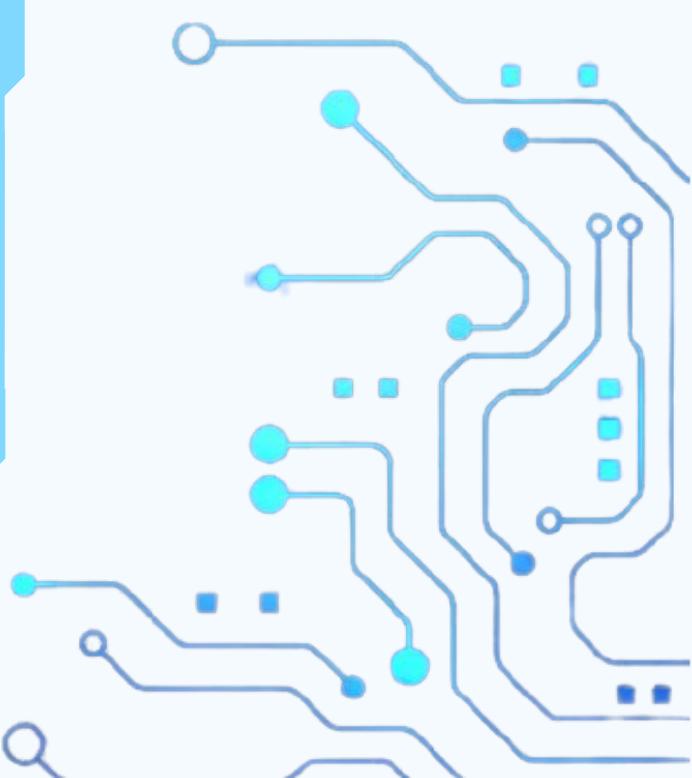
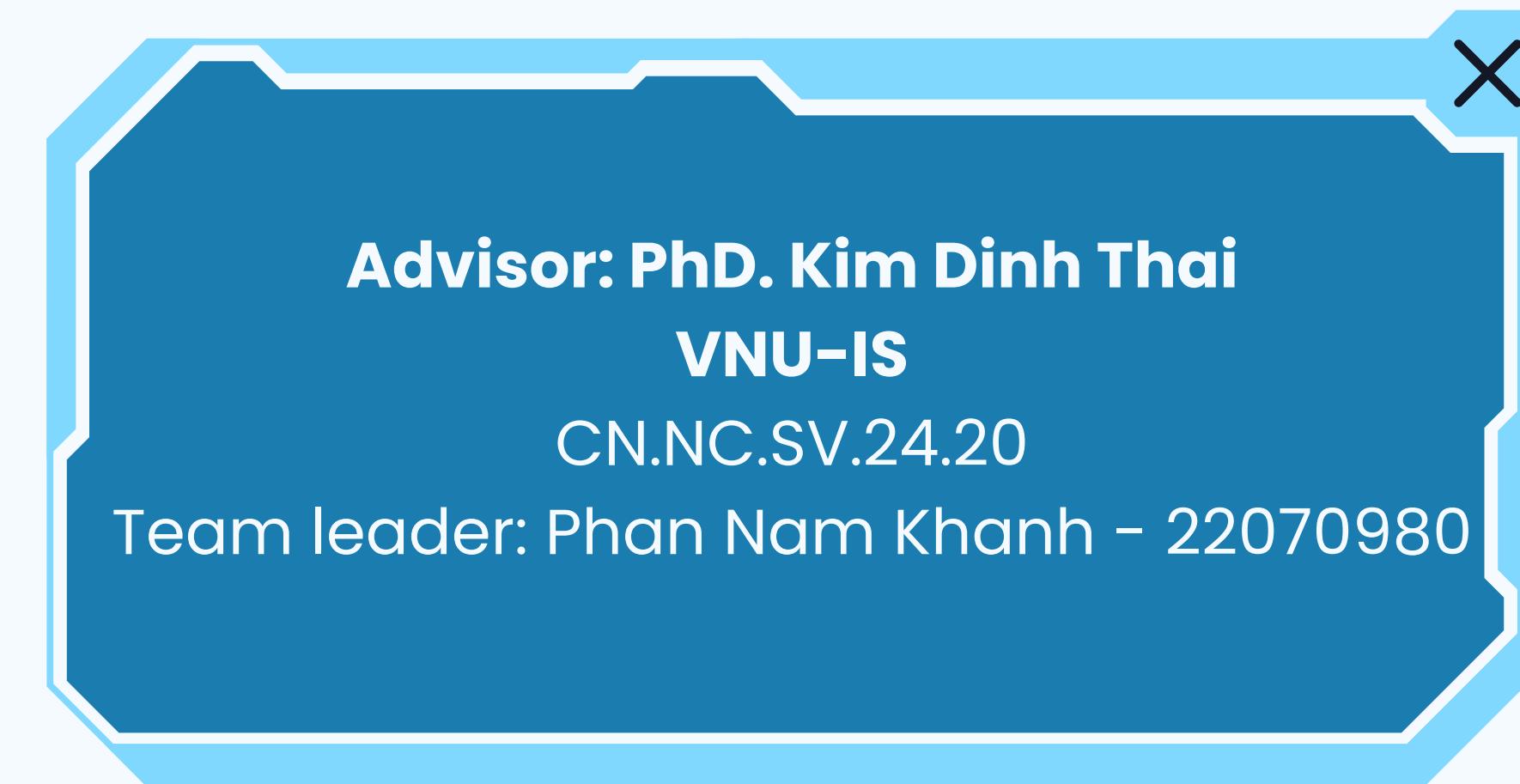


CROSS-PLATFORM EMOTION RECOGNITION FOR USERS USING YOLOV12 OBJECT DETECTION



INTRODUCTION

This study introduces an emotion recognition system using YOLOv12 and the Gemini 1.5 Flash Pro API, combining facial analysis with an AI chatbot to enable natural emotional interaction.

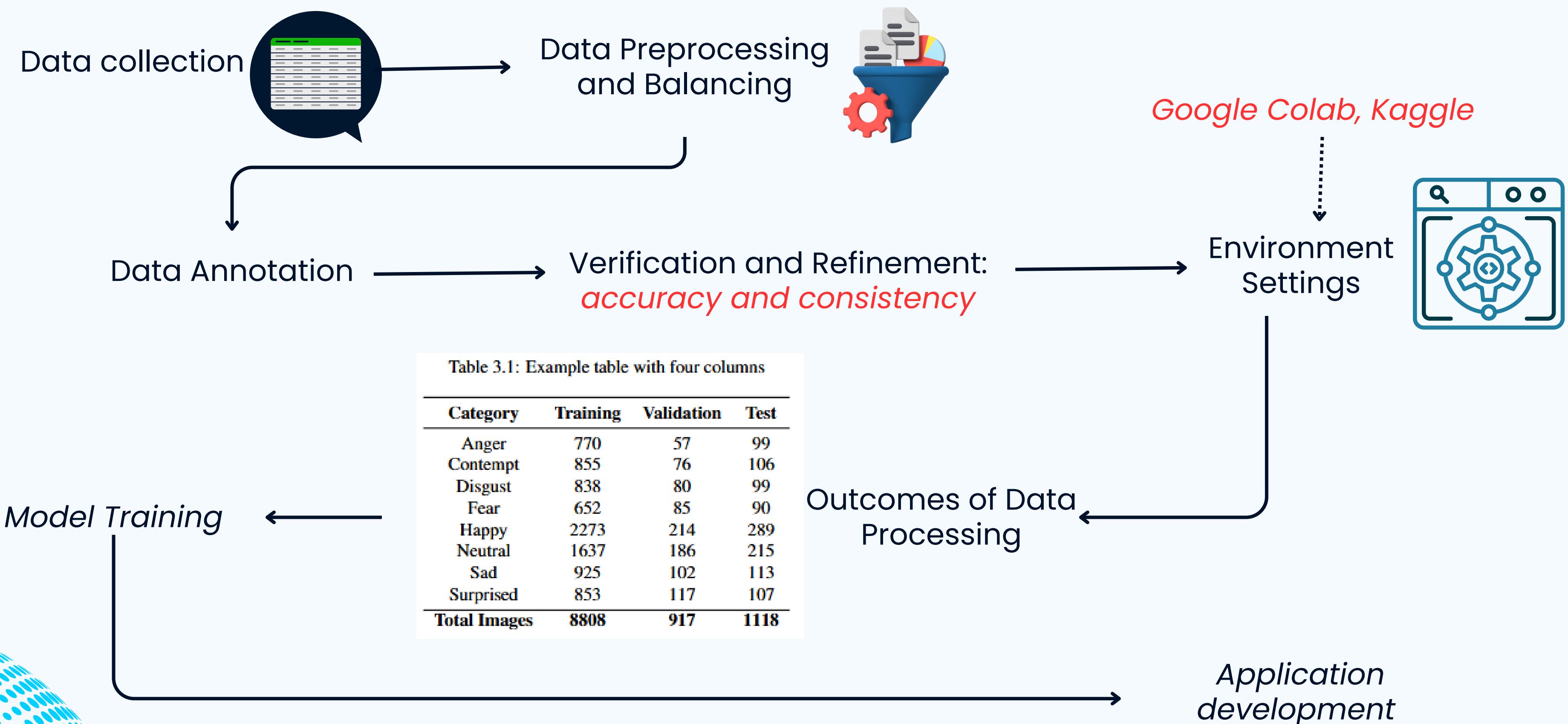
Motivation:

- Address rising mental health needs with accessible, AI-driven emotional support.
- Enhance human-computer interaction through empathetic, real-time emotion recognition.

Research Objective:

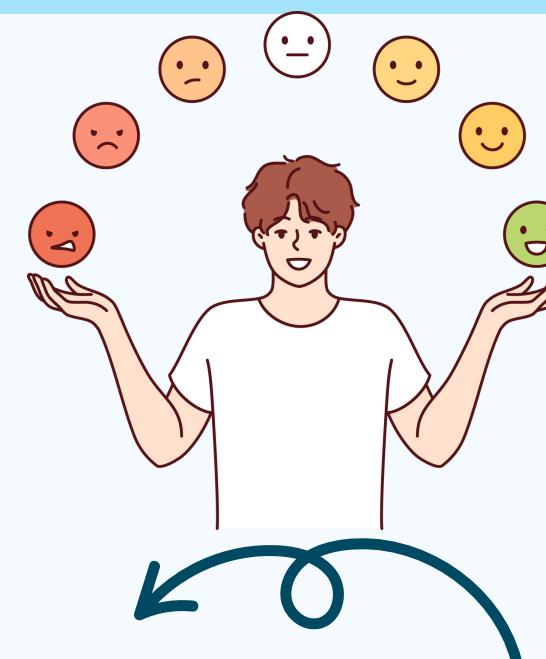
- Analyze facial expressions to recognize emotions
- Support medical diagnosis & human-computer interaction
- Use YOLOv12 + attention for real-time accuracy & speed

DATA & METHODOLOGY



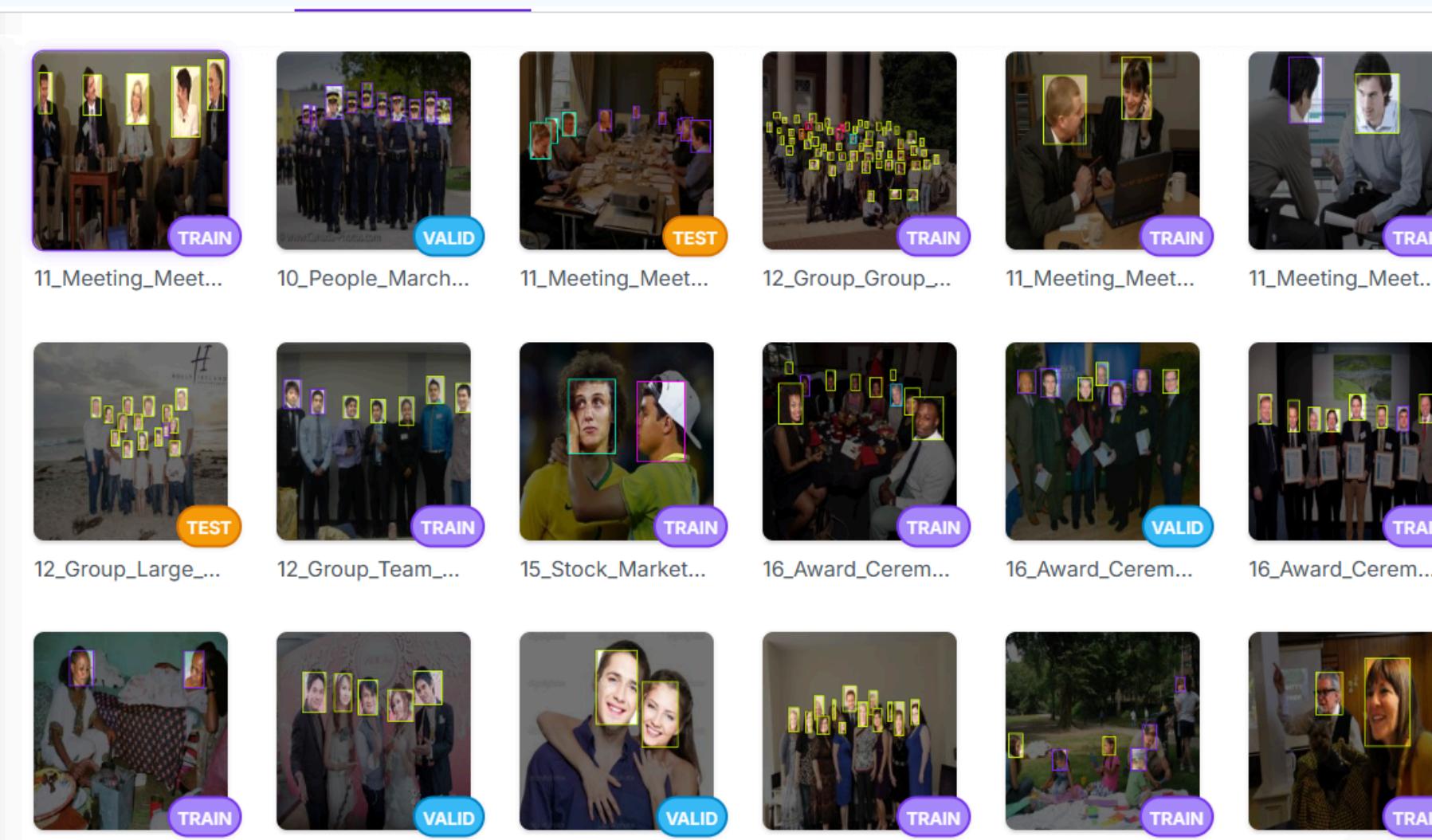
DATA PROCESS

- The dataset about **8821 images annotated by ourselves**
- **Sources:** Roboflow, the internet, and self-captured images
- **Expressions:** 8 emotion (*Angry, Contempt, Disgust, Fear, Surprise, Happy, Sad, Neutral*)
- **Images:** Focus on center of face



Data cleaning and label data:

- **Step 1:** Filter out blur and noise
- **Step 2:** Label images with appropriate emotions
- **Step 3:** Automatically extract faces (align, resize, store)



Multi-face image labeling

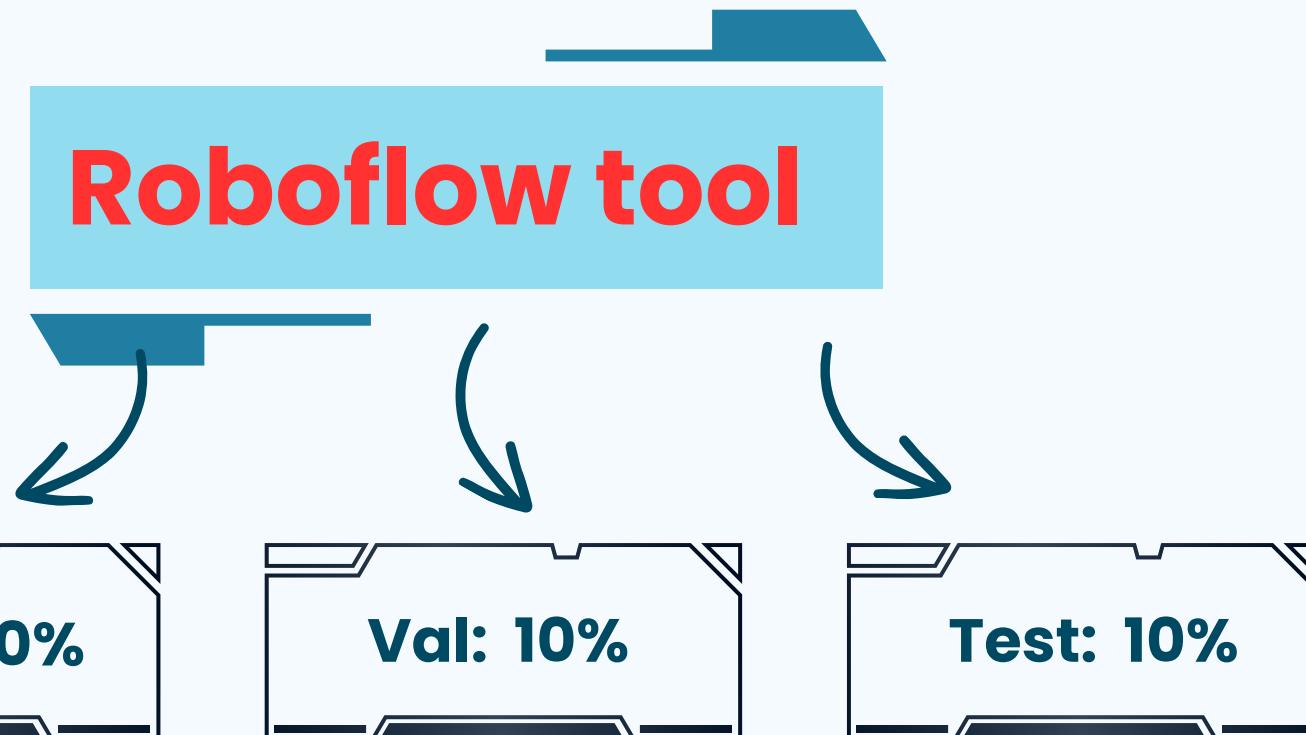


Roboflow tool

Train: 80%

Val: 10%

Test: 10%



DATA & METHODOLOGY

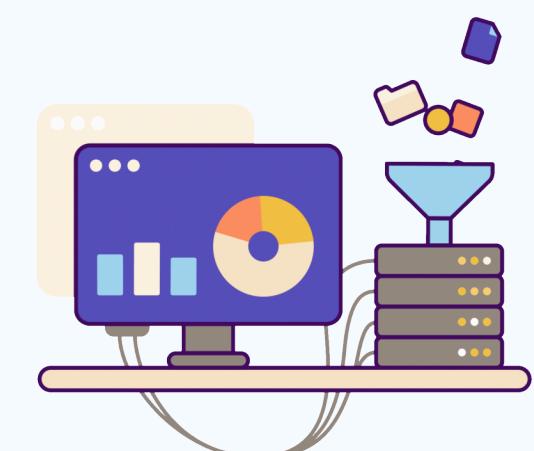
Emotion	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Anger	94.3	85.2
Contempt	96.4	87.3
Disgust	95.1	86.0
Fear	92.8	83.5
Happy	91.7	75.1
Neutral	92.5	82.9
Sad	98.1	89.4
Surprised	93.6	84.7
Overall	94.1	85.1

Model Evaluation Table. YOLOv12 performance on test set

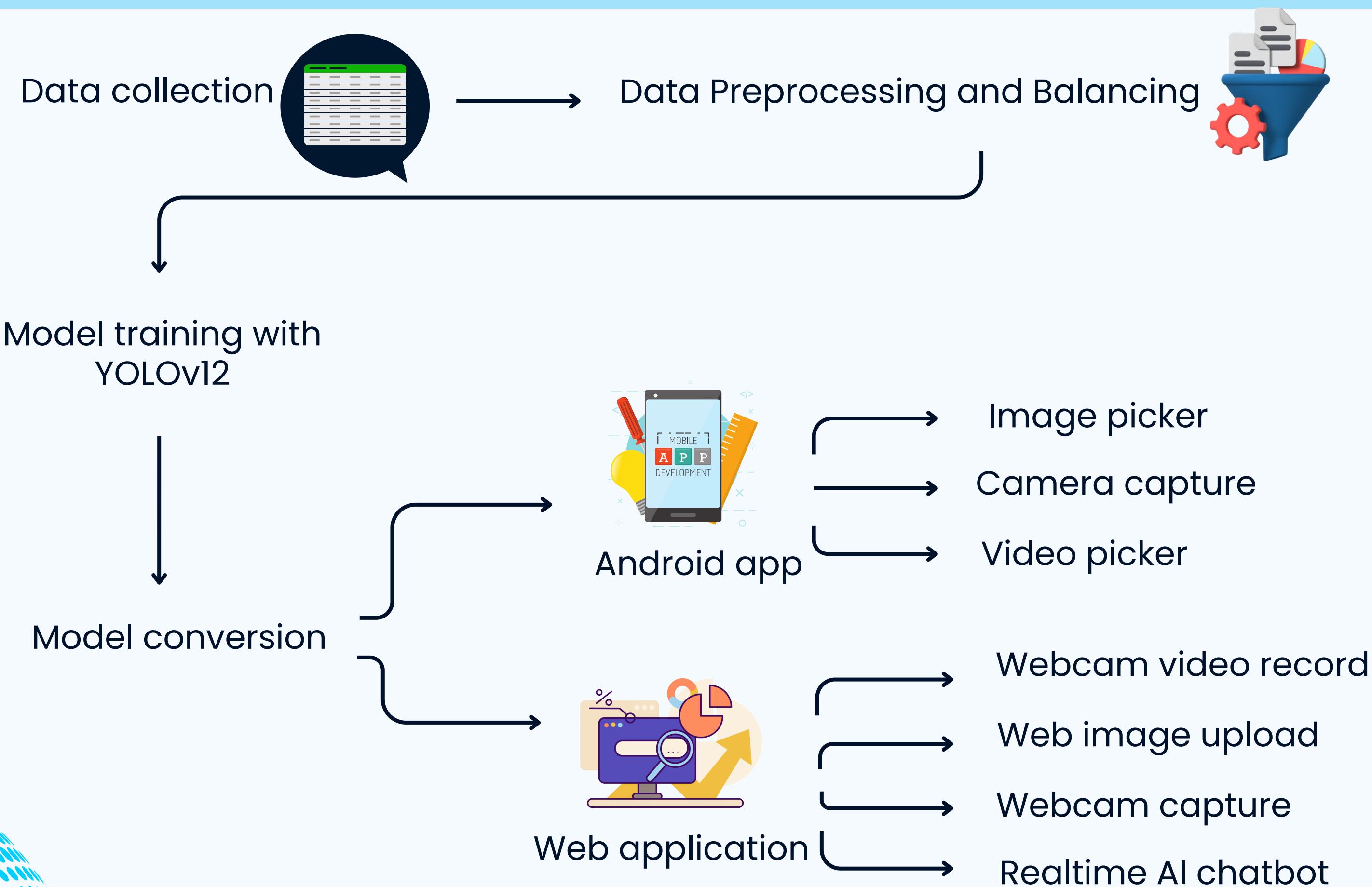
Performance Highlights (YOLOv12):

- **Sad:** Highest performance – mAP@0.5: 98.1%, mAP@0.5:0.95: 89.4%, indicating **strong detection** and **localization capabilities**.
- **Happy:** Lowest mAP@0.5:0.95 (75.1%) due to expression diversity; still strong at mAP@0.5: 91.7%.
- **Contempt:** Robust results – mAP@0.5: 96.4%, mAP@0.5:0.95: 87.3%.

Overall: mAP@0.5: 94.1%, mAP@0.5:0.95: 85.1%, Recall: 0.908 – indicating high accuracy and reliability.



DEVELOPMENT PROCESS



YOLOV12 MODEL



AREA ATTENTION MODULE (A2)

A2 divides the feature map into regions to reduce computational complexity and maintain a large receptive field.



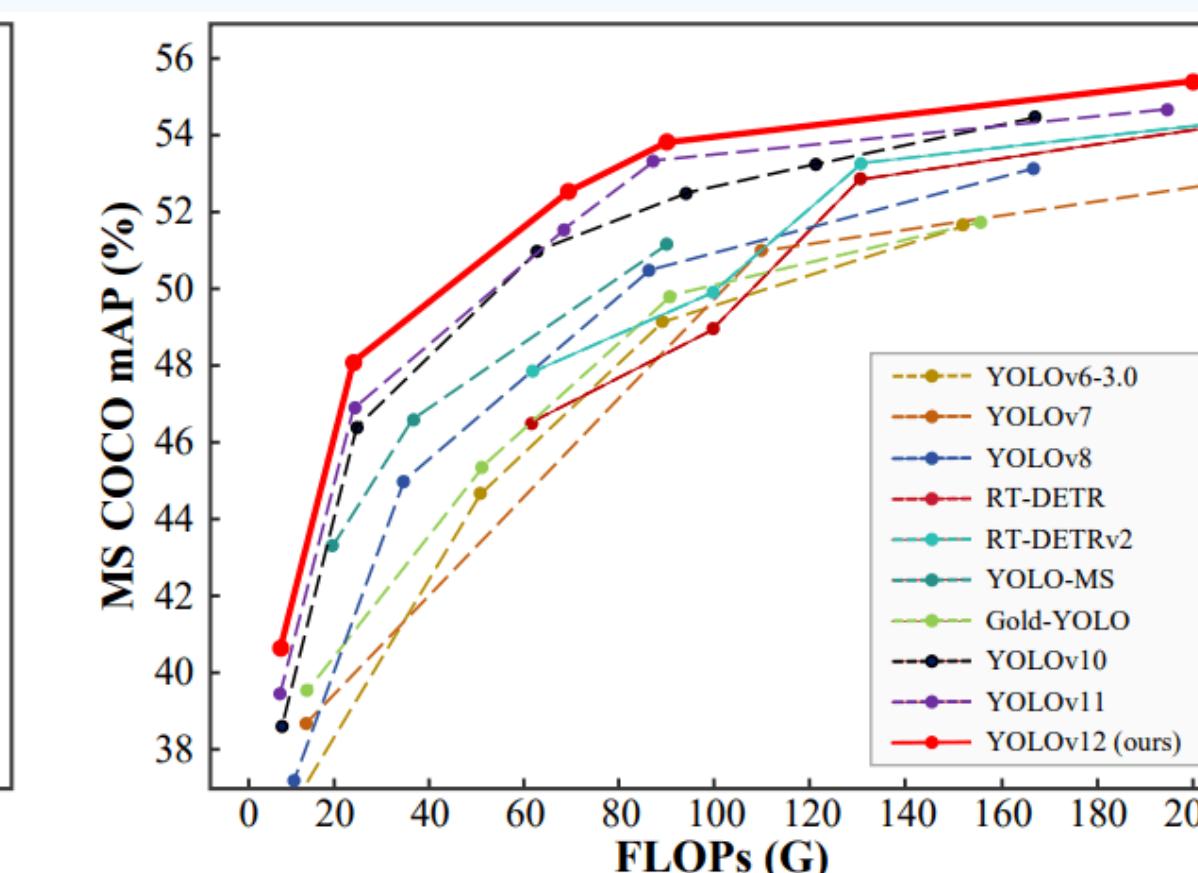
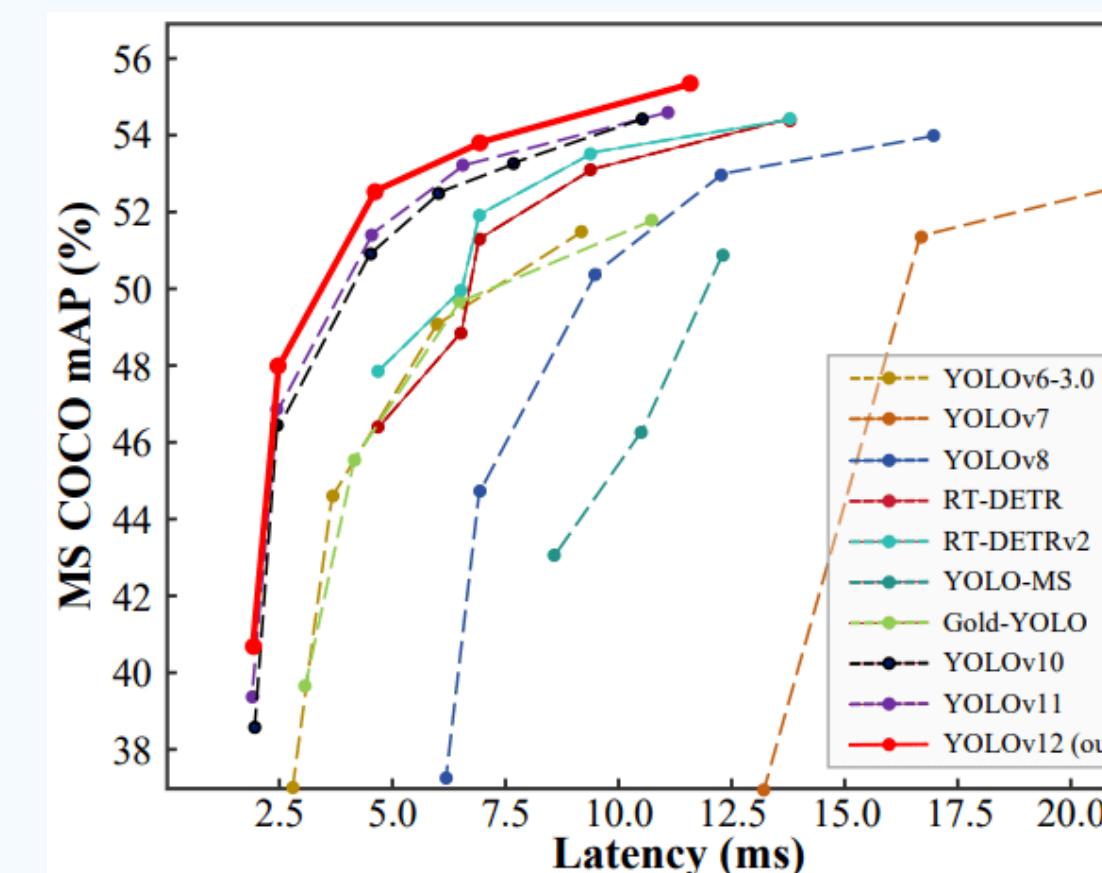
RESIDUAL EFFICIENT LAYER AGGREGATION NETWORKS (R-ELAN)

R-ELAN enhances feature aggregation and addresses optimization challenges when integrating the attention mechanism into YOLO.

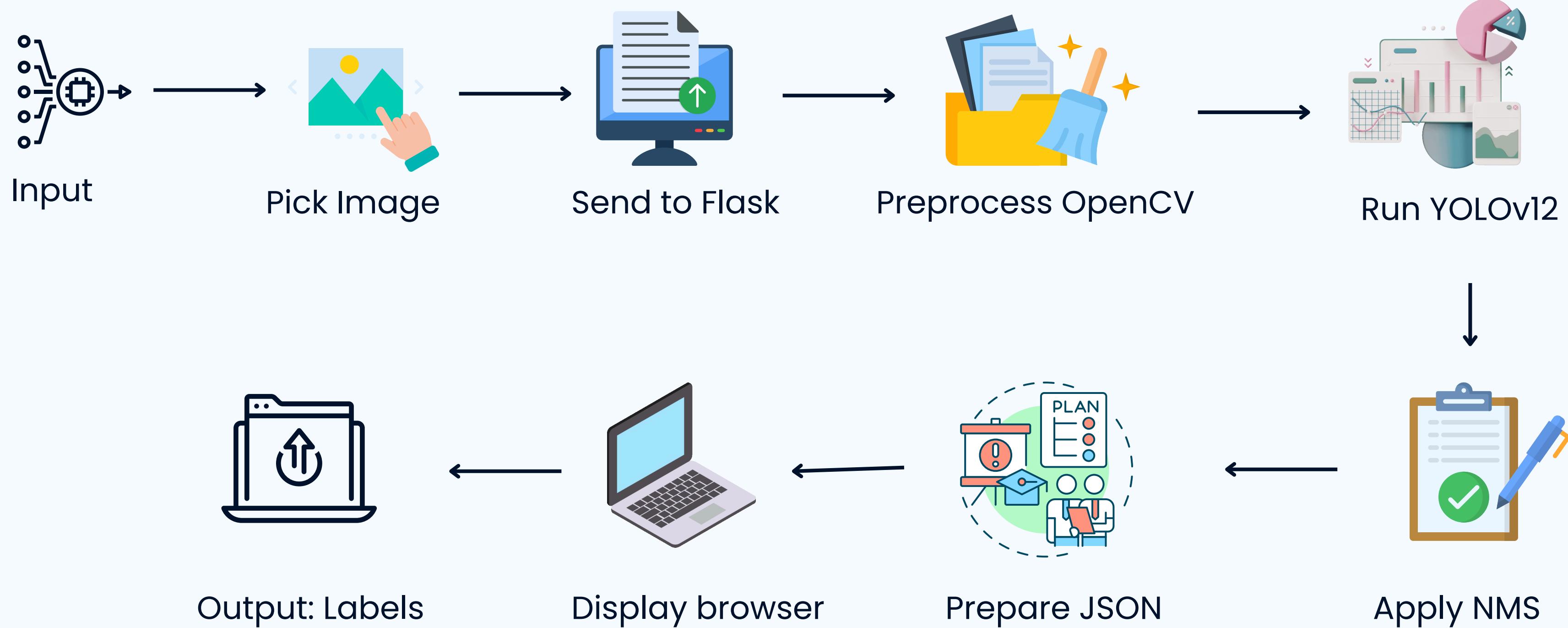


ARCHITECTURAL IMPROVEMENTS

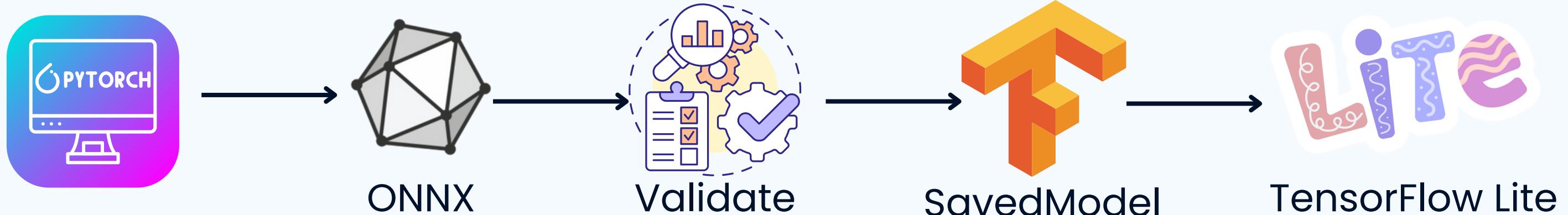
Incorporates Flash Attention, removes positional encoding, and adds Position Perceiver convolution for speed and accuracy balance.



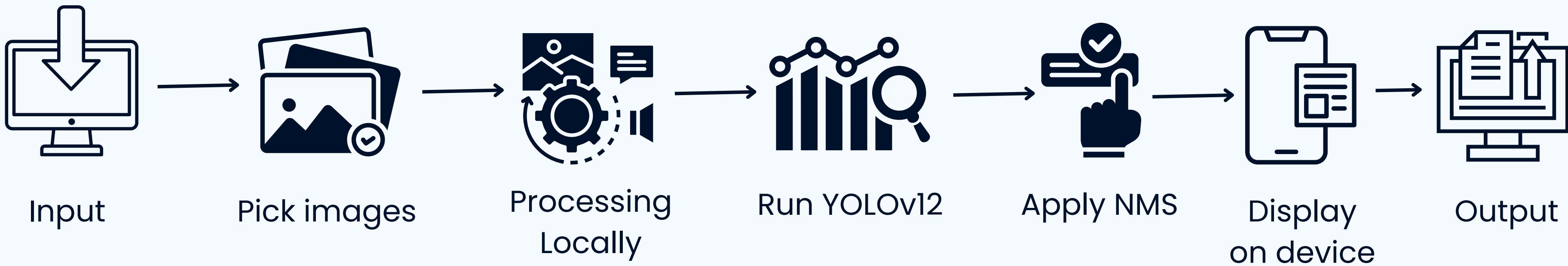
WEB FLOW



MODEL CONVERSION FLOW



ANDROID FLOW



CHATBOT AI

- **Operation:** Client sends webcam frames via Socket.IO every 100ms. Server processes frames using OpenCV and YOLOv12, then returns results to client.
- **Work:** In the first 5s after enabling the camera, an emotion buffer identifies the dominant emotion, ensuring the chatbot's response aligns with it and reflects reality.
- **Technology:** Uses Gemini 1.5 Pro API, supports multiple languages (English, Vietnamese).

The screenshot displays the EVision AI application interface, which integrates real-time emotion detection and AI-based conversation.

Realtime Emotion Detection (Left Panel):

- Header: "Realtime Emotion Detection".
- Image: A placeholder image with a camera icon and the text "Webcam is not active". Below it is a button "Turn On Webcam".
- Text: "Click 'Turn On Webcam' to start".

Chat with AI (Right Panel):

- Header: "Chat with AI".
- Text from AI: "Hey! I'm EVision AI, ready to chat! 😊"
- User Input: "hi brooo"
- Text from AI: "Hey brooo! 🙌😊 What's up? Been ages! Spill the tea 🍵... U been up to anything cool lately? 🌟 Tell me EVERYTHING! 😊"
- User Input: "im sad because of my deadlines :((
- Text from AI: "Hey bro! :(Deadlines got u down? That totally sucks! What kinda deadlines r we talkin' about? Exams? Projects? Smth else entirely? 🤔 Tell me all about it! Maybe we can brainstorm some solutions, or at least I can distract u with some totally awesome memes 😅. Don't worry, u got this! 🌟 We'll figure it out together. Just spill the tea 🍵 and let's chill 😊"
- User Input: "Can you recommend some music"
- Text from AI: "Hey bro! 😊 Music recs, u say? What kinda vibe u feelin' rn? 🥺 Smth chill to relax to, or are we pumpin' up the jams? 🎶🔥 Tell me ur fav artists or genres, and I can throw some stellar 🌟 suggestions ur way! Let's get this playlist poppin'! 🎶"

Bottom Status Bar:

- "Chatbot is active! Response sent! 🚀 ✅"

Emotion Detection Results (Bottom Center):

- Image: A video feed of a person wearing glasses, with a yellow bounding box highlighting their face.
- Text: "Contempt: 87%"
- Buttons: "Turn Off Webcam" and "Confidence Threshold" slider set to 50%.
- Metrics: FPS: 9.8, Duration: 13s, Result: Emotion: Contempt 😰 (Confidence: 87%)

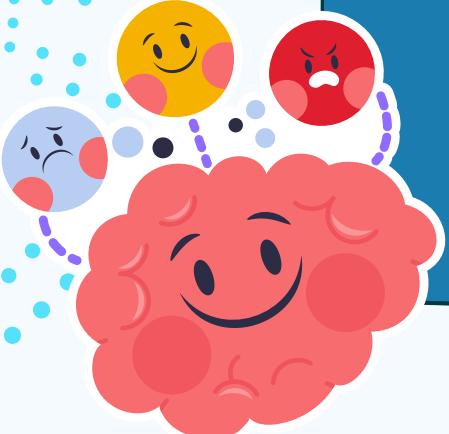
Input Fields:

- "Type a message..." input field.
- "Type a message..." input field in the Chat with AI panel.

CONCLUSION

OVERVIEW

- The “Evision” chatbot app leverages the YOLOv12 model to detect emotional states on Android and web platforms.
- Supports real-time video analysis and emotion-driven chatbot interactions.

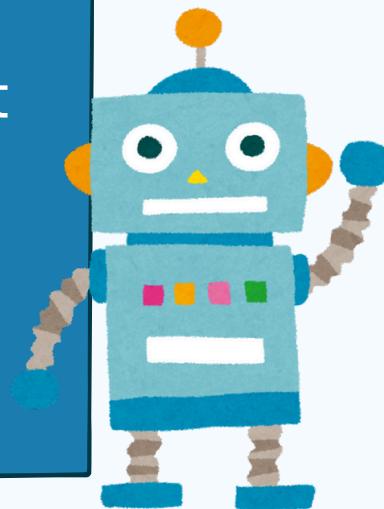


FUTURE DEVELOPMENT DIRECTIONS

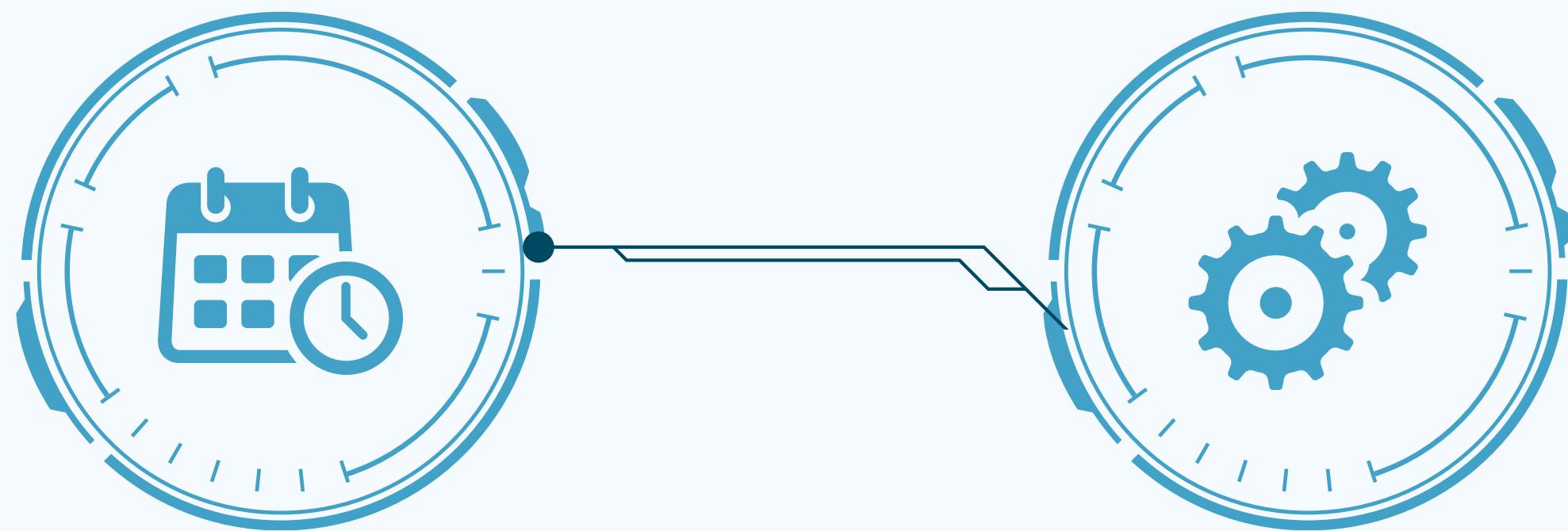
- Enhance recognition accuracy
- Upgrade interface, features such as voice to chatbot,...
- Recommend and autoplay songs

LIMITATION

- The Android version does not yet fully integrate the AI chatbot
- No iOS version is available

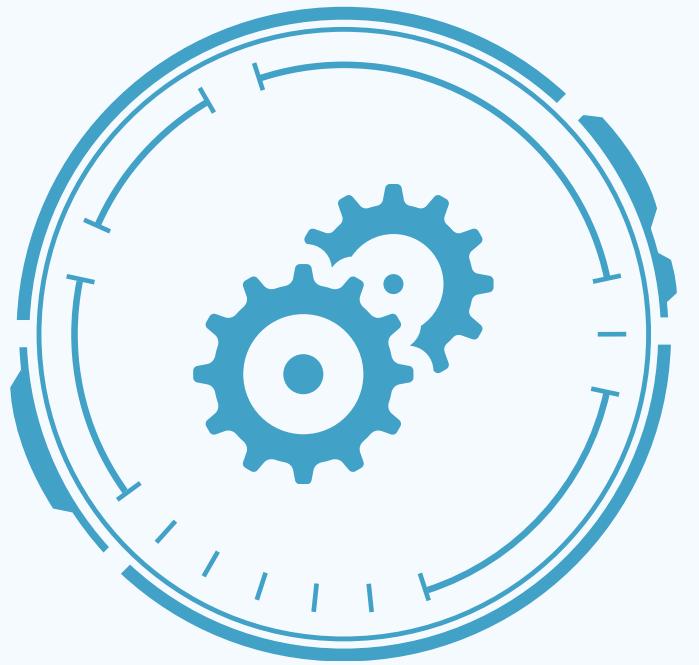


APP DEMO

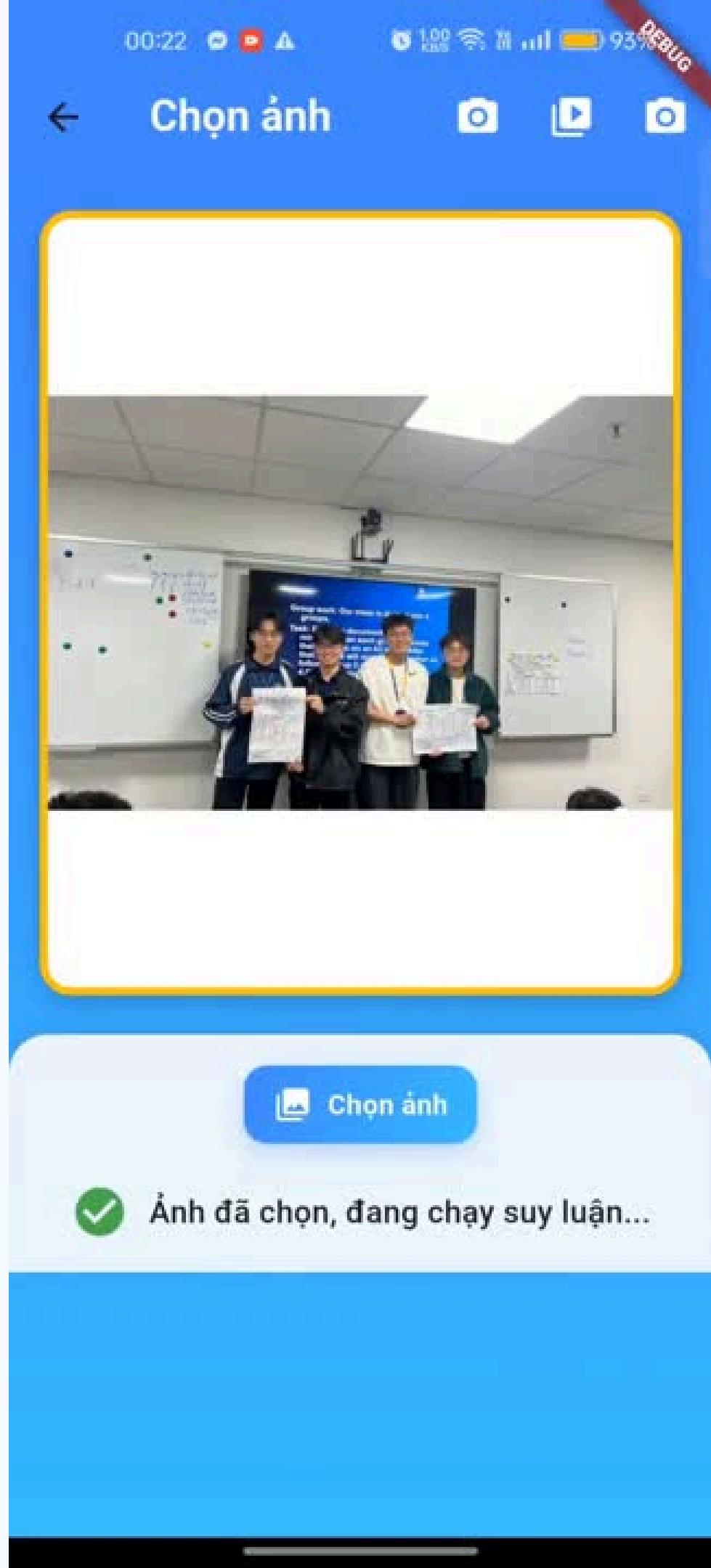


WEB

APP - ANDROID



APP – ANDROID



The image shows a screenshot of a web browser displaying the "EVision - Emotion Detection AI" application. The page has a light blue background with a central purple callout box containing text and a smiley face icon. Below the callout is a white rounded rectangle containing four colored buttons: blue, pink, green, and purple, each with an icon and text. The browser's header bar at the top shows the title and the URL "127.0.0.1:5000".

EVision

WEB

EVision - Emotion Detection AI

127.0.0.1:5000

Upload Capture Record Realtime

Discover Emotions with EVision AI 😊

Experience cutting-edge emotion detection through images, videos, and real-time webcam analysis. Connect with our friendly AI to explore your feelings!

Choose Your Experience

Explore the power of emotion detection with our intuitive tools.

Upload Image

Capture Photo

Record Video

Realtime Detection



ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG QUỐC TẾ
VNU-INTERNATIONAL SCHOOL

Thanks you
for listening!

