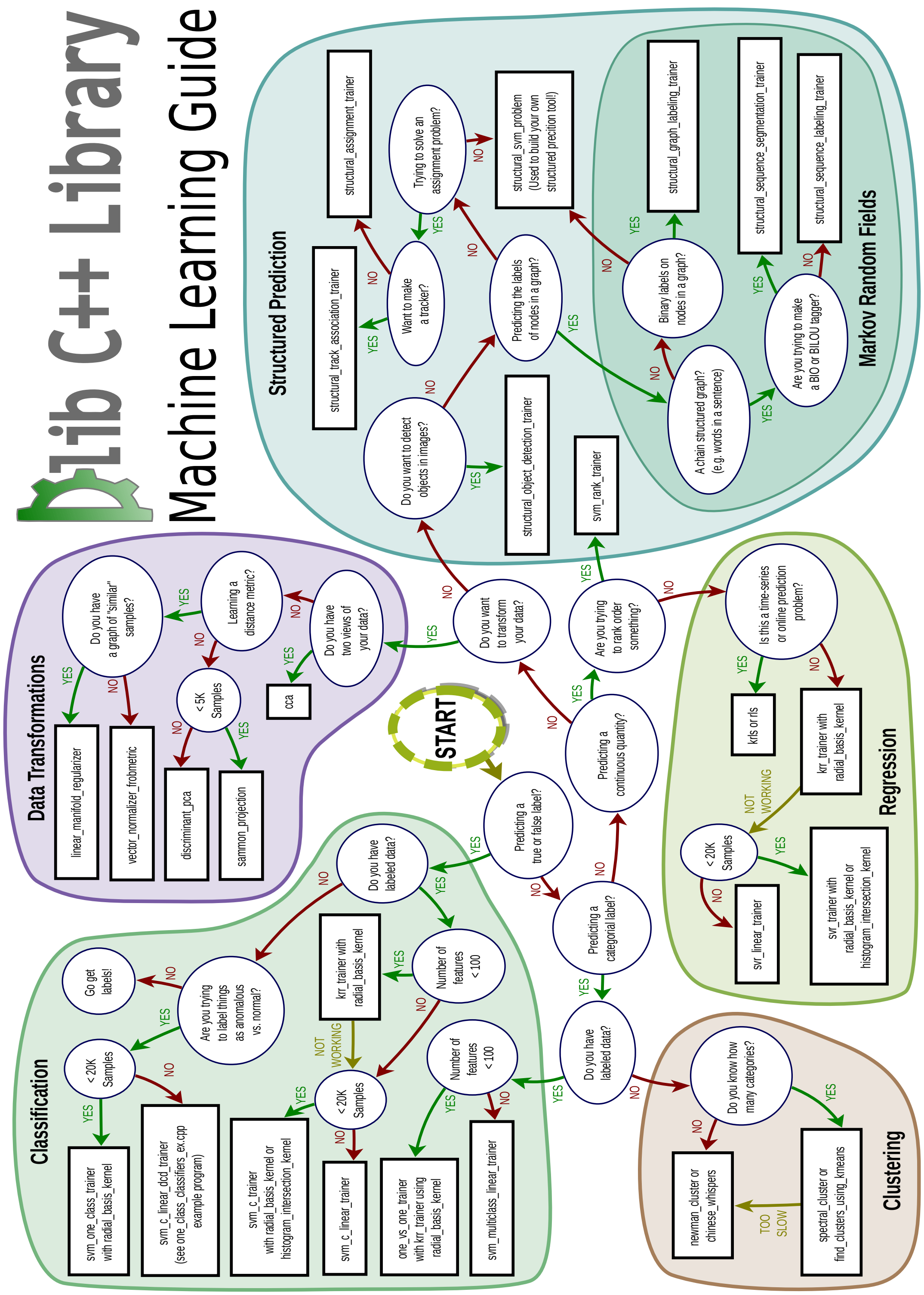
In both cases we can also express the norm as the square root of the inner product of the vector and itself:

$$\|\boldsymbol{x}\| := \sqrt{\boldsymbol{x}^* \ \boldsymbol{x}} \qquad \textit{(3.9)}$$

3.6) OpenCV library

The OpenCV library is a free software developed by Intel designed to offer a more coherent and rich library for image processing and by that extend to Robotic Vision. The development of human-computer interface, the detection, isolation and identification of objects, detection and identification of faces, the prediction and tracking of movement are some of the fields Robotic Vision covers.
The library is cross-platform since it can fully operate , on Windows OS , Mackintosh , Linux , Android ,Solaris and HP-UX. It is also multi-lingual since it is supported by Python and Java apart from C/C++. It’s a popular library widely used that offers a sense of high level programming language and can utilize in a great degree the hardware of the computer, offering high execution speed of complex programs. OpenCV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive though extensive older C interface. There are bindings in Python, Java and MATLAB/OCTAVE. The API for these interfaces can be found in the online documentation. Wrappers in other languages such as C#, Perl, Ch, Haskell and Ruby have been developed to encourage adoption by a wider audience.
All of the new developments and algorithms in OpenCV are now developed in the C++ interface.

3.7) Dlib library



Picture 3.5 - machine learning guide algorithm.

Dlib is an open code library that contains Machine Learning algorithms and tools for the creation of complex software based around Robotic Vision. Main writer of this library was Davis King from the very first version but also during her development the year 2002. This library is offered free on the internet with plenty of examples of how to use it' algorithms and tools. As mentioned it was initially written in C++ language, while in it's latest versions it also supports Python, in addition to that it's cross-platform since it can fully run on Windows OS, Mackintosh, Linux, Android, Solaris and HP-UX. Today it contains software elements for networks, thread and process handling, Graphical User Interfaces (GUI's), complex data structures, linear Algebra, Machine Learning, image processing, data mining, optimization of arithmetic operations and many more programming elements.

3.8) Face Landmark Detection (I-Bug 300-W)

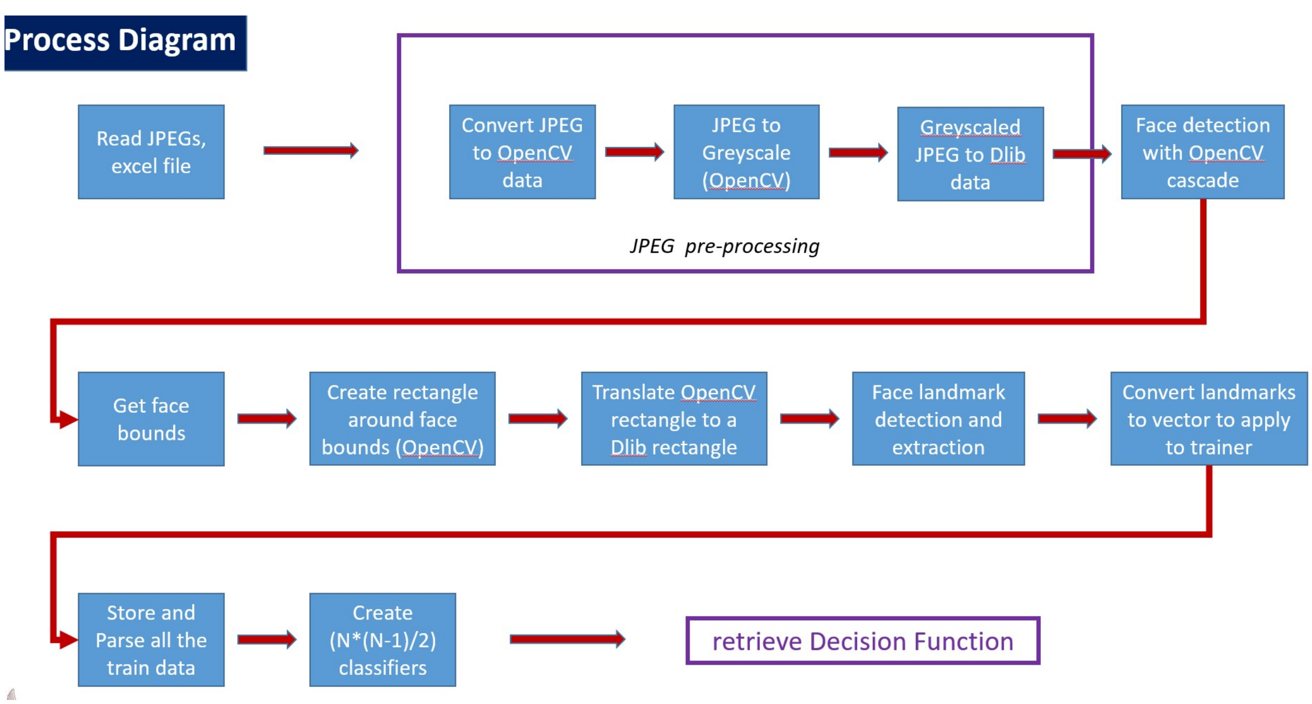
The face detector we use is made using the classic Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid, using dlib's implementation of the paper: One Millisecond Face Alignment with an Ensemble of Regression Trees byVahid Kazemi and Josephine Sullivan, CVPR 2014 and was trained on the iBUG 300-W face landmark data set. The 300-W, full name :Faces in-the-Wild Challenge: The first facial landmark localization Challenge held in conjunction with the International Conference on Computer Vision 2013, Sydney. The main challenge of the competition was to localize a set of 68 fiducial points in a newly collected test set with 2x300 facial images captured in real-world unconstrained settings (300 'Indoor' and 300 'Outdoor'). As a part of the challenge the most well known databases XM2VTS, LFPW, HELEN, and AFW were reannotated using the same mark-up.

CHAPTER 4

Proposed method

4.1) Input data (image pre-proccessing)

For our SVM Machine to work first we have to train it and in order to do that we need to provide it with some data, a data set of images each containing a single face from frontal perspective and an excel file with information (name and emotion code) about each picture used as input for the training process of the SVM. The pictures need an extra processing-modification before they can be used efficiently as input for the training process and also afterwards for the evaluation and correct prediction of each emotion shown on the images. To do that we need to take the info we want by throwing away any unnecessary data so that we can do the process of guessing fast so that it's "Live" in case of video feed



Picture 4.1 – proposed method process diagram.

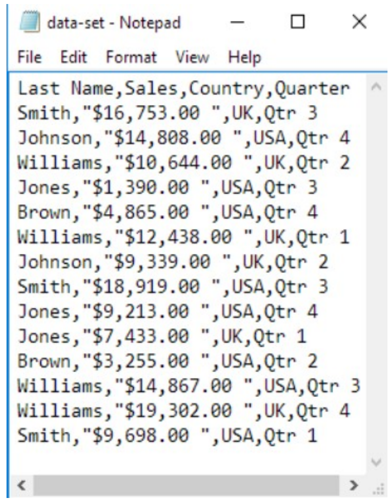
4.2) Analyzing experimental procedure

The experimental procedure of exporting an emotional state with the use of computer vision from a human face is explained below. The goal of this process, is to create a simple tool, which can be executed from the command line and takes as input a group of pictures (each of them showing a frontal view of a human face) in order to “train” the support vector machine, which generates as output a Decision Function(SVM Machine) which can be tested in different sets of pictures not only for its efficiency and evaluation but also for its practical use. The pattern which the above report is based on, comes next:

$$\{Group\,of\;pictures\,K\,,K\in Z\}\rightarrow SVM\;Classifier\rightarrow Machine$$
$$Machine\rightarrow SVM\;Classifier(picture)\rightarrow\;prediction$$

Grouping of input data

To start with, we must make a simple mechanism to import data to the program of the whole process. That is why a simple *.csv file is being used where all the program’s necessary input data are organized, meaning the names of each picture file that will be used for the Machine’s learning process and the emotional state present in each picture. We choose this format because a .csv file has quite a simple structure for grouping and presenting data (being read by a program). Each line of that kind of file is a new entry of data and each row (separated by the character “,”) is a new data element.




```
Last Name,Sales,Country,Quarter
Smith,"$16,753.00 ",UK,Qtr 3
Johnson,"$14,808.00 ",USA,Qtr 4
Williams,"$10,644.00 ",UK,Qtr 2
Jones,"$1,390.00 ",USA,Qtr 3
Brown,"$4,865.00 ",USA,Qtr 4
Williams,"$12,438.00 ",UK,Qtr 1
Johnson,"$9,339.00 ",UK,Qtr 2
Smith,"$18,919.00 ",USA,Qtr 3
Jones,"$9,213.00 ",USA,Qtr 4
Jones,"$7,433.00 ",UK,Qtr 1
Brown,"$3,255.00 ",USA,Qtr 2
Williams,"$14,867.00 ",USA,Qtr 3
Williams,"$19,302.00 ",UK,Qtr 4
Smith,"$9,698.00 ",USA,Qtr 1
```

Picture 4.2 - structure example of a .csv file.

In the present thesis the process is based on a file that in it's first row lies the name of the picture file and next to each name there is another row with a number that corresponds to the feeling presented in that picture. The final structure of the file is described next:

$¿ Image_{Fil} e_1 . jpg>,<emotional\ state\ code>¿$

$¿ Image_{Fil} e_2 . jpg>,<emotional\ state\ code>¿$

$\ldots$

$¿ Image_{Fil} e_{N-1} . jpg>,<emotional\ state\ code>¿$

$¿ Imag\, e_{Fil\, e_N} . jpg>,<emotional\ state\ code>¿$

At this stage it is worth mentioning the four different emotional states that are being recognized and are being used to train the SVM, which are being presented on the board below.

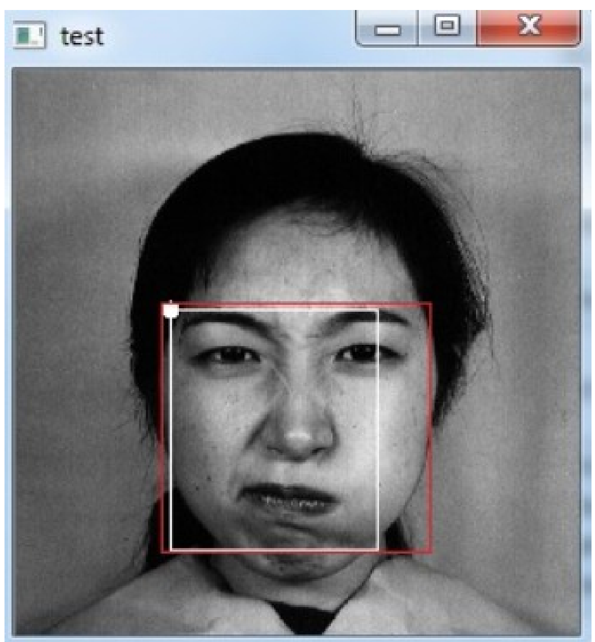
Emotion	Example	Code number
Anger		0
Joy		1
Tranquility		2
Sadness		3

Table 4.1 – code numbers of each emotion.

Based on that pattern, a data catalog is being created with the above structure, by the User with all the correct data which in turn is given as input to our Machine. When the file is being read, two Vectors are created in which the information of the picture (name) and the emotion code are being saved. At this point it should be noted that every group of data of this project is defined by the above format.

4.3) Input data analysis

To analyze the program's input data, the OpenCV and Dlib libraries are being used. Initially we want to read and load the real data for each picture file from all the input data to the OpenCV format. The OpenCV library contains all the tools for loading pictures(images) in real time (like real time video feed from a camera) and storing them in registries , which the library can identify and use. Each picture is being loaded based on the above procedure and then is modified from colored to grayscale , in order to optimize the results during the face scanning phase (with the OpenCV face detectors before is sent to the Dlib) where the color factor is being removed. We use the OpenCV library to modify each colored picture to a grayscale one and then the data registries described above are being translated in Dlib library registries, the Dlib library is then used to map-scan the face. After that we use the OpenCV library and its Haar Classifier to specify(pinpoint) the area of the face inside the picture.

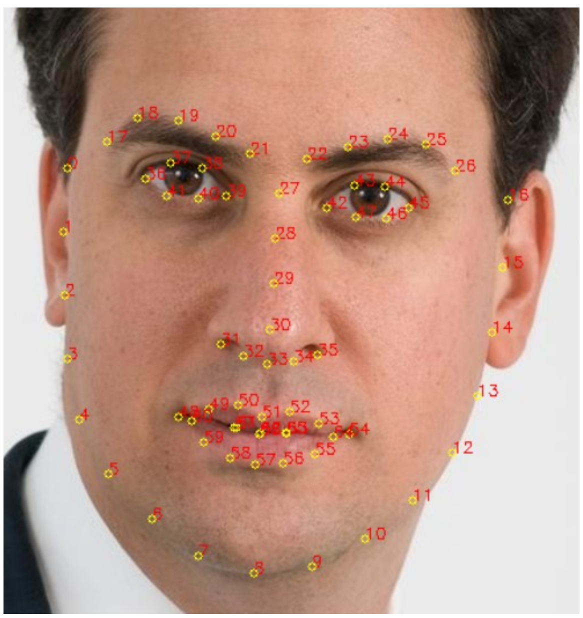


Picture 4.3 – example of face isolation procedure.

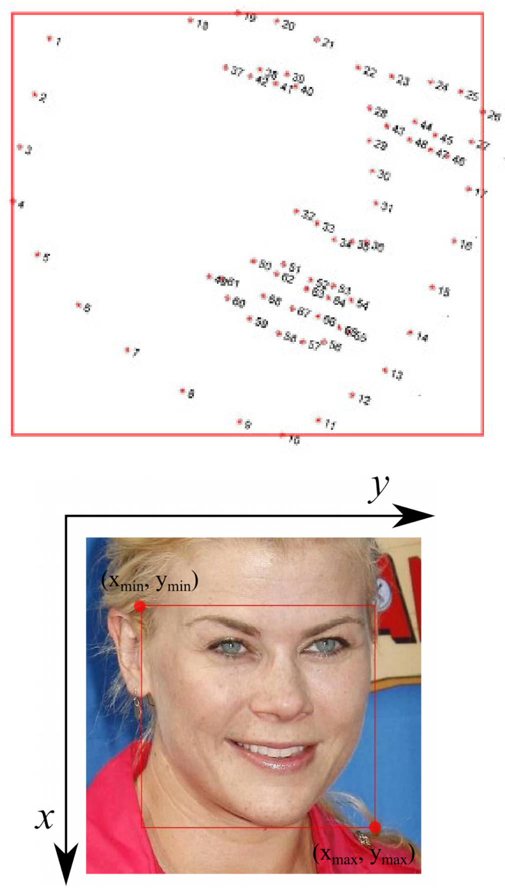
We do these procedures so that we can isolate the area of the face we actually need in order to accelerate program's execution speed. Should be noted that in the Machine's learning phase each

picture in the input data must contain only one face. The reason we use the OpenCV classifiers is that they are faster and more efficient the classifiers of Dlib.
The results of the face isolation process is to cut-off the area of the face we need from the whole picture and then we use the Dlib library to landmark the characteristics of the face being scanned.

4.4) Understanding Dlib's library facial landmark detector

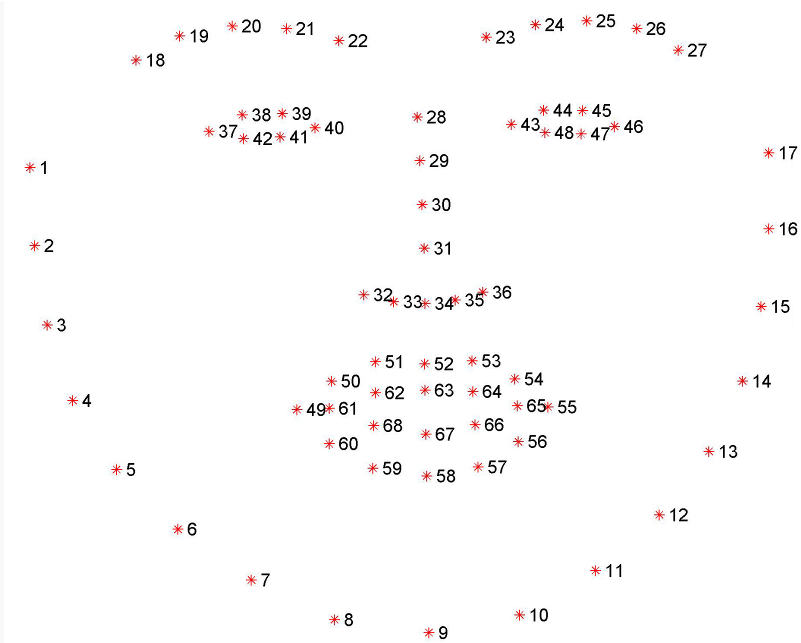


Picture 4.4 – example of landmark placement method.



Picture 4.5 – examples of face detector (with and without landmarks) method.

In any case, as long as the face detector (from the OpenCV Library) is successful, the pre-trained facial landmark detector inside the Dlib library is used to estimate the location of **68 (x, y)-coordinates** that map to facial structures.
The indexes of the 68 coordinates can be visualized on the image below:



Picture 4.5 - visualizing the 68 facial landmark coordinates from the iBUG 300-W dataset.

These annotations are part of the 68 point iBUG 300-W, a dataset of 300 faces provided for research purposes only. Was created for the first Automatic Facial Landmark Detection in-the-Wild Challenge (300-W 2013) to be held in conjunction with International Conference on Computer Vision 2013, Sydney, Australia. The Dlib facial landmark predictor was trained on iBUG 300-W. In the output we get 68 overall flags around the basic and fundamental features of the human face such as eyes, nose, eyebrows, mouth and jawline. These 68 marks, constitute the essential analysis data, which we later use to train the SVM Machine and it’s predictions. That data is then analyzed into all possible couples of unique combinations between them and their distance is being calculated based on the following formula(norm):

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}. \qquad \textbf{\textit{(4.1)}}$$

The value of each calculation, is saved in a vector among the total vectors, from which the SVM Machine’s training is based on.

4.5) Support vector machine (SVM) training

After the completion of the Input Data Analysis, for all the input data there is a training taking place for the SVM. For the Machine, initially a Gaussian Kernel (Rarial Basis Function) is selected due to its performance and non-linear separation of the input data and afterward because we have a multi-class problem, where each class represents a specific emotion, we choose an one-vs-one classifier. Based on that mechanism, the Machine internally with the input of these four classes (n = 4) creates all the possible combinations between them according to the following correlation:

$$\frac{n!}{2(n-2)!}=\frac{n*(n-1)}{2}=6 \tag{4.2}$$

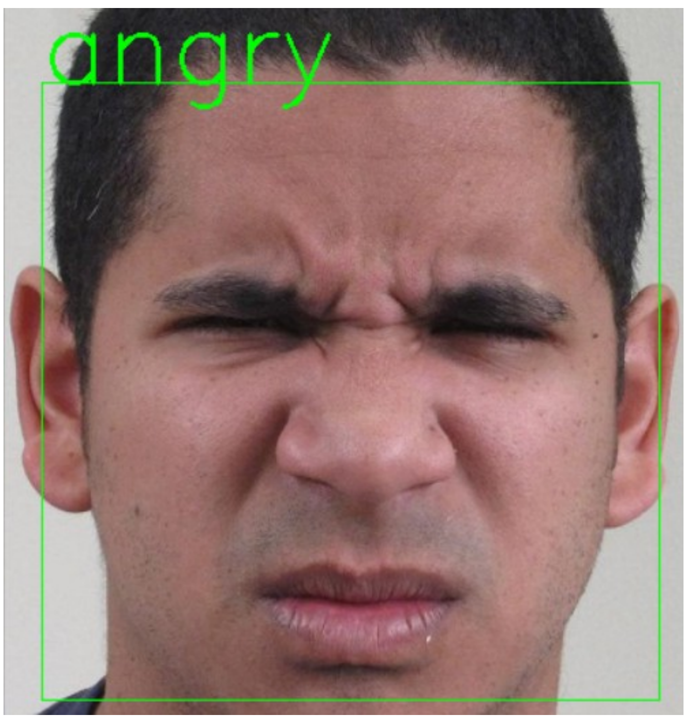
So it can find the dividing hyper-plane of each set(couple) of classes. All the necessary tools for the creation of such a Machine are offered in the Dlib library.

4.6) Support vector machine (SVM) prediction

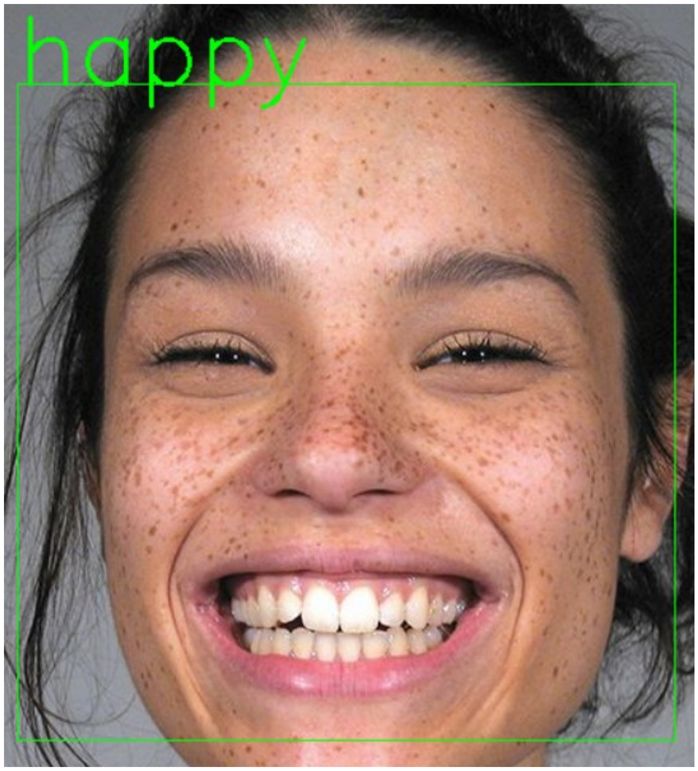
After the SVM Machine's training is finished ,the user can now advance to its evaluation. For the Evaluation of the procedure in the SVM's training we need to apply the grouping and analyzing of the input data from the start, as described before. Each of the new data is given as input to the now trained SVM Machine and its decision function and then we get its prediction as output. The results of the Machine's prediction show up on the user's screen with the indication of the picture and its prediction. For each of the above predictions (one for each picture in the input data) there is also a count of all the successful predictions in order to have a statistical analysis of the overall success of the SVM Machine's ability to identify one of the four emotions we are interested in from the input data.

4.7) Support vector machine (SVM) execution results

Here I'll present some instances of the overall Machine's execution, after the *Training process* during the *Emotion Prediction phase.* We notice that in the pictures where the expression of the face and angle of the photograph's shot is clear and straight the prediction results are pretty good. On the contrary there are enough examples where the extraction-prediction of the emotion being shown is difficult and in some cases beyond the emotional classes that we are interested in this Thesis and additionally the training's process four classes.



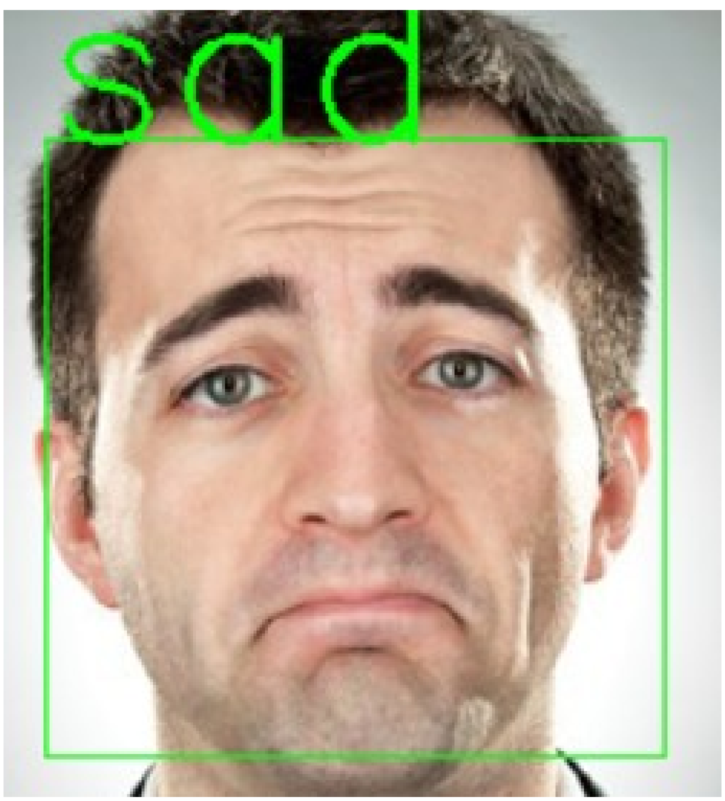
Picture 4.6 – example of successful prediction.



Picture 4.7 – example of successful prediction.

Picture 4.8 – example of failed prediction.

Picture 4.9 – example of failed prediction.



Picture 4.10 – *e*xample of successful prediction.

CHAPTER 5

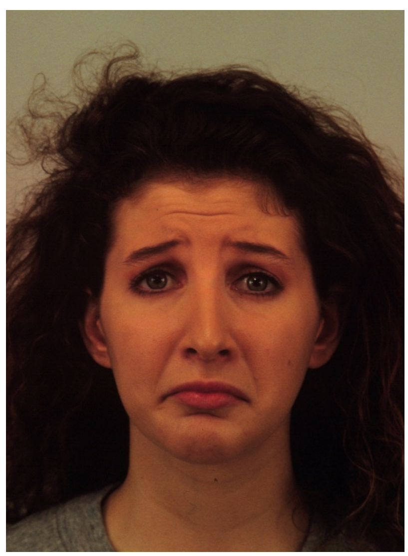
Datasets and results analysis

5.1) Data sets

Two data sets are being used for the training and validation of the SVM.

I) Patras A.I.nD.M. data set of 84 directed facial poses (portrait angle)

This dataset is from the A.I and Data Mining Lab's data base of The University of Patras. A collection of 85 photos 562 x 762 pixels resolution taken in medium ligthing of a single person each in close-up frontal view pre-cropped to only show the face example:



Picture 5.1 – picture from data set Patras.

From these 84 pictures (21 belong to each of the four emotional states we are looking for (Anger, Joy, Tranquility and Sadness). Because of its limited material for both training and verification we made two tests, one with 10% of the dataset's pictures as verification and 90% for training, and the second one 20% of it's pictures for verification and 80% for training.

Patras (A) The first *Test* consists of 19 pictures of each emotion for *Training* the SVM and 2 (10%) pictures of each emotion for *Verification.*

Patras (B) The second *Test* consists of 17 pictures of each emotion for *Training* the SVM and 4 (20%) pictures of each emotion for *Verification.*

II) Fer2013 data set of 960 undirected facial poses (random angle)

The second data set was prepared by Pierre-Luc Carrier and Aaron Courville, as part of an ongoing research project and was given freely for a Robotic Vision Contest for Kaggle.com and it was free for downloading. The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. Both data sets are accompanied by a files.csv that contains two columns, "PicName" and "Emotion Code". The "PicName" column contains the name of the picture (JPEG file) that the "Emotion Code" column on it's right corresponds to. The "Emotion Code" column contains a numeric code ranging from 0 to 3, inclusive, for the emotion that is present in the image.

The *Training Set* consists of 960 examples (240 pictures for each emotion) and the *Verification Set of 60* examples (20 %) of the dataset.

Example 1 (Actual size):

Picture 5.2 – pics from data set Fer2013.

5.2) Testing and use of data sets

The Usual practice in using a data set for training and Validation is by using approximately the 90% of a data set as training material and the rest 10% for Validation and that's what we are also going to abide by.

So for the 1st one (the smaller one) from Patras University I used 17 pictures of each emotion for Training the SVM Machine and 4 (20%) pictures of each emotion for Verification.

And for the 2nd one (the big one) called Fer2013 from Kaggle I used 240 of each emotion for Training the SVM and 60 for Verification.

Now we will analyze them based on the **3.4) Sensitivity, Specificity and Accuracy (evaluation stats)** for each different emotion (clasifier) so since we have 4 different emotions we will do it 4 times

from Patras University Data sets

Patras (A)

1st emotion : **Anger (code "0")**

True positive (TP) = 2 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 1 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 5 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 0 (the number of cases *incorrectly* identified as ***Missed)***

$Accuracy = (TP + TN) / total \quad = (2 + 5) / 8 = 0.875$

Sensitivity = TP / actual yes = 2 / 2 = 1

Specificity = TN / actual no = *5 / 6* = 0.83

2nd emotion : <u>Joy (code "1")</u>

True positive (TP) = 2 (the number of cases <u>*correctly*</u> identified as ***Match)***

False positive (FP) = 0 (the number of cases <u>*incorrectly*</u> identified as ***Match)***

True negative (TN) = 6 (the number of cases <u>*correctly*</u> identified as ***Missed)***

False negative (FN) = 0 (the number of cases <u>*incorrectly*</u> identified as ***Missed)***

Accuracy = (TP + TN) / total = (2 + 6) / 8 = 1

Sensitivity = TP / actual yes = 2 / 2 = 1

Specificity = TN / actual no = 6 / 6 = 1

3rd emotion : <u>Tranquility (code "2")</u>

True positive (TP) = 1 (the number of cases <u>*correctly*</u> identified as ***Match)***

False positive (FP) = 0 (the number of cases <u>*incorrectly*</u> identified as ***Match)***

True negative (TN) = 6 (the number of cases <u>*correctly*</u> identified as ***Missed)***

False negative (FN) = 1 (the number of cases <u>*incorrectly*</u> identified as ***Missed)***

Accuracy = (TP + TN) / total = (1 + 6) / 8 = 0.875

Sensitivity = TP / actual yes = 1 / 2 = 0.5

Specificity = TN / actual no = 6 / 6 = 1

4th emotion : Sadness (code "3")

True positive (TP) = 1 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 1 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 5 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 1 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (1 + 5) / 8 = 0.75

Sensitivity = TP / actual yes = 1 / 2 = 0.5

Specificity = TN / actual no = 5 / 6 = 0.83

Patras (B)

1st emotion : Anger (code "0")

True positive (TP) = 2 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 1 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 11 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 2 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (2 + 11) / 16 = 0.81

Sensitivity = TP / actual yes = 2 / 4 = 0.5

Specificity = TN / actual no = *11 / 12* = 0.92

2nd emotion : *Joy (code "1")*

True positive (TP) = 4 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 0 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 12 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 0 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (4 + 12) / 16 = 1

Sensitivity = TP / actual yes = 4 / 4 = 1

Specificity = TN / actual no = 12 / 12 = 1

3rd emotion : Tranquility (code "2")

True positive (TP) = 2 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 1 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 11 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 2 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (2 + 11) / 16 = 0.81

Sensitivity = TP / actual yes = 2 / 4 = 0.5

Specificity = TN / actual no = 11 / 12 = 0.92

4th emotion : Sadness (code "3")

True positive (TP) = 3 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 3 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 9 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 1 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (3 + 9) / 16 = 0.75

Sensitivity = TP / actual yes = 3 / 4 = 0.75

Specificity = *TN / actual no = 9 / 12 = 0.75*

from Fer2013 data set

1st emotion : **Anger (code "0")**

True positive (TP) = 20 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 21 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 91 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 16 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (20 + 91) / 142 = 0.78

Sensitivity = TP / actual yes = 20 / 36 = 0.56

Specificity = TN / actual no = *91 / 110* = 0.83

2nd emotion : <u>Joy (code "1")</u>

True positive (TP) = 28 (the number of cases <u>*correctly*</u> identified as ***Match)***

False positive (FP) = 12 (the number of cases <u>*incorrectly*</u> identified as ***Match)***

True negative (TN) = 93 (the number of cases <u>*correctly*</u> identified as ***Missed)***

False negative (FN) = 11 (the number of cases <u>*incorrectly*</u> identified as ***Missed)***

Accuracy = (TP + TN) / total = (28 + 93) / 142 = 0.85

Sensitivity = TP / actual yes = 28 / 39 = 0.72

Specificity = TN / actual no = 93 / 124 = 0.75

3rd emotion : <u>Tranquility (code "2")</u>

True positive (TP) = 18 (the number of cases <u>*correctly*</u> identified as ***Match)***

False positive (FP) = 22 (the number of cases <u>*incorrectly*</u> identified as ***Match)***

True negative (TN) = 70 (the number of cases <u>*correctly*</u> identified as ***Missed)***

False negative (FN) = 23 (the number of cases <u>*incorrectly*</u> identified as ***Missed)***

Accuracy = (TP + TN) / total = (18 + 70) / 142 = 0.62

Sensitivity = TP / actual yes = 18 / 40 = 0.45

Specificity = TN / actual no = 70 / 104 = 0.67

4th emotion : Sadness (code "3")

True positive (TP) = 6 (the number of cases *correctly* identified as ***Match)***

False positive (FP) = 17 (the number of cases *incorrectly* identified as ***Match)***

True negative (TN) = 83 (the number of cases *correctly* identified as ***Missed)***

False negative (FN) = 22 (the number of cases *incorrectly* identified as ***Missed)***

Accuracy = (TP + TN) / total = (6 + 83) / 142 = 0.62

Sensitivity = TP / actual yes = 6 / 29 = 0.21

Specificity = TN / actual no = 83 / 104 = 0.8

5.3) Analyzing results

Patras (A)	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Anger	2	5	1	0	1	0.83	0.875
Joy	2	6	0	0	1	1	1
Tranquility	1	6	0	1	0.5	1	0.875
Sadness	1	5	1	1	0.875	0.5	0.83

Patras (B)	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Anger	2	11	1	2	0.5	0.92	0.81
Joy	4	12	0	0	1	1	1
Tranquility	2	11	1	2	0.5	0.92	0.81
Sadness	3	9	3	1	0.75	0.75	0.75

Fer2013	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Anger	20	91	21	16	0.56	0.85	0.78
Joy	28	93	12	11	0.72	0.75	0.85
Tranquility	18	70	22	23	0.45	0.67	0.62
Sadness	6	83	17	22	0.21	0.8	0.62

Table 5.1 – statistic results of the datasets' validation tests.

At first glance we see that the two tests from Patras data set did overall better than the Fer2013 data set and there are two reasons behind it. The first reason, the Patras data set is very small only 84 pictures (22 for each emotion) detailed clean and focused with good resolution 562 x 762 and nearly each person presented in its pictures has a pose for each emotion so its very specific and very narrow it was expected to do well.
The second reason to add to the first one, the Fer2013 data set is a lot bigger 1200 examples (300 for each emotion) and because it was created for a Facial Recognition competition it is made to test the recognition algorithms to their limits with very small resolution 48x48 pixels and full with out of focus, obscured (either by hair, hands, sunglasses, hats, bad angle etc), some of them are not even from real pictures rather portraits or drawings of faces, with wide variety of different people from around the world.

Of all the four emotions *Joy* was the overall *best* distinguished with 100% in both Patras data set tests and 72%, 75% and 85% in Sensitivity, Specificity and Accuracy in the Fer2013 data set accordingly. The reason being that the happy facial expression has a very distinguishable mouth formation compared to the other three (it seems mouth formation plays a more heavy role than eye area formation in decision making).

Anger comes second with the highest Sensitivity and Specificity compared to Tranquility and *Sadness* emotions in all three tests with 100%, 83%, 87.5% in Sensitivity, Specificity and Accuracy in Patras(A) data set, 50%, 92% and 81% (Sen, Spe, Acc) in Patras(B) dataset and 56%, 85%, 78% in Fer2013 data set. That’s mainly to the distinctive eyes and eyebrow area formation and its even more distinguishable when accompanied with an open mouth (when swearing for example).

Tranquility is third with 50%, 100%, 87.5% (Sen, Spe, Acc) in Patras(A) data set, 50%, 92%, 81% in Patras(B) data set and 45%, 67% and 62% (Sen, Spe, Acc) in Fer2013 data set. Quite close with *Sadness* with 87.5%, 50% and 83% in Patras(A) data set, 75% in all three in Patras(B) data set and 21%, 80% and 62% in Fer2013 data set because the tranquility and sadness facial expression don’t always have extreme facial formation usually more subtle especially when not crying so they are not easy to distinguish one another.

CHAPTER 6

Future work and proposed method improvements

Recognizing emotions from facial expressions is challenging sometimes even for humans even from personal experience and you must have a good clear look on somebody's face (no hands, hair, hats angle, lighting etc obscuring their face) and we also analyze the body language of that person in order to be more accurate on our reading. Another thing is that humans are complex beings with complex and combined emotions a lot of times so one of the ways to improve the proposed method is by checking percentage of each emotions presented in each picture or analyzed face in a frame so that its more accurate and detailed about its readings. Voice analysis and tone can also be add a lot of times we can understand someones emotional state just by the tone of his voice without the need of understanding what he says. The eyes pupil’s can also tell a lot, if the method is enhanced with pupil analysis and movement could benefit in recognizing Anger for example where the pupils widen. Another improvements is to add more emotion classifiers so that it can have a more complete gamma of emotional states for a better analysis. Thermal Vision or techniques that can detect the psychosomatic changes in the body since most of the emotions have specific psychosomatic traits like blood pressure rising, heart beat rising, tears, faster breathing or the lack there of.

ACRONYMS

Facial Expression Recognition	FER
Fusiform Face Area	FFA
Facial Action Coding System	FACS
Action Units	Aus
Obsessive–Compulsive Disorder	OCD
Support Vector Machines	SVM
Graphical User Interfaces	GUI
Radial Basis Function (kernel)	rbf

LINKS – REFERENCES

[1] Muscles of facial expression. https://www.kenhub.com/en/library?sequence=facial-muscles. Accessed: 2016

[2] David Matsumoto and Hyi Sung Hwang, *Reading facial expressions of emotion* ,a Science Brief from the site of American Psychological Association http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx May 2011

[3] Radom´ır Cˇ IHA´ K. Anatomie 2. 2. vyd. Praha. Grada Publishing, 2002

[4] Victoria Contreras. Artnatomya. http://www.artnatomia.net/, 2005

[5] Charles Darwin, Paul Ekman, and Phillip Prodger. The expression of the emotions in man and animals. Oxford University Press, USA, 1998

[6] Paul Ekman. Darwin, deception, and facial expression. Annals of the New York Academy of Sciences, 1000(1):205–221, 2003

[7] Paul Ekman. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. Macmillan, 2007

[8] Paul Ekman and Wallace V Friesen. Facial action coding system. 1977

[9] Paul Ekman andWallace V Friesen. Unmasking the face: A guide to recognizing emotions from facial clues. Ishk, 2003

[10] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. Emotion in the human face: Guidelines for research and an integration of findings. Elsevier, 2013

[11] Paul Ekman and Dacher Keltner. Universal facial expressions of emotion. California Mental Health Research Digest, 8(4):151–158, 1970

[12] Margaret J Fehrenbach and Susan W Herring. Illustrated anatomy of the head and neck. Elsevier Health Sciences, 2015

[13] Wallace V Friesen and Paul Ekman. Emfacs-7: Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco, 2(36):1, 1983

[14] Carl-Herman Hjortsj¨o. Man's face and mimic language. Studen litteratur, 1969

[15] David Matsumoto, Seung Hee Yoo, and Sanae Nakagawa. Culture, emotion regulation, and adjustment. Journal of personality and social psychology, 94(6):925, 2008

[16] Jason Brownlee in Machine Learning Algorithms (Support Vector Machines for Machine Learning) 2016

[17] Theodoridis, Sergios; and Koutroumbas, Konstantinos; Pattern Recognition, 4th Edition, Academic Press, , 2009

[18] Kecman, Vojislav; Learning and Soft Computing — Support Vector Machines, Neural Networks, Fuzzy Logic Systems, The MIT Press, Cambridge, MA, 2001

[19] Schölkopf, Bernhard; Burges, Christopher J. C.; and Smola, Alexander J. (editors); Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 1999.

[20] Shawe-Taylor, John; and Cristianini, Nello; Kernel Methods for Pattern Analysis, Cambridge University Press, 2004

[21] Steinwart, Ingo; and Christmann, Andreas; Support Vector Machines, Springer-Verlag, New York, 2008

[22] Campbell, Colin; and Ying, Yiming; Learning with Support Vector Machines, Morgan and Claypool, 2011

[23] Fradkin, Dmitriy; and Muchnik, Ilya; "Support Vector Machines for Classification" in Abello, J.; and Carmode, G. (Eds); Discrete Methods in Epidemiology, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, volume 70, pp. 13–20, 2006

[24] Jean-Philippe Vert, Koji Tsuda, and Bernhard Schölkopf. "A primer on kernel methods". Kernel Methods in Computational Biology 2004

[25] Shashua, Amnon. "Introduction to Machine Learning: Class Notes 67577" 2009

[26] Andreas Müller. Kernel Approximations for Efficient SVMs (and other feature extraction methods) 2012

[27] Ali Rahimi and Benjamin Recht. "Random features for large-scale kernel machines" 2007

[28] Alireza Baratloo, Mostafa Hosseini, Ahmed Negida and Gehad El Ashal Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. Emerg (Tehran) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614595/ 2015

[29] Simple guide to confusion matrix terminology by Kevin Markham Machine Learning 2014

[30] Bourbaki, Nicolas. "Chapters 1–5". Topological vector spaces. Springer 1987

[30] Trèves, François. Topological Vector Spaces, Distributions and Kernels. Academic Press, Inc. pp. 136–149, 195–201, 240–252, 335–390, 420–433. 1995

[31] OpenCV Tutorials https://docs.opencv.org/3.4/d9/df8/tutorial_root.html 2018

[32] OpenCV Guide https://docs.opencv.org/2.4/doc/user_guide/user_guide.html 2018

[33] Dlib Tutorial http://dlib.net/intro.html 2018

[34] Ms. Aswathy.R, Facial Expression Recognition using SVM classification in Perceptual Color Space *Department of Computer Science & Engineering, Nehru College of Engineering & Research Center, Thrissur, Kerala, India* **https://ijcsmc.com/docs/papers/June2013/V2I6201393.pdf** 2013

[35] P. Visutsak, Emotion classification using Adaptive SVMs, *International Journal of Computer and Communication Engineering, Vol. 1, No. 3* **http://www.ijcce.org/papers/72-P20002.pdf** 2012

[36] *Kotsia and I. Pitas*, Real-time Facial Expression Recognition from image sequences using Support Vector Machines *I., Aristotle University of Thessaloniki, Department of Informatics* **http://poseidon.csd.auth.gr/papers/PUBLISHED/CONFERENCE/pdf/Kotsia05a.pdf**

[35] Hend Ab. ELLaban, A. A. Ewees, Elsaeed E. AbdElrazek, a real-time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor Classifier by *Computer Inst. Prep. Dep. Damietta University Egypt* **https://pdfs.semanticscholar.org/e9c7/2b160e697e635e7a16c25857a229b4ed5725.pdf** 2017

[34] Nazia Perveen, Nazir Ahmad, M. Abdul Qadoos Bilal Khan, Rizwan Khalid, Salman Qadri, Facial Expression Recognition through machine learning (*International Journal of Scientific & Technology Research Volume 5, Issue 03)* **http://www.ijstr.org/final-print/mar2016/Facial-Expression-Recognition-Through-Machine-Learning.pdf** *2016*

[34] Spiros V. Ioannou, Amaryllis T. Raouzaiou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, Stefanos D. Kollias, Article:Literature review in Neural Networks

https://www.researchgate.net/publication/7779057_Emotion_recognition_through_facial_expression_analysis_based_on_a_neurofuzzy_network 2005

[34] Shruti Bansal, Pravin Nagar Emotion recognition from facial expression based on Benzier curve International Journal of Advanced Information Technology (IJAIT) Vol. 5, No. 3/4/5/6 **http://aircconline.com/ijait/V5N6/5615ijait01.pdf** 2015

[35] Fatima Zahra Salmam, Abdellah Madani, Mohamed Kissi, Emotion recognition from facial expression based on fiducial points detection and using Neural Network Article in International Journal of Electrical and Computer Engineering. Available from: **https://www.researchgate.net/publication/324790426_Emotion_recognition_from_facial_expression_based_on_fiducial_points_detection_and_using_Neural_Network** 2018

[36] Robert Fischer, Automatic Facial Expression Analysis and Emotional Classification Submitted to the Department of Math and Natural Sciences MIT – Thesis Available from: **http://cbcl.mit.edu/publications/theses/thesis-masters-fischer-robert.pdf** 2004

[37] Galateia Iatraki, Emotional Facial Expression Recognition & Classification by Master of Science in Media and Knowledge Engineering TUDelft University – Thesis from: **https://pdfs.semanticscholar.org/e70c/2ab8d046588dd551872d009f6b4aee78f59c.pdf** 2009

[38] Santosh Kumar, Shubam Jaiswal, Rahul Kumar, Sanjay Kumar Singh, Emotion Recognition Using Facial Expression by Innovative Research in Attention Modeling and Computer Vision (book), *(PDF) Emotion Recognition Using Facial Expression*. Available from: **https://www.researchgate.net/publication/301770606_Emotion_Recognition_Using_Facial_Expression** 2018

[39] C. Sagonas, E. Antonakos, G, Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 faces In-the-wild challenge: Database and results. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild". 2016.

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

[41] Belhumeur, P., Jacobs, D., Kriegman, D., Kumar, N.. ‘Localizing parts of faces using a consensus of exemplars’. In Computer Vision and Pattern Recognition, CVPR. (2011).

[42] X. Zhu, D. Ramanan. ‘Face detection, pose estimation and landmark localization in the wild’, Computer Vision and Pattern Recognition (CVPR) Providence, Rhode Island, June 2012.

[43] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Boudev, Thomas S. Huang. ‘Interactive Facial Feature Localization’, ECCV2012.

[44] Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G. ‘Xm2vtsdb: The ex- tended m2vts database’. In: 2nd international conference on audio and video-based biometric person authentication. Volume 964. (1999).

[45] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge. Proceedings of IEEE Int’l Conf. on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W). Sydney, Australia, December 2013.

[46] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. A semi-automatic methodology for facial landmark annotation. Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013). Oregon, USA, June 2013.