**1. Introduction**  
The objective of this project is to build an emotion recognition system using deep learning, specifically through transfer learning with a modified ResNet50 model. The system classifies human emotions into seven categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The primary dataset used is FER 2013, a widely recognized benchmark for facial emotion recognition, consisting of grayscale facial images. A pre-trained ResNet50 model is adapted to accept single-channel grayscale images, and its final classifier is replaced to suit the 7-class emotion classification task.

**2. Proposed System and Flow Diagram**

**Overview of the Proposed System**  
The system is composed of three key stages:

* **Model Definition and Modification:**  
  A pre-trained ResNet50 is adapted for grayscale input by modifying the first convolutional layer. The default classifier is replaced with a custom classifier tailored for the seven emotion classes.
* **Training:**  
  The model is trained on the FER 2013 dataset using standard deep learning techniques. Preprocessing includes grayscale confirmation, resizing, tensor conversion, and normalization. Data augmentation is also used to improve model generalization. Training uses the Adam optimizer with cross-entropy loss.
* **Real-Time Inference:**  
  The system captures real-time video via webcam. Faces are detected using Haar cascades (OpenCV), cropped, preprocessed, and then passed to the trained model. The predicted emotion is displayed on the video frame along with a bounding box.

**3. Explanation**

**Data Preprocessing and Augmentation**

* **Dataset (FER 2013):**  
  The FER 2013 dataset is already in grayscale, aligning with the model's modified input layer.  
  Images are resized to 48×48 pixels and normalized to improve training stability and convergence.  
  Data is converted to tensors for compatibility with the deep learning framework.
* **Augmentation Techniques:**  
  Common techniques such as random horizontal flipping and rotation are applied to prevent overfitting and enhance model generalization.

**Model Modification: ResNet50 Adaptation**

* **Pre-trained Backbone:**  
  ResNet50, originally trained on ImageNet, provides robust feature extraction capabilities.
* **Input Layer Adaptation:**  
  The first convolutional layer is modified to accept single-channel (grayscale) input instead of the original three-channel (RGB) format.
* **Custom Classifier:**  
  The final fully connected layer is replaced with a new classifier. This consists of a dense layer, activation function, dropout (for regularization), and an output layer with 7 neurons—one for each emotion class.

**Training Pipeline**

* **Setup:**  
  The training setup includes setting hyperparameters such as batch size, number of epochs, and learning rate. A GPU is used if available.
* **Optimizer and Loss Function:**  
  Adam optimizer is used for its efficiency in handling sparse gradients.  
  Cross-entropy loss is employed as it is standard for multi-class classification.
* **Training Loop:**  
  During training:
  + The model learns from batches of data.
  + Loss is computed and backpropagated.
  + The model is updated accordingly.
  + After each epoch, a validation pass checks model performance on unseen data.

**Real-Time Emotion Recognition**

* **Face Detection:**  
  The system uses OpenCV’s Haar Cascade to detect faces in the video stream.
* **Preprocessing Before Inference:**  
  Detected faces are cropped, resized, and normalized using the same steps as training to maintain consistency.
* **Inference:**  
  Preprocessed face images are passed through the trained model.  
  The model outputs emotion probabilities, and the label with the highest probability is displayed on the video frame.

**4. Summary**

This project presents a robust, real-time emotion recognition system that:

* Adapts a pre-trained ResNet50 for grayscale image input.
* Uses the FER 2013 dataset for training.
* Employs data preprocessing and augmentation for better model performance.
* Integrates a webcam-based real-time inference pipeline using OpenCV.
* Predicts and overlays detected emotions on live video.