

Multi-modal Emotion Recognition for Determining Employee Satisfaction

by

Saadat Hussain
15301042

Farhan Uz Zaman
16101300

Maisha Tasnia Zaman
15201019

Nahian Kabir
16101016

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science
Department of Computer Science and Engineering BRAC University December
2019 c 2019. Brac University All rights reserved.

Department of Computer Science and Engineering
BRAC University
April 2020

© 2020. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



Saadat Hussain
15301042



Maisha Tasnia Zaman
15201019



Farhan Uz Zaman
16101300



Nahian Kabir
16101016

Approval

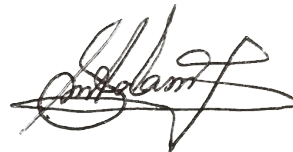
The thesis/project titled “Multi-modal Emotion Recognition for Determining Employee Satisfaction” submitted by

1. Saadat Hussain (15301042)
2. Farhan Uz Zaman (16101300)
3. Maisha Tasnia Zaman (15201019)
4. Nahian Kabir (16101016)

Of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc.in Computer Science on April05, 2020.

Examining Committee:

Supervisor:
(Member)



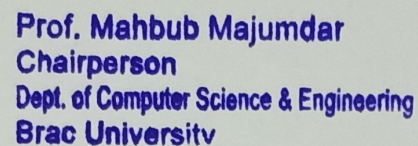
Md. Ashraful Alam, PhD
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)



Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)



Prof. Mahbub Majumdar
Chairperson
Dept. of Computer Science & Engineering
Brac University

Mahbubul Alam Majumdar, PhD
Professor and Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

Emotion Detection has been very popular in the field of research for a couple of years. In the past, emotion recognition has been studied and applied in order to detect the overall emotional state of a person using individual modalities such as facial recognition from images. However, in order to ensure the authenticity of the real time emotional state detected from the data that is received, it is required to use multiple modes. In our research, we have classified emotional states into six specific entities which are: Happiness, Sadness, Neutral, Disgust, Anger and Surprise. The real time emotional state of the candidate is classified into one of these entities according to the candidate's response. We have used two important modes to detect the real time emotion of the candidate, Emotion recognition from facial expression as well as emotion recognition from sentimental analysis. For our research, we have mainly used the Convolutional Neural Network(CNN). We have trained and tested both the facial recognition and sentimental analysis datasets with all the six entities. Therefore, results can be obtained from both the modes in order to justify the candidate's emotional state in real time. The two separate results from the individual independent artificial neural networks are then fed into a machine learning algorithm called Support Vector Machine (SVM) so that the final emotional state can be achieved. Our goal is to apply this multimodal emotion detection technique on employees in various offices and workplaces by asking them questions regarding their work so that their genuine emotions can be obtained from their answers. This is important because in this way, employee satisfaction in workplaces can be recognized which is vital for mental health as well as productivity. In fact, the mental health of the employees not only affects their individual well-being but it affects the overall productivity and environment of the workplace. In order to improve certain aspects of a workplace for better performance along with employee satisfaction and productivity, determining the emotional condition of each employee is vital.

Keywords: Convolutional Neural Network, (CNN), Facial Recognition, Sentimental Analysis, Multimodal Emotion Detection, Feedforward Neural Network

Dedication

We would like to dedicate this thesis to our loving parents.

Acknowledgement

We want to dedicate our acknowledgement of gratitude to our thesis supervisor Md. Ashrafal Alam, PhD Assistant Professor, Department of Computer Science and Engineering of BRAC University for his guidance for the completion of our thesis. We are thankful to CSE department, BRAC University for providing us the necessary equipment for the completion of this project

Table of Contents

Declaration	i
Approval	ii
Abstract	iii
Dedication	iv
Acknowledgment	v
Table of Contents	vi
List of Figures	viii
List of Tables	x
Nomenclature	xi
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	2
2 Literature Review and Related works	3
2.1 Previous Research	3
3 Algorithms	5
3.1 Haar Cascade	5
3.1.1 Selecting the features	6
3.1.2 Integral Image	8
3.1.3 Cascade Classifier	8
3.2 Convolutional Neural Network	9
3.2.1 Convolutional Neural Network Layers	11
3.3 Long Short Term Memory	15
3.4 Gated Recurrent Unit	16
3.5 Extreme Gradient Boosting	17
3.6 Support Vector Machine	19
4 Methodology	20
4.1 Experimental Setup	20
4.1.1 Data Processing	21

4.2	Questionnaire Development	23
4.3	Creating the dataset for the Final Model	25
4.4	Emotion Detection from Sentimental Analysis	25
4.5	CNN for Facial Expression Classification	33
4.5.1	CNN	33
4.5.2	Pre-Processing	33
4.5.3	Training Object Classifier	34
5	Results	35
5.1	Result of Neural Network	35
5.2	Result of XG-Boost	37
5.3	Result of Support Vector Machine	37
6	Conclusion	39
6.1	Conclusion	39
6.2	Future Work	40
6.3	Conclusion	40
	Bibliography	44

List of Figures

3.1	Example of an Haar-Cascade	6
3.2	There are more than thousands of haar features	6
3.3	The AdaBoost algorithm for classifier learning [17]	7
3.4	How the features are selected	7
3.5	The value of the integral image at position 1 is the sum of pixel values in A. The value of the integral image at position 2 A+B, at position 3 is A+C and at position 4 is A+B+C+D	8
3.6	From left to right, a normal image to an integral image	8
3.7	Cascade Classifier	9
3.8	How a positron works and its equations	9
3.9	Architecture of LeNet-5 [21]	10
3.10	Gradient descent illustrated[25]	10
3.11	CNN Layers[26]	11
3.12	Values of each pixel of an image and how it is taken in by an input layer [26]	12
3.13	The convolution operation[28]	12
3.14	The max pooling operation [29]	13
3.15	The average-pooling operation [30]	13
3.16	The ReLu function[31]	14
3.17	The ReLu Operation on the ReLu layer[31]	14
3.18	Fully Connected Layer [32]	15
3.19	Example of CART model	17
3.20	Sequential Tree Structure	18
3.21	Support Vector Machine classification	19
4.1	Video dataset of interviews	21
4.2	Sentiments from each question	21
4.3	Emotions from the videos	22
4.4	22
4.5	22
4.6	One hot encoding for sentiment	22
4.7	Preprocessing for XG-Boost and SVM, part-1	23
4.8	Word cloud of the trainable tokens	27
4.9	LSTM training graph of Cross-Entropy Loss against Epochs	28
4.10	LSTM training graph of Accuracy against Epochs	28
4.11	LSTM on pre trained word2vec training graph of Cross-Entropy Loss against Epochs	29

4.12	LSTM on pre trained word2vec training graph of Accuracy against Epochs	29
4.13	GRU training graph of Cross-Entropy Loss against Epochs	30
4.14	GRU training graph of Accuracy against Epochs	30
4.15	Training graph of Cross-Entropy Loss against Epochs	32
4.16	Training graph of Accuracy against Epochs	32
4.17	Samples of the emotion dataset from the interview	33
4.18	Samples of Different Entities of Emotion	33
4.19	VGG-16 Architecture	34
5.1	Feed-Forward Neural Network	36
5.2	Training Graph Of Our Feed Forward Neural Network	36
5.3	Training Graph Of Our Feed Forward Neural Network	37
5.4	Training Graph Of Our Feed Forward Neural Network	38

List of Tables

4.1	Raw data from crowdflower_data dataset CSV file	26
4.2	Dataset after the removal of punctuations and digits	26
4.3	Accuracy of different models of sentiment analysis	32

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ANN Artificial Neural Network

CNN Convolutional Neural Network

GRU Gated Recurrent Unit

LSTM Long Short-Term Memory

ML Machine Learning

MLR Multivariate Linear Regression

RNN Recurrent Neural Network

SVM Support Vector Machine

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

Emotion detection can make our life easier in various ways. One of the most significant modes of detecting emotions is from facial expressions which can be applied in several crucial circumstances such as verification when it comes to the need for access and security check, criminal identification, verifying a payment, advertising, facial biometrics and so on. Facial recognition is successfully brought about using Deep learning's field convolutional Neural Network(CNN)[1] [2] [3] [4]. CNN is a multi-layer network trained to perform a particular task using classification. We have used CNN in our research for facial recognition in order to serve the purpose of detecting real-time emotions from employees when asked questions regarding their workplaces. This is one of the modes that we have used in order to identify the real time emotional state of employees. Another very important mode we have used is sentimental analysis. It is the process which is used to identify and categorize opinions expressed in the form of words. We have used Convolutional Neural Network for sentiment analysis as well[5]. CNN makes practical use of layers with convolving filters that are applied to local features. As CNN has been effective for Natural Language Processing(NLP), it has been of use. The CNN is trained with one layer of convolution on top of word vectors which have been found from a neural language that wasn't supervised. Both facial recognition as well as sentimental analysis have been used so that the results from both of these modes can be fed into Support Vector Machine algorithms in order to finally obtain the final result which provides the overall satisfaction of the employees.

1.1 Motivation

Neural network has been dramatically improving both in terms of accuracy and optimized performance and has been very useful in a variety of important areas. Neural network and deep learning recently solved many problems in image recognition, speech recognition and natural language processing. As we have been eager to do our research in a field that is not only expanding but also on great demand due to the day to day helpful applications in real life. Therefore, we intended to do research on neural networks, in particular Convolutional Neural Network (CNN)[6] and come up with ideas to detect real time emotion using multiple modes so that the final result that is found by both facial recognition and sentimental analysis combined is ensured as authentic. However, our topic is quite challenging, considering the fact that it is quite difficult for a machine to detect human emotions

even after training and testing since humans have the capability of hiding emotions through both facial expressions as well as the words which they speak. Therefore, there were some ethical obligations. Nevertheless, we have managed to collect data from numerous employees through both modes of facial recognition and sentimental analysis in order to identify their emotional state towards their work. The urge to come up with solutions for something quite new like ensuring employee satisfaction for proper mental health and productivity in workplaces using multiple modes of emotion has motivated us to do our research with utmost effort. We took it as a challenge and accomplished it. We have also considered this as an opportunity for us to contribute to the overall employment sector and therefore, the society.

1.2 Objectives

Our main objective is to find out if the standard and popular machine learning algorithms along with the Neural Network that is used in supervised learning are able to maintain their performance when they operate in multiple modalities and the results are passed through a feed forward neural network to get the actual real time emotional state. Additionally, if the renowned image processing algorithm CNN is able to perform the best with expected accuracy to be nearly perfect. Our intention was to also figure out which algorithms can be most suitable for emotion recognition so that the most accurate results are obtained. Our goal also focused on understanding which modalities of emotions are more easily recognized than the others so that the genuine real time emotions of the candidate can be shown. Our proposal is based on the real time emotions of the employees while they answer the questions asked regarding their workplace and its overall environment. As spoken words cannot be reliable enough, our intention is to emphasize on making sure to capture their true real time facial expressions along with their sentiments expressed in the words they speak. Therefore, we planned to classify the emotions into six basic entities: Happiness, Sadness, Anger, Surprise, Disgust and Neutral. By classifying, it is easier to significantly distinguish between the entities. We have trained and tested numerous data with facial expressions containing these six entities as well as sentences derived from tweets to describing these six entities. It is our mission to be able to use multiple modes in detecting emotions and finally reach a conclusion from multiple modes so that authenticity is ensured. In order to fulfill our target, we have followed the guidelines and research papers that have done research on emotional recognition prior to us.

Chapter 2

Literature Review and Related works

In this section, a detailed overview of previous research of similar works will be discussed. Mostly it will be about application of machine learning algorithms along with artificial neural networks used in emotional recognition.

2.1 Previous Research

In recent years, researchers have made significant progress in developing emotion recognition using EEG with Deep Recurrent Neural Networks, critical frequency bands and channels has been investigated using EEG-based emotion recognition models for three emotions: positive, neutral and negative[7] ECG signals have also been used effectively in application of emotional recognition. In terms of valence, arousal and dominance, EEG and ECG signals were recorded during effect elicitation by means of audio-visual stimuli[8]. In order to represent the characteristics associated with emotional states, a new effective EEG feature referred to as “Differential Entropy” was introduced and it was confirmed that EEG signals on frequency band gamma related to emotional states more closely compared to other frequency bands. [9]

As emotion is a subjective matter, gathering knowledge or the science behind labeled data has been quite challenging for several years. However, with the evolution of deep learning, it has been in the research for a long time. Facial recognition using CNN has been the most efficient model and has given quite good results on image classification problems[10]. Therefore, it has proven to be successful mainly in terms of detecting facial expressions from images. CNN layers are structured using multiple layers. On each layer, groups are formed based on the features that they work on and these groups feed forward to the next layer. At the end, it connects to the fully connected layer [10] [11].

Nonetheless, using only facial expressions to identify emotions is not authentic enough. Speech emotion recognition systems have been used. A recurrent neural network was initially used to classify six different emotions. Their performances were then compared to multivariate linear regression (MLR) and support vector machine techniques[12]. Machine learning algorithms have been proven to be effective and reliable for opinion mining as well as sentimental classification [13]. The use of emotion recognition has been so far used in different applications [10] [9]

[12]. Unfortunately, emotional states of employees and its impacts have not significantly been highlighted in most researches. It has been determined that job crafting predicted intrinsic need satisfaction which as a result predicted the mental well-being of employees in a research [14]. The importance of employee mental health has been highlighted by World Health Organization multiple times and its consequences leading to not only very poor performance in workplace but also an overall global impact [15]. Therefore, it is vital to verify the emotional state of each employee in a workplace and ensure their well being.

Chapter 3

Algorithms

This chapter talks about the different types of artificial neural networks and machine learning models, algorithms and elements which we have studied. For our purpose we have studied and used two types of popular neural networks, convolutional neural network and recurrent neural network. We will discuss the use of two particular types of recurrent neural networks for the purpose of our sentiment analysis, firstly the Long Short Term Memory, LSTM and the Gated Recurrent Unit, GRU. This chapter will also discuss an information extraction algorithm called Haar Cascade which has been used for the extraction of features.

3.1 Haar Cascade

Computer vision is the field in computer studies that studies how computer extract multiple level of information in digital format and how humans make use of that and perhaps automate their daily lives.

Haar Cascades is a type of cascade classifier, where multiple clustered classifiers form the system [16]. Haar Cascades is one of the earliest algorithms for grading images. Viola and Jones[17] initially proposed a way of using Haar wavelets to remove features in an image in their face detection algorithm[16]. They proposed a face detection algorithm using Haar Cascades. The principal concept behind the algorithm was the intrinsic structure present in all faces. For example, the area of the eye is darker than the forehead, and the area of the nose bridge boundary is darker than the eyes.

The HAAR Cascade Classifier is a simple but yet powerful algorithm to detect visual images quickly, many HAAR features are used to rapidly detect objects. Essentially a HAAR feature is one bit of an image subtracted from another bit of an image (fig-3.1), on its own these features are not effective at detecting and classifying images but when multiple (perhaps thousands or more) of them are used it becomes a powerful algorithm. They are similar to convolutional kernels and all possible sizes and locations of each kernel is used to calculate plenty of features[18]. There are multiple rectangular images called features (Fig-3.2), for example the middle of the nose is lighter than the side of the nose and the eyebrows are darker than the top of the eye brows, multiple such features are compared to the images of people to find which features work best. A dataset of positive images(that contain images we want to detect) and negative images(images that do not have the image we want to detect).



Figure 3.1: Example of an Haar-Cascade

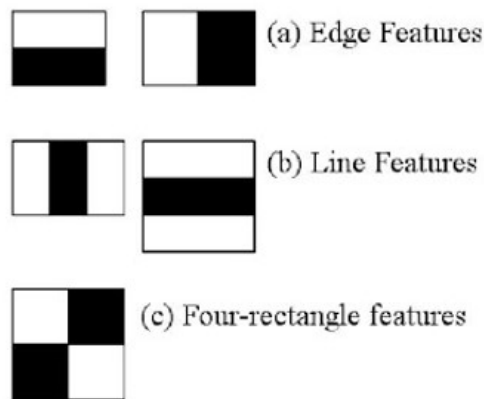


Figure 3.2: There are more than thousands of haar features

3.1.1 Selecting the features

A dataset of positive images (that which contain images we want to detect) and negative images (images that do not have the object we want to detect) is required for using AdaBoost to training the classifier and choose a reduced selection of features[16], selecting and training features from a variety of features which could be good features. For the creation of a strong classifier, AdaBoost permutes and combines several weak classifiers[19] for all training step.

From hundreds of thousands the total number of features are reduced to a few thousands. The features are found and arranged in the priority of classifying a given image.

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \dots, T$:
 - Normalize the weights,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$
 so that w_t is a probability distribution.
 - For each feature, j , train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t , $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
 - Choose the classifier, h_t , with the lowest error ϵ_t .
 - Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$
 where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
 where $\alpha_t = \log \frac{1}{\beta_t}$

Figure 3.3: The AdaBoost algorithm for classifier learning [17]

Haar-features

0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1

ideal Haar-feature
pixel intensities
0: white
1: black

0.1	0.2	0.8	0.8
0.2	0.3	0.8	0.8
0.2	0.1	0.8	0.8
0.2	0.1	0.8	0.9

these are real values
detected on an image

Δ for ideal Haar-feature is **1**

Δ for the real image: **0.74 - 0.18 = 0.56**

The closer the value to **1**, the more likely we have found a **Haar-feature** !!!
(of course we will never get 0 or 1: there are thresholds)

Viola-Jones algorithm will compare how close the real scenario is to the ideal case

- let's sum up the white pixel intensities
- calculate the sum of the black pixel intensities

$$\Delta = \text{dark} - \text{white} = \frac{1}{n} \sum_{\text{dark}} I(x) - \frac{1}{n} \sum_{\text{white}} I(x)$$

Figure 3.4: How the features are selected

3.1.2 Integral Image

Haar features of all sizes and locations have to be computed with the image, to reduce the time complexity of the mathematical operation the original image is turned to an integral image which we get by cumulative adding the values of subsequent pixels in both the horizontal and vertical axis(fig-3.5 [16]). Fig-3.6 shows a fully converted integral image.

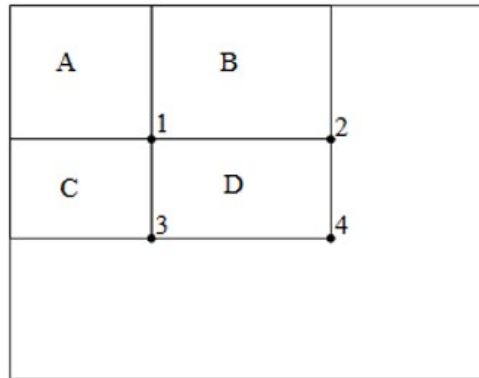


Figure 3.5: The value of the integral image at position 1 is the sum of pixel values in A. The value of the integral image at position 2 is $A+B$, at position 3 is $A+C$ and at position 4 is $A+B+C+D$

4	1	2	2
0	4	1	3
3	1	0	4
2	1	3	2

4	5	7	9
4	9	12	17
7	13	16	25
9	16	22	33

Figure 3.6: From left to right, a normal image to an integral image

3.1.3 Cascade Classifier

A decision tree like structure called a cascade is used, each process of this cascade is a classifier trained on AdaBoost to detect a negative image as soon as possible. If an image fails to pass any one of the cascade classifiers then the classification stops entirely and it is declared to not contain a face and if it passes through all the cascade classifiers then it is said to have a face.

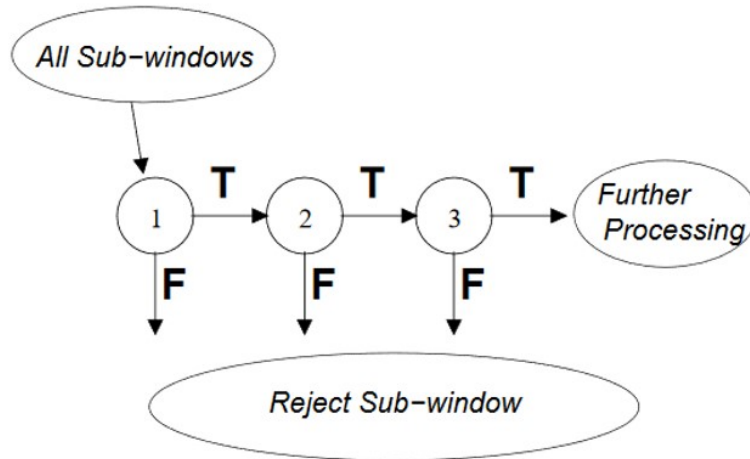


Figure 3.7: Cascade Classifier

3.2 Convolutional Neural Network

One of the first Artificial Neural Networks ever designed was based loosely on biological neural network by a psychologist Frank Rosenblatt in 1958[20]. It was a single neuron model with multiple inputs, each input has with itself an associated weight that is learnt during training. All the input values are multiplied with their associated weights and then added together, if the cumulated value is larger than a certain threshold value the neuron gives an output of 1 (fig-3.8)[21].

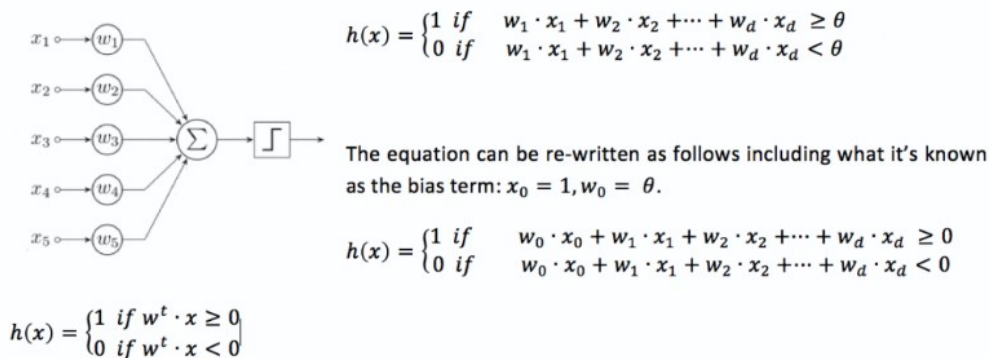


Figure 3.8: How a positron works and its equations

One of the biggest breakthroughs to come for deep learning in 1994 with the LeNet-5 architecture (Fig-3.9) which was the first convolution neural network [21]. A Convolutional Neural Network turns the input image into a feature and maps using convolutions and learns important features automatically, before this the features

needed to be hand crafted. The CNN was inspired by Hubel and Wiesel's[22] early work on the cat's visual cortex[23].

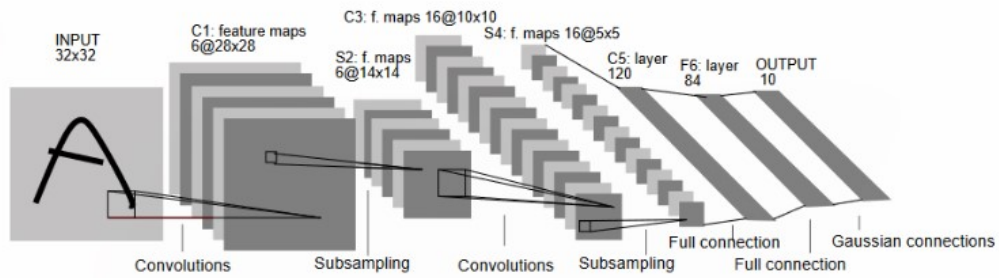


Figure 3.9: Architecture of LeNet-5 [21]

CNNs learn by using 2 processed, Gradient descent and backpropagation. Backpropagation repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector[24]. By how much the weights should be moved is calculated by a mathematical process called gradient descent. Gradient descent (fig-3.10) is the concept that the derivative of a function y measures the sensitivity to change of the function value (output value) with respect to a change in its argument x (input value). In other words, the derivative tells us the direction C is going. The gradient shows how much the x needs to change and in which direction to minimize y .

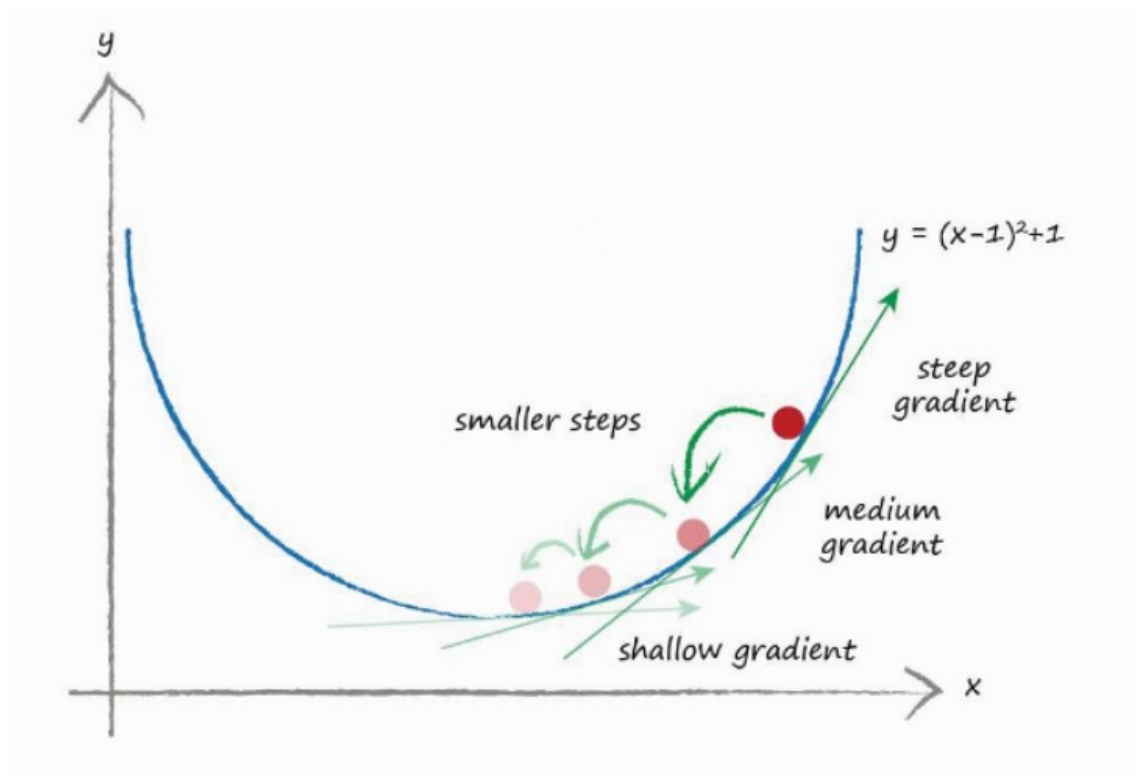


Figure 3.10: Gradient descent illustrated[25]

3.2.1 Convolutional Neural Network Layers

The CNN model achieves its object classification by using multiple mathematical operations through multiple layers. Through these layers it learns and classifies objects.

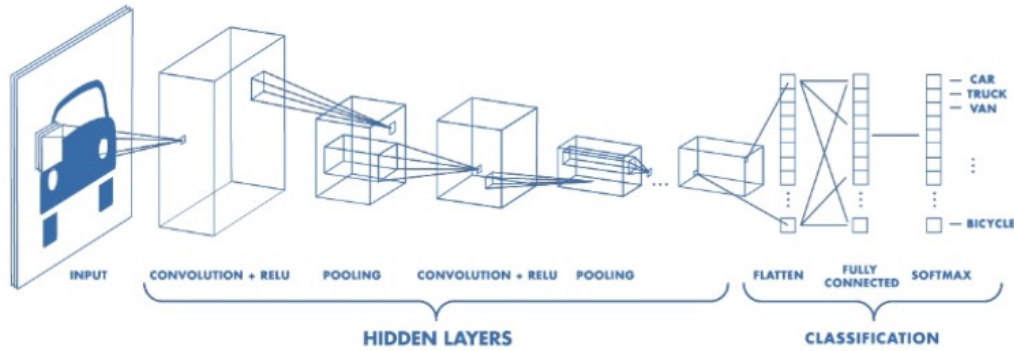


Figure 3.11: CNN Layers[26]

All the layers that general CNNs have are the Input layer, Relu(Rectified Linear Unit) layer, Pooling Layer, Fully Connected Layer and Output Layer. All of which will be further discussed below.

3.2.1.1 Input Layer

This layer represents all the values in all the pixels in the images about to be trained. An image is after all a matrix of values. Figure 3.12 illustrates how the input layer takes in values from all the pixels on an image.

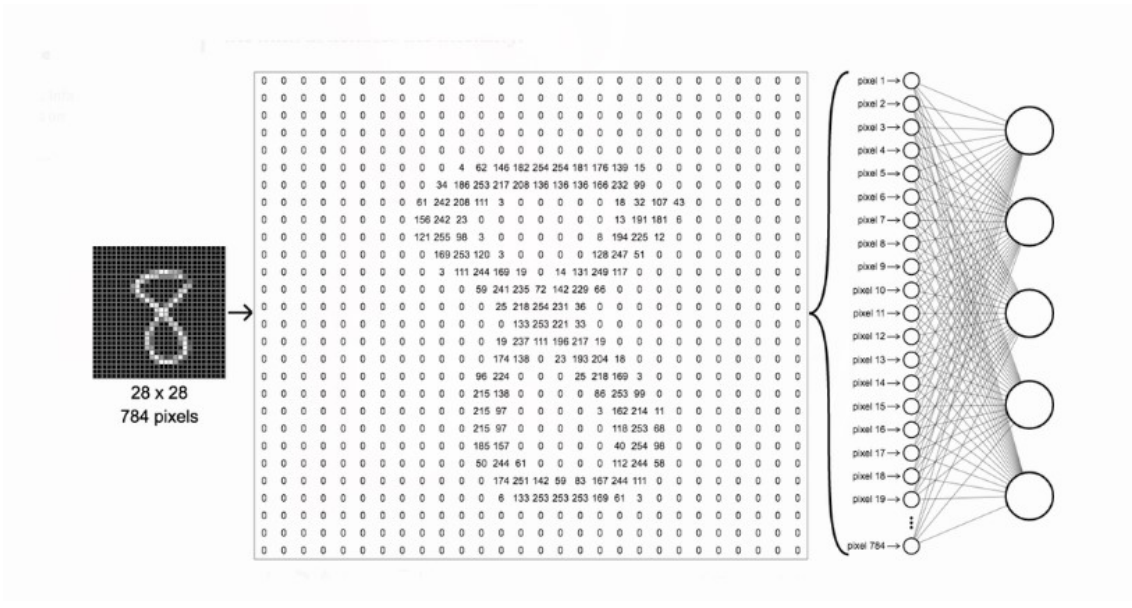


Figure 3.12: Values of each pixel of an image and how it is taken in by an input layer [26]

3.2.1.2 Convolution Layer

To understand the convolution we need to understand what a convolution is. In its most general form, convolution is an operation on two functions of a real-valued argument[27]. is a function derived from two given functions by integration which expresses how the shape of one is modified by the other[28]. It's equation is given below on fig-00sth. How the convolution operation is implemented in a CNN is shown on fig-3.13.

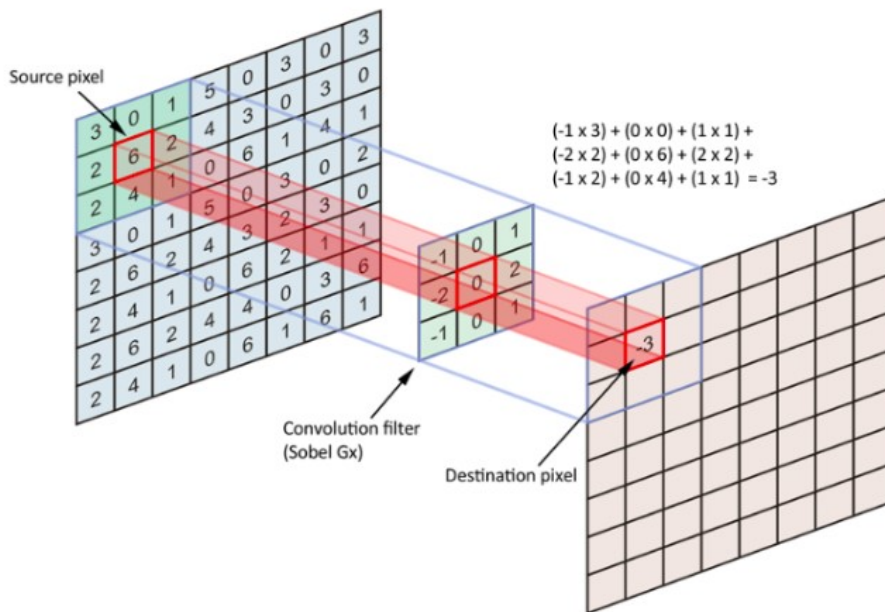


Figure 3.13: The convolution operation[28]

The main function of this layer is to gather features from images given to it. This does so by applying the convolution operation by sliding the kernel filter over the image and applying the convolution operation, the filter is a visual feature that helps recognize objects. The convolution between the image and filter is called a feature map. Higher the value is of a feature map the more the image reassembles the feature. The kernel isn't hand crafted it is learned using back propagation and gradient descent.

3.2.1.3 Pooling Layer

This layer receives the feature maps from the previous convolution layer and applies the pooling operation to the feature maps. It down-samples the image, reducing the image dimension while attempting to keep the important features.

There are 2 types of pooling operations used:

Max-Pool (fig-3.14): Calculate the average value for each patch on the feature map.

Average-Pool (fig-3.15): Calculate the maximum value for each patch of the feature map.

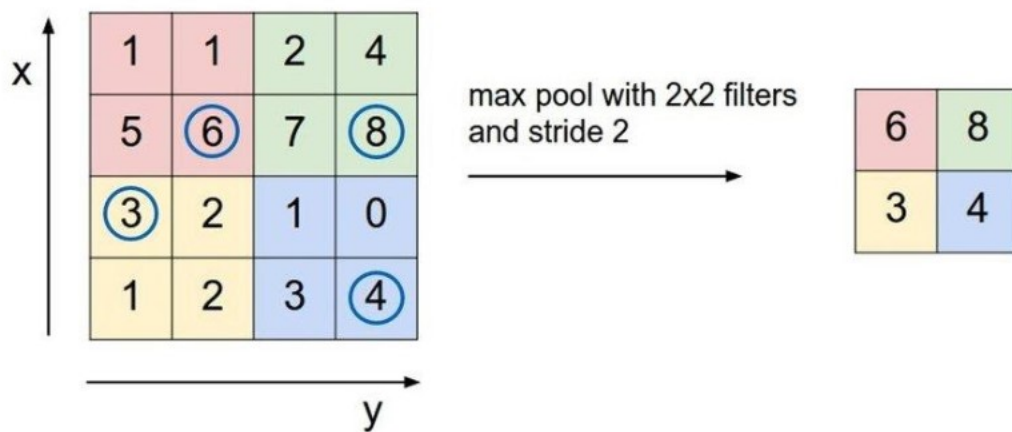


Figure 3.14: The max pooling operation [29]

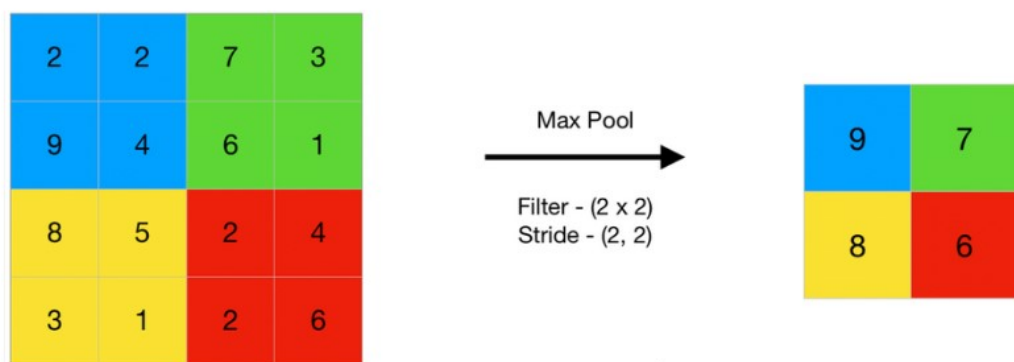


Figure 3.15: The average-pooling operation [30]

3.2.1.4 ReLu

ReLU (Rectified Linear Units) is a non-linear function(fig-15), that turns all the negative values it receives to zero without changing the positive values(fig-16). It is represented by the function $f(x) = \max(0,x)$.

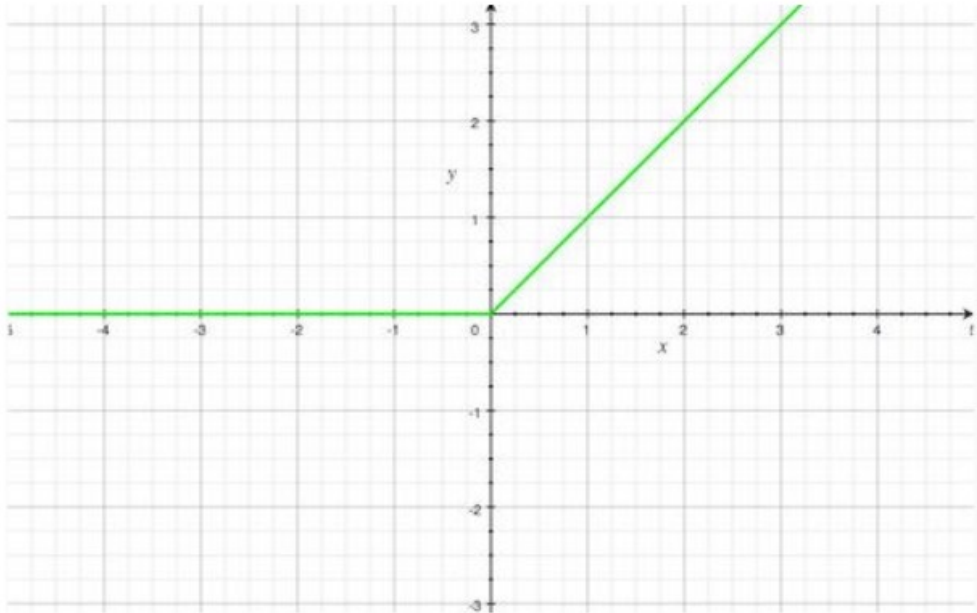


Figure 3.16: The ReLu function[31]

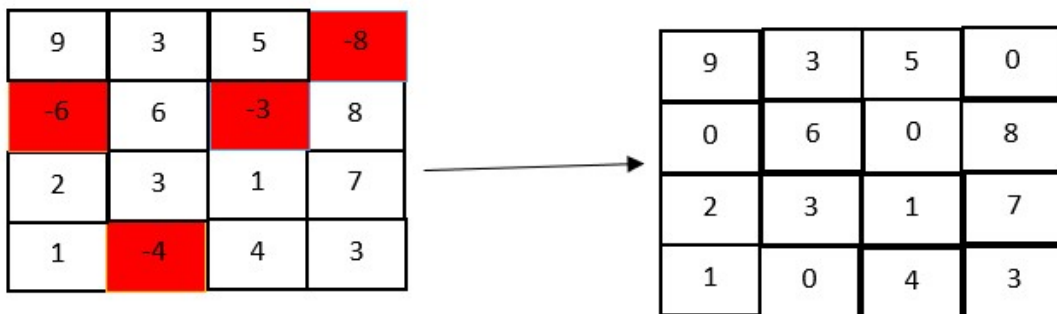


Figure 3.17: The ReLu Operation on the ReLu layer[31]

3.2.1.5 Fully Connected Layer

This layer basically means all the neurons are connected to all the other neurons on the layer after it. This receives an input and returns a vector of size N(number of labels in the classification). This classifies the original image into a label, each element in the vector is a label. The output vector is the probability of the original image being in a particular class. It is calculated by multiplying each input with its associated weight, adding them and then applying an activation function.

The fully connected layer's weights are also learnt like the filters in the convolutional layer using backpropagation and gradient descent.

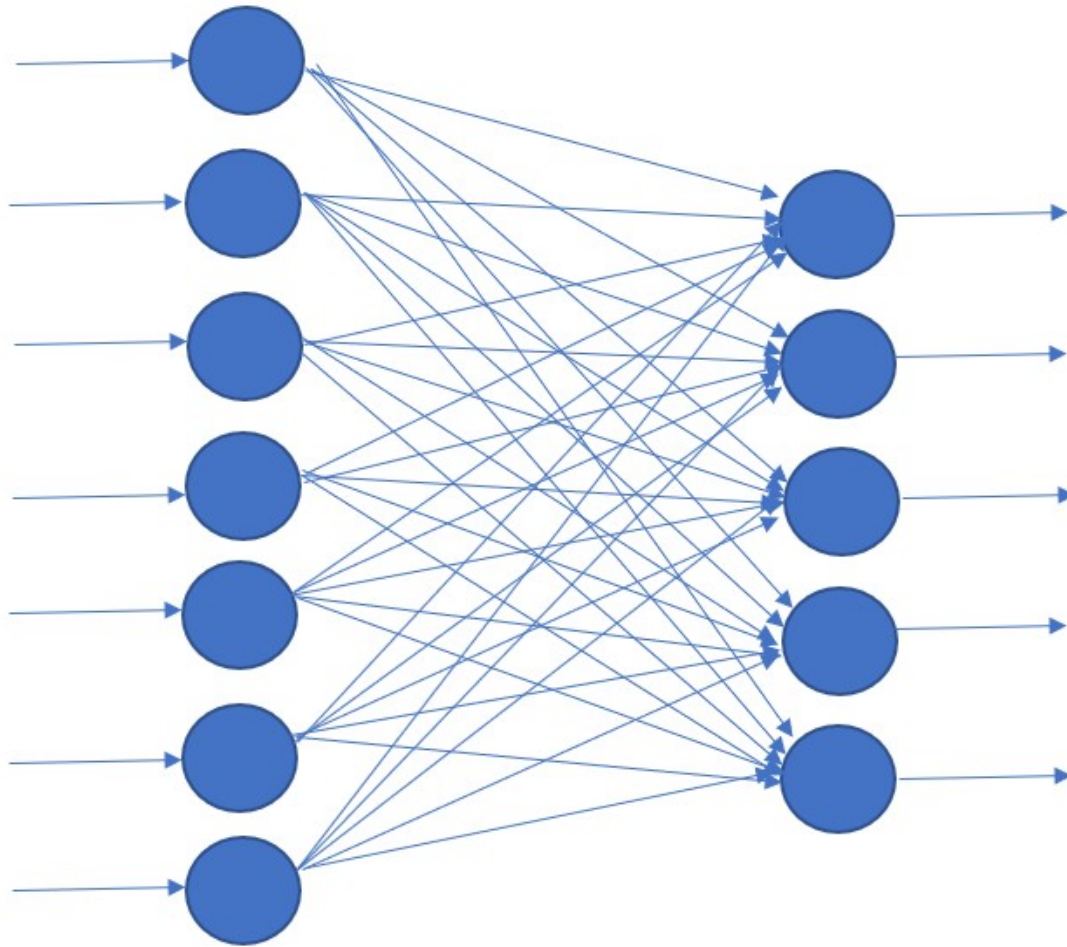


Figure 3.18: Fully Connected Layer [32]

3.3 Long Short Term Memory

Long short term memory, LSTM[33], is a variant of recurrent neural network having a very good reputation in the field of sentiment analysis. It consists of internal contextual state cells that act as long-term or short-term memory cells. This makes them capable of learning long-term dependencies. It was designed because recurrent neural networks cannot propagate data from previous time steps which are bigger in size because of its smaller memory. It is so effective, that it can process single information as well as a group of information, utilizing its input association structure. Its capabilities run from computer vision, speech recognition to time arrangement examination. LSTM has a very well connected structure along with a four-layer neural network. LSTM cells have input, next likely input, forget gate and output. Two types of non-linear functions, sigmoid and hyperbolic tangent, are used in the gates. These functions avoid exploding gradient problem by having inputs between -1 to 1 respectively.

The forget gate uses sigmoid activation function to forget values that range from 0

and 1.

$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

Here, W_f is weight function over h_{t-1} and x_t , the current input. Both the input and next likely input determine new cell state C_t . The input gate uses sigmoid activation function just like the forget gate, however, the latter uses hyperbolic tangent function giving it and C_t .

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$C_{t\sim} = \tanh (W_C \cdot [h_{t-1}, x_t] + b_C)$$

Subsequently, the output gate determines the current cell state output h_t by fusing the current state \tanh of C_t with function o_t .

$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$C_t \cdot f_t + i_t \cdot C_{t\sim}$$

$$h_t = o_t \cdot \tanh(C_t)$$

This output then becomes the input for the next cell state and the above-mentioned steps are replicated for all the cell states.

3.4 Gated Recurrent Unit

Gated Recurrent Unit (GRU) algorithm is a gating mechanism in recurrent neural networks. It has been introduced [34][35] in the year 2014. The GRU is like a long-short-term memory(LSTM)[33] with forget gate. However, it has fewer parameters than LSTM since it does not have an output gate. On smaller datasets, GRUs tend to perform better than LSTMs. In recent years, researchers have made significant progress when it comes to developing emotional recognition using GRU. As normal feedforward neural networks are unable to persist information, they are not the right choice for speech recognition[36]. LSTM networks and GRU networks have been evaluated so that their individual performance on speech data can be compared. The results showed that LSTM performed good but compared to LSTM, GRU achieved results close to those of LSTM in less time[36]. GRU Neural Network has mainly shown success in many applications that involve sequential or temporal data. In particular, they have so far proved to be effective in speech recognition, Natural Language Processing as well as Machine Translation[37]. By reducing parameters in the update gates and reset gates, three variants of the GRU in RNN have been evaluated. In fact, the computational expense also gets reduced[37]. The results of the experiment has shown that both GRUs and LSTMs are advanced recurrent neural networks which perform much better than traditional recurrent units, for

example tanh units. Furthermore, GRUs can be compared to LSTMs and can often work as replacement for LSTMs[35]. GRU networks have performed well mainly with long sequence applications. This is due to the fact that gating network signals control how the present input and the previous memory are used to update the current activation and and therefore produce the current state. During the training and evaluation process, the sets of weights of the gates are updated[37].

3.5 Extreme Gradient Boosting

XGBoost[38][39],the strong gradient boost, is a machine learning algorithm. They are used to improve tree algorithms. This algorithm is widely used for supervised learning problems, where several features of the training data are used to predict a goal outcome. This is a very effective powerful one as it uses the predictive power of various learners to achieve the answer. It is composed of both liner model solver and tree learning algorithms. It also supports different objective functions, such as regression, classification and ranking as well as additional functionality for cross validation and finding significant variables. It has several parameters that need to be managed for model optimization.

Initially the learners are weak. These weak learners add information for predicting. The sum of all the information added create a strong learner. This strong learner can now bring down bias and variance. The tree ensemble model has a set classification and regression trees (CART). Unlike decision trees, where the decision values are only available in the leaf, each leaf of CART holds a real score, resulting in better interpretation.

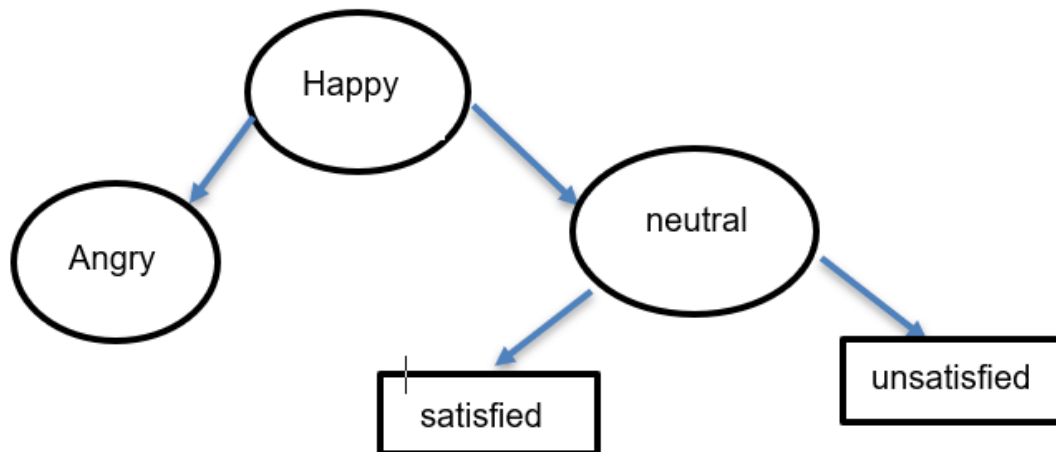


Figure 3.19: Example of CART model

The model gets trained to find the result , which is closest to the proper fitting for the training data x_i and labels y_i . This is achieved with the objective function which measures how well the model fits the training data. The function is the summation of training loss and regularization term.

$$obj(\theta) = L(\theta) + \lambda \sum_{j=1}^p \omega_j^2$$

Here, L represents the loss function and λ represents the regularization term that

controls the complexity of the model and helps to avoid overfitting.

Trees are being built parallelly in bagging. Boosting trees in an ordered manner so that each following tree reduces the error of the previous tree. Hence, the subsequent tree is always the updated version of the previous one. This is called additive strategy.

The entire process can be broken as follows [12].

1. Fitting a model to the data: $f_1(X)=Y$
2. Fitting a model to the residuals: $h_1(x) = y - F_1(x)$
3. Creating a new model: $F_2(x) = F_1(x) + h_1(x)$

This can be generalized as:

$$F(x) = F_1(x) \quad F_2(x) = F_1(x) + h_1(x) \dots \quad F_M(x) = F_{M-1}(x) + h_{M-1}(x)$$

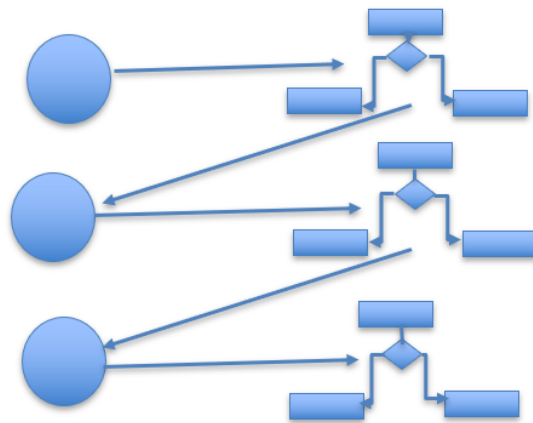


Figure 3.20: Sequential Tree Structure

In every step, the residual gets calculated: $h_m(x) = y - F_m(x)$ here $h_m(x)$ can be any model. Suppose rather than training h_0 on the residuals of F_0 , train h_0 on the gradient of the loss function, $L(y, F_0(x))$ with respect to the prediction values produced by $F_m(x)$. With samples in h_m clustered into leaves, an average gradient can be calculated and then scaled by some factor, η , so that $F_m + \eta h_m$ can reduce the loss function for the samples in each leaf. In practice, a non-identical factor is chosen for each leaf. For iteration $m = 1$ to M :

- Calculate the gradient of L at the point s_{m-1}
- “Step” in the direction of the highest descent (the negative gradient) with step size η . Which means, $s_m = s_{m-1} - \eta \nabla L(s_{m-1})$. If η is small and M is large enough then s_m will be the minimum value of L at s position. XGBoost also has incredible features[12] making it one of a kind that includes Handling sparse data: Missing values or data processing steps like one-hot encoding can render data sparse. XGBoost implements a sparsity-aware algorithm for split finding that can take care of various sparsity patterns in the data.
- Weighted quantile sketch: Tree based algorithms that find split points when the data points are of equal weights (using quantile sketch algorithm). Weighted data cannot, however be treated. XGBoost has a distributed weighted quantile which

manages weighted data effectively.

- Block structure for parallel learning: XGBoost can use several cores on the CPU for faster computations. Unlike other algorithms, this allows the structure of the data to be reused iterations, rather than computing it again.
- Cache awareness: XGBoost requires non-continuous memory access by row index to get the gradient. Therefore, it was planned to allow maximum use of hardware.
- Out-of-core computing: This feature optimizes the usable disk space and maximizes its use when computing non-memory compatible datasets.

3.6 Support Vector Machine

Support Vector Machine, SVM is a supervised machine learning algorithm. It is highly useful for classification of data[13]. This algorithm plots all the information item in n-dimensional area where n is scope of choices. The worth of the feature being the value of a specific coordinate, classification is done by discovering the hyper-plane that can recognize the different categories. It is a non-probabilistic linear classifier but it can also perform nonlinear probabilistic classification by designing a versatile algorithmic program. In SVM, instances are set as points in a mapped area which gets labelled using the transparent gap between the points. Unseen specimen are then mapped into the identical area and gets anticipated in which class it would be supported and which aspect of the gap are included. It is highly effective in high dimensional area[14][15]. Further, it is memory efficient. Also it is fast at evaluating the learned target functions.

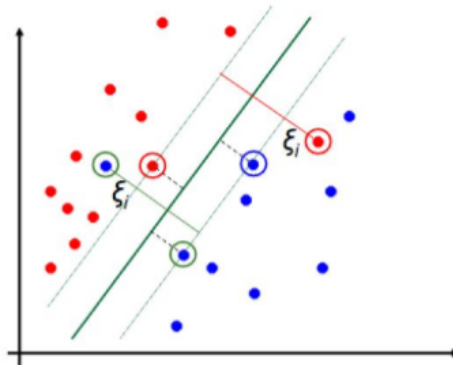


Figure 3.21: Support Vector Machine classification

Chapter 4

Methodology

This chapter describes how emotion recognition has been implemented to find employee satisfaction using Convolutional Neural Network, CNN, and Recurrent Neural Network, RNN. It includes the describing the data and how it has been processed. It also contains how the derived results from the neural networks have been compared and the ones with the most accuracy have been selected to train different Machine Learning algorithms. Also, the overall process of achieving the end result using a Likert Scale is explained.

4.1 Experimental Setup

Multiple neural networks are used for strongly separate architectures, each network operates on its own domain independently. In our case we have used VGG16 to classify facial expressions using an accumulation of multiple datasets and Yoon Kim's CNN to train sentiment analysis, on top of that we have added a feed-forward neural network(fig-18). Each network is built and trained for a different task. The VGG16 just identifies the facial expression and Yoon Kim's CNN gives the sentiment with the highest probability. Both run independently and their output is given to the Feed Forward Neural Network after them. The final decision is made from the outputs of their previous neural network(s), which are commonly called expert networks or agents. The modular architecture combines the two different outputs from the neural networks, this enhances the overall generalization which could be significant in a high dimensional space[40].

We have also decided to use XG-Boost and Support Vector Machine on the output data that we got from sentiment analysis and facial expression classification to classify and predict the job satisfaction. We needed to change the data structure that need to be given to XG-boost.

We decided to conducted interviews(fig-4.1) of different employees from various profession. The interviewees were of different age groups. This was done to make the data generalized and not focused on any particular profession or age range. The interviews were consisted of questions where we have asked people regarding their jobs according to the questionnaire which is discussed in details below. The interviews were recorded as mp4 videos, each video represents the answer of each of the questions in the questionnaire and the file name of the video represents the sequence of the answer according to the question. In total we conducted 160 interviews. The 80% of the interviews were then used to train and the rest 20% are used for testing

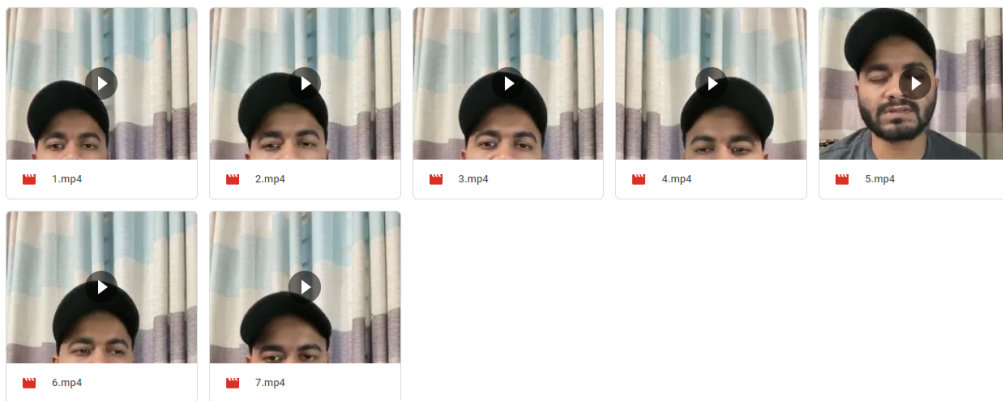
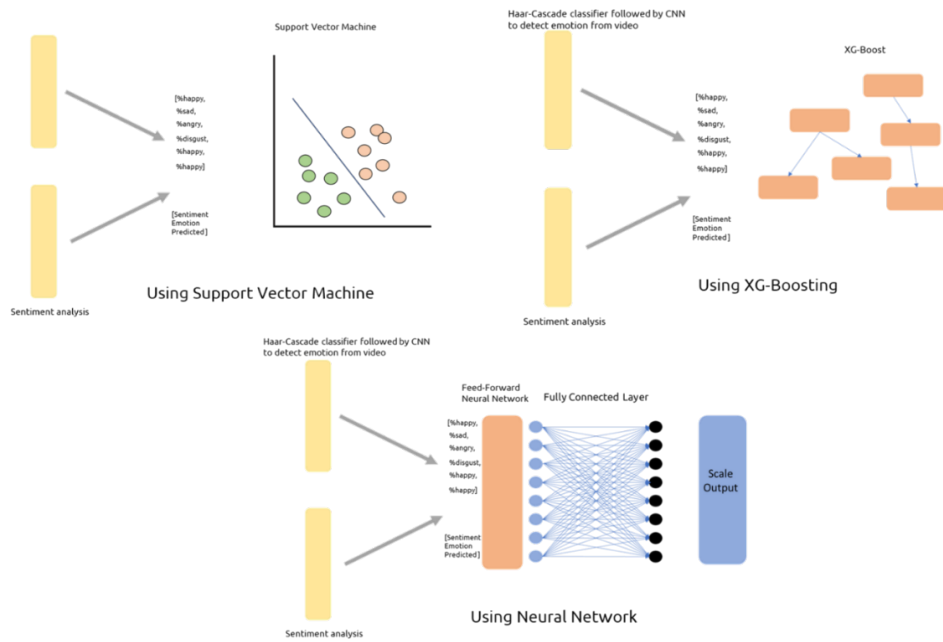


Figure 4.1: Video dataset of interviews

4.1.1 Data Processing

For sentiment analysis we have used Google’s speech to text API to convert the speech into text and run it through Yoon Kim’s CNN. All the videos of an individual get to get a vector of sentiment emotions(fig-4.2).

```

Q1 [angry]
Q2 [disgust]
Q3 [angry]
Q4 [sad]
Q5 [neutral]
Q6 [angry]
Q7 [disgust]

```

Figure 4.2: Sentiments from each question

All the videos are then run through VGG-16. All the frames of each video are

classified to an emotion and for each video we get a vector that illustrated the proportion of emotions shown in each frame(fig-4.3).

```

Q1 ["50% angry", "30% disgust", "5% happy", "5% sad", "5% surprise", "5% neutral"]
Q2 ["10% angry", "70% disgust", "5% happy", "5% sad", "5% surprise", "5% neutral"]
Q3 ["19% angry", "20% disgust", "55% happy", "2% sad", "4% surprise", "0% neutral"]
Q4 ["1% angry", "7% disgust", "5% happy", "12% sad", "55% surprise", "20% neutral"]
Q5 ["12% angry", "3% disgust", "5% happy", "10% sad", "15% surprise", "55% neutral"]
Q6 ["1% angry", "70% disgust", "5% happy", "15% sad", "8% surprise", "1% neutral"]
Q7 ["13% angry", "4% disgust", "17% happy", "39% sad", "21% surprise", "6% neutral"]

```

Figure 4.3: Emotions from the videos

```

angry: [0, 0, 0, 0, 0, 0, 1]
disgust: [0, 0, 0, 0, 0, 1, 0]
scared: [0, 0, 0, 0, 1, 0, 0]
happy: [0, 0, 0, 1, 0, 0, 0]
sad: [0, 0, 1, 0, 0, 0, 0]
surprised: [0, 1, 0, 0, 0, 0, 0]
neutral: [1, 0, 0, 0, 0, 0, 0]

```

Figure 4.4

```

[[ 0.68965517 0.68965517 0.68965517 1.37931034 0.        0.68965517
 95.86206897]
 [ 2.81456954 0.        4.8013245 6.45695364 0.99337748 0.33112583
 84.60264901]
 [ 0.17094017 0.        0.17094017 1.02564103 0.        0.
 98.63247863]
 [ 0.22573363 0.        0.        0.        0.        0.
 99.77426637]
 [ 0.7518797 0.        0.        2.63157895 0.        0.
 96.61654135]
 [ 1.27591707 0.15948963 3.18979266 4.94417863 0.79744817 0.
 89.63317384]
 [ 0.22421525 0.        0.        2.69058296 0.        0.44843049
 96.6367713 ]
 [ 6.        3.        6.        4.        3.        4.
 4.        ]]

```

Figure 4.5

```

1: [0, 0, 0, 0, 1]
2: [0, 0, 0, 1, 0]
3: [0, 0, 1, 0, 0]
4: [0, 1, 0, 0, 0]
5: [1, 0, 0, 0, 0]

```

Figure 4.6: One hot encoding for sentiment

To prepare data for the neural network we one hot encoded the sentiment data from each individual as shown in fig-4.4 and combined the data from the Facial expression classification and sentiment analysis to get an 8x7 matrix (fig 4.5) to be trained. The labels data classify on a scale of 1-5 the satisfaction level and we have also used one hot encoding according to fig-4.6.

For XG-Boost and Support Vector Machine we turned our data into 9 columns (Fig-4.7), ques1-ques7 column represents the maximum sentiment shown in the video of each video and the columns senti1-senti7 represents the sentiment analysis of what the interviewee said. Then using one Hot encoding each of the string categories were given a

	ques1	ques2	ques3	ques4	ques5	ques6	ques7	senti1	senti2	senti3	senti4	senti5	senti6	senti7
0	neutral	neutral	neutral	neutral	neutral	neutral	neutral	neutral	happy	neutral	sad	happy	sad	sad
1	neutral	neutral	neutral	neutral	neutral	neutral	neutral	happy	neutral	happy	happy	happy	neutral	happy
2	angry	angry	sad	neutral	sad	angry	angry	sad	neutral	sad	sad	angry	neutral	neutral
3	angry	neutral	neutral	neutral	neutral	neutral	neutral	happy	happy	happy	angry	neutral	happy	neutral
4	neutral	neutral	neutral	neutral	neutral	neutral	neutral	neutral	sad	neutral	sad	angry	neutral	happy

Figure 4.7: Preprocessing for XG-Boost and SVM, part-1

4.2 Questionnaire Development

As it has been discussed earlier in this paper, the real time emotional states of employees working in various workplaces have been significantly taken into consideration in this research. Although emotional recognition has already been applied in several different sectors, the importance of comprehending the emotional states of employees has not been emphasized on. Therefore, this research focuses on capturing the genuine emotional states of employees regarding their work. The mental health as well as the mental attitude that a certain employee possesses towards their work has a major impact on their individual performance and as a result the overall success of an organization. Most mental health professionals have studied that the workplace environment plays a vital role in an individual’s mental well-being and vice versa [41]. According to the World Health Organization (WHO), mental health is a state of well-being that causes every individual to understand his/her potential. It depends on his/her mental health whether they are able to adjust to regular stresses of everyday life, productivity and thus their contribution in the workplace is determined. WHO also found out that due to poor mental health of over 300 million employees around the globe, the global economy loses \$1 trillion in productivity every year [42].

With the intention of capturing the genuine real time emotional states of employees in a variety of workplaces, in this research a set of questionnaires has been composed having studied the different reasons and areas of a workplace that are capable of affecting the emotional state of an employee [42]. One of the most obvious factors is the environment of the workplace. The environment of the workplace can generally refer to a lot of things including the health and safety policies. It is also essential that the individual finds their work to be enjoyable and meaningful as finding their work monotonous or feeling that they are performing unpleasant tasks results in poor emotional state. An employee’s emotional state is largely impacted by how he/she is treated by their supervisor/manager [41]. Moreover, whether the individual is being able to cope with the work pressure or not is an important question. Another factor that demotivates and deteriorates an employee’s mental state is if their system of promotion is not satisfactory. Additionally, if an employee is

not pleased with his/her co-workers/colleagues, this majorly affects their emotional state. Last but not the least, the salary of every employee turns out to either work as a huge inspiration to work or a discouragement otherwise depending on whether or not they are satisfied with it [43]. Based on these criteria that have been mentioned, this research prepared the following set of questionnaire:

- How is your working environment?
- How enjoyable and meaningful do you find your work to be?
- Are you satisfied with how your supervisors treat you?
- How stressful do you find your work to be?
- How satisfied are you with your system of promotion and your content of work?
- Do you have complains regarding your co-workers and/or working in groups?
- What do you feel when you come to the office?
- Are you satisfied with your salary?

The real time emotional state of the employee is judged according to the answers that are received. The answers are received in the form of mp4 video. The speeches are converted to texts and the videos are converted into stream of frames. Those inputs are then used for facial recognition using a neural network and sentimental analysis using another neural network. The final result of the emotional state is derived by judging the outputs of the two neural networks according to the Likert Scale range [4] with the help of a Support Vector Machine algorithm. In addition to measuring statements of agreement, Likert scales can measure other variations such as frequency, quality, importance, and likelihood, etc. It is required in order to understand the intensity of the emotional state expressed by the employee. Likert scale is mainly used to measure the level of mental satisfaction of the employee. The scale ranges from 1 to 5 as follows:

- Highly dissatisfied
- Dissatisfied
- Neutral
- Satisfied
- Highly Satisfied

The six entities of emotions that are used in this research are Happiness, Sadness, Anger, Disgust, Neutral and Surprise. According to the emotions detected through both facial recognition as well as sentimental analysis, the emotional state in terms of both facial recognition as well as sentimental analysis is decided and fed into a Support Vector Machine algorithm in order to detect the employee satisfaction.

4.3 Creating the dataset for the Final Model

To achieve our goal, we have created our own dataset by conducting interviews. We have interviewed 160 different employees in order to determine emotional states of employees working in various job sectors. We have made sure to collect data from people working in a wide variety of workplaces starting from employees working in grocery markets as salespeople to employees who work as managers in their company. Therefore, our datasets include answers to the work-related questions of doctors, teachers, engineers, bankers and people from multiple other different professions. Not only have we created a dataset from people of different professions, but we have also ensured that this created dataset is from people of diverse backgrounds who live in different places around the globe so that divergent perspectives of emotions was achieved. The interviews of people living abroad were conducted over the internet through video calling which was recorded. One thing we have ensured is that our video recordings are done only after getting the consent of the interviewees. Through this we did not violate anybody's privacy or confidentiality. And for this purpose, our dataset cannot be exposed and is being kept among the members of this research group only, since, some of our interviewees are insecure about their information being leaked in wrong terms.

The questions that we have designed, as discussed above were based on a vast research done on employee satisfaction and factors that affect the mental health of employees. We have constructed our set of questionnaire ensuring that all the factors of job satisfaction were covered including an employee's working environment, work stress, salary, whether or not they find their job enjoyable and meaningful, how they are treated by their supervisors, a question regarding their co-workers, the system of promotion of their job, how their workplace makes them feel and so on. Precisely, we made eight separate videos of each employee answering the eight different questions in real time. From each video, we analyzed their answers using both facial recognition as well as sentimental analysis. Finally, the two results obtained from sentimental analysis and facial recognition were then passed through a third Feedforward neural network, XGboost and an SVM so that the authentic emotional state of the employee can be captured from these two important modes of emotion detection. From all the videos, we have classified the emotions of the employees into six basic groups: Happiness, Sadness, Anger, Neutral, Surprised and Disgust. In this way, we were able to find out the emotions distinctively. Unfortunately, we were not able to conduct interviews of more than 160 people, but we do intend on making the dataset larger, in our future proposed research so that we can work in further details and obtain a higher level of accuracy in our research.

4.4 Emotion Detection from Sentimental Analysis

Sentimental analysis [44] [45] [36] [9] is carried out to find the emotions of the employees towards their work, office environment and pay scale. Two types of neural networks have been used in this process, both CNN [5] [1] [2] and GRU and LSTM of RNN. The neural networks have been trained using a dataset called

crowdflower_data, which contains 40,000 tweets[47] of different sentiments. This dataset is public and can be downloaded by using this link, The tweets have been screened and only those which belong to the emotion categories that match with the classifications used in this have been chosen. As mentioned earlier this paper classifies the six basic emotions of humankind which are happy, sad, angry, disgusted, surprised and neutral. The results from both the neural networks are then compared and the one with the higher accuracy is used in the feedforward neural network to derive the actual result. Different models of these two different neural networks have been used, especially in the case of the RNN models. The first few models of LSTM and GRU proved to have very low accuracy which could not be used to determine the sentiments. However, eventually the model with a decent accuracy was derived from lstm.in total four different methods to determine sentiments from texts have been used. Below complete description of the models for sentiment analysis are being provided.

Serial no.	Tweets
1	@tiffanylue i know i was listenin to bad habit earlier and i started freakin at his part =[
2	Layin n bed with a headache ughhhh...waitin on your call...
3	Funeral ceremony...gloomy friday...
4	wants to hang out with friends SOON!
5	@dannycastillo We want to trade with someone who has Houston tickets, but no one will
6	Re-pinging @ghostidah14: why didn't you go to prom? BC my bf didn't like my friends
7	I should be sleep, but im not! thinking about an old friend who I want. but he's married now. damn, & he wants me 2! scandalous!
8	Hmmm. http://www.djhero.com/ is down
9	@charviray Charlene my love. I miss you
10	@kelcouch I'm sorry

Table 4.1: Raw data from crowdflower_data dataset CSV file

i. Data Pre Processing: Before the data are fed into the neural networks, they are processed [46] [25] so that proper results can be drawn from them. In terms of computer science, the data which is being used, the tweets, are known as strings. Each string contains a sequence of characters. The length of the strings have varied from being a single word to long sentence. The very first step of data processing for sentiment analysis was converting all the English characters in lowercase. Once that was done, all the punctuations in the tweets were removed. Punctuations are written signs which are used to express the meaning of sentences vividly. There are fifteen punctuations in English Grammar which are period, comma, exclamation point, question mark, colon, semicolon, bullet point, dash, hyphen, parenthesis, bracket, brace, ellipsis, quotation mark, and apostrophe. After the removal of the punctuations, digits were removed from the tweets. Digits are symbols that represent numbers according to some positional numeral systems used for counting and calculating.

Serial no.	Tweets after removing signs and digits
4	i should be sleep but not thinking about an old friend who i want but hes married now damn amp he wants me scandalous
5	charlene my love i miss you
6	sorry at least its friday
7	ugh i have to beat this stupid song to get to the next rude
8	if u watch the hills in london u will realise what tourture it is because were weeks and weeks late i just watch it online lol
9	the storm is here and the electricity is gone
10	so sleepy again and its not even that late i fail once again

Table 4.2: Dataset after the removal of punctuations and digits

ii. Data Processing: Once the data have been pre processed, final processing of data was carried out after which the data was fed into the neural networks for

training and testing. The pre-processed data is then converted into arrays of tokens or lexicons [44] [47] where each token consists of only a single word from the English language, for example “can”. The tokens then went through further processing, where the stop words of the English language are removed. The stop words are the articles, pronouns and the Be verbs, which are not useful in determining the emotions hidden in the sentences. After removing the stop words from the data, the data now contained tokens which expressed certain emotions.

For better visualisation of the processed data we used wordcloud. A wordcloud is a visualization wherein the most frequent words appear in large size and the less frequent words appear in smaller sizes. Here, no logo or image was used for the purpose of simplicity. No color has been imposed over the wordcloud. Then recolor the words from the dataset to the image’s color. The background of the image was kept black so that it can be easily visualised Interpolation has been used to smoothen the generated image. These tokens then go through embedding during the training process which is described below.

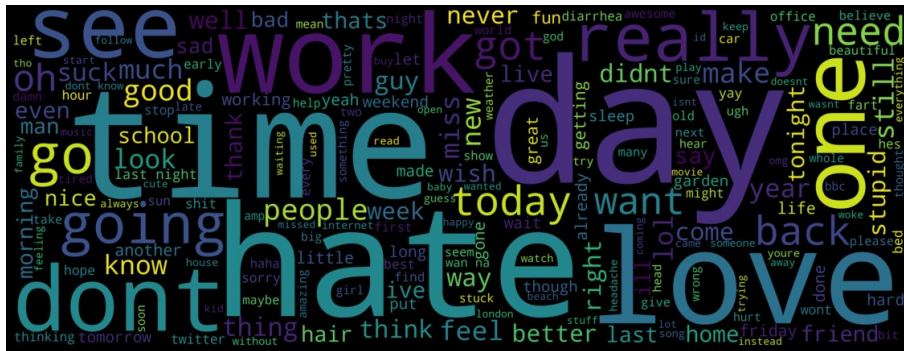


Figure 4.8: Word cloud of the trainable tokens

iii. Emotion Detection Using LSTM: Long Short Term Memory [48] [49] [33] has feedback connections that allows it to process an entire sequence of data. However, this methodology is very time consuming while training a large dataset. But, LSTM is renowned for its reliability. Therefore, this artificial recurrent neural network architecture was implemented for the sentimental analysis. The LSTM model which was used in this paper is comparatively a small model but was enough to generate a decent accuracy.

For the implementations of LSTM, the tokenized data is then converted into sequences of integers and gets embedded [3] [39] [50].

The LSTM model that has been built for the sentimental analysis in this paper consists of an instance of the Sequential class. The first layer of this model is the embedding layer which receives as input an integer matrix of size (34,54) and a vocabulary of size 3000. This layer gives an output of shape (*, 54, 32). Then a layer of LSTM is added to the model. This layer has 1080 units. The unit number has been changed many times during the experiment. If the units are increased or decreased, the accuracy gets significantly low mainly because of overfitting or underfitting. The return_sequence has been set to false, so that the last hidden state output captures an abstract representation of the input sequence. Linear activation was used for this model so no activation argument was passed. The LSTM layer

was specified to expect 54 steps and 30 features. After that a dropout layer was added to the model. The dropout was set at a rate of 0.2 to force other weights to help generalize the network. Subsequently, the last layer was added to the model. This last layer was a dense layer. This layer took a sigmoid activation function as an argument. The output from this layer is an array of shape $(*,6)$. No dense layer was added to this model.

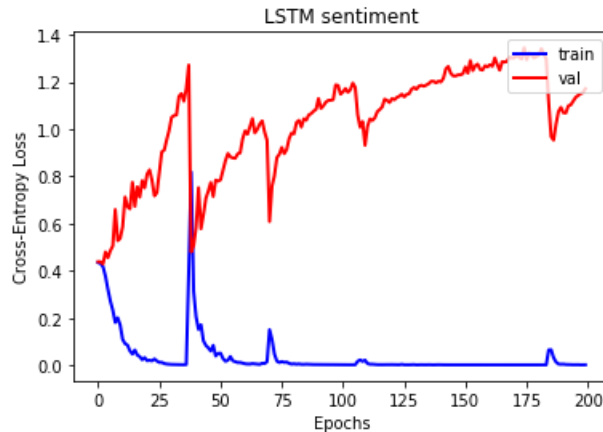


Figure 4.9: LSTM training graph of Cross-Entropy Loss against Epochs

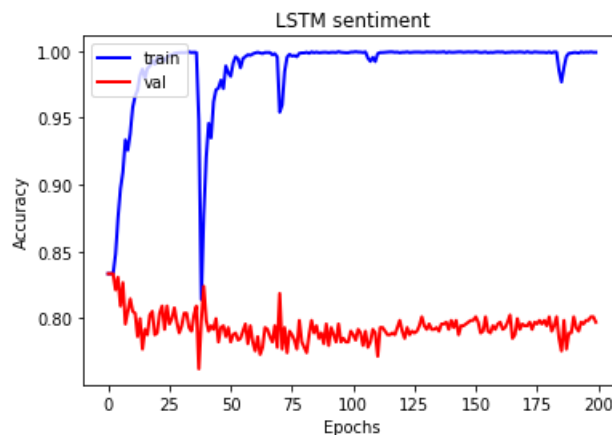


Figure 4.10: LSTM training graph of Accuracy against Epochs

iv. Emotion Detection Using LSTM over pre-trained Word Vector: As a part of the experiment, a different approach was taken to increase the accuracy of the sentiment analysis. This time an LSTM model was developed which would get trained on a pre trained word vector [51]. This type of training is a form of semi supervised learning [45]. The LSTM model of this method was kept identical to that of the previous model. However, since a pre-trained word vector is being used in this model, the model is expected to have better accuracy than the previous model. Here, the model has been kept identical to the previous LSTM model so that valid comparison can be made about the improvements in accuracy. The pre trained word vector which was used in this model provides word embeddings[39] and is done completely randomly in the embedding layer of the model. Google News corpus word

vector model[51] has been used as the pre-trained word2vec. This model of word vector consists of 3 million words and phrases. The model was trained on roughly 100 billion words from a Google News dataset and the vector length is 300 features. The model was built with an instance of the Sequential class. The Embedding layer was added to the model, which produces a vector of the optimum size. The weight of this layer has been kept constant. This layer creates an output of shape (*, 50, 300). Subsequent to this layer the only LSTM layer was added. This LSTM layer, just like the previous model, had 1080 units and its return_sequence has been set to false. Linear activation was used for this model to emulate the first model of LSTM, so no activation argument was passed and the layer was specified to expect 54 steps and 30 features. A dropout layer having a dropout of 0.2 was provided in an attempt to generalise the network. And finally, a dropout layer of sigmoid activation argument was set as the last layer of the model. Again, no dense layer was added to the LSTM model.

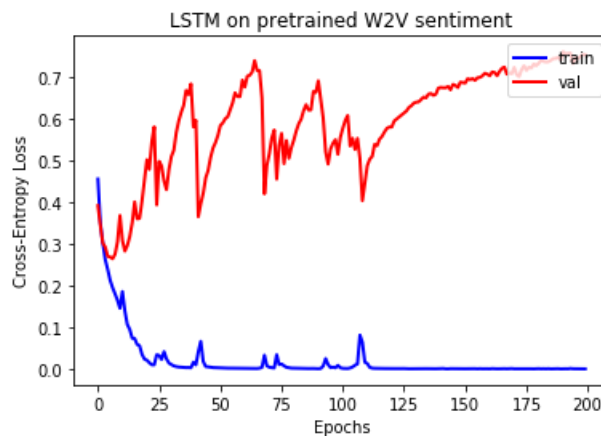


Figure 4.11: LSTM on pre trained word2vec training graph of Cross-Entropy Loss against Epochs

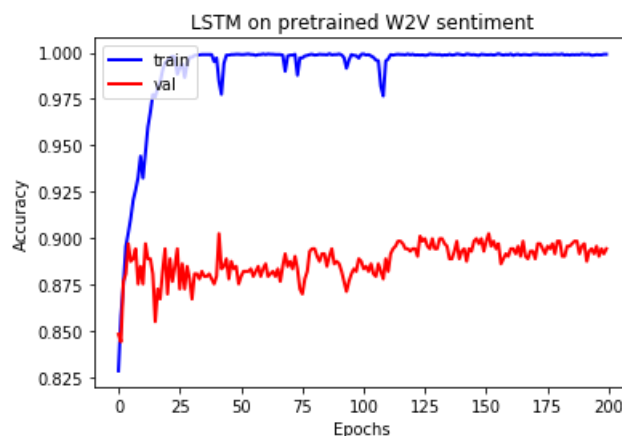


Figure 4.12: LSTM on pre trained word2vec training graph of Accuracy against Epochs

v. Emotion Detection Using Gated Recurrent Unit: Another form of recurrent neural network, RNN, was used for sentimental analysis. Here, the Gated

Recurrent Unit, GRU [36] [37] [52] has been utilised for the purpose of sentimental analysis. This model was built, despite being very similar to LSTM, since it is a modern version of recurrent neural networks. This model, just like the previous two models, is a small neural network. This model is also a linear stack of layers and has been built with an instance of the sequential class. The first layer of this model is the embedding layer which receives as input an integer matrix of size (34,54) and a vocabulary of size 4000. This layer gives an output of shape (*, 54, 32). A GRU layer has been added followed by this. The GRU layer has 100 units, which proved to be sufficient enough for the purpose of producing a better accuracy than the LSTM layer. The unit of the GRU layer was determined on a trial and error basis. The final layer added to this small but efficient model was the dense layer. The layer had a sigmoid activation argument passed to it. The output from this layer is an array of shape (*,6). This model had no dropout layer unlike the LSTM models which were discussed above.

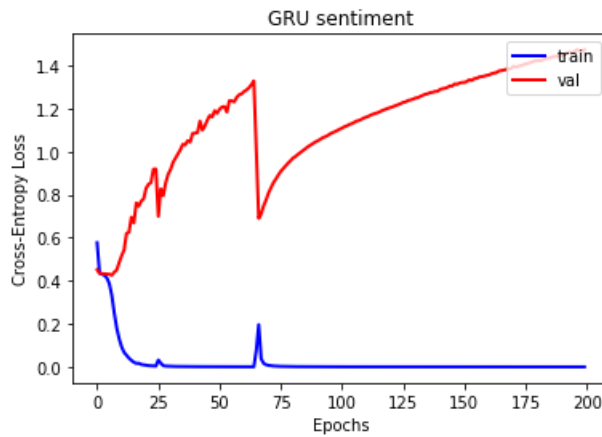


Figure 4.13: GRU training graph of Cross-Entropy Loss against Epochs

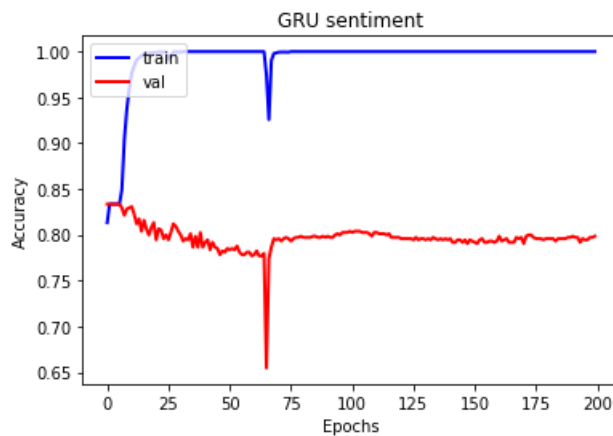


Figure 4.14: GRU training graph of Accuracy against Epochs

vi. Emotion Detection Using YOON KIM'S Model of CNN: Aiming to have better accuracy, another class of artificial neural network has been used. This time Convolutional Neural Network [1] [2] [53] was used. The model which is being used is known as the Yoon Kim's model [5], where the convolutional neural network gets

trained on top of a pre-trained word vector. The CNN used in this model consists of little hyper parameters. The most interesting fact about this model is that it takes very little time to achieve a very high accuracy. The approach that has been used is the CNN-static approach, where a word2vec is used to provide word embeddings. The embedding of each word is provided completely randomly in this case. Google News corpus word vector model [51] has been utilised here as well as the pre-trained word2vec, since this model of word vector consists of 3 million words and phrases and was trained on roughly 100 billion words from a Google News dataset and its vector length is 300 features which makes this very reliable word vector.

The model consists of the embedding layer as the first layer just like any other convolutional neural network. It compresses data into a much smaller size and tries to find the optimum mapping of the input sequences into a new and compressed vector. The weights of the layer are prevented from being updated during the training. The desired dimension of the dense layer provided was 50. At the end of this layer a 3D tensor can be derived. The output of the embedding layer is then passed onto 1D convolutional layer. Here 1D convolutional layer is being used instead of 2D convolutional layer which are more popular because 2D layer would filter partial widths and cut words into pieces. This layer consists of a set of filters. The filters take a subset of the input data at a time, but are applied across the full input. The operations performed by this layer are still linear or matrix multiplications. This process is done repeatedly for kernel sizes 2, 3, 4, 5 and 6 respectively. We have experimented with various different kernel sizes. The output from the aforementioned kernel sizes were the expected output with decent accuracy. Throughout our experiment we have tried different filters, eventually we settled with a filter number of 5. All of the 1D convolutional layers of the 5 different kernels were connected to the embedding layer. Relu activation was used in this layer. The outputs from each of the convolutional layers having the aforementioned filters are then connected with a 1D global maxpool layer. This gives us a set of five convolutional layers individually passing their output to 5 global maxpool layers. The set of 5 convolutional layers connected to 5 global maxpool layers are then concatenated with an axis of positive one. Then the first dropout layer was added to reduce overfitting. This layer had a dropout of 0.1 which was initially set at 0.5 but it was found that the previous one provided better accuracy. This helped us to maintain proper convergence. The next layer added was the dense layer. Each neuron receives input from all the neurons in the previous layer. This layer will provide an output arrays of shape (*,128). Relu activation has also been used in this layer. This was followed by another dropout layer which had a dropout of 0.1. Finally, the last layer was added to the model which was a dense layer. This layer took a sigmoid activation as an argument. The output from this layer is an array of shape (*,6).

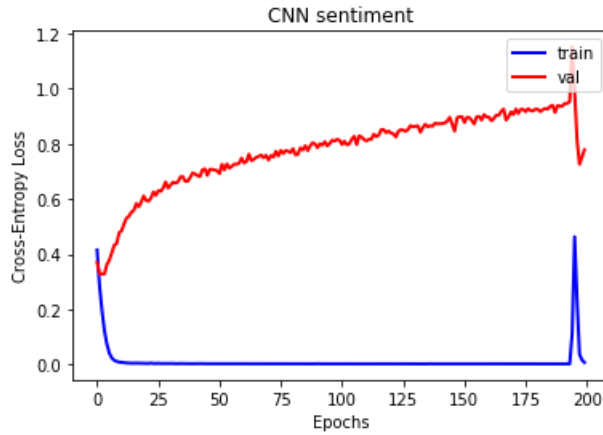


Figure 4.15: Training graph of Cross-Entropy Loss against Epochs

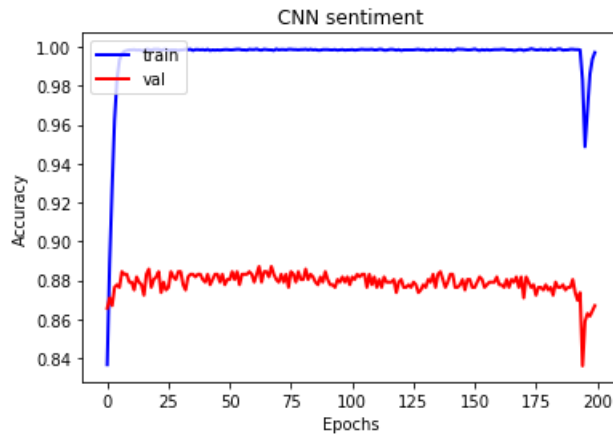


Figure 4.16: Training graph of Accuracy against Epochs

vii. Model Selection: After training four different models, it was time to select the one which would be used for sentimental analysis. All the models were trained with an epoch number of 200 which can be seen from the training graphs. This was done so that valid comparison could be drawn easily and effectively. The comparison was made by looking at the accuracy of the models. The one with the highest accuracy has been chosen and put into the final model.

Models	Accuracy in %
LSTM	77
LSTMon pre-trained word2vec	80
GRU	79
Yoon Kim's CNN	83

Table 4.3: Accuracy of different models of sentiment analysis

From the table it was concluded that Yoon Kim's CNN would be used in the final model since it has higher accuracy than the rest.

4.5 CNN for Facial Expression Classification

4.5.1 CNN

Convolutional Neural Networks extract specific features of objects given to it, unlike classical Machine Learning Algorithms. It does so using the convolution operation, after multiple such operation is carried out enough times, the neural network learns how to differentiate the images it has been trained upon. We have trained our CNN using the Japanese Female emotion dataset[54], Extended Cohn-Kanade Dataset (CK+)[55], Multimedia Understanding Group (MUG) Dataset[40], USTC-NVIE spontaneous-based Dataset[56], Indian Movie Face Database (IMFDB)[57], Acted Facial Expressions in the Wild Database (AFEW)[58], using the labels ang “angry”, “disgust”, “happy”, “sad”, “surprise”, “neutral”. There are a total of 28,709+x1 images for training and 3589+x2 number of test images.



Figure 4.17: Samples of the emotion dataset from the interview

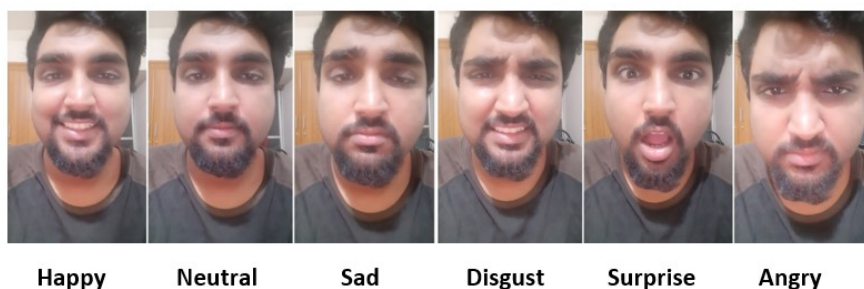


Figure 4.18: Samples of Different Entities of Emotion

4.5.2 Pre-Processing

We have used datasets of images of varying size, so using [specific photo editor] we downsized the JAFFE dataset to ensure a single homogenous dataset we have saved all the JAFFE images in csv format using python.

4.5.3 Training Object Classifier

The purpose of our experiment is to determine the employee satisfaction level using neural network. We have chosen the VGG16 Neural network architecture to train on the many facial expression dataset. After training for 6 hours on an NVidia GTX1060 we got around 88% accuracy. Then using Yoon Kim's CNN architecture, we have trained [sentiment dataset] to detect text sentiment with an accuracy of 83%.

We have chosen the VGG16 architecture(fig-4.19) to train a bunch of datasets(elaborate) with 7 different emotion("angry", "disgust", "happy", "sad", "surprise", "neutral").

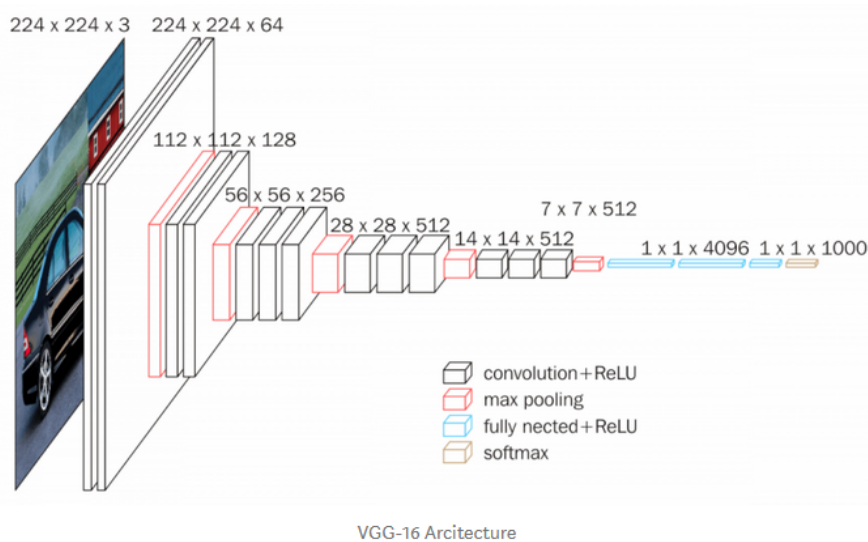


Figure 4.19: VGG-16 Architecture

Chapter 5

Results

The data-set of the interviews was split into train and test set. The data are then put into the different models. Here, the original intention was to feed the data to the feed forward neural network. But to achieve a better result by having a better accuracy, the data was fed to a machine learning algorithm, XG Boost, so that the accuracy can be compared and the one with the better accuracy can be selected. The accuracy indicated the performance of the model. This section of the chapter will discuss the accuracy of our proposed model.

The output for the input video should be one of the five answers which can be derived from our Likert scale. Here the scale varies from 1, which means highly dissatisfied, to 5, which means highly satisfied.

5.1 Result of Neural Network

The accuracy indicated the performance of the model. This section of the chapter will discuss the accuracy of our proposed model. The output for the input video should be one of the five answers which can be derived from our Likert scale. Here the scale varies from 1, which means highly dissatisfied, to 5, which means highly satisfied. The result is then compared with the overall satisfaction on the manual Likert scale which the interviewees have given during each interview. The comparison would determine the accuracy of the model.

While we did receive a high accuracy of sentiment analysis and facial expression separately, our overall detection of employee job satisfaction had a comparatively lower accuracy. This could be because the interviewees whose interview we took gave contradictory facial expression compared to what they said. Many people are shy in nature and so tried to say something whereas their expression showed something else, perhaps completely the opposite.

which were provided as input in the third neural network had contradictory results from the first two neural networks. We have used a feed forward Neural Network(fig-sthN1) to train the data and after 20 epochs we got a validation accuracy of 60% (fig-sthN2) after 20 epochs, giving 15 of our data for validation.

Layer (type)	Output Shape	Param #
dense_96 (Dense)	(None, 8, 110)	880
dense_97 (Dense)	(None, 8, 150)	16650
dense_98 (Dense)	(None, 8, 300)	45300
dense_99 (Dense)	(None, 8, 500)	150500
dense_100 (Dense)	(None, 8, 1000)	501000
dense_101 (Dense)	(None, 8, 1500)	1501500
dense_102 (Dense)	(None, 8, 2500)	3752500
dense_103 (Dense)	(None, 8, 4500)	11254500
dense_104 (Dense)	(None, 8, 1000)	4501000
flatten_38 (Flatten)	(None, 8000)	0
dense_105 (Dense)	(None, 5)	40005

Total params: 21,763,835
 Trainable params: 21,763,835
 Non-trainable params: 0

Figure 5.1: Feed-Forward Neural Network

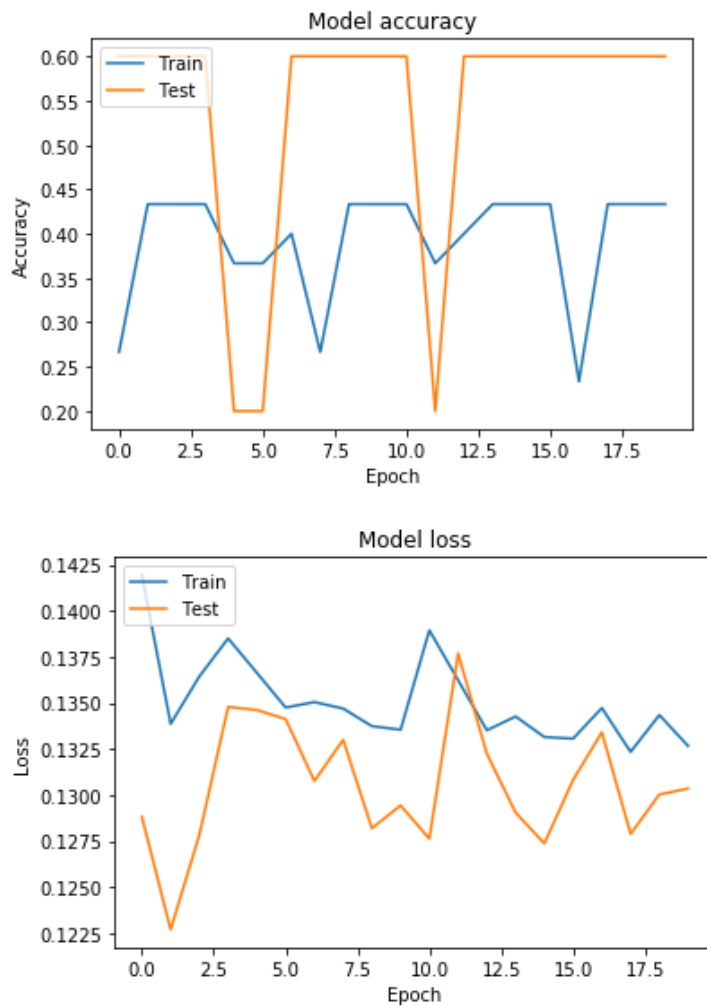


Figure 5.2: Training Graph Of Our Feed Forward Neural Network

5.2 Result of XG-Boost

We have also implemented XG-Boost on the data that came from Facial Expression Recognition and sentiment analysis, getting an accuracy of 72%. With a `max_depth=2`, `n_estimators=10`, `min_child_weight=5`, `learning_rate=0.01`, `colsample_bylevel=.4`, `colsample_bytree=.5`. Keeping the `max_depth`, `sub_sample`, `colsample_bylevel` and `colsample_bytree` low helps in avoiding over-fitting, `min_child_weight` and `lamda` on the other hand helps in regularizing, the learning rate rate doesn't cause it to overestimate but having one that is too low can increase the time required for learning. Fig-sthN3 shows some of the trees formed by the XG-Boost.

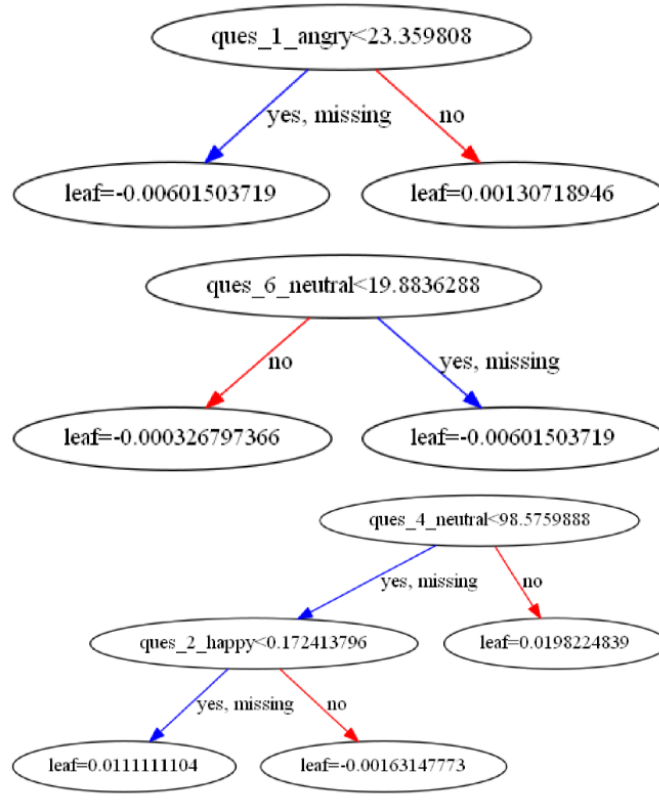


Figure 5.3: Training Graph Of Our Feed Forward Neural Network

5.3 Result of Support Vector Machine

State Vector Machines work surprisingly with high dimension data, our data having a moderate amount of dimension saw the highest accuracy, with a gamma of 0.01 our cross validation accuracy came to an 82.3% and validation accuracy of 85%, fig-sthN5 shows the confusion matrix of the validation.

```
[[ 1  0  0  0  0]
 [ 0  2  0  0  0]
 [ 0  0 10  3  0]
 [ 0  0  1 18  0]
 [ 0  0  0  2  3]]
accuracy: 0.85
```

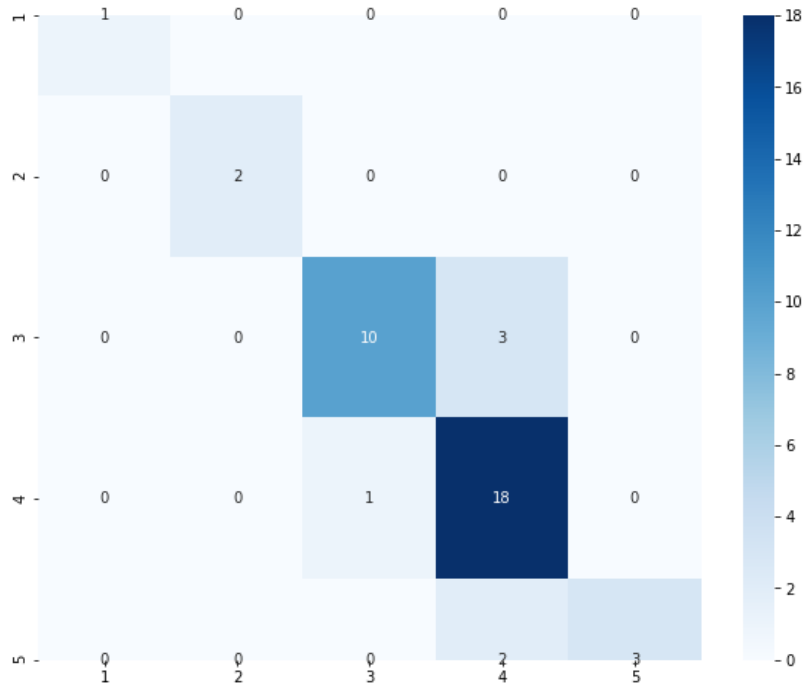


Figure 5.4: Training Graph Of Our Feed Forward Neural Network

Chapter 6

Conclusion

6.1 Conclusion

Despite the fact that emotion detection is a very easy process, the research had to face a lot of challenges because it had to take into account the availability and the willingness of the employees who are the major part of this research, as it was completely dependent on the interviews which were being conducted.

The first challenge which we faced was deciding proper places to conduct the interviews. This was because most of the places in the city are crowded which creates problems for the interviewees, as the interviewees were shy, busy and wanted to keep their opinion confidential. It was immensely hard to get female employees as interviewees. Female employees had various obligations such as returning home on time, not spending time outside the office, etc. However, we tried to focus on female employees because throughout the research we were receiving different reports which said female employees faced different problems such as harassment, discriminations, etc. The next challenge was finding a proper way to initiate the interviews. Since the interviews were conducted on people from various backgrounds it was very important to make them feel at ease to extract the proper answers to the questions. And because of that we could not jump right into the interview part, rather had to be patient and understanding, spending a good amount of time for each interview. The other problem which we faced was gaining the trust of the employees as some of them thought their information might get leaked which would affect them professionally. The next challenge was the stage fright or camera fright of the interviewees as some of them were nervous hearing about the interview and few were not comfortable being recorded over the camera. But the biggest challenge was the authenticity of the interviews, as we don't know if the information provided was true or not. However, we tried to extract the authentic data through our interviews. Again, creating the best neural network architecture to get the most accurate result in each of the two independent neural networks was also difficult. Moreover, the output from the neural network had to match with the Likert Scale. The scale ranges from 1 to 5. The algorithm selection was very important because without proper algorithms the result will not be satisfactory. And it was very difficult as there are different algorithms of machine learning and neural network and finding out the best model by trial and error method has been too time consuming.

Our major limitation was reaching the mass people who employees. This was partly because most of the employees are discouraged to speak regarding their office and

even when they do at times the truth is being fabricated.

6.2 Future Work

Through this research we have established that using Artificial Neural Network and Machine Learning, employee satisfaction can be detected. There are a lot of scope for improvements since this is perhaps the only research for employee satisfaction using neural network and machine learning. Our goal in the future would be to modify the questionnaire and to upgrade into a better one which can be used to get better understanding of the job and employee mental health. Further, we would like to conduct more interviews to increase the number of data in our datasets with the existing questionnaire. This would help us to achieve a better accuracy. Again, in this research, a simple feed forward neural network has been used which can be further optimized. We would try to use complex neural networks such as convolutional neural network and recurrent neural network to reach a better performance by the model. Moreover, other algorithms for the final comparison can also be done, since there are a lot of other algorithms which are being intended for better accuracy. Furthermore, this research was more generalized since interviews were conducted of different people from different workplaces, as it was mentioned above. No particular industry or workplace was highlighted here. For better research the industries can be segmented and interviews and questionnaire can be designed according to the industry. This would provide better understanding of employee mental health at a particular industry.

6.3 Conclusion

Emotion detection for employee satisfaction is a major application in the real world since mental health of employees at workplaces is a key factor. This paper has thrown light upon how modern technology can be used for the study of this aspect. Machine Learning and Neural Networks which have provided answers to various other questions can now be used to answer “How satisfied an employee is at his or her respective workplace.” This will help us to have better understanding of human psychology and can be used for better productivity.

The benefits from this research will surely improve the working environment as well as other aspects which are affecting the mental health of employees all over the world. Since this research is the first of its kind, it will give rise to many more of such research which will be used for several necessary improvements.

Bibliography

- [1] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences”, *arXiv preprint arXiv:1404.2188*, 2014.
- [2] C. Dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts”, in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.
- [3] T. H. Nguyen and R. Grishman, “Relation extraction: Perspective from convolutional neural networks”, in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015, pp. 39–48.
- [4] N. Aifanti, C. Papachristou, and A. Delopoulos, “The mug facial expression database”, in *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, IEEE, 2010, pp. 1–4.
- [5] Y. Kim, “Convolutional neural networks for sentence classification”, *arXiv preprint arXiv:1408.5882*, 2014.
- [6] A. Schmidt, “A modular neural network architecture with additional generalization abilities for high dimensional input vectors”, *Manchester Metropolitan University, Department of Computing*, 1996.
- [7] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, “Differential entropy feature for eeg-based emotion classification”, in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, IEEE, 2013, pp. 81–84.
- [8] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. A. Mahjoub, and C. Cleder, “Automatic speech emotion recognition using machine learning”, in *Social Media and Machine Learning*, IntechOpen, 2019.
- [9] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, “Machine learning techniques for sentiment analysis: A review”, *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, p. 27, 2017.
- [10] B. Zhang, C. Quan, and F. Ren, “Study on cnn in the recognition of emotion in audio and images”, in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, IEEE, 2016, pp. 1–5.
- [11] G. Harnois, P. Gabriel, W. H. Organization, *et al.*, *Mental health and work: Impact, issues and good practices*, WHO/MSD/MPS/00.2. World Health Organization, 2000.
- [12] S. der Kinderen and S. N. Khapova, “Positive psychological well-being at work: The role of eudaimonia”, *The Palgrave Handbook of Workplace Well-Being*, pp. 1–28, 2020.

- [13] Y.-c. I. Chang, “Boosting svm classifiers with logistic regression”, *See www.stat.sinica.edu.tw/library/c_tec_rep/2003-03.pdf*, 2003.
- [14] N. Cristianini, J. Shawe-Taylor, *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [15] T. Joachims, “Making large-scale support vector machine learning practical, advances in kernel methods”, *Support vector learning*, 1999.
- [16] S. Kapur, *Computer Vision with Python 3*. Packt Publishing Ltd, 2017.
- [17] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, IEEE, vol. 1, 2001, pp. I–I.
- [18] *Face detection using haar cascades*. [Online]. Available: https://docs.opencv.org/3.3.0/d7/d8b/tutorial_py_face_detection.html.
- [19] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting”, in *European conference on computational learning theory*, Springer, 1995, pp. 23–37.
- [20] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.”, *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [22] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex”, *The Journal of physiology*, vol. 195, no. 1, pp. 215–243, 1968.
- [23] *Convolutional neural networks (lenet)*. [Online]. Available: <http://deeplearning.net/tutorial/lenet.html>.
- [24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors”, *nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [25] J. Turian, L. Ratinov, and Y. Bengio, “Word representations: A simple and general method for semi-supervised learning”, in *Proceedings of the 48th annual meeting of the association for computational linguistics*, Association for Computational Linguistics, 2010, pp. 384–394.
- [26] C. Heipke and F. Rottensteiner, “Deep learning for geometric and semantic tasks in photogrammetry and remote sensing”, *Geo-spatial Information Science*, vol. 23, no. 1, pp. 10–19, 2020. DOI: 10.1080/10095020.2020.1718003. eprint: <https://doi.org/10.1080/10095020.2020.1718003>. [Online]. Available: <https://doi.org/10.1080/10095020.2020.1718003>.
- [27] I. Goodfellow, Y. Bengio, and A. Courville, “Deep learning. book in preparation for mit press”, *URLj http://www.deeplearningbook.org*, vol. 1, 2016.
- [28] G. Surma, *Image classifier*, Jan. 2019. [Online]. Available: <https://towardsdatascience.com/image-classifier-cats-vs-dogs-with-convolutional-neural-networks-cnns-and-google-colabs-4e9af21ae7a8>.

- [29] I. Aniemeka, *A friendly introduction to convolutional neural networks*, Aug. 2017. [Online]. Available: <https://hashrocket.com/blog/posts/a-friendly-introduction-to-convolutional-neural-networks>.
- [30] H. Yar, T. Jan, A. Hussain, and S. Din, *Real-time facial emotion recognition and gender classification for human robot interaction using cnn*.
- [31] G. AinomugishaGerald, G. Ainomugisha, and Gerald, *Why employee mental health should be every manager's concern*, Jan. 2018. [Online]. Available: <https://inside.6q.io/employee-mental-health-managers-concern/?fbclid=IwAR2R9u5hXEyYLkFbIrE5fkZ3ABBQbk1WHFGHzLXRuWipTT5NB.TizL20AM0>.
- [32] J. Parkinson, "Measuring positive mental health: Developing a new scale", *NHS Health Scotland, Glasgow*, 2006.
- [33] J. Schmidhuber and S. Hochreiter, "Long short-term memory", *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [34] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks", in *International conference on machine learning*, 2015, pp. 2067–2075.
- [35] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling", *arXiv preprint arXiv:1412.3555*, 2014.
- [36] A. N. Shewalkar, "Comparison of rnn, lstm and gru on speech recognition data", 2018.
- [37] R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (gru) neural networks", in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, IEEE, 2017, pp. 1597–1600.
- [38] Tqchen, *Tqchen/xgboost*, Jul. 2018. [Online]. Available: <https://github.com/tqchen/xgboost>.
- [39] D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao, *et al.*, "Relation classification via convolutional deep neural network", 2014.
- [40] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference", *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682–691, 2010.
- [41] G. R. Slep and D. A. Vella-Brodrick, "Optimising employee mental health: The relationship between intrinsic need satisfaction, job crafting, and employee well-being", *Journal of Happiness Studies*, vol. 15, no. 4, pp. 957–977, 2014.
- [42] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression", in *2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, IEEE, 2010, pp. 94–101.
- [43] K. Matzler and B. Renzl, "The relationship between interpersonal trust, employee satisfaction, and employee loyalty", *Total quality management and business excellence*, vol. 17, no. 10, pp. 1261–1271, 2006.
- [44] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks", *arXiv preprint arXiv:1412.1058*, 2014.

- [45] —, “Semi-supervised convolutional neural networks for text categorization via region embedding”, in *Advances in neural information processing systems*, 2015, pp. 919–927.
- [46] X. Zhang and Y. LeCun, “Text understanding from scratch”, *arXiv preprint arXiv:1502.01710*, 2015.
- [47] P. Wang, J. Xu, B. Xu, C. Liu, H. Zhang, F. Wang, and H. Hao, “Semantic clustering and convolutional neural network for short text categorization”, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, pp. 352–357.
- [48] Y. Chen, J. Yuan, Q. You, and J. Luo, “Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm”, in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 117–125.
- [49] A. Sherstinsky, “Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network”, *arXiv preprint arXiv:1808.03314*, 2018.
- [50] J. Weston, S. Chopra, and K. Adams, “# tagspace: Semantic embeddings from hashtags”, in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1822–1827.
- [51] *Google code archive - long-term storage for google code project hosting*. [Online]. Available: <https://code.google.com/archive/p/word2vec/>.
- [52] S. Chen, J. Wen, and R. Zhang, “Gru-rnn based question answering over knowledge base”, in *China Conference on Knowledge Graph and Semantic Computing*, Springer, 2016, pp. 80–91.
- [53] Y. Zhang and B. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification”, *arXiv preprint arXiv:1510.03820*, 2015.
- [54] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets”, in *Proceedings Third IEEE international conference on automatic face and gesture recognition*, IEEE, 1998, pp. 200–205.
- [55] S. Setty, M. Husain, P. Beham, J. Gudavalli, M. Kandasamy, R. Vaddi, V. Hemadri, J. Karure, R. Raju, B. Rajan, *et al.*, “Indian movie face database: A benchmark for face recognition under wide variations”, in *2013 fourth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*, IEEE, 2013, pp. 1–5.
- [56] W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks”, *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.
- [57] S. Katsigiannis and N. Ramzan, “Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices”, *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.
- [58] P. R. Dachapally, “Facial emotion detection using convolutional neural networks and representational autoencoder units”, *arXiv preprint arXiv:1706.01509*, 2017.