

DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING

Submitted in partial fulfillment of the requirements for the award of Bachelor of
Engineering Degree in Electronics and Communication Engineering

by

JYOTHI VENKATESH (37130172)
AMIRISETTY VENU (37130021)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

MARCH 2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

www.sathyabama.ac.in



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of JYOTHI VENKATESH (37130172) and AMIRISETTY VENU (37130021) who carried out the project entitled “**DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING**” under our supervision from September 2020 to March 2021.

Internal Guide

DR.G.D ANBARASI JEBASELVI M.E.,Ph.D.

Associate Professor, Department of ECE

Head of the Department

Dr. T. RAVI, M.E., Ph.D.

Submitted for Viva voce examination held on _____

Internal Examiner

External Examiner

DECLARATION

We, **JYOTHI VENKATESH (37130172)** and **AMIRISETTY VENU (37130021)** here by declare

that the Project Report entitled “**DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING**” done by us under the guidance of **DR.G.D ANBARASI JEBASELVI M.E.,Ph.D. Associate Professor.** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

1)

2)

DATE:

PLACE:

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr. N. M. NANDHITHA, M.E., Ph.D. Dean, School of Electrical and Electronics Engineering** and **Dr. T. RAVI, M.E., Ph.D. Head of the Department, Department of Electronics and Communication Engineering** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **DR.G.D ANBARASI JEBASELVI M.E.,Ph.D. Associate Professor, Department of Electronics and Communication Engineering** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of the project work.

We wish to express our thanks to all teaching and Non-teaching staff members of the **Department of Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

We express our gratitude to our parents for their constant encouragement and support for the completion of the project.

Abstract

A Facial demeanor is the obvious sign of the full of feeling state, intellectual action, expectation, character and psychopathology of an individual and assumes an open part in relational relations. Customized affirmation of outward appearances can be a critical section of regular human-machine interfaces; it may in like manner be used in direct science and in clinical practice. A programmed Facial Expression Recognition framework needs to perform identification and area of countenances in a jumbled scene, facial element extraction, and outward appearance order.

Outward appearance acknowledgment framework is executed utilizing Convolution Neural Network (CNN). CNN model of the task depends on LeNet Architecture. Kaggle outward appearance dataset with seven outward appearance names as cheerful, pitiful, shock, dread, outrage, sicken, and nonpartisan is utilized in this undertaking. The framework accomplished 56.77 % exactness and 0.57 accuracy on testing dataset.

Keywords: Outward appearance Recognition, Convolutional Neural Network, Deep Learning, Theano

Abbreviations

CNN	Convolution Neural Network
FACS	Facial Action Coding System
FER	Facial Expression Recognition
GPU	Graphics Processing Unit
JAFFE	Japanese Female Facial Expressions
LDA	Linear Discriminant Analysis
PCA	Principle Component Analysis
ReLU	Rectified Linear Unit
SIANN	Space Invariant Artificial Neural Network

Table of Contents

Abstract.....	ii
Acknowledgement.....	iii
Abbreviations.....	iv
Table of Contents.....	v
List of Table and Figure	vi
1. Overview.....	1
1.1 Background.....	1
1.2 Problem definition	6
1.2 Objective.....	6
1.3 Scope of the Project.....	6
2. Literature Review	7
3. Methodology.....	9
3.1 Dataset 10	
3.2 Architecture of CNN	10
4. Results and Analysis.....	13
5. Conclusion	17
References	
Appendix	

List of Table and Figure

Figure 1.1: A model of CNN	2
Figure 1.2: Max Pooling	4
Figure 3 (a): Training Phase	9
Figure 3 (b): Testing Phase	9
Figure 3.2: Architecture of CNN	11
Figure 4.1 : Cost Function in Training and Validation... ..	13
Figure 4.2 : Error Rate in Training and Validation	14
Table 4.1: Confusion matrix for facial expression recognition... ..	14
Table 4.2: Normalized confusion matrix for facial expression recognition... ..	15
Table 4.3: Precision, Recall and F1-Score.....	15

CHAPTER 1

INTRODUCTION

1.1 Background

A Facial demeanor is the noticeable indication of the full of feeling state, intellectual action, goal, character and psychopathology of an individual and assumes an informative part in relational relations. Human outward appearances can be effectively grouped into 7 essential feelings: cheerful, dismal, shock, dread, outrage, nauseate, and impartial. Our facial feelings are communicated through actuation of explicit arrangements of facial muscles. These occasionally unpretentious, yet unpredictable, signals in an articulation regularly contain a plentiful measure of data about our perspective.

Programmed acknowledgment of outward appearances can be a significant part of common human-machine interfaces; it might likewise be utilized in conduct science and in clinical practice. It have been read for a significant stretch of time and acquiring the advancement late many years. In spite of the fact that much advancement has been made, perceiving outward appearance with a high precision stays to be troublesome because of the intricacy and assortments of outward appearances [1].

On an everyday nuts and bolts people ordinarily perceive feelings by trademark highlights, shown as a piece of an outward appearance. For example satisfaction is verifiably connected happily or an upward development of the sides of the lips. Likewise different feelings are portrayed by different disfigurements ordinary to a specific articulation. Examination into programmed acknowledgment of outward appearances tends to the issues encompassing the portrayal and classification of static or dynamic qualities of these misshapenings of face pigmentation [2].

In AI, a convolutional neural organization (CNN, or ConvNet) is a kind of feed-forward fake neural organization where the availability design between its neurons is roused by the association of the creature visual cortex. Individual cortical neurons react to boosts in a limited area of room known as the responsive field. The responsive fields of various neurons halfway cover with the end goal that they tile the visual field. The reaction of an individual neuron to improvements inside its open field can be approximated numerically by a convolution operation.[3] Convolutional networks were motivated by organic processes[4]

what's more, are varieties of multi-facet perceptron intended to utilize insignificant measures of preprocessing.[5]

They have wide applications in picture and video acknowledgment, recommender frameworks and common language handling. The convolutional neural organization is otherwise called move invariant or space invariant fake neural organization (SIANN), which is named dependent on its common loads design and interpretation invariance attributes.

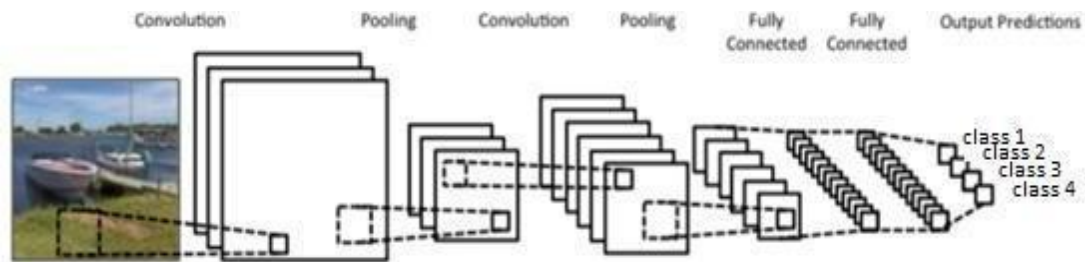


Figure 1.1: A model of CNN [5]

LeNet is one of the absolute first convolutional neural organizations which moved the field of Deep Learning. This spearheading work by Yann LeCun was named LeNet5 was utilized fundamentally for character acknowledgment assignments, for example, perusing postal divisions, digits, and so on The essential engineering of LeNet can be appeared as beneath [5]:

There are four main operations in the Convolution Neural Network shown in Figure 2.2 above:

1. Convolution:

The main role of Convolution if there should arise an occurrence of a CNN is to extricate highlights from the info picture. Convolution safeguards the spatial connection between pixels by learning picture highlights utilizing little squares of info information. The convolution layer's boundaries comprise of a bunch of learnable channels. Each channel is little spatially (along width and stature), yet reaches out through the full profundity of the info volume. For instance, a regular channel on a first layer of a CNN may have size 3x5x5 (for example pictures have profundity 3 for example the shading channels, 5 pixels width and stature). During the forward pass, each channel is convolved across the width and stature of the information volume and process dab items between the sections of the channel and the contribution at any position. As the channel convolve over the width and tallness of the info volume it creates a 2-dimensional enactment map that gives the reactions of that channel at each spatial position. Naturally, the organization will learn

channels that enact when they see some kind of visual component like an edge of some orientation or a blotch of some color on the first layer, or eventually entire honeycomb or wheel-like patterns on higher layers of the network. Now, there will be an entire set of filters in each convolution layer (e.g. 20 filters), and each of them will produce a separate 2-dimensional activation map.

The 2-dimensional convolution between image A and Filter B can be given as:

$$C(i,j) = \sum_{m=0}^{M_A-1} \sum_{n=0}^{N_A-1} A(m,n) * B(i-m, j-n) \quad (2.1)$$

where size of A is $(M_A \times N_A)$, size of B is $(M_B \times N_B)$, $0 \leq i < M_A - M_B + 1$ and $0 \leq j < N_A - N_B + 1$

A channel convolves with the input picture to deliver a component map. The convolution of another channel over a similar picture gives an alternate element map. Convolution activity catches the nearby conditions in the first picture. A CNN learns the estimations of these channels all alone during the preparation interaction (in spite of the fact that boundaries like number of channels, channel size, design of the organization and so forth actually expected to indicate before the preparation cycle). The more number of channels, the more picture highlights improve network becomes at perceiving designs in inconspicuous pictures.

The size of the Feature Map (Convolved Feature) is obliged by three limits Depth: Depth identifies with the amount of channels we use for the convolution action.

- **Stride:** Stride is the size of the channel, assuming the size of the channel is 5x5, step is 5.
- **Zero-cushioning:** Sometimes, it is advantageous to cushion the information lattice with zeros around the boundary, so that channel can be applied to lining components of information picture network. Utilizing zero cushioning size of the component guide can be controlled.

2. Rectified Linear Unit:

An extra activity called ReLU has been utilized after each Convolution activity. A Rectified Linear Unit (ReLU) is a cell of a neural organization which utilizes the accompanying initiation capacity to figure its yield given x:

$$R(x) = \text{Max}(0, x) \quad (2.2)$$

Utilizing these cells is more productive than sigmoid and still advances more data contrasted with parallel units. While instating the loads consistently, half of the loads are negative. This

helps making an inadequate element portrayal. Another positive viewpoint is the moderately modest calculation. No remarkable capacity must be determined. This capacity likewise forestalls the disappearing slope blunder, since the angles are direct capacities or zero yet for no situation non-straight capacities.

3. Pooling (sub-sampling)

Spatial Pooling (likewise called subsampling or downsampling) diminishes the dimensionality of each element map however holds the main data. Spatial Pooling can be of various kinds: Max, Average, Sum and so on. If there should arise an occurrence of Max Pooling, a spatial area (for instance, a 2×2 window) is characterized and the biggest component is taken from the corrected element map inside that window. If there should be an occurrence of normal pooling the normal or amount of all components in that window is taken. Practically speaking, Max Pooling has been appeared to work better.

Max Pooling reduces the commitment by applying the most limit over the data x_i . Let m be the size of the channel, by then the yield figures as follows:

$$M(\quad) = \{ +, + \mid \leq \quad / 2, \mid \leq \quad / 2, \quad \mathbb{N} \} \quad (2.3)$$

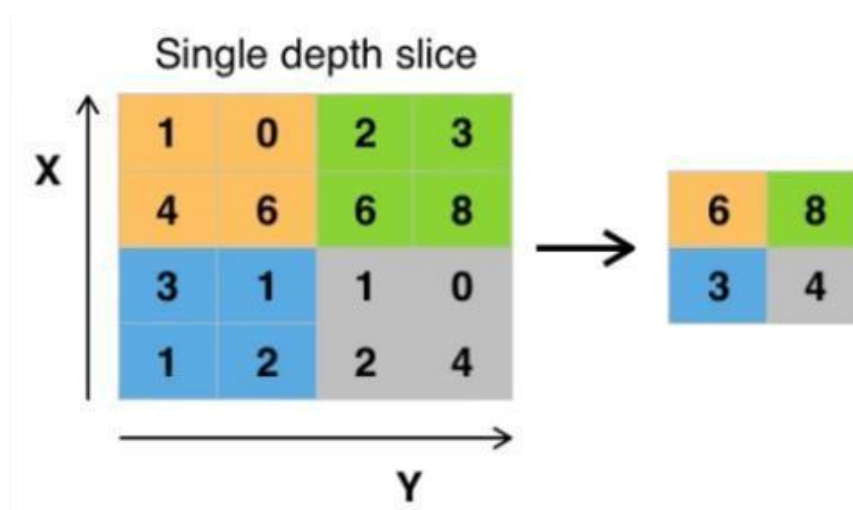


Figure 1.2 : Max Pooling

The capacity of Pooling is to continuously diminish the spatial size of the info portrayal. Specifically, pooling

- Makes the info portrayals (include measurement) more modest and more reasonable
 - Reduces the quantity of boundaries and calculations in the organization, in this manner, controlling over-fitting
 - Makes the organization invariant to little changes, twists and interpretations in the information picture (a little contortion in info won't change the yield of Pooling).
 - Helps us show up at a nearly scale invariant portrayal. This is incredible since articles can be recognized in a picture regardless of where they are found.

4. Classification (Multilayer Perceptron):

The Fully Connected layer is a conventional Multi-Layer Perceptron that utilizes a softmax actuation work in the yield layer. The expression "Completely Connected" suggests that each neuron in the past layer is associated with each neuron on the following layer. The yield from the convolutional and pooling layers address undeniable level highlights of the information picture. The motivation behind the Fully Connected layer is to utilize these highlights for characterizing the info picture into different classes dependent on the preparation dataset.

Softmax is utilized for actuation work. It regards the yields as scores for each class. In the Softmax, the capacity planning remained unaltered and these scores are deciphered as the un-standardized log probabilities for each class. Softmax is determined as:

where j is file for picture and K is number of all out outward appearance class.

Aside from characterization, adding a completely associated layer is likewise a (normally) modest method of learning non-straight mixes of these highlights. A large portion of the highlights from convolutional and pooling layers might be useful for the arrangement task, however blends of those highlights may be far and away superior.

The amount of yield probabilities from the Fully Connected Layer is 1. This is guaranteed by utilizing the as the enactment work in the yield layer of the Fully Connected Layer. The Softmax work takes a vector of self-assertive genuine esteemed scores and squashes it to a vector of qualities somewhere in the range of nothing and one that entirety to one.

1.2 Problem definition

Human feelings and expectations are communicated through outward appearances and inferring a productive and viable element is the principal part of outward appearance framework. Outward appearances pass on non-verbal signals, which assume a significant part in relational relations. Programmed acknowledgment of outward appearances can be a significant part of common human-machine interfaces; it might likewise be utilized in social science and in clinical practice.

A programmed Facial Expression Recognition framework needs to tackle the accompanying issues: identification and area of countenances in a jumbled scene, facial element extraction, and outward appearance arrangement.

1.2 Objective

The objective of the project is:

1. To implement Convolutional Neural Networks for classification of facial expressions.

1.3 Scope of the Project

In this task outward appearance acknowledgment framework is executed utilizing convolution neural organization. Facial pictures are characterized into seven outward appearance classes specifically Anger, Disgust, Fear, Happy, Sad, Surprise and 'Nonpartisan. Kaggle dataset is utilized to prepare and test the classifier.

CHAPTER 2

REVIEW OF LITERATURE

2. Literature Review

Two unique methodologies are utilized for outward appearance acknowledgment, the two of which incorporate two distinct techniques, exist [6]. Partitioning the face into discrete activity units or saving it in general for additional preparing seems, by all accounts, to be the first and the essential differentiation between the fundamental methodologies. In both of these methodologies, two unique procedures, in particular the 'Mathematical based' and the 'Appearance-based' definitions, can be utilized.

Utilizing the entire front facing face picture and handling it to wind up with the characterizations of 6 all inclusive outward appearance models: nauseate, dread, happiness, shock, pitiful ness and outrage; diagrams the main methodology. Here, it is accepted that every one of the previously mentioned feelings have trademark appearances on face and that is the reason acknowledgment of them is fundamental and adequate. Rather than utilizing the face pictures in general, isolating them into some sub-segments for additional handling structures up the principle thought of the second methodology for outward appearance investigation. As articulation is more related with inconspicuous changes of some discrete highlights like eyes, eyebrows and lip corners; these fine-grained changes are utilized for breaking down mechanized acknowledgment.

There are two principle strategies that are utilized in both of the above clarified approaches. Mathematical Based Parameterization is an old way which comprises of following and handling the movements of certain spots on picture arrangements, first and foremost introduced by Suwa et al to perceive outward appearances [7]. Cohn and Kanade later on attempted mathematical displaying and following of facial highlights by guaranteeing that every AU is given a particular arrangement of facial muscles [8]. The disservices of this technique are the forms of these highlights and parts must be changed physically in this casing, the issues of strength and troubles turn out in instances of posture and light changes while the following is applied on pictures, as activities and articulations will in general change both in morphological and in dynamical faculties, it turns out to be difficult to appraise general boundaries for development and removal. Consequently, winding up with powerful choices for facial activities under these differing conditions becomes to be troublesome. Maybe than following spatial focuses and utilizing situating and development boundaries that shift inside time, shading (pixel) data of related areas of face are prepared in Appearance Based

Parameterizations; to get the boundaries that will frame the component vectors. Various highlights, for example, Gabor, Haar wavelet coefficients, along with include extraction

and determination techniques, for example, PCA, LDA, and Adaboost are utilized inside this system.

For characterization issue, calculations like Machine learning, Neural Network, Support Vector Machine, Deep learning, Naive Bayes are utilized.

Raghuvanshi A. et al have assembled a Facial demeanor acknowledgment framework upon late examination to group pictures of human appearances into discrete feeling classes utilizing convolutional neural organizations [9].

CHAPTER 3

METHODOLOGY AND DESCRIPTION

3. Methodology

The facial expression recognition system is implemented using convolutional neural network.

The block diagram of the system is shown in following figures:

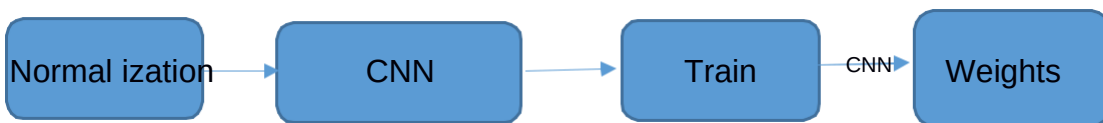


Figure 3 (a): Training Phase

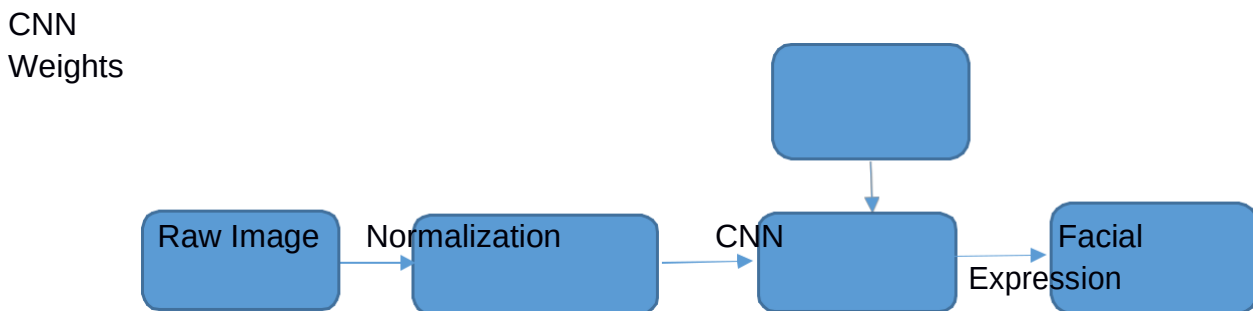


Figure 3 (b): Testing Phase

During preparing, the framework got a preparation information involving grayscale pictures of appearances with their individual articulation name and learns a bunch of loads for the organization. The preparation step took as info a picture with a face. From that point, a power standardization is applied to the picture. The standardized pictures are utilized to prepare the Convolutional Network. To guarantee that the preparation execution isn't influenced by the request for introduction of the models, approval dataset is utilized to pick the last best arrangement of loads out of a bunch of trainings performed with tests introduced in various orders. The yield of the preparation step is a bunch of loads that accomplish the best outcome with the preparation information. During test, the framework got a grayscale picture of a face from test dataset, and yield the anticipated appearance by utilizing the last organization loads mastered during preparing. Its yield is a solitary number that addresses one of the seven essential articulations.

3.1 Dataset

The dataset from a Kaggle Facial Expression Recognition Challenge (FER2013) is utilized for the preparation and testing. It contains pre-trimmed, 48-by-48-pixel grayscale pictures of countenances each marked with one of the 7 feeling classes: outrage, appall, dread, joy, misery, shock, and unbiased. Dataset has preparing set of 35887 facial pictures with outward appearance marks.. The dataset has class awkwardness issue, since certain classes have huge number of models while some has not many. The dataset is adjusted utilizing oversampling, by expanding numbers in minority classes. The decent dataset contains 40263 pictures, from which 29263 pictures are utilized for preparing, 6000 pictures are utilized for testing, and 5000 pictures are utilized for approval.

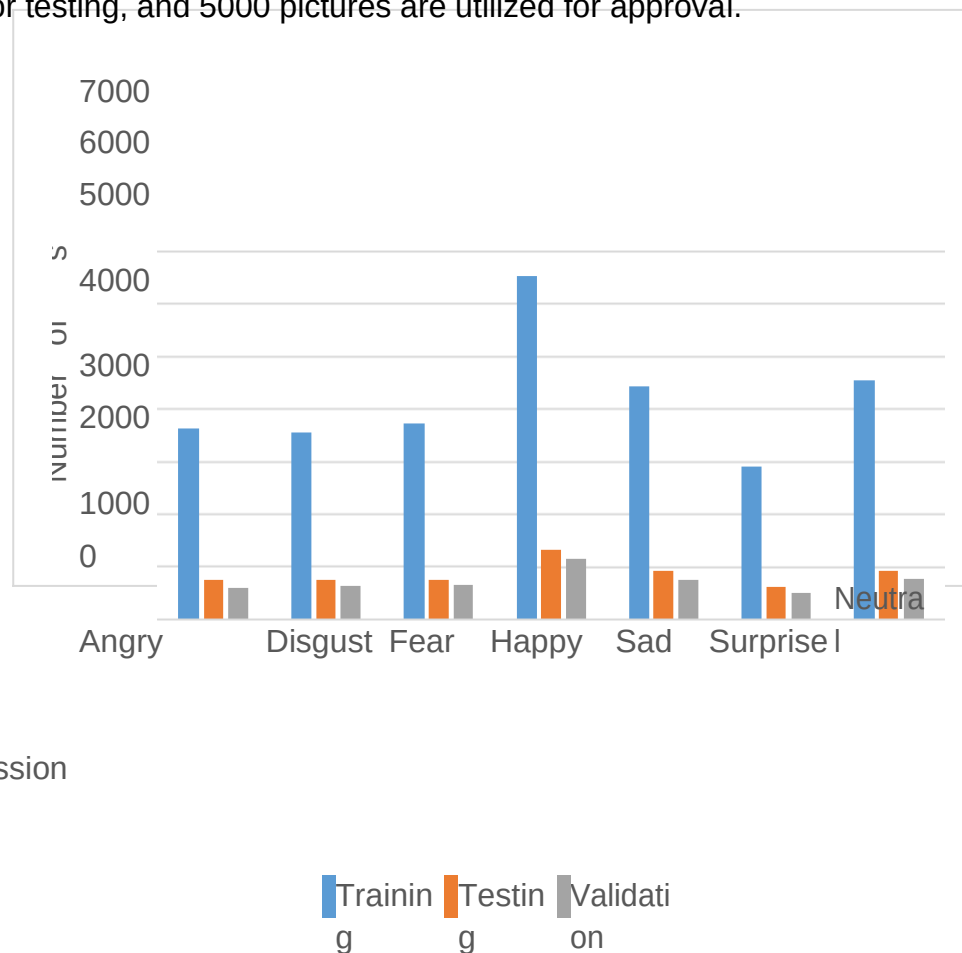


Figure 3.1: Training, Testing and Validation Data distribution

3.2 Architecture of CNN

A normal design of a convolutional neural organization contains an information layer, some convolutional layers, some completely associated layers, and a yield layer. CNN is planned with some alteration on LeNet Architecture [10]. It has 6 layers without thinking about info and yield. The design of the Convolution Neural Network utilized in the venture is appeared in the Accompanying figure.

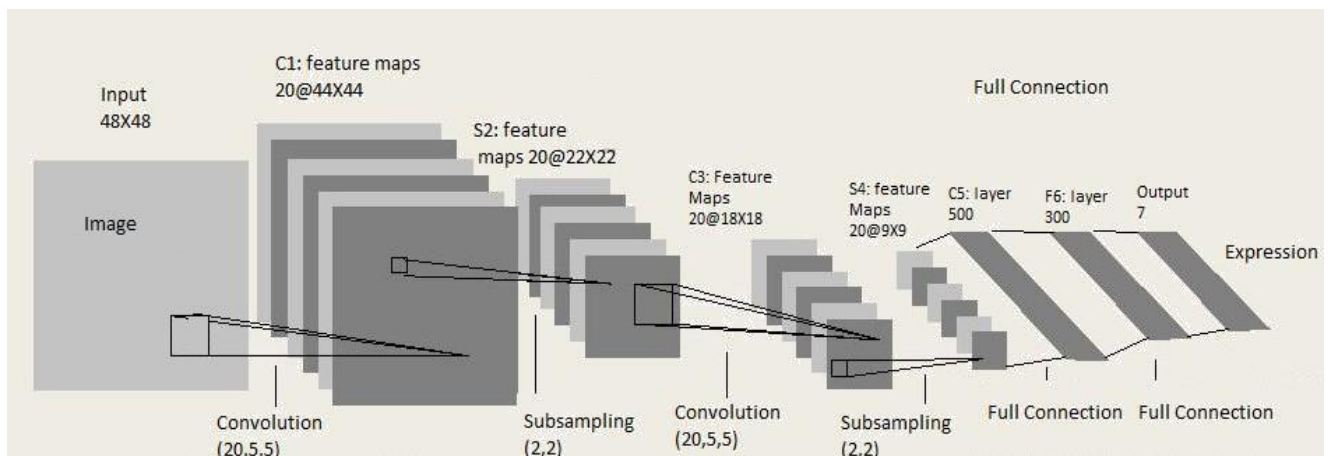


Figure 3.2: Architecture of CNN (Modified from LeNet Architecture)

1. Input Layer:

The information layer has pre-decided, fixed measurements, so the picture should be pre-handled before it very well may be taken care of into the layer. Standardized dark scale pictures of size 48 X 48 pixels from Kaggle dataset are utilized for preparing, approval and testing. For testing propose PC webcam pictures are additionally utilized, in which face is distinguished and edited utilizing OpenCV Haar Cascade Classifier and standardized.

2. Convolution and Pooling (ConvPool) Layers:

Convolution and pooling is done dependent on group handling. Each clump has N pictures and CNN channel loads are refreshed on those clusters. Every convolution layer takes picture group contribution of four measurement N x Color-Channel x width x stature. Highlight guide or channel for convolution are likewise four dimensional (Number of highlight maps in, number of highlight delineates, channel width, channel stature). In every

convolution layer, four dimensional convolution is determined between picture bunch and highlight maps. After convolution just boundary that change is picture width and stature.

9. New picture width = old picture width – channel width + 1 New picture stature = old picture tallness – channel tallness + 1 After every convolution layer downsampling/subsampling is accomplished for dimensionality decrease. This cycle is called Pooling. Max pooling and Average Pooling are two celebrated pooling technique. In this venture max pooling is done after convolution. Pool size of (2x2) is taken, what parts the picture into matrix of squares every one of size 2x2 and takes limit of 4 pixels. Subsequent to pooling just stature and width are influenced.

12. Two convolution layer and pooling layer are utilized in the design. From the outset convolution layer size of info picture clump is $N \times 1 \times 48 \times 48$. Here, size of picture group is N, number of shading channel is 1 and both picture stature and width are 48 pixel. Convolution with include guide of $1 \times 20 \times 5 \times 5$ outcomes picture group is of size $N \times 20 \times 44 \times 44$. After convolution pooling is finished with pool size of 2x2, which results picture cluster of size $N \times 20 \times 22 \times 22$. This is trailed by second convolution layer with include guide of $20 \times 20 \times 5 \times 5$, which results picture cluster of size $N \times 20 \times 18 \times 18$. This is trailed by pooling layer with pool size 2x2, which results picture group of size $N \times 20 \times 9 \times 9$.

This layer is roused by the manner in which neurons send signals through the cerebrum. It takes countless info includes and change highlights through layers associated with teachable loads. Two secret layers of size 500 and 300 unit are utilized in completely associated layer. The loads of these layers are prepared by forward engendering of preparing information then in reverse spread of its mistakes. Back engendering begins from assessing the distinction among expectation and genuine worth, and back ascertains the weight change expected to each layer previously. We can handle the preparation speed and the intricacy of the design by tuning the hyper-boundaries, for example, learning rate and organization thickness. Hyper-boundaries for this layer incorporate learning rate, force, regularization boundary, and rot.

The yield from the second pooling layer is of size $N \times 20 \times 9 \times 9$ and contribution of first secret layer of completely associated layer is of size $N \times 500$. In this way, yield of pooling layer is leveled to $N \times 1620$ size and took care of to first secret layer. Yield from first secret layer is taken care of to second secret layer. Second secret layer is of size $N \times 300$ and its yield is taken care of to yield layer of size equivalent to number of outward appearance classes.

13. Output Layer

Yield from the second secret layer is associated with yield layer having seven unmistakable classes. Utilizing Softmax initiation work, yield is acquired utilizing the probabilities for every one of the seven class. The class with the most elevated likelihood is the anticipated class.

CHAPTER 4

RESULT AND ANALYSIS

4. Results and Analysis

CNN design for outward appearance acknowledgment as referenced above was executed in Python.

Alongside Python programming language, Numpy, Theano and CUDA libraries were utilized.

Preparing picture bunch size was taken as 30, while channel map is of size 20x5x5 for both convolution layer. Approval set was utilized to approve the preparation cycle. In last bunch of each age in approval cost, approval blunder, preparing cost, preparing mistake are determined. Information boundaries for preparing are picture set and comparing yield names. The preparation interaction refreshed the loads of highlight maps and covered up layers dependent on hyper-boundaries like learning rate, energy, regularization and rot. In this framework clump insightful learning rate was utilized as $10e-5$, force as 0.99, regularization as $10e-7$ and rot as 0.99999.

The correlation of approval cost, approval mistake, preparing cost, preparing blunder are appeared In figures underneath.

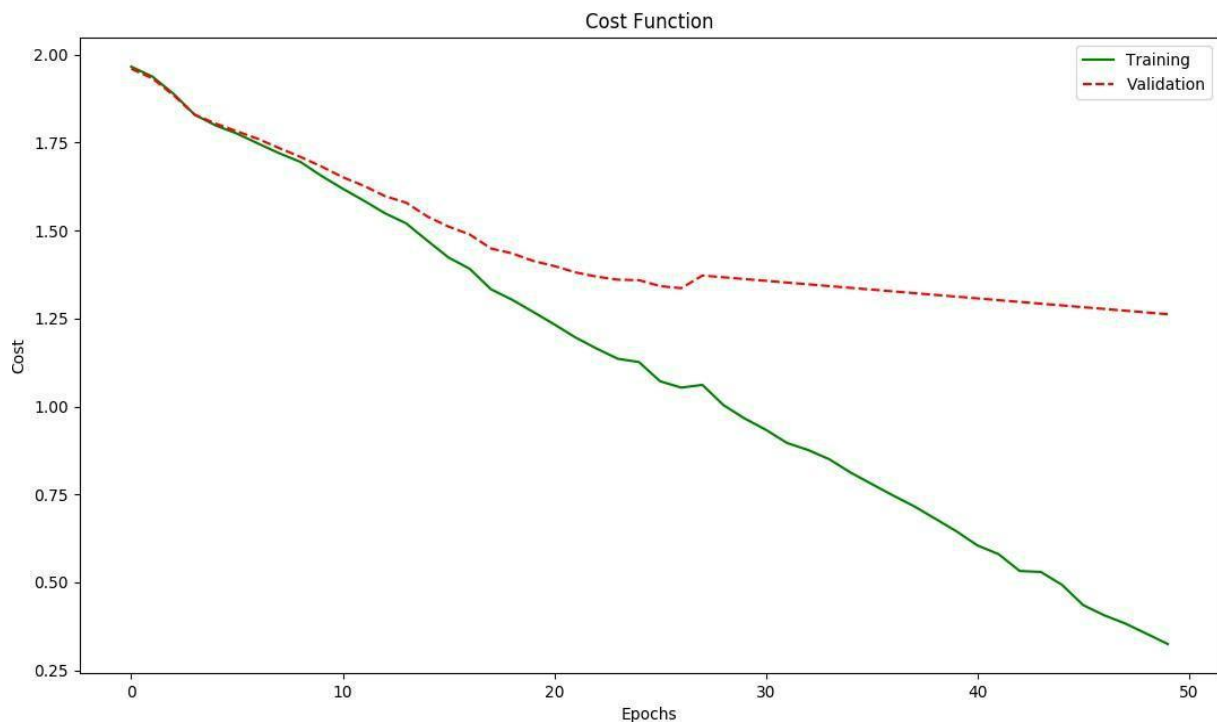


Figure 4.1 : Cost function in training and validation

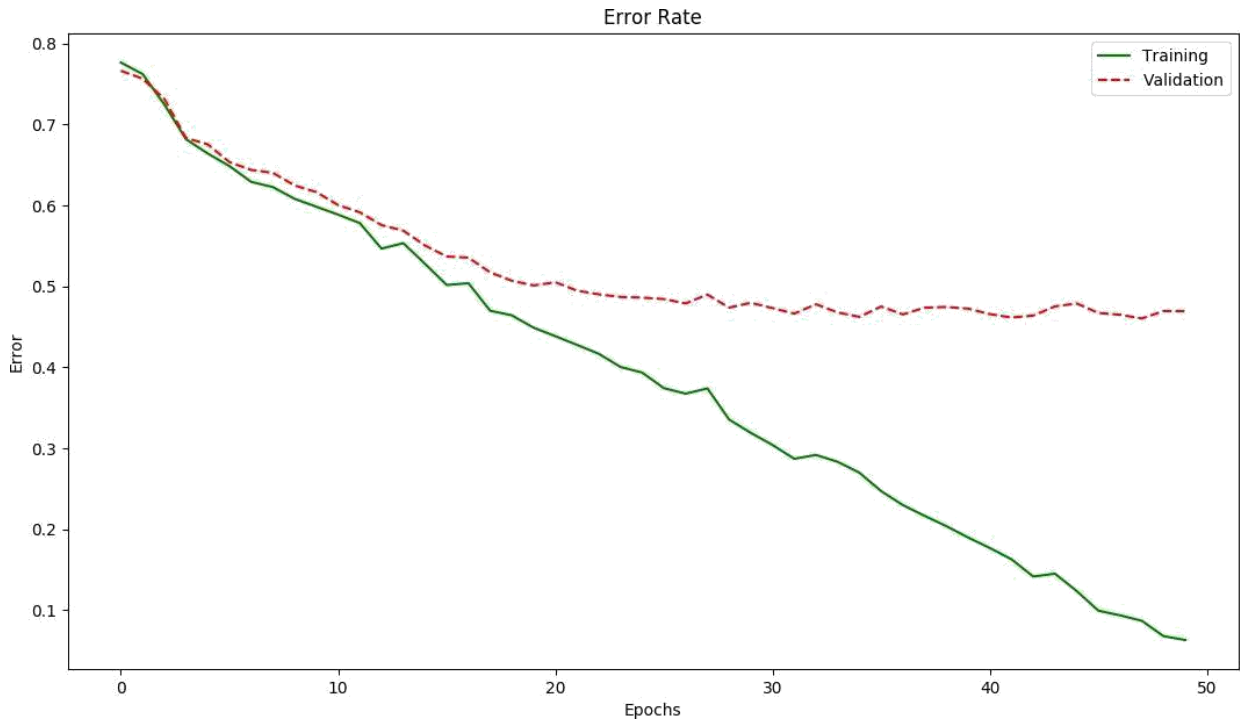


Figure 4.2 : Error Rate in Training and Validation

Above chart shows that the in 50 ages preparing cost and mistake rate is diminished 0.35 and 0.1 individually while approval cost and blunder is 1.25 and 0.45 separately.

The testing of the model is done utilizing 6000 pictures. The classifier gave 56.77 % precision. The disarray framework for seven outward appearance classes is appeared beneath:

Table 4.1 : Confusion matrix for facial expression recognition

Predicted \ Actual	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral
Anger	312	16	93	89	97	27	106

Disgust	35	70	0	0	0	0	0
Fear	12	7	30	69	10	67	11
Happy	85	5	74	88	80	29	12
Sad	13	6	12	89	33	27	16
Surprise	27	4	75	45	24	40	46
Neutral	11	4	91	11	12	32	46

Table 4.2 : Normalized confusion matrix for facial expression recognition

Pred ict ed Actual	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.42 1622	0.0 217 69	0.1 187 74	0.06 9099	0.10 9729	0.04 3339	0.11 193
Disgust	0.04 7619	0.9 523 8	0	0	0	0	0
Fear	0.16 6216	0.0 095 24	0.3 831 42	0.05 3571	0.11 3122	0.10 7544	0.12 3548
Happy	0.11 4865	0.0 068 03	0.0 945 08	0.68 8665	0.09 0498	0.04 6549	0.13 5164
Sad	0.18 5135	0.0 081 63	0.1 558 11	0.06 9099	0.38 2353	0.04 3339	0.17 4234
Surprise	0.03 6486	0.0 054 42	0.0 957 85	0.03 4938	0.02 7149	0.64 5265	0.04 8574
Neutral	0.15 4054	0.0 054 42	0.1 162 2	0.09 0062	0.1 391 4	0.05 1364	0.49 3136

Above tables showed that this model most noteworthy precision for sicken feeling with 95.23 %, trailed by content with 68.86 %, shock with 64.52 %, impartial with 49.31 %, outrage with 42.16

%, dread with 38.31 % and most reduced exactness for tragic feeling as 38.23 %.

The precision, recall and F1-score of each facial expression class is shown below:

Table 4.3 : Precision, Recall and F1-Score

	Precision	Recall	F1-score
Anger	0.39	0.42	0.41
Disgust	0.95	0.99	0.97
Fear	0.45	0.38	0.39
Happy	0.68	0.69	0.69
Sad	0.44	0.38	0.41
Surprise	0.69	0.65	0.67
Neutral	0.45	0.49	0.47
Average	0.57	0.57	0.57

The general exactness and review are 0.57 and 0.57 individually. The model performs truly well on characterizing positive feelings bringing about generally high accuracy scores for glad and shocked. Nauseate has most noteworthy accuracy and review as 0.95 and 0.99 as pictures in this class were oversampled to address class unevenness. Cheerful has an accuracy of 0.68 and review of 0.69 which could be clarified by having the most models (6500) in the preparation set. Curiously,

shock has an exactness of 0.69 and review of 0.65 having minimal models in the preparation set.

There should be extremely solid signs in the unexpected articulations.

Model execution appears to be more fragile across negative feelings by and large. In especially, the feeling miserable has a low exactness of just 0.44 and review 0.38. The model oftentimes misclassified irate, dread and unbiased as miserable. Furthermore, it is most befuddled while anticipating pitiful and impartial countenances in light of the fact that these two feelings are likely the most un-expressive (barring crying appearances).

The generally F1-score is additionally 0.57 . F1-score is most elevated for sicken due to oversampling of pictures. Cheerful and shock have higher F1-score as 0.69 and 0.67 separately. Dread has least F1-score as 0.39 and dismal, outrage and unbiased additionally have low F1-score.

CNN Classifier is then used to order picture taken from webcam in Laptop. Face is recognized in webcam outlines utilizing Haar course classifier from OpenCV. At that point distinguished face is edited and standardized and taken care of to CNN Classifier. Some grouping results utilizing webcam are recorded in Appendix A.

CHAPTER 5

SUMMARY AND CONCLUSION

5. Summary

5.1 Conclusion

In this task, a LeNet engineering based six layer convolution neural organization is executed to arrange human outward appearances for example cheerful, pitiful, shock, dread, outrage, appall, and unbiased. The framework has been assessed utilizing Accuracy, Precision, Recall and F1-score. The classifier accomplished exactness of 56.77 % , accuracy of 0.57, review 0.57 and F1-score 0.57.

5.2 Enhancement

Later on work, the model can be reached out to shading pictures. This will permit to examine the

viability of pre-prepared models, for example, AlexNet[11] or VGGNet [12] for facial feeling acknowledgment.

References

- [1] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In *Image Processing, 2005. ICIP 2005*. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.
- [2] Chibelushi, C. C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. *CVonline: On-Line Compendium of Computer Vision*, 9.
- [3] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 31 August 2013.
- [4] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). *Neural Networks*. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013.
- [5] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013
- [6] C. Zor, "Facial expression recognition," Master's thesis, University of Surrey, Guildford, 2008.
- [7] Suwa, M.; Sugie N. and Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression, *Proc. International Joint Conf, Pattern Recognition*, pages 408-410, 1978
- [8] Recognizing action units for facial expression analysis YI Tian, T Kanade, JF Cohn *IEEE Transactions on pattern analysis and machine intelligence* 23 (2), 97- 115
- [9] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with
- [10] Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016
- [11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet

classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[12] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

Appendix A

CNN classifier result using webcam

1. Examples with Correct Outputs

2. Examples with Incorrect Output

- a) Neutral expression is incorrectly recognized as surprise
- b) Anger expression is incorrectly recognized as surprise
- c) Sad expression is incorrectly recognized as anger

Coding Snapshots

```
real_time_video.py - C:\Users\jvenk\OneDrive\Desktop\Emotion-recognition-master\Emotion-recognition-master\real_time_video.py (3.7.8)
File Edit Format Run Options Window Help
from keras.preprocessing.image import img_to_array
import imutils
import cv2
from keras.models import load_model
import numpy as np

# parameters for loading data and images
detection_model_path = "haarcascade_files\haarcascade_frontalface_default.xml"
emotion_model_path = "models\mini_XCEPTION.102-0.66.hdf5"

# hyper-parameters for bounding boxes shape
# loading models
face_detection = cv2.CascadeClassifier(detection_model_path)
emotion_classifier = load_model(emotion_model_path, compile=False)
EMOTIONS = ["angry", "disgust", "scared", "happy", "sad", "surprised", "neutral"]

# feelings_faces = []
# for index, emotion in enumerate(EMOTIONS):
#     feelings_faces.append(cv2.imread('emojis/' + emotion + '.png', -1))

# starting video streaming
cv2.namedWindow('your_face')
camera = cv2.VideoCapture(0)
while True:
    frame = camera.read()[1]
    # reading the frame
    frame = imutils.resize(frame, width=300)
    gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
    faces = face_detection.detectMultiScale(gray, scaleFactor=1.1, minNeighbors=5, minSize=(30, 30), flags=cv2.CASCADE_SCALE_IMAGE)

    canvas = np.zeros((250, 300, 3), dtype="uint8")
    frame_clone = frame.copy()
    if len(faces) > 0:
        faces = sorted(faces, reverse=True,
            key=lambda x: (x[2] - x[0]) * (x[3] - x[1]))[0]
        (fx, fy, fw, fh) = faces
        # Extract the ROI of the face from the grayscale image, resize it to a fixed 28x28 pixels, and then prepare
        # the ROI for classification via the CNN
        roi = gray[fy:fy + fh, fx:fx + fw]
        roi = cv2.resize(roi, (64, 64))
        roi = roi.astype("float") / 255.0
        roi = img_to_array(roi)
        roi = np.expand_dims(roi, axis=0)

        preds = emotion_classifier.predict(roi)[0]
        emotion_probability = np.max(preds)
        label = EMOTIONS[preds.argmax()]
        else: continue

    for (i, (emotion, prob)) in enumerate(zip(EMOTIONS, preds)):
        # construct the label text
        text = "{}: {:.2f}%".format(emotion, prob * 100)

        # draw the label + probability bar on the canvas
        emoji_face = feelings_faces[np.argmax(preds)]

        w = int(prob * 300)
        cv2.rectangle(canvas, (7, (i * 35) + 5),
            (w, (i * 35) + 35), (0, 0, 255), -1)
        cv2.putText(canvas, text, (10, (i * 35) + 23),
            cv2.FONT_HERSHEY_SIMPLEX, 0.45,
            (255, 255, 255), 2)
        cv2.putText(frame_clone, label, (fx, fy - 10),
            cv2.FONT_HERSHEY_SIMPLEX, 0.45, (0, 0, 255), 2)
        cv2.rectangle(frame_clone, (fx, fy), (fx + fw, fy + fh),
            (0, 0, 255), 2)

    for c in range(0, 3):
        frame[200:320, 10:130, c] = emoji_face[:, :, c] * \
            (emoji_face[:, :, 3] / 255.0) + frame[200:320,
            10:130, c] * (1.0 - emoji_face[:, :, 3] / 255.0)

    cv2.imshow('your_face', frame_clone)
    cv2.imshow("Probabilities", canvas)
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

camera.release()
cv2.destroyAllWindows()
```

```
real_time_video.py - C:\Users\jvenk\OneDrive\Desktop\Emotion-recognition-master\Emotion-recognition-master\real_time_video.py (3.7.8)
File Edit Format Run Options Window Help
# Extract the ROI of the face from the grayscale image, resize it to a fixed 28x28 pixels, and then prepare
# the ROI for classification via the CNN
roi = gray[fy:fy + fh, fx:fx + fw]
roi = cv2.resize(roi, (64, 64))
roi = roi.astype("float") / 255.0
roi = img_to_array(roi)
roi = np.expand_dims(roi, axis=0)

preds = emotion_classifier.predict(roi)[0]
emotion_probability = np.max(preds)
label = EMOTIONS[preds.argmax()]
else: continue

for (i, (emotion, prob)) in enumerate(zip(EMOTIONS, preds)):
    # construct the label text
    text = "{}: {:.2f}%".format(emotion, prob * 100)

    # draw the label + probability bar on the canvas
    emoji_face = feelings_faces[np.argmax(preds)]

    w = int(prob * 300)
    cv2.rectangle(canvas, (7, (i * 35) + 5),
        (w, (i * 35) + 35), (0, 0, 255), -1)
    cv2.putText(canvas, text, (10, (i * 35) + 23),
        cv2.FONT_HERSHEY_SIMPLEX, 0.45,
        (255, 255, 255), 2)
    cv2.putText(frame_clone, label, (fx, fy - 10),
        cv2.FONT_HERSHEY_SIMPLEX, 0.45, (0, 0, 255), 2)
    cv2.rectangle(frame_clone, (fx, fy), (fx + fw, fy + fh),
        (0, 0, 255), 2)

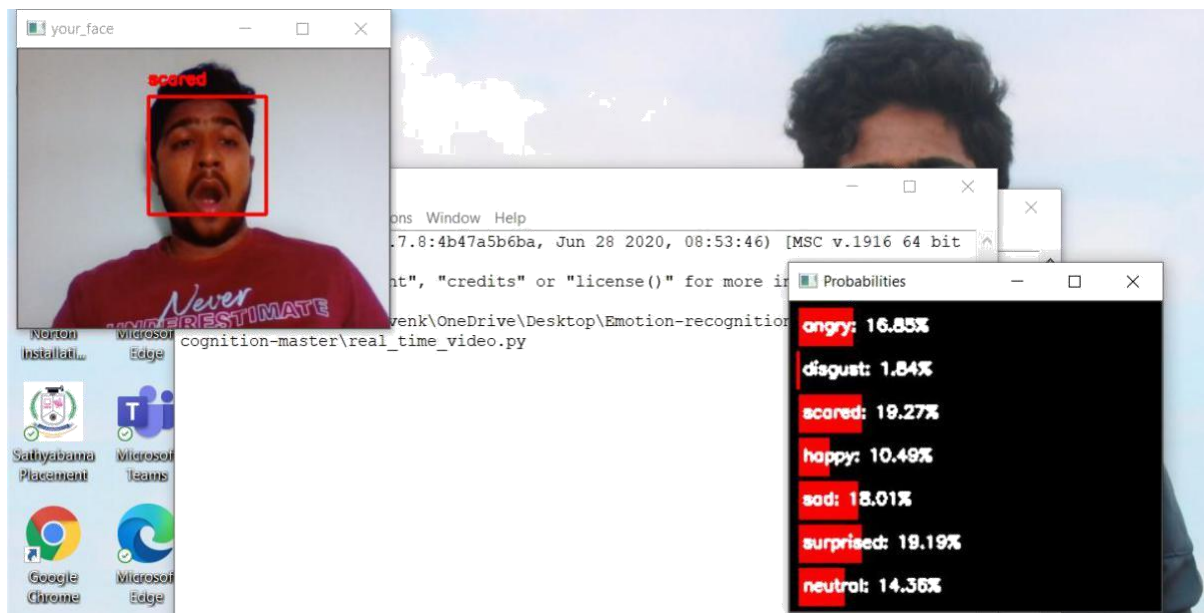
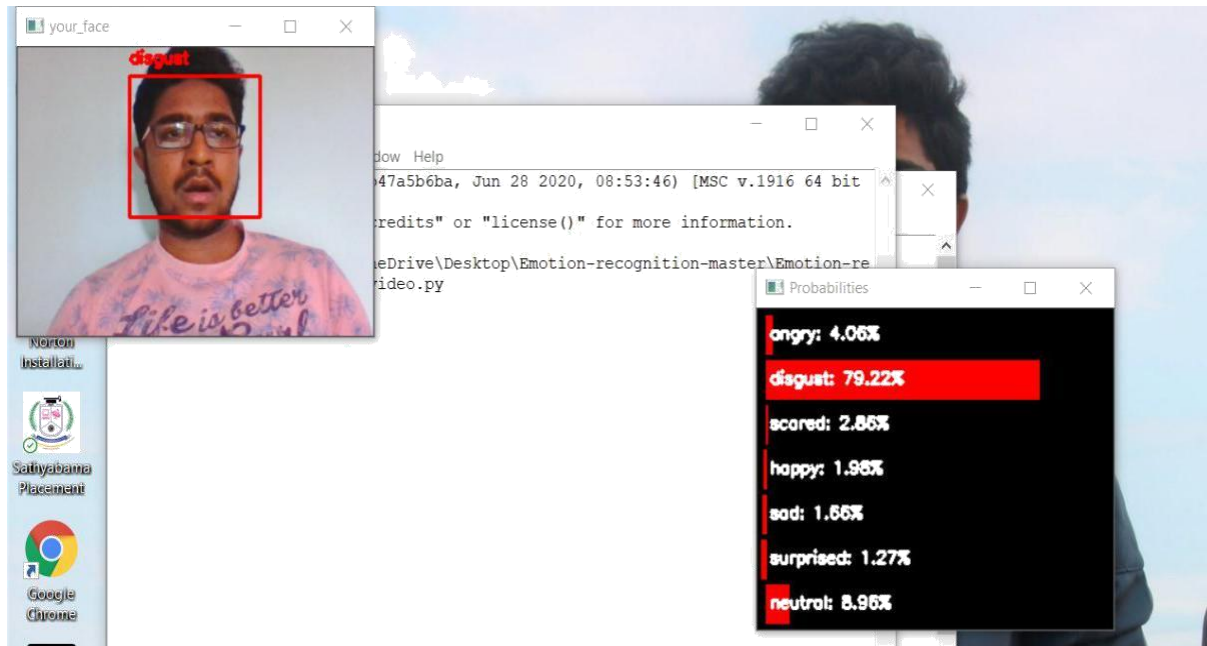
    for c in range(0, 3):
        frame[200:320, 10:130, c] = emoji_face[:, :, c] * \
            (emoji_face[:, :, 3] / 255.0) + frame[200:320,
            10:130, c] * (1.0 - emoji_face[:, :, 3] / 255.0)

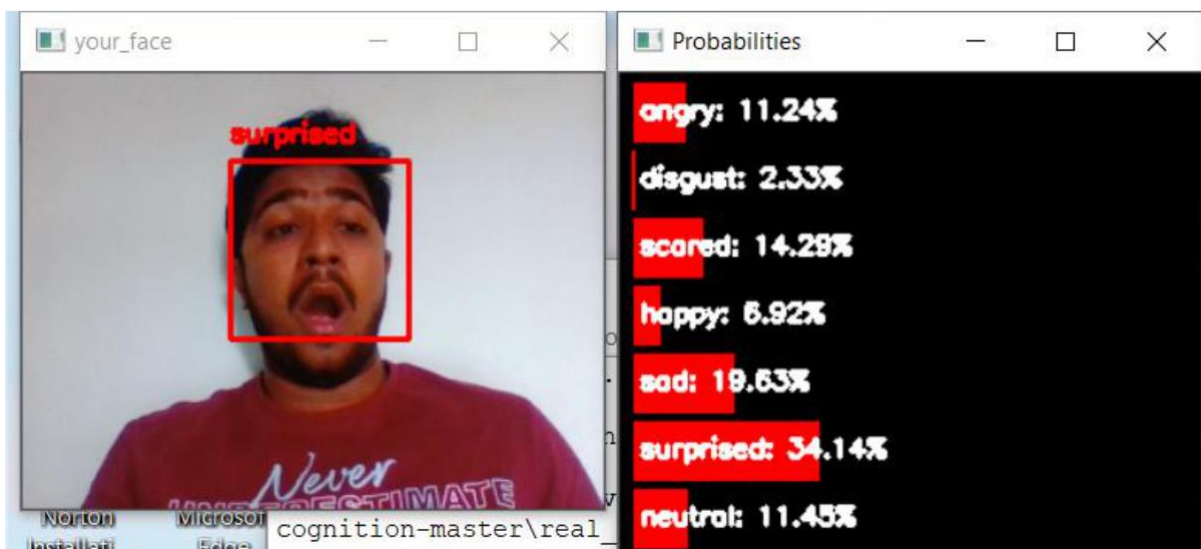
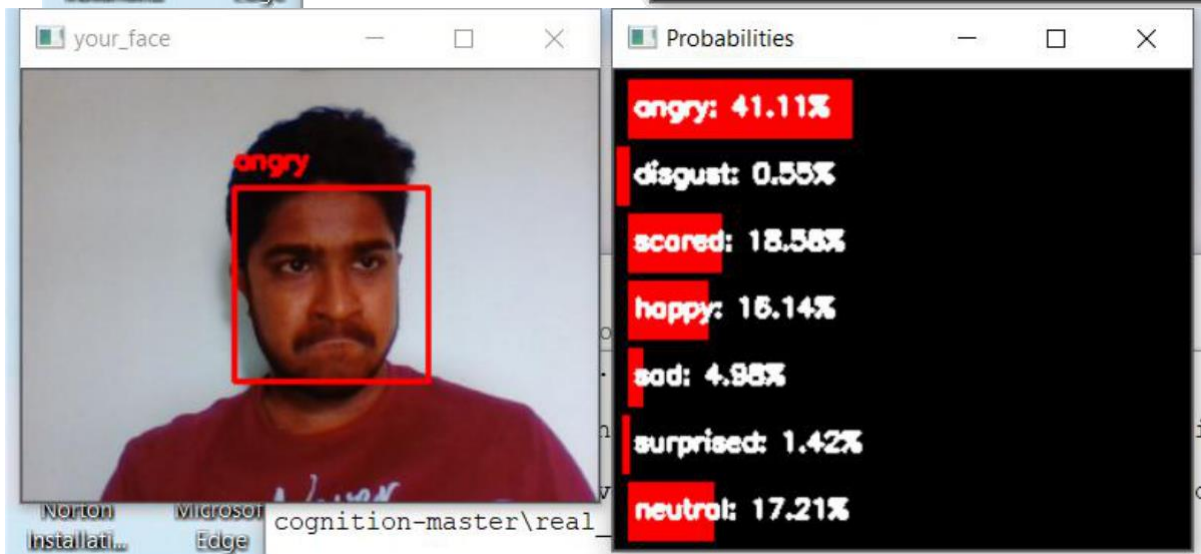
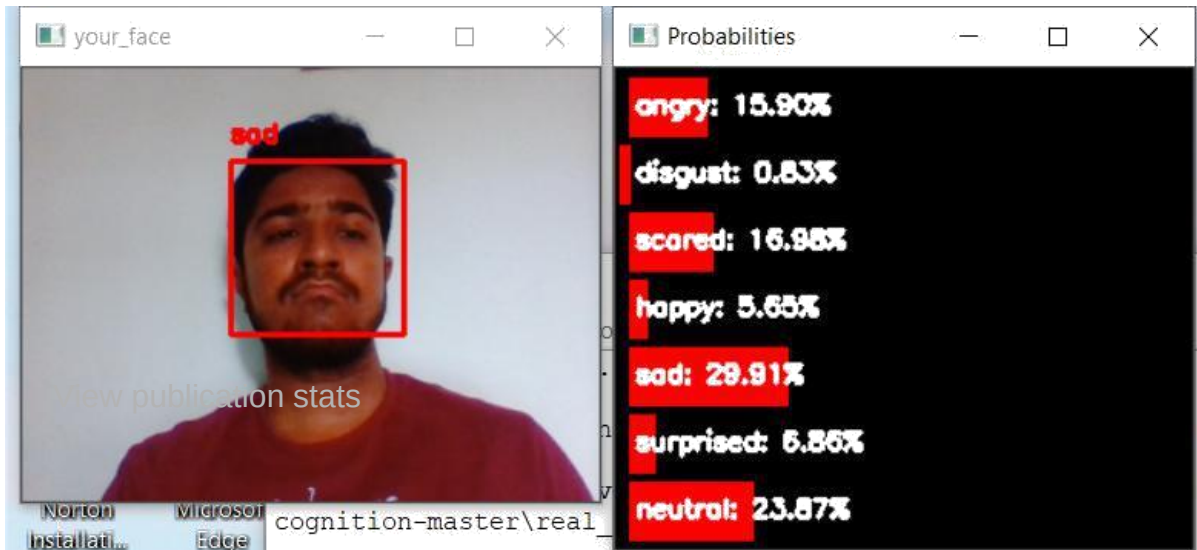
    cv2.imshow('your_face', frame_clone)
    cv2.imshow("Probabilities", canvas)
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

camera.release()
cv2.destroyAllWindows()
```

Emotion Snapshots









Plagiarism Checker X Originality Report

Similarity Found: 9%

Date: Thursday, April 15, 2021

Statistics: 459 words Plagiarized / 5365 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Electronics and Communication Engineering by Jyothi Venkatesh (37130172) Amirisetty Venu(37130021) DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE UNIVERSITY) Acreith GrdeA" bAAC JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119 MARCH 2021 ii SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE UNIVERSITY) Acreith GrdeA" bAAC JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119 www.sathyabama.ac.in DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING BONAFIDE CERTIFICATE This is to certify that this Project Report is the bonafide work of JYOTHI VENKATESH (37130172) and AMIRISETTY VENU (37130021) who carried out the project entitled " DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING " under our supervision from September 2020 to March 2021. Internal Guide DR.G.D ANBARASI JEBASELVI Associate Professor, Department of ECE Head of the Department Dr. T. RAVI, M.E., Ph.D.

Submitted for Viva voce examination held on Internal Examiner External Examiner iii DECLARATION We, J. VENKATESH (37130172) and A. VENU (37130021) hereby declare that the Project Report entitled " DETECTION OF FACIAL EXPRESSION THROUGH MACHINE LEARNING USING IMAGE PROCESSING " done by us under the guidance of DR.G.D ANBARASI JEBASELVI Associate Professor. is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

1) 2) DATE: PLACE: SIGNATURE OF THE CANDIDATES iv ACKNOWLEDGEMENT We are pleased to acknowledge our sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. We are grateful to them. We convey our thanks to Dr. N. M. NANDHITHA, M.E., Ph.D. Dean, School of Electrical and Electronics Engineering and Dr. T. RAVI, M.E., Ph.D. Head of the Department, Department of Electronics and Communication Engineering for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide DR.G.D ANBARASI JEBASELVI Associate Professor, Department of Electronics and Communication Engineering for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of the project work. We wish to express our thanks to all teaching and Non-teaching staff members of the Department of Electronics and Communication Engineering who were helpful in many ways for the completion of the project.

We express our gratitude to our parents for their constant encouragement and support for the completion of the project. i Abstract A Facial demeanor is the obvious sign of the full of feeling state, intellectual action, expectation, character and psychopathology of an individual and assumes an open part in relational relations. Customized affirmation of outward appearances can be a critical section of regular human-machine interfaces; it may in like manner be used in direct science and in clinical practice.A

programmed Facial Expression Recognition framework needs to perform identification and area of countenances in a jumbled scene, facial element extraction, and outward appearance order. Outward appearance acknowledgment framework is executed utilizing Convolution Neural Network (CNN). CNN model of the task depends on LeNet Architecture. Kaggle outward appearance dataset with seven outward appearance names as cheerful, pitiful, shock, dread, outrage, sicken, and nonpartisan is utilized in this undertaking. The framework accomplished 56.77 % exactness and 0.57 accuracy on testing dataset.

Keywords: Outward appearance Recognition, Convolutional Neural Network, Deep Learning, Theano ii Abbreviations CNN Convolution Neural Network FACS Facial Action Coding System FER Facial Expression Recognition GPU Graphics Processing Unit JAFFE Japanese Female Facial Expressions LDA Linear Discriminant Analysis PCA Principle Component Analysis ReLU Rectified Linear Unit SIANN Space Invariant Artificial Neural Network iii Table of Contents

Abstract..... ii

Acknowledgement	iii
Abbreviations	
iv Table of Contents	v List of
Table and Figure	vi 1. Overview
.....	1 1.1
Background.....	1 1.2
Problem definition	6 1.2
Objective.....	
6 1.3 Scope of the Project	6 2.
Literature Review	7 3.
Methodology.....	9 3.1
Dataset 10 3.2 Architecture of CNN	
.....	10 4. Results and
Analysis.....	13 5. Conclusion
.....	
17 References Appendix iv List of Table and Figure	
.....	2 Figure 1.1: A model of CNN
.....	4 Figure 1.2: Max Pooling
.....	9 Figure 3 (a): Training Phase
.....	9 Figure 3 (b): Testing Phase
.....	9 Figure 3.2: Architecture of CNN
.....	11 Figure 4.1 : Cost Function in Training
and	13 Figure 4.2
: Error Rate in Training and Validation	14 Table 4.1:
Confusion matrix for facial expression	14 Table 4.2: Normalized
confusion matrix for facial expression	15 Table 4.3: Precision, Recall and
F1-Score	15 5 1. Overview 1.1 Background A Facial
demeanor is the noticeable indication of the full of feeling state, intellectual action, goal,	
character and psychopathology of an individual and assumes an informative part in	
relational relations.	
Human outward appearances can be effectively grouped into 7 essential feelings:	
cheerful, dismal, shock, dread, outrage, nauseate, and impartial. Our facial feelings are	
communicated through actuation of explicit arrangements of facial muscles. These	
occasionally unpretentious, yet unpredictable, signals in an articulation regularly contain	
a plentiful measure of data about our perspective. Programmed acknowledgment of	
outward appearances can be a significant part of common human-machine interfaces; it	

might likewise be utilized in conduct science and in clinical practice.

It have been read for a significant stretch of time and acquiring the advancement late many years. In spite of the fact that much advancement has been made, perceiving outward appearance with a high precision stays to be troublesome because of the intricacy and assortments of outward appearances [1]. On an everyday nuts and bolts people ordinarily perceive feelings by trademark highlights, shown as a piece of an outward appearance. For example satisfaction is verifiably connected happily or an upward development of the sides of the lips.

Likewise different feelings are portrayed by different disfigurements ordinary to a specific articulation. Examination into programmed acknowledgment of outward appearances tends to the issues encompassing the portrayal and classification of static or dynamic qualities of these misshapenings of face pigmentation [2]. In AI, a convolutional neural organization (CNN, or ConvNet) is a kind of feed-forward fake neural organization where the availability design between its neurons is roused by the association of the creature visual cortex.

Individual cortical neurons react to boosts in a limited area of room known as the responsive field. The responsive fields of various neurons halfway cover with the end goal that they tile the visual field. The reaction of an individual neuron to improvements inside its open field can be approximated numerically by a convolution operation.[3] Convolutional networks were motivated by organic processes[4] what's more, are varieties of multi-facet perceptron intended to utilize insignificant measures of preprocessing.[5] They have wide applications in picture and video acknowledgment, recommender frameworks and common language handling.

The convolutional neural organization is otherwise called move invariant or space invariant fake neural organization (SIANN), which is named dependent on its common loads design and interpretation invariance attributes. Figure 1.1: A model of CNN [5] 6 LeNet is one of the absolute first convolutional neural organizations which moved the field of Deep Learning. This spearheading work by Yann LeCun was named LeNet5 was utilized fundamentally for character acknowledgment assignments, for example, perusing postal divisions, digits, and so on The essential engineering of LeNet can be appeared as beneath [5]: There are four main operations in the Convolution Neural Network shown in Figure 2.2 above: 1. Convolution: The main role of Convolution if there should arise an occurrence of a CNN is to extricate highlights from the info picture.

Convolution safeguards the spatial connection between pixels by learning picture

highlights utilizing little squares of info information. The convolution layer's boundaries comprise of a bunch of learnable channels. Each channel is little spatially (along width and stature), yet reaches out through the full profundity of the info volume. For instance, a regular channel on a first layer of a CNN may have size 3x5x5 (for example pictures have profundity 3 for example the shading channels, 5 pixels width and stature).

During the forward pass, each channel is convolved across the width and stature of the information volume and process dab items between the sections of the channel and the contribution at any position. As the channel convolve over the width and tallness of the info volume it creates a 2-dimensional enactment map that gives the reactions of that channel at each spatial position. Naturally, the organization will learn channels that enact when they see some kind of visual component like an edge of some orientation or a blotch of some color on the first layer, or eventually entire honeycomb or wheel-like patterns on higher layers of the network. Now, there will be an entire set of filters in each convolution layer (e.g. 20 filters), and each of them will produce a separate 2-dimensional activation map.

The 2-dimensional convolution between image A and Filter B can be given as: $(A * B)(i, j) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} A(k, l) B(i-k, j-l)$ (2.1) where size of A is $(M \times N)$, size of B is $(M_b \times N_b)$, $0 \leq i < M - M_b + 1$, $0 \leq j < N - N_b + 1$. A channel convolves with the info picture to deliver a component map. The convolution of another channel over a similar picture gives an alternate element map. Convolution activity catches the nearby conditions in the first picture.

A CNN learns the estimations of these channels all alone during the preparation interaction (in spite of the fact that boundaries like number of channels, channel size, design of the organization and so forth actually expected to indicate before the preparation cycle). The more number of channels, the more picture highlights improve network becomes at perceiving designs in inconspicuous pictures. The size of the Feature Map (Convolved Feature) is obliged by three limits

- Depth: Depth identifies with the amount of channels we use for the convolution action.

- Stride: Stride is the size of the channel, assuming the size of the channel is 5x5, step is 5.
- Zero-cushioning: Sometimes, it is advantageous to cushion the information lattice with zeros around the boundary, so that channel can be applied to lining components of information picture network. Utilizing zero cushioning size of the component guide can be controlled.
- 2. Rectified Linear Unit: An extra activity called ReLU has been utilized after each Convolution activity.

A Rectified Linear Unit (ReLU) is a cell of a neural organization which utilizes the

accompanying initiation capacity to figure its yield given x : $R(x) = \text{Max}(0, x)$ (2.2) Utilizing these cells is more productive than sigmoid and still advances more data contrasted with parallel units. While instating the loads consistently, half of the loads are negative. This helps making an inadequate element portrayal. Another positive viewpoint is the moderately modest calculation. No remarkable capacity must be determined.

This capacity likewise forestalls the disappearing slope blunder, since the angles are direct capacities or zero yet for no situation non-straight capacities. 3. Pooling (sub-sampling) Spatial Pooling (likewise called subsampling or downsampling) diminishes the dimensionality of each element map however holds the main data. Spatial Pooling can be of various kinds: Max, Average, Sum and so on. If there should arise an occurrence of Max Pooling, a spatial area (for instance, a 2×2 window) is characterized and the biggest component is taken from the corrected element map inside that window. If there should be an occurrence of normal pooling the normal or amount of all components in that window is taken. Practically speaking, Max Pooling has been appeared to work better.

Max Pooling reduces the commitment by applying the most limit over the data x_i . Let m be the size of the channel, by then the yield figures as follows: $M() = \{ +, + \mid = /2, \mid = /2, N \}$ (2.3) Figure 1.2 : Max Pooling 8 The capacity of Pooling is to continuously diminish the spatial size of the info portrayal. Specifically, pooling • Makes the info portrayals (include measurement) more modest and more reasonable • Reduces the quantity of boundaries and calculations in the organization, in this manner, controlling over-fitting • Makes the organization invariant to little changes, twists and interpretations in the information picture (a little contortion in info won't change the yield of Pooling. • Helps us show up at a nearly scale invariant portrayal.

This is incredible since articles can be recognized in a picture regardless of where they are found. 4. Classification (Multilayer Perceptron): The Fully Connected layer is a conventional Multi-Layer Perceptron that utilizes a softmax actuation work in the yield layer. The expression "Completely Connected" suggests that each neuron in the past layer is associated with each neuron on the following layer. The yield from the convolutional and pooling layers address undeniable level highlights of the information picture.

The motivation behind the Fully Connected layer is to utilize these highlights for characterizing the info picture into different classes dependent on the preparation dataset. Softmax is utilized for actuation work. It regards the yields as scores for each class. In the Softmax, the capacity planning remained unaltered and these scores are deciphered as the un-standardized log probabilities for each class. Softmax is

determined as: $(j) = (j) / (2.4) = 1 / (j)$ where j is file for picture and K is number of all outward appearance class. Aside from characterization, adding a completely associated layer is likewise a (normally) modest method of learning non-straight mixes of these highlights.

A large portion of the highlights from convolutional and pooling layers might be useful for the arrangement task, however blends of those highlights may be far and away superior. The amount of yield probabilities from the Fully Connected Layer is 1. This is guaranteed by utilizing the as the enactment work in the yield layer of the Fully Connected Layer. The Softmax work takes a vector of self-assertive genuine esteemed scores and squashes it to a vector of qualities 9 somewhere in the range of nothing and one that entirety to one. 10 1.2

Problem definition Human feelings and expectations are communicated through outward appearances and inferring a productive and viable element is the principal part of outward appearance framework. Outward appearances pass on non-verbal signals, which assume a significant part in relational relations. Programmed acknowledgment of outward appearances can be a significant part of common human-machine interfaces; it might likewise be utilized in social science and in clinical practice.

A programmed Facial Expression Recognition framework needs to tackle the accompanying issues: identification and area of countenances in a jumbled scene, facial element extraction, and outward appearance arrangement. 1.2 Objective The objective of the project is: 1. To implement Convolutional Neural Networks for classification of facial expressions. 1.3 Scope of the Project In this task outward appearance acknowledgment framework is executed utilizing convolution neural organization.

Facial pictures are characterized into seven outward appearance classes specifically Anger, Disgust, Fear, Happy, Sad, Surprise and 'Nonpartisan. Kaggle dataset is utilized to prepare and test the classifier. 11 2. Literature Review Two unique methodologies are utilized for outward appearance acknowledgment, the two of which incorporate two distinct techniques, exist [6]. Partitioning the face into discrete activity units or saving it in general for additional preparing seems, by all accounts, to be the first and the essential differentiation between the fundamental methodologies.

In both of these methodologies, two unique procedures, in particular the 'Mathematical based' and the 'Appearance-based' definitions, can be utilized. Utilizing the entire front facing face picture and handling it to wind up with the characterizations of 6 all inclusive outward appearance models: nauseate, dread, happiness, shock, pitiful ness and outrage; diagrams the main methodology. Here, it is accepted that every one of the

previously mentioned feelings have trademark appearances on face and that is the reason acknowledgment of them is fundamental and adequate.

Rather than utilizing the face pictures in general, isolating them into some sub-segments for additional handling structures up the principle thought of the second methodology for outward appearance investigation. As articulation is more related with inconspicuous changes of some discrete highlights like eyes, eyebrows and lip corners; these fine-grained changes are utilized for breaking down mechanized acknowledgment. There are two principle strategies that are utilized in both of the above clarified approaches.

Mathematical Based Parameterization is an old way which comprises of following and handling the movements of certain spots on picture arrangements, first and foremost introduced by Suwa et al to perceive outward appearances [7]. Cohn and Kanade later on attempted mathematical displaying and following of facial highlights by guaranteeing that every AU is given a particular arrangement of facial muscles [8]. The disservices of this technique are the forms of these highlights and parts must be changed physically in this casing, the issues of strength and troubles turn out in instances of posture and light changes while the following is applied on pictures, as activities and articulations will in general change both in morphological and in dynamical faculties, it turns out to be difficult to appraise general boundaries for development and removal.

Consequently, winding up with powerful choices for facial activities under these differing conditions becomes to be troublesome. Maybe than following spatial focuses and utilizing situating and development boundaries that shift inside time, shading (pixel) data of related areas of face are prepared in Appearance Based 12 Parameterizations; to get the boundaries that will frame the component vectors. Various highlights, for example, Gabor, Haar wavelet coefficients, along with include extraction and determination techniques, for example, PCA, LDA, and Adaboost are utilized inside this system.

For characterization issue, calculations like Machine learning, Neural Network, Support Vector Machine, Deep learning, Naive Bayes are utilized. Raghuvanshi A. et al have assembled a Facial demeanor acknowledgment framework upon late examination to group pictures of human appearances into discrete feeling classes utilizing convolutional neural organizations [9]. Alizadeh, Shima, and Azar Fazel have created Facial Expression Recognition framework utilizing Convolutional Neural Networks dependent on Torch model [10].

13 Raw Image Normalization CNN Train CNN Weights CNN Weights Raw Image Normalization CNN Facial Expression 3. Methodology The facial expression recognition system is implemented using convolutional neural network. The block diagram of the system is shown in following figures: Figure 3 (a): Training Phase Figure 3 (b): Testing Phase During preparing, the framework got a preparation information involving grayscale pictures of appearances with their individual articulation name and learns a bunch of loads for the organization. The preparation step took as info a picture with a face. From that point, a power standardization is applied to the picture. The standardized pictures are utilized to prepare the Convolutional Network.

To guarantee that the preparation execution isn't influenced by the request for introduction of the models, approval dataset is utilized to pick the last best arrangement of loads out of a bunch of trainings performed with tests introduced in various orders. The yield of the preparation step is a bunch of loads that accomplish the best outcome with the preparation information. During test, the framework got a grayscale picture of a face from test dataset, and yield the anticipated appearance by utilizing the last organization loads mastered during preparing.

Its yield is a solitary number that addresses one of the seven essential articulations. 14 3.1 Dataset The dataset from a Kaggle Facial Expression Recognition Challenge (FER2013) is utilized for the preparation and testing. It contains pre-trimmed, 48-by-48-pixel grayscale pictures of countenances each marked with one of the 7 feeling classes: outrage, appall, dread, joy, misery, shock, and unbiased. Dataset has preparing set of 35887 facial pictures with outward appearance marks.. The dataset has class awkwardness issue, since certain classes have huge number of models while some has not many.

The dataset is adjusted utilizing oversampling, by expanding numbers in minority classes. The decent dataset contains 40263 pictures, from which 29263 pictures are utilized for preparing, 6000 pictures are utilized for testing, and 5000 pictures are utilized for approval. Figure 3.1: Training, Testing and Validation Data distribution 3.2 Architecture of CNN A normal design of a convolutional neural organization contains an information layer, some convolutional layers, some completely associated layers, and a yield layer. CNN is planned with some alteration on LeNet Architecture [10].

It has 6 layers without thinking about info and yield. The design of the Convolution Neural Network utilized in the venture is appeared in the accompanying figure. 7000 6000 5000 4000 3000 2000 1000 0 Angry Disgust Fear Happy Sad Surprise Neutral Facial Expression Training Testing Validation Number of Images 15 Figure 3.2: Architecture of CNN (Modified from LeNet Architecture) 1. Input Layer: The information layer has

pre-decided, fixed measurements, so the picture should be pre-handled before it very well may be taken care of into the layer.

Standardized dark scale pictures of size 48 X 48 pixels from Kaggle dataset are utilized for preparing, approval and testing. For testing propose PC webcam pictures are additionally utilized, in which face is distinguished and edited utilizing OpenCV Haar Cascade Classifier and standardized. 2. Convolution and Pooling (ConvPool) Layers: 3. Convolution and pooling is done dependent on group handling. Each clump has N pictures and CNN channel loads are refreshed on those clusters. Every convolution layer takes picture group contribution of four measurement $N \times \text{Color-Channel} \times \text{width} \times \text{stature}$.

Highlight guide or channel for convolution are likewise four dimensional (Number of highlight maps in, number of highlight delineates, channel width, channel stature). In every convolution layer, four dimensional convolution is determined between picture bunch and highlight maps. After convolution just boundary that change is picture width and stature. 4. 5. New picture width = old picture width – channel width + 1 New picture stature = old picture tallness – channel tallness + 1 6. 7. After every convolution layer downsampling/subsampling is accomplished for dimensionality decrease.

This cycle is called Pooling. Max pooling and Average Pooling are two celebrated pooling technique. In this venture max pooling is done after convolution. Pool size of (2x2) is 8. 9. taken, what parts the picture into matrix of squares every one of size 2x2 and takes limit of 4 pixels. Subsequent to pooling just stature and width are influenced. 10. 11. Two convolution layer and pooling layer are utilized in the design. From the outset convolution layer size of info picture clump is $N \times 1 \times 48 \times 48$. Here, size of picture group is N, number of shading channel is 1 and both picture stature and width are 48 pixel.

Convolution with include guide of $1 \times 20 \times 5 \times 5$ outcomes picture group is of size $N \times 20 \times 44 \times 44$. After convolution pooling is finished with pool size of 2x2, which results picture cluster of size $N \times 20 \times 22 \times 22$. This is trailed by second convolution layer with include guide of $20 \times 20 \times 5 \times 5$, which results picture cluster of size $N \times 20 \times 18 \times 18$. This is trailed by pooling layer with pool size 2x2, which results picture group of size $N \times 20 \times 9 \times 9$. 16 12. Fully Connected Layer This layer is roused by the manner in which neurons send signals through the cerebrum. It takes countless info includes and change highlights through layers associated with teachable loads.

Two secret layers of size 500 and 300 unit are utilized in completely associated layer. The loads of these layers are prepared by forward engendering of preparing information

then in reverse spread of its mistakes. Back engendering begins from assessing the distinction among expectation and genuine worth, and back ascertains the weight change expected to each layer previously. We can handle the preparation speed and the intricacy of the design by tuning the hyper-boundaries, for example, learning rate and organization thickness. Hyper-boundaries for this layer incorporate learning rate, force, regularization boundary, and rot.

The yield from the second pooling layer is of size $N \times 20 \times 9 \times 9$ and contribution of first secret layer of completely associated layer is of size $N \times 500$. In this way, yield of pooling layer is leveled to $N \times 1620$ size and took care of to first secret layer. Yield from first secret layer is taken care of to second secret layer. Second secret layer is of size $N \times 300$ and its yield is taken care of to yield layer of size equivalent to number of outward appearance classes. 13. Output Layer Yield from the second secret layer is associated with yield layer having seven unmistakable classes. Utilizing Softmax initiation work, yield is acquired utilizing the probabilities for every one of the seven class.

The class with the most elevated likelihood is the anticipated class. 17 4. Results and Analysis CNN design for outward appearance acknowledgment as referenced above was executed in Python. Alongside Python programming language, Numpy, Theano and CUDA libraries were utilized. Preparing picture bunch size was taken as 30, while channel map is of size $20 \times 5 \times 5$ for both convolution layer. Approval set was utilized to approve the preparation cycle. In last bunch of each age in approval cost, approval blunder, preparing cost, preparing mistake are determined. Information boundaries for preparing are picture set and comparing yield names.

The preparation interaction refreshed the loads of highlight maps and covered up layers dependent on hyper- boundaries like learning rate, energy, regularization and rot. In this framework clump insightful learning rate was utilized as $10e-5$, force as 0.99, regularization as $10e-7$ and rot as 0.99999. The correlation of approval cost, approval mistake, preparing cost, preparing blunder are appeared in figures underneath. Figure 4.1 : Cost function in training and validation 18 Figure 4.2 : Error Rate in Training and Validation Above chart shows that the in 50 ages preparing cost and mistake rate is diminished 0.35 and 0.1

individually while approval cost and blunder is 1.25 and 0.45 separately. The testing of the model is done utilizing 6000 pictures. The classifier gave 56.77 % precision. The disarray framework for seven outward appearance classes is appeared beneath: Table 4.1 : Confusion matrix for facial expression recognition

	Pre dict ed	Act ual	An ger	Dis gus t	Fea r	Ha pp y	Sad	Su rp ris e	Ne utr al
Anger	31	2	16	93	89	97	27	10	6
Disgust	35	70	0	0	0	0	0	0	0
Fear	12	3	7	30	0	69	10	0	67
Happy	85	5	74	88	7	80	29	12	8
Sad	13	7	6	12	2				

89 33 8 27 16 5 Surprise 27 4 75 45 24 40 2 46 Neutral 11 4 4 91 11 6 12 3 32 46 7 19
Table 4.2 : Normalized confusion matrix for facial expression recognition

	Actual	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	0.42	1622	0.0	217	69	0.1	187	74
Disgust	0.06	9099	0.10	9729	0.04	3339	0.11	193
Fear	0.04	7619	0.9	523	8	0	0	0
Happy	0.16	6216	0.0	095	24	0.3	831	42
Sad	0.05	3571	0.11	3122	0.10	7544	0.12	3548
Surprise	0.11	4865	0.0	068	03	0.0	945	08
Neutral	0.68	8665	0.09	0498	0.04	6549	0.13	5164
Average	0.18	5135	0.0	081	63	0.1	558	11
Overall	0.06	9099	0.38	2353	0.04	3339	0.17	4234
Surprise	0.03	6486	0.0	054	42	0.0	957	85
Neutral	0.03	4938	0.02	7149	0.64	5265	0.04	8574
Overall	0.15	4054	0.0	054	42	0.1	162	2
Overall	0.09	0062	0.1	391	4	0.05	1364	0.49

Above tables showed that this model most noteworthy precision for sicken feeling with 95.23 %, trailed by content with 68.86 %, shock with 64.52 %, impartial with 49.31 %, outrage with 42.16 %, dread with 38.31 % and most reduced exactness for tragic feeling as 38.23 %. The precision, recall and F1-score of each facial expression class is shown below: Table 4.3 : Precision, Recall and F1-Score

	Precision	Recall	F1-score
Anger	0.3	9	0.4
Disgust	0.9	0.4	2
Fear	0.4	5	0.3
Happy	0.6	8	0.6
Sad	0.4	4	0.3
Surprise	0.6	9	0.6
Neutral	0.4	5	0.4
Average	0.5	7	0.5

The general exactness and review are 0.57 and 0.57 individually. The model performs truly well on characterizing positive feelings bringing about generally high accuracy scores for glad and shocked. Nauseate has most noteworthy accuracy and review as 0.95 and 0.99 as pictures in this class were oversampled to address class unevenness. Cheerful has an accuracy of 0.68 and review of 0.69 which could be clarified by having the most models (6500) in the preparation set. Curiously, shock has an exactness of 0.69 and review of 0.65 having minimal models in the preparation set. There should be extremely solid signs in the unexpected articulations. Model execution appears to be more fragile across negative feelings by and large.

In especially, the feeling miserable has a low exactness of just 0.44 and review 0.38. The model oftentimes misclassified irate, dread and unbiased as miserable. Furthermore, it is most befuddled while anticipating pitiful and impartial countenances in light of the fact that these two feelings are likely the most un-expressive (barring crying appearances). The generally F1-score is additionally 0.57 . F1-score is most elevated for sicken due to oversampling of pictures. Cheerful and shock have higher F1-score as 0.69 and 0.67 separately. Dread has least F1-score as 0.39 and dismal, outrage and unbiased additionally have low F1-score.

CNN Classifier is then used to order picture taken from webcam in Laptop. Face is recognized in webcam outlines utilizing Haar course classifier from OpenCV. At that point distinguished face is edited and standardized and taken care of to CNN Classifier. Some grouping results utilizing webcam are recorded in Appendix A. 21 5. Summary 5.1

Conclusion In this task, a LeNet engineering based six layer convolution neural organization is executed to arrange human outward appearances for example cheerful, pitiful, shock, dread, outrage, appall, and unbiased.

The framework has been assessed utilizing Accuracy, Precision, Recall and F1-score. The classifier accomplished exactness of 56.77 % , accuracy of 0.57, review 0.57 and F1-score 0.57. 5.2 Enhancement Later on work, the model can be reached out to shading pictures. This will permit to examine the viability of pre-prepared models, for example, AlexNet[11] or VGGNet [12] for facial feeling acknowledgment. 22 References [1] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE. [2] Chibelushi, C. C., & Bourel, F. (2003).

Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9. [3] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 31 August 2013. [4] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555 – 559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 17 November 2013. [5] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013 [6] CFpressin recoion," Mastethesis, University, 2008. [7] Suwa, M.; Sugie N. and Fujimora K.

A Preliminary Note on Pattern Recognition of Human Emotional Expression, Proc. International Joint Conf, Pattern Recognition, pages 408-410, 1978 [8] Recognizing action units for facial expression analysis YI Tian, T Kanade, JF Cohn IEEE Transactions on pattern analysis and machine intelligence 23 (2), 97- 115 [9] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford University, 2016 [10] Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition."

Stanford University, 2016 [11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. "a classificational neural networks in information processing systems. 2012. [12] Simonyan, Karen, and Andrew Zisserman. "y deep convolutional networks for large-scale image recognition" arXiv preprint arXiv: (2014). Appendix A CNN classifier result using webcam 1. Examples with Correct Outputs 2. Examples with Incorrect Output a) Neutral expression is incorrectly recognized as surprise b) Anger expression is incorrectly recognized as surprise c) Sad expression is incorrectly recognized as anger 21 View publication stats

INTERNET SOURCES:

<1% -

<https://www.scribd.com/document/446002605/Souvenir-Volume-8-Issue-1S4-June-2019-pdf>

<1% -

https://www.researchgate.net/publication/284020001_Fabrication_of_Humanoid_robot_and_Implementation_of_Gaits

<1% -

<https://www.coursehero.com/file/74249805/Machine-Learning-training-report-karan-dhawan-ecedocx/>

<1% -

<https://dokumen.pub/progress-in-advanced-computing-and-intelligent-engineering-proceedings-of-icacie-2019-volume-2-1st-ed-9789811563522-9789811563539.html>

<1% -

<https://www.ssn.edu.in/college-of-engineering/electronics-and-communication-engineering-department-ssn-institutions/>

<1% -

https://www.researchgate.net/publication/329844331_Project_Report_on_STUDY_OF_THE_EFFECTS_OF_STENTER_MACHINE_ON_KNIT_FABRICS_PROPERTIES_LETTER_OF_APPROVAL_ASSISTANT_PROFESSOR_DEPARTMENT_OF_TEXTILE_ENGINEERING_FACULTY_OF_ENGINEERING_DAFFODIL_INTERNATIO

<1% -

<https://studylib.net/doc/8833318/self-study-report---egs-pillay-engineering-college>

<1% - <https://www.coursehero.com/file/87866421/CNF-351-Project-report-sampledocx/>

<1% - <https://core.ac.uk/download/pdf/53188325.pdf>

<1% - <https://irjet.academia.edu/IRJET?swp=tc-au-35628387>

1% -

https://www.researchgate.net/profile/Deepesh-Lekhak/publication/322103128_Facial_Expression_Recognition_System_using_Convolutional_Neural_Network/links/5a453623458515f6b0547796/Facial-Expression-Recognition-System-using-Convolutional-Neural-Net-work.pdf

1% - [https://nvbdcp.gov.in/Doc/JE-AES-Prevention-Control\(NPPCJA\).pdf](https://nvbdcp.gov.in/Doc/JE-AES-Prevention-Control(NPPCJA).pdf)

<1% - <https://www.coursehero.com/file/87048285/CHEM-43docx/>

<1% -

https://afni.nimh.nih.gov/pub/dist/doc/html/doc/background_install/install_instructs/overview.html

<1% -

<https://era.library.ualberta.ca/items/4f8f4562-28f7-4e6d-9fdd-bd0d83ddcd91/download/336677f8-fc78-491a-a8a0-ddd6f37fd817>

<1% - https://www.london.gov.uk/sites/default/files/draft_hra_report.pdf
 <1% - <https://core.ac.uk/download/pdf/161888987.pdf>
 <1% - <https://www.atlantis-press.com/journals/ijcis/125945494/view>
 1% - http://www.ijtimes.com/papers/special_papers/AESMT17.pdf
 <1% -
https://www.researchgate.net/publication/349162145_PLANT_DISEASE_DETECTION_USING_CONVOLUTIONAL_NEURAL_NETWORK
 <1% - <https://acadpubl.eu/hub/2018-119-17/4/458.pdf>
 <1% -
https://www.researchgate.net/publication/318960019_Review_of_Topical_Enlargement_in_Laser_Beam_Machining
 <1% - <https://www.sciencedirect.com/science/article/pii/S2589871X20300176>
 <1% - <https://patents.justia.com/patent/20190361454>
 <1% -
<https://medium.com/@rathna211994/convolution-neural-networks-cnn-b6fe90214b1e>
 <1% - <https://www.sciencedirect.com/science/article/pii/S0377221719307581>
 <1% - https://www.ripublication.com/ijaer19/ijaerv14n3_24.pdf
 <1% - <https://www.sciencedirect.com/science/article/pii/B9780128222263000076>
 <1% - <https://www.sciencedirect.com/science/article/pii/S0950061819328193>
 <1% -
<https://dokumen.pub/probabilistic-deep-learning-with-python-1nbsped-1617296074-9781617296079.html>
 <1% -
<https://dokumen.pub/intelligent-data-engineering-and-automated-learning-ideal-2019-20th-international-conference-manchester-uk-november-1416-2019-proceedings-part-i-1st-ed-2019-978-3-030-33606-6-978-3-030-33607-3.html>
 <1% - <https://www.aclweb.org/anthology/events/acl-2018/>
 <1% -
<https://www.datacamp.com/community/tutorials/convolutional-neural-networks-python>
 2% -
https://www.researchgate.net/publication/322103128_Facial_Expression_Recognition_System_using_Convolutional_Neural_Network
 <1% - http://tasbeha.org/content/hh_books/svnwrds/
 <1% - https://issuu.com/tarp/docs/notnature_finaldraft_041012
 <1% - <http://www.me.cs.scitec.kobe-u.ac.jp/~takigu/pdf/2017/W167.pdf>
 <1% -
<https://dokumen.pub/evolving-technologies-for-computing-communication-and-smart-world-proceedings-of-etccs-2020-1st-ed-9789811578038-9789811578045.html>
 <1% -
<https://dokumen.pub/soft-computing-techniques-and-applications-proceeding-of-the-i>

nternational-conference-on-computing-and-communication-ic3-2020-1st-ed-97898115
73934-9789811573941.html

<1% - <https://www.sciencedirect.com/science/article/pii/S1389128621000438>

<1% - <http://www.acadpubl.eu/hub/2018-119-15/2/391.pdf>

1% - <https://www.worldscientific.com/doi/abs/10.1142/S0219843619410020>