PROJECT REPORT

ON

"Human Face Emotion Recognition using Machine Learning"

Submitted to

SANT GADGE BABA AMRAVATI UNIVERSITY, AMRAVATI

For

Partial Fulfilment of the degree in Bachelor of Engineering

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

Mr. Prajwal Wankhede

Miss. Ankita Patil

Mr. Bhushan Kharat

Mr. Kunal Jadhav

Under Guidance Of

Dr. A. S. Kapse

Department of Computer Science and Engineering,

ANURADHA ENGINEERING COLLEGE, CHIKHLI

DIST. BULDHANA (M.S.) INDIA

2022-2023

ANURADHA ENGINEERING COLLEGE, CHIKHLI

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the project entitled

"Human Face Emotion Recognition using Machine Learning"

*which is being submitted here with for the award of the **'Bachelor of Engineering'** in **'Computer Science and Engineering'** of Sant Gadge Baba Amravati University, Amravati. This is the result of the Project work and contribution by 'Mr. Prajwal Wankhede, Ms. Ankita Patil, MBhushan Kharat, Mr. Kunal Jadhav' under my supervision and guidance.*

Place:

Date:

Dr. A. S. Kapse
Guide

Dr. A. S. Kapse
Head of Department

Dr. A. N. Nanhai
Principal

DECLARATION

Mr. Prajwal Wankhede, Miss. Ankita Patil, Mr. Bhushan Kharat, Mr. Kunal Jadhav, *hereby declare that we have completed the project towards the awards of Bachelor of Engineering Degree of Sant Gadge Baba Amravati University, Amravati in Computer Science and Engineering discipline on the topic entitled* ***"Human Face Emotion Recognition using Machine Learning"*** *under the supervision of* ***Dr. A.S. Kapse***

This project report is outcome of our own research work. This project embodies the original research work in partial fulfilment of the requirements of the ***Bachelor of Engineering Degree*** *of Sant* ***Gadge Baba Amravati University, Amravati in computer Science and Engineering.***

***Place*:**

Mr. Prajwal Wankhede
Miss. Ankita Patil
Mr. Bhushan Kharat
Mr. Kunal Jadhav

BE Final Year
Computer Science and Engineering

PROJECT REPORT APPROVAL SHEET

Project report Entitled

"Human Face Emotion Recognition using Machine Learning"

By

Mr. Prajwal Wankhede

Miss. Ankita Patil

Mr. Bhushan Kharat

Mr. Kunal Jadhav

Is approved for the degree of **Bachelor of Engineering**

In

Computer Science and Engineering

Of

Sant Gadge Baba Amravati University, Amravati

Internal Examiner:	External Examiner
Name:	Name:
Designation:	Designation:
Sign:	Sign:
Date:	Date:

ACKNOWLEDGEMENT

The making of the project needed co-operation and guidance of a number of people. We therefore consider it our prime duty to thank all those who had helped us through their venture. It is our immense pleasure to express our gratitude to **Dr. A. S. Kapse** as guide who provided us constructive and positive feedback during the preparation of this project. We express our sincere thanks to the head of department **Dr. A. S. Kapse** and all other staff members of CSE department for their kind co-operation.

We would like to thanks ***Dr. A. N. Nanhai***, Principal of our institution for providing necessary facility during the period of working on this report. We are thankful to our friends and library staff members whose encouragement and suggestion helped us to complete our Work. We are also thankful to our parents whose best wishes are always with us.

Mr. Prajwal Wankhede

Miss. Ankita Patil

Mr. Bhushan Kharat

Mr. Kunal Jadhav

BE Final Year

Computer Science and Engineering

LIST OF FIGURES

LIST OF TABLES

Abstract

Presently large amount of data is available on social networking sites, product review sites, blogs, forums etc. This data holds expressed opinions and sentiments. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyze it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis and opinion mining. Expression of any sentiment is a mixture of text, prosody, facial expression, body posture etc. Thus only text input cannot fully represent a sentiment. A multimodal system uses a combination of input modes e.g. video. This system analyses the techniques used for multimodal sentiment real time data and also demonstrate that how to predict the emotion. Extracting the sentiments from different input modes is achieved by different classifying techniques. In this system, firstly we will discuss the different input modes. In the end we analyse challenges of our proposed system. Also we have discussion on integration of different modes and its effect on emotion reorganization system

Keywords- Opinion Mining, Sentiment Analysis, Multimodal Sentiment Analysis, Big Data, Data Mining.

INDEX

CHAPTER 1

INTRODUCTION

1.1 Introduction

In recent years, with the popularity of social media, users are increasingly keen to express their feelings and opinions in the form of pictures and text, which makes multimodal data with text and pictures the con tent type with the most growth. Most of the information posted by users on social media has obvious sentimental aspects, and multimodal sentiment analysis has become an important research field. Previous studies on multimodal sentiment analysis have primarily focused on extracting text and image features separately and then combining them for sentiment classification. These studies often ignore the interaction between text and images. Therefore, this project proposes a new multimodal sentiment analysis model. The model first eliminates noise interference in textual data and extracts more important image features. Then, in the feature-fusion part based on the attention mechanism, the text and images learn the internal features from each other through symmetry. Then the fusion features are applied to sentiment classification tasks. The experimental results on two common multimodal sentiment datasets demonstrate the effectiveness of the proposed model.

The goal of image classification is to decide whether an image belongs to a certain category or not. Different types of categories have been considered in the literature, e.g. defined by presence of certain objects, such as cars or bicycles, or defined in terms of scene types, such as city, coast, mountain, etc. To solve this problem, a binary classifier can be learned from a collection of images manually labeled to belong to the category or not. Increasing the quantity and diversity of hand-labeled images improves Tags: desert,nature,landscape,sky Tags: rose, pink Labels: clouds, plant life, sky, tree Labels: flower, plant life Tags: india Tags: aviation, airplane, airport Labels: cow Labels: aeroplane. Example images from MIR Flickr (top row) and VOC'07 (bottom row) data sets with their associated tags and class labels. The performance of the learned classifier, however, labeling images is a time consuming task. Although it is possible to label large amounts of images for many categories for research purposes, this is often unrealistic, e.g. in personal photo organizing applications. This motivates our interest in using other sources of information that can aid the learning process using a limited amount of labeled images. With the increasing popularity of social media, people are increasingly keen to express their views or opinions on social media platforms. In social media, hundreds of millions of data records are generated every day. A large volume of data is in the form of text and image combinations, which constitute a huge volume of multimodal data. Rich sentimental information exists in the multimodal data. The data are helpful to understand people's attitudes and views on some events by analyzing the sentiment of multimodal data. Multimodal sentiment analysis has great application value in box office predictions, political elections, stock market predictions, and so on. In the graphic data of social media, text and images contain sentimental information, which is different and complementary to each other. Compared with the single-mode data of text or an image, multimodal data contains more information and can better reveal the real feelings of users. However, multimodal sentiment analysis remains a particularly challenging task. First, sentimental information in

different modal data is different, so the sentimental feature representation of the modal data must be obtained effectively for sentiment analysis. From the human viewpoint, not all image areas in a picture are related to sentimental expression, and not all words in text data are related to sentiment. Therefore, in feature extraction, we should highlight the sentimental key areas of the image and eliminate the noise interference in the textual data. Second, different modal data must express the underlying features of different dimensions and attributes. In previous studies, the sentiment analysis of single-mode texts primarily used traditional statistical methods, which are highly dependent on the quality of the extracted features. For example, some used data envelopment analysis (DEA) as a Word Sense Disambiguation (WSD) tool to automatically identify the appropriate meaning of words in sentences, so as to determine the inherent meaning of voter intentions regarding possible political candidates. The sentiment analysis of single-mode images also depends on the quality of features selected by hand-designed extraction rules. These images often contain redundant sentimental information. With the continuous development of machine learning and deep learning, researchers have developed many new methods for sentiment analysis of multimodal data and have achieved good results. Multimodal sentiment analysis is an increasingly important research field, and many researchers have conducted studies in this field. For multimodal sentiment analysis tasks, the core challenge is how to make better use of modal internal information and modal interaction information. The internal information of the mode is information that can be mined and used by a single independent mode, and the interaction between modes provides useful information. The interaction between modes can be determined by the correlation and connection between different modes. Determing how to use the interactive information between different modes is also the biggest difference between multimodal tasks and single-modal tasks. Therefore, multimodal affective analysis must determine effective ways to combine the information of each mode. By capturing the similar structure between different modal data, effective information missing in traditional tasks is captured to maximize the fusion of modal information and the interaction between the modes. Presently large amount of data is available on social networking sites, product review sites, blogs, forums etc. This data holds expressed opinions and sentiments. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyze it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis and opinion mining. Expression of any sentiment is a mixture of text, prosody, facial expression, body posture etc. Thus only text input cannot fully represent a sentiment.

A multimodal system uses a combination of input modes e.g. text and audio or text and video or all of these three. This paper analyses the techniques used for multimodal sentiment data and also demonstrate that how individual model works. Extracting the sentiments from different input modes is achieved by different classifying techniques. In this paper, firstly we will discuss the different input modes. In the end

we analyse challenges of our proposed system. Also we have discussion on integration of different modes and its affect on emotion reorganization system

1.2 Problem Statement

In order to identify potential risks, it is important for companies to collect and analyze information about their competitors' products and plans. Sentiment analysis find a major role in competitive intelligence to extract and visualize comparative relations between products from customer reviews, this information can be used to improve product, marketing strategy and potential risk can be identified in early days. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyses it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form much of the current research is focusing on the area of sentiment analysis and opinion mining The Opinion mining provides functions over to data, no matter how great its volume, velocity or variability or where it lives. That data is often out-of-date, missing or stored in disparate systems, so analyzer relies on their talent alone to make strategic decisions.

1.3 Multi-Modal Modeling

Multi-Modal Modeling of images and text combines semantic knowledge extracted from text with knowledge of spatial structures extracted from images. Models of this type learn joint representations of images and text. These joint representations have been used to relate images and text to improve search-and-retrieval, classification, and self-supervised learning. Additionally, training data from the web has been shown to yield more generalizable models. This study was focused on using multi-modal data to augment image classification tasks. Facial expression, Body movement, posture, gestures etc. are analysed to extract the sentiment features. Facial expressions are among the most universal forms of body language which are studied in emotion analysis

Table 1.1 Facial Cues and Emotion

Emotion	**Observed Facial Cues**
Surprise	Brows raised (curved and high) Skin below brow stretched Horizontal wrinkles across forehead Eyelids opened and more of the white of the eye is visible Jaw drops open without tension or stretching of the mouth
Fear	Brows raised and drawn together

	Forehead wrinkles drawn to the center Upper eyelid is raised and lower eyelid is drawn up Mouth is open Lips are slightly tense or stretched and drawn back
Disgust	Upper lip is raised Lower lip is raised and pushed up to upper lip or it is lowered Nose is wrinkled Cheeks are raised Lines below the lower lid, lid is pushed up but not tense Brows are lowered
Anger	Brows lowered and drawn together Vertical lines appear between brows Lower lid is tensed and may or may not be raised Upper lid is tense and may or may not be lowered due to brows' action Eyes have a hard stare and may have a bulging appearance Nostrils may be dilated (could occur in sadness too) unambiguous only if registered in all three facial areas
Happiness	Corners of lips are drawn back and up Mouth may or may not be parted with teeth exposed or not A wrinkle runs down from the nose to the outer edge beyond lip corners Cheeks are raised Lower eyelid shows wrinkles below it, and may be raised but not tense Crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness	Inner corners of eyebrows are drawn up Skin below the eyebrow is triangulated, with inner corner up Upper lid inner corner is raised Corners of the lips are drawn or lip is trembling

1.4 Self-Supervised Learning

As an alternative to fully human-supervised algorithms, recently, there has recently been a growing interest in self-supervised or naturally-supervised. These approaches make use of non-visual signals, intrinsically correlated to images, as a form of supervision for visual feature learning. The prevalence of websites with images and loosely-related human annotations provide a natural opportunity for self-supervised learning. This differs from previous image-text embedding methods in that the goal is to learn generic and discriminative features in a self-supervised fashion without making use of any annotated dataset.

1.5 Generalizability of Learning's From the Web.

Research has lately focused on joint image and text embeddings. Merging different kinds of data has motivated the possibilities of learning together from different kinds of data, which put more focus on the field of study where both general and applied research has been done. A Deep Visual-Semantic Embedding Model proposes a pipeline that, instead of learning to predict ImageNet classes, learns to infer the Word2Vec representations of their labels. By exploiting distributional semantics of a text corpus of every word associated with an image provides inferences of previously unseen concepts in the training set. Semantically relevant predictions make this model valuable even when it makes errors. These errors are generalized to a class outside the labeled training set.

1.6 Motivation

Presently large amount of data is available on social networking sites, product review sites, blogs, forums etc. This data holds expressed opinions and sentiments. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyze it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis and opinion mining. Expression of any sentiment is a mixture of text, prosody, facial expression, body posture etc. Thus only text input cannot fully represent a sentiment. In image categorization the goal is to decide if an image belongs to a certain category or not. A binary classifier can be learned from manually labeled images; while using more labeled examples improves performance, obtaining the image labels is a time consuming process.

1.7 Objectives

A proposed multimodal system uses a combination of input modes e.g. text and audio or text and video or all of these three.

- To analysed the techniques used for multimodal sentiment data and also demonstrate that how individual model works. Extracting the sentiments from different input modes is achieved by different classifying techniques.
- To find out different input modes to generate the model for analysis.
- To analysed challenges of our proposed system. Also we have discussion on integration of different modes and its effect on emotion reorganization system

Sentiment analysis is a method, which classifies the given data due to having positive or negative opinion and being subjective or objective in general. In this case, the concept of emotions and opinions is comes in focus, the opinionated data is deeply analyzed to determine the strength of opinions, which is closely related

to the intensity of emotions such as happy, fear, sad, anger, surprise etc. It is claimed that people's sentiments can be identified by examining their language expressions, and can be classified, according to the level of their strength. It is mostly used in advertisement placement, product benchmarking and market intelligence, and detection of company reputation or brand popularity and identification of fake or misinforming comments

- **Importance to implement the system**

  Multimodal sentiment analysis is computational study of mood, sentiments, views, affective state etc. from the text and audio, video data. Furthermore to make multimodal effective and realistic other parameters can be considered with respect to human entity such as age, gender.

- **Scope of the system**

  Multi-modal Sentiment Analysis problem is a machine learning problem that has been a research interest for recent years. Though lot of work is done till date on sentiment analysis, there are many difficulties to sentiment analyser since Cultural influence, linguistic variation and differing contexts make it highly difficult to derive sentiment. The reason behind this is unstructured nature of natural language. The main challenging aspects exist in use of other modes; dealing with Multi-Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features.

CHAPTER 2

LITERATURE REVIEW

1. Literature Survey

Alejandra Sarahi Sanchez-Moreno et al state that facial recognition is fundamental for a wide variety of security systems operating in real-time applications. Recently, several deep neural networks algorithms have been developed to achieve state-of-the-art performance on this task. The present work was conceived due to the need for an efficient and low-cost processing system, so a real-time facial recognition system was proposed using a combination of deep learning algorithms like FaceNet and some traditional classifiers like SVM, KNN, and RF using moderate hardware to operate in an unconstrained environment. Generally, a facial recognition system involves two main tasks: face detection and recognition.

Proposed System:

It uses the YOLO-Face method for the face detection task which is a high-speed real-time detector based on YOLOv3, while, for the recognition stage, a combination of FaceNet with a supervised learning algorithm, such as the support vector machine (SVM), is proposed for classification. Extensive experiments on unconstrained datasets demonstrate that YOLO-Face provides better performance when the face under an analysis presents partial occlusion and pose variations; besides that, it can detect small faces. The face detector was able to achieve an accuracy of over 89.6% using the Honda/UCSD dataset which runs at 26 FPS with darknet-53 to VGA-resolution images for classification tasks. The experimental results have demonstrated that the FaceNet+SVM model was able to achieve an accuracy of 99.7% using the LFW dataset. On the same dataset, FaceNet+KNN and FaceNet+RF achieve 99.5% and 85.1%, respectively; on the other hand, the FaceNet was able to achieve 99.6%. Finally, the proposed system provides a recognition accuracy of 99.1% and 49 ms runtime when both the face detection and classifications stages operate together. All people commonly have the ability to instantly recognize many objects in an automatic manner because the brain is capable of performing millions of extraordinarily complex operations to perform high quality pattern recognition tasks. To allow computers to be able to emulate some pattern recognition tasks performed by the human brain, several efficient algorithms have been proposed during the last few decades to operate in constrained as well as unconstrained environments, such as video sequences. These efforts began with the work, which established the principles of understanding neuronal activity and the brain operation. Face recognition is a classic problem that has received significant attention in recent years. Nevertheless, face recognition accuracy is influenced by the environment in which it is used. Particularly in unconstrained face recognition, numerous factors affect the accuracy; for example, the face images present different variations such as: pose variation, scale variation, partial occlusion, and complex illumination, resulting in low recognition accuracies. Because of the success of the deep learning architecture, several fields of artificial intelligence have developed deep learning based solutions, such as face recognition. Recognition methods based on deep neural networks have shown many advantages in terms of learning ability, high variability, and generalization. However, efficient algorithms still present several limitations when real-time operation is required, as well as in an unconstrained environments because it requires

achieving high accuracy and computational efficiency. Therefore, face recognition still represents an important challenge in real-time applications, and it is an active research field in the context of computer vision, deep learning, real-time systems, etc.

Findings:

In the system proposed by the author the complexity of face recognition systems depends on the interaction of several less complex sub-systems jointly operating to solve more complex tasks; in particular, we can generalize two fundamental operations involved in the facial recognition tasks: face detection and face recognition. A face recognition system is limited in minor conditions and required to detect faces in images (or videos) regardless of the facial object appearance. Secondly, face images are then processed; subsequently, face features are extracted with a feature extractor. Finally, the system compares the extracted features with the enrolled faces to make face matching. This is why proposed system is evolved to benefit the system performance and achieve high recognition rates. [1]

Yassin Kortli et al. mentioned that over the past few decades, interest in theories and algorithms for face recognition has been growing rapidly. Video surveillance, criminal identification, building access control, and unmanned and autonomous vehicles are just a few examples of concrete applications that are gaining attraction among industries. Various techniques are being developed including local, holistic, and hybrid approaches, which provide a face image description using only a few face image features or the whole facial features.

Proposed System:

The main contribution of this survey is to review some well-known techniques for each approach and to give the taxonomy of their categories. A detailed comparison between these techniques is exposed by listing the advantages and the disadvantages of their schemes in terms of robustness, accuracy, complexity, and discrimination. One interesting feature mentioned by the researcher is about the database used for face recognition. An overview of the most commonly used databases, including those of supervised and unsupervised learning, is given. Numerical results of the most interesting techniques are given along with the context of experiments and challenges handled by these techniques. Finally, a solid discussion is given in the paper about future directions in terms of techniques to be used for face recognition. The objective of developing biometric applications, such as facial recognition, has recently become important in smart cities. In addition, many scientists and engineers around the world have focused on establishing increasingly robust and accurate algorithms and methods for these types of systems and their application in everyday life. All types of security systems must protect all personal data. The most commonly used type for recognition is the password. However, through the development of information technologies and security algorithms, many systems are beginning to use many biometric factors for recognition task. These biometric factors make it possible to identify people's identity by their physiological or behavioral characteristics. They also provide several advantages, for example, the presence of a person in front of the sensor is sufficient, and

there is no more need to remember several passwords or confidential codes anymore. In this context, many recognition systems based on different biometric factors such as iris, fingerprints, voice, and face have been deployed in recent years.

Findings:

Application of the system in real world is difficult to achieve due to limited data and this system tries to identify people based on their biological characteristics which are very attractive because they are easy to use. The human face is composed of different structures and characteristics. For this reason, in recent years, it has become one of the most widely used biometric authentication systems, given its potential in many applications and fields (surveillance, home security, border control, and so on). In our proposed system we try to figure out the real time application location where we can install the system and used it as a biometric identification in future with high authentication procedure. Facial recognition system as an ID (identity) is already being offered to consumers outside of phones, including at airport check-ins, sports stadiums, and concerts. In addition, this system does not require the intervention of people to operate, which makes it possible to identify people only from images obtained from the camera. In addition, many biometric systems that are developed using different types of search provide good identification accuracy. However, it would be interesting to develop new biometric systems for face recognition in order to reach real-time constraints.[2]

HANG DU, et all state that Face recognition (FR) is an extensively studied topic in computer vision. Among the existing technologies of human biometrics, face recognition is the most widely used one in real-world applications. With the great advance of deep convolutional neural networks (DCNNs), the deep learning based methods have achieved significant improvements on various computer vision tasks, including face recognition. In this survey, we focus on 2D image based end-to-end deep face recognition which takes the general images or video frames as input, and extracts the deep feature of each face as output. We provide a comprehensive review of the recent advances of the elements of end-to-end deep face recognition. Specifically, an end-to-end deep face recognition system is composed of three key elements: face detection, face alignment, and face representation.

Proposed system:

In the following, they give a brief introduction of each element. In the face representation stage, the discriminative features are extracted from the aligned face images for recognition. This is the final and core step of face recognition. In early studies, many approaches calculate the face representation by projecting face images into low-dimensional subspace, such as Eigenfaces and Fisherfaces. Later on, handcrafted local descriptors based methods prevail in this area for a detailed review of these traditional methods. In the last few years, the face representation benefits from the development of DCNNs and witnesses great improvements for high performance face recognition. Face detection is the first procedure of the face

recognition system. Given an input image, the face detection aims to find all the faces in the image and give the coordinates of bounding box with a confidence score.

Findings:

The major challenges of face detection contain varying resolution, scale, pose, illumination, occlusion, etc. Some provide a categorization of the deep learning based face detection methods from multiple dimensions, which includes multi-stage, single-stage, anchor-based, anchor-free, multi-task learning, CPU real-time and problem-oriented methods. It is worth noting that there exist overlapping techniques between the categories, because, the categorization is built up from multiple perspectives. In our proposed system we make use of deep learning techniques to rectify the issue of resolution, illumination by making sing of classifiers which utmost try to reduce the error for face recognition. [3]

Madan Lal et al state that with the rapid growth in multimedia contents, among such content face recognition has got much attention especially in past few years. Face as an object consists of distinct features for detection; therefore, it remains most challenging research area for scholars in the field of computer vision and image processing.

Proposed system:

In this proposed system, we have tried to address most endeavoring face features such as pose invariance, aging, illuminations and partial occlusion. They are considered to be indispensable factors in face recognition system when realized over facial images. This paper also studies state of the art face detection techniques, approaches, viz. Eigen face, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Principal Component Analysis (PCA), Independent Component Analysis (ICA), Gabor Wavelets, Elastic Bunch Graph Matching, 3D morphable Model and Hidden Markov Models. In addition to the aforementioned works, we have mentioned different testing face databases which include AT & T (ORL), AR, FERET, LFW, YTF, and Yale, respectively for results analysis. However, aim of this research is to provide comprehensive literature review over face recognition along with its applications. And after in depth discussion, some of the major findings are given in conclusion. The 21st century is a modern and scientific era in which a lot of progress has been achieved as to expedite humans for accomplishing their tasks. In support of above statement, nowadays use of computer technology has been an integral part of life. Computers are being used in pyramids of applications, which range from simple to complex problem solving methods. Among such contributions face recognition technology has emerged as useful tool to recognize features of faces through their inherent traits. And it has been one of the most researched areas in the field of pattern recognition and computer vision. However, due to its wide use in multitude of applications such as in biometrics, information security, law enforcement access control, surveillance system and smart cards. But it possesses many challenges for researcher that needs to be addressed.

Findings:

Face an object depends on facial expressions, which constitute meaningful features. For instance, pose invariance, illuminations and aging which are potential areas that require further investigation over previous work. The result of previous researches reveals that facial expressions are changing with respect to aging; therefore, they could not be permanently modeled in face recognition. The face recognition problem can be categorized into two main phases: 1) face verification and 2) face identification. For example, in real time system, face verification identifies the same person in the scene, and face identification who is this person in that scene. In the first phase it locates a face in an image. Similarly, in the second stage, it extracts features from an image for discrimination. However, some existing recognition methods for authentication suffer from lack of reliability. For instance, smart cards, wallets, keys, tokens use PINs and passwords which are very difficult to remember. In addition, these passwords and codes can easily be forgotten; also these magnetic cards can be misplaced, even robbed or reproduced. As a result this makes them illegible. Contrary to biological characteristics and traits of an individual person they cannot be stolen, forgotten or misplaced. In our proposed system, we find out the loop holes which makes the system working inefficiently with factors such as model creation, dataset evaluation accuracy, which in turn help to sort out the errors at maximum levels. [4]

Zhigang Yu, et al. mentioned that face recognition is a relatively mature technology, which has some applications in many aspects, and now there are many networks studying it, which has indeed brought a lot of convenience to mankind in all aspects. This paper proposes a new face recognition technology. First, a new GoogLeNet-M network is proposed, which improves network performance on the basis of streamlining the network. Secondly, regularization and migration learning methods are added to improve accuracy. The experimental results show that the GoogLeNet-M network with regularization using migration learning technology has the best performance, with a recall rate of 0.97 and an accuracy of 0.98. Finally, it is concluded that the performance of the GoogLeNet-M network is better than other networks on the dataset, and the migration learning method and regularization help to improve the network performance This have been in the era of big data, which has brought about an explosive increase in the amount of information, and in some access control and other aspects, people often use biometrics for identity authentication for a reason because people's faces or fingerprints are unique. In this regard, face recognition is the main recognition method, which brings great convenience to people's life. It mainly uses optical imaging of human faces to perceive and recognize people.

Proposed system:

At present, this technology is mainly applied to criminal investigation, surveillance systems, and secure payment. The traditional face recognition technology is mainly to extract feature points for face recognition and now the main application of deep learning technology. Due to the large amount of data and high computing power, the precision aspects of deep learning have been greatly improved. An improved additive

cosine interval loss function was proposed to improve the additive cosine interval loss function. By subtracting a value from the cosine value of the angle between the feature and the target weight and adding a value to the cosine value of the angle between the feature and the nontarget weight, the value is a number between 0 and 1, and select the best value through experiments to achieve the purpose of reducing the distance between classes and increasing the distance between classes.

Findings:

A face recognition model combining singular value face and attention convolutional neural network is proposed. The algorithm first uses a normalized singular value matrix to represent facial features, then inputs the features into the deep convolutional neural network with the attention module added, and improves the robustness of the network through cross-channel and spatial information fusion. Finally, the classification and recognition of face images is completed through the iterative training of the network. Through experiments on two commonly used databases, it is confirmed that the algorithm proposed in this paper has better recognition performance and better lighting robustness. Use of algorithm in the system elaborates the feature accuracy with higher efficiency. [5]

Palli Suryachandra et al. sentiment analysis uses information retrieval and computational linguistics. Sentiment analysis has advantages in various forms such as in marketing or for business purposes. In marketing, it is used to notice about the favorable or negative points about their new product which helps to determine how successful the new product is. A specific view or notion can be depicted as ideas prompted, opinions, judgements or coloured by emotions or emotions. In Computational Linguistics, the core is on feelings instead of sentiments, opinions or perceptions. The terms opinions and sentiments are frequently availed substitutable.

Proposed system:

The proposed system makes use of the information of a text which is divided into two categories. 1. Based on text 2. Based on persuasion. Whereas actualities or facts are observational utterances about events, entities and their opinions, characteristics are particular utterances that depict opinions of people, events and their properties, feelings towards entities or appraisals. A persuasion can be depicted by the following four terms: Sentiment, Claim, Holder and Topic . The Holder affirms a fact about a Topic, and frequently relates a persuasion, such as `bad' or `good', with the affirm. It depicts a persuasion as an implicit or explicit aspect in text of the holder's negative, positive or neutral notice into the requirement about the topic. Sentiment analysis suits with computational operation of persuasion, sentiment, opinion and individuality in text. The document inception is likely in the pattern of unstructured data. Opinion mining or Sentiment analysis hit an important role in our regular selection construction progress.

Proposed system:

The proposed system mentioned selections may has scope from acquisition a commodity such as mobile phone to scrutinize the movie to constructive contribution; all the choices will have a great favor on the

regular life. Now-a-days folk before purchasing a service/commodity will accomplish a glimpse to scrutinize websites, online shopping websites and social media websites to get an assessment related about the required commodity or utility in merchandise.

Findings:

Opinion Analysis or Sentiment Analysis handles many technological objections such as feature extraction, opinion orientation classification and object identification. Generally Opinion analysis is implemented by using unsupervised learning and supervised learning, those are generally Support Vector Machine, naive Bayes and Neural Networks. Among the given existing techniques which are in above, Support Vector Machine is recognized to be further applicable for Opinion Analysis. Use of Naives Bayes helps to modify the proposed in future for opinion analysis and helps to feature extraction in dominant way. [6]

Janyce Wiebe et al state that there has been a recent swell of interest in the automatic identification and extraction of opinions, emotions, and sentiments in text. Motivation for this task comes from the desire to provide tools for information analysts in government, commercial, and political domains, who want to automatically track attitudes and feelings in the news and on-line forums. How do people feel about recent events in the Middle East? Is the rhetoric from a particular opposition group intensifying? What is the range of opinions being expressed in the world press about the best course of action in Iraq? A system that could automatically identify opinions and emotions from text would be an enormous help to someone trying to answer these kinds of questions. The private states in these sentences are expressed entirely by the words and the style of language that is used. Although the writer does not explicitly say that he hates Sharon, his choice of words clearly demonstrates a negative attitude toward him. Describing the election as “daylight robbery” clearly reflects the anger being experienced by the speaker, Tsvangirai. As used in these sentences, the phrases “The time has come,” “gentlemen,” “the assassin,” “injustice cannot last long,” “fraud,” and “daylight robbery” are all expressive subjective elements.

Proposed system:

Expressive subjective elements are used by people to express their frustration, anger, wonder, positive sentiment, mirth, etc., without explicitly stating that they are frustrated, angry, etc. Sarcasm and irony often involve expressive subjective elements. The annotation scheme includes an agent frame for noun phrases that refer to sources of private states and speech events, i.e., for all noun phrases that act as the experience of a private state, or the speaker/- writer of a speech event. Each agent frame generally has two slots. The text anchor slot includes a pointer to the span of text that denotes the noun phrase source. The source slot contains a unique alpha-numeric ID that is used to denote this source throughout the document. The agent frame associated with the first informative (e.g., non-pronominal) reference to this source in the document includes an id slot to set up the document-specific source-id mapping.

Findings:

The source of a speech event is the speaker or writer. The source of a private state is the experience of the private state, i.e., the person whose opinion or emotion is being expressed. Obviously, the writer of an article is a source, because he or she wrote the sentences composing the article, but the writer may also write about other people's private states and speech events, leading to multiple sources in a single sentence. Proposed system make use of emotion sentiments with video framing and the future aspects can be combined with text and video to achieve better accuracy than the previous one.[7]

Scott Brave et al. mentioned that Emotion is a fundamental component of being human. Joy, hate, anger, and pride, among the plethora of other emotions, motivate action and add meaning and richness to virtually all human experience. Traditionally, human–computer interaction (HCI) has been viewed as the "ultimate" exception; users must discard their emotional selves to work efficiently and rationality with computers, the quintessentially unemotional artifact. Emotion seemed at best marginally relevant to HCI and at worst oxymoronic. Recent research in psychology and technology suggests a different view of the relationship between humans, computers, and emotion. After a long period of dormancy and confusion, there has been an explosion of research on the psychology of emotion. Emotion is no longer seen as limited to the occasional outburst of fury when a computer crashes inexplicably, excitement when a video game character leaps past an obstacle, or frustration at an incomprehensible error message. It is now understood that a wide range of emotions plays a critical role in every computer-related, goal-directed activity, from developing a three-dimensional (3D) CAD model and running calculations on a spreadsheet, to searching the Web and sending an e-mail, to making an online purchase and playing solitaire. Indeed, many psychologists now argue that it is impossible for a person to have a thought or perform an action without engaging, at least unconsciously, his or her emotional systems.

Proposed system:

The literature on emotions and computers has also grown dramatically in the past few years, driven primarily by advances in technology. Inexpensive and effective technologies that enable computers to assess the physiological correlates of emotion, combined with dramatic improvements in the speed and quality of signal processing, now allow even personal computers to make judgments about the user's emotional state in real time. Multimodal interfaces that include voices, faces, and bodies can now manifest a much wider and more nuanced range of emotions than was possible in purely textual interfaces. Indeed, any interface that ignores a user's emotional state or fails to manifest the appropriate emotion can dramatically impede performance and risks being perceived as cold, socially inept, untrustworthy, and incompetent. One of the most important effects of emotion lies in its ability to capture attention. Emotions have a way of being completely absorbing. Functionally, they direct and focus our attention on those objects and situations that

have been appraised as important to our needs and goals so that we can deal with them appropriately.

Findings:

Emotion-relevant thoughts then tend to dominate conscious processing—the more important the situation, the higher the arousal, and the more forceful the focus. In an HCI context, this attention-getting function can be used advantageously, as when a sudden beep is used to alert the user, or can be distracting, as when a struggling user is frustrated and can only think about his or her inability. A mood's effect on judgment, combined with its effect on memory, can also influence the formation of sentiments. Sentiments are not necessarily determined during interaction with an object; they often are grounded in reflection. This is important to consider when conducting user tests, as the mood set by the interaction immediately prior to a questionnaire may bias like/dislike assessments of earlier interactions. Thus, varying order of presentation ensures both that later stimuli do not influence the assessment of earlier stimuli, and that earlier stimuli do not influence the experience of later stimuli (as discussed earlier). There are also a host of more abstract needs underlying, and often adjacent to, application-specific goals. A user may have a strong need to feel capable and competent, maintain control, learn, or be entertained. A new user typically needs to feel comfortable and supported, while an expert is more focused on aesthetic concerns of efficiency and elegance. Acknowledging these more abstract goals in interface design can be as instrumental in determining a user's affective state as meeting or obstructing application-specific goals. Maslow's hierarchy presents a useful starting place for considering the structure of these more abstract user needs. In the further work grouped an individual's basic needs into eight categories:

- Physiological: hunger, thirst, bodily comforts, etc.
- Safety/security: being out of danger
- Social: affiliate with others, be accepted
- Esteem: to achieve, be competent, gain approval and recognition
- Cognitive: to know, to understand, and explore
- Aesthetic: symmetry, order, and beauty
- Self-actualization: to find self-fulfillment and realize one's potential
- Transcendence: to help others find self-fulfillment and realize their potential.

Facial expression with video content can be able to be useful to find out the various expressions in context with the input text value and proposed system wasdesigned to extend the terms in future. [8]

Bo Pang et al. in his study, we focused on features based on unigrams (with negation tagging) and bigrams. Because training MaxEnt is expensive in the number of features, we limited consideration to

(1) the 16165 unigrams appearing at least four times in our 1400- document corpus (lower count cutoffs did not yield significantly different results), and

(2) the 16165 bigrams occurring most often in the same data (the selected bigrams all occurred at least seven times). Note that we did not add negation tags to the bigrams, since we consider bigrams (and n-grams in general) to be an orthogonal way to incorporate context.

Proposed System:

On the other hand, we were not able to achieve accuracies on the sentiment classification problem comparable to those reported for standard topic-based categorization, despite the several different types of features we tried. Unigram presence information turned out to be the most effective; in fact, none of the alternative features we employed provided consistently better performance once unigram presence was incorporated. Interestingly, though, the superiority of presence information in comparison to frequency information in our setting contradicts previous observations made in topic-classification work. Today, very large amounts of information are available in on-line documents. As part of the effort to better organize this information for users, researchers have been actively investigating the problem of automatic text categorization. The bulk of such work has focused on topical categorization, attempting to sort documents according to their subject matter (e.g., sports vs. politics). However, recent years have seen rapid growth in on-line discussion groups and review sites (e.g., the New York Times' Books web page) where a crucial characteristic of the posted articles is their sentiment or overall opinion towards the subject matter— for example, whether a product review is positive or negative. Labeling these articles with their sentiment would provide succinct summaries to readers; indeed, these labels are part of the appeal and value-add of such sites as www.rottentomatoes.com, which both labels movie reviews that do not contain explicit rating indicators and normalizes the different rating schemes that individual reviewers use.

Findings:

Sentiment classification would also be helpful in business intelligence applications and recommender systems, where user input and feedback could be quickly summarized; indeed, in general, free-form survey responses given in natural language format could be processed using sentiment categorization. Moreover, there are also potential applications to message filtering; for example, one might be able to use sentiment information to recognize and discard "flames". In our proposed system keeping an idea of business applications our system make use of this recognition for higher levels of authentication from management level to the end user. [9]

Pravesh Kumar Singh et al thinks that generally individuals and companies are always interested in other's opinion like if someone wants to purchase a new product, then firstly, he/she tries to know the reviews i.e., what other people think about the product and based on those reviews, he/she takes the decision. Similarly, companies also excavate deep for consumer reviews. Digital ecosystem has a plethora for same in the form of blogs, reviews etc. A very basic step of opinion mining and sentiment analysis is feature extraction. In text classification sometimes data are linearly divisible, for very high dimensional problems and for multi-dimensional problems data are also separable linearly. Generally (in maximum cases) the opinion mining

solution is one that classifies most of the data and ignores outliers and noisy data. If a training set data say D cannot be separated clearly then the solution is to have fat decision classifiers and make some mistake. Non-linear classifiers are given by the using kernel to max margin hyperplanes and given a kernel trick i.e., every dot product is replaced by non-linear kernel function. When this case is apply then the effectiveness of SVM lies in the selection of kernel and soft margin parameters. Basically SVM relevant for two class tasks but for the multiclass problems there is multiclass SVM is available. In the multi class case labels are designed to objects which are drawn from a finite set of numerous elements. These binary classifiers might be built using two classifiers:

- Distinguishing one versus all labels and
- Among each pair of classes one versus one.

Proposed System:

Clustering is an unsupervised learning method and has no labels on any point. Clustering technique recognizes the structure in data and group, based on how nearby they are to one another. So, clustering is process of organizing objects and instances in a class or group whose members are similar in some way and members of class or cluster is not similar to those are in the other cluster. This method is an unsupervised method, so one does not know that how many clusters or groups are existing in the data. Using this method one can organize the data set into different clusters based on the similarities and distance among data points.

Findings:

The important part of gathering information always seems as, what the people think. The rising accessibility of opinion rich resources such as online analysis websites and blogs means that, one can simply search and recognize the opinions of others. One can precise his/her ideas and opinions concerning goods and facilities. These views and thoughts are subjective figures which signify opinions, sentiments, emotional state or evaluation of someone. In this paper, different methods for data (feature or text) extraction are presented. Every method has some benefits and limitations and one can use these methods according to the situation for feature and text extraction. The main intension in the above research work is to collect the authenticate information and data for further processing and this makes our system more convenient since it not only can able to predict the train data but can be modified to train the real time data also. [10]

Some reviews or comments are classified into positive and negative. Traditionally the document classification was performed on the topic basis but later research started working on opinion basis. Following machine learning methods Naive Bayes, Maximum Entropy Classification (MEC), and Support Vector Machine (SVM) are used for sentiment analysis. The conventional method of document classification based on topic is tried out for sentiment analysis. The major two classes are considered i.e. positive and negative and classify the reviews according to that. Naïve Bayes is best suitable for textual classification, clustering for consumer services and Support Vector Machine for biological reading and interpretation. The four methods discussed in the paper are actually applicable in different areas like

clustering is applied in movie reviews and Support Vector Machine (SVM) techniques is applied in biological reviews & analysis. Though field of opinion mining is latest technology, but still it provides diverse methods available to provide a way to implement these methods. Some researcher has presented experiments on prosody-based Automatic Personality Perception, i.e., on automatic prediction of personality traits attributed by human listeners to unknown speakers. The APP results show an accuracy ranging between 60 and 72 percent (depending on the trait) in predicting whether a speaker is perceived to be high or low with respect to a given trait. The most probable reason is that the corpus includes two categories of speakers (professional and nonprofessional ones) that differ in terms of characteristics typically related to the trait e.g. efficiency, reliability etc.. Some researchers investigated the role of the context information (e.g. subject, gender, features representing local and global aspects of prior dialogue) on audio affective recognition. Still the natural interaction of human beings has subtle emotions and six basic emotions (described above) seldom occur. Thus, mostly researchers include acoustic features with linguistic features (language dependent) to improve recognition process. Typical examples of linguistic-paralinguistic-fusion methods are used vocal features, spoken contents (words) for sentiment analysis over speech signals. This paper focuses on recognizing emotions from spoken language. The importance of emotion recognition from human speech has increased significantly with the need to improve both the naturalness and efficiency of spoken language human-machine interfaces Another challenge is how to reliably extract these linguistic and paralinguistic feature from the audio channel. Several prosodic features are taken into account from the emotion detection within a speech signal but the research activities are not up to mark. The automatic extraction of high-level underlying semantic linguistic information such as dialogue act, corrections, repetitions, and syntactic information has to be further studied. Some paper shows that valence ratings and skin conductance were able to characterize the emotional nature of the content. A repeated measure ANOVA was run on valence and skin conductance (normalized separately for each participant using z-scores). Two participants were removed from the analysis due to a lack of variation in their raw skin conductance. The mean of valence and arousal successfully differentiated the emotional nature of different stimuli, especially negative as opposed to nonnegative, and detected cultural biases. The analysis of simultaneous agreement in both valence and arousal led to finding moments of universal experiences. The moments where a culture had high cohesion in their valence and arousal were moments were there were smaller differences between the emotional experiences between cultures.

Two concepts are fundamental to this study: image classification and natural language understanding. Since the success of AlexNet in 2012, the application of convolutional neural network models in image processing have been a dominant area of research. Similarly, sequence neural network models have dominated recent research in NLU. This study combines these two areas of research, focusing on improving image classification models with joint learned representations with text. In addition, exceptionally large models are required to train modern neural network models.

CHAPTER 3

SYSTEM DEVELOPMENT

3.1 Existing System

Traditionally the document classification was performed on the topic basis but later research started working on opinion basis. Following machine learning methods Naive Bayes, Maximum Entropy Classification (MEC), and Support Vector Machine (SVM) are used for sentiment analysis. The conventional method of document classification based on topic is tried out for sentiment analysis. The major two classes are considered i.e. positive and negative and classify the reviews according to that. In [5], Naïve Bayes is best suitable for textual

classification, clustering for consumer services and Support Vector Machine for biological reading and interpretation. The four methods discussed in the paper are actually applicable in different areas like clustering is applied in reviews and Support Vector Machine (SVM) techniques is applied in biological reviews & analysis. Though field of opinion mining is latest technology, but still it provides diverse methods available to provide a way to implement these methods.

3.2 Existing Technology or Algorithms

There is a large literature on semi-supervised learning techniques. For sake of brevity, we discuss only two important paradigms, and we refer to [5] for a recent book on the subject. When using generative models for semi-supervised learning a straightforward approach is to treat the class label of unlabeled data as a missing variable, see e.g. [1, 15]. The class conditional models over the features can then be iteratively estimated using the EM algorithm. In each iteration the current model is used to estimate the class label of unlabeled data, and then the class conditional models are updated given the current label estimates. This idea can be extended to our setting where we have variables that are only observed for the training data. The idea is to jointly predict the class label and the missing text features for the test-data, and then marginalize over the unobserved text features. These methods are known to work well in cases where the model fits the data distribution, but can be detrimental in cases where the model has a poor fit. Current state-of-the-art image classification methods are discriminative ones that do not estimate the class conditional density models, but directly estimate a decision function to separate the classes. However, using discriminative classifiers, the EM method of estimating the missing class labels used for generative models does not apply: the EM iterations immediately terminate at the initial classifier. Co-training [4] is a semi-supervised learning technique that does apply to discriminative classifiers, and is designed for settings like ours where the data is described using several different feature sets. The idea is to learn a separate classifier using each feature set, and to iteratively add training examples for each classifier based on the output of the other classifier. In particular, in each iteration the examples that are most confidently classified with the first classifier are added as labeled examples to the training set of the second classifier, and vice-versa. A potential drawback of the co-training is that it relies on the classifiers over the separate feature sets to be accurate, at least among the most confidently classified examples. In our setting we find that for most categories one of the

two feature sets is significantly less informative than the other. Therefore, using the classifier based on the worse performing feature set might provide erroneous labels to the classifier based on the better performing feature set, and its performance might be deteriorated. In the next section we present a semi-supervised learning method that uses both feature sets on the labeled examples, and we compare it with co-training in our experiments.

3.3 Hardware and Software Requirements

Software Requirements

- Python
- Software libraries

  Pickle, Argparse, Sys, Numpy, Tensorflow, Tqdm
- Datasets

  Possibly to refer MOSI, MOSEI and IEMOCAP datasets.

- Interface standards

  Dnjango Framework

Hardware Requirements

- CORE I5 PROCESSOR
- 8 GB Ram / 500GB Hardisk

3.4 Proposed System

Speech signals convey not only words but also emotions. Various analysis and models had been submitted and explored for Textual Analysis but this analysis is incomplete due to ignorance of Sentiments involved and result may not be reliable and in addition Textual Analysis only focus on word content and thereby ignores the acoustic features of speech . Thus it needs analysis of Sentiments as well as text simultaneously.

The present schemes are dealing with only facial image which holds good quality of resolution. But to be practical this is not the case indeed existing system may fail due to several reasons such as low resolution, fast movement of subject etc. Another issue with respect to present system in how to reliable extract there linguistic and paralinguistic features from the image data with many features that has been from the emotion detection within a certain image quality, but such research activity were not satisfied to overcome the traditional challenges. There is need of an automatic extraction of linguistic information from the image data. Spoken contents can be extracted from speech signals with the help of speech to text converter. In

order to implement an effective multimodal, the fusion of different modes must be done with suitable joint feature vectors which is composed of various features of different modalities with differ in time scales. Furthermore to make multimodal effective and realistic other parameters can be considered with respect to human entity such as age, gender. One mode doesn't give sufficient solution. There is need to consider other modes also such as audio/video. Social media are a huge untapped source of user opinion for various products and services. Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. Speech signals convey not only words but also emotions. Various analysis and models had been submitted and explored for Textual Analysis but this analysis is incomplete due to ignorance of Sentiments involved and result may not be reliable and in addition Textual Analysis only on focus is on word content and thereby ignores the acoustic features of speech . Thus it needs analysis of Sentiments as well as text simultaneously.

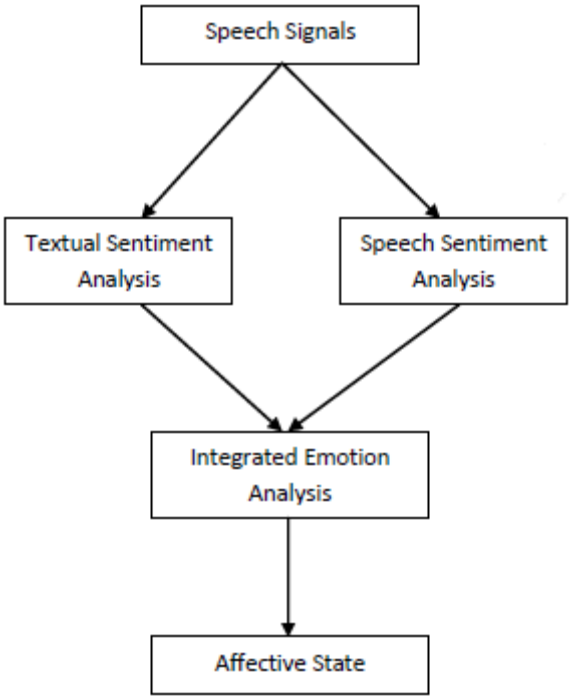

Fig. 3.1 Flow of the proposed system

In integration, we can integrate the results of all these modes. Facial expressions helps to understand the present emotion of human thus various methods have been proposed to identify typical part of face and movement of specific points associated with different emotional states. Often techniques used are sign judgement in which method describes the appearance in spite the meaning of shown behaviour, message judgement which addresses the interpretation of shown behaviour. One more technique is known as FACS (Facial Action Coding System) which does manual labelling of facial behaviour, with respect to sign judgement. The movement of certain parts of facial muscles are encoded by FACS.

3.5 Dataset

User's opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, product review sites, micro blogs provide a good understanding of thereception level of the products and services.

A. Blogs contains reviews of many product and issue. These blogs contains large amount of opinionated text. Blogger opens a topic in discussion and records daily events for opinions, feelings, emotions etc, so it is very important to apply sentiment features over blog data [15].

B. Review sites site contains views of people about product or topic. These views are in form of comments which are in unstructured format. S, people read those comments and makes purchasing decision. E-commercewebsites1, 2, 3, 4 hosts millions of product reviews given by customer such data is used in sentiment analysis. Other than these the available are professional review sites5, 6 and consumer opinion sites on different topics.

C. Dataset- Most of the work in the field uses movie reviews data for classification. Movie review data is available asdataset7. Other dataset which is available online is multidomain sentiment (MDS) dataset8.The MDS dataset includes four different types of product reviews which are summarized from Amazon.com including Books, cloths, Electronics and Decoration1 www.amazon.com2 www.yelp.com3 www.CNETdownload.com4 www.reviewcentre.com5 www.dpreview.com6 www.zdnet.com7 www.cs.cornell.edu/People/pabo/movie-review-data8 www.cs.jhu.edu/mdredze/datasets/sentimentappliances, with 1000 positive and 1000 negative reviews for each domain.

D.Micro-blogging- These are sites which allow user post small messages as a status e.g. tweets of tweeter. Sometimes these twits express opinions. So, Twitter messages are studied to classifysentiments.4.3 Technique usedThere are different algorithms you can implement in sentiment analysis models, depending on how much data you need to analyze, and how accurate you need your model to be. We'll go over some of these in more detail, below.

3.6 Python

Given the context of processing large amounts of data, memory management became prioritary. Consequently, a programming language which is capable to handle processing and storage of these amounts of data was imperative. An iterative system approach, when processing a list of items, has to firstly store it, which requires memory. In these regards, Python provides generators, which are particularly useful when processing large amounts of data, passing the source data through the processing chain, one item at a time, storing only the results of the processing chain [11].

Considering the above argument, Python proves capable of efficiently managing memory, a task crucial for the „real time" component of the project. A drawback of Python is that it is an interpreted

language, which by contrast, is slower than compiled languages (such as C, Java, etc.). The developers community considered the disadvantage, and proposed different ways to improve Python‟s speed. As such, projects like Numba and PyPy are viable solutions. The creator of Python, Guido van Rossum recognizes the improvements added to Python and states that PyPy is the best way to obtain high-performance systems while using Python. Furthermore, advanced libraries for data processing such as NumPy and SciPy were developed by the scientific community and domain experts. Such tools proved to be helpful during the development stage, and reinforced the reasoning of choosing Python.

3.7 Libraries and packages used

3.7.1 Numpy

NumPy is a basic level external library in Python used for complex mathematical operations. NumPy overcomes slower executions with the use of multi-dimensional array objects. It has built-in functions for manipulating arrays. It can convert different algorithms to can into functions for applying on arrays.NumPy has applications that are not only limited to itself. It is a very diverse library and has a wide range of applications in other sectors.

Numpy can be put to use along with Data Science, Data Analysis and Machine Learning. It is also a base for other python libraries. These libraries use the functionalities in NumPy to increase their capabilities.

3.7.2 Tensorflow

ensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

TensorFlow provides a collection of workflows to develop and train models using Python or JavaScript, and to easily deploy in the cloud, on-prem, in the browser, or on-device no matter what language you use.

3.7.3 CV2

TensorFlow provides a collection of workflows to develop and train models using Python or JavaScript, and to easily deploy in the cloud, on-prem, in the browser, or on-device no matter what language you use.

OpenCV has a function to read video, which is cv2.VideoCapture(). We can access our webcam using pass 0 in the function parameter. If you want to capture CCTV footage then we can pass RTSP url in the function parameter, which is really useful for video analysis.

3.7.4 vaderSentiment

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is. Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker.

3.7.5 Keras

Keras is an open-source high-level Neural Network library, which is written in Python is capable enough to run on Theano, TensorFlow, or CNTK. It was developed by one of the Google engineers, Francois Chollet. It is made user-friendly, extensible, and modular for facilitating faster experimentation with deep neural networks. It not only supports Convolutional Networks and Recurrent Networks individually but also their combination. Keras can be developed in R as well as Python, such that the code can be run with TensorFlow, Theano, CNTK, or MXNet as per the requirement. Keras can be run on CPU, NVIDIA GPU, AMD GPU, TPU, etc. It ensures that producing models with Keras is really simple as it totally supports to run with TensorFlow serving, GPU acceleration (WebKeras, Keras.js), Android (TF, TF Lite), iOS (Native CoreML) and Raspberry Pi. Keras being a model-level library helps in developing deep learning models by offering high-level building blocks. All the low-level computations such as products of Tensor, convolutions, etc. are not handled by Keras itself, rather they depend on a specialized tensor manipulation library that is well optimized to serve as a backend engine. Keras has managed it so perfectly that instead of incorporating one single library of tensor and performing operations related to that particular library, it offers plugging of different backend engines into Keras.

3.8 Algorithm used

3.8.1 SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM can be of two types:

- Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

"Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate.

3.9 Classifier Used

3.9.1 Haar Cascade

Haar Cascade is a machine learning-based approach where a lot of positive and negative images are used to train the classifier. Positive images – These images contain the images which we want our classifier to identify. Negative Images – Images of everything else, which do not contain the object we want to detect.
The algorithm can be explained in four stages:

- Calculating Haar Features
- Creating Integral Images
- Using Adaboost
- Implementing Cascading Classifiers

It's important to remember that this algorithm requires a lot of positive images of faces and negative images of non-faces to train the classifier, similar to other machine learning models.

- **Calculating Haar Features**

  The first step is to collect the Haar features. A Haar feature is essentially calculations that are performed on adjacent rectangular regions at a specific location in a detection window. These features can be difficult to determine for a large image. This is where **integral images** come into play because the number of operations is reduced using the integral image

- **Creating Integral Images**

  Without going into too much of the mathematics behind it (check out the paper if you're interested in that), integral images essentially speed up the calculation of these Haar features. Instead of computing at every pixel, it instead creates sub-rectangles and creates array references for each of those sub-rectangles. These are then used to compute the Haar features. It's important to note that nearly all of the Haar features will be **irrelevant** when doing object detection, because the only features that are important are those of the object.

- **Adaboost Training**

  Adaboost essentially chooses the best features and trains the classifiers to use them. It uses a combination of "weak classifiers" to create a "strong classifier" that the algorithm can use to detect objects. Weak learners are created by moving a window over the input image, and computing Haar features for each subsection of the image. This difference is compared to a learned threshold that separates non-objects from objects. Because these are "weak classifiers," a large number of Haar features is needed for accuracy to form a strong classifier.

- **Implementing Cascading Classifiers**

  The cascade classifier is made up of a series of stages, where each stage is a collection of weak learners. Weak learners are trained using boosting, which allows for a highly accurate classifier from the mean prediction of all weak learners. Based on this prediction, the classifier either decides to indicate an object was found (positive) or move on to the next region (negative). Stages are designed to reject negative samples as fast as possible, because a majority of the windows do not contain anything of interest. It's important to maximize a **low false negative rate**, because classifying an object as a non-object will severely impair your object detection algorithm. A video below shows Haar cascades in action. The red boxes denote "positives" from the weak learners.

3.10 Sentiment analysis algorithms fall into one of three buckets:

Rule-based: these systems automatically perform sentiment analysis based on a set of manually crafted rules.

Automatic: systems rely on machine learning techniques to learn from data.

Hybrid systems combine both rule-based and automatic approaches.

Rule-based Approaches Usually, a rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion. These rules may include various NLP techniques developed in computational linguistics, such as: *Stemming*, *tokenization*, *part-of-speech tagging* and *parsing*. Lexicons (i.e.

lists of words and expressions).Here's a basic example of how a rule-based system works: Defines two lists of polarized words (e.g. negative words such as *bad*, *worst*, *ugly*, etc and positive words such as *good*, *best*, *beautiful*, etc).Counts the number of positive and negative words that appear in a given text.If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment. Rule-based systems are very naive since they don't take into account how words are combined in a sequence. Of course, more advanced processing techniques can be used, and new rules added to support new expressions and vocabulary. However, adding new rules may affect previous results, and the whole system can get very complex. Since rule-based systems often require fine-tuning and maintenance, they'll also need regular investments.

Automatic Approaches Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral. Queries of products could be associated with the images, titles, and descriptions, returning a much richer set of data. Continued improvements in image classification model development have progressed the realm of computer vision centered on deep learning. Approaches to enhancing deep learning models such as leveraging statistical methods to distribute spatial characteristics of images within convolutional layers have provided significant impact to this effort. Additionally, the advancement of deep learning tasks to solve Natural Language Processing problems using expansive lexical and semantic representations of language structures has been increasingly and reliably implemented for extracting meaning from vectorized character and word embeddings within dimensional space. Overlap in the foundational implementations of these two branched technologies has enabled the shared learning from each to impact the results of the other, in collaboration.

CHAPTER 4

IMPLEMENTATION AND RESULT

4.1 Screenshot

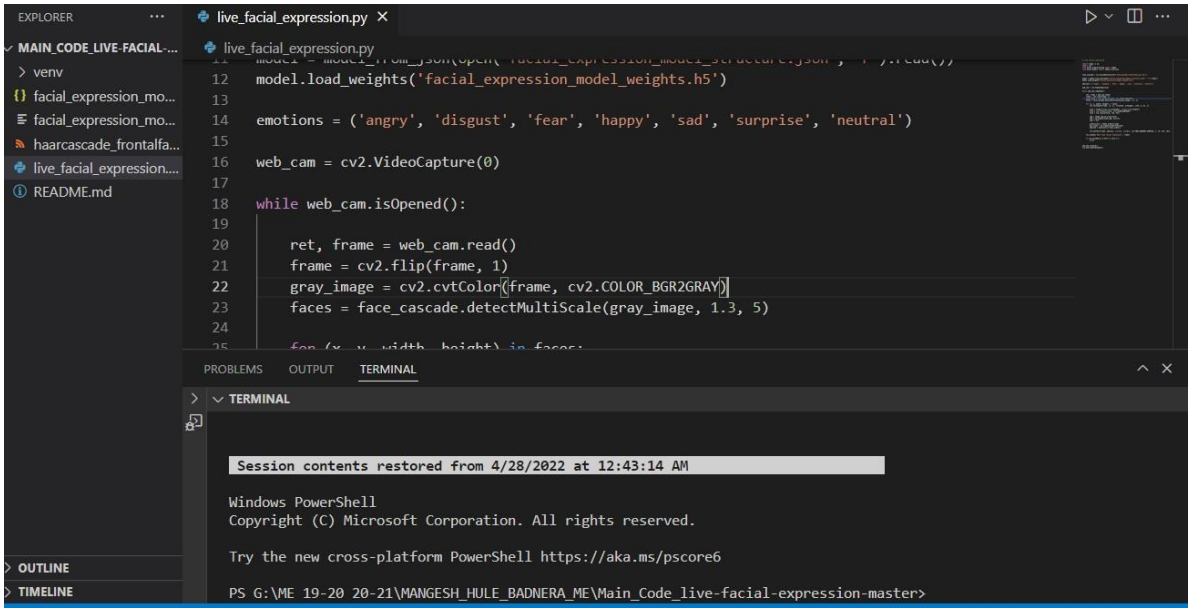

Fig. 4.1 Editor Window for coding

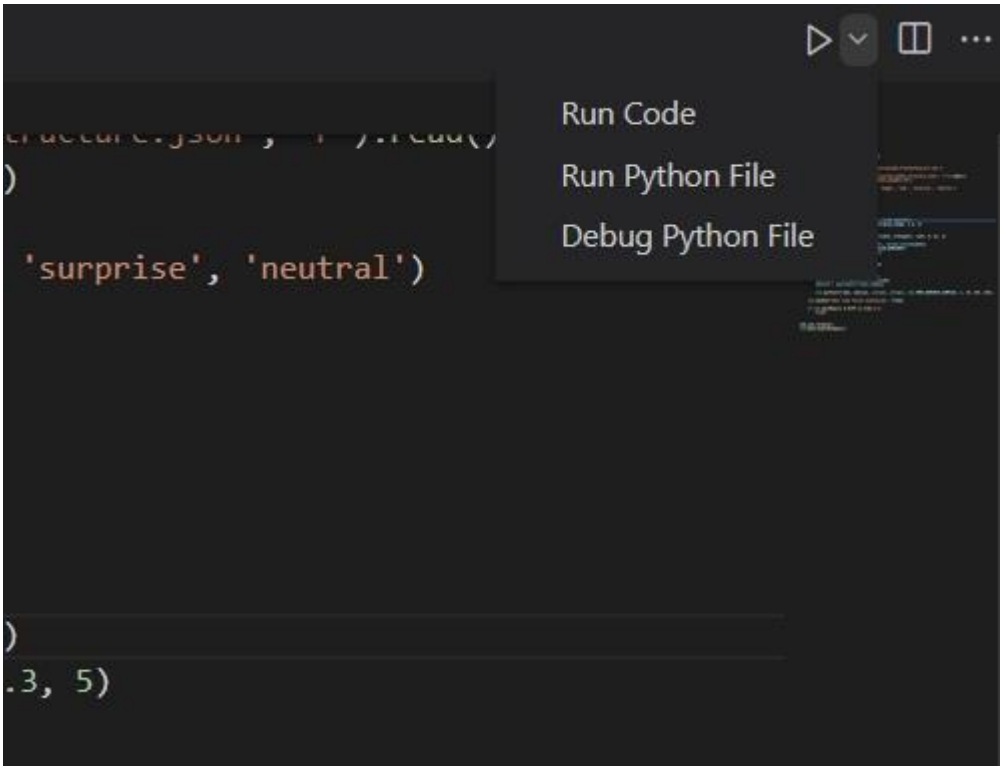

Fig. 4.2 Run with python code

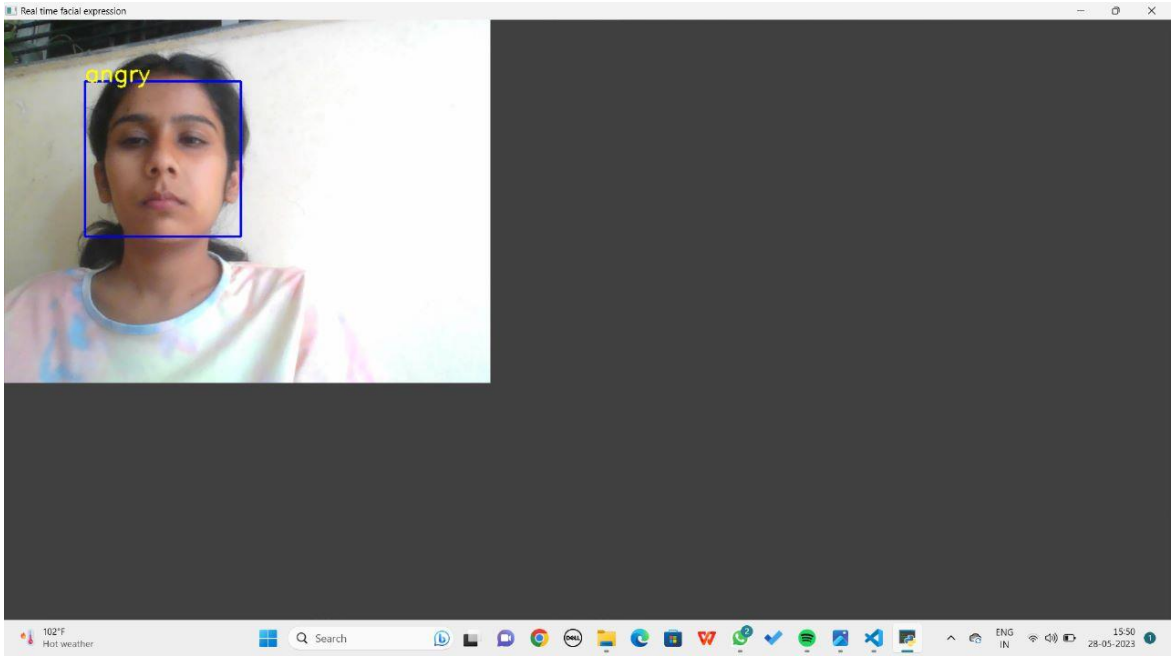

Fig. 4.3 Angry Face detection

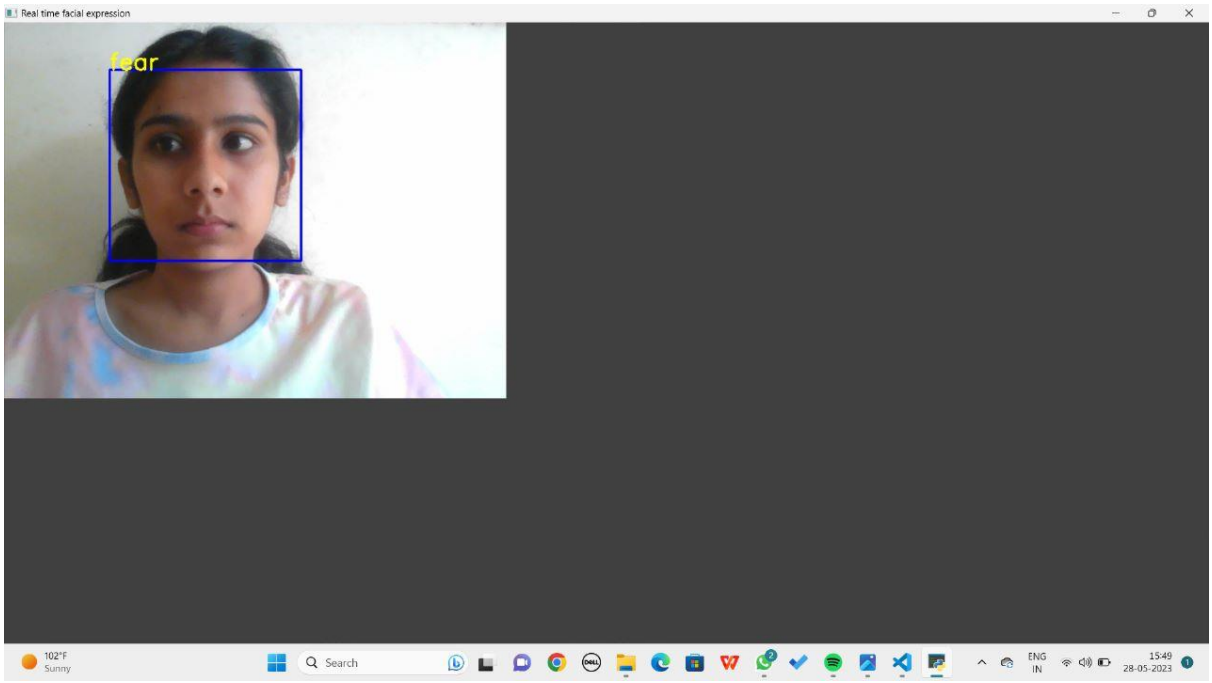

Fig. 4.4 Fear face detection

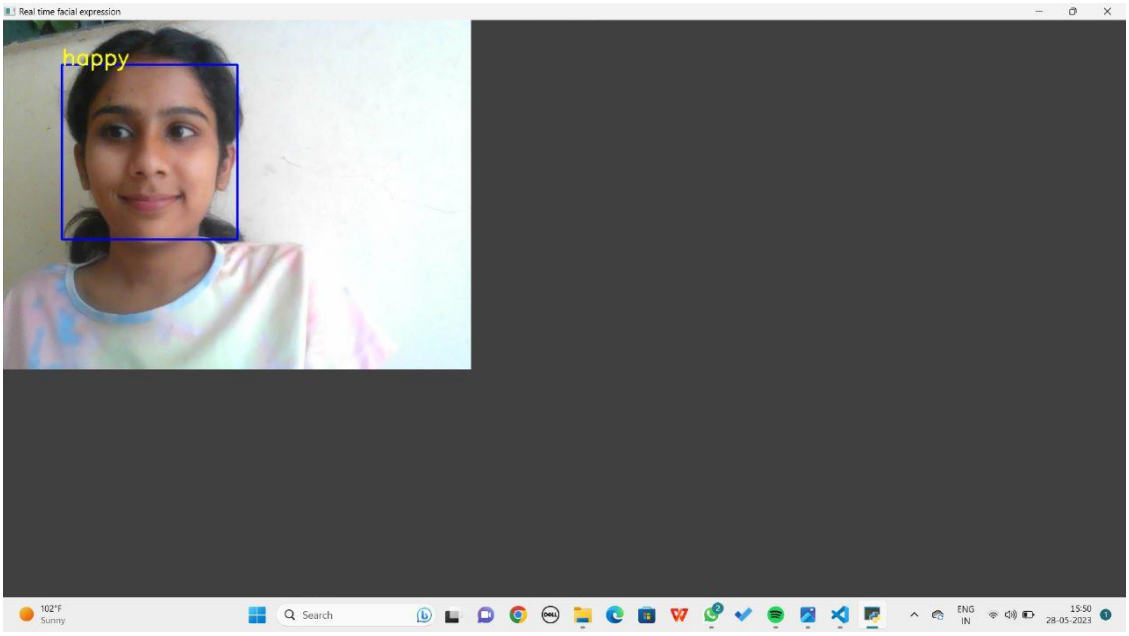

Fig. 4.5 Happy face detection

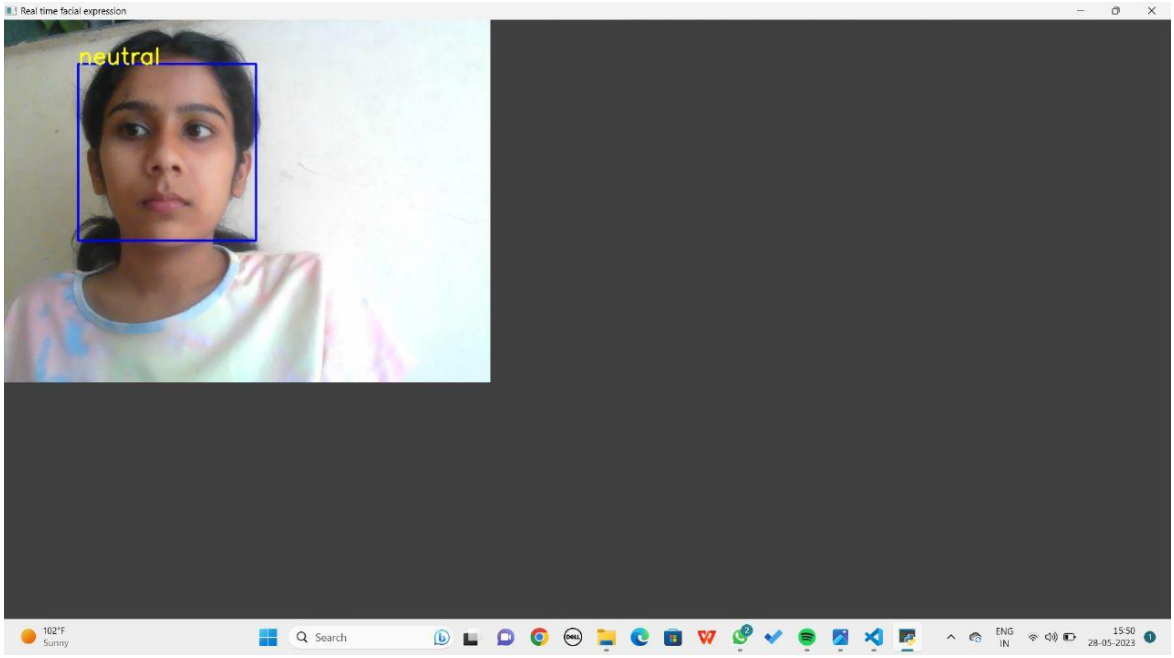

Fig.4.6 Neutral state of sentiment

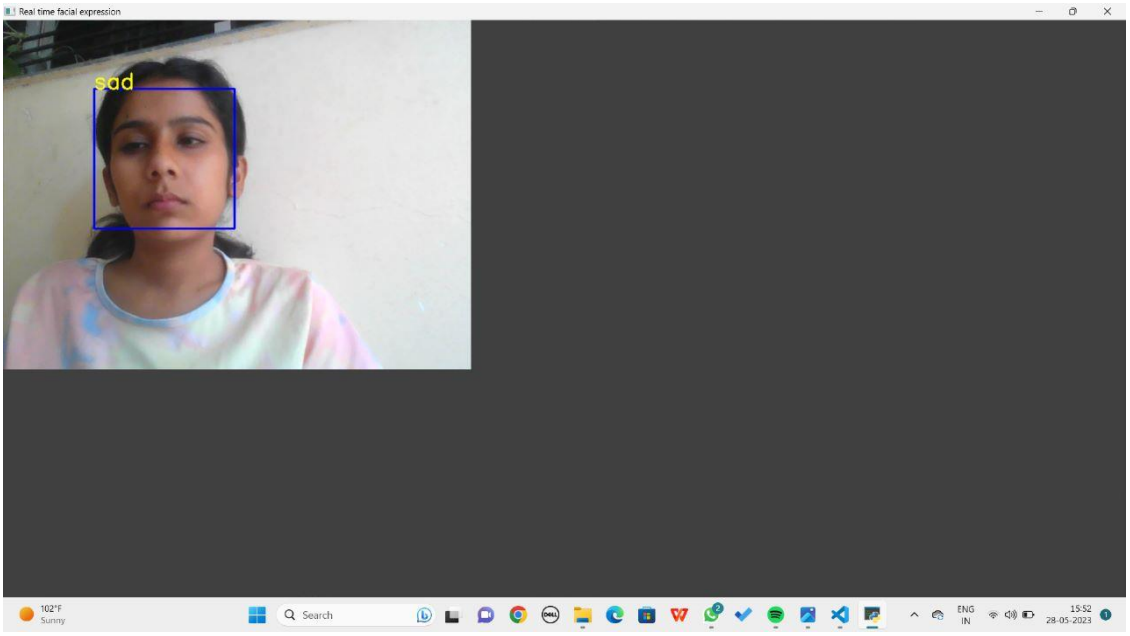

Fig. 4.7 Sad Face of detection

CHAPTER 5

APPLICATION OF THE SYSTEM

While face recognition has been around in one form or another since the 1960s, recent technological developments have led to a wide proliferation of this technology. This technology is no longer seen as something out of science fiction movies like Minority Report. With the release of the iPhone X, millions of people now literally have face recognition technology in the palms of their hands, protecting their data and personal information. While mobile phone access control might be the most recognizable way face recognition is being used, it is being employed for a wide range of use cases including preventing crime, protecting events and making air travel more convenient.

5.1 Prevent Retail Crime

Face recognition is currently being used to instantly identify when known shoplifters, organized retail criminals or people with a history of fraud enter retail establishments. Photographs of individuals can be matched against large databases of criminals so that loss prevention and retail security professionals can be instantly notified when a shopper enters a store that prevents a threat. Face recognition systems are already radically reducing retail crime. According to our data, face recognition reduces external shrink by 34% and, more importantly, reduces violent incidents in retail stores by up to 91%.

5.2 Unlock Phones

A variety of phones including the latest iPhone are now using face recognition to unlock phones. This technology is a powerful way to protect personal data and ensure that, if a phone is stolen, sensitive data remains inaccessible by the perpetrator.

5.3 Smarter Advertising

Face recognition has the ability to make advertising more targeted by making educated guesses at people's age and gender. Companies like Tesco are already planning on installing screens at gas stations with face recognition built in. It's only a matter of time before face-recognition becomes an omni-present advertising technology.

5.4 Find Missing Persons

Face recognition can be used to find missing children and victims of human trafficking. As long as missing individuals are added to a database, law enforcement can become alerted as soon as they are recognized by face recognition—be it an airport, retail store or other public space. In fact, 3000 missing children were Discovered In Just Four Days Using Face Recognition In India!

5.5 Help the Blind

Listerine has developed a ground breaking facial recognition app that helps the blind using face recognition. The app recognizes when people are smiling and alerts the blind person with a vibration. This can help them better understand social situations.

5.6 Protect Law Enforcement

Mobile face recognition apps, like the one offered by FaceFirst, are already helping police officers by helping them instantly identify individuals in the field from a safe distance. This can help by giving them contextual data that tells them who they are dealing with and whether they need to proceed with caution. As an example, if a police officer pulls over a wanted murderer at a routine traffic stop, the officer would instantly know that the suspect may be armed and dangerous, and could call for reinforcement.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

Multi-modal Sentiment Analysis problem is a machine learning problem that has been a research interest for recent years. Though lot of work is done till date on sentiment analysis, there are many difficulties to sentiment analyser since Cultural influence, linguistic variation and differing contexts make it highly difficult to derive sentiment. The reason behind this is unstructured nature of natural language. The main challenging aspects exist in use of other modes; dealing with Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. Infusion, we can integrate the results of all these modes; to get more accuracy. Future research could be dedicated to these challenges. So we are moving from uni-modal to multi-modal.

6.2 Future Scope

Multimodal sentiment analysis of social media is a challenging task. This system proposes a multimodal sentiment analysis model based on the attention mechanism. This model can effectively eliminate noise interference in the textual data of social media and obtain more accurate text features. Combined with the attention mechanism, the image features that are more important to sentiment classification are extracted. In terms of feature fusion, the attention mechanism is introduced again to fuse features in different modes effectively, which learns the interactive information between text and images. The model used modal internal information and modal interaction information to effectively obtain the sentimental feature representation of multimodal data, accurately judged the sentimental polarity of users' tweets, better revealed users' real feelings, and helped us understand people's attitudes and views towards certain events on social media. The experimental results on two open datasets demonstrate the feasibility and superiority of our proposed model. In future work, we aim to improve the existing models and methods and study more modalities, including audio, video, and so on. We intend to find more effective methods of feature extraction and feature fusion to provide more effective information for sentiment analysis. The main challenging aspects exist in use of other modes; dealing with Multi Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. In fusion, we can integrate the results of all these modes; to get more accuracy. Future research could be dedicated to these challenges. So we are moving from uni-modal to multi-modal.

References

[1] Efficient Face Recognition System for Operating in Unconstrained Environments Alejandra Sarahi Sanchez-Moreno 1, Jesus Olivares-Mercado 1 , Aldo Hernandez-Suarez 1 , Karina Toscano-Medina 1, Gabriel Sanchez-Perez 1 and Gibran Benitez-Garcia 2,* J. Imaging 2021, 7, 161. https://doi.org/10.3390/jimaging7090161

[2] The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, Tao Mei Accepted for publication in ACM Computing Surveys Computer Vision and Pattern Recognition (cs.CV)

[3] The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances HANG DU, HAILIN SHI, DAN ZENG, XIAO-PING ZHANG, TAO MEI, JD AI arXiv:2009.13290v4 [cs.CV] 27 Dec 2021

[4] (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018 42 | P a g e www.ijacsa.thesai.org Study of Face Recognition Techniques: A Survey Madan Lal, Kamlesh Kumar Department of Computer Science Sindh Madressatul Islam University, Karachi, Sindh, Pakistan Rafaqat Hussain Arain, Abdullah Maitlo, Sadaquat Ali Ruk, Hidayatullah Shaikh

[5] Research on Face Recognition Classification Based on Improved GoogleNet Zhigang Yu, Yunyun Dong, Jihong Cheng, Miaomiao Sun , and Feng Su Hindawi Security and Communication Networks Volume 2022, Article ID 7192306, 6 pages https://doi.org/10.1155/2022/7192306

[6] Boiy, E. Hens, P., Deschacht, K. &Moens, M.F., “Automatic Sentiment Analysis in Online Text", In Proceedings of the Conference on Electronic Publishing(ELPUB-2007).

[7] J. Wiebe, T. Wilson, and C. Cardie. “Annotating expressions of opinions and emotions in language”, Language Resources and Evaluation, 2005.

[8] Scott Brave and Clifford Nass, Emotion in Human- Computer Interaction. Retrieved from http://lrcm.com.umontreal.ca/dufresne /COM7162/EmotionHumanInteraction.pdf

[9] S. V. Bo Pang, Lillian Lee, “Thumbs up? Sentiment classifcation using machine learning techniques”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), ACL, pp. 79– 86,July 2002.

[10] Pravesh Kumar Singh and Mohd Shahid Husain “Methodological study of opinion mining and sentiment analysis techniques” International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014

[11] Gelareh Mohammadi and Alessandro Vinciarelli “Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features”, IEEE transactions on affective computing, vol. 3, no. 3, july-september 2012

[12] Litman, D.J. and Forbes-Riley, K., “Predicting Student Emotions in Computer-Human Tutoring Dialogues”. In Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), July 2004

[13] Lee C M Narayanan, S.S., “Toward detecting emotions in spoken dialogs”. IEEE Tran. Speech and Audio Processing, Vol. 13 NO. 2, March 2005

[14] Mozziconacci, S., “Prosody and Emotions”. Int. Conf. on Speech Prosody. 2002 Danielle Lottridge, Mark Chignell, and Michiaki Yasumura,” Identifying Emotion through Implicit and Explicit Measures: Cultural Differences, Cognitive Load, and Immersion”, IEEE transactions on affective computing, vol. 3, no. 2, april-june 2012

[15] Mohammad Soleymani, Maja Pantic, Thierry Pun,” Multimodal Emotional Recognition in Response to Videos”, IEEE transactions Affective Computing , Vol.3,No.2, April-June 2012

[16] Gaurav Vasmani and Anuradha Bhatia “A Real Time Approach with Big data- A Survey”, International Journal of Engg Sciences & Research Technology, Vol 3, Issue 9, September 2013

[17] LouisPhilippe, Morency Rada Mihalcea and Payal Doshi “Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web”, ICMI’11, November 14–18, 2011, Alicante, Spain.

[18] Andrea Kleinsmith and Nadia Bianchi-Berthouze ”Affective Body Expression Perception And Recognition:A Survey”, IEEE transactions Affective Computing , Vol.4,No.1, January-March 2013

[19] Gérard Dray, Michel Plantie , Ali Harb, Pascal Poncelet Mathieu Roche and François Trousset, “Opinion Mining From Blogs” International Journal of Computer Information Systems and Industrial Management Applications – IJCISIM Vol. 1 2009

[20] Charles B. Ward, Yejin Choi, Steven Skiena, “Empath: A Framework for Evaluating Entity-Level Sentiment Analysis”, IEEE 2011

[21] Georgios Paltoglou and Michael Thelwall, “Seeing Stars of Valence and Arousal in Blog Posts” IEEE transactions on affective computing, vol. 4, no. 1, January-March 2013

[22] Shangfei Wang, Zhilei Liu, Zhaoyu Wang, Guobing Wu, Peijia Shen, Shan He, and Xufa Wang “Analyses of a Multimodal Spontaneous Facial Expression Database” IEEE transactions on affective computing, vol. 4, no. 1, January-March 2013

[23] Ashish Tawari and Mohan Manubhai Trivedi, “Speech Emotion Analysis: Exploring the Role of Context” IEEE transactions on multimedia, vol. 12, no. 6, October 2010

[24] Tal Sobol-Shikler and Peter Robinson, “Classification of Complex Information: Inference of Co-Occurring Affective States from Their Expressions in Speech”, IEEE transactions on pattern analysis and machine intelligence, vol. 32,no.7,July 2010

[25] Felix Weninger, Jarek Krajewski, Anton Batliner, and Bjorn Schuller, “The Voice of Leadership:Models and Performances of Automatic Analysis in Online Speeches”, IEEE transactions on affective computing, vol. 3, no. 4, October-December.2012