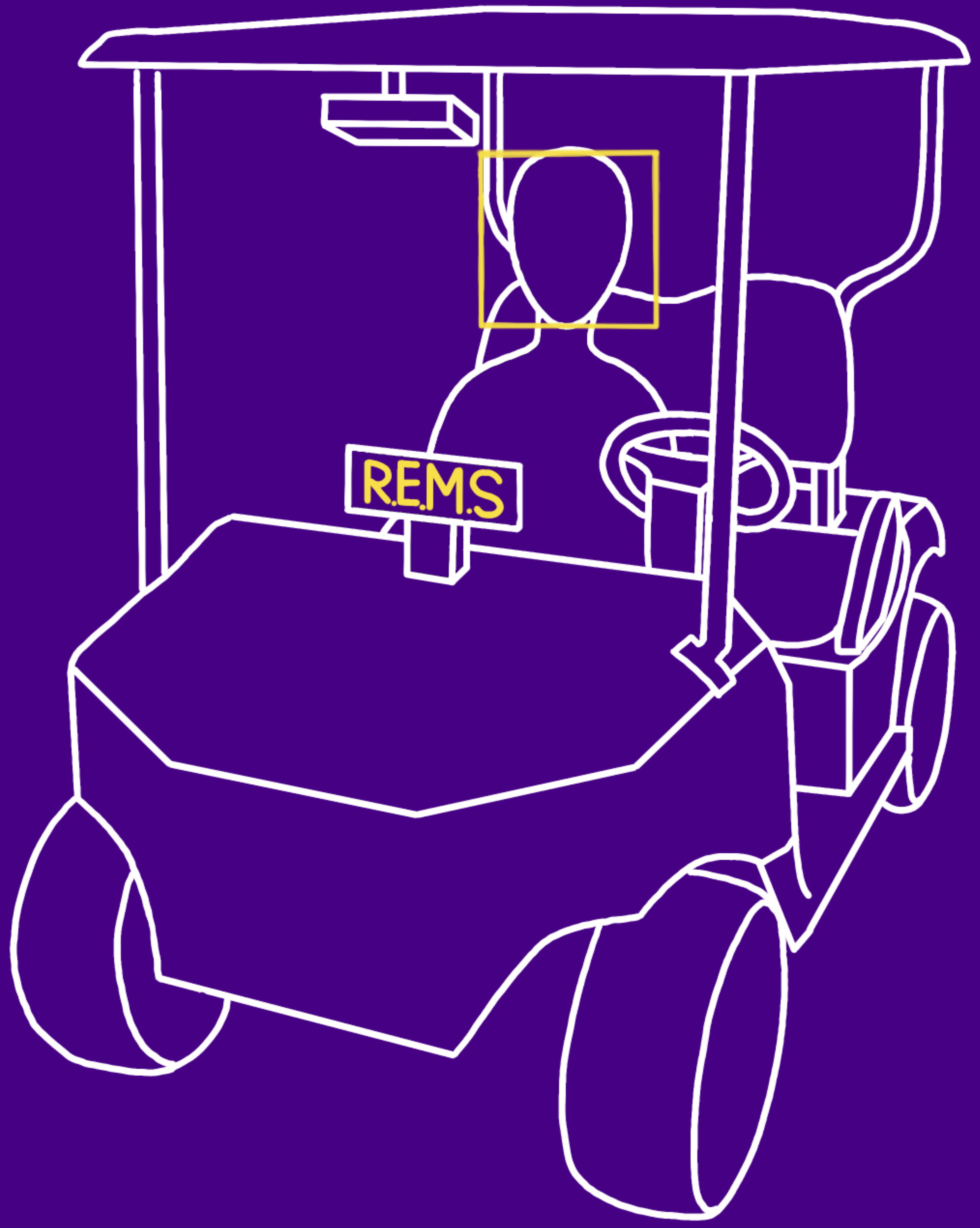
Real-Time Emotion Monitoring System (REMS) for Autonomous Vehicles

Enhancing Safety and Passenger Experience



AIU. × JMU

Content

Introduction

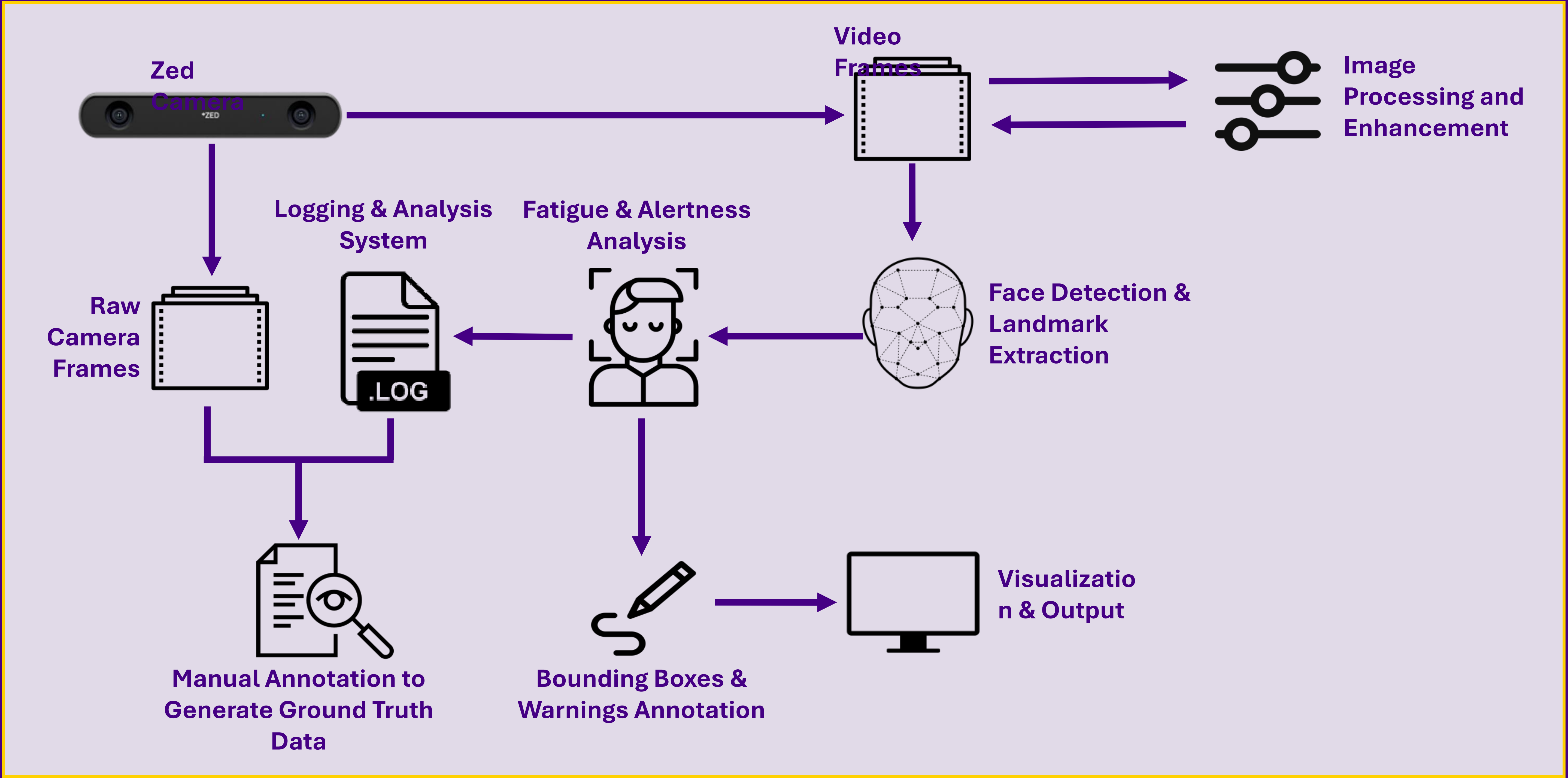
What is R.E.M.S?

AIU. × JMU

Introduction: What is R.E.M.S?

- Real-Time Emotion Monitoring System for Autonomous Vehicles.
- Enhances safety and passenger experience in autonomous carts, specifically for elderly individuals in retirement homes.
- Leverages computer vision and deep learning to analyze camera feeds for drowsiness, yawning, and emotional states.
- Provides immediate visual feedback and warnings.
- Designed for modularity and robustness.

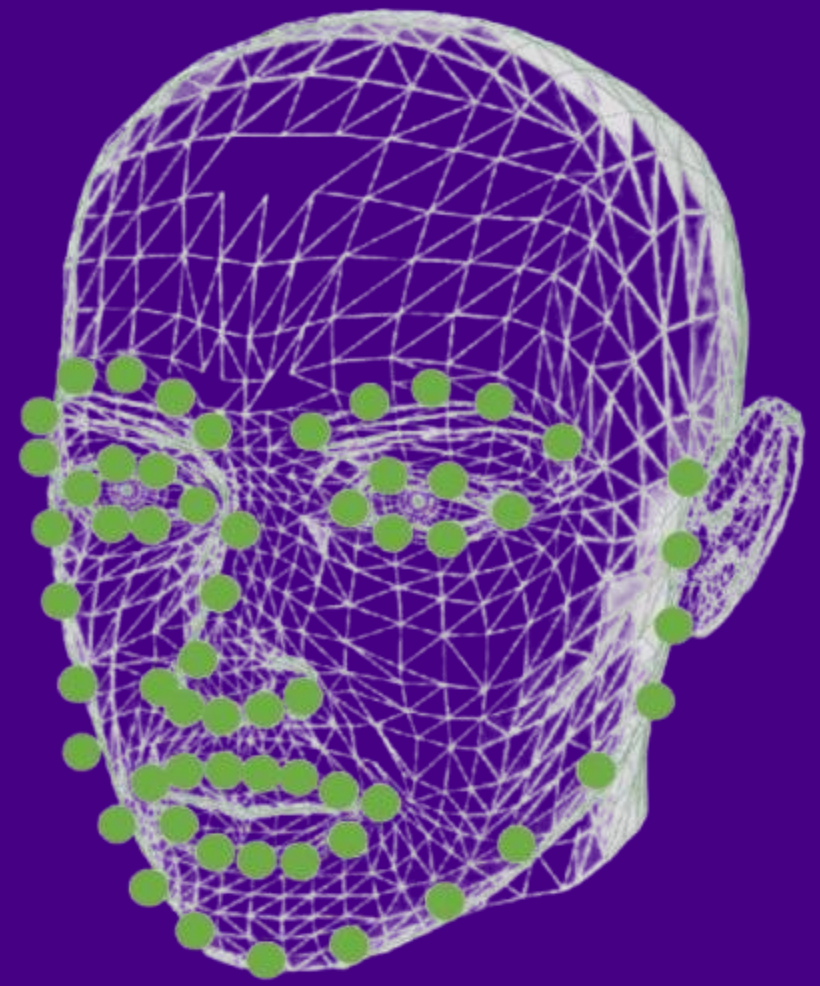
Introduction: System Architecture Diagram



Section 1: Old Model (DeepFace + dlib68)

Pipeline:

- Face detection: dlib68 (68-point landmark detector)
- Emotion classification: DeepFace (pretrained)



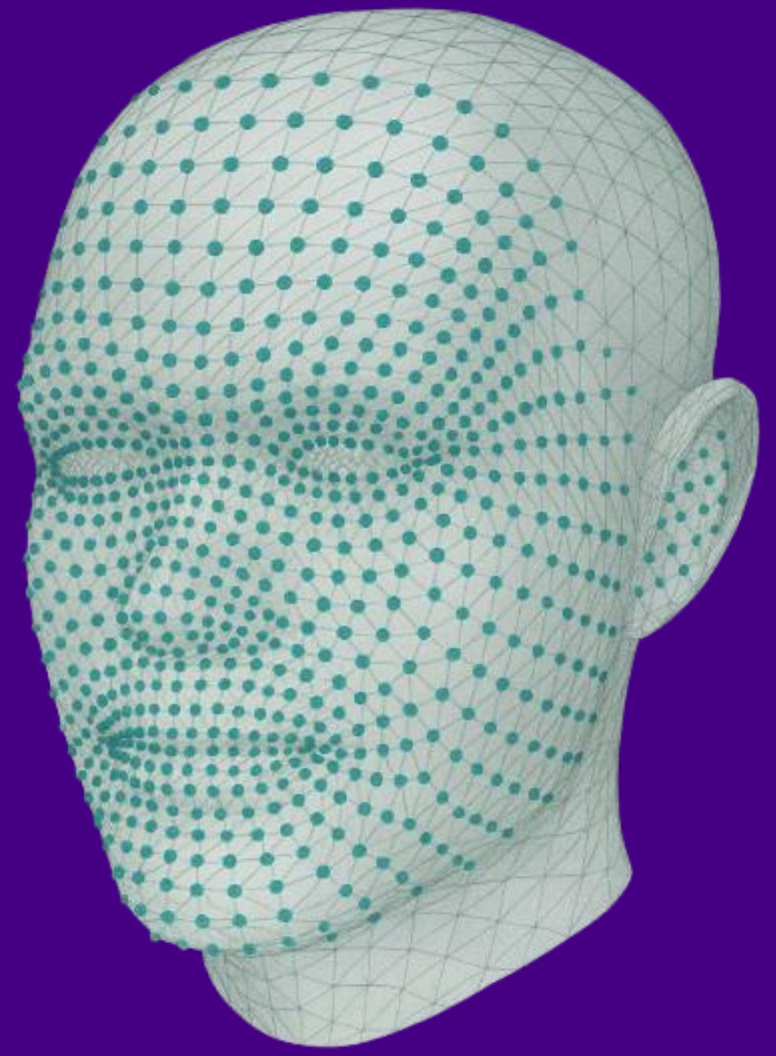
Section 1.1: Problems Identified

- **Inconsistent detection with glasses or head turns.**
- Weak performance in **low lighting.**
- No temporal smoothing → **unstable predictions.**
- Only **single-face** detection.

Section 2: New Model (DeepFace + MediPipe)

- **Accurate** tracking of **eyes, mouth, nose**
- **Robust** to partial face coverage (e.g., glasses)
- **Multi-face** support and stable tracking

•MediaPipe's Face Mesh solution estimates **468 3D face landmarks** in real-time.



Result:

Much more reliable system for in-motion carts with variable lighting and multiple passengers.

Section 2.1: New Model's Core Functionality

Tiredness Detection:

- Eye Aspect Ratio (EAR) for blinking/eye closure.
- Mouth Aspect Ratio (MAR) for yawning.
- Head Nod Detection: Compares the vertical displacement of the nose tip.

Emotion Correction & Smoothing:

- Applies smoothing to avoid flickering/unstable predictions.
- Logic to override fear or sad when facial context contradicts (e.g., fear downgraded to neutral if calm, sadness reclassified if mouth corners raised or EAR high).
- Uses rolling emotion history for stability.

Visual Output:

Real-time bounding box, emotion label, drowsiness/ yawning/ head nodding warnings, facial landmarks.

Section 3: The Bad Lighting Problem & Solution

Challenge: Emotion misclassification in low lighting or poor face angles.

Solution: Advanced Image Processing & Cropping:-

•A new, tunable **image processing function** has been developed and integrated.

•Enables **enhancement of video frames** for improved detection accuracy in all lighting conditions.

•**Can be adjusted** as needed for different environments.

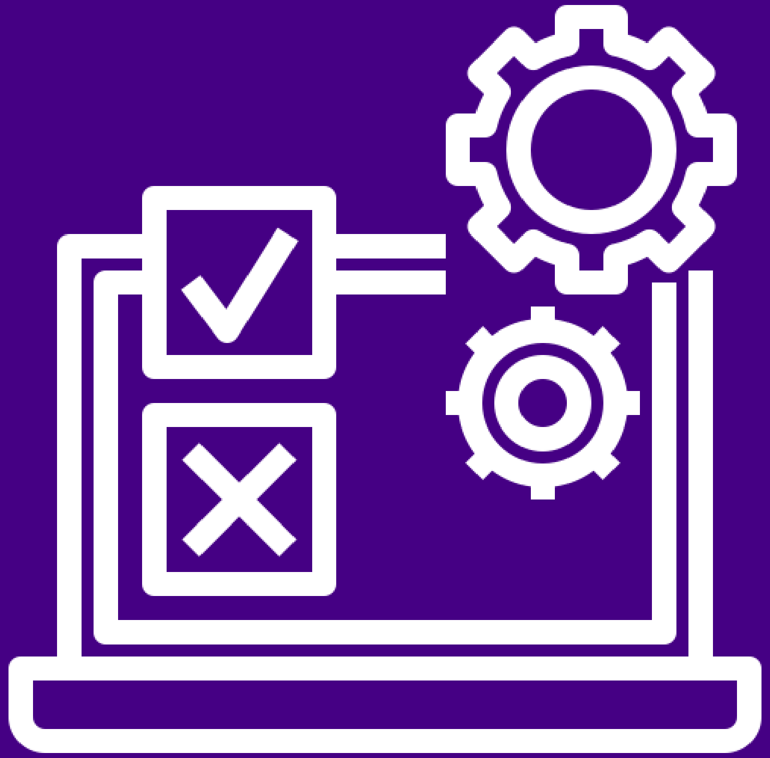
•`crop_and_enhance_passenger.py` script specifically processes passenger camera images, cropping and enhancing them.

•More **reliable** detection and analysis when **lighting conditions change**.

Section 3.1 : The Bad Lighting Problem & Solution Example

After

Before



Section 4

Testing Methodologies

Section 4.1: Testing Methodology - Overview

Comprehensive Analysis

Performed a comprehensive analysis of the model's performance using **all available log files.**

Automated Logging & Manual Annotation

- Detected emotions and states are **automatically logged** at regular intervals during live tests.
- After each test, raw video footage is **manually reviewed** and annotated to provide ground truth labels for each person at each timestamp.
- This enables **direct comparison** between predicted and true emotions for quantitative analysis.

AIU. × JMU

Section 4.2
Testing With Cameras

Section 4.2.1: Testing With Both Front And Back Cameras

Was done to assess the system's ability to detect passenger emotions and understand potential environmental influences from different camera perspectives.

AIU. × JMU

Section 4.2.2: Testing With Only The Back Camera

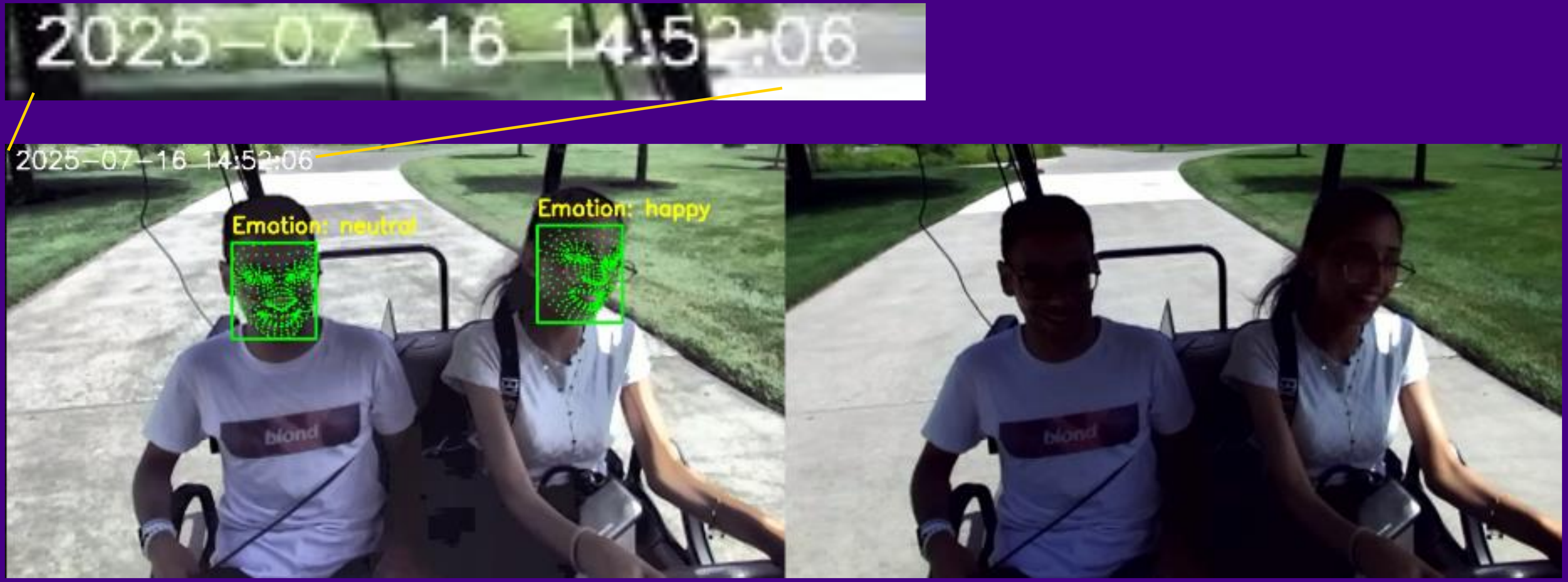
Helped with focusing on the model's performance, this is implied by the **focus on the passenger monitoring system**, where the "back camera" would be the primary one observing the passenger. The overall performance metrics are derived from this.

AIU. × JMU

Section 4.3: Testing With Original Footage Next To The Detected Footage

Dual-Frame Video Recording with Logging:

- The system outputs videos with both raw and processed frames side by side.
- This makes it easier to compare detection results with the original footage and validate detections




```
timestamp,emotion_1,emotion_2,true_emotion1,true_emotion2
2025-07-16 14:52:06,happy,neutral
```

Section 4.4: Manually Logging The Actual Passenger's Emotion Per Frame

This is a core part of the testing methodology for ground truth validation.

```
timestamp,emotion_1,emotion_2,true_emotion1,true_emotion2
2025-07-16 14:52:06,happy,neutral
```

```
2025-07-16 14:52:06,happy,neutral,happy,neutral
```

Section 5

Current Model's Performance

Section 5.1: Current Model's Accuracy

Overall Accuracy: 85% (Person 1), 84% (Person 2), 85% (combined). This demonstrates strong and consistent performance across different individuals.

Per-Emotion Accuracy

Happy	Sad	Angry	Neutral	Fear
91%	96%	95%	76%	96%

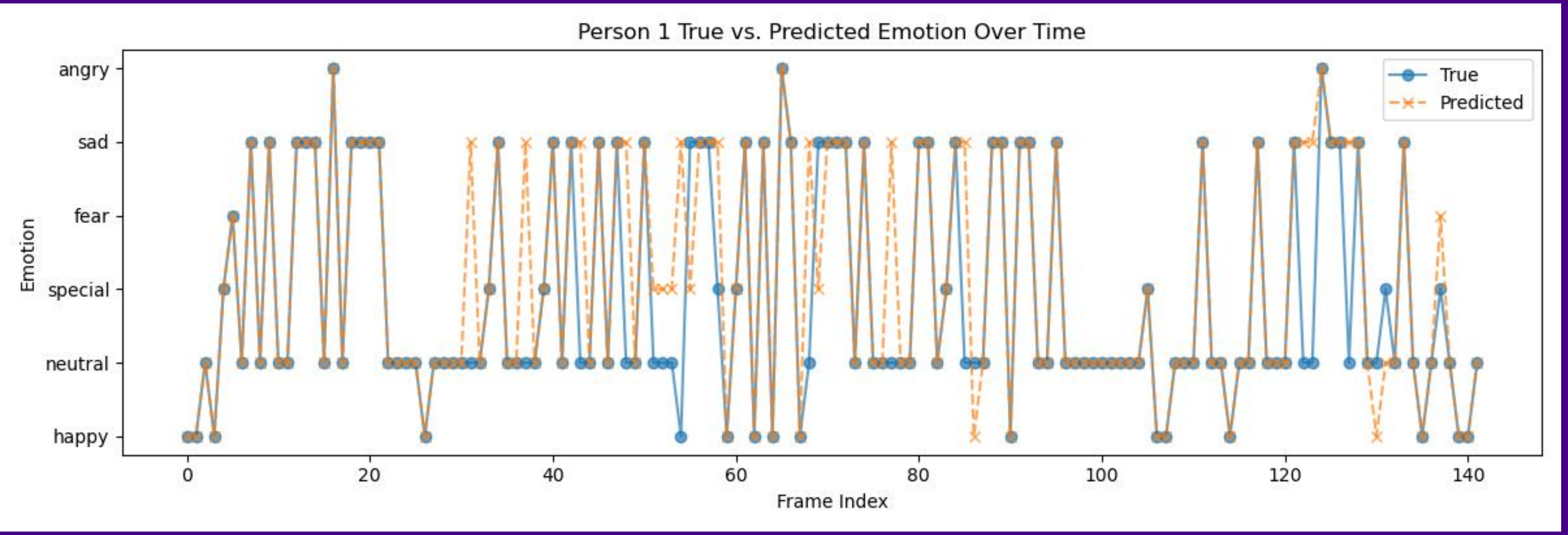
AIU. × JMU

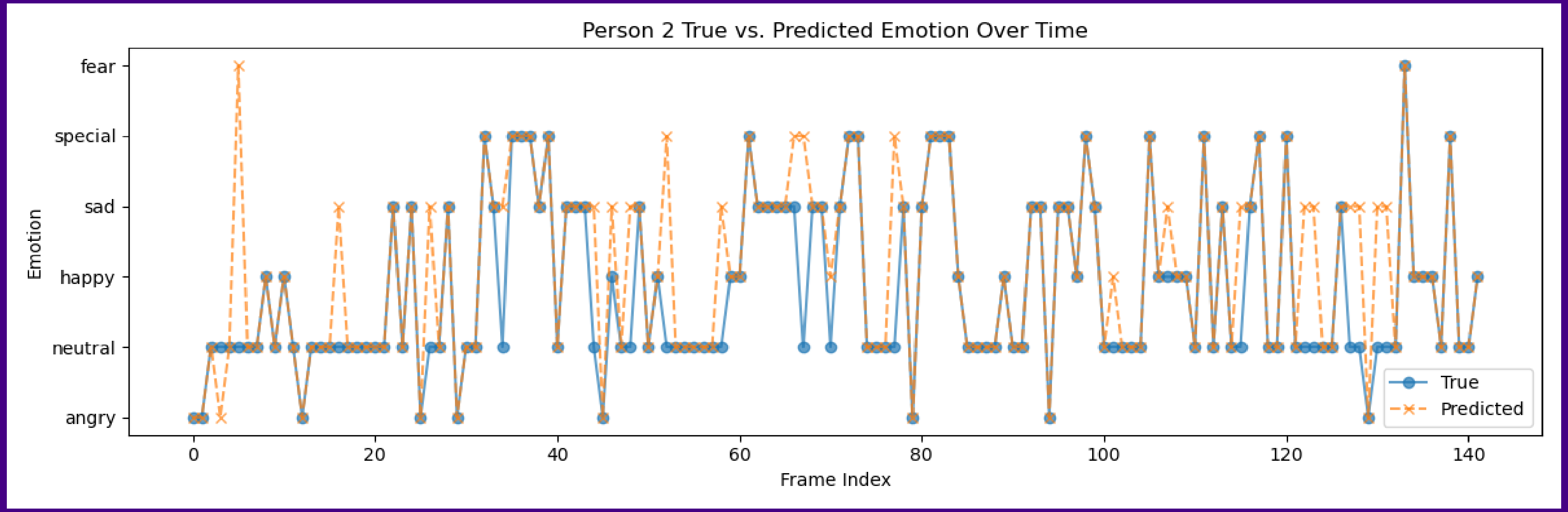
Section 5.2: Current Model's Metrics & Visualizations

Precision, Recall, F1-score (Examples):

Our model exhibits particularly high performance for **'Happy'**, **'Sad'**, and **'Angry'** emotions, indicating robust detection in these categories.

Happy: Precision 0.88, Recall 0.91, F1 0.90
Sad: Precision 0.71, Recall 0.96, F1 0.81
Angry: Precision 0.92, Recall 1.00, F1 0.96
Neutral: Precision 0.99, Recall 0.76, F1 0.86
Fear: Precision 0.50, Recall 1.00, F1 0.67





Section 6
Our Custom Model

Section 6.1: Model's Architecture

- **Base: ResNet-18**
  A lightweight and efficient CNN backbone for feature extraction.
- **Enhancement: CBAM** (Convolutional Block Attention Module)
  Adds attention mechanisms to focus on the most relevant or important features
- **Dataset: FER2013** (7 emotions) (35,887 Images)
- **Preprocessing:** Convert grayscale → RGB, resize to 224x224 and **normalize** pixel values using ImageNet mean & std
- **Augmentation: Albumentations** library, random transforms like flip, blur, contrast, dropout to improve generalization.

Section 6.2: Model's Training

Multi-stage Training

Stage 1:
Classifier head
Train only the last layers first to help the model adapt gradually

Stage 2:
End-to-end fine-tuning Unfreeze the full model and fine-tune all layers for better accuracy

Section 6.3: Model's Techniques

Label smoothing: Helps prevent the model from becoming overconfident, so it leads to better generalization on unseen data.

Class weights: Handles class imbalance by giving more importance to rare classes.

Data augmentation: Applies random changes like flipping, cropping, and brightness adjustment to generate diverse training samples helping the model perform better on new, unseen images.

Section 6.4: Model's Challenges

Emotion Similarity

Similar expressions (e.g., sad vs. neutral) caused confusion.
Solution: Used CBAM and smoothing to improve distinction.

Overfitting Risk

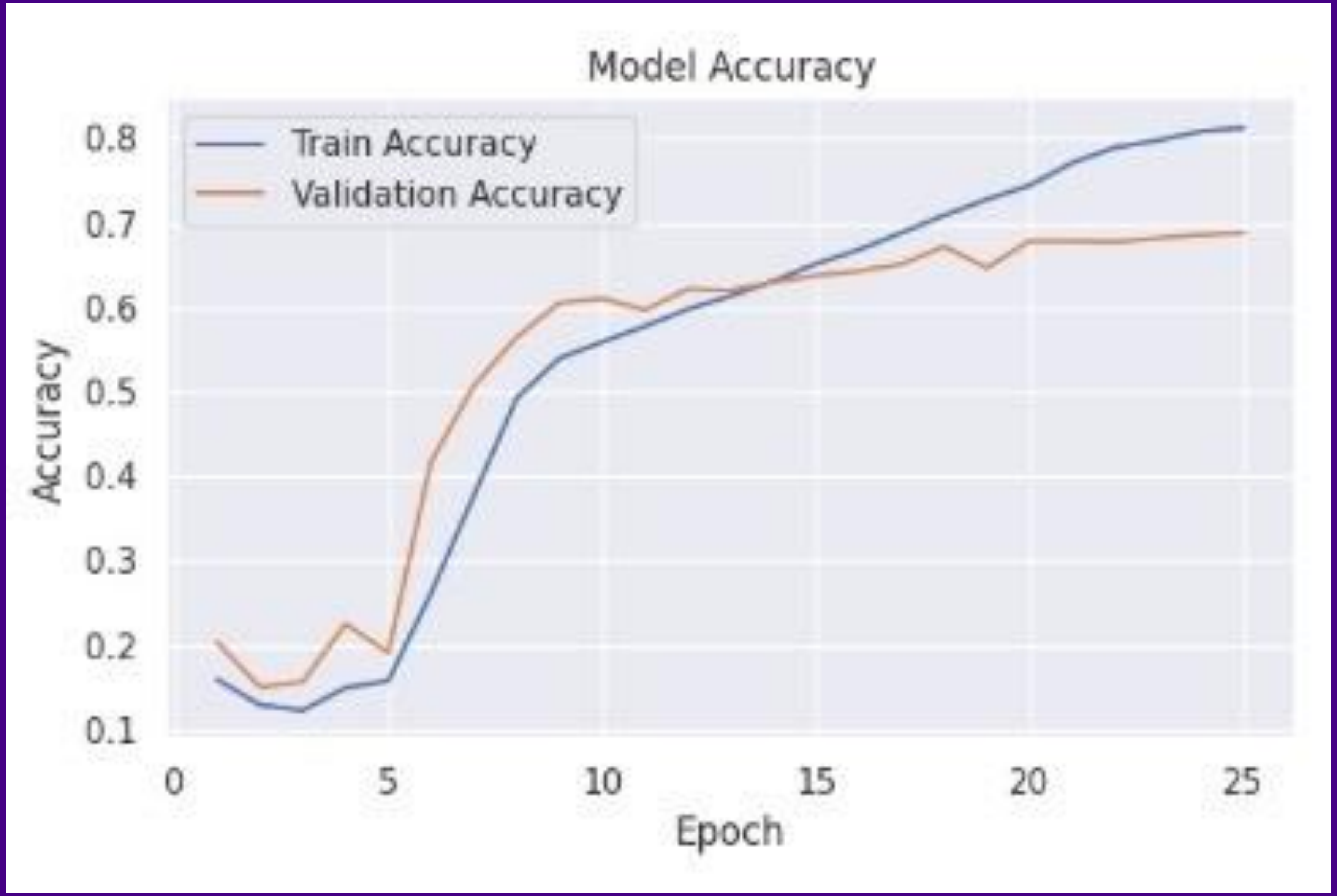
Small dataset size made the model likely to memorize training data.
Solution: Used **data augmentation, dropout**, and **early stopping**.

AIU. × JMU

Section 6.5: Results

Total Accuracy is 70.0%

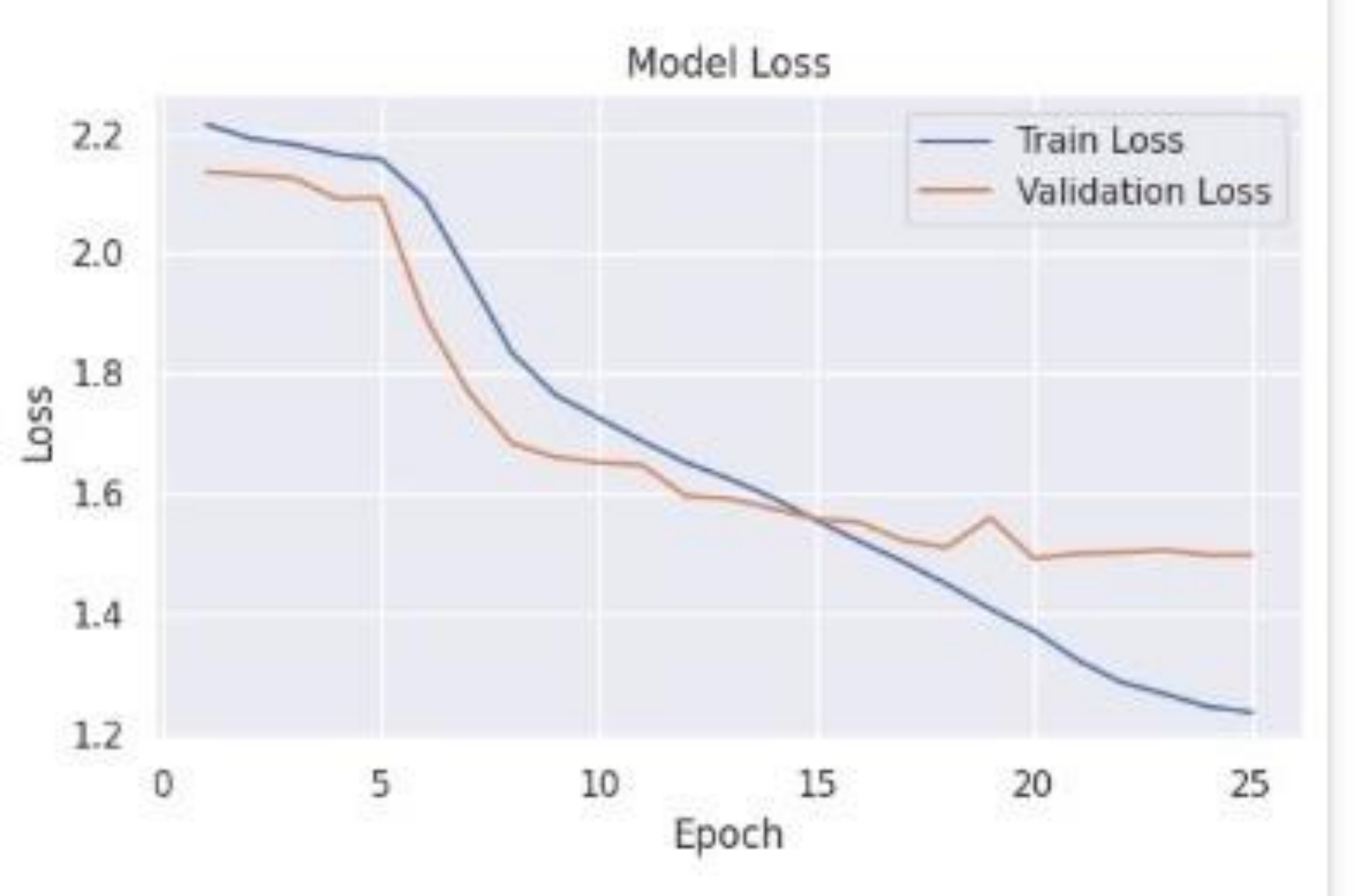
- Accuracy improved steadily (Train & Val)
- Validation accuracy peaked at 70%



Model Accuracy over Epochs

Section 6.5: Results

- Loss **decreased** consistently
- **No** signs of overfitting



Model Loss over Epochs

Section 6.5: Results

- **Best F1** (happy): **0.89**
- **Macro F1** (7 classes): **0.68**

Confusion Matrix: Shows how well the model predicted each emotion

Confusion Matrix – Predicted vs. Actual Emotions

Section 7: Future Work and Next Steps

- **Test the new pre-built (custom) model** on live webcam/video data and evaluate its real-time performance.
- **Run both models** (current and custom) side by side in real-time and directly compare their accuracy, stability, and robustness in practical scenarios.
- **Integrate additional sensors** (e.g., heart rate, steering input) for multi-modal monitoring.
- **Expand emotion and fatigue detection** to include more subtle cues and longer-term trends.
- **Deploy the system** in real vehicles and collect more data for validation and improvement.

Conclusion

- The REMS project has achieved significant advancements in **image processing**, **emotion detection** (with 85% accuracy for the current model), and overall system robustness.
- It significantly enhances safety and passenger experience in autonomous carts, particularly for elderly individuals, by providing real-time monitoring of emotional and physical states.
- The system ensures a secure and comfortable journey through immediate visual feedback and proactive warnings.

Thank you!

QA Time!