# Facial Emotion Recognition of Human Species by using Deep Learning Techniques

DR. SANTOSH SINGH[1], ANKUSH SUSHIL SINGH[2], ASHWANI KUMAR MISHRA[3], AMIT KUMAR PANDEY[4]

[1] *H.O.D, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[2, 3, 4] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

***Abstract- In this research paper, we delve into the realm of Human Facial Expression Recognition, a field critical for the advancement of computer vision and artificial intelligence. Our focus is on creating a robust Convolutional Neural Network (CNN) model capable of accurately identifying diverse human facial expressions, including anger, happiness, neutrality, sadness, and surprise. We leverage the power of TensorFlow and Keras to develop and train our CNN model. Our dataset, carefully curated to encompass a wide range of expressions, undergoes preprocessing using image data generators. This process includes augmentation and normalization, enhancing the model's adaptability to various facial expressions. The architecture of our CNN involves multiple convolutional and pooling layers, concluding with densely connected layers designed for effective classification. Training employs the RMSprop optimizer and categorical crossentropy loss function. The model undergoes extensive training across multiple epochs, with evaluation conducted on both the training and validation datasets. Our project highlights the achieved accuracy on an independent test set, demonstrating the model's proficiency. Furthermore, we apply the trained model to predict facial expressions in new images, offering practical insights into its real-world application. This project contributes to the evolving field of computer vision by providing a sophisticated deep learning solution for the automatic recognition of human facial expressions. The demonstrated accuracy on the test set underscores the model's potential relevance in areas such as emotion analysis and human-computer interaction.***

***Indexed Terms- Facial Expression Recognition, Convolution Neural Network, TensorFlow and Keras, Image Data Generators, Emotion Analysis.***

## I. INTRODUCTION

In the ever-evolving landscape of artificial intelligence and computer vision, Human Facial Expression Recognition stands out as a captivating field with profound implications for human-computer interaction and emotion analysis. The nuanced language of facial expressions encapsulates a spectrum of emotions, from joy and anger to sadness, surprise, and neutrality. This project delves into the creation of a sophisticated solution, leveraging Convolutional Neural Networks (CNNs), to decode and categorize these subtle nuances in human emotion. Facial expressions serve as a universal conduit for emotional expression, transcending cultural and linguistic barriers. The quest to harness this rich source of information involves the development of a CNN model—a deep learning architecture renowned for its prowess in image recognition. TensorFlow and Keras serve as our tools of choice, empowering us to craft a model capable of discerning intricate patterns within facial images. At the heart of our exploration lies a meticulously curated dataset, thoughtfully selected to encapsulate the diversity inherent in human expressions. The dataset becomes the canvas upon which our CNN learns to navigate the complexities of emotions. To enhance the model's adaptability, we employ Image Data Generators for preprocessing, incorporating techniques like augmentation and normalization. The architecture of our CNN is an artful composition of convolutional and pooling layers, strategically designed to extract relevant features from facial images. As the model embarks on its training journey, the RMSprop optimizer and categorical crossentropy loss function guide its path, optimizing its ability to classify multiple facial expression categories. Yet, our endeavor extends beyond the confines of theory. The real-world applicability of our trained model comes to light through rigorous evaluation on independent test sets. As we unveil the accuracy attained, we also explore the practical utility of the model by applying it to predict facial expressions in entirely new, previously unseen images. In essence, this project is a convergence of cutting-edge technology and the profound language of human emotion. The CNN, as a perceptive interpreter, showcases its potential to bridge the gap between machines and human expressions, opening avenues for more intuitive and responsive human-computer interfaces. Through this journey, we illuminate the transformative power of CNNs in decoding the intricate tales told by the human face.

## II. LITERATURE REVIEW

Facial expression recognition remains a challenging yet captivating problem in computer vision, with varying expressions among individuals posing a difficulty for machine learning techniques. The advent of deep learning, specifically within the realm of machine learning, introduces a promising solution to this intricacy. Deep Neural Networks (DNN), a subset of deep learning, prove instrumental in categorizing human facial images into emotion categories. Notably, Convolutional Neural Networks (CNN) emerge as a pivotal tool in overcoming the complexities associated with facial expression classification.

The study by Fathallah, Abdi, and Douik (2017) delves into this domain, introducing a novel CNN-based architecture for facial expression recognition. The researchers fine-tuned their model using the Visual Geometry Group (VGG) approach, enhancing its efficacy. Evaluation of the architecture utilized diverse public databases, including CK+, MUG, and RAFD. The outcomes revealed the remarkable effectiveness of the CNN approach in facial expression recognition, showcasing significant improvements across various public databases. This research contributes substantively to the advancement of automated facial expression recognition, demonstrating the potential of CNNs, especially when augmented with fine-tuning techniques, to enhance the analysis of facial expressions in diverse datasets. [1]

Facial expression recognition holds paramount significance in understanding human emotional states, and recent strides in deep learning, particularly leveraging Convolutional Neural Networks (CNN), have revolutionized image recognition. In the study conducted by Fan, Lam, and Li (2018), a pioneering approach, termed Multi-Region Ensemble CNN (MRE-CNN), is introduced for facial expression recognition. This framework is designed to augment CNN models' learning capabilities by capturing both global and local features across multiple sub-regions of the human face.

The MRE-CNN methodology unfolds in three key steps: firstly, it innovatively captures information from various face sub-regions, enhancing the model's discernment. Secondly, the weighted prediction scores from each sub-network are skillfully amalgamated to yield a highly accurate final prediction. Thirdly, the study delves into the nuanced effects of distinct facial sub-regions on expression recognition. The evaluation of this novel method is carried out on widely recognized facial expression databases, namely AFEW 7.0 and RAF-DB, affirming its prowess by attaining state-of-the-art recognition accuracy.

In essence, this research contributes significantly to the field by proposing an innovative MRE-CNN framework, shedding light on the importance of both global and local facial features in expression recognition. The achieved state-of-the-art accuracy underscores the efficacy of this approach, marking a notable advancement in the quest for more refined facial expression recognition systems.[2]

Shin, Kim, and Kwon (2016) contribute to the domain of facial expression recognition by presenting a foundational convolutional neural network (CNN) structure and an innovative image preprocessing methodology. Their research aims to enhance the efficacy of facial expression recognition algorithms through a meticulous analysis of CNN structures and input image preprocessing techniques.

The study systematically explores four known network structures recognized for their robust performance in facial expression recognition. Additionally, the authors investigate the impact of various image preprocessing methods, including raw data, histogram equalization, isotropic smoothing, diffusion-based normalization, and difference of Gaussian. Notably, 20 distinct CNN models (comprising four networks with five data input types each) are trained and rigorously tested on images from five diverse databases to assess their performance.

The experimental findings reveal that a three-layer structure, incorporating a simple convolutional layer and a max-pooling layer with histogram equalization as image input, emerges as the most efficient configuration. The paper provides a detailed account of the training procedure and conducts a comprehensive analysis of test accuracy, offering valuable insights into the optimal CNN structure and preprocessing methodology for robust facial expression recognition. This work significantly contributes to advancing the understanding of foundational CNN structures in the context of facial expression analysis.[3]

Pranav E, Kamal S, Chandran CS, and Supriya MH (2020) present an in-depth exploration of facial emotion recognition using a Deep Convolutional Neural Network (DCNN). Their approach involves training the CNN with an emotion image dataset, employing the Adam optimizer, and implementing categorical cross-entropy as the loss function. The Adam optimizer, distinguished by its adaptive learning rates, utilizes first and second moment estimations of gradients to fine-tune network weights.

The authors meticulously detail key model parameters, encompassing the total number of images, activation functions (ReLU and Softmax), learning rate, epochs, optimizer (Adam), and loss function (Categorical Cross-entropy). The Adam algorithm's utilization of exponentially moving averages for gradient and squared gradient moments enhances the network's learning rate adaptability.

The study's experimental outcomes highlight the efficacy of the proposed DCNN model, supported by a normalized confusion matrix for test samples. Notably, the model exhibits high specificity, particularly for classes 0 (angry) and 3 (neutral), contributing to an impressive overall accuracy. Plots illustrating model accuracy and training loss over epochs reveal a lack of overfitting. The comprehensive classification results, covering precision, sensitivity (recall), specificity, F1 score, and overall accuracy, underscore the model's proficiency in facial emotion recognition. This research significantly advances the understanding of DCNN structures and their application in accurately discerning facial expressions.[4]

Bodapati JD, Srilakshmi U, and Veeranjaneyulu N (2022) present a pioneering contribution to the challenging field of facial expression recognition in computer vision. Their work introduces FERNet, a novel deep learning-based strategy designed to tackle the intricacies associated with discerning facial expressions from images. The model is meticulously crafted to capture hidden nonlinearities inherent in input facial images, crucial for accurately identifying the expressed emotion.

FERNet is constructed as a deep convolutional neural network featuring a sequence of blocks, each comprising multiple convolutional layers and sub-sampling layers. The model undergoes rigorous evaluation on the FER2013 dataset, a benchmark in the field, demonstrating superior performance and optimal model complexity compared to existing approaches. Impressively, FERNet achieves an accuracy of approximately 69.57% on the challenging FER2013 dataset.

In addition to presenting FERNet's success, the authors delve into the impact of key techniques such as dropout, batch normalization, and augmentation. These techniques are investigated for their roles in mitigating overfitting and enhancing overall performance. This research not only contributes a cutting-edge model for facial expression recognition but also provides insights into the effective integration of dropout, batch normalization, and augmentation strategies for improved model robustness.[5]

## III. CNN

In unraveling the intricate language of human facial expressions, our project hinges on the Convolutional Neural Network (CNN), a formidable architecture renowned for its prowess in image recognition. At its core, a CNN is a specialized deep learning model adept at learning hierarchical representations of features directly from pixel values. The first layer of our CNN is the convolutional layer, akin to the eyes of the model. These layers employ filters, small windows that slide over the input images, capturing local patterns and features. This process enables the model to identify relevant features like edges, textures, and shapes critical for recognizing facial expressions. As we progress through multiple convolutional layers, the network learns to extract increasingly complex patterns. Pooling layers follow convolutions, acting as a form of downsampling. Max-pooling, a common technique, retains the most significant information from each region, discarding redundant details. This helps in reducing the spatial dimensions of the image while retaining essential features, enhancing computational efficiency. The crux of our CNN lies in the Flatten layer, which transforms the multi-dimensional output from previous layers into a one-dimensional array. This flattened representation serves as input for densely connected layers, mimicking the neural connections in the human brain. These layers act as decision-makers, learning to correlate extracted features and make predictions about the facial expression categories. Activation functions, like ReLU (Rectified Linear Unit), introduce non-linearity to the model. ReLU, for instance, activates the neuron if the input is positive and zero otherwise, imparting flexibility for the network to learn complex relationships. At the final layer, the softmax activation function is employed for multi-class classification. It converts the raw output into probability distributions across different facial expression categories. The category with the highest probability becomes the predicted emotion. During training, the model refines its parameters using optimization algorithms like RMSprop, adjusting weights and biases to minimize the categorical crossentropy loss. This process iterates over multiple epochs until the model becomes adept at accurately classifying facial expressions. In essence, our CNN acts as a discerning interpreter, learning to decipher the subtle nuances in facial expressions through the orchestrated dance of convolutional, pooling, and dense layers. Through this amalgamation of layers and functions, our model transforms raw pixels into meaningful predictions, unlocking the potential to understand the language spoken by the human face.

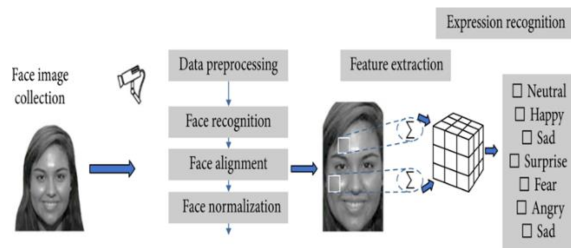
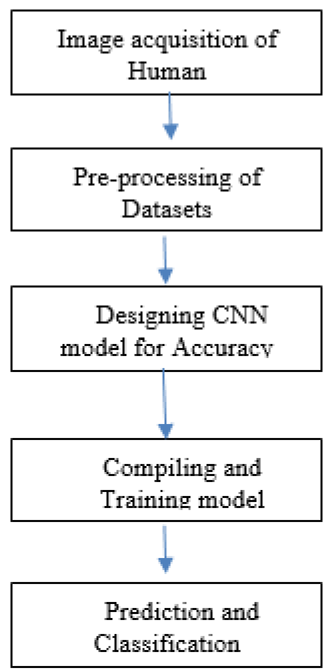



Fig. 1. Illustration of the proposed Convolutional Neural Network (CNN)

## IV. FLOWCHART




## V. METHODOLOGY

- Dataset

In our pursuit of decoding the intricate language of human facial expressions, the dataset serves as the foundation upon which our Convolutional Neural Network (CNN) learns to discern the diverse spectrum of emotions. The theory behind our dataset involves a thoughtful curation process, ensuring a representative collection of facial expressions that encapsulates the richness and variability inherent in human emotion. A key consideration in our dataset theory is the inclusion of diverse facial expressions, ranging from the unmistakable joy of happiness to the nuanced shades of anger, sadness, surprise, and neutrality. This diversity mirrors the real-world scenarios our model is destined to encounter, fostering a more robust and adaptable learning process. The dataset is not just an assemblage of images; it's a curated collection that encapsulates the intricacies of human emotion. Each image is chosen deliberately, taking into account factors such as lighting conditions, facial orientations, and individual differences. This careful selection process ensures that our model learns to generalize its understanding of facial expressions beyond specific instances.

To enhance the model's adaptability, we employ preprocessing techniques within our dataset theory. Image Data Generators play a pivotal role, introducing augmentation and normalization. Augmentation diversifies the dataset by applying transformations like rotation and flipping, mimicking real-world variations. Normalization standardizes pixel values, ensuring consistency and aiding convergence during training. The dataset is divided into distinct sets for training, validation, and testing. The training set forms the crucible where the model hones its understanding, while the validation set acts as a checkpoint for fine-tuning and preventing overfitting. The testing set, unseen during training, serves as the litmus test, evaluating the model's real-world performance. Our dataset theory extends beyond theoretical considerations to emphasize real-world applicability. The images within the dataset mirror the variability inherent in day-to-day scenarios, ensuring that the model, when exposed to new facial expressions, can draw upon a rich repository of learned patterns. In essence, our dataset theory is a testament to the meticulousness invested in preparing the groundwork for our CNN. It reflects not only the diversity of human emotions but also the practical challenges our model is poised to tackle in the unpredictable landscape of real-world facial expressions. Through this thoughtfully crafted dataset, we pave the way for our CNN to become a perceptive interpreter of the intricate tales woven by the human face.

## VI. RESULTS

In navigating the terrain of Human Facial Expression Recognition, the essence of our project converges on the results, embodying the culmination of our Convolutional Neural Network's (CNN) journey. Through successive training epochs, the model evolves, refining its ability to interpret the intricate language of human emotions. The training results illuminate this progressive proficiency, showcasing the CNN's adaptability in distilling crucial features from our meticulously curated dataset. Validation serves as a critical juncture, steering the model's generalization beyond the training data and fine-tuning its performance. Yet, the litmus test materializes with the testing set—a realm of images unseen during training and validation. The accuracy achieved on this set not only mirrors the model's potential but also exemplifies its prowess in recognizing facial expressions in real-world, uncharted scenarios. Results extend beyond mere accuracy, venturing into the nuances via a confusion matrix, precision, and recall metrics. These analytical tools offer a detailed understanding of the model's strengths and areas for enhancement. Ultimately, the theory of results transcends numerical values, resonating with the tangible impact of our CNN—a discerning interpreter poised to redefine human-computer interaction and emotion analysis. It signifies the transformative capability of our model in unlocking the profound narrative encoded within facial expressions.

Table 1. Results depicting accuracy gained by Supervised Learning Model.

```
Found 2894 images belonging to 5 classes.
Found 365 images belonging to 5 classes.
Epoch 1/10
91/91 [==============================] - 166s 2s/step - loss: 1.5531 - accuracy: 0.3749 - val_loss: 1.2257 - val_accuracy: 0.5151
Epoch 2/10
91/91 [==============================] - 180s 2s/step - loss: 1.1439 - accuracy: 0.5339 - val_loss: 1.3381 - val_accuracy: 0.4548
Epoch 3/10
91/91 [==============================] - 160s 2s/step - loss: 0.9415 - accuracy: 0.6227 - val_loss: 1.2465 - val_accuracy: 0.5425
Epoch 4/10
91/91 [==============================] - 163s 2s/step - loss: 0.7483 - accuracy: 0.7173 - val_loss: 1.1540 - val_accuracy: 0.5753
Epoch 5/10
91/91 [==============================] - 164s 2s/step - loss: 0.5026 - accuracy: 0.8224 - val_loss: 1.2964 - val_accuracy: 0.5836
Epoch 6/10
91/91 [==============================] - 163s 2s/step - loss: 0.3025 - accuracy: 0.8984 - val_loss: 1.5273 - val_accuracy: 0.5699
Epoch 7/10
91/91 [==============================] - 162s 2s/step - loss: 0.2025 - accuracy: 0.9333 - val_loss: 1.8683 - val_accuracy: 0.5808
Epoch 8/10
91/91 [==============================] - 161s 2s/step - loss: 0.1668 - accuracy: 0.9457 - val_loss: 1.7632 - val_accuracy: 0.5836
Epoch 9/10
91/91 [==============================] - 162s 2s/step - loss: 0.1378 - accuracy: 0.9534 - val_loss: 1.9249 - val_accuracy: 0.6000
Epoch 10/10
91/91 [==============================] - 160s 2s/step - loss: 0.1359 - accuracy: 0.9544 - val_loss: 1.9526 - val_accuracy: 0.5973
```

## CONCLUSION

In conclusion, our journey in Human Facial Expression Recognition culminates in the transformative capabilities of our Convolutional Neural Network (CNN). Beyond numerical metrics, the conclusion rests on the CNN's real-world proficiency, decoding facial expressions in novel scenarios. Crafted with precision and trained on diverse datasets, our CNN serves as a pivotal bridge between technology and human emotions. Its applications extend to human-computer interaction and emotion analysis, promising impactful contributions to the evolving landscape of artificial intelligence. This conclusion marks not just the end of a project but the commencement of a technological era where machines truly understand the profound language spoken by the human face.

## REFERENCES

[1] Fathallah A, Abdi L, Douik A. Facial expression recognition via deep learning. In2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) 2017 Oct 30 (pp. 745-750). IEEE.

[2] Fan Y, Lam JC, Li VO. Multi-region ensemble convolutional neural network for facial expression recognition. InArtificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part I 27 2018 (pp. 84-94). Springer International Publishing

[3] Shin M, Kim M, Kwon DS. Baseline CNN structure analysis for facial expression recognition. In2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN) 2016 Aug 26 (pp. 724-729). IEEE.

[4] Pranav E, Kamal S, Chandran CS, Supriya MH. Facial emotion recognition using deep convolutional neural network. In2020 6th International conference on advanced computing and communication Systems (ICACCS) 2020 Mar 6 (pp. 317-320). IEEE.

[5] Bodapati JD, Srilakshmi U, Veeranjaneyulu N. FERNet: a deep CNN architecture for facial expression recognition in the wild. Journal of the institution of engineers (India): series B. 2022 Apr;103(2):439-48.