
SYNTHEIA
SENSING WHAT WORDS CAN'T SAY

By: Gowtham Kumar Kamuni (A20549435)

Instructor: Dr. Jafar Saniie

ECE 501- Artificial Intelligence and Edge Computing

ABSTRACT:

In a world where emotions are often hidden behind silence or forced smiles, we wanted to build something that could see through the noise, something that truly understands people. This idea gave rise to **SYNTHEIA**: an AI powered emotion recognition system that detects and interprets human emotions through facial expressions in real time. Built on the NVIDIA Jetson Orin Developer Kit, it uses lightweight computer vision models to classify emotional states such as happiness, sadness, anger, fear, and neutrality, all while respecting user privacy by avoiding the storage of personal data.

While pre-trained models will help us build a reliable foundation, we also plan to experiment with a small, custom-collected dataset to enhance SYNTHEIA's performance and personalize its predictions. The current focus is purely on facial expressions, but the system is designed with future expansion in mind, including voice patterns or biometric signals. SYNTHEIA isn't about judgment or lie detection. It's about understanding, a new way to feel seen when words fall short.

INTRODUCTION:

Emotions often exist between words, subtle expressions, brief hesitations, and silent cues that reveal far more than spoken language. In everyday life, especially in fields like education, mental health, research, and interviews, being able to read these emotional cues can make communication more meaningful. However, most machines and digital systems fail to recognize this layer of human interaction.

SYNTHEIA is our attempt to bridge that gap. It is an AI-powered, embedded system designed to detect and interpret facial expressions in real time to classify emotional states such as happiness, sadness, anger, fear, and neutrality. By using computer vision models on the NVIDIA Jetson Orin Developer Kit, we aim to create a portable and privacy conscious emotion recognition tool that operates fully on-device, without cloud dependency or data storage.

Problem Statement

Despite the progress in AI, emotion recognition systems are often cloud-based, bulky, or require multiple input modalities like voice and biometrics, making them impractical for real-time use in sensitive or mobile settings. There is a clear need for a lightweight, standalone emotion recognition system that can be deployed on edge devices to interpret human emotion using only facial expressions.

Scope

This project will focus solely on facial expression recognition using image data. While we recognize the potential of incorporating voice patterns or biometric signals, such additions are considered out of scope for the current timeline. We will use a hybrid model approach, starting with pre-trained models and optionally fine-tuning them with a small, self-collected dataset. The goal is to deliver a real-time, efficient, and user-friendly prototype capable of interpreting visible emotions with reasonable accuracy and speed.

SYNTHEIA is not a lie detector, nor is it meant to judge. It is designed to enhance emotional awareness, to give users a deeper, more empathetic layer of understanding when words aren't enough.

SYSTEM VIEW AND DESIGN ARCHITECTURE:

A. System Flow Overview:

The operational pipeline of SYNTHEDIA consists of the following stages:

- **Video Input:**
A USB or CSI camera captures live video frames of a human subject.
- **Preprocessing:**
The frames are resized, normalized, and face-cropped using OpenCV’s Haar Cascades or Dlib-based face detectors.
- **Emotion Recognition Model:**
A CNN-based model (e.g., MobileNet, Mini-Xception, or fine-tuned FERNet) processes the face image to classify the emotion into one of the defined categories: happy, sad, angry, surprised, fearful, neutral.
- **Output Display:**
The detected emotion and its confidence score are displayed on the screen using a simple graphical interface — either through an LCD panel or a Flask-based web UI.
- **Local Logging (Optional):**
Detected emotions may be logged with timestamps for offline analysis if required (with user consent).

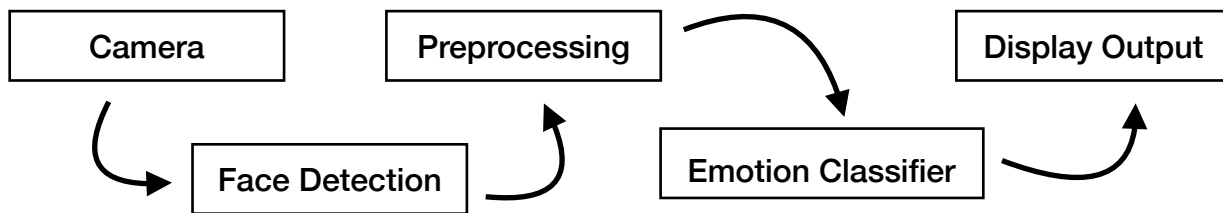
B. Hardware Components:

Component	Purpose
Jetson Orin Dev Kit	Edge AI platform to run the entire pipeline
Camera Module	Capture real-time facial data
Optional Display	On-device display of emotion output
Power Supply	Portable or wall-powered, depending on use case

C. Software Stack:

Layer	Tools/Frameworks
OS & Drivers	Ubuntu for Jetson, NVIDIA Jetpack SDK
Preprocessing	OpenCV, NumPy, face-detection algorithms
Model Inference	TensorFlow Lite / PyTorch
GUI/Web Interface	Python + Tkinter / Flask / OpenCV window

D. Real-Time Processing Loop



E. Flexibility for Future Extensions:

The modular design allows for future additions such as:

- Voice input stream via microphone
- Audio signal processing (MFCC features)
- Biometric sensor integration (heart rate, GSR)
- Multimodal emotion fusion using ensemble models

These are not part of the current scope but have been accounted for in the design for potential expansion.

DATASET AND MODEL TRAINING:

We are exploring the option to collect a small, controlled dataset of our own, using images of group members or volunteers exhibiting specific facial expressions. These images would be used:

- For fine-tuning a pretrained model
- For performance evaluation in real-world settings

SYNTHEIA's model pipeline will follow one of the following strategies:

1. Transfer Learning on Pretrained Models
 - Base model: Mini-Xception, MobileNet, or Custom CNN
 - Pretrained on FER-2013
 - Fine-tuned on a small portion of collected images (if available)
2. Training a Lightweight CNN from Scratch (if performance permits)
 - Designed for real-time inference on Jetson Orin
 - Includes dropout and batch normalization for generalization

All models will be optimized for low latency, compact size, and accuracy, suitable for real-time processing on edge devices.

Training Tools and Frameworks

- TensorFlow / Keras or PyTorch
- OpenCV for image preprocessing
- Matplotlib / TensorBoard for training visualization
- ONNX or TensorRT for model optimization on Jetson Orin

EXPECTED OUTPUT AND EVALUATION PLAN:

Expected Outputs:

The core deliverable of this project is a working prototype of SYNTHEIA that can:

- Detect a human face from live video input (via connected camera)
- Classify facial expression into one of the selected emotion categories:
 - Happy, Sad, Angry, Fearful, Surprised, Neutral
- Display the predicted emotion and confidence score on-screen in real time
- Operate fully on-device using the Jetson Orin Developer Kit without cloud dependence
- (Optional) Save logs of emotion predictions with timestamps for analysis

The system is expected to operate at a frame rate of 10–20 FPS with lightweight models, depending on the resolution and complexity of the classifier used.

Evaluation Plan:

To measure the effectiveness of SYNTHEIA, we will evaluate the model and system using the following metrics:

1. Model-Level Evaluation

- Accuracy on validation data (FER-2013 or custom test split)
- Confusion matrix to analyze performance across emotion classes
- Precision, Recall, F1-Score (optional, for selected test samples)

2. System-Level Evaluation

- Real-time performance (measured in FPS on Jetson Orin)
- Inference latency (ms per frame)
- Qualitative Testing with sample images and live video to visually confirm:
 - Correct face detection
 - Appropriate emotion label output
 - Stability and responsiveness

Optional Human Evaluation

If time permits, we may test the prototype with real users and collect subjective feedback on:

- Whether the emotion classification “feels accurate”
- Perceived smoothness of interaction
- Suggestions for improvement

Success Criteria

The system will be considered successful if it meets the following:

- Consistent, real-time emotion classification on-device
- 60–70%+ validation accuracy across emotion classes
- Smooth visual output (no major lag or crashing)
- Reliable detection in typical lighting and camera angles

FUTURE SCOPE AND IMPROVEMENTS:

While the current version of SYNTHETIA focuses on facial expression-based emotion recognition, there are several directions for expanding its capabilities. These improvements can enhance accuracy, usability, and versatility across various real-world applications.

Planned future enhancements include:

- Integration of voice-based emotion detection using speech features
- Addition of biometric sensors (e.g., heart rate, GSR) for multimodal emotion analysis
- Multi-person detection for group scenarios such as classrooms or meetings
- Adaptive learning to personalize emotion prediction over time
- Development of context-specific modes (e.g., Interview Mode, Therapy Mode)
- Strengthening of privacy and ethical safeguards

TEAM CONTRIBUTION:

This project is being developed individually by:

Gowtham Kumar Kamuni (A20549435)

- Proposed and structured the overall project idea
- Designed the system architecture and selected appropriate hardware (Jetson Orin)
- Researched datasets and model training strategies
- Will implement facial expression recognition pipeline, integrate all components, and perform real-time testing
- Responsible for writing the full report and preparing the presentation
- Will handle all deployment, documentation, and submission tasks

As a solo project, all aspects, from design to execution, will be completed independently, while ensuring quality, functionality, and completeness.