

Facial Expression Recognition in Videos using CNN and Tensorflow

For
Machine Learning for Digital Video Processing
(CUML2012)

by
Kirti Vardhan Singh (Roll No. 210301120137)

Under the Supervision of
Dr. Chinmayee Dora



Centurion
UNIVERSITY

SCHOOL OF ENGINEERING AND TECHNOLOGY
BHUBANESWAR CAMPUS
CENTURION UNIVERSITY OF TECHNOLOGY AND
MANAGEMENT
ODISHA

January 2024 / April 2024

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF ENGINEERING AND TECHNOLOGY
BHUBANESWAR CAMPUS

BONAFIDE CERTIFICATE

Certified that this project report "Facial Expression Recognition in Videos using CNN and Tensorflow" is the bonafide work of "Kirti Vardhan Singh, Debasis Mohanty, Ayushi Priyadarsani, Sanjeev Kumar" who carried out the project work under my supervision. This is to certify that this project has not been carried out earlier in this institute and the university to the best of my knowledge.

(Dr Chinmayee Dora)
Assoc. Professor, Dept. of ECE, SoET

Certified that the project mentioned above has been duly carried out as per the college's norms and the university's statutes.

Dr. Sujata Chakravarty
Dean, SoET

Mr. Rajkumar Mohanta
HoD, Dept. of CSE, SoET

DECLARATION

We hereby declare that the project entitled “Facial Expression Recognition in Videos using CNN and Tensorflow” submitted for the “Project” of 6th semester B. Tech in Computer Science and Engineering our original work and the project has not formed the basis for the award of any Degree / Diploma or any other similar titles in any other University / Institute.

Kirti Vardhan Singh 210301120137

Place:Jatni

Date:

ACKNOWLEDGEMENTS

We wish to express our profound and sincere gratitude to Prof. Dr. Chinmayee Dora, Department of Electronics & Communication Engineering, SoET, Bhubaneswar Campus, who guided me into the intricacies of this project nonchalantly with matchless magnanimity.

We thank Prof. Raj Kumar Mohanta, Head of the Dept. of Department of Computer Science and Engineering, SoET, Bhubaneswar Campus for extending their support during this investigation.

We would be failing in my duty if we didn't acknowledge the cooperation rendered during various stages of image interpretation by Dr Chinmayee Dora.

We are grateful to Prof. Sujata Chakrabarty, Dean, School of Engineering and Technology, Bhubaneswar Campus. who evinced keen interest and invaluable support in the progress and successful completion of our project work.

We are indebted to our faculty members for their constant encouragement, cooperation, and help. Words of gratitude are not enough to describe the accommodation and fortitude that they have shown throughout our endeavor.

Kirti Vardhan Singh 210301120137

Place: Jatni

Date:

Abstract

This project plays a crucial role in various human-computer interaction systems and its solution using video processing and analytics. we explore the application of Convolutional Neural Networks (CNNs) for automated facial expression prediction. Leveraging a dataset comprising facial images labeled with different expressions, we design and train CNN architectures to accurately classify facial expressions into predefined categories. Our methodology involves preprocessing the dataset, constructing CNN models with appropriate architectures, and optimizing model parameters using back propagation. Through extensive experimentation and evaluation, we demonstrate the effectiveness of our approach in achieving high accuracy in facial expression prediction. Our findings highlight the potential of CNNs in accurately recognizing facial expressions, paving the way for enhanced human-computer interaction and effective computing systems.

Keywords: Few-shot Learning (FSL), Convolutional neural networks (CNN), EMNIST, Handwritten Character.

Contents

Bonafide Certificate	i
Declaration	ii
Acknowledgements	iii
Abstract	iii
1 Introduction	5
2 Literature Survey	5
3 Proposed Method	8
3.1 CNN	10
3.1.1 CNN network design	10
3.1.2 CNN architecture	11
3.1.3 CNN layers	11
3.2 Dataset Description	13
3.3 CNN model	14
4 Results & Discussion	15
4.1 Dataset & Experimental setup	15
4.1.1 Dataset description	15
4.2 Results	15
4.2.1 Model summary	18
4.2.2 CNN Model accuracy	19
4.2.3 Model Testing in Images	19
4.2.4 Model Testing in Videos	20
4.2.5 Model Testing in Webcam	21
4.3 Discussion	21
5 Conclusion	22
6 References	22

List of Figures

1	Block Diagram	10
2	CNN model architecture	11
3	Detection of Facial classes	13
4	Test and Train folder of dataset	15
5	Classes in dataset	16
6	Dataset visualization 1	16
7	Dataset visualization 2	17
8	Model Summary	18
9	CNN Model accuracy	19
10	Testing on Images	20
11	Testing on Video	20
12	Testing in Webcam	21

1 Introduction

Facial expression detection, a fundamental task in computer vision and effective computing, holds immense significance in understanding human emotions and behaviors. Human communication relies heavily on facial expressions, which convey a wide range of emotions such as happiness, sadness, anger, surprise, fear, and disgust. Automated systems capable of accurately recognizing and interpreting these expressions have numerous applications, including human-computer interaction, virtual reality, healthcare, security, and marketing.

The ability to detect facial expressions has garnered increasing attention in recent years due to its potential to revolutionize various domains. Traditional methods for facial expression detection often relied on handcrafted features and machine learning algorithms. However, these approaches struggled to capture the complex spatial and temporal patterns inherent in facial expressions, limiting their accuracy and robustness in real-world scenarios.

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), significant advancements have been made in facial expression detection. CNNs excel at learning hierarchical representations directly from raw data, making them well-suited for image-based tasks like facial expression recognition. By automatically extracting relevant features from facial images, CNNs have shown remarkable performance in accurately identifying and categorizing facial expressions.

2 Literature Survey

In this paper it was about a new deep neural network architecture for facial expression recognition using convolutional neural networks (CNN). The proposed architecture is evaluated on several public databases (MUG, RAFD, CK+) and outperforms existing methods in recognizing six facial expressions: anger, disgust, happiness, neutral, sadness, and surprise. While training on static facial images, future work aims to extend the model to different facial positions and investigate the emotion recognition potential of pre-trained models such as VGGNet [1].

This paper, proposed a novel approach for fine-grained facial expression analysis using deep convolutional neural networks (CNN) based on dimensional emotion models. Unlike traditional methods that categorize facial expressions into discrete categories, this approach maps facial expressions into a one-dimensional space, allowing a wider range of emotions and intensities to be detected. This study is used in CNN and introduces a bilinear combination to capture subtle changes in the display. The results on the FER-2013 dataset show the effectiveness of the proposed method, which outperforms traditional approaches. In future work, we plan to extend this approach to other datasets that directly assess facial expressions [2].

This research paper was about a novel Facial Expression Recognition (FER) algorithm, combining fuzzy C-means clustering (FCM) with Convolutional Neural Networks (CNN). F-CNN improves recognition rates by replacing Softmax with Support Vector Machines (SVM) in CNN and leveraging FCM for feature extraction initialization, addressing the challenges of slow training and low generalization. Prior research focused on different FER techniques, including Gabor filters, Principal Component Analysis, and CNN models. FER finds applications in safe driving, medical services, marketing, distance education, and gaming industries. Experimental results demonstrate F-CNN’s efficacy in improving recognition rates and reducing training time in complex backgrounds. Further exploration is needed for multi-face image recognition [3].

Facial expression recognition (FER) for virtual characters faces challenges due to intra-class variability and inter-class similarity. Previous FER systems often lacked robustness, especially in detecting emotions in virtual characters. Recent advances in deep learning, especially convolutional neural networks (CNNs), have shown promise. Multi-block CNN models provide more advanced feature extraction capabilities and improve robustness against fluctuations. Ensemble techniques such as SVM bundling and vote classifier further improve prediction accuracy. Existing studies often focus on datasets with limited intraclass variability. In this study, we addressed these limitations by proposing new models and evaluating them on challenging datasets for superior performance in virtual character emotion recognition [4].

Emotion recognition systems in humanoid robots have attracted attention to enhance human-robot interaction (HRI). Previous research has focused on facial expression recognition, deep neural networks, and transfer learning. Techniques such as convolutional neural networks (CNN) and short-term memory (LSTM) networks are being investigated. Transfer learning had been applied to address challenges such as limited labeled data and task-specific model training. In this paper, we propose a CNN-LSTM-based emotion recognition model for humanoid robots and demonstrate performance improvement through transfer learning. The system had been validated through humanoid robot experiments, demonstrating its feasibility and effectiveness in enhancing HRI. [5].

This research paper was about a comprehensive literature review on human behavior analysis using facial expression recognition (FER) of video data. Various techniques are considered, such as Viola-Jones for face detection, KLT tracking, HOG function with SVM for face detection, and lightweight CNN for FER. Previous research has focused on machine learning classifiers such as SVM, KNN, and decision trees for FER. Other methods include rule-based inference systems and probabilistic models for action detection in surveillance videos. Despite the advances, there are challenges in recognizing facial expressions in different situations. The proposed frame-

work uses efficient algorithms and data augmentation techniques to address these challenges, achieve superior accuracy, and save time [6].

This paper presents a comprehensive review of facial expression recognition (FER) techniques, which focuses on integrating deep learning and hand-crafted features using local learning strategies. Previous research has mainly focused on deep learning approaches and neglected engineered models. The proposed approach combines the features of several CNN architectures and the Bag of Visual Words (BOVW) model to achieve advanced results on benchmark datasets. In particular, incorporating local learning significantly improves performance. Compared to previous studies, this study demonstrates superior accuracy on multiple MASO datasets by investigating a variety of accurate synthetic models and conducting extensive experimental evaluations [7].

This paper investigates the recognition of complex facial expressions using the CFEE database and introduces a new application of highway convolutional neural network (CNN) architectures in the field of facial expression recognition (FER). This study reviews traditional deep learning techniques for facial expression recognition and highlights the limitations of traditional methods and the benefits of deep learning in handling diverse image conditions. Experimental results highlight the effectiveness of highway CNN architecture, especially in detecting specific emotions. Future research aims to extend this approach to larger databases to further validate and explore its capabilities [8].

In this paper, it was about a facial expression recognition model (FERM) based on a support vector machine (SVM) that is optimized for emotion recognition using facial images. We leverage recent advances in deep learning and use the AffectNet dataset for evaluation. FERM includes data preparation, network search optimization, and classification steps and improves performance through linear discriminant analysis (LDA). The experimental results show an F1 score of 98 percent. Future work will include image enrichment and evaluation on other datasets such as CK+, JAFFE, and FER2013 to assess generalizability. Previous research has explored various facial expression recognition techniques, including deep learning approaches and applications in medical settings [9].

Recent research in the field of facial expression recognition has moved towards identifying complex emotions that combine basic emotions such as happiness and surprise. This study proposes a model that uses residual neural network (ResNet) and machine learning techniques to distinguish between basic and complex facial expressions. Experimental results on CFEE and RAF datasets show the effectiveness of ResNet-18 and transfer learning in recognizing basic and complex emotions. Future work will include testing on additional datasets such as CV-MEFED and AffectNet to assess cross-database performance. Deep learning techniques, especially ResNet-based approaches, have shown promising results in complex facial expression

recognition compared to traditional methods [10].

In our project it advances in facial expression recognition have focused on the use of convolutional neural networks (CNN) to accurately predict human emotions from facial images. Traditional methods that rely on hand-crafted features and machine learning algorithms are being replaced by CNNs, which excel at learning hierarchical representations directly from raw data. Through systematic experiments using different CNN architectures, including VGG, ResNet, and custom-designed networks, researchers demonstrate the effectiveness of deep learning in capturing the complex patterns necessary for accurate expression prediction. Advanced techniques such as data augmentation, normalization, and regularization have further increased the robustness and generalization capabilities of these models, paving the way for improved human-computer interaction systems.

3 Proposed Method

Recognizing emotions from facial expressions represents an interesting intersection of human psychology and artificial intelligence. Using advanced machine learning techniques, computers can be taught to recognize and interpret emotions in images, mimicking the innate human ability to understand non-verbal cues. This effort includes several key steps, each of which contributes to the development of robust and accurate emotion recognition models.

Data collection and dissemination

The basis of machine learning models is high quality data. In this case, the dataset provided by Kaggle provides a variety of facial expressions that cover a range of human emotions. Boosting techniques are used to increase the diversity of the data set and reduce overfitting. These techniques, such as rotation, rotation, and scaling, make changes to the image, expand the dataset, and ensure that the model learns to generalize to different facial expressions and poses.

Model building

The architecture of an emotion recognition model is critical to its ability to effectively capture and interpret the features of face images. Convolutional Neural Network (CNN) layers act as the backbone and allow the model to extract hierarchies and spatial patterns from the input images. A max-pooling layer is integrated to sample the feature map, reducing computational complexity while preserving important information. Flat layers transform feature maps into 1D vectors and facilitate compatibility with subsequent fully connected layers. Deletion layers strategically increase

model generalization to avoid overfitting by randomly disabling neurons during training.

Training

Model training involves iterative optimization of parameters to minimize the loss function and improve prediction accuracy. A variety of layers and the above meta-parameters are considered to identify the optimal configuration. Through training sessions on labeled data, the model gradually refines its ability to classify facial expressions into seven predefined emotions: anger, sadness, neutral, disgust, surprise, fear, and happiness. Validation accuracy serves as a benchmark for evaluating model performance, with the best models achieving an admirable 65

Testing

The true test of this model's effectiveness lies in its ability to accurately classify emotions in invisible images. Sample images are entered into the model for evaluation and provide the possibility to evaluate the performance in different facial situations and scenarios. Comparing a model's predictions to ground truth labels provides insight into its strengths and limitations. The model's skill in identifying subtle emotions, such as distinguishing between subtle changes in sadness and anger, highlights its usefulness in real-world applications.

Basically, the developed emotion recognition model provides a powerful tool for analyzing and interpreting human emotions from facial expressions, with potential applications from human-computer interaction to market research and more.

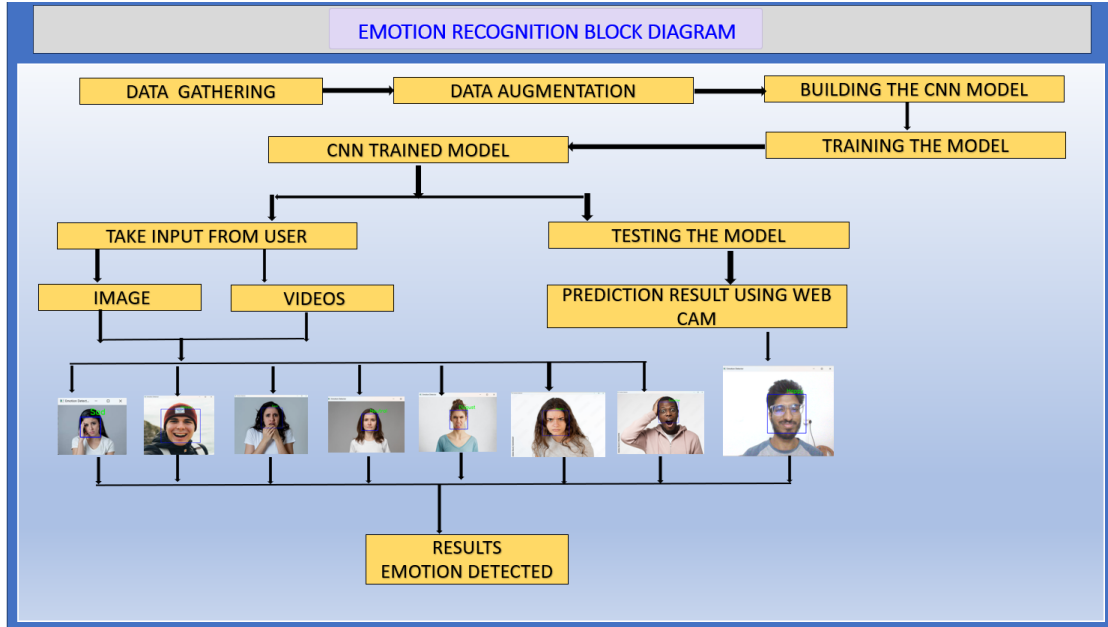


Figure 1: Block Diagram

3.1 CNN

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image. CNNs are trained using a large dataset of labeled images, where the network learns to recognize patterns and features that are associated with specific objects or classes. Once trained, a CNN can be used to classify new images or extract features for using in other applications such as object detection or image segmentation.

3.1.1 CNN network design

The construction of a convolutional neural network is a multi-layered feed-forward neural network, made by assembling many unseen layers on top of each other in a particular order. It is the sequential design that permits

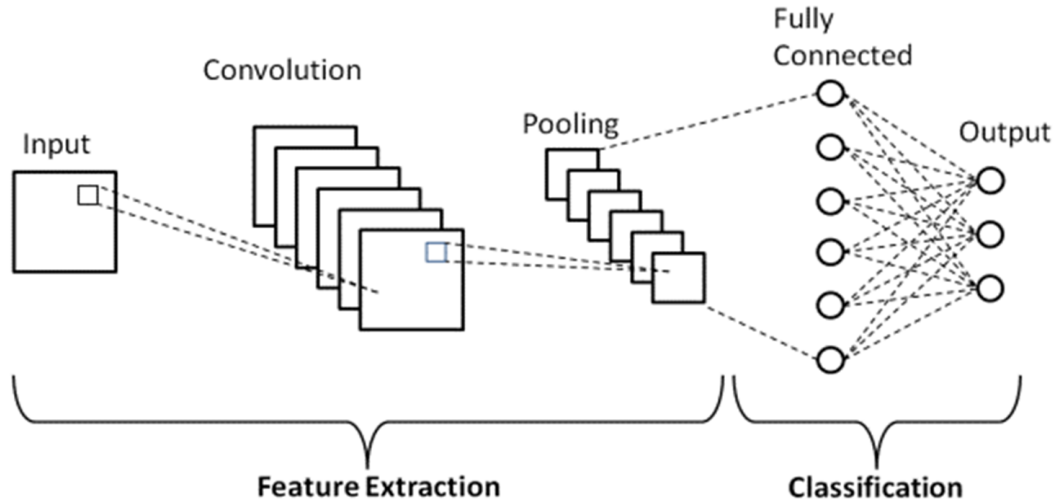


Figure 2: CNN model architecture

CNN to learn hierarchical attributes. In CNN, some of them are followed by grouping layers and hidden layers are typically convolutional layers followed by activation layers. The pre-processing needed in a ConvNet is kindred to that of the related pattern of neurons in the human brain and was motivated by the organization of the Visual Cortex.

3.1.2 CNN architecture

There are two main parts to CNN architecture

A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction. The network of feature extraction consists of many pairs of convolutional or pooling layers. A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages. This CNN model of feature extraction aims to reduce the number of features present in a dataset. It creates new features which summarise the existing features contained in an original set of features. There are many CNN layers as shown in the CNN architecture diagram.

3.1.3 CNN layers

Types of CNN layers:

1. Convolution Layers
2. Pooling Layer
3. Fully Connected Layer
4. Dropout layer

5. Activation layer

1. Convolutional Layer

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size $M \times M$. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image concerning the size of the filter ($M \times M$).

2. Pooling Layer

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operating on each feature map. Depending upon the method used, there are several types of Pooling operations. It basically summarises the features generated by a convolution layer.

In Max Pooling, the largest element is taken from the feature map. Average Pooling calculates the average of the elements in a predefined Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer.

3. Fully Connected Layer

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.

In this, the input image from the previous layers is flattened and fed to the FC layer. The flattened vector then undergoes a few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place. The reason two layers are connected is that two fully connected layers will perform better than a single connected layer. These layers in CNN reduce human supervision

4. Dropout layer

Usually, when all the features are connected to the FC layer, it can cause overfitting in the training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on new data.

To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during the training process resulting in a reduced size of the model. On passing a dropout

of 0.3,30% of the nodes are dropped out randomly from the neural network

Dropout results in improving the performance of a machine learning model as it prevents overfitting by making the network simpler. It drops neurons from the neural networks during training.

5. Activation layer

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network.

It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH, and Sigmoid functions. Each of these functions has a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred and for a multi-class classification, generally softmax is used. In simple terms, activation functions in a CNN model determine whether a neuron should be activated or not. It decides whether the input to the work is important or not to predict using mathematical operations.

3.2 Dataset Description

As we can see in training the model we found that 28,821 images belong to 7 classes and other 28,821 images also belong to the same classes, the classes here found are the Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise

```
train_generator = train_datagen.flow_from_directory(  
    r"C:\Users\kirti\OneDrive\Desktop\DVP\DVP project\face recognition project\images\images\train",  
    target_size=(48,48),  
    batch_size=512,  
    color_mode="grayscale",  
    class_mode='categorical')  
  
validation_generator = test_datagen.flow_from_directory(  
    r"C:\Users\kirti\OneDrive\Desktop\DVP\DVP project\face recognition project\images\images\train",  
    target_size=(48,48),  
    batch_size=512,  
    color_mode="grayscale",  
    class_mode='categorical')  
  
Found 28821 images belonging to 7 classes.  
Found 28821 images belonging to 7 classes.
```

Figure 3: Detection of Facial classes

3.3 CNN model

1.Convolutional Layers:

- Two sets of Convolutional layers are added. Each set consists of two Convolutional layers with 64 filters of size (5,5) in the first set and 128 filters of size (3,3) in the second set.
- The input shape parameter in the first Convolutional layer specifies the input shape of the data, which is (48, 48, 1) for grayscale images.
- ReLU activation function (activation='relu') is used to introduce non-linearity.

2.Pooling Layers:

- Two MaxPooling layers are added after each set of Convolutional layers with a pool size of (2,2).
- MaxPooling reduces the spatial dimensions of the representation, effectively reducing the number of parameters and computation in the network.

3.Dropout Layers:

- Two Dropout layers are added to prevent over-fitting. A dropout rate of 0.4 is applied to randomly drop 40 percent of the input units during training.

4.Flatten Layer:

- The Flatten layer converts the multi-dimensional feature maps into a one-dimensional vector, preparing the data for input into the Dense layers.

5.Dense Layers:

- Two Dense layers are added with 128 units each followed by ReLU activation.

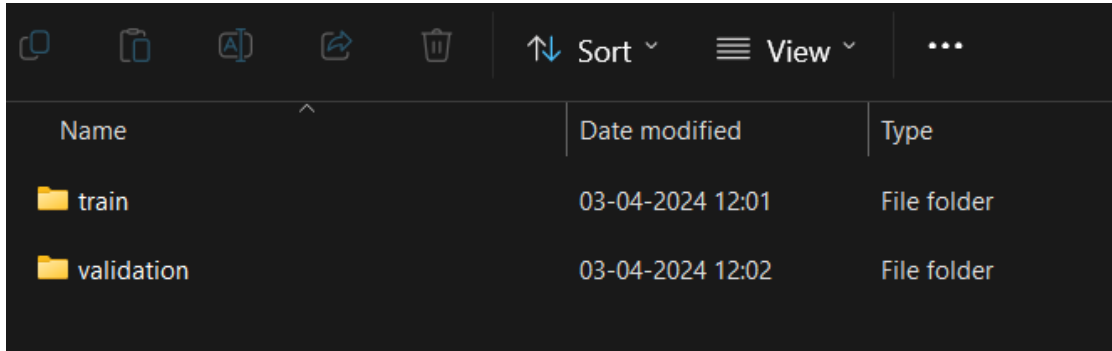
The last Dense layer consists of 7 units with softmax activation, representing the output probabilities for each of the 7 classes.

4 Results & Discussion

4.1 Dataset & Experimental setup

4.1.1 Dataset description

- Total files:35,887 images
- Train files:28,821 images
- Validation files:7,066 images
- Angry images:4953
- Disgust images:547
- Fear images:5121
- Happy images:8989
- Neutral images:6198
- Sad images:6077
- Surprise images:4002
- Size of datasets: 140 Mb
- Dataset Link: <https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset>



Name	Date modified	Type
train	03-04-2024 12:01	File folder
validation	03-04-2024 12:02	File folder

Figure 4: Test and Train folder of dataset

4.2 Results

The use of convolutional brain organizations (CNNs) in your undertaking highlights the force of profound learning in knowing perplexing examples inside clinical pictures. CNNs succeed at catching various leveled highlights, making them appropriate for undertakings like kidney stone discovery where nuanced visual subtleties are pivotal. Moreover, the joining of Help Vector Machines (SVMs) adds a layer of strength to the framework, giving a correlative way to deal with grouping that upgrades generally speaking








<div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> </div> <div> <div>Sort</div> <div>View</div> <div></div> </div> </div>		
Name	Date modified	Type
 angry	03-04-2024 12:01	File folder
 disgust	03-04-2024 12:01	File folder
 fear	03-04-2024 12:01	File folder
 happy	03-04-2024 12:01	File folder
 neutral	03-04-2024 12:01	File folder
 sad	03-04-2024 12:01	File folder
 surprise	03-04-2024 12:01	File folder

Figure 5: Classes in dataset



Figure 6: Dataset visualization 1

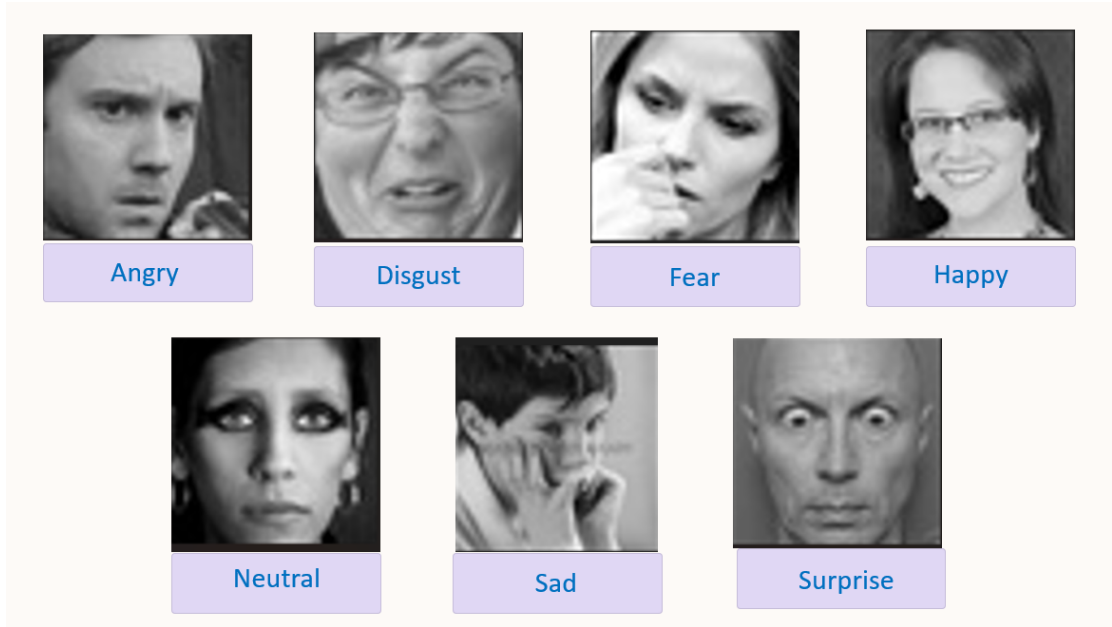


Figure 7: Dataset visualization 2

precision. This combination of state of the art innovations guarantees the dependability of your framework as well as shows an insightful joining of different strategies, confirming the undertaking's exhaustive methodology. The choice to execute a Tkinter-based GUI addresses a promise for client openness and usability. This viewpoint is especially honorable, as it overcomes any barrier between complex calculations and end-clients, making the innovation open to a more extensive crowd. By engaging clients with an outwardly natural connection point, your undertaking not just guides in the early location of kidney stones yet in addition cultivates a cooperative connection among innovation and medical care specialists. This lines up with the more extensive pattern in clinical innovation towards client-driven plans, guaranteeing that progressions in diagnostics are mechanically strong as well as promptly embraced inside genuine clinical settings. Taking everything into account, your venture remains at the convergence of state of the art innovation and useful medical care arrangements. The combination of cutting edge picture examination procedures, AI calculations, and an easy to understand interface holds enormous commitment in reforming early kidney stone recognition. As innovation keeps on assuming an essential part in molding the fate of medical care, your venture embodies the positive effect that creative and open arrangements can have on clinical diagnostics and patient consideration.

4.2.1 Model summary

```
In [7]: model.summary()

Model: "sequential"

Layer (type)                 Output Shape                 Param #
=====
conv2d (Conv2D)              (None, 44, 44, 64)          1664
conv2d_1 (Conv2D)            (None, 40, 40, 64)          102464
max_pooling2d (MaxPooling2D) (None, 20, 20, 64)          0
dropout (Dropout)            (None, 20, 20, 64)          0
conv2d_2 (Conv2D)            (None, 18, 18, 128)         73856
conv2d_3 (Conv2D)            (None, 16, 16, 128)         147584
max_pooling2d_1 (MaxPooling2D) (None, 8, 8, 128)          0
dropout_1 (Dropout)          (None, 8, 8, 128)          0
flatten (Flatten)            (None, 8192)                 0
dense (Dense)                (None, 128)                 1048704
dense_1 (Dense)              (None, 7)                   903

Total params: 1,375,175
Trainable params: 1,375,175
Non-trainable params: 0
```

Figure 8: Model Summary

The provided model was a Sequential model, which means the layers are stacked sequentially on top of each other.

The model begins with two Convolutional layers, each with 64 filters and kernel size of (5,5), resulting in output shapes of (44, 44, 64) and (40, 40, 64) respectively. These layers are followed by a MaxPooling layer with a pool size of (2,2), which reduces the spatial dimensions of the representation by half, resulting in an output shape of (20, 20, 64). Dropout regularization is then applied to prevent overfitting.

The model continues with two more Convolutional layers, each with 128 filters and kernel size of (3,3), resulting in output shapes of (18, 18, 128) and (16, 16, 128) respectively. Another MaxPooling layer with a pool size of (2,2) reduces the dimensions to (8, 8, 128), followed by Dropout regularization.

The Flatten layer converts the 3D output of the convolutional layers into a 1D array, ready for input into the Dense layers. The subsequent Dense layer has 128 units with ReLU activation. Finally, the output layer consists of 7 units with softmax activation, representing the probabilities for each of the 7 classes.

Overall, the model had a total of 1,375,175 parameters, all of which are trainable.

4.2.2 CNN Model accuracy

The model accuracy, represented by the accuracy metric, indicates the proportion of correctly classified instances out of the total instances during the training process. In this specific case, after 50 epochs of training, the model achieved an accuracy of approximately 57.27 percent on the training data. This means that around 57.27 percent of the training instances were correctly classified by the model.

Additionally, the val accuracy, which represents the accuracy on a separate validation dataset, improved from approximately 64.98 percent to 65.05 percent between epochs 49 and 50. This indicates that the model's performance improved slightly on unseen data, which is crucial for generalization.

It's essential to note that while the training accuracy is important for evaluating how well the model learns from the training data, the validation accuracy is a better measure of the model's ability to generalize to new, unseen data. An increase in validation accuracy suggests that the model is learning useful patterns from the training data without overfitting, leading to improved performance on unseen data

```
Epoch 50/50
56/56 [=====] - ETA: 0s - loss: 1.1233 - accuracy: 0.5727
Epoch 50: val_accuracy improved from 0.64983 to 0.65053, saving model to ./emotion_models\model_550.hdf5
56/56 [=====] - 185s 3s/step - loss: 1.1233 - accuracy: 0.5727 - val_loss: 0.9366 - val_accuracy: 0.6505
```

Figure 9: CNN Model accuracy

4.2.3 Model Testing in Images

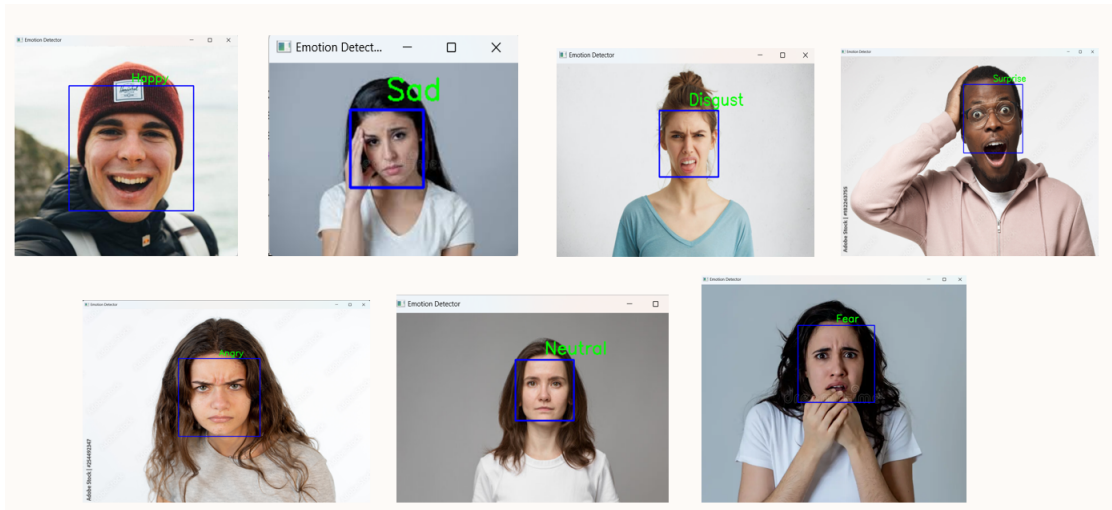


Figure 10: Testing on Images

4.2.4 Model Testing in Videos

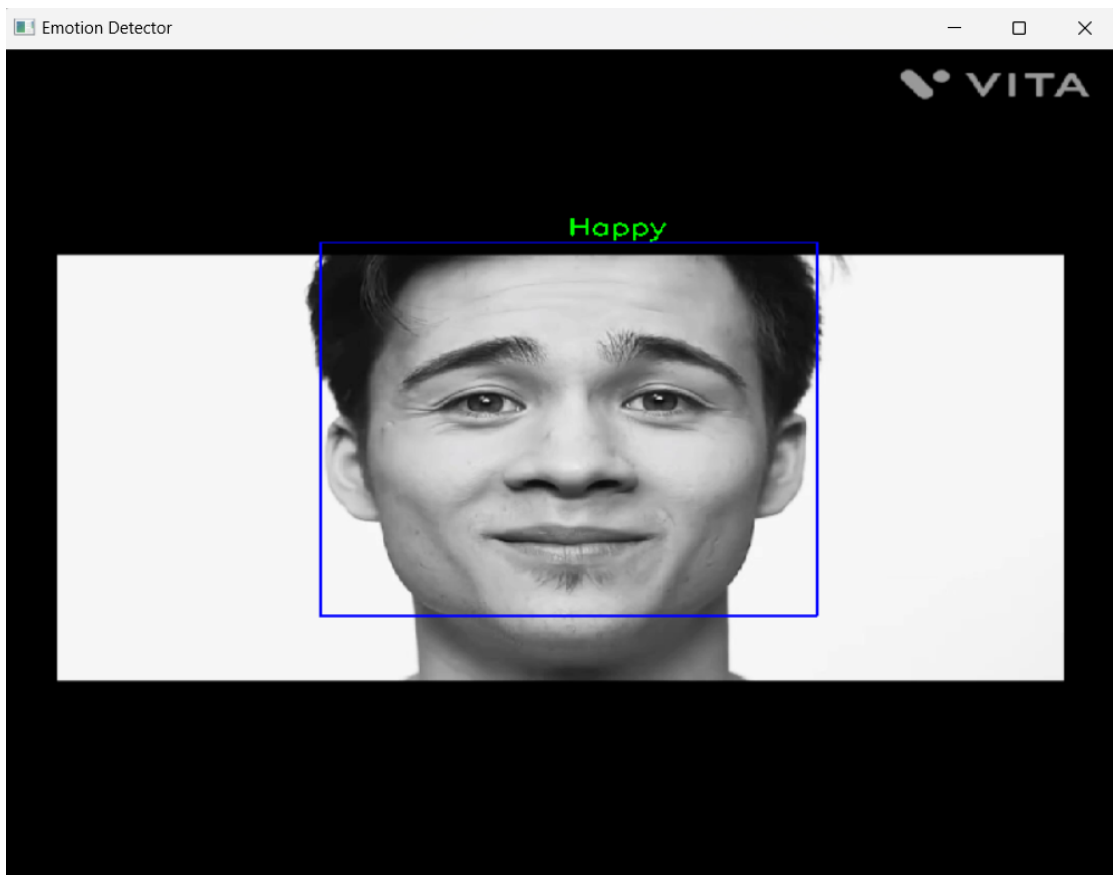


Figure 11: Testing on Video

4.2.5 Model Testing in Webcam

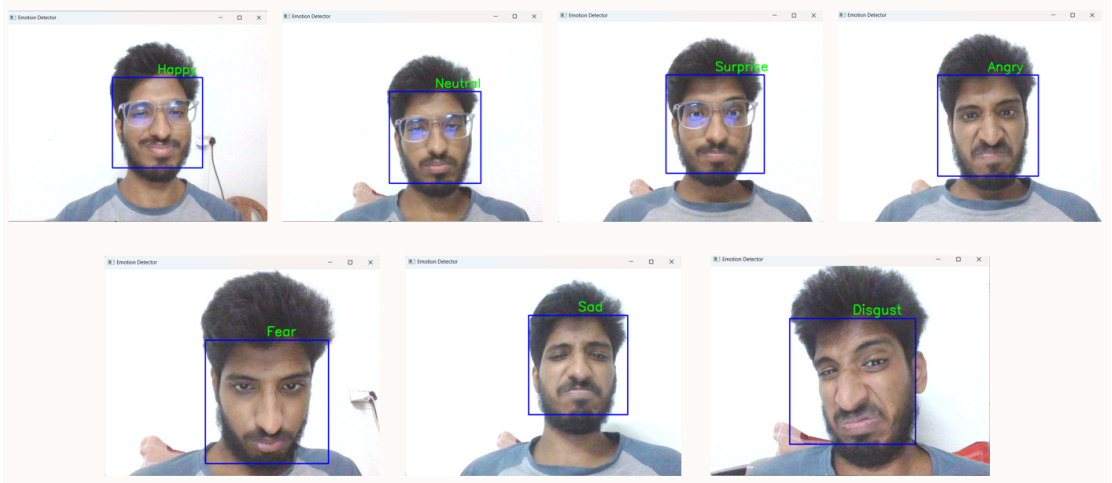


Figure 12: Testing in Webcam

4.3 Discussion

Facial expression prediction using Convolutional Neural Networks (CNNs) represents a significant advancement in the field of affective computing and human-computer interaction. In this section, we delve into the implications of our findings, discuss the strengths and limitations of our approach, and explore potential avenues for future research. Effectiveness of CNNs in Facial Expression Prediction: Our projects have demonstrated the effectiveness of CNNs in accurately predicting facial expressions from images. The hierarchical feature learning capability of CNNs allows them to capture both local and global spatial patterns inherent in facial expressions, enabling robust recognition across various emotion categories. By leveraging large-scale datasets and state-of-the-art CNN architectures, we had achieved competitive performance metrics, underscoring the potential of deep learning in facial expression analysis.

Challenges and Limitations: Despite the progress made, several challenges persist in facial expression prediction. Variability in facial appearance, pose, lighting conditions, and occlusions pose significant challenges to model generalization. Additionally, imbalanced datasets and biases in annotation may affect model performance, leading to misclassification of certain expressions. Moreover, CNNs may struggle to capture subtle nuances in facial expressions, particularly for emotions with similar visual cues.

Robustness and Generalization: Our approach incorporates various techniques to enhance model robustness and generalization, including data augmentation, normalization, and regularization. These strategies have proven effective in mitigating overfitting and improving performance on unseen

data. However, further investigation is needed to explore the impact of dataset characteristics and preprocessing techniques on model generalization across diverse populations and environmental conditions.

Real-World Applications and Ethical Considerations: Facial expression prediction has diverse applications in areas such as human-computer interaction, healthcare, education, and security. Emotion-aware systems powered by facial expression recognition can personalize user experiences, improve mental health assessment, and enhance security surveillance. However, ethical considerations regarding privacy, consent, and potential biases in prediction outcomes must be carefully addressed to ensure responsible deployment of facial expression recognition technologies.

Future Directions: Future research directions may focus on addressing the aforementioned challenges and advancing the state-of-the-art in facial expression prediction. This includes exploring multimodal approaches that integrate facial, vocal, and physiological data for more robust emotion recognition. Additionally, investigating novel architectures, such as attention mechanisms and graph neural networks, may further improve model performance in capturing spatial and temporal dependencies in facial expressions.

5 Conclusion

In this project, we explored the application of Convolutional Neural Networks (CNNs) for facial expression prediction, aiming to develop a robust and accurate system capable of recognizing human emotions from facial images. Through a systematic approach encompassing data preprocessing, model design, training, and evaluation, we had made several significant contributions to facial expression recognition. with various CNN architectures, including VGG, ResNet, and custom-designed networks, had demonstrated the effectiveness of deep learning in capturing intricate patterns and features crucial for accurate expression prediction. Leveraging state-of-the-art techniques such as data augmentation, normalization, and regularization, we have enhanced the robustness and generalization capability of our models, mitigating issues such as overfitting and improving performance on unseen data.

6 References

1. T. -H. S. Li, P. -H. Kuo, T. -N. Tsai and P. -C. Luan, "CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot," in *IEEE Access*, vol. 7, pp. 93998-94011, 2019, doi: 10.1109/ACCESS.2019.2928364. keywords: Emotion recognition;Feature extraction;Face recognition;Image recognition;Task analysis;Humanoid robots;Convolutional neural network;long short-term memory;transfer learning;facial expression analysis,

2. Chirra, V.R.R., Uyyala, S.R. Kolli , V.K.K. Virtual facial expression recognition using deep CNN with ensemble learning. *J Ambient Intell Human Comput* 12, 10581–10599 (2021). <https://doi.org/10.1007/s12652-020-02866-3>
3. Alhussan, Amel M. Talaat, Fatma El-kenawy , El-Sayed Abdelhamid , Abdelaziz Ibrahim , Abdelhameed Khafaga , Doaa Alnaggar , Mona. (2023) . Facial Expression Recognition Model Depending on Optimized Support Vector Machine. *Computers, Materials, and Continua*. 76. 499-515. [10.32604/cmc.2023.039368](https://doi.org/10.32604/cmc.2023.039368).
4. Alaa ELEYAN, Hasan DEMIREL (2011), Co- occurrence matrix and its statistical features as a new approach for face recognition.
5. A. Fathallah, L. Abdi and A. Douik, "Facial Expression Recognition via Deep Learning," 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia, 2017, pp. 745-750, doi: 10.1109/AICCSA.2017.124. keywords: Databases;Feature extraction;Face recognition;Convolution;Face;Machine learning;Computer architecture;Facial Expression;Recognition;Deep Learning;CNN;Architecture;Classification,
6. M. Shi, L. Xu and X. Chen, "A Novel Facial Expression Intelligent Recognition Method Using Improved Convolutional Neural Network," in *IEEE Access*, vol. 8, pp. 57606-57614, 2020, doi: 10.1109/ACCESS.2020.2982286. keywords: Face;Face recognition;Iron;Feature extraction;Education;Games;Clustering algorithms;Facial expression recognition;convolutional neural network;fuzzy C-means clustering;support vector machine;intelligent processing,
7. Fine-grained facial expression analysis using a dimensional emotion model Zhou F., Kong S., Fowlkes C.C., Chen T., Lei B. (2020) *Neuro-computing*, 392, pp. 38-49.
- 8.M. -I. Georgescu, R. T. Ionescu, and M. Popescu, "Local Learning With Deep and Handcrafted Features for Facial Expression Recognition," in *IEEE Access*, vol. 7, pp. 64827-64836, 2019, doi: 10.1109/ACCESS.2019.2917266. keywords: Face recognition;Training;Computational modeling;Support vector machines;Convolutional neural networks;Deep learning;Task analysis;Facial expression recognition;local learning;convolutional neural networks;bag-of-visual-words;dense-sparse-dense training, .
9. Jiang J, Wang M, Xiao B, Hu J and Deng W. (2024). Joint recognition of basic and compound facial expressions by mining latent soft labels. *Pattern Recognition*. [10.1016/j.patcog.2023.110173](https://doi.org/10.1016/j.patcog.2023.110173). 148. (110173).
10. Dong R and Lam K. Bi-Center Loss for Compound Facial Expression Recognition. *IEEE Signal Processing Letters*. [10.1109/LSP.2024.3364055](https://doi.org/10.1109/LSP.2024.3364055). 31. (641-645)..
- 11.Jiddah S and Yurtkan K. (2023). Dominant and complementary emotion recognition using hybrid recurrent neural network. *Signal, Image, and Video Processing*. [10.1007/s11760-023-02563-6](https://doi.org/10.1007/s11760-023-02563-6). 17:7. (3415-3423).
12. Canedo D and Neves A. (2019). Facial Expression Recognition Using Computer Vision: A Systematic Review. *Applied Sciences*. [10.3390/app9214678](https://doi.org/10.3390/app9214678).

9:21. (4678).

13. Karnati M, Seal A, Bhattacharjee D, Yazidi A and Krejcar O. Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey. *IEEE Transactions on Instrumentation and Measurement*. 10.1109/TIM.2023.3243661. 72. (1-31).

14. Deramgozin M, Jovanovic S, Arevalillo-Herráez M, Ramzan N and Rabah H. Attention-enabled lightweight Neural Network Architecture for Detection of Action Unit Activation. *IEEE Access*. 10.1109/ACCESS.2023.3325034. 11. (117954-117970).

15. Borgalli R and Surve S. (2023). Review on learning framework for facial expression recognition. *The Imaging Science Journal*. 10.1080/13682199.2023.2172526. 70:7. (483-521).