Executive Summary of DS Boot Camp (Fall 2024) Project

Facial Expression Recognition with Machine Learning and Deep Learning Methods

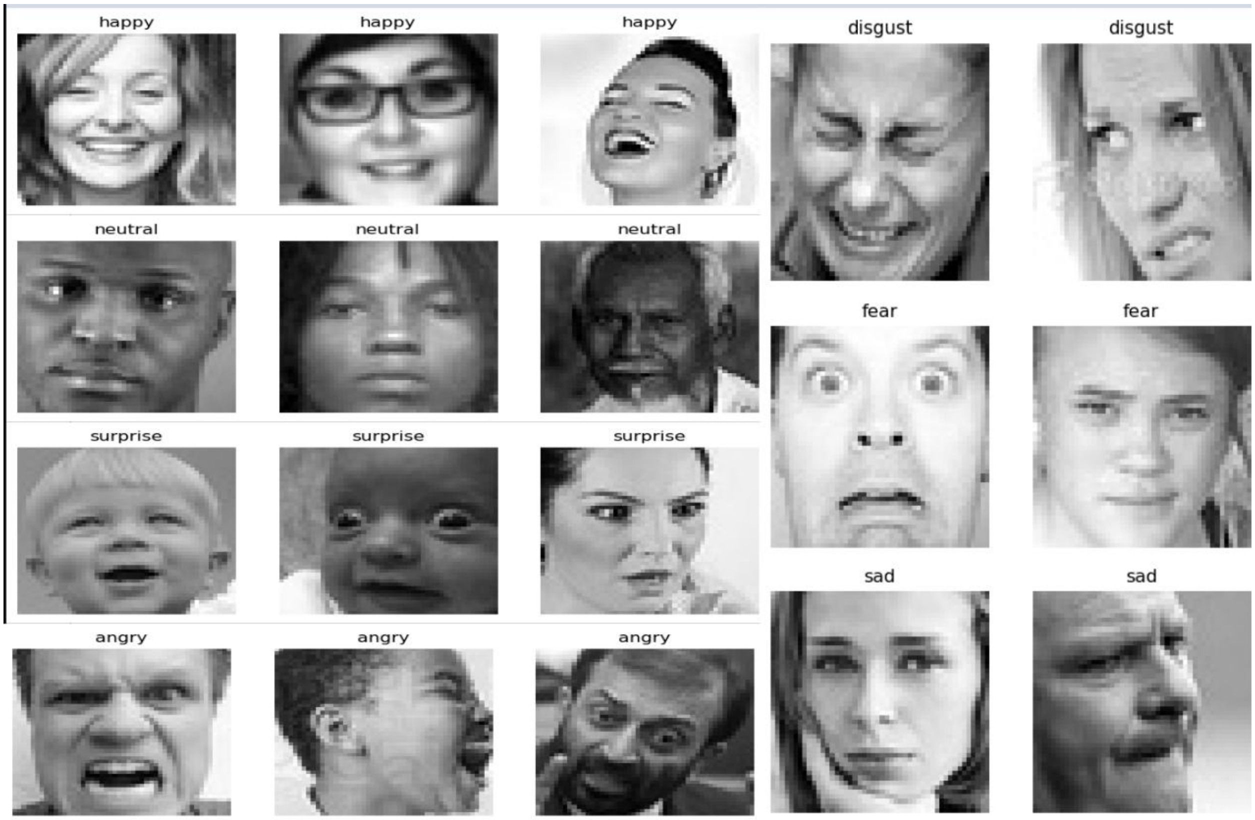
Keywords:
Facial Expression | Data Augmentation | Feature Extraction | Image Filters | Eigenfaces (PCA) | KNN | Regression | SVM | Random Forest | XGBoost | CNN | FNN | Cross Validation

Team members: Yuting Ma, Rui Shi, Menglei Wang, Jiayi Wang
GitHub: https://github.com/Rui-Shi/DS2024_facial_expression

Background and Introduction

Motivation: Facial expressions are a universal form of non-verbal communication. Recognizing these expressions can enhance human-computer interaction, improve accessibility technologies, and enable impactful applications such as mental health monitoring and customer sentiment analysis. Despite advances in the field, creating an accurate and computationally efficient **Facial Emotion Recognition (FER)** system remains challenging due to imbalanced datasets, real-world variability, and resource constraints.

Dataset: The project leverages the **FER-2013 dataset**, consisting of 48x48 pixel grayscale facial images across seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. This dataset, sourced from Kaggle, consists of 28,709 training images and 7178 testing images.



Goal: Develop a comprehensive FER system consisting of traditional machine learning and deep learning approaches, producing well-tuned models for real-world application.

KPIs: The primary performance metrics include **Accuracy, F1-score**, and **Confusion Matrix**

Related Work: The FER-2013 dataset has inspired numerous studies. Notable baselines include feature-based simple classifiers (e.g., logistic regression). Our project aims to expand on this by integrating advanced data preprocessing, augmentation, feature extraction, ensemble techniques, and Neural Networks.

Data Preprocessing and Modelling Approach

1. Data Preprocessing and Augmentation

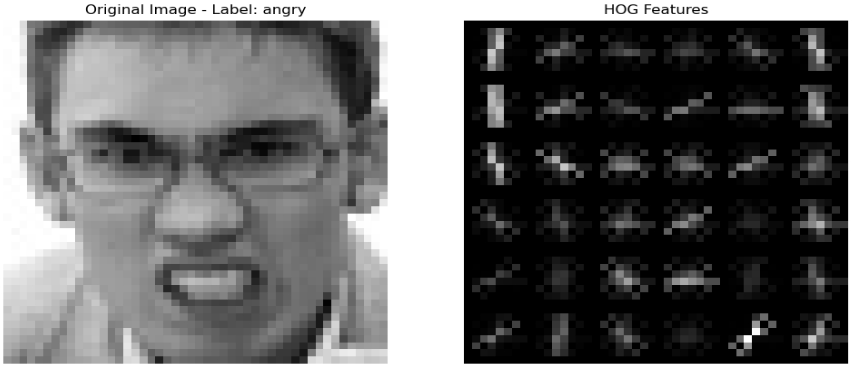
- **Standardization:** Images are resized to **48x48 pixels** and denoised to improve quality.
- **Data Augmentation:** Augment minority classes with transformation techniques like rotation, flipping, and zooming while downsampling some majority classes.

2. Feature Extraction Methods

We explore multiple **feature extraction techniques** and classifiers to analyze and process facial expressions effectively:

1. Histogram of Oriented Gradients (HOG):

- Extract gradient-based features to emphasize edges and textures.
- Parameters:
  - Cell size: 4-16
  - Block size: 2-4
  - Number of orientation bins: 4-10
- Output: Gradient histograms focusing on local intensity changes in the images.



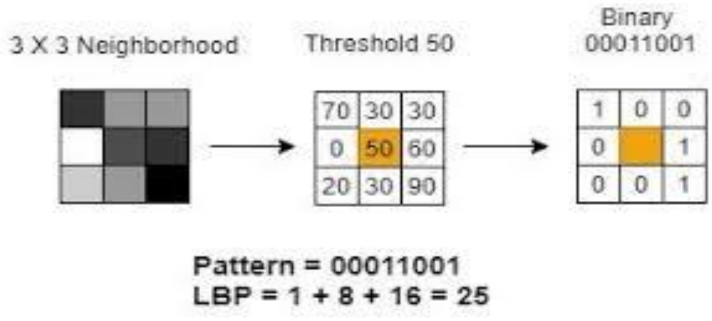
Original Image - Label: angry

HOG Features

2. Local Binary Patterns (LBP):

- Capture texture patterns using uniform LBP configurations.
- Parameters:
  - Radius: 1–3 pixels.

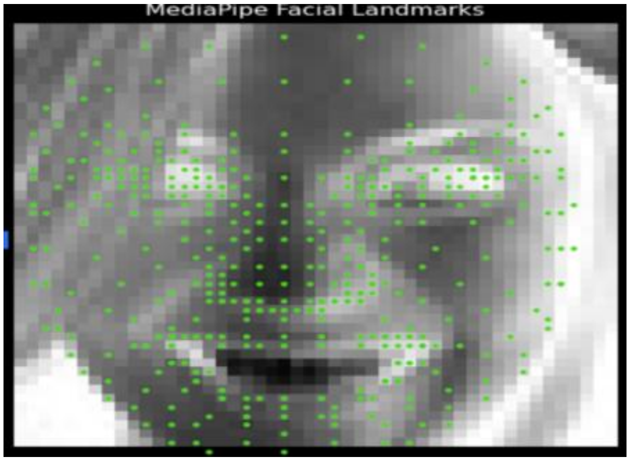
    - Neighbors: 8–24 points.
- Output: Histograms representing frequency distributions of patterns for efficient feature representation. However, the output is often binary for simplicity when using the uniform method.



Pattern = 00011001
LBP = 1 + 8 + 16 = 25

3. MediaPipe Facial Landmarks:

- Extract 478 3D facial landmark points (x, y, z coordinates) per image.
- Output: Landmark-based geometric features representing facial structure.

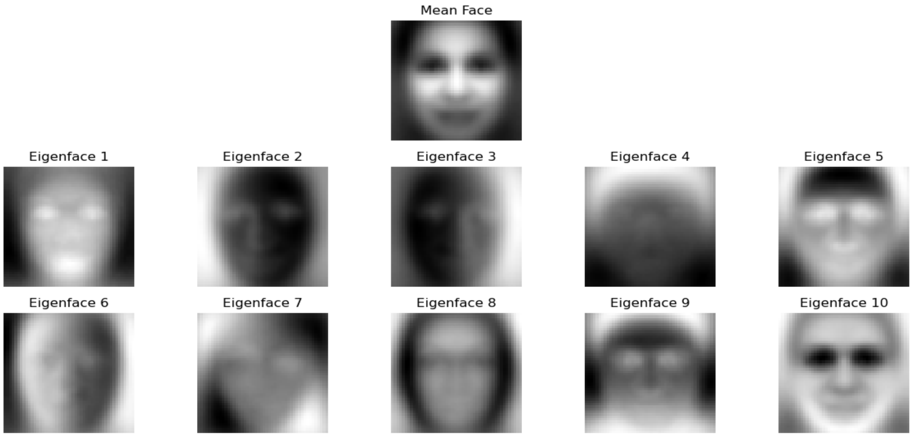


4. **Gabor Filter:**

- Extracts texture features by identifying specific frequencies and orientations.
- Parameters: Wavelength (λ), Orientation (θ), Phase offset (φ), Aspect ratio (γ), Bandwidth (σ)
- Output**:** Response map highlighting edges and textures.

4. Eigenface Analysis:

- A dimensionality reduction technique based on **Principal Component Analysis (PCA)**.
- **Implementation Steps:**
    - **Extract Principal Components**: Decompose normalized face images into orthogonal **eigenvectors (**i.e**., eigenfaces)** representing maximum variance directions.
    - **Variance Retention**: Retain eigenvectors explaining **90% of the variance** to reduce data dimensionality while preserving essential features.
    - **Projection onto Eigenvectors**: Map input data onto Eigenspace for a compact representation of facial features.


Mean Face
Eigenface 1
Eigenface 2
Eigenface 3
Eigenface 4
Eigenface 5
Eigenface 6
Eigenface 7
Eigenface 8
Eigenface 9
Eigenface 10

- Output: Compressed, variance-maximized feature vectors for classification.

3. Feature-Based Approaches

Feature-based approaches employ traditional classifiers such as **KNN**, **SVM**, **Random Forest**, and **Logistic Regression**, with ensemble methods (e.g., **XGBoost**) for enhanced performance.

4. Deep Learning approaches

- **Custom CNN:** Design and train **Convolutional Neural Networks (CNN)** tailored to FER-2013.
- **Training Techniques:**
    - Loss: Categorical cross-entropy.
    - Optimizer: Adam with dynamic learning rate scheduling.
    - Data Augmentation: Transformations applied dynamically during training.
    - Regularization: Early stopping and dropout to reduce overfitting.

5. Evaluation Metrics for the testing images

- **Quantitative Metrics**: Accuracy, precision, recall, F1-score, ROC-AUC.
- **Efficiency Metrics**: Inference time, model size, and resource usage.
- **Confusion Matrix Analysis**: Identify and address misclassifications across different classes.

Results and Preliminary Insights

Feature-Based Models: Performance Highlights

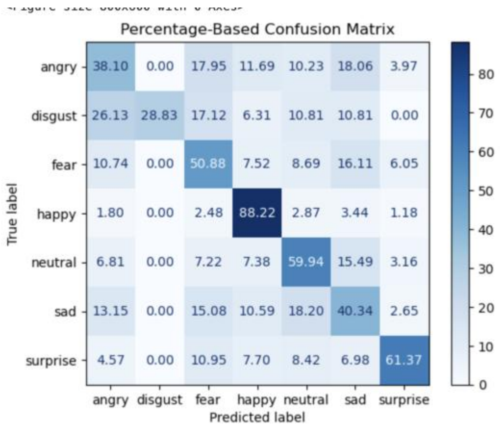
Overall Accuracy and weighted F1 scores

- **PCA + Logistic Regression (baseline)**: Obtained an **accuracy of 37%** and a **F1-score of 0.35**.
- **PCA + KNN:** When k=1, it achieved the best performance with **40% accuracy** and **0.40 F1-score**.
- **PCA + Randomforest**: Got **45% accuracy** and **0.41 F1-score** after intensive grid search and cross-validation.
- **PCA + XGBoost**: **48% accuracy** and **0.46 F1-score** after intensive grid search and cross-validation.
- **LBP + KNN**: **29.38% accuracy** and **0.3 F1-score**.
- **LBP + SVM**: **33.73% accuracy** and **0.3 F1-score.**
- **LBP + Randomforest**: **37.21% accuracy** and **0.34 F1-score**.
- **LBP + XGBoost**: **35.68% accuracy** and **0.33 F1-score**.
- **MediaPipe + PCA + KNN**: **34% accuracy** and **0.34 F1-score**
- **MediaPipe + PCA + Random Forest**: **52% accuracy** and **0.50 F1-score**
- **MediaPipe & PCA+ Gabor Filter & PCA+Randomforest: 49% accuracy** and **0.48 F1-score**

- **HOG & PCA + Gabor Filter & PCA+ XGBoost**: Got **54% accuracy** and **0.53 F1-score** after intensive grid search and cross-validation. Highlighting the potential of combining gradient-based (HOG) and texture-rich (Gabor) features with ensemble learning techniques.
- **Gabor Filter & PCA + MediaPipe & PCA + XGBoost**: Delivered the best **F1-score of 0.59** and **60% accuracy,** demonstrating the effectiveness of integrating geometric facial landmarks (MediaPipe) with Gabor-based texture features.

Selected Confusion Matrices of Some Classifiers:

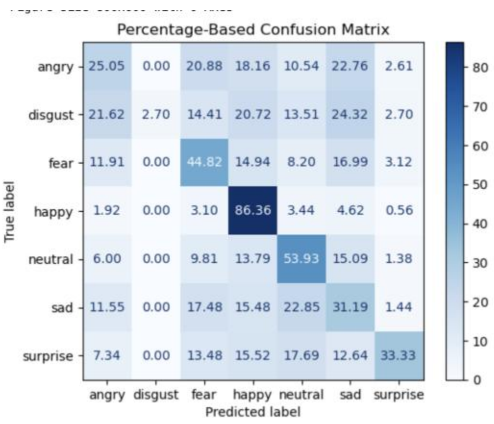
MediaPipe & PCA + Gabor Filter & PCA + RandomForest



Percentage-Based Confusion Matrix

True label \ Predicted label	angry	disgust	fear	happy	neutral	sad	surprise
angry	38.10	0.00	17.95	11.69	10.23	18.06	3.97
disgust	26.13	28.83	17.12	6.31	10.81	10.81	0.00
fear	10.74	0.00	50.88	7.52	8.69	16.11	6.05
happy	1.80	0.00	2.48	88.22	2.87	3.44	1.18
neutral	6.81	0.00	7.22	7.38	59.94	15.49	3.16
sad	13.15	0.00	15.08	10.59	18.20	40.34	2.65
surprise	4.57	0.00	10.95	7.70	8.42	6.98	61.37

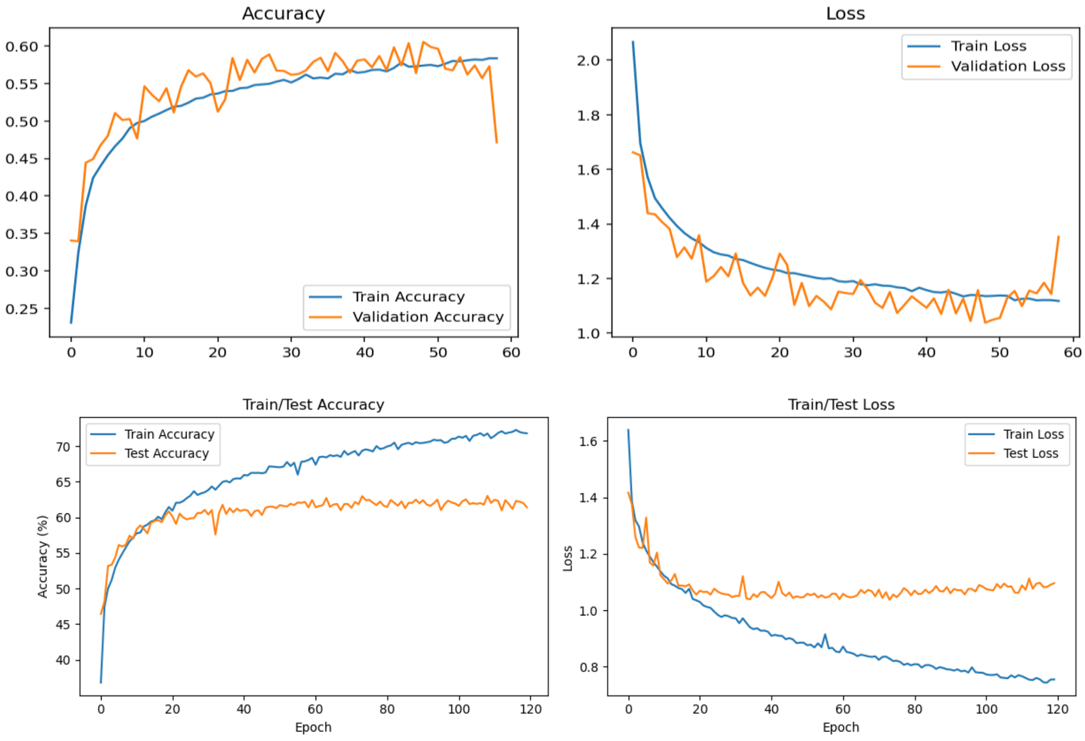
MediaPipe & PCA + Gabor Filter & PCA + XGBoos



Percentage-Based Confusion Matrix

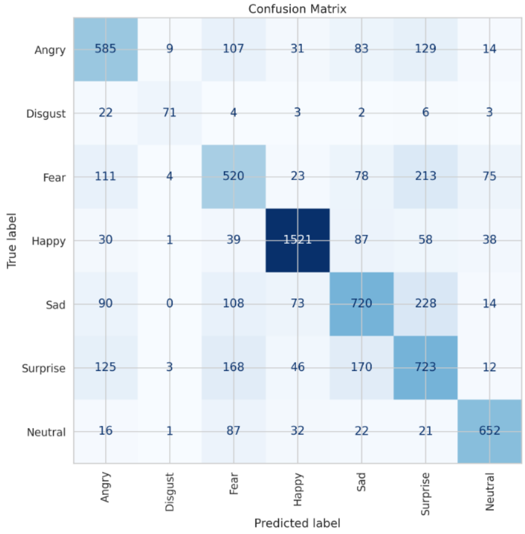
True label \ Predicted label	angry	disgust	fear	happy	neutral	sad	surprise
angry	25.05	0.00	20.88	18.16	10.54	22.76	2.61
disgust	21.62	2.70	14.41	20.72	13.51	24.32	2.70
fear	11.91	0.00	44.82	14.94	8.20	16.99	3.12
happy	1.92	0.00	3.10	86.36	3.44	4.62	0.56
neutral	6.00	0.00	9.81	13.79	53.93	15.09	1.38
sad	11.55	0.00	17.48	15.48	22.85	31.19	1.44
surprise	7.34	0.00	13.48	15.52	17.69	12.64	33.33

- **CNN:** Delivers the test accuracy of 63.39% accuracy and 1.038 test lost. We set the kernel of (3,3), number of filters (32,64,128), learning rate of 0.0005, and 60-120 epochs.



- **TransferLearning Approach(ResNet50):** The learning approach with ResNet50 leverages this 50-layer deep convolutional neural network, pre-trained on ImageNet, for facial emotion recognition with the FER-2013 dataset. Replacing the connected layers designed for ImageNet classification with a task-specific layer outputting the number of emotion classes, the model is fine-tuned to recognize emotions.

- This approach achieves **over 65% accuracy**, combining the robustness of pre-trained features with the adaptability of fine-tuning for emotion classification.



Deep Learning Models

- **Performance**: Our **ResNet50** experiments achieved **over 65% accuracy**, beating Kaggle's record using neural networks on the FER2013 dataset (Kaggle reference). This highlights the potential of deep learning for FER tasks, particularly when leveraging advanced augmentation and optimization techniques.

Efficiency Trade-offs

- **Feature-Based Methods**: Offered higher computational efficiency but were constrained by lower accuracy.
- **Deep Learning Models**: Provided superior accuracy at the expense of greater computational resources, making hardware optimization a key consideration for real-world deployment.

Future Works

1. **Model Generalization:** Train and validate models on additional datasets (e.g., CK+, JAFFE) to improve real-world performance.
2. **Edge Device Deployment:** Optimize CNN architectures for mobile and edge deployment using model compression techniques like pruning and quantization.

3. **Emotion Grouping:** Test broader emotion categories (e.g., positive vs. negative) to simplify classification and improve accuracy.
4. **Multi-Modal Integration:** Combine facial features with vocal or physiological data for a comprehensive emotion recognition system.

Concluding Remarks

The proposed FER system will provide a robust framework for recognizing emotions with a balance of accuracy and efficiency. We aim to uncover valuable insights for deploying FER solutions in real-world scenarios by comparing the traditional and deep learning approaches. The project will serve as a benchmark for future developments in facial emotion recognition.