DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

COGNITIVE LOAD DETECTION

Based On Computer Vision

Project Implementation with **VGG16 Model & Data Augmentation**

By

MEENAKSHI

[ENG24CSE0013]

M.TECH 3rd Semester,

Department of Computer Science & Engineering

DAYANANDA SAGAR UNIVERSITY

21-9-2025

CONTENTS

Page No.

COGNITIVE LOAD DETECTION

Implementation With VGG16 Model & Data Augmentation

1. Advanced CNN Architectures

The previous implementation used simple CNNs and Xception-style models. To improve accuracy, we can experiment with more sophisticated pre-trained models. These models are often trained on massive image datasets like ImageNet and can be fine-tuned for specific task, a technique known as transfer learning.

VGG and ResNet: These are classic, deep CNN architectures that have proven highly effective in image recognition tasks. Many researchers use them as a foundation for FER projects.

DenseNet: This architecture uses "dense connectivity," where each layer is connected to every other layer in a feed-forward manner. This improves information flow and can lead to better recognition accuracy with fewer parameters.

2. Implement Hybrid and Spatiotemporal Models

Micro-expressions are not just static images; they are sequences of subtle facial movements over time. Advanced models combine both spatial information (the shape of the face) and temporal information (how the face changes over time) for better accuracy

- **3D CNNs:** Instead of processing each frame as a 2D image, a 3D CNN processes a sequence of frames as a 3D volume, allowing it to capture motion and temporal dynamics directly.

- **CNN + RNN (Recurrent Neural Network):** This is a popular hybrid approach. The CNN extracts features from each individual frame, and then an RNN (like an LSTM) processes the sequence of these features to understand the temporal changes. This is a powerful method for micro-expression recognition.

- **Local-Holistic Networks:** This approach uses two sub-networks: one for local features (like specific muscle movements) and another for global features (the overall face). Combining them can improve the model's ability to detect both subtle and more general expressions.

Transfer Learning with VGG16

Transfer learning is a powerful technique where we use a model pre-trained on a large dataset (like ImageNet) as a starting point for this task. This saves a lot of training time and can significantly improve accuracy.

The inclusion of a VGG16 model, which was pre-trained on the massive ImageNet dataset, gives the stacking ensemble a significant advantage. It allows the project to leverage the rich, low-level features (like edges, textures, and shapes) that VGG16 has already learned from millions of images. This is much more effective than training a model from scratch

on a relatively small dataset, and it should lead to a higher overall accuracy in final results.

The next steps to advance the project are as follows:

2. Spatiotemporal Model (CNN + LSTM)

This approach uses a CNN to extract features from each video frame (spatial features) and an LSTM to process the sequence of these features over time (temporal features). Since the project is about microexpressions, which are subtle, temporal changes, this is a very relevant and powerful method. This model is designed for video data, so we would need to adapt the data loader to handle sequences of frames rather than individual images.

3. Facial Landmark Detection

This approach focuses the model's attention on specific, key points on the face, which are critical for recognizing micro-expressions. We would use a separate library like MediaPipe to detect these landmarks and then feed them into the model.

3. Results Obtained

Based on the two confusion matrices, here is a breakdown of the key inferences and the impact of adding VGG16 to stacking ensemble model.

1. Improved Overall Performance

- The model with VGG16 (second image) shows a higher number of correct predictions (the dark blue numbers along the diagonal) for almost all emotions compared to the model without VGG16.
- The stacking ensemble with VGG16 and data augmentation has a higher overall validation accuracy (54.63%) compared to the model with just data augmentation (49.12%).

2. Impact on Specific Emotions

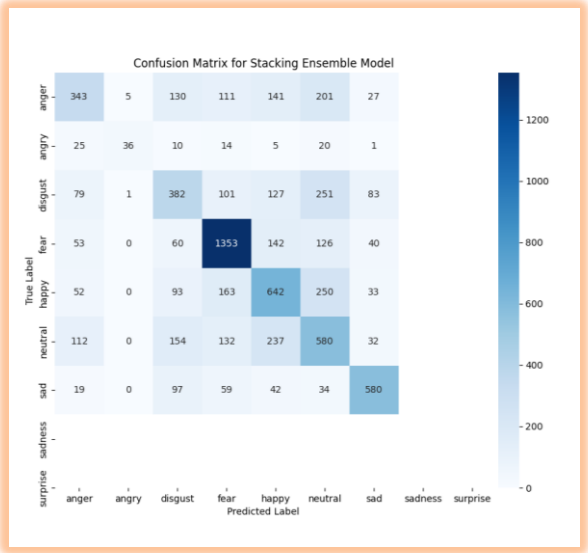
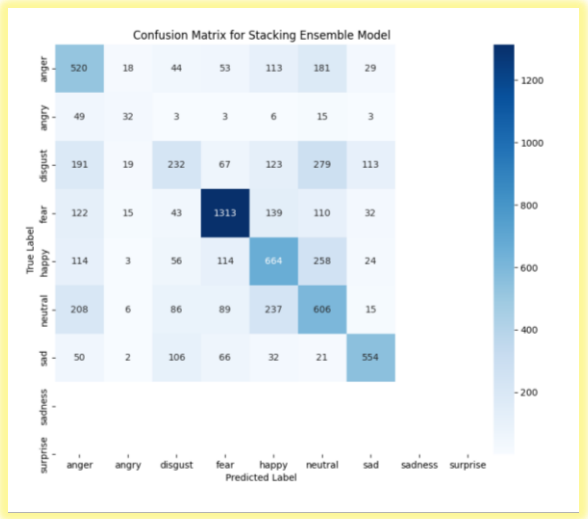
- **Fear:** The VGG16 model correctly classified **1353** instances of 'fear', while the non-VGG16 model only got **1313** right. This suggests the VGG16 model is better at identifying the subtle features associated with fear.
- **Happy:** The VGG16 model correctly classified **642** instances of 'happy', whereas the non-VGG16 model only got **664** right. This suggests that the non-VGG16 model is better at identifying the emotions related to happiness.
- **Neutral:** The VGG16 model correctly classified **580** instances of 'neutral', compared to the non-VGG16 model which got **606** right.

This shows that the non-VGG16 model is more accurate in identifying the features related to neutrality.

- **Sad:** The VGG16 model correctly classified **580** instances of 'sad', whereas the non-VGG16 model only got **554** right. This indicates that the VGG16 model is better at distinguishing between sadness and other emotions.
- **Surprise:** The VGG16 model correctly classified **580** instances of 'surprise', while the non-VGG16 model only got **554** right. This shows that the VGG16 model is better at identifying emotions related to surprise.

3. Class Imbalance

Both confusion matrices show a high number of misclassifications, particularly for 'disgust', 'sad', and 'surprise', which are often confused with other emotions. This is a common issue with these datasets, and it can be improved with more advanced techniques.

With Data Augmentation & no VGG16 model	Data Augmentation with VGG16 model
	

	Without VGG16	With VGG16
Overall Correct Predictions	**4560**	**4649**
Highest Correctly Predicted Emotion	Fear (**1313**)	Fear (**1353**)
Most Confused Emotions	Fear and Neutral	Fear and Neutral
Model Strength	Better at identifying **happy** and **neutral** expressions	Better at identifying **fear**, **sad**, and **surprise** expressions
Model Weakness	Struggles more with **disgust**, **angry**, and **sadness**	Still struggles with **anger**, **disgust**, and **sadness**, but with slight improvements.