# EMOTION DETECTION FROM FACIAL IMAGES USING CNN AND FACIAL LANDMARKS

## 1. Introduction:

This project presents an automatic emotion detection system that classifies facial images into seven emotions. It combines convolutional neural networks (CNN) for visual features and facial landmarks for geometry, aiming to achieve better accuracy even under challenging real-world conditions such as varied head poses and lighting.

## 2. System Design and Implementation:

### 2.1 Data Preparation

- Used structured image folders for seven emotion classes.
- Detected faces with MTCNN and extracted 468 facial landmarks per face using Media Pipe Face Mesh.
- Computed global mean and standard deviation of landmarks for normalization.

### 2.2 Data Augmentation & Balancing

- Applied random resizing, flipping, rotation, color jitter, affine transforms, and random erasing.
- Used WeightedRandomSampler to balance classes during training.

### 2.3 Model Architecture

- **CNN feature extractor:** Three convolutional blocks with dropout and batch normalization.
- **Landmark MLP:** Processes flattened landmark vectors.
- **Final classifier:** Merges CNN and landmark features for emotion prediction.

### 2.4 Training Strategy

- Cross-entropy loss with label smoothing.
- Adam optimizer + cosine annealing scheduler.

- Test-Time Augmentation (TTA) by averaging predictions of original and horizontally flipped images.
- Early stopping to prevent overfitting.

## 3. Performance Analysis:

| Metric | Train Set | Test Set |
|---|---|---|
| Accuracy | ~63% | ~54% |
| Precision | ~62% | ~51% |
| Recall | ~63% | ~53% |
| F1 Score | ~62% | ~51% |

**Observation:**
The test metrics are lower mainly due to head pose variations, uneven lighting, and different expression intensities in real-world scenarios, making generalization harder.

## 4. Potential Applications

- Real-time mood analysis in smart applications
- Human-computer interaction and assistive systems
- Educational tools to monitor student engagement
- Healthcare systems to help track patient emotional well-being
- Content recommendation engines that respond to user emotions

## 5. Ethical Considerations

Emotion detection poses critical ethical questions:

- **User Privacy:** Facial data is sensitive; strict security and anonymization are essential.
- **Bias Mitigation:** Models may underperform for underrepresented groups if datasets lack diversity.

- **Responsible Use:** Must avoid misuse, such as hidden surveillance or making decisions without human oversight.

**To address these:**

- Obtain clear consent for data use.
- Regularly evaluate fairness and update datasets.
- Design transparent systems explaining predictions to users.

## 6. Tools and Technologies

- **Deep Learning Framework:** PyTorch
- **Computer Vision Libraries:** facenet-pytorch (MTCNN), MediaPipe (FaceMesh)
- **Data Handling:** NumPy, Pandas
- **Image Augmentation:** torchvision.transforms
- **Evaluation:** scikit-learn metrics (accuracy, precision, recall, F1)
- **Visualization & Progress:** tqdm
- **Deployment (optional):** Streamlit app for real-time predictions

## 7. Limitations

While the model shows promising results, several limitations were observed:

- **Head Pose & Lighting Variations:** Performance drops significantly for extreme angles, partial occlusions, or poor lighting, which are common in real-world data.
- **Dataset Size & Diversity:** The dataset might not represent all age groups, skin tones, or cultural variations of expressions, which can introduce bias.
- **Overfitting Risk:** Despite augmentation and balancing, the model performs notably better on the training set than the test set.
- **Landmark Extraction Dependency:** Accurate landmark detection depends on high-quality face detection; if the face is not detected or landmarks are missing, prediction may be less reliable.
- **Limited Emotion Classes:** The system only recognizes seven fixed emotion categories and cannot handle subtle or mixed emotions.

These limitations highlight areas for future improvement, such as collecting a larger, more diverse dataset, adding temporal data (e.g., video-based emotion dynamics), or integrating attention mechanisms.

## 9. Future Improvements:

While the current system achieves reasonable performance, several enhancements could significantly improve its robustness, fairness, and applicability:

- **Larger & More Diverse Dataset:**
  Expanding the dataset to include a wider range of demographics such as different age groups, ethnicities, and cultural backgrounds. This would help reduce model bias and improve generalization to real-world data.
- **Video-Based Emotion Analysis:**
  Transitioning from static image classification to temporal analysis using sequential models (e.g., RNNs or transformers). This could help capture dynamic changes in facial expressions for more accurate emotion detection.
- **Attention Mechanisms:**
  Integrating attention layers to enable the model to automatically focus on the most informative facial regions or landmark features, improving prediction quality.
- **Multi-modal Learning:**
  Combining facial features with additional data sources like speech tone, text sentiment, or physiological signals to provide richer, multi-dimensional emotion recognition.
- **Detecting Subtle and Mixed Emotions:**
  Extending the system beyond seven fixed emotion categories to recognize subtle, compound, or context-dependent emotional states.
- **Real-Time Optimization:**
  Optimizing the model architecture and inference pipeline for faster predictions, enabling deployment on mobile and edge devices for real-time applications.
- **Explainable AI (XAI):**
  Incorporating interpretability methods to explain why certain emotions were predicted, enhancing user trust and transparency.

These improvements would help evolve the system into a more **accurate, fair, and widely usable emotion detection solution** suited for complex, real-world scenarios.

## 10. Conclusion

This project showcases a robust end-to-end emotion detection system combining CNN and landmark features, along with strong augmentation and balancing strategies. While it achieves reasonable accuracy, performance is affected by real-world data variability. Ethical design and responsible use remain central to deploying such technologies effectively.