Facial Emotion Detection – Capstone Project

This presentation delves into a comprehensive deep learning capstone project that explored the performance of seven distinct models for human emotion classification task from the images. We meticulously analyzed each model's architecture, training strategies, and evaluation metrics, providing valuable observations and insights into them. The aim is to identify the most effective model for the given task and to highlight the key factors influencing their performance.

Atif Saleem

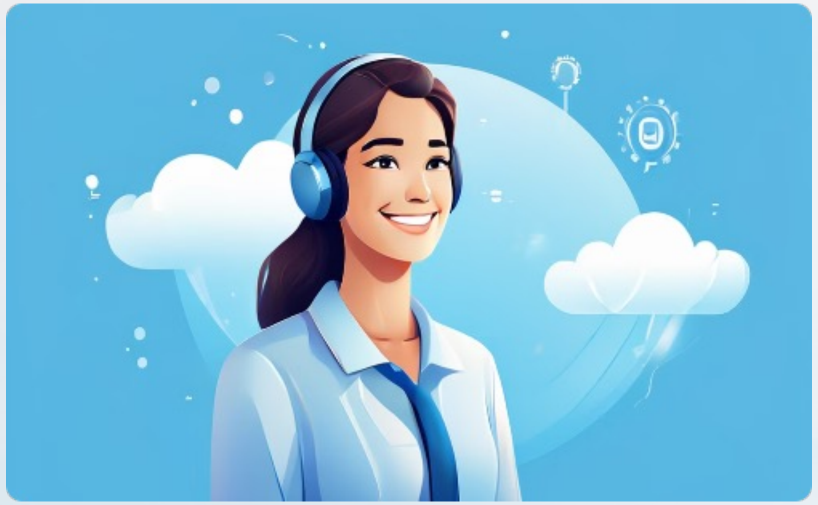
MIT-PE: Applied Data Science Program

March 2024 Cohort

Project Introduction – Facial Emotion Detection

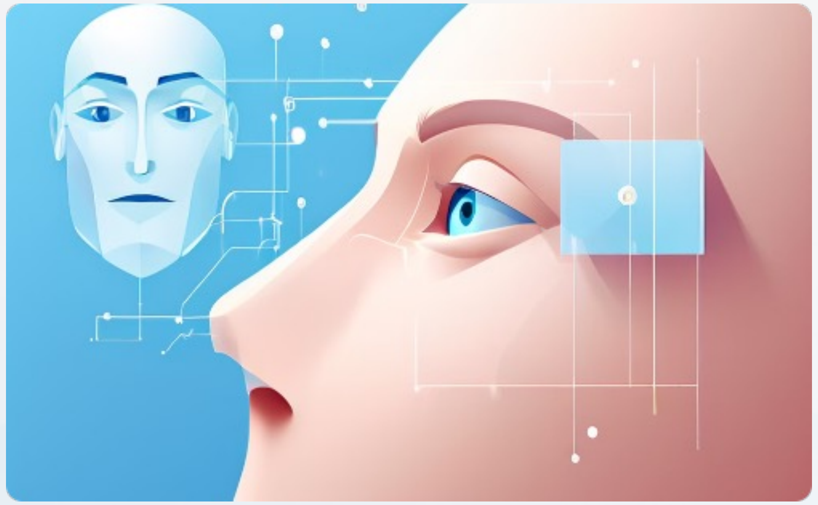
Detecting Emotional States

Facial Emotion Detection (FED) is the ability to automatically detect and classify the different emotional states expressed through a person's facial features. We'll detect and classify grayscale images of 4 classes of emotions in our case.



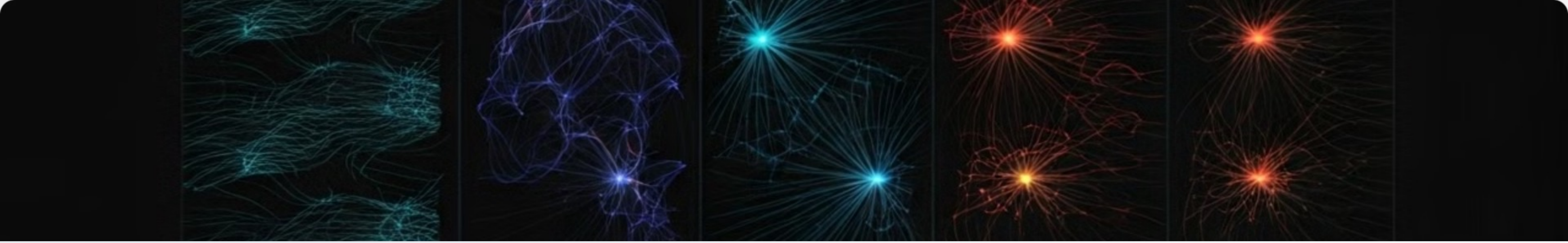
Wide-Ranging Applications

FED holds immense potential for a wide range of applications, from customer service and education to healthcare, market research, security and beyond.
It can revolutionize the human-computer interaction (HCI)



Leveraging Advanced Algorithms

By harnessing the power of advanced algorithms and deep neural networks, FED systems can provide valuable insights into the emotional needs of users in multiple fields of HCI.



Problem Statement & Project Objectives

Problem

Classify grayscale images (48x48 pixels) into four classes of human emotions

Goal

Achieve best accuracy with lower loss on the validation dataset.

Approach

Explore and evaluate, performance of, different deep learning models.

Exploratory Data Analysis

Before training our models, it's important to perform an EDA on the dataset.

Following are some of the observations from our dataset’s EDA:

- The dataset contains about 20,214 grayscale images.
- Images are 48x48 pixels and classified into 4 emotion classes.
- Diversity in facial features, ages, and expressions.
- Variations in lighting conditions and image quality.
- Cropped images and ambiguous expressions are potential challenges to the models training.
- Dataset is balanced for the most part to avoid biases in the data
- Most of the images in validation dataset belong to “Happy” class

Data Pre-processing Techniques

Loading and Sizing

- Images loaded
- Image shape size set to 48x48 pixels to be consistent with the dataset.
- Batch size set to 32 for balanced data processing by models.

Data Augmentation

- Rotation
- Rescaling
- Flipping
- Zooming
- Brightness Adjustments

One-hot Encoding

The four categorical classes of happy, sad, neutral, and surprise were converted into numerical representations, enabling the models to process them effectively.

Data Augmentation Strategies

Rotation

Rotation was applied by random angles to create variations and improve the model's robustness to different orientations.

Rescaling

Rescaling pixel values to the range of 0-1 to improve convergence and generalization. And, also to help models reduce overfitting.

Flipping

Horizontal flipping was set to True that enhanced its ability to generalize to different viewpoints.

Zooming

Zooming in and out on the images allowed the model to learn features at different scales, improving its understanding of object sizes and details.

Brightness

Adding brightness range to the images helped the model to become more tolerant of the lighting variations.

Model Techniques

1

Technique 1: Basic CNN

We experimented with two basic traditional CNN models, known for their effectiveness in image processing, at first. Where second model had slightly more blocks than the first one.

2

Technique 2: Transfer Learning

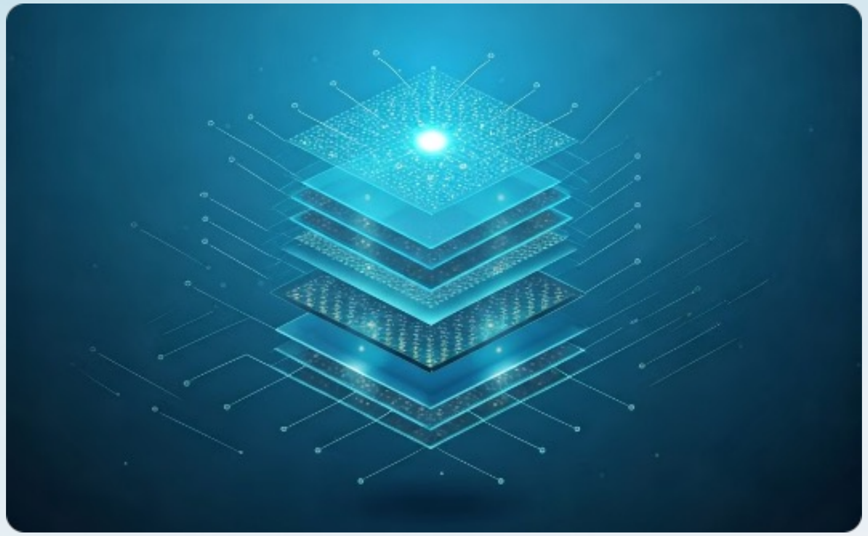
We also evaluated pre-trained models that have been trained on large, general datasets and can be fine-tuned for our specific task.

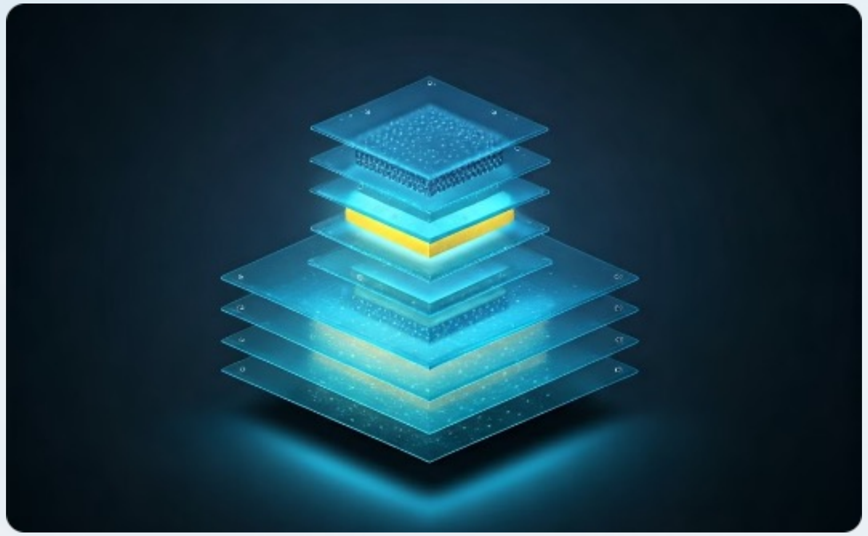
Pre-trained Models: VGG16, ResNet50V2B2, EfficientNetV2

3

Technique 3: Complex CNN

We created more complex custom CNN models. First complex model with 5 convolution blocks along with normalization. Second model with 6 convolution blocks and additional refinements such as L2 regularization and noise addition etc.





Model Techniques – Training Approaches

1

Batch Normalization

It was applied to normalize the inputs for faster convergence, generalization and stability during the training.

2

Down Sampling

Down sampling was leveraged for dimensionality reduction, handling variations in facial expressions, and focusing on key features.

3

Optimization

The adaptive optimization algorithm (Adam) used to adjust the learning rate during training, allowing for more efficient and faster convergence.

4

Regularization

Regularization techniques were applied to prevent overfitting and making models robust to noise and variations in facial features.

Models Comparison

- Models were evaluated mostly based on accuracy, loss, and training overhead
- CNN Models performed well, including basic and complex models
- Pre-trained transfer learning models performed poor overall
- Custom **Complex CNN Model-1** stood out with highest validation accuracy and lowest loss though Basic CNN Model-1 also performed good

Model	Parameters	Validation Loss	Validation Accuracy
cnn_model-1: Baseline (Gray)	684,036	0.62	0.76
cnn_model_2: Baseline (RGB)	2,765,316	0.67	0.75
vgg_model_1: VGG16 (RGB)	14,714,688	1.05	0.54
resnet_model-1: ResNet50V2 (RGB)	25,703,876	0.99	0.58
efficient_model_1: EfficientNetV2B2 (RGB)	10,253,090	1.36	0.37
complex_cnn_model_1: Complex CNN (Gray - Chosen Model)	1,839,876	0.60	0.76
complex_cnn_model_2: Complex CNN (Gray)	3,412,100	0.67	0.74

Final Chosen Model – Complex CNN Model 1

1 Final Chosen Model

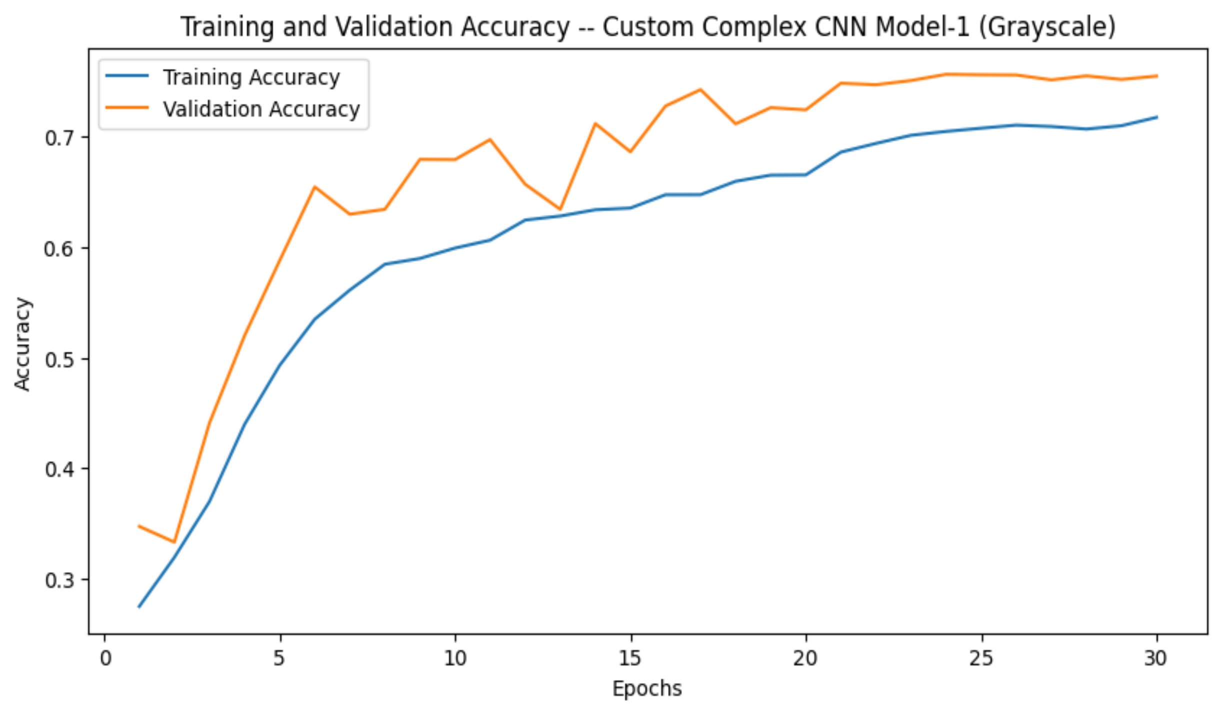
Based on the comparison, I propose using **Complex CNN Model 1** as the final chosen model.

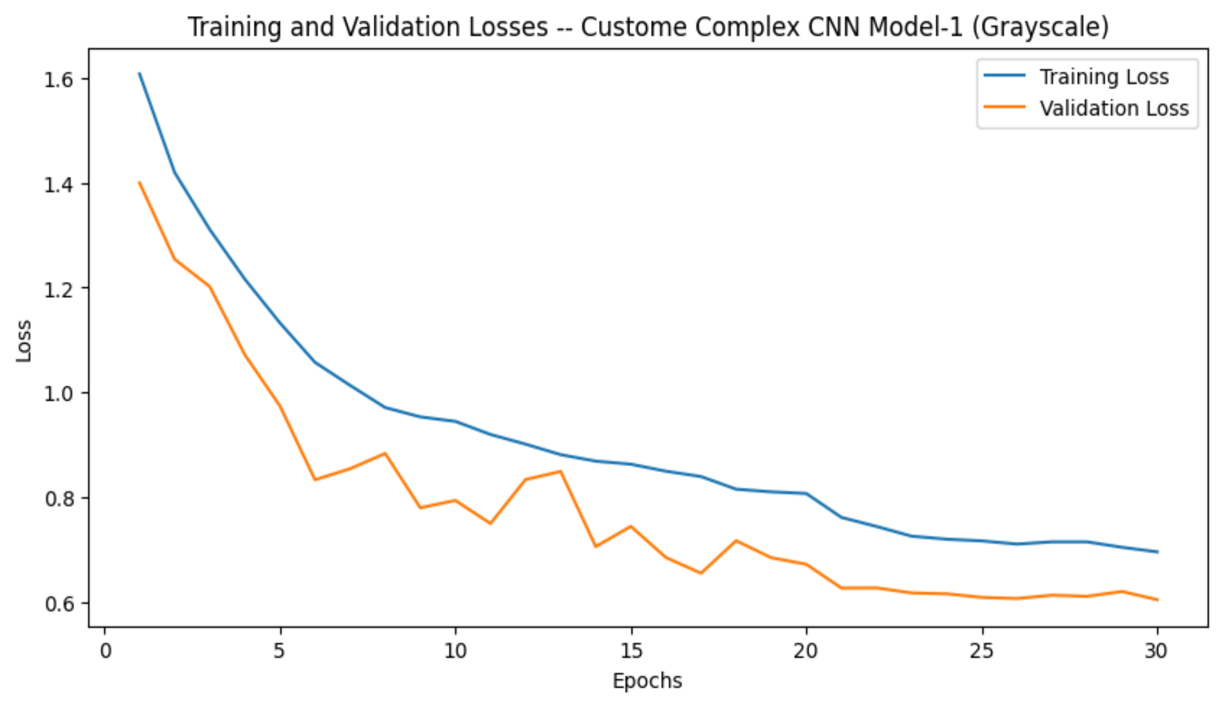
2 Rationale

- This model offers a balance of high accuracy, lower loss, computational efficiency, and flexibility for the real-world applicability.
- Its deeper architecture allows the model to learn more complex features from the input images.

3 Final Model Architecture

- 5 convolution blocks with normalization and regularization.
- Multiple dropout layers with varying values to prevent overfitting.
- Flattening for dense layers
- 2 dense layers precede the output layer.





Implementation Recommendations

1 Deployment Strategies

- Combine cloud and on-premise/edge for a balanced approach of scalability and security.
- Leverage serverless functions to reduce operational overhead.

2 Optimization Strategies

- Quantization (size, speed)
- Fine-tuning (accuracy)
- Explore new architectures.
- Reduce model size for faster loading and smaller storage footprint.

3 Ongoing Evaluation

- Continuous monitoring and fine-tuning as needed.
- Periodic retraining and updating of the model as user demographics and data inputs evolve.

4 Real-world Usability

- More emotions and cleaner data
- Bias Conduct user testing to gather feedback and iterate on the design.
- Reinforcement learning with human feedback for fine-tuning
- Real-world robustness for different deployments
- Consider ethical implications and address potential biases.

Conclusion - Project Summary and Future Directions

1

Key Findings

- We explored the task of grayscale image classification
- Successfully implemented and trained seven different models.
- By analyzing the performance of each model, we were able to identify Complex CNN Model 1 as the best-performing model, achieving high accuracy.

2

Complex CNN Model 1 Success

- Complex CNN Model 1 demonstrated a balance of high accuracy and reasonable computational requirements, making it a promising candidate for real-world deployment.
- The model's depth and use of data augmentation and regularization techniques were key factors in its strong performance.

3

Project Achievements

Over the course of this project, we were able to accomplish the following:

- Successfully implemented and trained seven different deep learning models
- Carefully analyzed the performance of each model and selected the best one
- Visualized the results and gained valuable insights from the data

4

Potential Benefits

- Accurate facial emotion recognition can provide valuable insights that can improve customer experiences, enhance learning outcomes, and support mental health assessments.
- By understanding people's emotional states, organizations can tailor their interactions and services to better meet the needs of their users.

5

Technical Challenges

- Developing reliable facial emotion recognition systems is challenging due to the wide variation in facial features, lighting conditions, and other environmental factors that can affect the appearance of facial expressions.
- Overcoming these challenges is crucial for enabling the technology to reach its full potential.