**Artificial Intelligence Techniques PG (6685)**



**Emotion Recognition using image classification**

Submitted By:

Birat Jung Thapa (u3245672)

Prahlad Panthi (u3248319)

Table of Contents

[1. Abstract 3](#_Toc165641354)

[2. Introduction 3](#_Toc165641355)

[3. Literature Review 4](#_Toc165641356)

[4. Methodology 5](#_Toc165641357)

[4.1. Dataset Description 5](#_Toc165641358)

[4.2. Algorithms Used 5](#_Toc165641359)

[4.2.1. HaarCascade 5](#_Toc165641360)

[4.2.2. Convolution Neural Network 6](#_Toc165641361)

[4.3. Steps Involved 6](#_Toc165641362)

[4.4. Outcome Achieved 9](#_Toc165641363)

[5. Conclusion 10](#_Toc165641364)

[6. References 12](#_Toc165641365)

# **Abstract**

The paper aims to leverage computer vision techniques to classify seven distinct emotions: Angry, Disgust, Fearful, Happy, Sad, Surprised, and Neutral, from facial expressions. This classification can have wide-ranging applications across human-computer interactions, healthcare, education, and other sectors, potentially improving understanding and support for emotional states in various contexts. A model is developed by utilizing Convolution Neural Network that detects emotions in real time. The trained model along with Haar Cascade Algorithm uses webcam to identify faces and detect emotions. The results are displayed via a graphical interface.

# **Introduction**

Emotions are pivotal in human interaction, revealing themselves through shifts in both physical and mental behavior. Some cues like facial expressions, actions, and speech serve as signs of emotional states, which machines can scrutinize through mediums such as images and videos. Face detection lays the groundwork for a broad spectrum of facial analysis techniques, which empowers computers to go deeper into human sentiments. Emotion classification typically consists of extracting features, refining them through filtering and transformation, and ultimately categorizing them using trained algorithms and datasets.(Ashu, et al., 2018)

This project primarily focuses on CNN for emotion classification along with the HaarCascade classifier for face detection. Haar feature-based Cascade classifiers are best at object detection, including facial recognition. They employ a cascade structure of classifiers to efficiently identify objects within images, effectively outlining the facial region when faces are detected. Whereas, deep neural networks like CNN, have sparked significant advancements across various domains, notably in computer vision. CNNs demonstrate the capacity to identify patterns and features directly from data, rendering them potent tools for tasks like facial recognition and emotion classification. While CNNs and HaarCascade classifiers share similarities in processing image data, CNNs excel, particularly in managing high network depth and intricate algorithmic processes. (Bhavya, et al., 2021)

This report will center around the classification of seven facial expressions: angry, disgust, fearful, happy, sad, surprised, and neutral, utilizing the FER2013 dataset for training.

# **Literature Review**

# A study presents a facial expression recognition system tailored to tackle challenges arising from limited data and single-person facial images, utilizing the FER2013 dataset. The research outlines both the training and testing phases of the model. During the training phase, techniques such as data augmentation are employed, revealing that 'Disgust' exhibits higher sparsity compared to other emotions. The system leverages HaarCascade for precise identification of facial regions crucial for feature extraction during model training. To effectively process facial features, a convolutional neural network (CNN) is deployed, enabling efficient expression classification. Experimental findings showcase promising results, with the model achieving a notable accuracy rate of 69.47% post data augmentation. Furthermore, the paper highlights the model's superiority over several existing methods in terms of performance. However, it acknowledges trade-offs in time complexity when compared to certain approaches.(Jui-Feng, et al., 2023)

The study offers a comprehensive analysis of a multi-person facial expression recognition system, with a primary focus on enhancing accuracy and efficiency. Initially, the algorithm's performance is assessed using the extended Cohn–Kanade expression dataset, revealing accuracy limitations stemming from sparse data. To address this issue, the researchers augment the dataset by integrating additional internet sources and personal images, resulting in notable accuracy enhancements. Employing a traditional CNN module, the study extracts the primary expressional vector (EV) from facial images, pivotal for tracking expression changes. This EV, crucially derived from identifying key facial points, is generated by a basic perceptron unit operating on a background-removed face image. Furthermore, the proposed Facial Expression Recognition and Classification (FERC) model integrates a non-convolutional perceptron layer as the final stage. Each convolutional layer receives input data, transforms it, and subsequently passes it to the next level, facilitating feature extraction and classification. While the model achieves high accuracy at 96%, there's a decline in accuracy observed with an increasing number of images, indicative of overfitting.(Mehendale, 2020)

The study, titled "Real-Time Facial Emotion Classification and Recognition Using Deep Learning Models," aims to develop a system capable of real-time detection, recognition, and categorization of human facial expressions into seven distinct states. The research leverages Anaconda and Python 3.5, incorporating Viola-Jones and Haar cascade algorithms for face detection, alongside the KDEF dataset and the VGG16 CNN model for face recognition and classification. System evaluation demonstrates an impressive accuracy of 88%, surpassing the performance of previous methodologies. This paper follows a similar methodology as the above-mentioned studies. Another research effort, focusing on enhancing emotion classification in face recognition, explores model performance by implementing and comparing two modified models for emotion classification. Various metrics such as precision, loss, accuracy, recall, and AUC are utilized for comparison purposes. Data augmentation techniques are applied to augment the dataset size. Findings indicate that the VGG16 model outperforms MobileNet, achieving higher accuracy (0.9238 compared to 0.9009), thereby demonstrating its superior capability in accurately classifying emotions within the dataset.(Yang, 2023)

# **Methodology**

## **Dataset Description**

* 1. The dataset utilized in this project originates from Kaggle and is known as FER2013. It consists of grayscale facial images, each with dimensions of 48x48 pixels. These images are uniformly aligned to ensure that the face is approximately centered and occupies a consistent amount of space within each image. In total, the dataset comprises 35,887 images distributed across 7 classes. During model training, 28,709 images are utilized, while 7,178 images are reserved for testing purposes. (Sambare, 2020)

## **Algorithms Used**

## **HaarCascade**

* + 1. The HaarCascade algorithm is engineered to identify faces within images or real-time video streams. Its operation involves employing a cascade classifier trained with Haar-like features to detect objects or patterns in these visual inputs. This classifier consists of multiple stages, each housing a set of weak classifiers that progressively refine the identification of potential regions of interest in the image or video frame. Haar-like features are fundamental rectangular patterns utilized to characterize visual attributes of objects, such as edges, corners, and textures.(Behera, 2020) By integrating the Haar cascade algorithm for face detection into our project, we'll be able to accurately locate faces in the live webcam feed, enabling subsequent emotion recognition based on the detected facial expressions.

## **Convolution Neural Network**

* + 1. A Convolutional Neural Network (CNN) stands as a deep learning technique mainly used in tasks related to image processing and recognition. Within CNNs, convolutional layers scrutinize input images by executing convolution operations, which discern crucial features such as edges, textures, and patterns. Subsequently, pooling layers condense the resultant feature maps, retaining vital information while reducing the spatial dimensions of the data. Ultimately, fully connected layers leverage these extracted features to produce output predictions.(Craig, 2024)

## **Steps Involved**

The steps that are involved in creating the model are:

1. **Data Preparation:** The dataset used for training and validating the model is obtained from Kaggle and is known as the FER-2013 dataset. It comprises grayscale images of 48x48 pixels, totaling 35,887 faces. These images are categorized into seven distinct emotion classes:
   1. Angry (0)
   2. Disgust (1)
   3. Fear (2)
   4. Happy (3)
   5. Sad (4)
   6. Surprise (5)
   7. Neutral (6)

The dataset is split into a training set containing 28,709 images and a validation set containing 7,178 images. This partitioning allows for training the model on one subset of the data and evaluating its performance on another, ensuring a reliable assessment of the model's generalization capability.

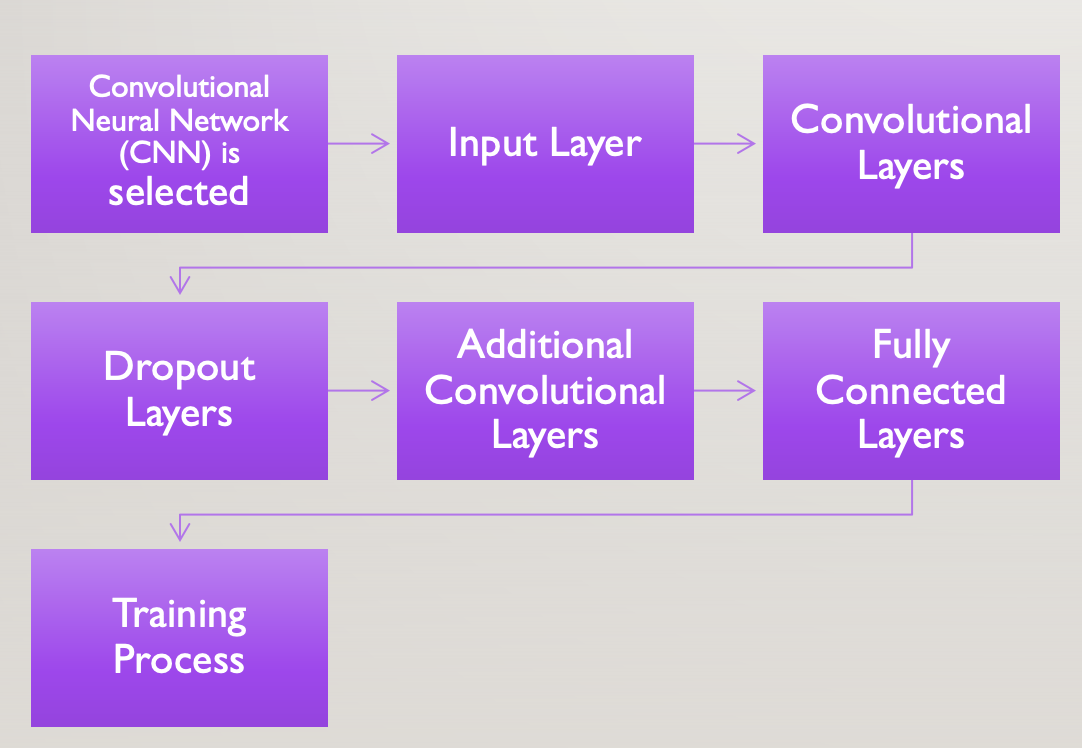
1. **Defining Model Architecture and Training**: For facial emotion recognition, a convolutional neural network (CNN) architecture is employed. The model consists of multiple layers, including convolutional layers, pooling layers, dropout layers, and fully connected layers.

The model architecture is defined using the Keras Sequential API. It comprises several convolutional layers with rectified linear unit (ReLU) activation, followed by max-pooling layers to down sample the feature maps. Dropout layers are inserted to prevent overfitting. After flattening the feature maps, dense fully connected layers are added to perform classification tasks. The final layer utilizes a SoftMax activation function to compute the probabilities of each emotion class.

The model is compiled with a categorical cross-entropy loss function and the Adam optimizer with a learning rate of 0.0001 and decay rate of 1e-6. Evaluation metrics, including accuracy, are specified to monitor the model's performance during training.

The model is trained using data generators created from the training and validation directories. These generators preprocess the images and yield batches of data during training. The training process involves iterating over the dataset for a specified number of epochs, updating the model's weights to minimize the loss function.

The validation data generator is used to evaluate the model's performance on the validation set after each epoch. This allows for monitoring the model's generalization ability and identifying potential overfitting. Here is a diagrammatic approach of the model development and training process:



**Model Architecture Overview:**

1. **Input Layer**: Grayscale images of size 48x48 pixels.
2. **Convolutional Layers**: We have taken two sets of convolutional layers. After each layer we have used a Rectified Linear Unit (ReLU) activation function. First set includes 32 filters, and the second set includes 64 filters. Max Pooling layers with a pool size of (2,2) are added after each convolutional layer to reduce spatial dimensions.
3. **Dropout Layers:** We have inserted a dropout rate of 0.25 after the first and third max pooling layers. This ensures that there is no overfitting by randomly dropping a fraction of input units.
4. **Additional Convolutional Layers**: We have inserted two more sets of convolutional layers, each with 128 filters, along with max pooling layers. Also, we have added a final set of convolutional layers with 256 filters, max pooling, and a dropout layer.
5. **Fully Connected Layers:** The output of the convolutional layers is flattened and then passed through two fully connected layers with 1024 units each, followed by ReLU activation and dropout. The output layer consists of 7 units with a softmax activation function which represents the probabilities of each facial emotion class.

**Training Process:**

* We have used the Adam optimizer to train the model with a learning rate of 0.0001 and a decay rate of 1e-6.
* A categorical cross-entropy is used as the loss function, and accuracy is monitored as the evaluation metric.
* Training is performed using a generator to feed batches of training data, with validation data also provided through a separate generator.

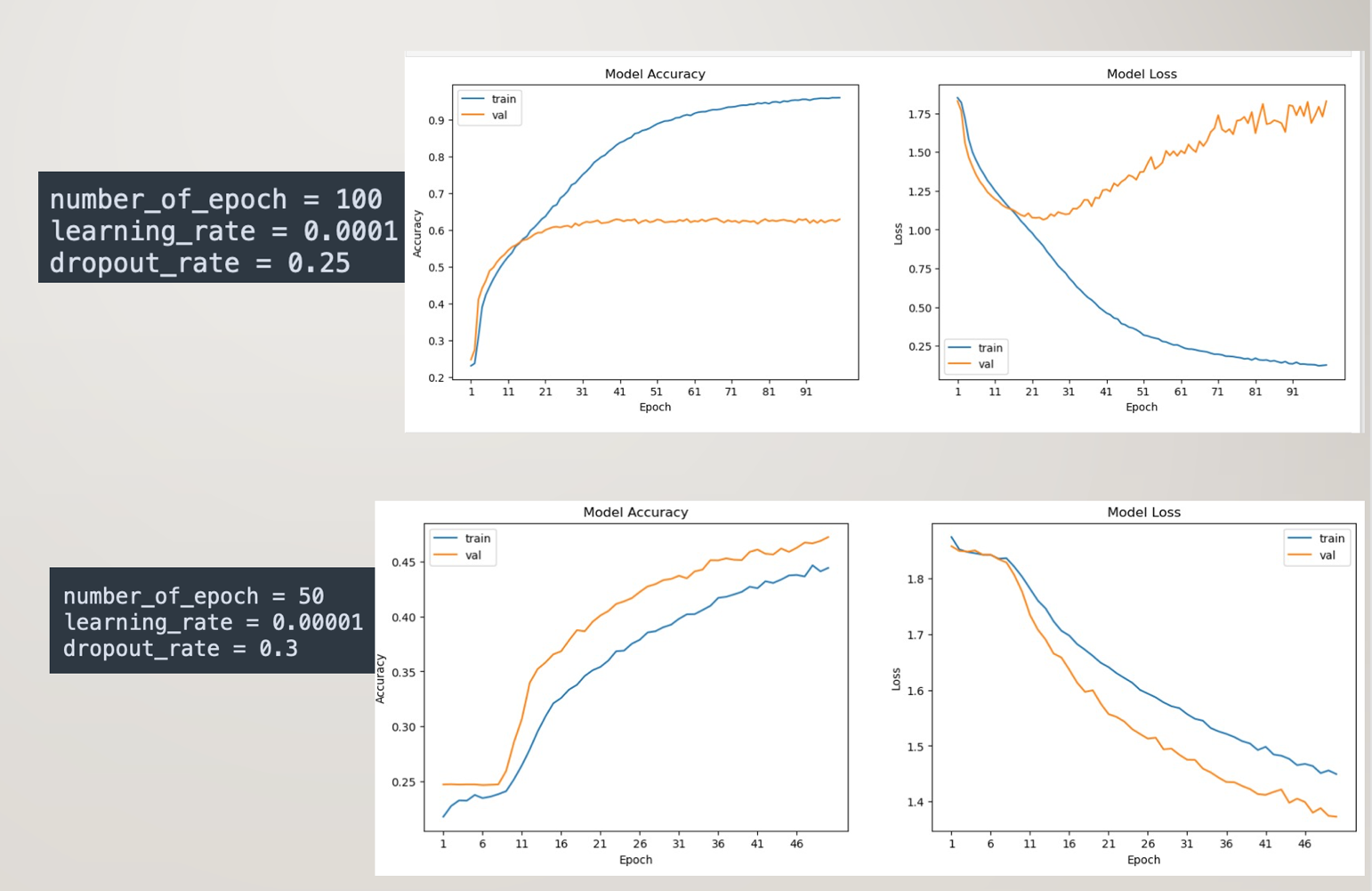
**Key Concepts:**

* *Convolutional Layer*: Applies filters to extract features from input images.
* *ReLU Activation Function*: Introduces non-linearity by setting negative inputs to zero.
* *Max Pooling*: Reduces spatial dimensions by retaining the maximum value in each window.
* *Dropout*: Prevents overfitting by randomly dropping input units during training.
* *Softmax Activation Function*: Converts raw scores into probability distributions over multiple classes.

1. **Real-time Emotion Detection from Webcam**: Finally, we use OpenCV library to initialize the webcam feed which continuously captures frames from the webcam and detects faces using a Haar Cascade Classifierand predicts the emotions on the detected faces using the trained CNN model. The predicted emotion label is then overlaid on the webcam feed using OpenCV drawing functions.

## **Outcomes Achieved**

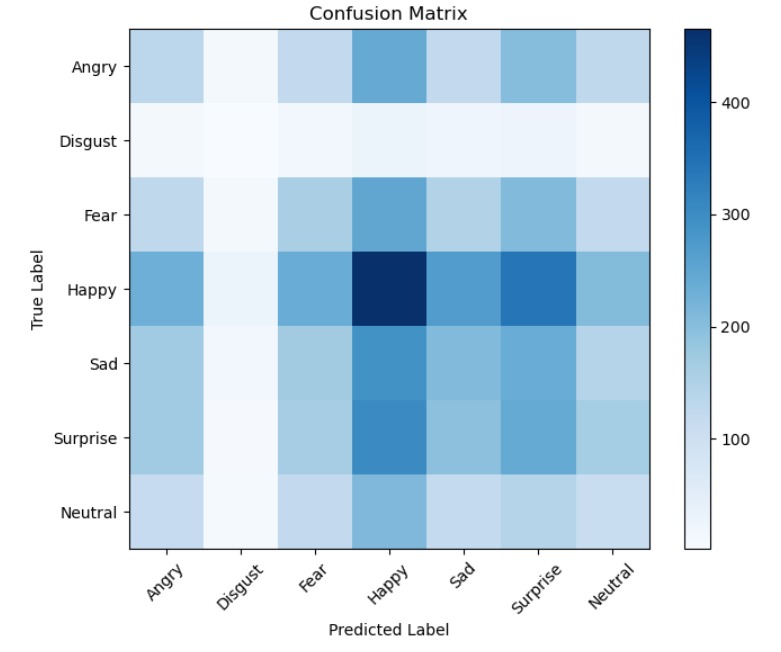
The final model was achieved after running the algorithm under different hyperparameters. The highest accuracy model was taken as the final model for emotion detection. The different models trained under different parameters are as follows:



From the first graph, we can see that the differences in the accuracy of our training set and validation set is very high. The accuracy of our validation set is significantly lower than the training set. Also, the loss of validation set is significantly higher than the training set. We can conclude it as being overfitted. Similarly, the second graph shows somewhat linear relation between the accuracy and loss of training and validation set, but the accuracy was too low for us to take it into account. A screenshot of a graph

Description automatically generated

Finally, we were able to get a fairly accurate model in comparison to the previous ones when number of epoch was set to 50 and learning rate was set to 0.0001. We obtained an accuracy of 88.3% and our model was able to distinguish emotions from the live webcam feed. Below is the confusion matrix obtained.



# For our model to be perfect, we would want all the values to be on the diagonal. However, there are off-diagonal values, and we can work on rectifying it in the future implementation.

# **Conclusion**

The study highlights the effectiveness of Convolution Neural Network in understanding human emotions. The model uses convolutional layers to extract hierarchical features directly from raw pixel data, enabling comprehensive understanding of complex patterns in facial expressions. Dropout layers are strategically integrated to mitigate overfitting by randomly deactivating units during training. Additionally, the SoftMax activation function in the output layer generates interpretable probabilities for each emotion class, aiding in effective classification. To optimize training, the model utilizes the Adam optimizer with a small learning rate and decay, enhancing performance and efficiency throughout the training process.

# **References**

Ashu, K., Amandeep, K. & Munish, K., 2018. Face detection techniques: a review. *Artificial Intelligence Review,* Volume 52, pp. 927-948.

Balaji, A., 2022. *Emotion Detection.* [Online]   
Available at: https://github.com/atulapra/Emotion-detection  
[Accessed 9 April 2024].

Behera, G. S., 2020. *Face Detection with Haar Cascade.* [Online]   
Available at: https://towardsdatascience.com/face-detection-with-haar-cascade-727f68dafd08  
[Accessed 9 April 2024].

Bhavya, A., Mohammad Sharay, A. & Harleen, K., 2021. Facial Emotion Detection Using Deep Learning and Haar Cascade Face Identification Algorithm. *Advances in Intelligent Computing and Communication,* Volume 202.

Craig, L., 2024. *convolutional neural network (CNN).* [Online]   
Available at: https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network#:~:text=Both%20RNNs%20and%20CNNs%20are,complex%20patterns%20in%20deeper%20layers.  
[Accessed 9 April 2024].

Jui-Feng, Y., Kuei-Mei, L., Chia-Chen, C. & Ting-Hao, W., 2023. Expression Recognition of Multiple Faces Using a Convolution Neural Network Combining the Haar Cascade Classifier. *Topic Applied System on Biomedical Engineering, Healthcare and Sustainability,* 23(13).

Lucey, P. et al., 2010. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition ,* pp. 94-101.

Mehendale, N., 2020. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences,* Volume 2, p. 446.

Sambare, M., 2020. *FER-2013.* [Online]   
Available at: https://www.kaggle.com/datasets/msambare/fer2013  
[Accessed 8 April 2024].

Shaik Asif, H. & Ahlam Salim Abdallah Al, B., 2020. A real time face emotion classification and recognition using deep learning model. *Journal of Physics,* Volume 1432.

Yang, F., 2023. Facial emotion classification based on the modified VGG and. *Journal of Physics,* Volume 2646.