RIPHAH
INTERNATIONAL UNIVERISTY
Department Of MATHEMATICS

Convolutional Neural Network

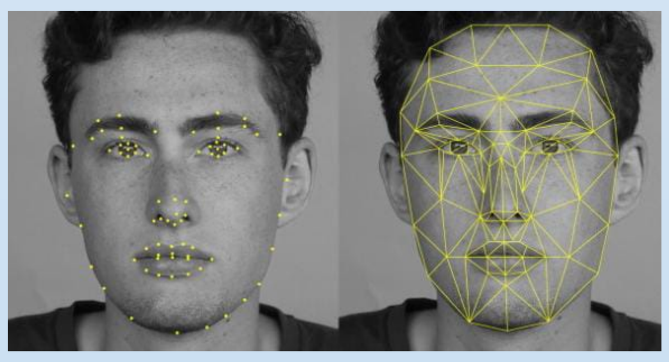
Facial Emotion Recognition

Rabia Riaz, Fiza Akram , Umm-e- Farwa, Kiran Iqbal
Email: rabiariaz.rr786@gmail.com, fizaakram87@gmail.com, ummefarwa0283@gmail.com, iqbalkiran043@gmail.com
Supervised by: M. Saqib Khan
Department of Mathematics, Riphah International University Islamabad (Lahore Campus)
2nd Workshop on Advancement of Mathematics & it's Applications(WAMA-2024)

CNNs: Facial Expression Recognition via Deep Learning

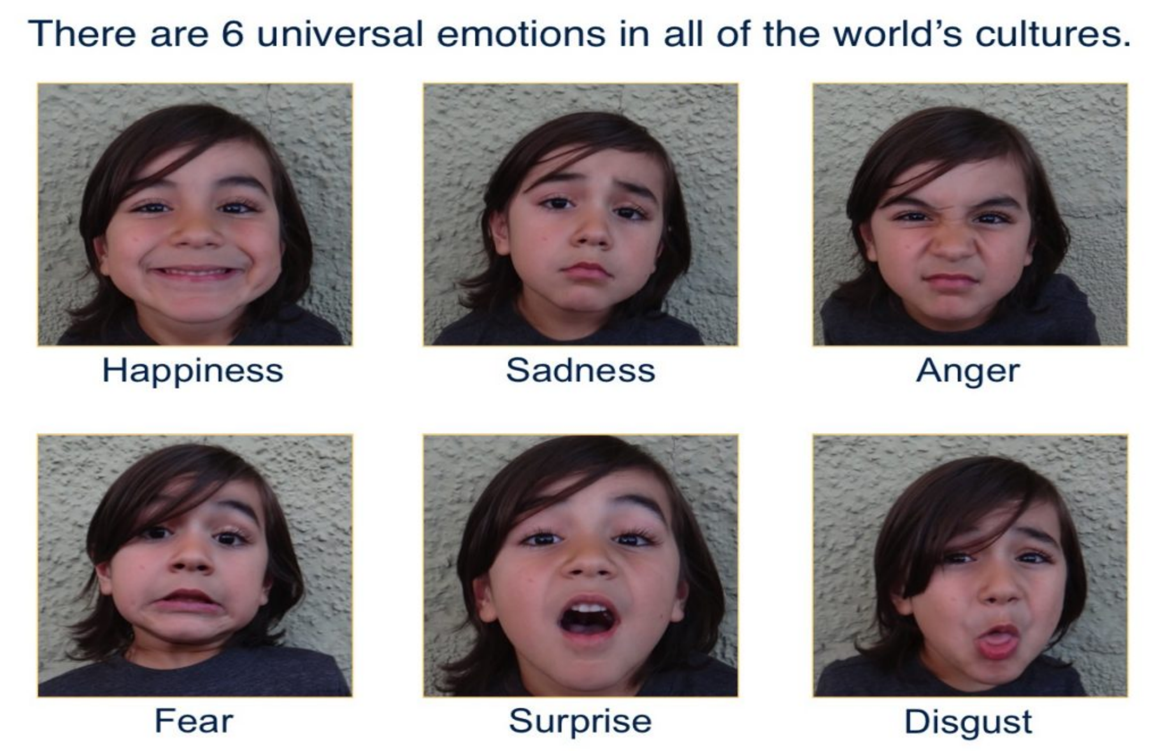
INTRODUCTION

Convolutional Neural Networks (CNNs) are a type of deep learning algorithm that is primarily used for image processing and recognition tasks. CNNs consist of multiple layers, each of which performs a specific operation to transform the input data.
The key component of CNNs is the convolutional layer, which applies convolution operations to the input image. These convolution operations involve sliding a small matrix called a kernel or filter over the input image to perform element-wise multiplication and then summing up the results to produce a feature map.



OBJECTIVE

➤ Our goal is to develop a neural network to recognize facial expressions and classify them into one of seven emotions - happy, sad, disgusted, surprised, angry, fearful and neutral.
➤ By identifying these emotions, we will then generate content in the style of a news feed.



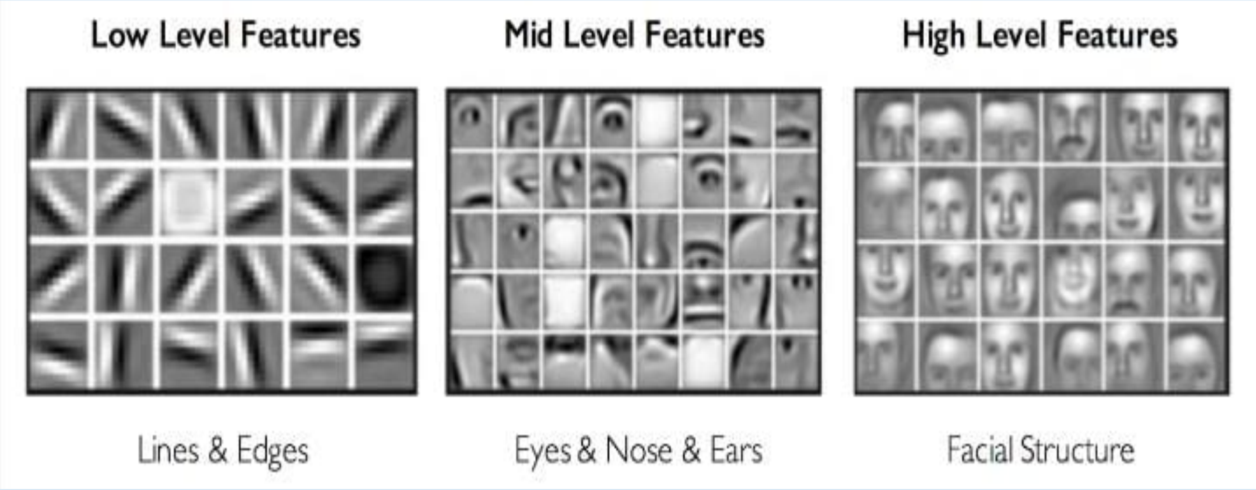
CHALLENGES

➤ Class Imbalance Problem.
➤ Moreover, a feedforward network generally predicts the same emotion all the time.
➤ Another issue is that images have to be well illuminated. Low light / highly exposed images produce poor results.

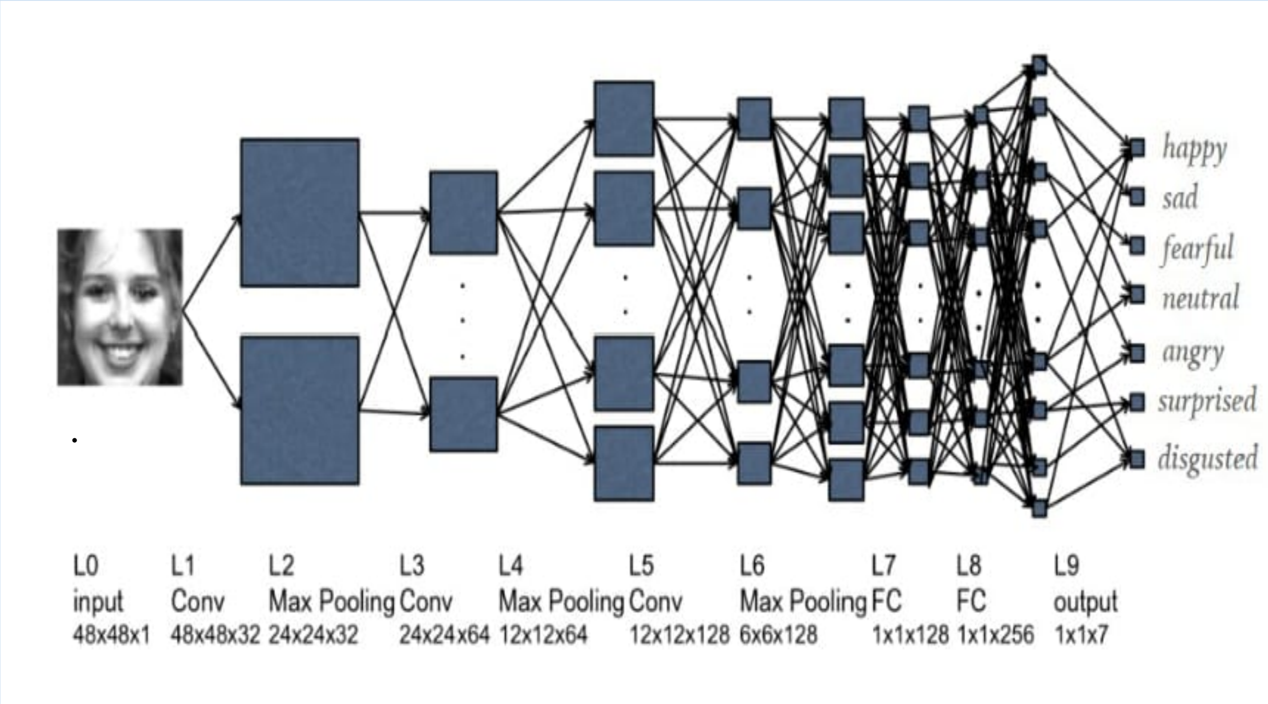
ADRESSING THE CHALLENGES

➤ We have used a deep neural network - a Convolutional Neural Network which is capable of overcoming this problem by spatial locality - detecting edges and extracting certain features.

➤ Plus, our model uses Haar Cascades to detect faces. A pre-trained cascade of classifiers that can detect faces. This addresses most low light issueș.

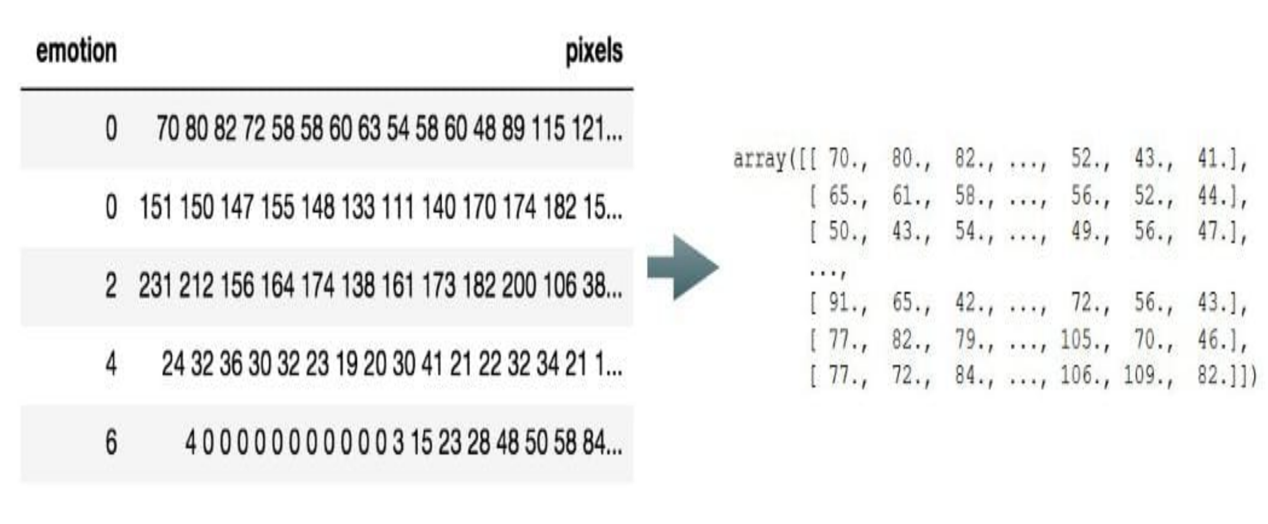


PROPOSED ARCHITECTURE



PREPROCESSING THE DATA

➤ The initial dataset was split into two sections - a string of 2304 numbers indicating pixel values for the image and a number from 1-7 indicating the emotion.
➤ We converted the string of numbers into a 48x48 matrix to feed into the neural network.



emotion	pixels
0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...
0	151 150 147 155 148 133 111 140 170 174 182 15...
2	231 212 156 164 174 138 161 173 182 200 106 38...
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...
6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...

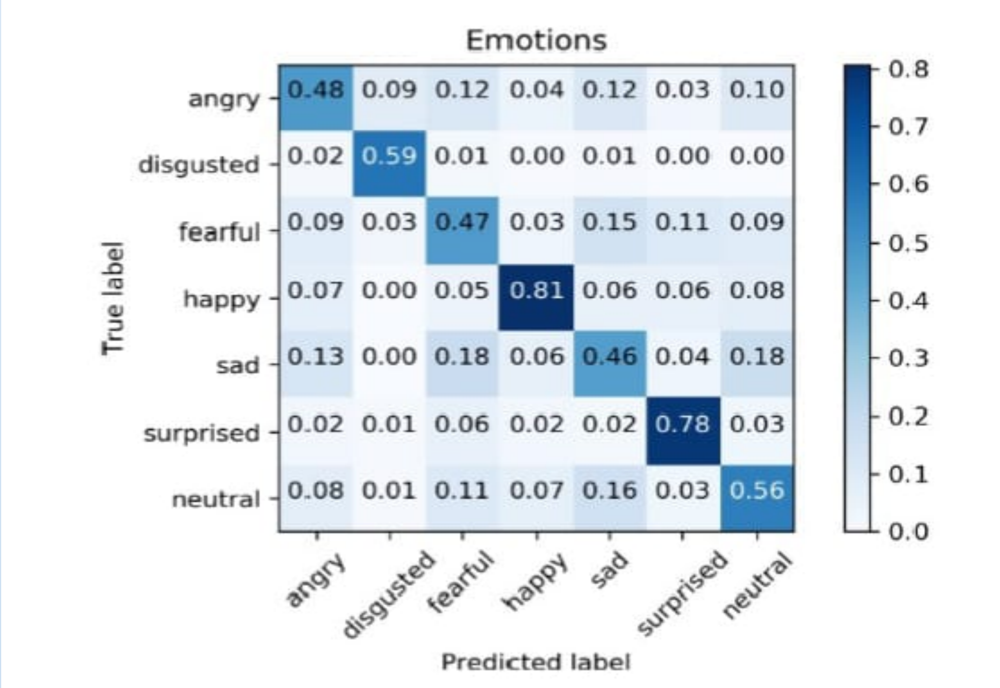
```
array([[ 70.,  80.,  82., ...,  52.,  43.,  41.],
       [ 65.,  61.,  58., ...,  56.,  52.,  44.],
       [ 50.,  43.,  54., ...,  49.,  56.,  47.],
       ...,
       [ 91.,  65.,  42., ...,  72.,  56.,  43.],
       [ 77.,  82.,  79., ..., 105.,  70.,  46.],
       [ 77.,  72.,  84., ..., 106., 109.,  82.]])
```

APPROACHES

Model	Batch Size	Optimizer	Epochs	Accuracy
Feed Forward	128	RMSProp	10	17.386
Simple CNN	128	RMSProp	10	24.728
Decision Tree	40	-	-	30.843
Model #1	96	RMSProp	100	57.397
Model #2	64	SGD	10	55.900
Model #3	128	Adam	20	60.587

CONFUSION MATRIX

➤ Adam optimizer gave the highest overall accuracy.
➤ It does drastically better for faces showing disgust and fearful - almost 30%



COMPARISON WITH OTHER SYSTEMS

➤ Similar systems that made use of emotion analysis did so on static images and not on real-time video streams.
➤ Some such models made use of using basic machine learning techniques such as Support Vector machine and Linear Discriminant Analysis, in combination with regular neural networks.
➤ A disadvantage of these systems is that they take a long time to train and their predictions are not instantaneous as required by a real-time system.
➤ These systems take a long time to train because of the complexity of the data and the network itself.
➤ Owing to the low computational requirements and faster training and prediction time, our model can be further developed for mobile use.

CONCLUSION

➤ Thus we conclude that a real-time system in which emotions can be detected is feasible and generating content based on these emotions is a viable proposal.
➤ Furthermore, the established average accuracy of 60% is competent considering the complex nature of a human face and the real-time constraints.

REFERENCES

➤ [1] Albert Mehrabian. Silent Messages, University of California Los Angeles, 1971.
➤ [2] P. Ekman and W. V. Friesen. Emotional facial action coding system. Unpublished manuscript, University of California at San Francisco, 1983.
➤ [3] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
➤ [4] Thai Hoang Le. Applying artificial neural networks for face recognition. Advances in Artificial Neural Systems, 2011:15, 2011.