Human Sentiment Analysis Using DenseNet-169: A Deep Learning Approach

Prateek Majumder Neha Roy Choudhury Yuvraj Singh Srinet

prateek.m25pgai@jioinstitute.edu.in neha.r25pgai@jioinstitute.edu.in yuvraj.s25pgai@jioinstitute.edu.in

Abstract

Human sentiment analysis has become the key to understanding emotions and opinions through applications in NLP and computer vision. This work examines, for the first time, the state-of-the-art application of DenseNet-169 for CNN-based image sentiment analysis, concerning the applications of facial sentiment expressions and extended to be integrated into NLP models for a holistic sentiment analysis. In this regard, we delve into the architecture of DenseNet-169, describe its capability regarding the extraction of deep hierarchical features, and demonstrate its effectiveness in sentiment prediction with an extensive experimental analysis. Pre-training of deep convolutional neural networks (DCNNs) plays a crucial role in the field of visual sentiment analysis (VSA)[1]. The results indicate that DenseNet-169 outperforms traditional models in image-based sentiment analysis and can be extended to multimodal approaches that combine textual data.

1. Introduction

Sentiment analysis, with the advent of social media and other digital platforms, has been a wide attractant in recent memory due to its applications relating to marketing, customer feedback, and public opinion mining. Deep learning algorithms have been proven to achieve remarkable outcomes across a broad spectrum of applications. Examining feelings conveyed by images is complex, but there is much space for development [2].

Conventionally, sentiment analysis makes use of NLP techniques that classify sentiments in text as positive, negative, or neutral. There is an increasing interest, however, in analyzing visual content, especially in the form of facial expressions, since these provide a more nuanced understanding of sentiment.

In this work, we employ DenseNet-169—a deep learning architecture recently developed among the top-performing methods for image recognition tasks—to analyze human sentiment from facial images. DenseNet-169 is a variation of a Dense Convolutional Network, Dense Net, where a densely connected structure is introduced to improve the diversity of information flowing between successive layers for better feature extraction. We aim to see how well DenseNet-169 will do in face sentiment classification.

2. Related Work

2.1 Sentiment Analysis in Text

Traditional approaches to sentiment analysis rely on NLP methodologies including but not limited to bag-of-words, TF-IDF, and word embeddings such as Word2Vec and GloVe. Advanced methods employ deep learning models, for instance, Long Short-Term Memory (LSTM) networks, Transformer-based architectures—for example, BERT or GPT—and attention mechanisms that may give information about context-dependent sentiments.

2.2 Image Sentiment Analysis

Facial expressions are one of the most revealing non-verbal signals that depict emotions. Traditional approaches were based on hand-crafted features, and Histogram of Oriented Gradients along with Local Binary Patterns were extensively used in this respect. Very recently, state-of-the-art results were achieved by using deep learning methods and especially CNNs. Pre-trained CNNs like VGG-16, ResNet, and DenseNet demonstrated good generalization capability on image classification tasks, including the challenge of recognizing facial emotions.

2.3 DenseNet-169

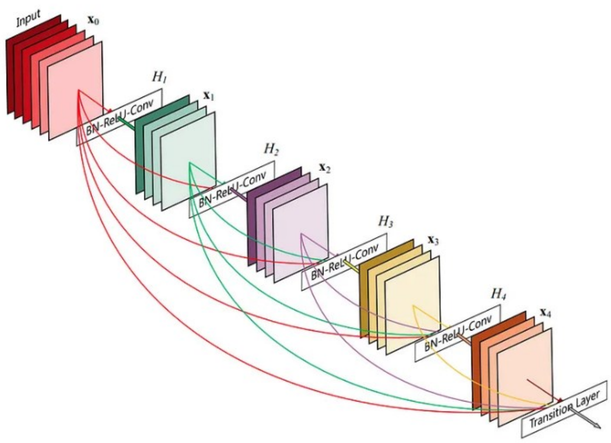
DenseNet, proposed by Huang et al. (2017), applies dense connectivity between layers for utmost information flow [3]. In this way, it helps in better flow of gradients and promotes feature re-use. Generally, this 169-layer network has been widely used in image classification and object detection regarding its efficiency in parameter and computational resource usage.

3. DenseNet-169 Architecture

DenseNet-169 has a total of four dense blocks, each with a set of convolutional layers that are connected with all preceding layers through direct connections [4]. This architecture contributes to acquiring multiple variations of features at deeper depths, hence making this model very useful for eventualities requiring complex image understanding and processing, as in the case of emotion recognition from facial images.

3.1 Dense Blocks and Transition Layers

Each dense block comprises batch normalization, ReLU activation, and 3x3 convolution layers. The transition layers between the dense blocks comprise 1x1 convolution layers followed by 2x2 average pooling layers, which further facilitate the dimensionality reduction of the feature maps with the preservation of critical information.



DenseNet architecture (Huang et al.)

The DenseNet169 architecture is composed of several types of layers including convolutional, maxpool, dense, and transition layers [5].

3.2 Feature Propagation

DenseNet-169 offers feature propagation by providing each layer with all the feature maps of the preceding layers. It improves parameter utilization and solves the problem of vanishing gradients that may happen in a very deep network.

3.3 Network Parameters

DenseNet-169 has 14.3 million parameters [6], while ResNet-152 contains 60.2 million parameters; hence, DenseNet-169 can be thought of as a light network as compared to other deep networks like ResNet-152. This efficiency enables DenseNet-169 to process high-resolution images with no significant additional computational cost and hence is very suitable for real-time applications of sentiment analysis.

4. Methodology

4.1 Dataset

We used, for experimentation, the FER-2013 dataset—the most well-known benchmark in facial emotion recognition—consisting of 35,887 grayscale images of size 48x48 pixels [7]. Overall, the dataset consists of seven categories of emotion: anger, disgust, fear, happiness, sadness, surprise, and neutral.

4.2 Data Preprocessing

To prepare the dataset for training, we utilized TensorFlow's **ImageDataGenerator** with the preprocessing function from DenseNet. This ensured that the images were standardized appropriately for input into the model.

4.3 Data Augmentation

For the training dataset, we implemented several augmentation techniques to enhance model robustness, including:

- **Horizontal Flipping:** Randomly flipping images horizontally.
- **Width Shift:** Shifting images horizontally up to 10%.
- **Height Shift:** Shifting images vertically up to 5%.
- **Rescaling:** Normalizing pixel values to the range of [0, 1].

4.4 Data Generators

We created three separate data generators: the training generator for loading and augmenting images from the training directory, the validation generator for loading images from the validation split of the training dataset, and the test generator for loading images from the test directory without any shuffling. The dataset comprised 22,968 training images, 5,741 validation images, and 7,178 test images, all belonging to 7 classes. This methodology aimed to improve the model's generalization and overall performance across various image classes.

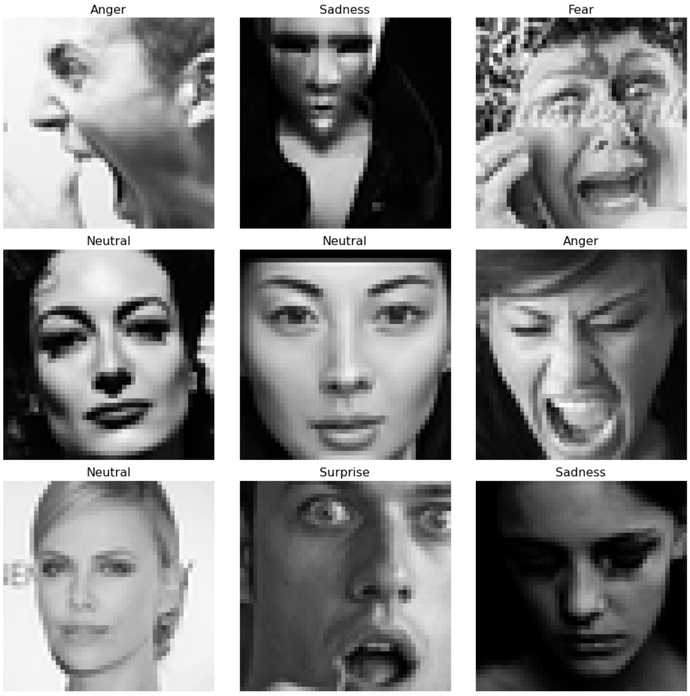


Image Classes and Data

4.5 Data Distribution

We visualized the distribution of training images across different emotions using a bar chart, with the x-axis representing emotion classes (labeled with emojis) and the y-axis showing the number of images. The chart effectively highlights the balance of data among the seven emotion categories in the training set.

4.6 DenseNet169 Transfer Learning

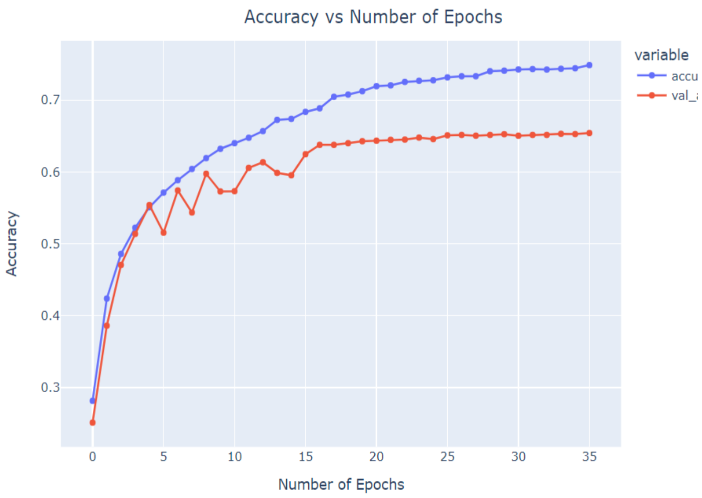
We employed DenseNet169 as a feature extractor for transfer learning. The architecture begins by defining the `feature extractor` function, which initializes DenseNet169 without the top layers and utilizes ImageNet weights. The output from this feature extractor undergoes global average pooling, followed by a series of fully connected layers with ReLU activations and L2 regularization to enhance model generalization. Specifically, we included three dense layers with 256, 1024, and 512 units, respectively, interspersed with dropout layers to mitigate overfitting.

The final output layer consists of SoftMax activation, catering to seven emotion classes. The model is compiled using Stochastic Gradient Descent (SGD) with a learning rate of 0.1 and categorical cross-entropy as the loss function. Initially, the feature extraction layers of DenseNet169 are frozen, preventing their weights from being updated during training. The model summary indicates a total of 13,860,679 parameters, of which 1,217,799 are trainable, providing a robust foundation for emotion classification tasks.

```
Model: "model"
_________________________________________________________________
Layer (type)                Output Shape              Param #
=================================================================
input_1 (InputLayer)        [(None, 48, 48, 3)]       0
_________________________________________________________________
densenet169 (Functional)    (None, 1, 1, 1664)        12642880
_________________________________________________________________
global_average_pooling2d (Gl (None, 1664)             0
_________________________________________________________________
dense (Dense)               (None, 256)               426240
_________________________________________________________________
dropout (Dropout)           (None, 256)               0
_________________________________________________________________
dense_1 (Dense)             (None, 1024)              263168
_________________________________________________________________
dropout_1 (Dropout)         (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)             (None, 512)               524800
_________________________________________________________________
dropout_2 (Dropout)         (None, 512)               0
_________________________________________________________________
classification (Dense)      (None, 7)                 3591
=================================================================
Total params: 13,860,679
Trainable params: 1,217,799
Non-trainable params: 12,642,880
```

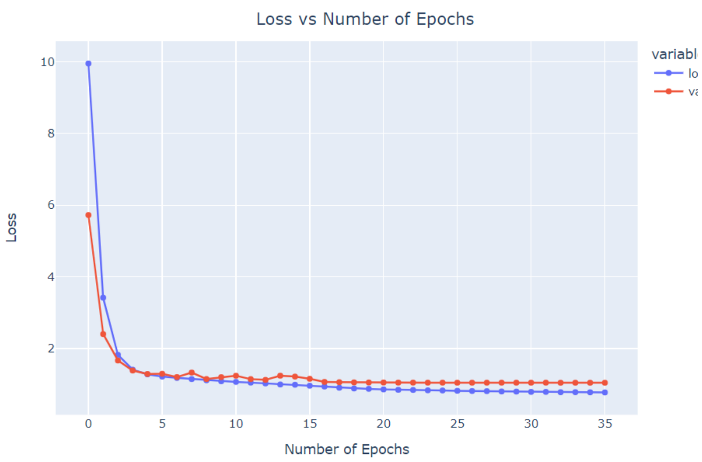
4.7 Training and Fine-Tuning

We started our training process using the DenseNet169 architecture, where we initially froze the feature extraction layers to make the most of the pre-trained weights from ImageNet. To ensure we didn’t overfit, we implemented an EarlyStopping callback, which monitored the validation loss and allowed us to halt training when improvements plateaued, preserving the best version of our model. Over 16 epochs, we saw a significant boost in accuracy, climbing from 28.17% in the first epoch to 62.46% by the end, while the validation loss steadily decreased.



Blue: Accuracy **Red**: Validation Accuracy

After completing the initial training phase, we moved on to fine-tuning. We un freezed the DenseNet169 layers to enable further optimization of the entire model, using a lower learning rate of 0.001 to avoid making any drastic changes to the weights. This fine-tuning phase lasted for an additional 20 epochs, during which our model achieved an accuracy of 74.86%. Validation accuracy stayed at 65.41 %This two-step training approach allowed us to gradually enhance the model's ability to classify emotions effectively, combining the strengths of transfer learning with targeted fine-tuning.



Blue: Loss **Red**: Val Loss

We tracked how the model's performance changed over time, specifically looking at the loss during training and validation. This helps us understand if the model is improving, and whether it's learning well without overfitting. The plot makes it easy to see these trends across the epochs.

4.8 Model Evaluation and Confusion Matrix

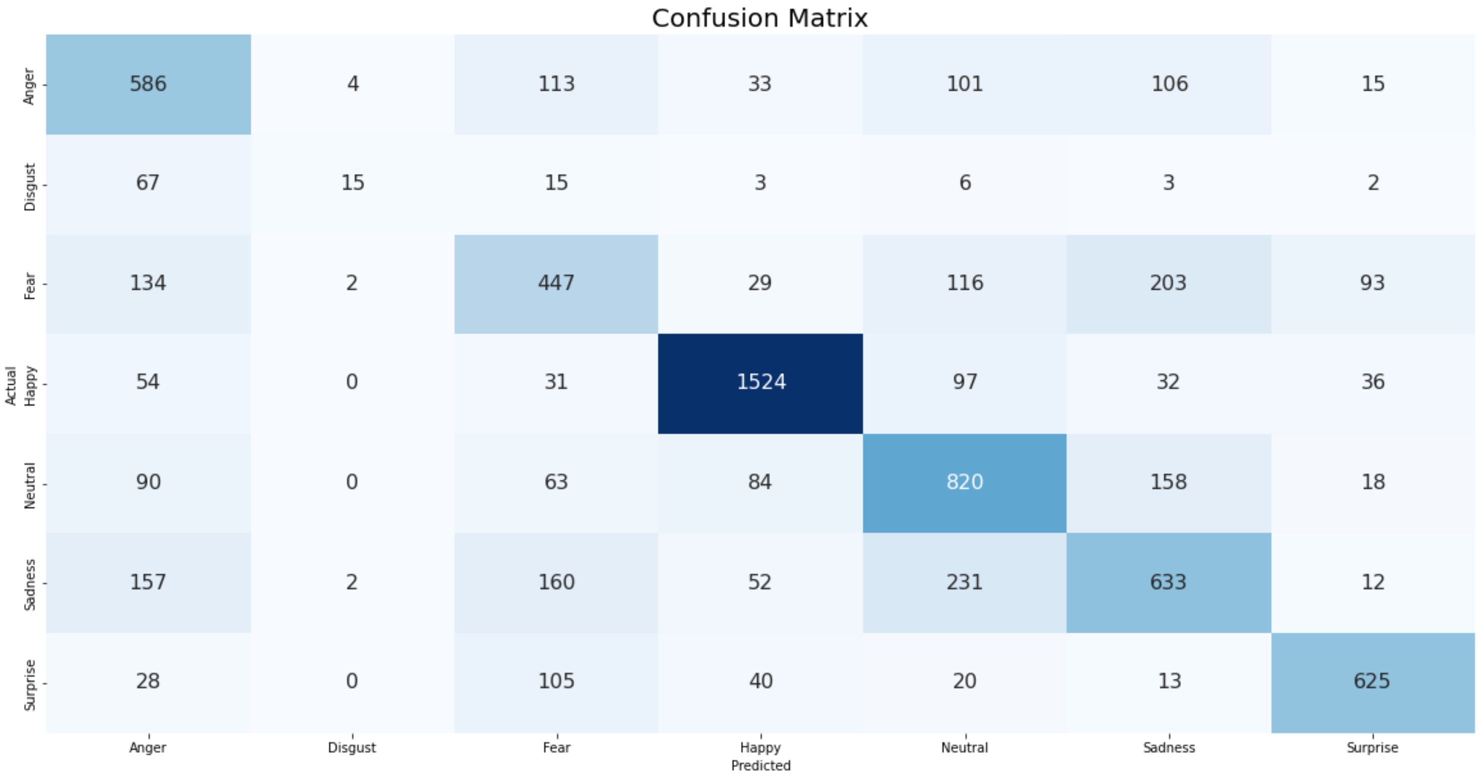
To evaluate the model's performance, we tested it on unseen data and achieved a 64.78% accuracy with a loss of 1.0661. To further understand how well the model predicted each class, we created a confusion matrix, which shows how often the model correctly predicted each class versus when it misclassified. The confusion matrix is visualized using a heatmap, where darker shades represent better accuracy, making it easier to spot where the model struggled and performed well across the different categories.

```
113/113 [==============================] - 29s 259ms/step
- loss: 1.0661 - accuracy: 0.6478
```

4.9 Classification Report

class	precision	recall	f1-score	support
0	0.53	0.61	0.57	958
1	0.65	0.14	0.22	111
2	0.48	0.44	0.46	1024
3	0.86	0.86	0.86	1774
4	0.59	0.67	0.62	1233
5	0.55	0.51	0.53	1247
6	0.78	0.75	0.77	831
accuracy			0.65	7178
macro avg	0.63	0.57	0.58	7178
weighted avg	0.65	0.65	0.64	7178

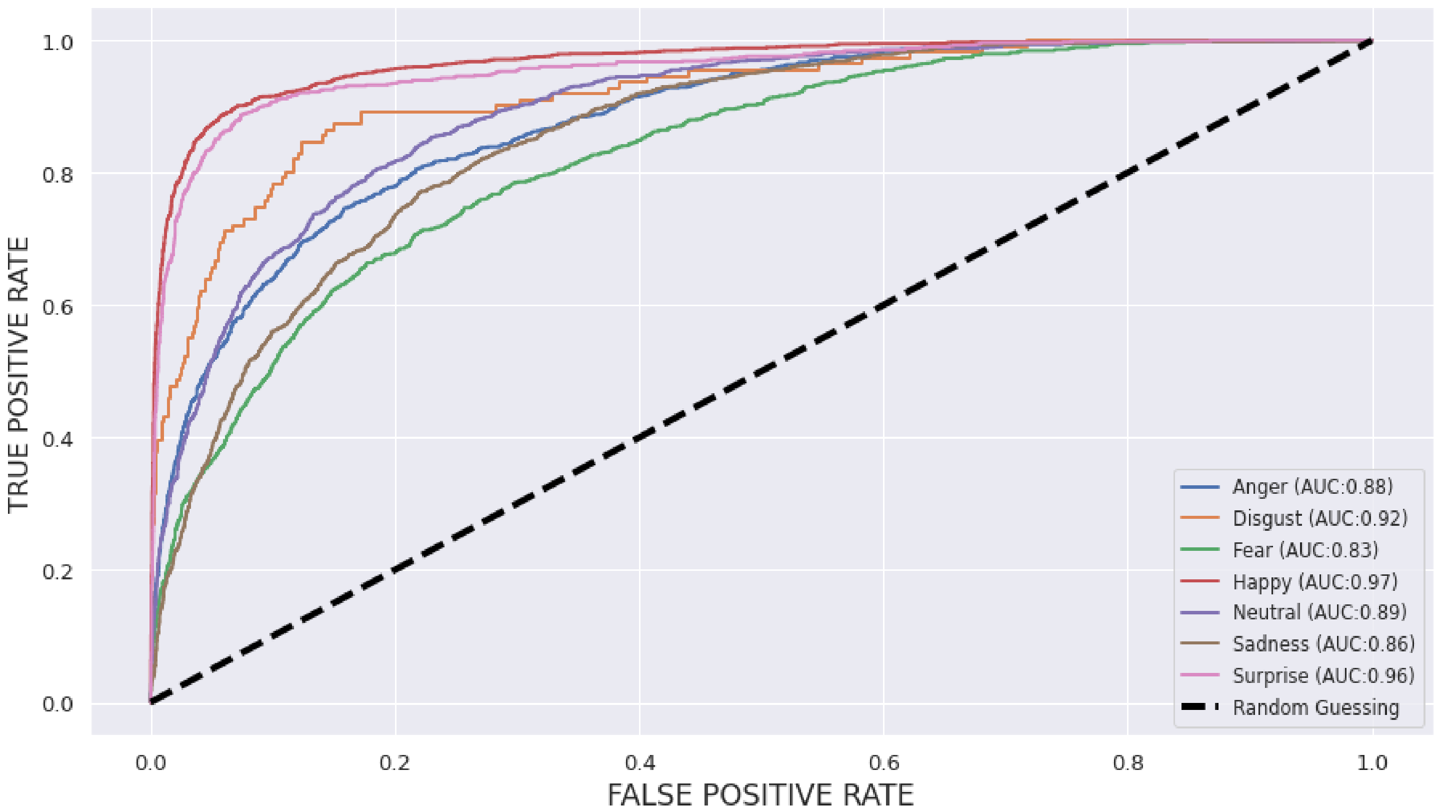
The model's evaluation on the test data shows a loss of 1.0661 and an accuracy of 64.78%. The confusion matrix provides a detailed view of the model's predictions compared to actual values, displaying its performance across multiple categories [8]. The classification report presents precision, recall, and F1-scores for each category, leading to an overall accuracy of 65% and a weighted average F1-score of 0.64. This suggests that the model performs reasonably well on the test set, though there is some variability in performance across different categories.



4.10 Multiclass ROC AUC

The multiclass ROC AUC curve provides a visual representation of the model's performance across various classes, displaying how well it can differentiate between them [9]. Each class is represented by its own curve, and the AUC values highlight the model's ability to accurately classify the emotions.

The overall ROC AUC score of about 0.92 indicates a robust performance, suggesting that the model is effective in distinguishing between different emotional states. A secondary calculation also yields a score of approximately 0.90, reinforcing the model's strong classification capabilities.



Discussion:

Our exploration of emotion classification using DenseNet169 has yielded valuable insights. The DenseNet architecture is known for its ability to mitigate the vanishing gradient problem and improve feature propagation through dense connections between layers [10]. This characteristic is particularly advantageous in tasks requiring nuanced understanding, such as emotion recognition from images. The model's overall accuracy of approximately 65% is commendable, especially considering the complexities involved in classifying subtle emotional expressions. The confusion matrix revealed specific misclassifications, particularly in emotions that may share similar visual features, highlighting the challenges the model faces in differentiating these nuanced categories. The high ROC AUC scores, around 0.92 and 0.90, indicate a strong ability to distinguish between classes, suggesting that the model has learned to recognize key features associated with different emotions.

Conclusion:

In conclusion, utilizing DenseNet169 for emotion classification demonstrates a solid foundation for effectively identifying and distinguishing between various emotions based on image data. The model's strengths lie in its capacity to learn intricate features, evidenced by its performance metrics. However, the evaluation revealed areas for improvement, particularly for specific emotional classes that the model struggled to classify accurately.

References:

1. T. Feng, J. Liu and J. Yang, "Probing Sentiment-Oriented PreTraining Inspired by Human Sentiment Perception Mechanism," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 2850-2860, doi: 10.1109/CVPR52729.2023.00279.
2. Gaurav Meena, Krishna Kumar Mohbey, Sunil Kumar, Sentiment analysis on images using convolutional neural networks-based Inception-V3 transfer learning approach, International Journal of Information Management Data Insights, Volume 3, Issue 1, 2023, 100174, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2023.100174.
3. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700-4708. https://doi.org/10.1109/CVPR.2017.243
4. Deep Transfer Learning Based Rice Plant Disease Detection Model R. P. Narmadha1,*, N. Sengottaiyan2 and R. J. Kavitha3 Intelligent Automation & Soft Computing DOI:10.32604/iasc.2022.020679
5. An Adaptive Early Stopping Technique for DenseNet169-Based Knee Osteoarthritis Detection Model. Bander Ali Saleh Al-rimy 1,*, Faisal Saeed 2, Mohammed Al-Sarem 3, Abdullah M Albarrak 4, Sultan Noman Qasem 4
6. COVID-DSNet: A novel deep convolutional neural network for detection of coronavirus (SARS-CoV-2) cases from CT and Chest X-Ray images, Hatice Catal Reis a,∗, Veysel Turk b
7. Facial Emotion Recognition: State of the Art Performance on FER2013 8 May 2021 · Yousif Khaireddin, Zhuofa Chen ·
8. How to interpret a confusion matrix for a machine learning model https://www.evidentlyai.com/classification-metrics/confusion-matrix
9. How to explain the ROC curve and ROC AUC score? https://www.evidentlyai.com/classification-metrics/explain-roc-curve
10. Huang, Gao & Liu, Zhuang & van der Maaten, Laurens & Weinberger, Kilian. (2017). Densely Connected Convolutional Networks. 10.1109/CVPR.2017.243.