“EMOTION RECOGNITION BY FACIAL EXPRESSIONS USING DEEP LEARNING”

In the partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering in Electronics and Communication Engineering

Submitted by

Batch No: 210

NAVYA CHALAMALASETTY (160117735071)

SAKETH REDDY DODDA (160117735107)

TEJA REDDY KOMMIDI (160117735110)

Under the Supervision of

Smt. D. Nagadevi
Assistant Professor
Dept. of ECE

Department of Electronics and Communication Engineering

Chaitanya Bharathi Institute of Technology

Hyderabad – 500075

Year 2020-2021

“EMOTION RECOGNITION BY FACIAL EXPRESSIONS USING DEEP LEARNING”

In the partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering in Electronics and Communication Engineering

Submitted by

Batch No: 210

NAVYA CHALAMALASETTY	**(160117735071)**
SAKETH REDDY DODDA	**(160117735107)**
TEJA REDDY KOMMIDI	**(160117735110)**

Under the Supervision of

Smt. D. Nagadevi
Assistant Professor
Dept. of ECE

Department of Electronics and Communication Engineering

Chaitanya Bharathi Institute of Technology

Hyderabad – 500075

Year 2020-2021

Declaration

We do declare that the project work **"EMOTION RECOGNITION BY FACIAL EXPRESSIONS USING DEEP LEARNING"** submitted in the department of Electronics and Communication Engineering (ECE), Chaitanya Bharathi Institute of Technology, Hyderabad in fulfilment of degree for the award of **Bachelor of Engineering** is a bonafide work done by us, which was carried under the supervision of "Smt. D. Nagadevi".

Also, we declare that the matter embedded in this report has not been submitted by us in full or partial thereof for the award of any degree/diploma of any other institution or University previously.

Ch. Navya

D S Reddy

Station: Hyderabad　　　　　　　　Signature of the candidates

Date:

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY(AUTONOMOUS)

HYDERABAD-500075

C E R T I F I C A T E

This is to certify that the Project work entitled "**EMOTION RECOGNITION BY FACIAL EXPRESSIONS USING DEEP LEARNING"** is a bonafide work carried out by **Navya Chalamalasetty (160117735071), Saketh Reddy Dodda (160117735107), Teja Reddy Kommidi (160117735110)** in partial fulfilment of the requirements for the degree of Bachelor of Engineering in Electronics and Communication Engineering, Osmania University, Hyderabad during the academic year 2019-20. The results embodied in this report have not been submitted to any other University or Institution for the award of any diploma or degree.

D. Nagadevi

Supervisor
Smt. D. Nagadevi
Assistant Professor
Department of ECE

Dr. D. Krishna Reddy
Professor and Head
Department of ECE

Acknowledgements

The satisfaction and euphoria that accompany the successful completion of the tasks would be incomplete without mentioning the people whose constant guidance and encouragement made it possible. We take the pleasure in presenting before you, our project which is a result of studied blend of both research and knowledge.

I am extremely grateful to my supervisor **SMT.D. NAGADEVI,** Assistant Professor, Dept of ECE, CBIT without whose insight and guidance, I would have not been able to make such a progress in this project.

I would like to extend my deep gratitude towards our project coordinators **Dr. P. NARAHARI SHASTRY**, Professor, **SRI. K. SUDERSHAN REDDY,** Assistant Professor, **SMT. A. SATYAVATI**, Assistant Professor, Dept of ECE, CBIT, for constant help and guidance to me, ensuring the project was always aligned with its right course of action. They have been always helpful and generous to us during the project.

I am fortunate to have been presented with such a wonderful opportunity to carry out this project and am greatly thankful to Head of department **PROF. Dr. D. KRISHNA REDDY** for providing excellent infrastructure.

Lastly, we would also like to thank everyone who directly or indirectly helped us in successful completion of the project.

Abstract

Facial expression is a major non-verbal means of expecting intentions in human communication. Facial expression was proved to play an important role in the entire data exchange process. The complication of automatic recognition of facial expressions is a study where research being carried out, and it depends on the improvements in Image Processing and Computer Vision techniques. Such systems have a wide range of interesting applications, from human-computer interaction, to robotics and computer animations to recognize and respond to human non-verbal communication such as emotions. Their aim is to provide robustness and high accuracy, but also to update with the changes occurring in the environment.

Automatic facial expression recognition is an actively emerging research in Emotion Recognition. This project extends the new approach to facial expression recognition task delivering the universal emotions like neutral, happiness, surprise, anger, sadness, fear and disgust. This project is implemented using Convolutional Neural Network (CNN). CNN model of the project is based on Alex Net Architecture. CK+ dataset is used, having thousands of images of different emotions and properties. Our model achieved an accuracy of 86.6%. As a result, the proposed model is presented by comparing with the previous techniques having better accuracy.

Contents

Abstract i
Contents ii
List of Figures v
List of Tables vii
Abbreviations viii

1 Introduction 1

1.1 Introduction 1

1.2 Aim of the thesis 3

1.3 Objectives 3

1.4 Motivation of the thesis 3

1.5 Literature survey 4

1.6 Technical approach 6

1.7 Application of thesis 7

1.8 Organization of the thesis 7

2 Theoretical Background 8

2.1 Affective Computing 8

2.2 Artificial Intelligence 9

2.3 Machine Learning 11

2.3.1 Advantages of Machine Learning 11

2.3.2 Disadvantages of Machine Learning 12

2.4 Deep Learning 13

2.4.1 Working 14

2.4.2 Advantages of Deep Learning over ML 15

2.4.3 Types of Deep Learning networks 16

2.4.4 Architecture of CNN 17

2.5 Image Processing involved in Deep Learning 19

2.5.1 Basic Block Diagram of Image Processing 20

2.6 Chapter Conclusion 21

3 Methodology 22

3.1 The Dataset 22

3.2 Alex Net Neural Network 22

3.2.1 Architectural Design 24

3.3 Implementation 29

3.3.1 Training the model 29

3.3.2 Validation of the model 33

3.3.3 Testing the model 36

3.4 Advantages 38

3.5 Chapter Conclusion 38

4 Results and Analysis 39

4.1 Analysis of the Result in Training Phase 39

4.2 Analysis of the Result in Validation Phase 41

4.3	Analysis of the Result in Testing Phase	44
4.4	Chapter Conclusion	45
5	**Conclusions and future scope**	**46**
5.1	Conclusion	46
5.2	Future Scope	46
	References	**48**

List of Figures

S.No	Description	Page No.
2.1	Overview of AI	14
2.2	Design of a Neural Network	15
2.3	Performance of DL over other algorithms	16
2.4	Architecture of CNN	17
2.5	Block Diagram of an Image Processing system	20
3.1	Part of the Dataset classified by different Emotions	22
3.2	Architecture of Alex Net Neural Network	23
3.3	Architectural Design of Alex Net	24
3.4	Methodology of Training phase	30
3.5	Importing the 'Train' dataset into the code	31
3.6	Pre-processing in Training	31
3.7	Performance metrics in Training	32
3.8	Designing the Layers of Alex Net	32
3.9	Training the network with Alex Net	32
3.10	Methodology of Validation phase	33
3.11	Importing 'Test' dataset in Validation phase	34
3.12	Pre-processing in Validation phase	34
3.13	Importing .mat files that are extracted from Training model	34
3.14	Classification between prediction and actual values	35
3.15	Obtaining accuracy	35
3.16	Plotting the confusion matrix	35
3.17	Methodology of the Testing phase	36
3.18	A Switch-case structure to print the output	37

4.1	Training phase result	39
4.2	Confusion matrix of Training phase	40
4.3	Training accuracy	41
4.4	Confusion matrix of Validation phase	42
4.5	Parameters obtained from confusion matrix	44
4.6	Folder used for testing	44
4.7	Final Output that recognizes the emotion	45

List of Tables

S.No	Description	Page No.
3.1	Design of Convolution Layers in the Architecture	25
3.2	Design of ReLU Layers in the Architecture	26
3.3	Design of Normalization Layers in the Architecture	26
3.4	Design of Pooling Layers in the Architecture	27
3.5	Design of Fully Connected Layers in the Architecture	28
3.6	Number of images in Training dataset	30
3.7	Number of images in Validation dataset	33

Abbreviations

S.No	Abbreviation	Description
1	MATLAB	Matrix Laboratory
2	CK+	Extended Cohn- Kanade
3	RGB	Red-Green-Blue
4	CNN	Convolutional Neural Network
5	R-CNN	Region-based Convolutional Neural Network
6	MLP	Multilayer Perceptron
7	FACS	Facial Action Coding System
8	AU	Action Units
9	FER	Facial Expression Recognition
10	AC	Affective Computing
11	CRM	Customer Relationship Management
12	HRM	Human Resource Management
13	AI	Artificial Intelligence
14	ML	Machine Learning
15	DL	Deep Learning
16	FC	Fully Connected
17	GPU	Graphical Processing Unit
18	ReLU	Rectified Linear Unit
19	SGDM	Stochastic Gradient Descent with Momentum

Spine

BE(ECE)

AY 2020-21

CHAPTER 1

1. INTRODUCTION

1.1 INTRODUCTION

A Facial expression is the visible demonstration of the affective state, intellectual activity, intention, personality of a person and plays an informative role in interpersonal relations. Human facial expressions are classified into 7 basic emotions: happy, sad, surprise, fear, anger, disgust, and neutral. Our facial emotions are depicted through activation of some particular sets of facial muscles. These sometimes be easy or complex, signals in each expression often contain ample amount of information about our state of mind.

Automatic facial expression recognition can be an important constituent of basic human- machine interfaces. Behavioral science and medical practice are primary areas that are being implemented. It has been studied for a long period of time and the progress have been seen in recent decades. Though there is a decent progress in the research, developing such model with utmost accuracy remains to be difficult due to the complexity and variations in facial expressions.

Emotion recognition is a technique utilized in software that allows a program to analyze the sentiments on a human face by using sophisticated image dispensation. From the last decade, firms have been testing with a combination of modern formulas with image processing techniques to appreciate more regarding a video or an image of a particular face that tells us how they are feeling at certain situations. We know that human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. Within no time, a technology will be developed that will be able to read emotions as precisely as our brains do. Expressions and emotions go hand in hand, i.e., special combinations of face muscular actions reflect a particular emotion. For certain emotions it is very hard, and maybe even impossible, to avoid its fitting facial expression. It has a big role in customer decision in many domains including e-commerce, restaurants, movies, interests, and satisfaction with a service or a product.

In research, emotional analysis is regarded as a sort of higher, evolved form of sentiment analysis. Sentiment analysis mainly classifies texts into positive, negative or neutral. Whereas Emotional analysis is a vast and deeper analysis of user’s emotions which tries to inspect the psychological behaviors revealing emotions such as neutral, joy, surprise, anger, sadness, fear and disgust. In order to interpret a person's emotional state in a nonverbal form, we should decode his/her’s facial expression. Most of the times, body languages especially facial expressions, describes more than words about one's state of mind. Face has an important role to play in social communication. This is a window to human personality, emotions and thoughts. According to the survey, verbal part contributes about 7% of the message, vocal - 34% and facial expression about 55%.

Importance of Facial Emotion Recognition:

Facial emotion recognition is significant because of its ability to imitate human coding skills. Facial expressions and other gestures convey nonverbal communication which plays an important role in interpersonal relations. These cues complement speech by helping the listener to interpret the intended meaning of spoken words. It is able to deliver exact and unbiased emotions, because it extracts and analyzes information from an image or video feed.

Human emotion recognition plays an important role in the interpersonal relationship. Many intelligent systems are evolving using this emotion recognition models in order to improve their interactions with humans. For example, in market research, interviews, gaming etc. This is important because the systems can adapt their responses and behavioral patterns according to the emotions of humans and make the interaction more natural.

In this project we intend to categorize the expressions on the basis of seven universal emotions: neutral, joy, surprise, anger, sadness, fear and disgust using Deep Learning which extends the Alex Net Neural Network approach to emotion recognition task. This can detect emotions by learning what each facial expression means and applying that approach to the new information given to it.

1.2 AIM OF THE THESIS

To recognize seven basic emotional states: neutral, joy, surprise, anger, sadness, fear and disgust based on facial expressions using Deep Learning.

1.3 OBJECTIVES OF THE THESIS

1. To perform this model on a bigger dataset i.e., CK+ dataset which is a collection of about 10,000 images.

2. To successfully train the model with properly designing the layers of Alex Net Architecture.

3. To develop a model which perfectly predicts not only RGB images but also the colored images.

4. To obtain the results with greater accuracy compared to the already existed models.

1.4 MOTIVATION OF THE THESIS

People in the society face a lot of difficulties in such various situations like Interviews, Automotive industries etc. Usually in interviews, the recruiter needs to evaluate based on the confidence levels of the interviewee, whether or not this candidate will be able to perform well at a client-facing job. Similarly, it will be able to recognize whether a person is replying genuinely by computing the change in emotions during his responses and relating it with the areas in this technology. So, using our model emotions of the interviewee can be recognized at each point of the interview and can help the recruiter to evaluate his/her performance.

One more such example is in the case of automobiles. Nowadays, car manufacturers around the world are increasingly focusing on making cars more personal and safer for us to drive. We also see that accidents are increasing in day-to-day life because of the drowsiness of the driver. In such situations, Emotion recognition plays a vital role in detecting emotions and alerts the driver when he is feeling drowsy.

1.5 LITERATURE SURVEY

The primary focus was on literature addressing the research on Facial Expression Recognition, Emotion Recognition and also the knowledge on different Neural Networks. Some of them are as follows:

- In order to avoid the complex explicit feature extraction process and the problem of low-level data operation involved in traditional facial expression recognition, Jiaxing Lia, Dexiang Zhanga, Jingjing Zhanga, Jun Zhanga, Teng Lia, Yi Xiaa, Qing Yana, and Lina Xuna proposed a method of Faster R-CNN for facial expression recognition in this paper. They developed a project named 'Facial Expression Recognition with Region-based CNN [1]'. Firstly, the facial image is normalized followed by feature extraction using the trainable convolution kernel. Then, max pooling is performed to decrease the dimensions of the extracted features. Later on, RPNs (Region Proposal Networks) are used to generate high-quality region proposals, which are used by Faster R-CNN for detection. Finally, the SoftMax classifier and regression layer are used to classify the facial expressions and predict the output of the test sample. The dataset is provided by Chinese Linguistic Data Consortium (CLDC), which is composed of multimodal emotional audio and video data. They got the mean average precision value of around 0.82 % and accuracy of 50.3% in detecting the emotions.

- Pawel Tarnowski along with his colleagues developed a project i.e., 'Emotion Recognition for facial expression recognition by MLP Neural Network [2]' using trainable convolutional kernel. They performed this experiment in Microsoft Kinect for 3D face modeling because of its small scanning resolution and also a relatively high rate of image registering i.e., 30 frames per sec. Changes observed in facial expressions that are resulted from the movement of certain muscles have been detected using FACS – facial action coding system in the form of special coefficients called Action units- AU. Kinect tool provides six Action Units (AU) derived from the FAC system. The Action Units may be used to describe emotions either separately or in combinations. AU take values between –1 and +1, and carry information about: AU0 - upper lip raising, AU1- jaw lowering, AU2 - lip stretching, AU3 - lowering eyebrows, AU4 - lip corner depressing, AU5 - outer brow raising. There are 6 action units which describes each emotion either separately or in combinations resulting in an accuracy of 73%.

- 'Facial emotion recognition on a dataset using convolutional neural network [3]' -This paper was proposed by V. Tumen, Ö. F. Söylemez and B. Ergen in the year and their study aims to build a CNN based Facial Expression Recognition System, in order to automatically classify expressions presented in Facial Expression Recognition database. This CNN achieved 57.1% success rate on FER2013 database.
- 'Celebrity Face Recognition using Deep Learning [4]' - This paper experiments a publicly available dataset that consists of 2000 images of celebrity faces. Deep Learning technique is achieving its popularity in computer vision and this paper applies this technique to analyze face recognition problem. This shows the promising results produced by CNN, Alex Net and GoogLe Net despite the differences in gender, face expressions, hair style, features, and background of the images in the dataset. One of the techniques under deep learning is Convolutional Neural Network (CNN). There are also pre-trained CNN models that are Alex Net and GoogLe Net, which produce excellent accuracy results. The experimental results indicate that Alex Net is better than basic CNN and GoogLe Net for face recognition.

1.6 TECHNICAL APPROACH

The technical approach of our project is as follows:

First, we created an account in MathWorks website to access all the features of MATLAB. Then we installed MATLAB, thereby logging into the account we created earlier. We have collected several images and merged into a single bigger dataset. Then we downloaded the deep learning tool box in MATLAB for accessing the neural networks.

DATASET COLLECTION:

- Collected different pre-existing datasets and merged into a bigger dataset from sources like Kaggle, CK+ etc.

MATLAB INSTALLATION:

- Creating an account in MathWorks website and logging into it.
- Installation of MATLAB 2020a.
- Downloading the deep learning tool in the MATLAB for importing neural networks.

The progress of the code is as follows:

TRAINING THE MODEL:

- Import Alex Net neural network from deep learning tool box.
- Load the dataset.
- Create the layers of the architecture of Alex Net.
- Perform Data Preprocessing options on the model.
- Train the model based on the architecture.

TESTING AND VALIDATION:

- Load the dataset.
- Create Validation sets and predict the labels using Alex Net.
- Obtaining some performance metrics and printing them.
- Obtaining the confusion matrix.
- Testing a part of the images from the dataset and predicting the emotion in those images.
- Printing that particular emotions in separate message boxes.

1.7 APPLICATION OF THE THESIS

The main objective of our model is to make a more innovative, easy to use, time saving and more efficient system than the existing models.

By measuring the accuracy, precision and some performance metrics for the performance and displaying all of them in a confusion matrix for each and every emotion, our model helps in detecting the emotion in every image it takes as an input, where the user can get all the information on a message box as an output. This model can be used in various applications according to the user's need. In today's world, this model is being used in applications like Health Care, Automotive Industry, in video games, interviews, lie detectors, and Market research.

1.8 ORGANIZATION OF THESIS

The project report is presented as five chapters:

CHAPTER 1: Gives a brief introduction to our project along with the technical approach and application of thesis.

CHAPTER 2: Gives detail information about architecture of Alex Net neural network and technical content related to the project.

CHAPTER 3: Deals with the methodology and implementation of our project.

CHAPTER 4: Deals with results of our project.

CHAPTER 5: Discusses the conclusions drawn from this project and about the future scope.

CHAPTER 2

2. THEORETICAL BACKGROUND

In this chapter, an explanation of relevant concepts for this project are presented. This chapter focuses to provide a background on the topics to be discussed during the rest of the report. By making use of chronological revision of techniques like affective computing and deep learning, the background of this chapter is achieved. So as to brief the hypothesis, a top-down approach is going to be used. In addition to this, introduction to the related research for this approach is explained as well.

2.1 AFFECTIVE COMPUTING

As described by Rosalind Picard, “Affective computing is the kind of computing that relates to, arises from, or influences emotions or other affective phenomena”. Affective computing aims to include emotions on the design of technologies since they are an essential part of tasks that define the human experience: communication, learning, and decision-making.

One of the main foundations behind affective computing is that without emotions, humans would not properly function as rational decision-making beings. Some researches show that there is no such a thing as “pure reason”. Emotions are involved in decision-making since the suggested approach would affect the whole process by consuming more time, and are not suitable for daily tasks. Researches in this recognition topic shows each probable option was not tested brain, but a quick decision can me made by an emotion.

An emotion is known as a class of face qualities that are internally associated to the nervous system. The nervous system will provide some set of instructions to emulate the particular modulations connected to that class whenever a specific emotional state is activated. So far, the importance of emotion has been discussed without taking human interaction into consideration. Empathy is a human behavior that that aware and gives us the understanding about what other humans are experiencing from their current’s position. Therefore, empathy allow us to conquer close relationships and strong communities. Moreover, it is fundamental towards a pro-social behavior, which includes social interaction and perception. In order to understand better about one's emotional state, it is very crucial for affective computing to create ways to measure these particular modulations properly. This can be achieved in two main ways by recognizing vocal and facial emotions. However, in this project, only facial emotions were used.

In affective computing, a device was introduced for human-computer interaction which has the capability to detect and respond to the user's emotions and other stimuli. Such computing device with this capacity could gather clues to user emotion from a variety of inputs. Facial expressions, posture, gestures, speech, pattern of key strokes and the change in hand temperature on a mouse can all impact the user's emotional state, which in turn are detected and interpreted by the computer.

Affective computing has the potential to humanize digital interactions and offer benefits in an almost limitless range of applications. For example, in an e-learning situation, an AC program could detect when a student is frustrated and offer expanded explanations or additional information. In medicinal field, AC programming can help physicians in understanding a patient's mood or look for emotion changes in their face i.e., sadness or depression. Other business applications currently being explored include customer relationship management (CRM), human resource management (HRM), marketing and entertainment.

2.2 ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) has direct relation to the simulation of human intelligence in machines that are designed to act like humans and mimic their actions. The name can also be applied to any device, that exhibits attributes associated with a person's mind such as learning and problem-solving.

The main idea of artificial intelligence is its ability to rationalize and make the actions that have the best chance of achieving a specific goal. Machine Learning, a constituent of Artificial Intelligence refers to the idea that, without human monitoring, computer programs can learn automatically and adapt to modifications made in data. On the other hand, Deep learning techniques is an automated learning that allows the unstructured data such as text, images, or video by absorption in huge amounts.

When we think of artificial intelligence, the first thing that triggers our mind is robots. Artificial intelligence is based on the principle that a machine can easily mimic human intelligence and execute tasks in a defined way, prioritizing from most simple more complex. One major goal of AI is to mimic human cognitive activity. In this field, we have seen huge developments by researchers and developers in areas such as learning, reasoning, and perception, to the extent where these can be perfectly defined. Whereas few believe that scientists will develop systems very soon that exceed the human capacity which can learn and summarize any subject. As advancements in the technology grows day-to-day, previous benchmarks made in artificial intelligence become outdated. For example, machines that recognize text through optical character recognition and calculate basic functions are no longer considered to manifest artificial intelligence, since this function is now taken for granted as an inherent computer function. AI is continuously evolving to benefit the technologies in different industries. Machines uses a weird non-disciplinary approach based on mathematics, linguistics, psychology, computer science, and more.

AI in the field of Emotion Recognition:

AI has a long way to go before it can perfectly recognize emotions. People are still debating Ekman's classic theory of emotions – which were used as a basis for future artificial intelligence and machine learning projects. The proposed theory, states that there are six universal human emotions and they can be found to express similarly in various cultures. Experts argue that the current AI technology have the training sets that are not diverse enough. So, they can process only exaggerated expressions.

Apart from these issues, there is still merit in using AI to detect facial expressions. The AI will try to recognize your emotion in the face based on several factors such as the location of your eyes, eyebrows, and how your mouth is positioned. These models detect 6 emotions: neutral, happy, sad, surprise, disgust and anger.

2.3 MACHINE LEARNING

Machine Learning (ML) is a subset of Artificial Intelligence. Where, a simple ML is the area of study that allows the computers to learn without being programmed explicitly. This statement provides a powerful insight in the particular approach of this field. It completely differs from other fields where any new feature has to be added by hand. For instance, in software development, a programmer has to update the software whenever a new requirement appears. In ML, this is not exactly the case. The ML algorithms create models, based on input data. These models generate an output with a set of predictions or decisions. Then, when a new requirement appears, the model might be able to handle it by itself without the need of additional code.

ML is broadly divided into 3 categories. Each category details about how the learning process is executed by the system. These categories are: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning comes into picture when a model receives the input as a set of labelled inputs, which means they belong to a particular class. The model tries to adapt itself in a way that can map each and every input with the corresponding output class. On the other hand, unsupervised learning receives a set of inputs that are not labelled. In that sense, the model obtains the output by reading the patterns on them. Finally, reinforcement learning is when an agent is rewarded or punished accordingly to the decisions that it took in order to achieve a goal.

2.3.1 ADVANTAGES OF MACHINE LEARNING:

- **Easily identifies trends and patterns**

  Machine Learning can audit huge amounts of data and innovate specific trends and patterns that would not be apparent to humans. For example, for an e-commerce sites like Amazon, in order to help users, cater the right products, deals, and reminders relevant to them, it is necessary to understand the browsing behaviours and purchase histories of their users. Such that these results are used to reveal relevant advertisements.

- **No human intervention needed (automation)**

  With ML, we don't need to observe the project in every step of the way. That means that we are giving the machines an ability to learn themselves, and letting them make their own predictions and also improving the algorithms on their own. An example for this is anti-virus software, which learns to filter new threats or problems that they recognize. ML is also good at recognizing spam.

- **Continuous Improvement**

  As ML algorithms develop immensely with the experience, they keep improving in accuracy and efficiency, which lets them make better decisions. For example, to develop a weather forecast model. As the amount of data that we are receiving keeps increasing, our algorithms learn to make more accurate predictions faster due to more training of the data.

- **Handling multi-dimensional and multi-variety data**

  Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

- **Wide Applications**

  ML can work for you if you are an e-tailer or a healthcare provider. It does apply, where it has a capability to deliver a personal experience not only to customers but also targeting the right ones.

2.3.2 DISADVANTAGES OF MACHINE LEARNING:

- **Data Acquisition**

  Machine Learning needs to be trained on bigger data sets, and these datasets should be unbiased in nature and are also of high quality. Sometimes you must also wait for the generation of new data.

- **Time and Resources**

  ML takes time to learn and build the algorithms sufficiently to meet its objectives with great accuracy and relevance. It also demands a vast amount of resources in order to function. This might entail further computing power needs for you.

- **Interpretation of Results**

  The capacity to appropriately interpret the output of the algorithms is another important difficulty. The algorithms for your objective must also be carefully selected.

- **High error-susceptibility**

  Machine learning is autonomous and is highly prone to errors. Assume you train a dataset algorithm small enough not to be inclusive. You finish with biased predictions from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of ML, such errors might trigger a sequence of mistakes, which cannot be discovered for extended times. And when they become aware, it takes quite a while to detect the source of the issue, and even longer to correct it.

Due to the above drawbacks of Machine Learning, and also having said that, our primary concern is to train on a larger data set and to get a higher accuracy, we have chosen Deep Learning in our project.

2.4 DEEP LEARNING

Deep learning is an artificial intelligence (AI) function that imitates the functioning of the human brain in data processing and in decision-making models. Deep learning is a subset of artificial intelligence machine learning, which comprises networks that are unable to learn from unstructured or non-labelled input. Also known as deep neural learning or deep neural network. Deep learning happens when decisions are taken without supervision on unstructured data. Emotion recognition, Object recognition, speech recognition, and language translation are some of the tasks accomplished through deep learning.

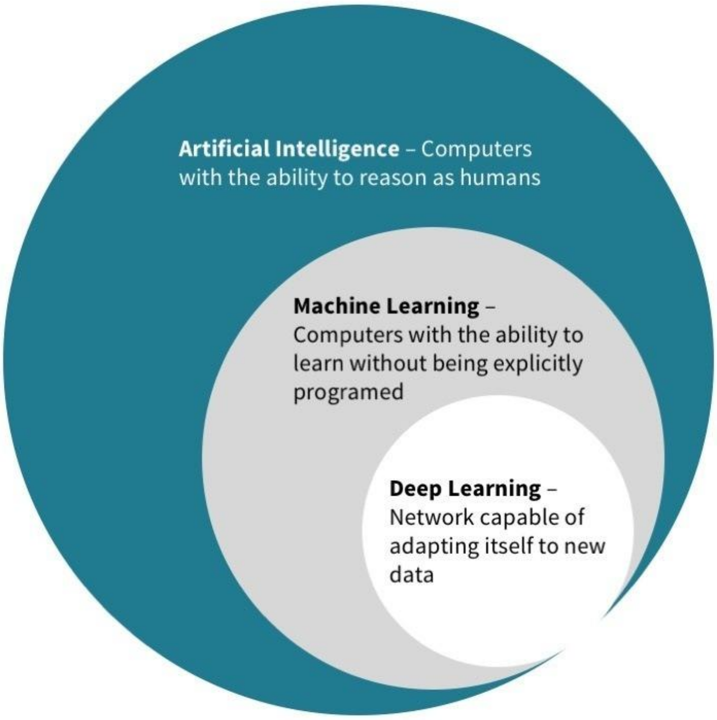

Fig 2.1: Overview of AI

2.4.1 WORKING:

Neural networks are layers of nodes, resembling much as the human brain consisting of neurons. Nodes within individual layers are connected to adjacent layers. Depending on the numbers of layers the network is supposed to be deeper. A single neuron receives thousands of signals from other neurons in the human brain. Signals pass between nodes in an artificial neural network and assign appropriate weights. Among all the nodes, a heavier weighted node will have more impact on the next layer of node. The final layer compiles the weighted inputs to produce an output.

Deep learning systems demand strong hardware since they handle a lot of data to be processed and perform multiple complicated mathematical computations. However, deep learning calculations might take weeks, even with such sophisticated equipment.

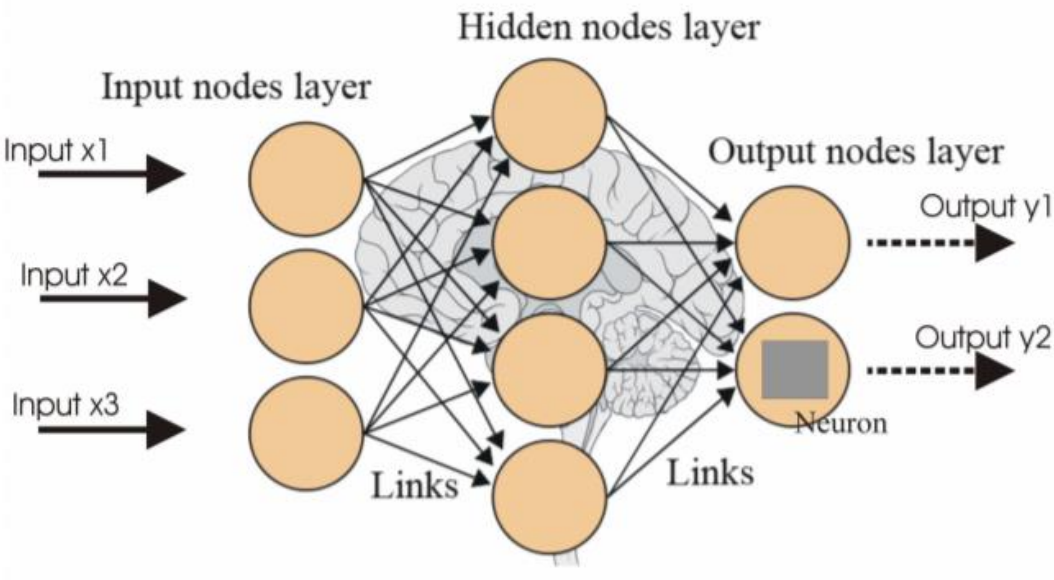

Fig 2.2. Design of a Neural Network

Deep learning algorithms require massive amounts of data in order to deliver accurate results, information is therefore fed into vast data sets. Artificial neural networks can categorize data as responses to a series of true or false binary questions, involving extremely complicated mathematical computations, while processing data. For example, a facial recognition program works by learning to detect and recognize edges and lines of faces, then more significant parts of the faces, and, finally, the overall representations of faces. The algorithm trains itself over time and raises the likelihood of correct answers.

2.4.2 ADVANTAGES OF DEEP LEARNING OVER ML:

Performance is the fundamental difference between the two algorithms. Although deep learning techniques do not work well when the data is small. That is the sole reason why deep learning algorithms require a large amount of data to understand it perfectly.

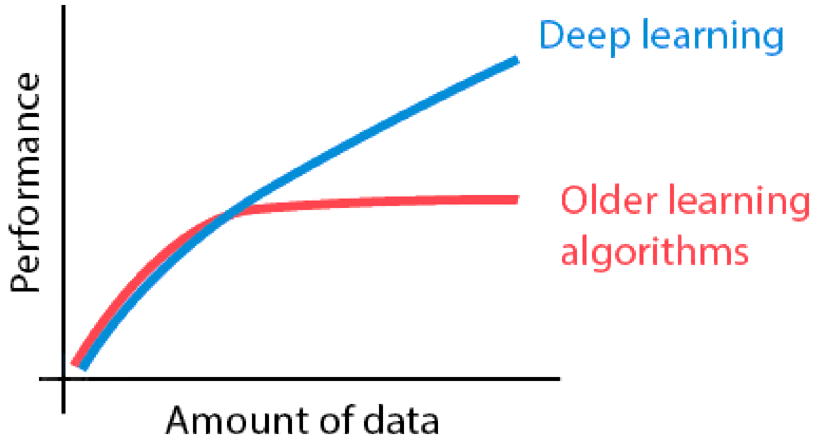

Fig 2.3 Performance of DL over other algorithms

- Deep Learning approaches require high-end training infrastructure to train in reasonable time.
- When there is lack of domain understanding for feature introspection, Deep Learning techniques outshines others as you have to worry less about feature engineering.
- Deep Learning particularly shines when it comes to complicated issues such as image classification, natural language processing, and speech recognition.

2.4.3 TYPES OF DEEP LEARNING NETWORKS:

These are some of the popular deep learning networks:

- Convolution neural network (CNN)
- Feedforward neural network
- Radial basis function neural networks
- Multi-layer perceptron
- Recurrent neural network
- Modular neural network

Among all these, a deep learning technique that is relevant for our project is presented i.e., Convolutional Neural Networks (CNN). However, there are some popular CNN's in deep learning like GoogLeNet, LeNet, AlexNet, ResNet and more. The recent availability of large datasets which consist of hundreds to thousands of images, has driven the demand for a highly effective model of deep learning. Then came AlexNet, which has higher performance rate compared to the rest. So, we implement our project using AlexNet Architecture and detail explanation about this will be explained in the next chapter.

2.4.4 ARCHITECTURE OF CNN:

There are two main parts to a CNN architecture.

- A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.
- A fully connected layer that uses the convolution process output and predicts the image class according to features extracted in earlier stages.

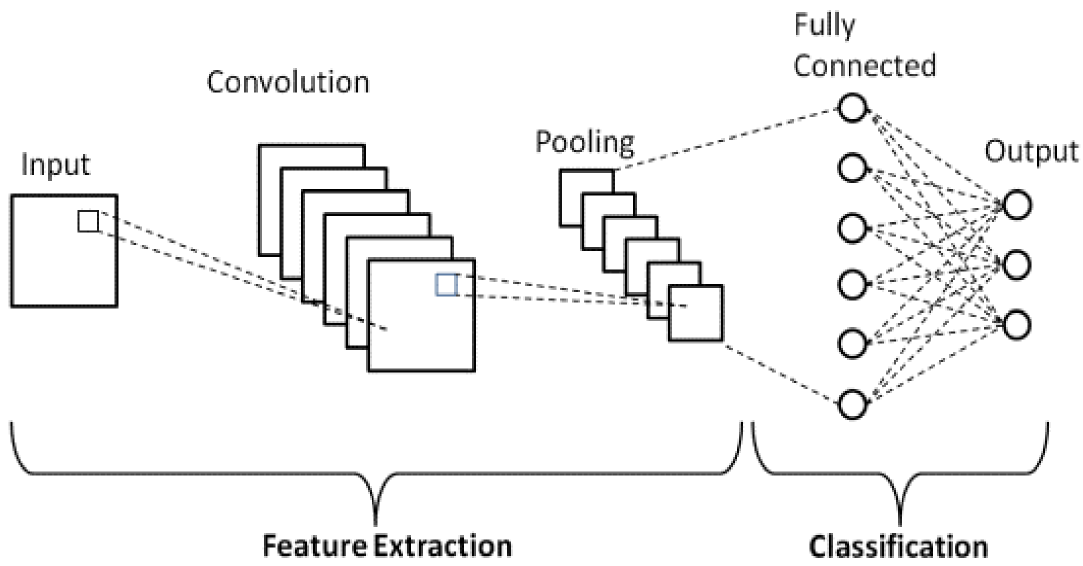

Fig 2.4 Architecture of CNN

Convolution Layers:

The CNN consists of three types of layers, the convolution layers, pooling layers and Fully connected (FC) layers. A CNN architecture will be constructed when these layers are stacked. In addition to these three layers, there are two more important parameters which are the dropout layer and the activation function which are defined below.

1. Convolutional Layer

This layer is the initial layer to extract the different features from the input images. The mathematical operation of convolution between the input image and a filter of a certain MxM size are carried out at this layer. The dot product is obtained between the filter and portion of the input image in relation to the filter size by sliding the filter over the input image (MxM).

The result is termed as the Feature Map that provides us with information of the image, such as corners and edges. This feature map is then fed to additional layers to learn more about the input image.

2. Pooling Layer

In the majority of situations, a pooling layer follows a convolutional layer. The main goal of this layer is to reduce computational costs by reducing the size of a convolved feature map. This is done by decreasing the layer connections and operates on each feature map individually. There are several types of pooling operations, depending on the method used.

In Max Pooling, the largest element is taken from feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. Sum pooling is calculated by total amount of the elements in the predefined section. The pooling layer generally serves as a link between the Convolutional layer and the FC layer.

3. Fully Connected Layer

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers normally lie in front of the output layer and comprise CNN architecture's final layers.

The input image is flattened from the previous layers and feds into the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place.

4. Dropout

Usually, it can produce overfitting on the training data set when all features are linked to the FC layer. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on a new data.

To overcome this problem dropout layer is used to drop few neurons from the neural network during the training phase, which in turn reduces the model size. 30 percent of the nodes are randomly dropped from the neural network when 0.3% of dropout is given.

5. Activation Functions

Finally, the activation function is one of the main parameters of the CNN model. They are used for learning and approximating any type of continuous and complex relationship between network variables. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network.

It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions. Each of these functions have a specific usage.

2.5 IMAGE PROCESSING INVOLVED IN DEEP LEARNING:

Image processing aims to transform an image into digital form and performs some process on it, to get an enhanced image or take some utilized information from it. It is a way of converting the image into digital form and carrying out several operations to obtain certain models or to extract relevant information from them. A video portion or an image, such as photograph, is the input to this method. The output matches the intended section of the picture or its focused version. In general, when using a predetermined signal processing methodology, the image processing system considers images as two-dimensional signals.

Fundamental steps in Digital Image Processing are:

1) Image Acquisition
2) Image Restoration
3) Image Enhancement
4) Pre-processing
5) Compression
6) Segmentation
7) Feature Extraction
8) Morphological processing
9) Representation and Description
10) Object Recognition

Based on our requirements in the project, the above steps are involved in developing the model.

Note: One of the main strengths of using DL techniques is that there is no need for feature extraction. The algorithms are able to learn features by themselves over basic representations.

2.5.1 BASIC BLOCK DIAGRAM OF IMAGE PROCESSING:

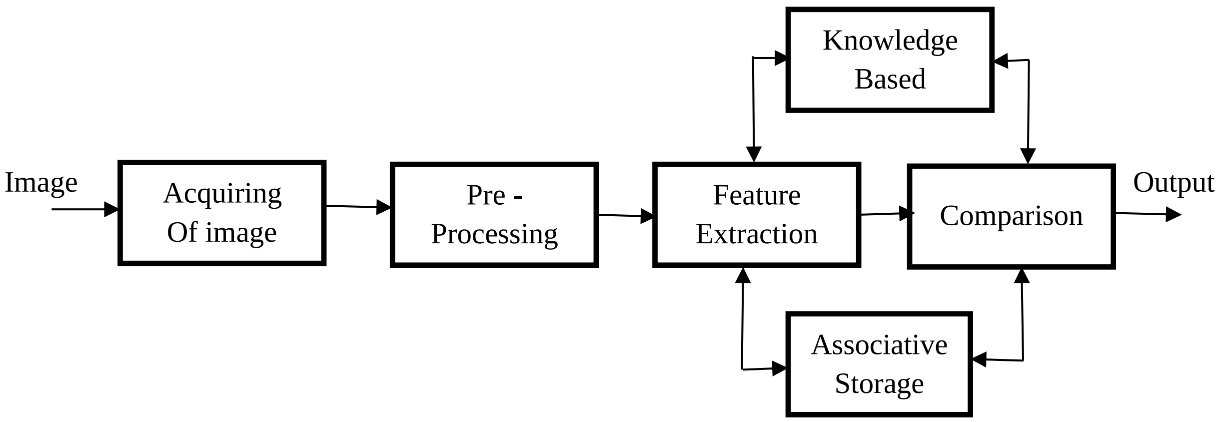

Fig 2.5 Block Diagram of an Image processing system

2.6 CHAPTER CONCLUSION:

In this chapter, we have discussed about the theoretical background of our project regarding Affective Computing, Machine Learning- its advantages and disadvantages, Deep Learning- its preference over ML, and a detail overview of Architecture of CNN, Image Processing techniques used in Deep Learning and its basic block diagram.

CHAPTER 3

3. METHODOLOGY

3.1 The Dataset

CK+ Dataset along with some other images have been merged forming a bigger dataset consisting of 8,746 images having different properties and comprising of 6 universal emotions. This has been used for training whereas 3,333 images of different emotions has been used for testing phase. Both training and testing datasets have different attributes. Different face expressions, views and background are the sample of attributes indicated in this dataset. Below figure shows the sample attributes includes in this dataset. There are different attributes in the datasets; gender, age are some of the examples of the attributes.

Fig 3.1 Part of the dataset classified by different emotions

3.2 Alex Net Neural Network

Alex Net was primarily designed by Alex Krizhevsky. It was published with Ilya Sutskever and Krizhevsky's doctoral advisor Geoffrey Hinton, and is a Convolutional Neural Network or CNN. Alex Net is the name of a convolutional neural network that has had a major impact on machine learning, particularly in the use of deep learning for machine vision. It famously won the 2012 ImageNet LSVRC-2012 competition by a large margin (15.3% VS 26.2% (second place) error rates). The primary result of the original paper was that the depth of the model was absolutely required for its high performance. This was quite expensive computationally but was made feasible due to GPUs or Graphical Processing Units, during training.

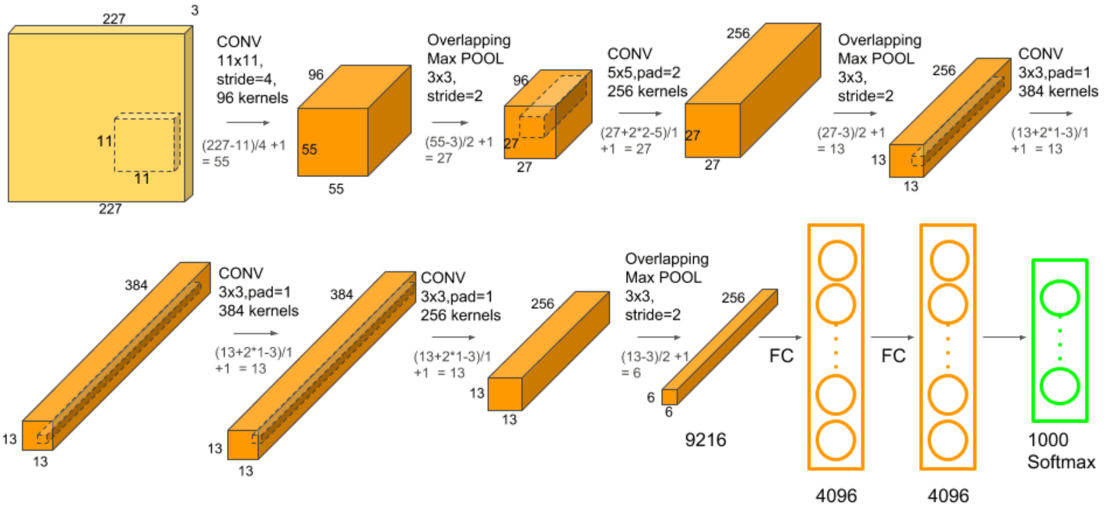

Fig 3.2 Architecture of Alex Net Neural Network

Alex Net Neural Network over CNN:

Drawback: CNNs had always been the go-to model for object recognition - they're strong models that are easy to control and even easier to train. But they experience overfitting when being used on millions of images. Also, they're hard to apply to high resolution images.

Solution: A few years ago, we have utilized tens of thousands of images in smaller data sets, such as CIFAR and NORB. These data sets were enough to master basic recognition tasks for machine learning models. But real life is never easy and has many more variables that are captured in these small datasets. The current availability of enormous datasets like as CK+, consisting of hundreds of thousands to millions of labelled images, has led to the necessity to develop an extremely capable deep learning model. Then came Alex Net. 60 million parameters were present in Alex Net, arose as a big problem in overfitting. Two methods were employed to reduce overfitting i.e., Data Augmentation and Dropout.

3.2.1 ARCHITECTURAL DESIGN

According to our need and performance as a major concern, AlexNet's architecture has been designed. The below figure illustrates the architecture of AlexNet used in our project which shows 25 layers consisting of input layer- 1, convolutional layers-5, ReLU layers -7, Normalization layers- 2, Pooling layers – 3, Dropout layers – 2, Fully- Connected layers- 3 and 1 Softmax layer, output layer -1.

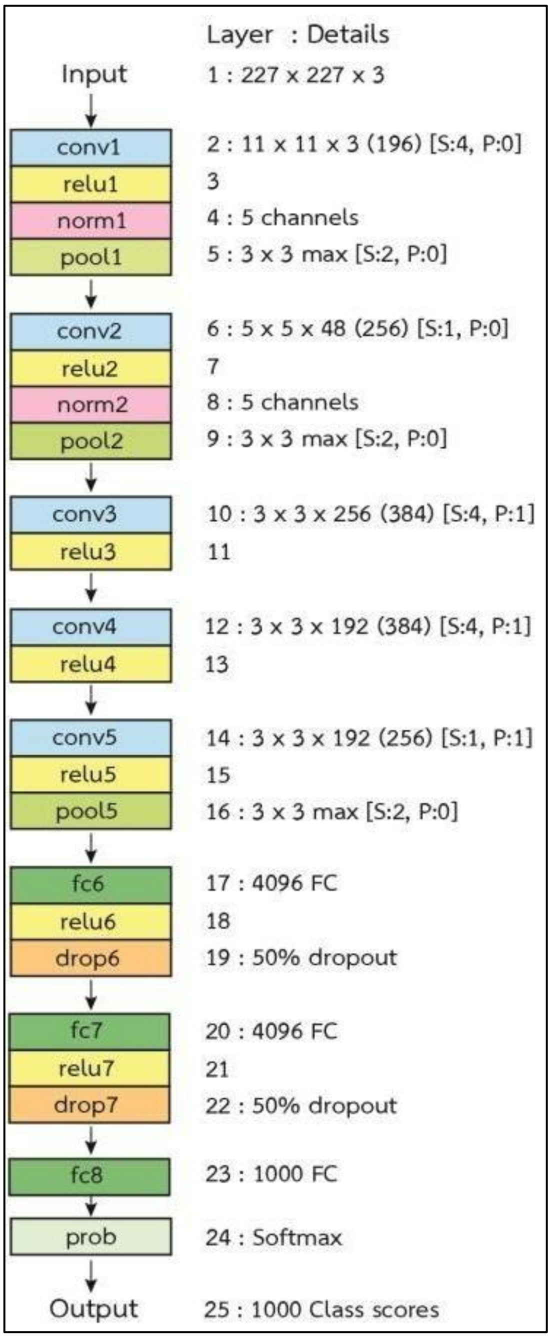

Fig 3.3 Architectural Design of AlexNet

Convolution Layers:

A convolutional layer contains a set of filters whose parameters need to be learned. Filters are smaller in height and weight than input volume. Each filter is convolved with the input volume to compute an activation map made of neurons i.e., the filter slides across the width and height of the input and dot product is calculated at every spatial position between the input and the filter. By stacking the activation maps of all filters along a depth dimension, we get the output volume of the convolutional layer. Since the width and height of each filter is designed to be smaller than the input, each neuron in the activation map is only connected to a small local region of the input volume.

Firstly, we applied convolutional layer with the filter size of 11 x 11, and a number of filters is 96. In this convolutional layer we used a stride of 4, where stride indicates a parameter of the neural network's filter that modifies the amount of movement over the image. So, after this first convolutional layer we will get 55×55×96 volume. Similarly, layers - 2,6,10,12, and 14 are designed with parameters as shown below.

Convolution Layers		
Layer No.	**Size**	**No. of filters**
C2	11 x 11	96
C6	5 x 5	48
C10	3 x 3	256
C12	3 x 3	192
C14	3 x 3	192

Table 3.1 Design of Convolution Layers in the Architecture

ReLU Activation Layer

The activation function in a neural network is used to transform the summed weighted input of the node to the activation of the node or output for that input.

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. For many types of neural networks, it has become the default activation function since the model is simple to train and often achieves better performance. Here, ReLU layer is designed with the size of 1 x 1.

ReLU Layers	
Layer No.	**Size**
R3	1 x 1
R7	1 x 1
R11	1 x 1
R13	1 x 1
R15	1 x 1
R18	1 x 1
R21	1 x 1

Table 3.2 Design of ReLU Layers in the Architecture

Normalization Layer

The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. There are two types of normalization techniques in deep learning i.e., Batch Normalization and Cross- channel Normalization. ReLUs have the desirable property that they do not require input normalisation to prevent them from saturating so the Cross- channel Normalization was applied after the ReLU. Cross- channel Normalization also tries to inhibits the input values because output of ReLU can be very large. The sum runs over the adjacent kernel maps at the same spatial position and can at-most include all the kernel(filter) maps at that position.

We used 2 normalization layers each having 5 channels. It creates a (5, ‘K’, 1) Cross-channel normalization layer for channel-wise normalization with a window size of 5 and K hyperparameter 1.

Normalization Layer	
Layer No.	N4, N8
Window channel	5
Alpha	1.0000 e-04
Beta	0.75
K	1

Table 3.3 Design of Normalization Layers in the Architecture

where **Alpha, Beta, K** are hyperparameters in normalization specified as numeric scalars. The value of Beta must be greater than or equals to 0.01 and the value of K must be greater than or equals to 0.00001.

Pooling Layer

We often use pooling to minimize the dimensions. This reduces the number of parameters, both shortening the training time and combats overfitting. Pooling layers performs down sampling on each feature map independently, thereby reducing height and width, and maintain the depth intact. MAX POOLING is an important type of pooling till date. It basically helps in preserving the most excited neurons information by storing the maximum value in the filter region. This means that by passing on only the important values on to the next layer, over-fitting can be prevented.

We have used pooling windows of size 3×3 with a stride of 2 between the adjacent windows.

Pooling Layer	
Layer No.	**P5, P9**
Pool size	3 x 3
Stride	[2,2]
Padding mode	manual
Padding size	[0,0,0,0]

Table 3.4 Design of Pooling Layers in the Architecture

Where **Pool Size** of [3,3] specifies pooling regions of height 3 and width 3,

Stride of [2,2] specifies a vertical step size of 2 and horizontal step size of 2,

Padding size of [0,0,0,0] indicates as no rows of padding to any of the columns.

Padding mode automatically sets to ‘manual’ because we set the ‘padding’ option to a scalar or vector.

Fully Connected Layer

The operations involved in FC layer are multiplying their inputs by trainable weight vectors, with a trainable bias sometimes summed to those results. The output of these layers was traditionally sent through ReLU activation function, and then on to next layer, similarly to convolution layers.

We designed the last fully-connected layer with 6 classes in which 6 depicts 6 emotions (output size). We designed 3 Fully- Connected layers with these parameters as shown below

Fully Connected Layer	
Layer No.	F17, F20, F23
Input size	4096
Output size	6
Weights	6 x 4096 single
Bias	4096 x 1 single
Weights Initializer	Glorot
Weight Learn Rate factor	20
Weight L2 factor	1

Table 3.5 Design of Fully Connected Layers in the Architecture

where, **glorot** initializes the weights with the Glorot initializer (also known as Xavier initializer). The Glorot initializer independently samples from a uniform distribution with zero mean and variance 2/(InputSize + OutputSize),

Weight Learn Rate Factor means the software multiplies this factor by the global learning rate to determine the learning rate for the weights in this layer. If Weight Learn Rate Factor is 20, then the learning rate for the weights in this layer is twenty times the current global learning rate,

Weight L2 Factor means the software multiplies this factor by the global L2 regularization factor to determine the L2 regularization for the weights in this layer.

Dropout Layer

Dropouts are a regularization method used to avoid overfitting of the model. Dropouts are added to randomly switching some percentage of neurons of the network. When neurons are turned off, their incoming and outgoing connections to those neurons are turned off. This is done to improve the learning of the model. Dropouts are often not utilized after the convolution layers. Instead, they are often used after the network's dense layers. It is always best to merely turn off the neurons to 50%. If we turned off more than 50%, then there is a probability that the model's leaning will be weak and the predictions will be bad.

We designed 2 dropout layers each having 0.5 probability.

SoftMax Layers

The SoftMax function converts a vector of K real values to a vector of K real values that sum to 1. Although the input values can be positive, negative, zero, or higher than one, the SoftMax turns them into values between 0 and 1, allowing them to be recognized as probabilities. Many multi-layer neural networks terminate with a penultimate layer that produces real-valued scores that are not easily scaled and can be challenging to deal with. In this case, the SoftMax comes in handy since it transforms the scores to a normalized probability distribution that may be shown to the user or used as input to other systems. For this reason, it is usual to append a SoftMax function as the final layer of the neural network.

3.3 IMPLEMENTATION

3.3.1 TRAINING THE MODEL

In order to predict an emotion in the input image, we should train the model. In deep learning, the neural network should be trained to improve the performance. For this, the model learns all the characteristics/ features of every image in the dataset iteratively. We should also make sure that the dataset we used for training should be large enough to get a decent accuracy. Because, the more the data trained, the more the performance will be. Below Figure illustrates the methodology of the training phase used in this project:

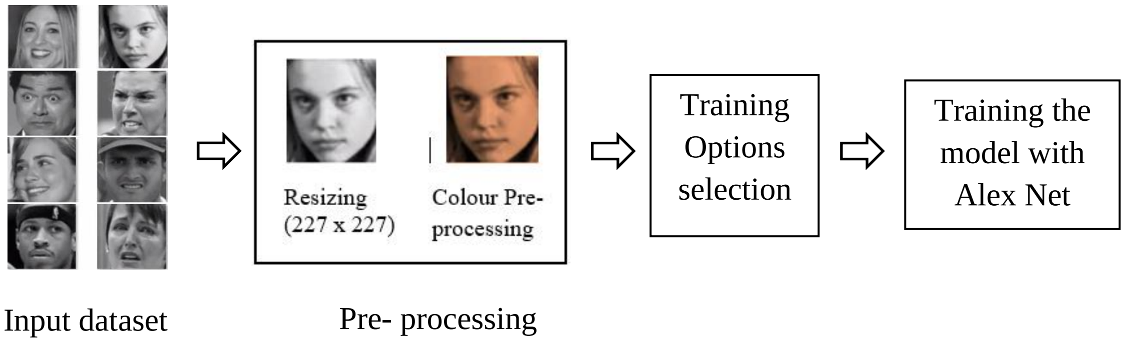

Fig 3.4 Methodology of the Training phase

Input dataset:

The training dataset that we have chosen for our model is large enough to yield statistically meaningful results. It also represents the dataset as a whole. The dataset should be different for training and testing.

- It consists of 8,746 images.
- This constitutes of 6 separate folders with each emotion labelled with their respective emotion.
- These images have different properties and also constitutes of colored images, also with different sizes.

CLASS	TRAINING
anger	154
disgust	55
happiness	3701
neutral	4532
sadness	80
surprise	224
TOTAL	**8746**

Table 3.6 Number of images in Training dataset

Here is the code snippet of the input phase in the training model to import dataset:

```
6 -    imds = imageDatastore(fullfile('Train'),...
7      'IncludeSubfolders',true,'FileExtensions','.jpg','LabelSource','foldernames');
```

Fig 3.5 Importing the 'Train' dataset into the code

As we can see in the image, line no. 6 indicates how the data is imported. 'Train' is the folder name that we have created, fullfile('Train') imports all the data in the folder. Line no.7 is the code that accepts all the files having the extension as .jpg in the 'Train' folder.

Pre- processing:

It is a process of preparing the raw data and making it suitable for machine learning model. While doing any operation on data, it is mandatory to clean it and put it in a formatted way. Here in our project, we perform data-pre-processing like resizing, formatting the image and to suppress the image distortions and helps in enhancing it. Other than this, we also performed color processing technique.

```
36     augimdsTrain = augmentedImageDatastore(inputSize(1:2),imdsTrain,
37     'ColorPreprocessing','gray2rgb');
```

Fig 3.6 Pre-processing in Training

Alex Net accepts only the input images having a size of 227 x 227 x 3, where 3 represents the RGB plane. So, it is mandatory for us to resize the input image. This is done in line no. 36, where inputSize (1:2) takes only 227 x 227 to resize and a built-in-function named 'ColorPreprocessing' with keyword 'gray2rgb' is used to convert greyscale to RGB colored images.

Training Options Selection:

While training a model, it is mandatory to select some training options to increase the performance of the model. We have to tune the options by trial-and-error method.

```
options = trainingOptions('sgdm', ...
    'MiniBatchSize',10, ...
    'MaxEpochs',50, ...
    'InitialLearnRate',1e-4, ...
    'Shuffle','every-epoch', ...
    'ValidationFrequency',50, ...
    'Verbose',false, ...
    'Plots','training-progress');
```

Fig 3.7 Performance metrics in Training

- **SGDM** - Stochastic gradient descent with momentum.
- **MiniBatchSize** – The no. of experienced trajectories in a batch
- **Epochs** - It corresponds to a full pass of the data.
- **Initial Learn Rate** - Time taken for training the data.
- **Shuffle** - Returns a datastore that contains a random ordering of dataset.
- **Validation Frequency** - No. of iterations.
- **Verbose –** Indicator to display training progress information.

Training the model with Alex Net:

Training the model has been done in 2 steps as mentioned below:

- Designing the layers of Alex Net as discussed in chapter 3.2 as shown in Fig. 3.7
- Training the whole model with the above mentioned layers and options as shown in the code below.

```
layers = [
    layersTransfer
    fullyConnectedLayer(numClasses,'WeightLearnRateFactor',20,'BiasLearnRateFactor',20)
    softmaxLayer
    classificationLayer];
```

Fig 3.8 Designing the layers of Alex Net

```
AlexnetTransfer = trainNetwork(augimdsTrain,layers,options);
```

Fig 3.9 Training the network with Alex Net

3.3.2 VALIDATION OF THE MODEL

Successively, the training model is used to predict the responses for the observations in a second dataset called the validation dataset**.** Validation datasets can be used for stopping training when the error on the validation dataset increases, as this is a sign of overfitting to the training dataset.

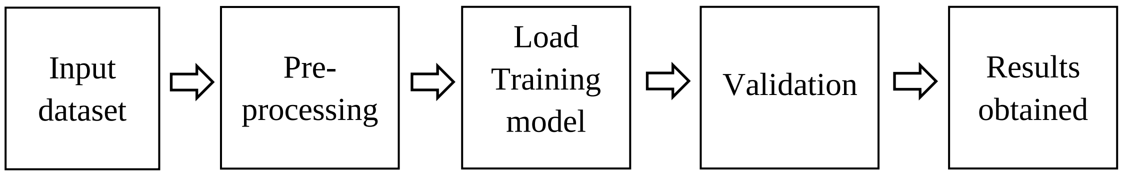

Fig 3.10 Methodology of Validation phase

Input dataset:

The dataset we used in validation and testing are same irrespective of training dataset.

- It consists of 3,333 images.
- This constitutes of 6 separate folders with each emotion labelled with their respective emotion.
- These images have different properties and also constitutes of colored images, also with different sizes.

CLASS	VALIDATION
anger	61
disgust	40
happiness	1233
neutral	1817
sadness	45
surprise	137
TOTAL	**3333**

Table 3.7 Number of images in Validation dataset

Here is the code snippet of the input phase in the validation model to import dataset:

```
imds = imageDatastore(fullfile('Test'),...
'IncludeSubfolders',true,'FileExtensions','.jpg','LabelSource','foldernames');
```

Fig 3.11 Importing 'Test' dataset in Validation phase

Syntax is same as explained in training model, except the fact that the folder that is imported is different i.e., 'Test'.

Pre-processing:

Pre-processing is also same as discussed earlier in the training phase.

```
imdsValidation = imds;
augimdsValidation = augmentedImageDatastore(inputSize(1:2),imdsValidation,'ColorPreprocessing','gray2rgb');
```

Fig 3.12 Pre-processing in Validation phase

Load Training model:

After training the model, we get two files namely, AlexnetTransfer.mat, inputSize.mat.

AlexnetTransfer.mat contains data regarding architectural design that we have designed earlier. Whereas, inputSize.mat contains only one cell regarding the size of the image that our network accepts.

```
load AlexnetTransfer.mat
load inputSize.mat
```

Fig 3.13 Importing .mat files that are extracted from training model

Validation:

This code represents the classification of AlexnetTransfer and the output obtained from the pre-processing section done on test data where the output is stored in the 'YPred' variable as shown.

```
[YPred,scores] = classify(AlexnetTransfer,augimdsValidation);
YValidation = imdsValidation.Labels;
```

Fig 3.14 Classification between prediction and actual values

Validation is done is such a way that the model predicts the score by comparing the actual data (training data) with the test data and gives it a score as 0's and 1's.

1 represents the correct prediction and 0 represents the wrong prediction.

Likewise, prediction is made on a whole dataset and accuracy is calculated based on the formula below.

Mean = Sum of the predicted scores / no. of predictions

The code snippet is as shown below:

```
accuracy = mean(YPred == YValidation);
fprintf('Accuracy of the Alexnet is: %f', accuracy*100);
```

Fig 3.15 Obtaining accuracy

Result Generation:

Result is obtained as a confusion matrix which gives a detailed view of the percentage of emotions that have predicted correctly.

```
cfmatrix2(YValidation,YPred);
% Confusion matrix
plotconfusion(YValidation,YPred);
 title('Alexnet');
```

Fig 3.16 Plotting the confusion matrix

This code plots the confusion matrix between YValidation and YPred with a title named 'Alexnet'.

3.3.3 TESTING THE MODEL

A test dataset is a dataset that is independent of the training dataset. If a model fits to the training dataset also fits the test dataset well with a minimal amount of overfitting. The test dataset is typically used to assess the final model that is selected during the validation process. A test set is therefore used only to assess the performance of a fully specified classifier.

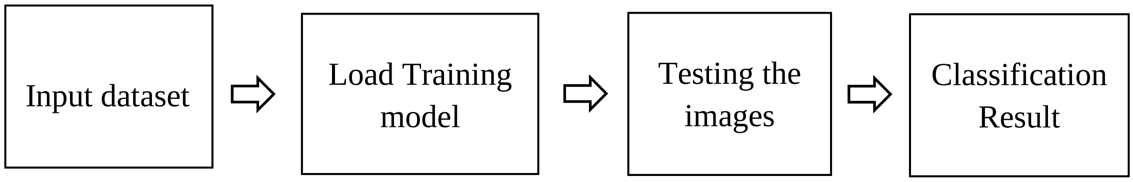

Fig 3.17 Methodology of the Testing phase

Input dataset, Loading the trained model is done same as in the above phases.

Testing the images:

Classification is done on the images that we have chosen for testing based on the Alexnet and the pre-processing techniques as shown in the line 15 of the code below. In line 16, 'pred' variable stores the output class (label name) of the tested images. In order to print the 'pred' variable, we used switch case structure.

```
predictedLabels = classify(AlexnetTransfer,imValidation)';
pred = char(predictedLabels);
switch pred
    case "anger"
        str = "Anger";
        figure,imshow(im);title(str);
        msgbox("Anger");
    case "disgust"
        str = "Disgust";
        figure,imshow(im);title(str);
        msgbox("Disgust");
    case "happiness"
        str = "Happiness";
        figure,imshow(im);title(str);
        msgbox("Happiness");
    case "neutral"
        str = "Neutral";
        figure,imshow(im);title(str);
        msgbox("Neutral")
    case "sadness"
        str = "Sadness";
        figure,imshow(im);title(str);
        msgbox("Sadness");
    case "surprise"
        str = "Surprise";
        figure,imshow(im);title(str);
        msgbox("Surprise");
    otherwise
        print('Incorrect image provided');
end
```

Fig 3.18 A switch-case structure to print the output

The above figure shows the cumulative code of testing and classification result of the model.

Classification Result:

Final result will be displayed in a Figure window showing the image titled with its emotion and also a message box popping up with the emotion name printed in it.

3.4 ADVANTAGES

1) It helps employees and HR team of any company to manage stress levels by recognizing the positive and negative moods of the employees and customers which help businesses to grow.

2) The technology does not require any additional expensive hardware to adopt. This recognition software will help in accomplishing the task of emotion sensing.

3) Helps companies to establish deep emotional connections with their consumers through virtual assistant devices in monitoring state of mind of the users in terms of mental and other health parameters.

4) This technology help children and elderly people by providing timely medical care and assistance by alerting to their caregivers or other family members.

5) Analysis of comments on social media is helpful for the country and the world.

3.5 CHAPTER CONCLUSION

A detailed description of the dataset used, Architectural design of each layer of the Alexnet Neural Network. The chapter also highlights the implementation of the project and briefs each phase in designing the model i.e., training, validation and testing.

CHAPTER 4

4. RESULTS AND ANALYSIS

Emotion Recognition by Facial Expression using Deep Learning has been implemented to detect an emotion of a person's face. Matlab R2020a is used as a tool to train and test the dataset. Our model accepts the input image with scale of the size and color defined as 227 x 227 x 3 which means that the size of the image is 100 x 100 pixels and the value 3 indicates that the training image is a color image. This chapter presents a detailed analysis of the system from the perspective of its performance and robustness.

4.1 ANALYSIS OF THE RESULT IN TRAINING PHASE

In training phase as discussed earlier in the chapter 3.3.1, by importing the Alex Net from math works website followed by the pre-processing phase, we designed the layers of Alex Net and applied the training options for the neural network. Then we started training our model on the training dataset. The results of this phase are shown below:

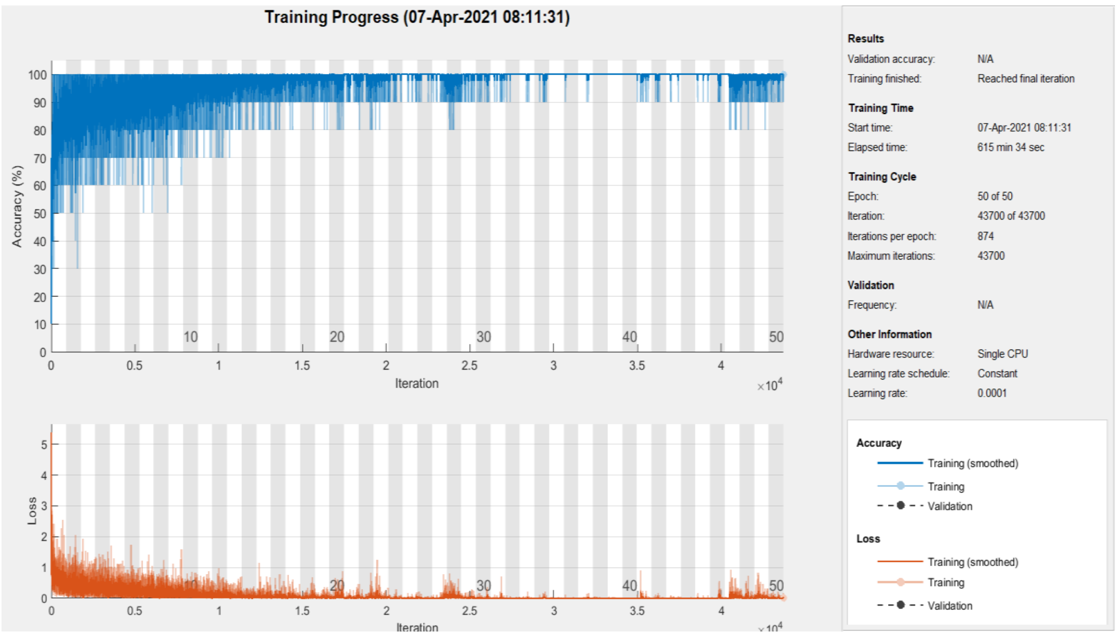

Fig 4.1 Training phase result

It took us around 10 hrs (615 minutes ,34 seconds) to train the whole dataset on Alex Net Neural network. In the above figure, it shows the graphical representation between the accuracy in percentage and loss of the trained data. Here, accuracy gives us the percentage of which the data has been trained and similarly the loss is the percentage of the data that has been failed to get trained. We obtained 100 percent accuracy in training the model. Finally, we also obtained a confusion matrix that describes the performance of the training phase as shown below.

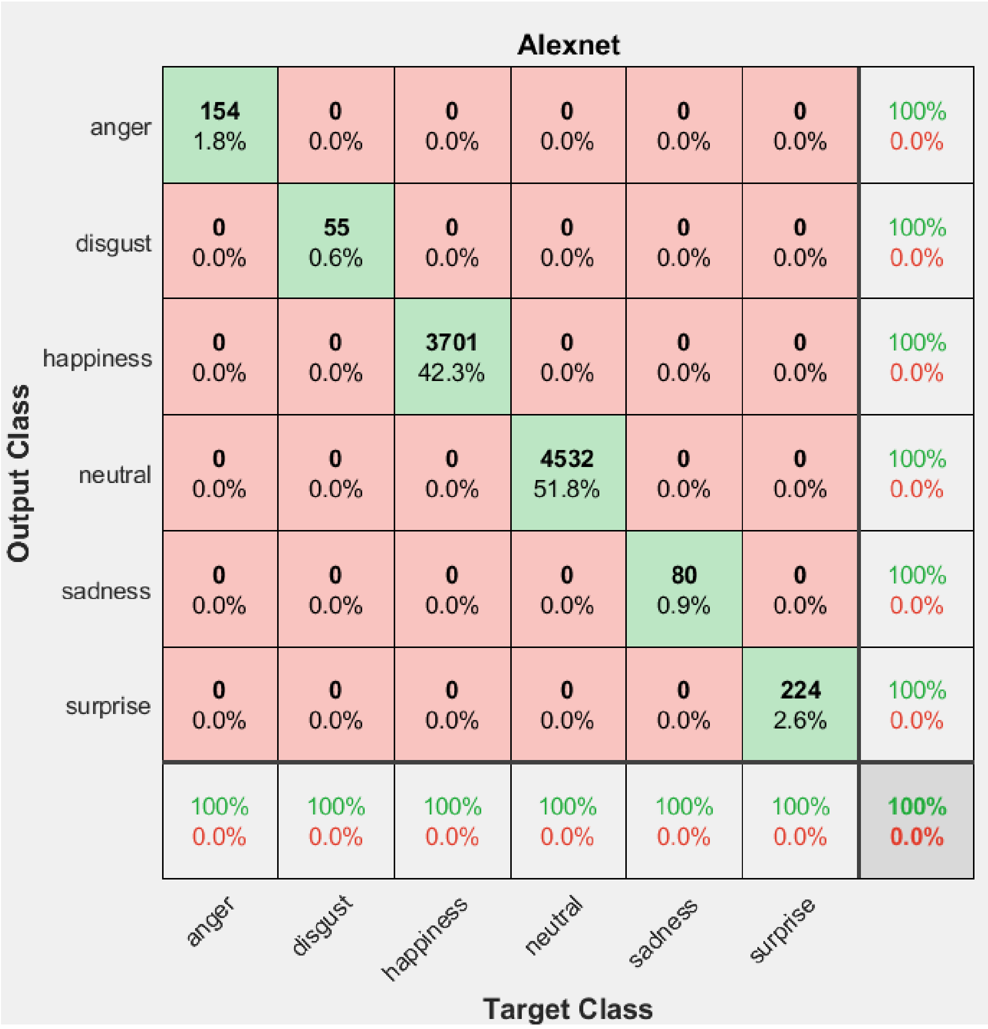

Fig 4.2 Confusion Matrix of Training phase

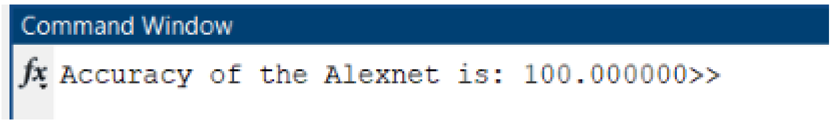

Fig 4.3 Training accuracy

Training time gives us the start time and time elapsed for training. In training cycle, we have few parameters:

- **Epochs** – It is the full pass of the training algorithm over the entire training set, we set it to 50.
- **Iteration per epoch** - It set to 874 automatically based on the size of the training dataset.
- **Iteration** – The total number of cycles elapsed during the training. It will be obtained by multiplying Epochs and Iteration per epoch. So, the total iterations would be 43700 (50 x 874).
- **Learning rate** - If the learning rate is too low, then training takes a long time. If the learning rate is too high, then training can reach a suboptimal result. So, we set it to 0.0001.

4.2 ANALYSIS OF THE RESULT IN VALIDATION PHASE

After completion of the training phase, we obtained AlexNettransfer.mat and inputSize.mat files from the training outputs and imported them into the validation code that we designed. These files contain the information about the Alex Net neural network. The more we train the data, more the prediction will be. After running the validation code, we obtained these results: A confusion matrix, accuracy for predictions, and some metric parameters that describes about the performance of validation phase.

The confusion matrix shows a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. It shows us how well the prediction was done on the test dataset. In this matrix, vertical data in each column gives the information about the target class (Top to Bottom) i.e., Test data and horizontal data in each row gives the information about the output class (left to Right) i.e., Predicted data. The obtained Confusion matrix is shown below.

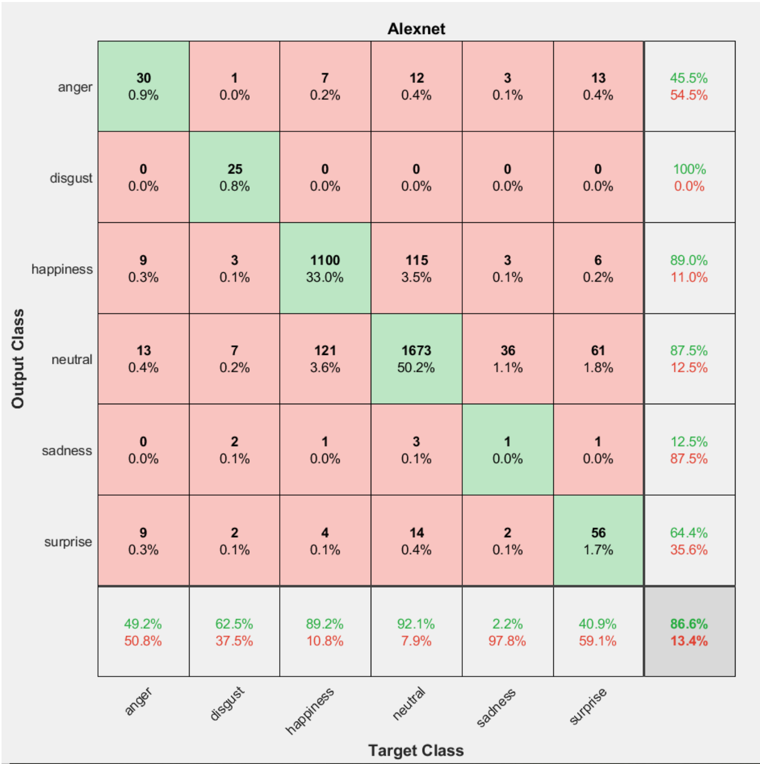

Fig 4.4 Confusion matrix of Validation phase

Some metric parameters will be obtained from the confusion matrix, those are:

- **tp (True positive)** – The number of images that correctly predicted in each emotion class in each class. These are shown as green boxes in the confusion matrix.
- **fp (False positive)** – The number of images that are incorrectly predicted regarding each emotion in each class. These are shown from left to right in each emotion class apart from the green boxes in the confusion matrix.
- **fn (False negative)** – The images which are actually a particular emotion class but are predicted as other emotions in each class. These are shown from top to bottom in each emotion class apart from the green boxes in the confusion matrix.

- **tn (True negative)** – The images which does not belong to a particular emotion and are correctly predicted when one emotion class is taken as reference. These can be depicted from the Subtraction of (tp+fp+tn) from the total no of test images.

- **Precision –** Precision is calculated as the number of correct positive predictions divided by the total number of positive predictions.

$$\mathrm{PREC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

- **Sensitivity –** Sensitivity is calculated as the number of correctly predicted divided by the total number of test images.

$$\mathrm{SN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = \frac{\mathrm{TP}}{\mathrm{P}}$$

- **Specificity –** Specificity is calculated as the number of true negative (tn) predictions divided by the total number of test images.

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} = \frac{\mathrm{TN}}{\mathrm{N}}$$

- **Accuracy -** Accuracy is calculated as the total number of correctly predictions divided by the total number of test images.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}}$$

All these parametric values obtained from the confusion matrix are shown in below figure:

```
    TP |   30.00 |   25.00 |   1100.00 |   1673.00 |   1.00 |   56.00
    FP |   36.00 |   0.00 |   136.00 |   238.00 |   7.00 |   31.00
    FN |   31.00 |   15.00 |   133.00 |   144.00 |   44.00 |   81.00
    TN |   3236.00 |   3293.00 |   1964.00 |   1278.00 |   3281.00 |   3165.00
Preci. |   0.45 |   1.00 |   0.89 |   0.88 |   0.13 |   0.64
Sensi. |   0.49 |   0.63 |   0.89 |   0.92 |   0.02 |   0.41
Speci. |   0.99 |   1.00 |   0.94 |   0.84 |   1.00 |   0.99
------------------------------------------
Model Accuracy is 0.87
------------------------------------------
```

Fig 4.5 Parameters obtained from confusion matrix

Finally, we got an overall validation accuracy of **86.55%** that is almost equal to **87%**, that means that the model has almost predicted correctly.

4.3 ANALYSIS OF THE RESULT IN TESTING PHASE

Finally, in the testing phase a folder with images which has to be tested is given as input to the testing code. We can add any number images to that folder to get tested and to predict the emotion in those particular images. As discussed earlier in the chapter 3.3.3, we used a ‘for loop’ in the code to show the predicted emotion of every image in the test folder. We printed a window with image in it and titled with the particular emotion and also a message box popping up with the emotion name printed on it.

The folder which we used for testing is shown below with 10(any number) test images:

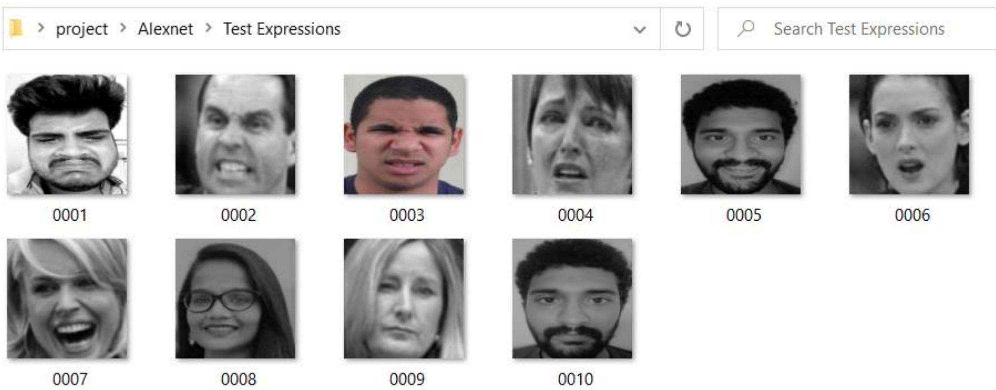

Fig 4.6 Folder used for testing

The output of this phase will be shown in the below figure:

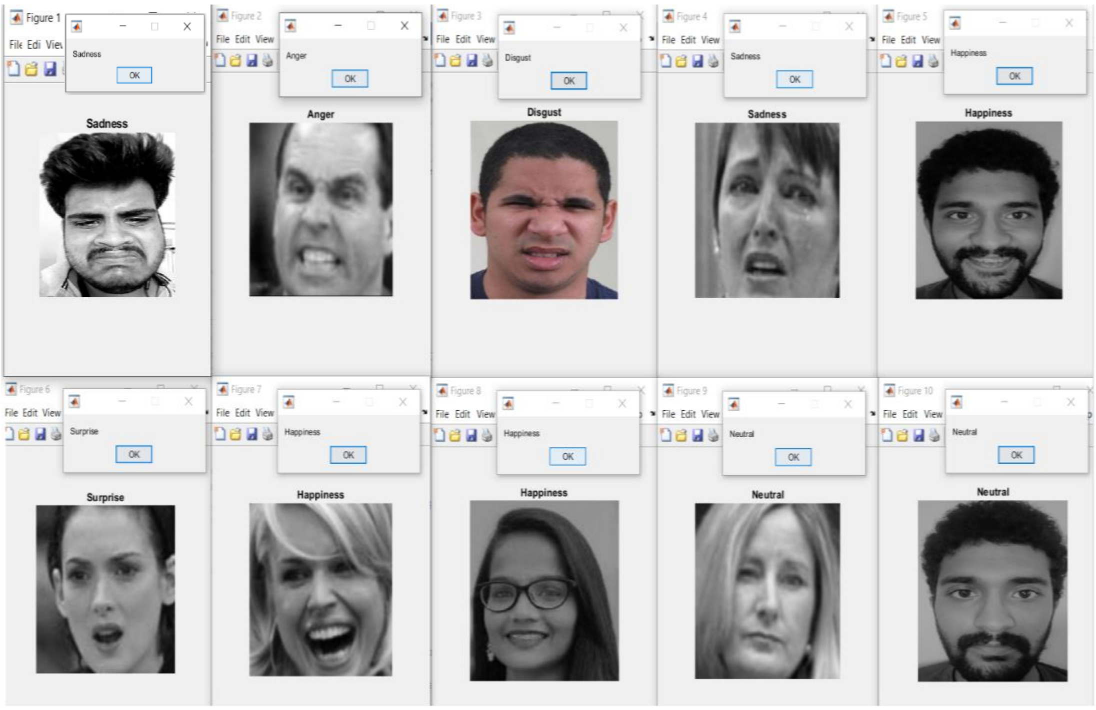

Fig 4.7 Final Output that recognizes the emotion

4.4 CHAPTER CONCLUSION

All the results obtained from every phase i.e., training, validation, testing, are explained in detail. Detailed analysis of the training output with the graphical representation is shown. Total accuracy along with other metric parameters like precision, sensitivity, specificity, and the confusion matrix are explained briefly. Finally, we illustrated how testing of images from a specified folder is done, and also seen the final output of the project where prediction of emotion in each image is done by printing all those images with a message box popping out and also as a title to the figure.

CHAPTER 5

5. CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

In this project, a research to classify facial emotions using deep learning techniques was developed. This is a complex problem that has already been approached several times with different techniques that have been achieved using feature engineering, this project focused on Deep Neural Network i.e., AlexNet Neural Network. This report demonstrates the achievements of the project, but also presents an assessment of the performance and reliability.

Overall, the proposed solution has delivered a system capable of classifying the six basic universal emotions, with an average accuracy of 86.55%. This is due to the training of thousands of images. It extensively makes use of Image Processing and Deep Learning techniques, to evaluate still images and derive suitable features, such that, it would be able to recognize the expressed emotion.

5.2 FUTURE SCOPE

- Deepfakes are images or videos that are synthetically manipulated or generated using deep learning. With some advancement in our project, we can use it for deepfake detection in order to decrease cyber-threats in our society.
- It has higher scope for improvement in the field of automotive industry in detecting subtle micro-expressions. The Emotion AI applications in automobiles are limitless: from detecting drowsiness for driver safety, to determine if a driver is focused or distracted while driving and to create a highly personalized and intimate in-cab experience. These features will not only improve road safety, but will also improve people's driving experiences in future generations of family vehicles and public transportation.

- With this project, we will be able to assist a person with autism in better managing their emotions. There is potential to build an app to test youngsters for autism by putting the subject's facial reactions to a video through a behavioural coding algorithm in order to detect the nature of their responses in contrast to a person who is not autistic.

- Facial emotion recognition AI can automatically detect facial expressions on user's faces and automate the video analysis completely. Market Research companies can use this technology to scale the data collection efforts and rely on technology to do consumer analysis quickly.

- Applications in surveillance and security. For instance, computer models obtained up to 71% correct classification of innocent or guilty participants based on the macro features extracted from the video camera footage.

REFERENCES

[1] Jiaxing Lia, Dexiang Zhanga, Jingjing Zhanga, Jun Zhanga, Teng Lia, Yi Xiaa, Qing Yana and Lina Xuna, “Facial expression Recognition with Faster R-CNN”, *Procedia Computer Science 107,* (2017) 135 – 140.

[2] Pawel Tarnowski, Marcin Kolodziej, Andrzej Majkowski, Remigiusz J. Rak, “Emotion recognition by facial expressions using MLP neural network, k-NN classifier and FACS”, *Procedia Computer Science 108C,* (2017) 1175-1184.

[3] V. Tümen, Ö. F. Söylemez and B. Ergen, “Facial emotion recognition on a dataset using convolutional neural network”, *International Artificial Intelligence and Data Processing Symposium (IDAP),* Malatya, 2017, IEEE Explore.

[4] Nur Ateqah Binti Mat Kasim, Nur Hidayah Binti Abd Rahman, Zaidah Ibrahim, Nur Nabilah Abu Mangshor, “Celebrity Face Recognition using Deep Learning” *Indonesian Journal of Electrical Engineering and Computer Science (2018),* Vol. 12, No. 2, pp.476~481.

[5] Illiana Azizan, Fatimah Khalid, “Facial Emotion Recognition: A Brief Review”, *International Conference on Sustainable Engineering, Technology and Management (ICSETM),* Malaysia, 2018.

[6] M. Rahman, “A comparative study on face recognition techniques and neural network”, vol. 3, no. 5, pp.155-160,2012.

[7] Rafael C. Gonzalez, Richard E. Woods. (2018). *Digital Image Processing,* 4th Edition. Pearson.

[8] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," *2020 International Conference for Emerging Technology (INCET),* 2020, pp. 1-5.

[9] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," *2018 14th Symposium on Neural Networks and Applications (NEUREL),* 2018, pp. 1-4