EMOTION DETECTION FROM FACE USING MACHINE LEARNING

BY

ABIR AHMED
ID: 181-15-10579
AND

SOURAV KUMAR SUR
ID: 181-15-10911
AND

BORNALY SAHA
ID: 181-15-11082

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

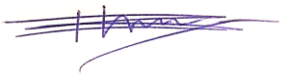
Ms. Nusrat Jahan
Sr. Lecturer
Department of CSE
Daffodil International University

APPROVAL

This Project titled **"Emotion Detection from Image Using Machine Learning"**, submitted by Abir Ahmed, ID No: 181-15-10579, Sourav Kumar Sur, ID No: 181-15-10911, and Bornaly Saha, ID No: 181-15-11082 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 6th January, 2022.

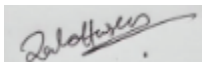
BOARD OF EXAMINERS

Chairman



Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Zahid Hasan (ZH)
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mohammad Monirul Islam (MMI)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Dewan Md. Farid
Professor
Department of Computer Science and Engineering
United International University

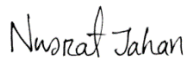
DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Zerin Nasrin Tumpa, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

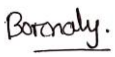


Ms. Nusrat Jahan
Sr. Lecturer
Department of CSE
Daffodil International University

Submitted by:

Abir Ahmed
ID: 181-15-10579
Department of CSE
Daffodil International University

Sourav Kumar Sur
ID: 181-15-10911
Department of CSE
Daffodil International University



Bornaly Saha
ID: 181-15-11082
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to **Zerin Nasrin Tumpa**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Professor Dr. Touhid Bhuiyan, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Human emotions are determined by one's action, behavior, voice tone, facial expression and body language. These are considered to be the channels to understand human emotion. Using modern technology, depth analysis has been allotted to enquire into the relationship between these emotions and the channels to understand human emotion. Human feeling recognition plays a vital role inside the social relationship. Automatic emotion recognition has been a vigorous analysis subject for quite a long period. Therefore, there unit some particular advances created throughout this range. Human emotion expresses from voice speech, voice tone, hand movement, through facial expressions and gestures of the body. Therefore, pulling out and classifying and fathoming of feeling options a great importance to the reciprocal action between the communication of human and machine. Our study intends to develop a system that mechanically identifies the feeling depicted on human face. Therefore, image process is employed to classify the universal emotions. These are happiness, sadness, anger, disgust, surprise, feared and neutral. The system takes human images of frontal face as its input. Feature purpose extraction methodology which is based on image process in this case, is employed to fetch a set of chosen feature characteristics, after the face is detected. And ultimately, when those fetched information points are given as input to the neural network in order to acknowledge the feeling ached, a set of values is generated as its output.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Feeling may be a physiological and psychological condition that's secular and individual. It embroils plenty of actions, behaviors, feelings and thoughts. Primary analysis assigned on emotions are traced in the book “The Expression of the Emotions in Man and Animals”, authored by naturalist. His believe was instead of plane-specific, emotions are species-specific [1]. Once identifying a catholicity between human emotions in various communities of individuals in spite the cultural variations, six universal emotional expressions were classified by “Ekman and Frisen”. These are happiness, sadness, disgust, surprise and concern, in 1969 [2, 1, 3, 4]. Facial expressions area unit typically thought-about not only entirely as a result of the foremost natural reasonably plotting human emotions but also as a numerical communication technique [5]. If economical ways in which area unit typically caused to mechanically acknowledge these facial expressions, hanging enhancements area unit typically achieved at intervals the house between machine and human intercommunication. Analyzing the facial emotion detection has been assigned in prospect of achieving these improvements [6]. Actually, various applications are available which can identify facial feeling automatically. In order to express the human emotions into robot truly, computing on the province of detecting facial emotions has been relied on for a long period. The researchers have been impressed by the modern enhancements throughout this area to augment the relevance of frontal expression detection to spaces like chatting area, making still avatars and live avatars for video conferencing. The capability to acknowledge facial feeling area unit is very important in face recognition applications also. Threat detection systems and intelligent

surveillance systems developed for teenagers with brain development disorders area unit another beneficiary [7].

1.2 Objective

Today's world is the world of modern science and technology. Technology has made our day-to-day life easy as we are using technology in every aspect of our life. Automatic emotion detection can help us build a better communication between human and machine. Our main objective is to classify different emotion state from facial image. So, we can describe our goals like this,

- Our objective is to study and learn the process in which we can identify different types of human emotion.
- To develop a system that can identify different human emotions from image input.
- To visualize some analytical analysis of human emotions classified by classifier algorithm.

1.3 Motivation

We were interested to try and do one thing totally different. So, we have a tendency to determined earlier that we are going to do analysis on computer science and machine learning field. So, we have a tendency to begin to hunt for some ideas. however, no plan could not satisfy us. Then we have a tendency to thought we will play a job in fighting against crime by forestalling or try and prevent it from happening. Now-a-days each vital place of our country is underneath police work camera that area unit observance any quite trespass activity 24/7. If we have a tendency to had a system that may determine the

emotions from external body part whether or not he/she is happy, sad, besieged or in concern, necessary steps would be taken. as a result of we have a tendency to don't recognize what hostile state of affairs is awaiting Us within the next number of minutes or moments. So, it's higher to be alert and cautious continually and be ready for any surprising events. From this thought, we have a tendency to get our plan. we have a tendency to finally found a subject that gave Us interest to review in additional. once we began to study on this specific field of analysis, we have a tendency to found that, automatic feeling detection system has far more blessings and that we will apply this technique in several alternative technological fields to boost them. Besides up the closed-circuit television, it also can be applied in health care [8]. It will facilitate to make a decision} once patients necessitate medication or to assist out physicians determine United Nations agency to check 1st. It also can be applied in automobile industry [8]. fashionable vehicles area unit automatic currently. So, automatic feeling detection will facilitate to forestall any quite dangerous state of affairs, dominant the speed or dynamic alternative parameters pro re nata by sensing the driver's feeling. Automatic feeling detection also can be utilized in testing video games [8]. Throughout the testing of a computer game, a player is given a hard and fast time to play the sport. His/her feedback is extremely vital for the sport developers to create the ultimate product. Automatic feeling detection of the players will facilitate the developers to observe the players period feelings whereas enjoying any specific section of the sport or to see the problem level and alternative necessary variables. These facts made us interested to try and do such a quite research on a basic work like this. Our work is totally connected with machine learning ways.

1.4 Rational of the study

Undoubtedly, there have been thousands of workshops done on image-processing before. In latterly a prolonged time, the convolutional systems are used to deal with the kind

bracket trip, still with clear datasets and approaches. Image- making ready could be a progress approach in which it'll be segmented into colorful orders. One amongst these is compression of the image and the another one is enhancement of the image. And thus, the final is that the recovery, and making hypothesis. It makes a distinction to drop the aggregate of data that's demanded to save a reused image into the memory space. The image will be abandoned. By digitization handle and by scarcities image will be lammed. The lammed image will be corrected exercising image upgrade procedures.

1.5 Expected Outcome

In this block, there are many agendas that are the main anticipated affair. The main ideal of this exploration- grounded design is to make a complete and effective procedure that can identify different mortal feelings from their face with image processing. Some expected outcome points are:

- Classify different human emotions
- Increase surveillance therefore security system
- Help to prevent upcoming hostile situations
- Reduce criminal activity
- Decrease crime rate
- Help in psychological treatment
- Help to treat children with special needs well

1.6 Research Questions

It turned into hence hard to complete the study and fulfil our purpose for us. So as to conquer a practical, economical, and precise reaction to the matter, we want to endorse the subsequent queries to express the intuitions and consequences of this drawback:

- Will we be predisposed to accumulate row photo facts for deep learning research?

- Is it doable to pre-technique the row facts victimization deep learning course?
- Is it possible to improve security system better than before with this new procedure?

1.7 Layout of the Report

In the first chapter, the objective, motivation, research questions, rationality and expected outcome have been discussed. In this section we will narrate the entire format of the report.

In the second chapter, we have discussed the previous works and researches which have been done on this particular topic or domain. The next part of this section shows their limitations of their research and how we studied and overcame their limitations with our research. And at last, the limitations that have been faced by us are discussed.

In chapter three, we have discussed the theoretical discussion related to our research area. To do so, the statistical methods of our work are described in this section. Moreover, the procedural approaches of Machine Learning classifier have been elaborated in this section. And at the last part of this chapter, confusion matrix is displayed to show the precision name of the classifier.

In chapter four, we have discussed the experimental result, evaluated the performance of the model and discussed on the output result. Images of a few experimental results are shown in this section for better understanding the model.

In chapter five, we have put the gist of our research, our future work on this field and concluded the research. At the last part of this section, we have discussed our limitations

which have appeared during our research. These limitations may help others or give scope to others who will do further research on this domain.

CHAPTER 2
BACKGROUND

2.1 Introduction

Here, linked works, evaluation define, and demanding situations concerning this evaluation is going to be discussed. inside the linked works section, we're going to speak about extraordinary evaluation papers and their studies, their strategies, and their exactness which can be applied into our study. inside the evaluation define section, we're going to offer a definition of our linked works. inside the demanding situations section, we're going to speak about but we have a tendency to increase the accuracy level.

2.2 Related Works

Lots of researchers have worked on this topic before. Previous works on this field can be largely classified in to three phases: detecting face, extracting facial features and classification of emotion.

In 2004, the Technical University of Munich published a review paper in which they classified six different human emotions [9]. In their system, they had a camera which records an ongoing video sequence. And later, the computer tried to match that recorded sequence with pre-recorded emotion sequence stored in their database. To match the facial expression, they extracted the coordinates of eyes, nose, mouth etc.

In the article published by Moon Hwan Kim, Young Hoon Joo and Jin Bae Park, they proposed a new algorithm for detecting the human emotion by dealing with frontal image [10]. Their new algorithm was composed of three stages. The stages were processing images, feature extraction and lastly emotion detection. These three Korean researchers also proposed a new algorithm to extract features from facial image which they believed to be worked better. These algorithms work on three different feature regions which are

eye, mouth and auxiliary region. With their new proposed methods, they were able to extract facial components from 124 images successfully [11]. They mainly worked on detecting five different human emotions which were happy, sad, angry, disgust and surprise. And their final emotion detection accuracy was 74.0%.

Four Researchers from Stockholm University worked with the six basic facial expressions in recent time. In 2013, they published a conference paper on automatic emotion recognition, analyzing facial expression based on artificial neural network [12]. They used images from the Cohn-Kanade database and merged them with their own manual dataset to train their model. They used 12 individuals who were chosen randomly. They were from different ethnic groups and gender. 8 of them were male and 4 were female. They also used multiple lighting condition while capturing the images of the model. 72 images were used in their research which were taken from the actors. Then they used these images with the merged ones from CK - database to train their model. Then they fed their input dataset in an ANN based model and tested the model. They were able to extract six emotional states of human. 58 test images out of 72 were successfully classified. The overall accuracy of their model was 80.6% [12]. Simulations were done on MATLAB platform for better results. Their research concluded that human feelings are mainly portraited by eyes and facial expression. So, only these two portions of human face had been considered for the emotion detection process. But other parts of human face can also express facial expression. That information was neglected in their studies.

In a very recent days, four Australian researchers, Lu chen, Tom Gedeon, Md Zakir Hossain and Sabrina Caldwell performed research only on anger. In 2017, they published a paper named "Are you really angry? Detecting emotion veracity as a proposed tool for interaction". According to them, anger can be detected by monitoring the pupil of human eyes [13]. The main goal of this study was to test the human ability to detect the veracity of anger in both of consciously and non-consciously, on the basis of their pupillary

response. Their main challenge was to collect the dataset. Because there was no benchmark dataset related to their work. So, they had to recruit 22 participants who were university students and record their pupillary response on various anger situations. There were 7 male and 15 female, among whom 18 were westerners and 4 were east Asians. All of them had valid pupillary responses. The participants sat on a chair, in front of a computer device with addition to an eye tribe that that measure the pupillary response. The instructor then gave oral instruction to the participants in various way. Some video sequences were also viewed to them. Participants performed an eye fixed gaze activity on increase the quality of the gaze pursuit knowledge recording, and were asked to perform the experiment on the pc. On completion of the computer tasks, they were in short interviewed by the experimenter on however they consciously known real or acted anger in their answers to those queries throughout the experiment. The average accuracy of their generated result was 60%. Their study shows that acted anger definitely differs from actual anger and it is surely be detected by machine. But they could not use benchmark dataset in their study. If they did, they could have generated a better result.

According to an IEEE journal published at 2006 published by Pritam pal, Ananth N. Lyer and Robert E. Yantorno it is possible to detect the reason for crying of a baby [14]. They worked with image processing as well as sound processing. And build a fusion system to detect the reason for cry. Infant Images and cry sound was captured for same cry event. In the image processing section, the model focused on eyebrows and mouths position. And in the sound processing section, K-means clustering was used to determine the reason of the cry. The decision level fusion system was able to detect six different reasons of baby cry: sad, anger, hunger, pain and fear. The result was based on the audio frequency and the position of eyebrows and mouth. The average accuracy of the image processing model was 64% and audio processing model was 74.2%. The fused decision was 75.2% accurate according to their study. They faced many issues while making the logic for the decision fusion unit. For example, for both "pain" and "fear", the eyes are

closed. So, it is truly difficult to distinguish the actual reason of crying of a baby in some cases.

Table 2.2.1: Previous works on related domain

Paper Name	Authors	Methodology Used	Accuracy
FGnet - Facial Expression and Emotion Database	Frank Wallhoff	Coordinate sequencing of facial components	Not mentioned
Emotion Detection Algorithm Using Frontal Face Image	Moon Hwan Kim, Young Hoon Joo and Jin Bae Park	Histogram segmentation algorithm and fuzzy classifier	74.0 %
Automatic emotion recognition through facial expression analysis in merged images based on an Artificial Neural Network	Javier G. Razuri, David Sundgren, Rahim Rahmani, Antonio Moran Cardenas	Artificial Neural Network	80.6 %
Are you really angry? Detecting emotion veracity as a proposed tool for interaction	Lu chen, Tom Gedeon, Md Zakir Hossain and Sabrina Caldwell	pupillary response	60.0 %
Emotion detection from infant facial expression and cries	pritam pal, Ananth N. Lyer and Robert E. Yantorno	Neural network and K-means clustering	75.2 %

2.3 Research Summary

CNN or convolution neural network is a type of "Artificial Neural Network", in the field of deep learning [15]. The most common uses of this methodology are to analyze image data. It is one of the basic systems of machine learning (ML). Convolutional networks square measure a specialized form of neural networks that use convolution in situ of general matrix operation in a minimum of one amongst their layers. Basically, a CNN is consisted of three layers: an input layer, hidden layers and an output layer. The creator of CNN is Yann Lecun. He was inspired by the structure and procedure of human brain. It is true that, CNN really works like a human visual cortex. This method really is a groundbreaking model in the field of image processing. In CNN we don't have to include assurance, however in other image classification and computations, we have to do it manually. There are bunch of filters that are used on CNN. These are also known as layers. In convolution layer, a moving bit is mixed up which goes through the input image. The bit goes through a 2D portrayal of the image and choose a tangible package and engages bit duplication and gathers it in a different grate.

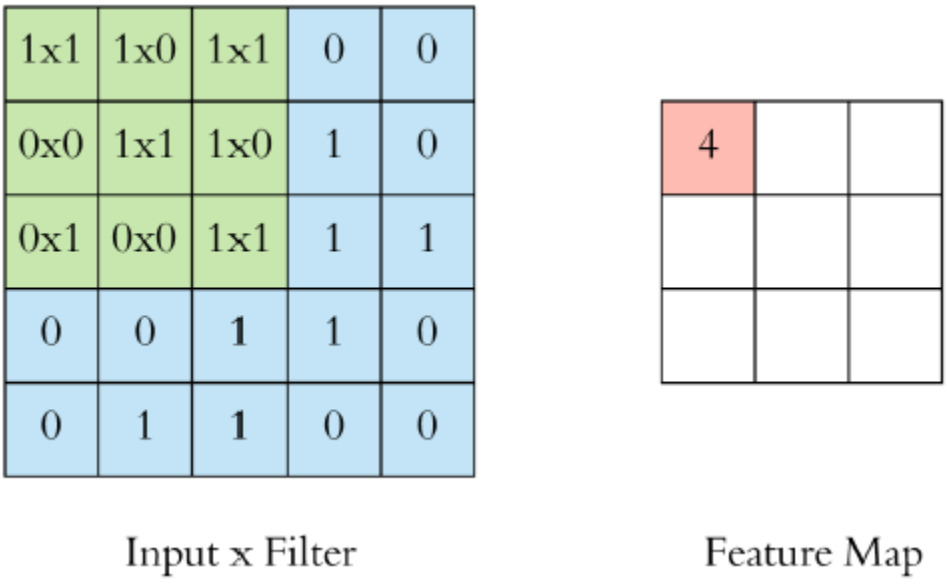


Figure 2.3.1: Convolution of filter on 2D image [16]

By using a condition, we can generate the output matrix. We will map an equation here. Here,

Output - Output dimension
Input - Input dimension
Size - Window size
Stride - Stride

$$\text{Output} = \text{floor}\left(\frac{input-size}{stride}\right) + 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(I)}$$

This condition is used to find Output of the measurement.

The most important layer in the convolution layers is the pooling layer. It helps to decrease memory consumption. As a result, a quick calculation can be done. I also decrease the volume. The pooling layer we have used in our research is the average-pooling layer. Because we have used the DenseNet CNN. And in DenseNet CNN, we used max-pooling layers instead of average-pooling [17](see Appendix A). The reason behind this is, max pooling can better represent the maximum strength of a feature by passing gradients through all indices, wherein average pooling, the gradient only flows through the average index. Max pooling is helps to fetch a smooth feature from an image. This method focuses on the bright pixels of an image. This is useful when the image background is black. And as we have used grayscale images to train our model, we used max-pooling technique.

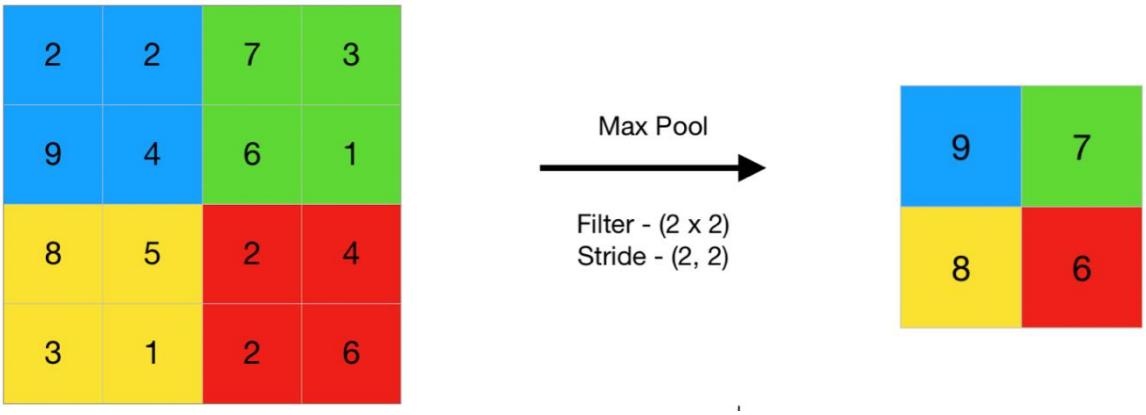


Figure 2.3.2: Max pooling [18]

Output generated from previous layers in form of 2D and 3D arrays are taken as input in a fully connected layer and then the arrays are converted into 1D array for further operation.

2.4 Challenges

The principal demanding situations of this research are gathering and mannering the dataset. Mannering the records set became too steep for us. To make it easy and generalize, we used some tricks and techniques. After these steps, we trained our machine with many layers and diacritic guises of epoch took a huge time for our model in total compilation, so we had to wait for a long time to get the very last output, maintaining patience. It took nearly 20 hours for our computer to compile and generate final output from our combined dataset. Moreover, we were under covid situation when we started our research study. So, we faced a lot of problems while collecting our primary data. Though several works on this topic have done before, we faced a lot of challenges to overcome the limitations of the researchers who have done their work before. Besides, collecting the necessary dataset that is suitable for our research was difficult. We got a very little help from the works that are done previously. So, we had to begin from our motivation.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

An explanation for the workflow of our novel technique to categorize seven specific human feelings routinely via machines has been attempted to provide here. There lay some important key factors such as collecting the data, processing them, proposed version to explain them along with the vital condition, chart, table, and description. The claim-created CNN-primarily based totally display related and our own dataset become applied on this work. The section is being ended via way of means of giving an evidence of our project's measurable theories and giving a clean idea of the execution requirements.

3.2 Research Subject and Instrumentation

In this study, the topic was inquired and studied which was reviewed for getting a clear concept on the research domain. It was not for implementing only, also to design the model, for collecting and processing the data, and training the model. In addition, instruments that are used are also discussed in this section. We used the "Windows" operating system, python programming language with a few necessary packages. We used NumPy, pandas, skit learn, matplotlib, etc libraries. The Anaconda navigator and JupyterLab were used for all training and testing processes. These are free and open-source platforms for developing python programming language-based projects which are used in information science and artificial intelligence applications.

3.3 Workflow

This study work has a sequence of workflow. Among them collecting data, processing them, prepare the data by applying data resize and augmentation, selecting model for evaluation etc.

Leap 1 – Collecting Data: Working data was collecting by clicking images of different people, with different emotions. To enrich our dataset, we also collected some images from website. Then we processed those data to recreate a combined data-set. It was a matter of tough challenge to collect and make a data set with these collected images so that we can feed them into our model, because we had to find the suitable data that can be used to fulfill our purpose.

Leap 2 – Processing Data: We prepared all information, classified those images after collecting them from different sources. There were noise and errors in many of the images. Processing those images manually was our first priority. We converted the raw images into grayscale. After that we implemented the combined dataset to the following leap.

Leap 3 – Resizing Data and Data Augmentation: Data have been augmented and resized after processing them class by class. All of the raw images were resized into 48 x 48 dimension before putting them in the dataset. For the purpose of training, we had to go through resize and data augmentation. Some overfitting occurs when augmented data is used in a dataset, that's why we have done only a few and those which were necessary.

Leap 4 – Selecting Model: To teach and validate our information for higher accuracy, we pick out a few models. There are many convolutional neural network models. To get higher accuracy with our device configuration, we ran some models in our system but there was no single model that could satisfy our purpose. So, we began from the sketch. We observed other machine learning models and recreated some models on our own,

tested them separately, observed their performance and eventually one version became suitable for the final process of training the model and testing it with our dataset.

Leap 5 – Evaluating the model: The outputs are represented with some graphs here in the section. After training the model and testing it, those processes gave us a few graph outputs with accuracy and loss. The graphs represented both training and validation accuracy and losses. The confusion matrix has also been calculated. And we have also calculated the precision, recall and f1 measure for the model's performance.

Leap 6 - Conclusion and Future Work: Here, we concluded the work with a future work plan.

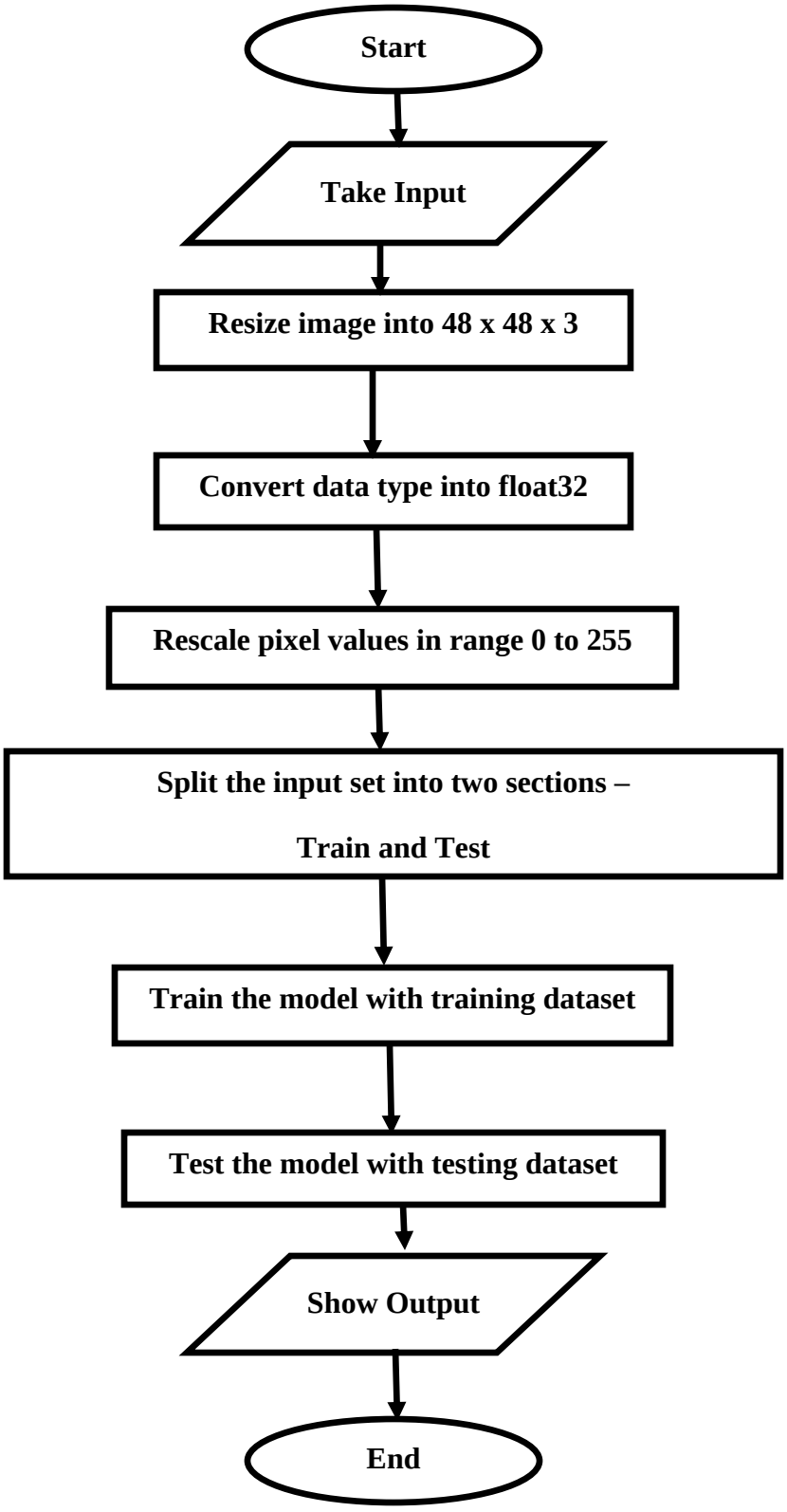


Figure 3.3.1: Workflow of the proposed approach

3.4 Data Collection Procedure

For our experiment, we created a dataset containing seven different human emotions. At first, we collected some primary images from our known people and friends. Their different emotions were captured. We were intended to collect seven kinds of emotions for our research study. But as these data were not enough for training a machine for evaluating "Artificial Intelligence", we have used a dataset from a public database [19]. The dataset included images of six different types of human emotions. The dataset includes more than 40 thousand of images. We have picked up 31,345 images from this dataset which were suitable for our work. These images contain six different facial expressions of human. These are happiness, sadness, anger, surprise, disgust, and fear. We searched in the internet for images of neutral emotion and included some suitable ones for our dataset. And later, augmented them for enlarging the dataset.

For our research purpose, we have divided our whole dataset images of seven different emotions (see Appendix B) into two sections: training and testing datasets. We have taken 30882 images from our modified dataset for training our model and 7775 images for testing the model (see Appendix C).

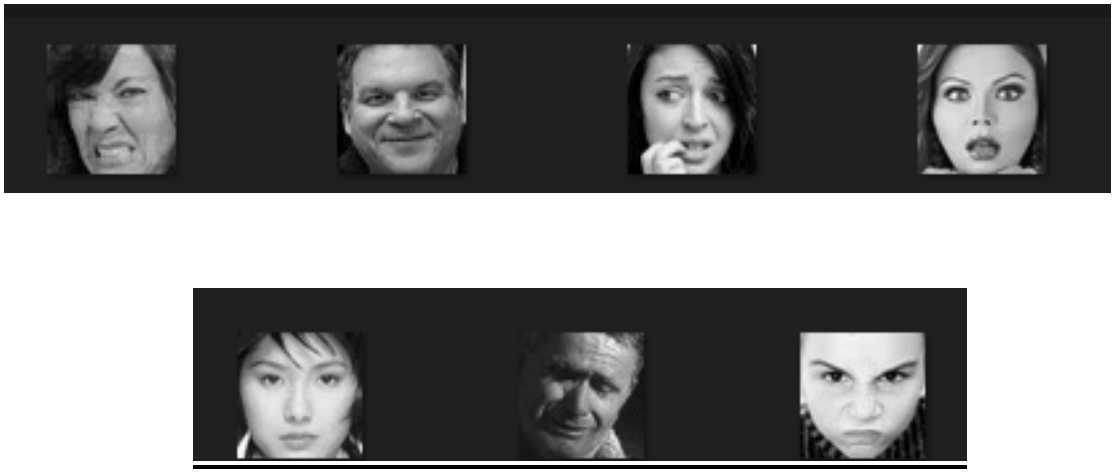


Figure 3.4.1: Images of different types of emotions from our dataset

3.5 Data Processing

The system of data process has been done in two phases. The first phase is augmenting the data and second phase is scheming the data. After the model was fed with the training dataset, generally the final result depends on the preprocessed images. The more effectively the images are pre-processed, the accurate the result will be. In a word, this is the challenge of first stage for such kind of research study about works which are based on machine learning.

3.5.1 Data Augmentation

For testing our model, we used a dataset from an online database. But we also wanted to add some primary data in our dataset. So, we collected some images of different facial expressions from different people. But the amount of our collected data was not enough to train the model we proposed. So, to extend our primary dataset, we augmented our data. We used five data augmentation techniques to extend our primary dataset. The techniques are given here:

- Cropping: We cropped the images with random image widths and heights.
- Salt and Pepper noises: We applied salt and pepper noise technique to add noise in the images.
- Rotation: We rotated our primary images by 40 degrees in clockwise direction.
- Scaling: We scaled the images with two scaling factors: s = 0.2 and s = 1.20
- Translation: We shifted the images along with X and Y axis with random translation units.
- Flipping: We flipped the actual image horizontally.

3.5.2 Data Preparation

We included our primary images with our previous dataset that we collected from online database. Since our demonstration requires a settled pixel for all pictures, we resize our dataset into 48 x 48 pixels.

Moreover, we have converted the images into grayscale. As we didn't have better GPU in our computer used to prepare the model as well as the whole process. For this reason, in order to prepare the model, we have used grayscale images.

3.6 Proposed Model

We proposed a possess Controversial Neural Network demonstration which has 10 layers.

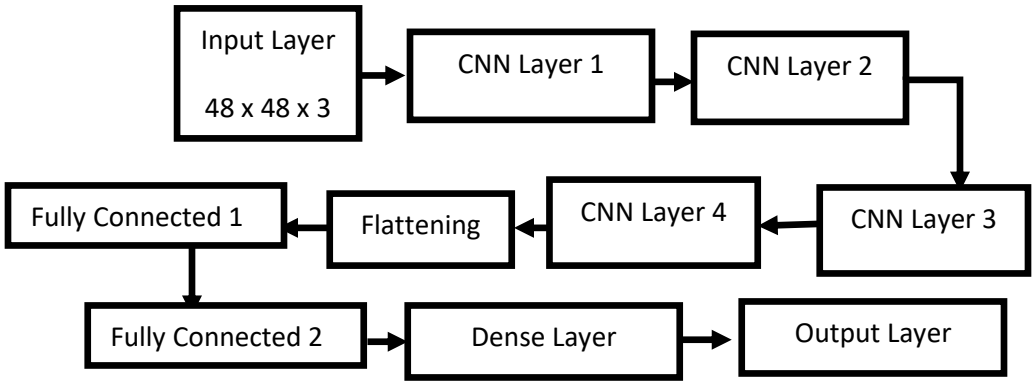


Figure 3.6.1: Proposed CNN Layer

There is an input layer of gauge 48 x 48 x 3. There are eight hidden layers. Among them four are convolution hidden layers, one layer for flattening and two fully connected layers. Each convolution hidden layer is build up with four layers for functionalities.

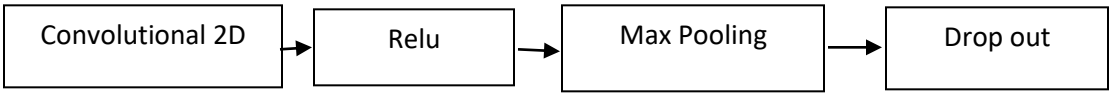


Figure 3.6.2: Internal Structure of Each CNN Layer

Moreover, there's the "Max pooling layer" within the network. As a result of max pooling will higher represent the strength of a feature by passing gradients through all indices. We should have a concise exchange about these layers' underneath.

3.6.1 Convolutional Layer

CNN's have a huge functionality and uses within the image and video acknowledgment, recommender frameworks, and traditional non-standard speech handling. Convolutional neural systems. seems like a peculiar combination of science and mathematics with a tiny low cesium wet in, however these systems are some of the foremost persuasive advancements among the sphere of pc vision [15]. 2012 was the first year that neural nets developed to conspicuousness as Alex Krizhevsky used them to win that year's ImageNet competition (essentially, the yearly Olympics of pc vision), dropping the classification mistake record from twenty sixth to fifteen, was a stunning advancement at the time. Once a PC takes an image as its input, it's going to see a cluster of constituent values. betting on the determination and estimate of the image, it's going to see a 32 x 32 x 3 cluster of numbers. These three numbers are the three attributes to the values of RGB. During a typical CNO (Convolution Neural Organize) style, there square measure alternative layers that square measure homogenized between these convolutional tiers. A classic CNN architecture looks like:

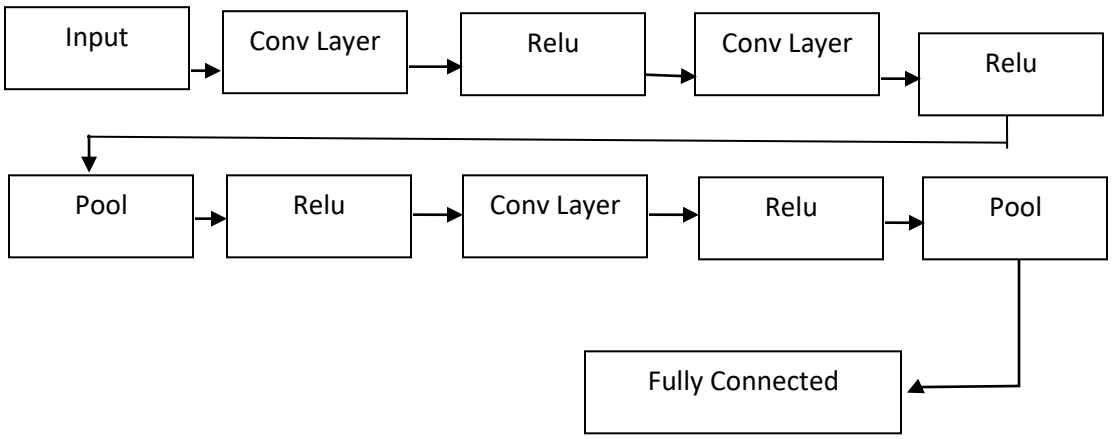


Figure 3.6.3: A Classic CNN Architecture

3.6.2 Feature extraction

One of the most common basic elements of CNN is convolution. In the dictionary of science, the word "convolution" refers to a combination of two capabilities to perform the third task. This process is combined with two flocks of knowledge. Inside the structure of CNN, the process of convolution is executed on the basis of an input data set. It generates a highlighted result output in a very short time period. Our target is to execute a convolution process by making glistening input data. At every step, the performance of a convolution network gets duplicated and sum up the results to produce a highlighted result.

3.6.3 Max Pooling

In Convolutional Neural Systems, max Pooling can be considered as a degenerating testing procedure. It's a pooling operation that considers the max price for patches of a feature map, and uses it to form a down sampled feature map. the target is to degenerate the sample associate input illustration. This process decreases its dimensionality and allow it for making highlighted assumptions contained at intervals the sub-territories

binned. It's basically turned to diminish the live from the image since the common pixel number offers associate overall info regarding the parameters. It adds a little quantity of translation changelessness - which means translating the image by a little quantity doesn't considerably have an effect on the values of most pooled outputs. during this method we have a tendency to need less parameters such that a CNN will still distinguish the image. In addition, max pooling technique is very useful if the images are in grayscale. It is because, max pooling method focuses on the bright pixels of a unit.

3.6.4 Dense Layer

The dense layer is truthful a different title of the fully connected layer. Comparable executions occur within the frizzed layer wherever every vegetative cell is related to one another. It deals with to a frizzed association of thick neurons. For this reason, it's to boot referred to as dense. A thick layer has great importance associated with every vegetative cell mix and with one in every of a sort value. Inside the neural prepare, various forms of work like SoftMax enactment work, SVM, and for high-level thinking other divertive area units are used here. However, for classification, we tend to adhere used SoftMax in our demonstration. We tend to get some good resolution highlights as input, when some convolutions and pooling layers are used. These input image highlights area unit used as classifying to analyze completely different categories. However, after we mix the convolution layer`s highlights and mensuration layer`s highlights it offers the method higher results of classifications. There is only a single yield chance in fully associated layers combination. One Convolution layer share values with different Convolution layers. It's exceptionally hard to hitch all hubs with a layer of SoftMax.

3.6.5 SoftMax

“Softmax” is a function which is used for activating the necessary functions in the output layer of neural network models. This process helps to predict a multinomial probability

distribution. In a word, “Softmax” is used as the activation function for multi-class classification problems.

Let us contemplate a classification model to classify with n categories. This model takes input datasets Associate in nursing an algorithmic program and produces a score of every category. The SoftMax activation operate converts from score to the likelihood between zero to one. The summation of all possibilities has one. We utilized this work to the final word layer of CNN systems to classify the categories. this work is delivered various lessons from Associate in Nursing input array [20]. the probability dissemination of SoftMax work is:

$$\sigma\ (x_j) = \frac{e^{xi}}{\sum_{j=1}^{n} e^{xj}}$$

Where i and j =1,2, 3……n

3.6.6 Dropout Layer

Dropout is also a way used to maneuver forward over-fit on neural systems. The data structure of a neural organize can be a coordinated chart wherever every hub speaks to Associate in Nursing inclination, whereas every edge speaks to a weight [22]. Just in case there are unity with predisposition b, and in edges e1, e2, e3 with weights w1, w2, w3, at that point, once the signals that comes into y on e1, e2, e3 are x1, x2, x3 severally, the yield of y is w1x1+w2x2+w3x3+b. As we tend to as of currently understand, a lot of profound the organize is, a lot of parameters it's. For illustration, VGGNet from ImageNet competition 2014, has a few 148 million parameters [21]. That’s apart. thereupon several parameters, the organize might effortlessly overfit, significantly with very little dataset. CNN should work within vigorous surroundings so dropout becomes basic. Dropout is basically chosen between 0.2 to 0.8. Dropout evacuates the neurons haphazardly supported the parameters given by the shopper like 0.4 etc.

3.6.7 Flattening Layer

Flattening refers to convert data in form of multi-dimensional array into one – dimensional array so that it can be imputed into the following layers of CNN. This layer is used to create a single array so that we can easily build our own CNN model for teaching the machine to perform AI based tasks.

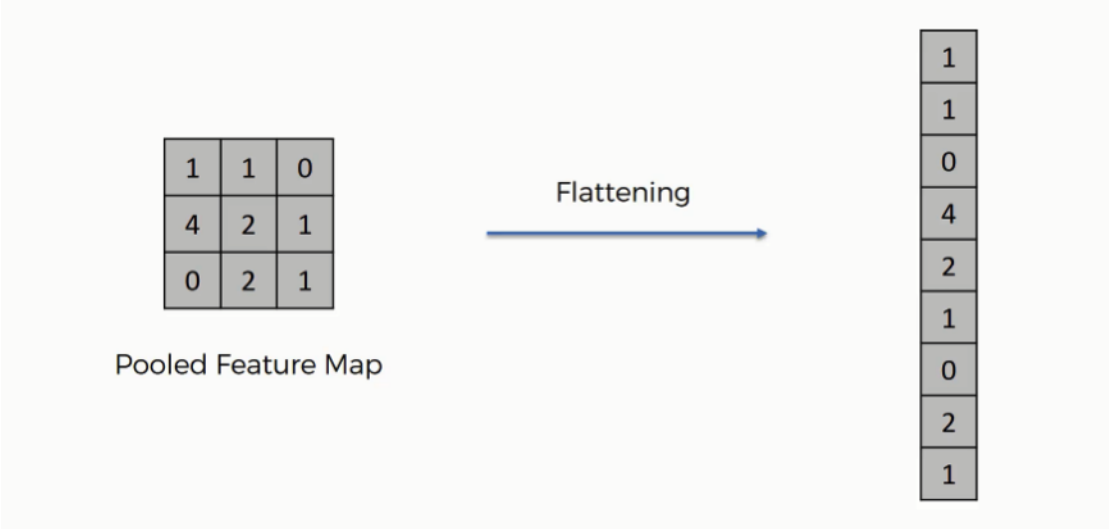


Figure 3.6.4: Data Flattening

3.7 Training the Model

We collected frontal face images of different expression from an online database. And merged that with our primary dataset expressing seven different emotional states of human. All together, we have 38,657 images to work with. We used 80 % of the images from the dataset to train our model and 20 % of the images of the dataset to test our model. We prepared 30 epochs to evaluate the model over our primary dataset and 40 epochs have been prepared for the convolution network testing over combined dataset.

3.8 Implementation Requirement

A list of necessary and compulsory tools has been identified after a long and effective analysis over all of the statistical and theoretical concepts and methods that were required for this study. These tools that are listed down are the basic requirement for our kind of image processing and classification. These necessary tools are:

Hardware and Software requirements

- Operating System: Windows 7 or above
- Hard Disk: 500 Giga Bytes (Recommended)
- RAM: 4 GB (Recommended)

Developing Tools

- Python Environment
- Anaconda Navigator
- JupyterLab

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

Here, our tendency is to delineated the event method of distinctive seven completely different human emotions in this section. They typically ready the demonstration isolated into many steps like dataset assortment, info coming up with, info increase, info size and projected.

4.2 Performance Evaluation

Preparing exactness is commonly the truth once the demonstration is connected to getting ready data. Once the demonstration is connected to randomly-selected footage from the distinctive lesson, it's referred to as approval exactness. In our model, we tend to ready the network for thirty epochs while performing on the primary dataset. And tested the model in keeping with that. The subsequent figure shows the model-generated graph of training and validation accuracy with relevance epochs.

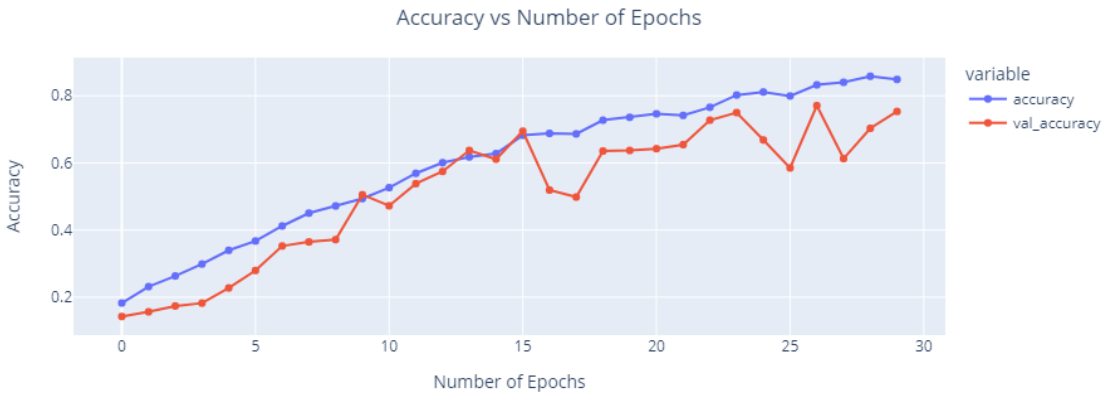


Figure 4.2.1: Training and Validation accuracy from primary dataset

Training loss is that the mistake of making ready a group of knowledge. Approval misfortune is that the blunder when running the approval set of knowledge through the ready network. The following figure shows the training and validation losses with respect to epochs.

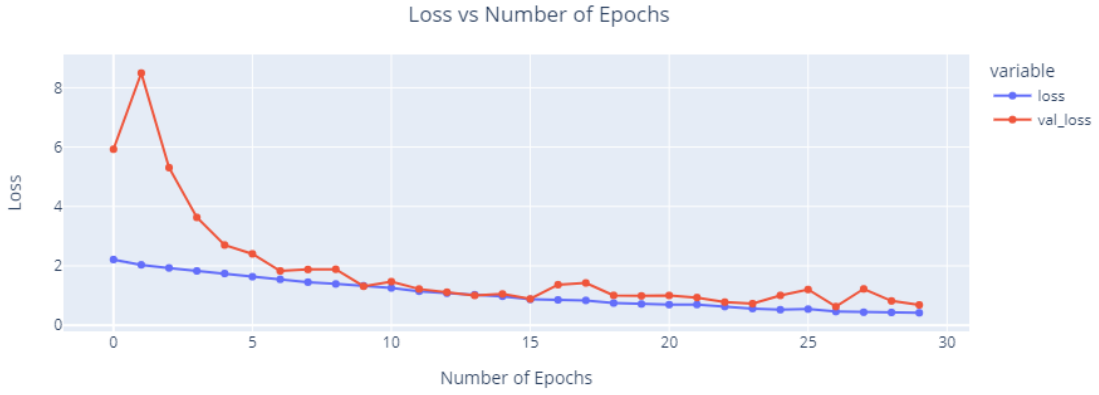


Figure 4.2.2: Training and Validation Losses from primary dataset

But when we applied our model on the combined dataset, we got our result which varies a little comparing to the output from primary dataset. We performed forty epochs while running the model on combined dataset. The combined dataset includes both sources of images, which are primary images and images collected from online. The following figure shows the model-generated graph of training and validation accuracy from our combined dataset with relevance epochs.

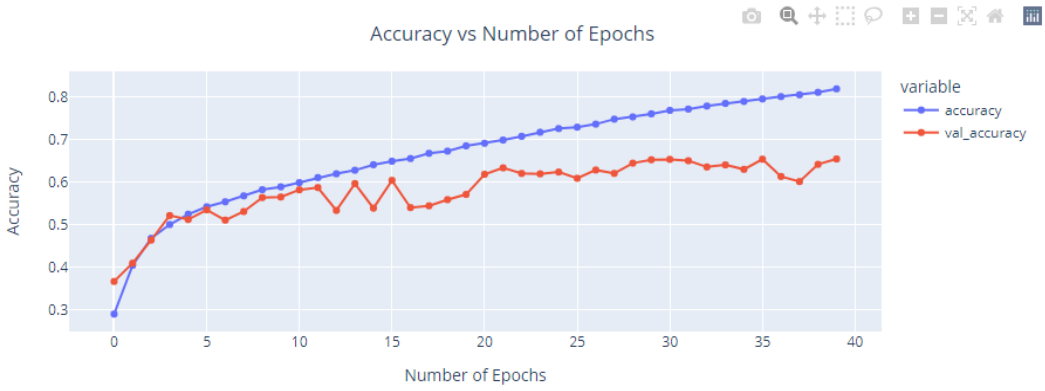


Figure 4.2.3: Training and Validation accuracy from combined dataset

We also observed slight changes in the reading of training and validation loss. The following figure of training and validation loss is generated while running the model on combined dataset.

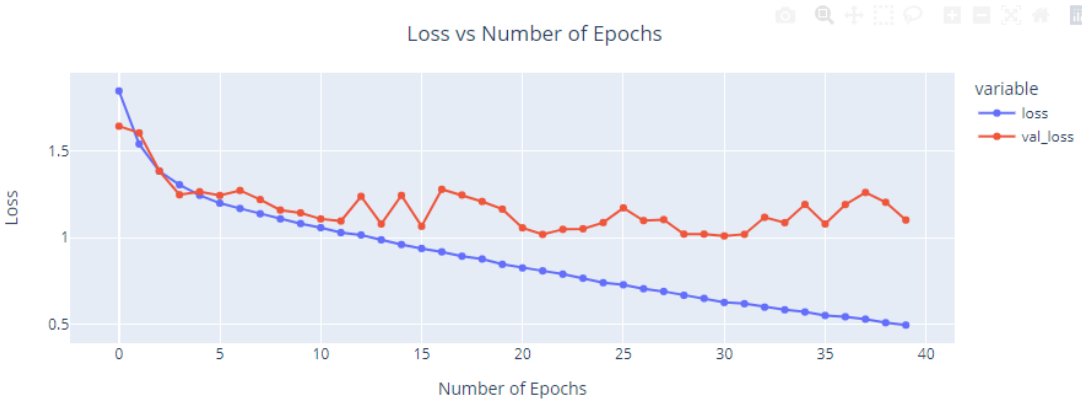


Figure 4.2.4: Training and Validation loss from combined dataset

4.3 Result Discussion

In this project we determined seven different emotion states from test dataset containing 7775 pictures. The total number of images used in our dataset is 38657. But at first, we evaluated our model with our primary dataset. Our primary dataset contains a total number of 2770 image. Our model showed 83 % accuracy when we ran it on the primary dataset. But as we evaluated our model on the basis of the combined dataset, it showed accuracy of 63 %.

The evaluation parameters such as precision, recall, support, f1 score and accuracy, all were calculated automatically by a bunch of line of coding used in the model. That is why we got our output result directly from the model.

Precision: One indicator of a machine learning model's performance is Precision. It refers to the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions:

$$\text{precision} = \frac{|\{relavent\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

Precision is used with the review, the percentage of all pertinent archives that is came by the design. Note that the that means and usage of "precision" among the sphere of information recovery contrasts from the definition of exactitude and exactitude within alternative branches of science and innovation.

Recall: The recall is that the piece of pertinent occasions that are recovered over the total quantity of relevant occurrences. The tall recall implies that a calculation came most of the many results.

$$\text{recall} = \frac{tp}{tp+fn}$$

Accuracy: Accuracy refers to the familiarity of the measured value to a known value.

$$\text{accuracy} = \frac{tp+tn}{tp+tn+fp+fn}$$

The accuracy for evaluating the model on primary dataset is 83 %. The complete performance parameters are shown here:

	precision	recall	f1-score	support
0	0.62	0.91	0.74	81
1	0.97	0.72	0.83	100
2	0.83	0.85	0.84	92
3	0.91	0.95	0.93	82
4	0.77	0.79	0.78	82
5	0.92	0.84	0.88	81
6	0.92	0.77	0.84	79
accuracy			0.83	597
macro avg	0.85	0.83	0.83	597
weighted avg	0.85	0.83	0.83	597

Figure 4.3.1: Performance parameters for primary dataset

The accuracy for evaluating the model on complete dataset is 63 %. The complete performance parameters are shown here:

	precision	recall	f1-score	support
0	0.36	0.69	0.48	1039
1	0.77	0.65	0.70	211
2	0.60	0.43	0.50	1116
3	0.89	0.81	0.85	1856
4	0.69	0.46	0.55	1315
5	0.53	0.56	0.54	1328
6	0.83	0.75	0.79	910
accuracy			0.63	7775
macro avg	0.67	0.62	0.63	7775
weighted avg	0.67	0.63	0.63	7775

Figure 4.3.2: Performance parameters for combined dataset

Therefore, the ultimate performance evaluation of our model stands:

Table 4.3.1: Comparison of accuracy over different dataset

Dataset	**Accuracy**
Primary	83 %
Combined	63 %

CHAPTER 5
SUMMARY AND CONCLUSION

5.1 Study Summary

It has little question that there's heaps of analysis work on this domain. essentially initial of all we've got projected a model for this project. we tend to had to gather such a lot information. Then our dataset was preprocessed for coaching and testing functions. we used associate algorithmic rule referred to as a “convolutional neural network” (CNN). At last, we discover our expected outcome in a brand-new domain with new parameters that accuracy is way higher than the remainder of the project-related work. the automated feeling recognition method is incredibly a lot of vital nowadays. This approach can facilitate in security police investigation, handling the hostile things yet as terrorist attacks, what is more in psychological treatment. It'll open the door for brand new trendy inventions that is our main goal.

5.2 Limitations

We faced many difficulties in the way of completing our research. We have always tried to overcome those difficulties. And we did our best to develop the system as we wanted it to be. But no system in the world is perfect., So are we. The automatic emotion detection model we have developed has a bunch of boundaries and limitations. We have tried to overcome these issues also. But due to technical problem, hardware limitations, lack of knowledge and shortage of time we could not find satisfactory solutions to those issues. They are as follows:

- Our system can only work with grayscale image.
- Accuracy gets decreased as we enlarge the dataset.
- System is taking too much time to train itself with big dataset.
- Evaluation parameters varies for multiple compilation.

These issues are unsolvable for us right now. But we believe that, we can overcome these problems too. We just need to stick to our idea, gain more and more knowledge, accept our failure and take lessons from our unsuccessfulness.

5.3 Conclusion

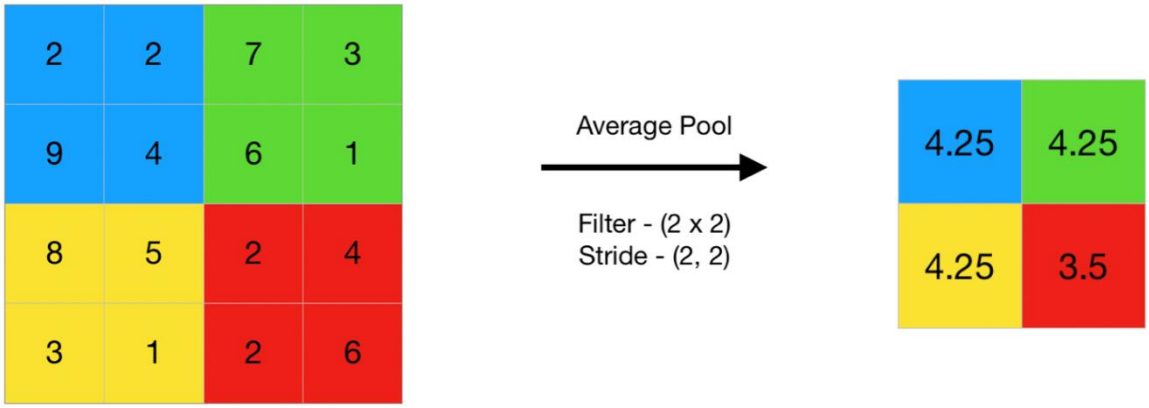
In this task, we've got given a strategy for detection emotions by our convolutional neural network. we have a tendency to use a complete of eleven convolution layers consisting of 3 dense layers, 3 dropout layers, and a pooling layer mistreatment "Average pool". For incrementing info, we have a tendency to took the assistance of knowledge augmentation. we have a tendency to used dropout layers to diminish overfitting. the result we've got accomplished is actually encouraging. this technique is going to be asked for and created within the future as a feature of any commitments in automatic feeling detection.

5.4 Future Works

In our proposed model, we have classified different facial emotions with convolution neural network. Our future objective is to frame a superior and more grounded precision. Our goal is to increase the accuracy level of our proposed model by training the model better and enriched information. We have a future objective to build a software or app which will be able to capture images and through image processing it will classify human emotions more accurately in real-time. We have classified seven emotions in our work. Our future target is to classify more human emotions through machine. We have faced some issues which are still unsolved. We will try to process colorful images and extract facial expression from real time captured images. We will try to fix the inconsistent accuracy problem of our proposed model. We can also try to make the algorithm efficient so that it becomes less time consuming while training and compiling. We hope it will bring another period in the era of modern technology.

APPENDICES

Appendix A



Appendix B

```
4332 angry images
737 disgusted images
4381 fearful images
7531 happy images
5254 neutral images
5147 sad images
3500 surprised images
```

Appendix C

```
Found 30882 images belonging to 7 classes.
Found 7775 images belonging to 7 classes.
```

REFERENCES

[1] Gu, H., Su, G., and Du, C., “Feature points extraction from faces”. 2003.

[2] Deng, X., Chang, C.-H., and Brandle, E., “A new method for eye extraction from facial image”, In DELTA ’04: Proceedings of the Second IEEE International Workshop on Electronic Design, Test and Applications, page 29, Washington, DC, USA, 2004. IEEE Computer Society

[3] Grimm, M., Dastidar, D. G., and Kroschel, K., “Recognizing emotions in spontaneous facial expressions”, 2008.

[4] Busso, C., Deng, Z., Yildirim, S., Bulut, M.,Lee, C. M., Kazemzadeh, A., Lee, S., Neumann,U., and Narayanan, S., “Analysis of emotion recognition using facial expressions, speech and multimodal information”, In ICMI ’2004: Proceedings of the 6th international conference on Multimodal interfaces, pages 205–211, New York, NY, USA, 2004. ACM.

[5] Te,o W. K., Silva, L. C. D., and Vadakkepat,P., “Facial expression detection and recognition system”, 2008.

[6] Cohen, I., Garg, A., and Huang, T. S., “Emotion recognition from facial expressions using multilevel hmm”, In Neural Information Processing Systems, 2000.

[7] Sanderson C., and Paliwal, K. K., “Fast feature extraction method for robust face verification. Electronics Letters”, 8:1648 – 1650, 2002.

[8] Everything About EMOTION RECOGNITION, available at <<https://sightcorp.com/knowledge-base/emotion-recognition/>>, last accessed on 03-12-2020 at 6:40 PM.

[9] Wallhoff, F., Facial expressions and emotion database, available at <<http://www.mmk.ei.tum.de/waf/fgnet/feedtum.html>>, last accessed on 01-12-2020 at 5:46 PM.

[10] Moon Hwan Kim, Young Hoon Joo and Jin Bae Park, “Emotion Detection Algorithm Using Frontal Face Image,” KINTEX, Gyeonggi-Do, Korea, vol. 1, pp. 1, June 2005.

[11] Moon Hwan Kim, Young Hoon Joo and Jin Bae Park, “Emotion Detection Algorithm Using Frontal Face Image,” KINTEX, Gyeonggi-Do, Korea, vol. 1, pp. 5-6, June 2005.

[12] Javier G. Razuri, David Sundgren, Rahim Rahmani, Antonio Moran Cardenas, “Automatic emotion recognition through facial expression analysis in merged images based on an Artificial Neural Network”,12th Mexican International Conference on Artificial Intelligence, vol. 1, pp. 1-13, November 2013.

[13] Lu chen, Tom Gedeon, Md Zakir Hossain and Sabrina Caldwel, "Are you really angry? Detecting emotion veracity as a proposed tool for interaction”,29th Australian Conference on Human-Computer Interaction, vol. 1, pp. 1-5, November 2017.

[14] pritam pal, Ananth N. Lyer and Robert E. Yantorno, "EMOTION DETECTION FROM INFANT FACIAL EXPRESSIONS AND CRIES”, IEEE, vol. 1, pp. 1-4, 2006.

[15] Convolutional Neural Network, available at <<https://en.wikipedia.org/wiki/Convolutional_neural_network>>, last accessed on 03-12-2020 at 6:55 PM

[16] Convolution of a filter over 2D image,Available at:<< https://www.allaboutcircuits.com/technical-articles/two-dimensional-convolution-in-image-processing/>> [Last accessed on 22-11-2020 at 9.02pm]

[17] Understanding and visualizing DenseNets, Available at: << https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a>> Last accessed on 29-11-2020 at 8.36 PM

[18] Introduction to pooling layer, Available at << https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/>> Last accessed on 30-11-2020 at 12.45 AM

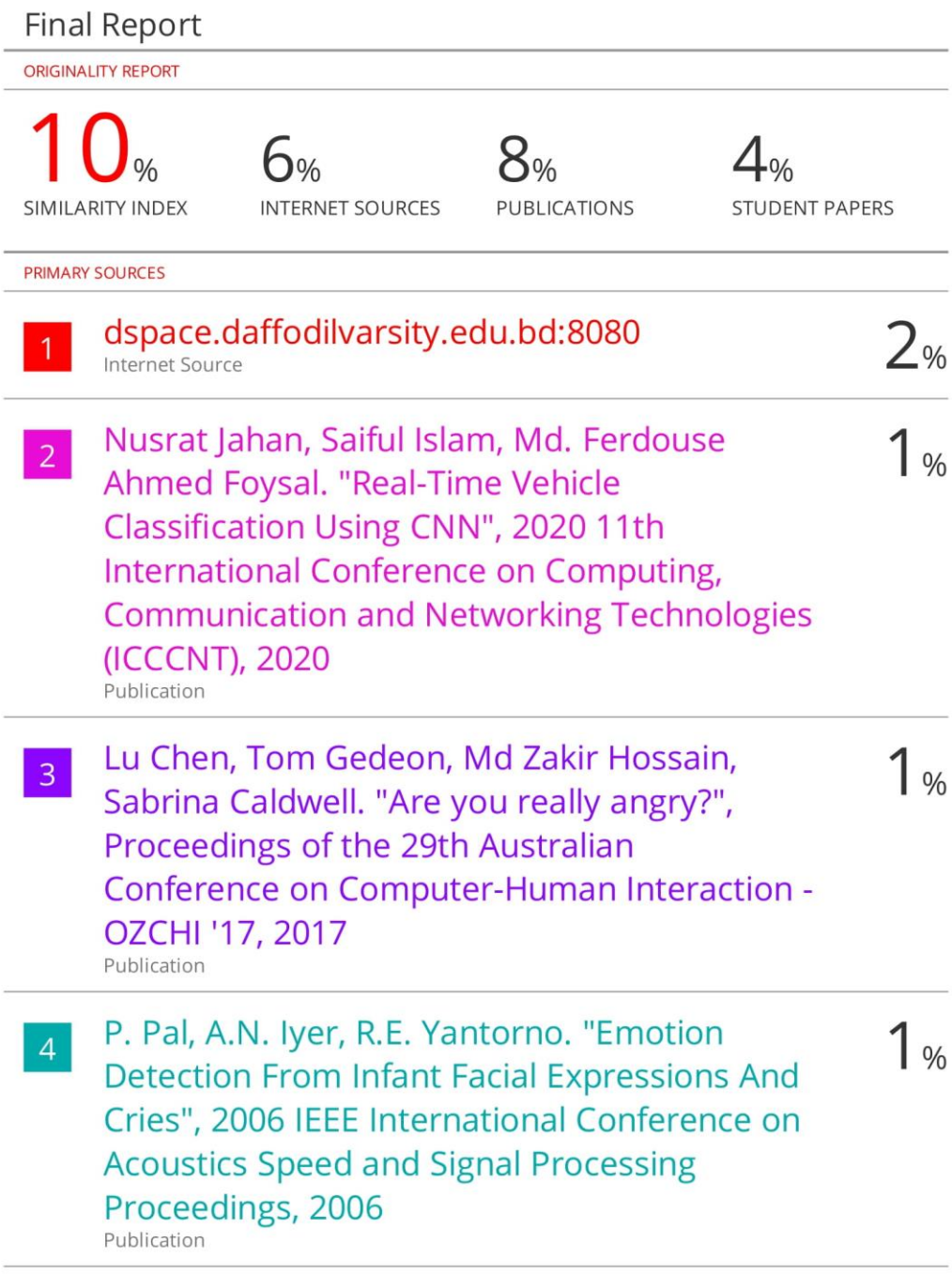
[19] FER-2013 Dataset, available at << https://www.kaggle.com/ananthu017/emotion-detection-fer>> Last accessed on 15-11-2020 at 01:03 AM.

[20] Softmax function, available at << https://deepai.org/machine-learning-glossary-and-terms/softmax-layer>> Last accessed on 03-12-2020 at 7:27 PM

[21] VGG net, available at << https://en.everybodywiki.com/VGG_Net>> Last accessed on 03-12-2020 at 7.31 PM

[22] Machine learning mastery, available at << https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/>> Last accessed on 03-12-2020 at 7.34 PM

PLAGIARISM REPORT



Final Report

ORIGINALITY REPORT

10%	6%	8%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
2	Nusrat Jahan, Saiful Islam, Md. Ferdouse Ahmed Foysal. "Real-Time Vehicle Classification Using CNN", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	1%
3	Lu Chen, Tom Gedeon, Md Zakir Hossain, Sabrina Caldwell. "Are you really angry?", Proceedings of the 29th Australian Conference on Computer-Human Interaction - OZCHI '17, 2017 Publication	1%
4	P. Pal, A.N. Iyer, R.E. Yantorno. "Emotion Detection From Infant Facial Expressions And Cries", 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, 2006 Publication	1%

5	Submitted to Shri Guru Gobind Singhji Institute of Engineering and Technology Student Paper	1 %
6	su.diva-portal.org Internet Source	1 %
7	www.ijert.org Internet Source	1 %
8	www.kdnuggets.com Internet Source	<1 %
9	dblp.dagstuhl.de Internet Source	<1 %
10	researchr.org Internet Source	<1 %
11	Submitted to Liverpool John Moores University Student Paper	<1 %
12	Submitted to SASTRA University Student Paper	<1 %
13	docplayer.net Internet Source	<1 %
14	Yongjie Wang, Wei Zhang, Yanyan Liu, Jianghua Zhu. "Multi-density map fusion network for crowd counting", Neurocomputing, 2020 Publication	<1 %

15	"Computational Intelligence in Pattern Recognition", Springer Science and Business Media LLC, 2020 Publication	<1 %
16	Submitted to Birla Institute of Technology and Science Pilani Student Paper	<1 %
17	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
18	air.ashesi.edu.gh Internet Source	<1 %
19	core.ac.uk Internet Source	<1 %
20	"Neural Information Processing", Springer Science and Business Media LLC, 2020 Publication	<1 %
21	Advances in Intelligent Systems and Computing, 2016. Publication	<1 %
22	uwspace.uwaterloo.ca Internet Source	<1 %
23	www.ncbi.nlm.nih.gov Internet Source	<1 %
24	Razuri, Javier G., David Sundgren, Rahim Rahmani, and Antonio Moran Cardenas.	<1 %

"Automatic Emotion Recognition through Facial Expression Analysis in Merged Images Based on an Artificial Neural Network", 2013 12th Mexican International Conference on Artificial Intelligence, 2013.
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		