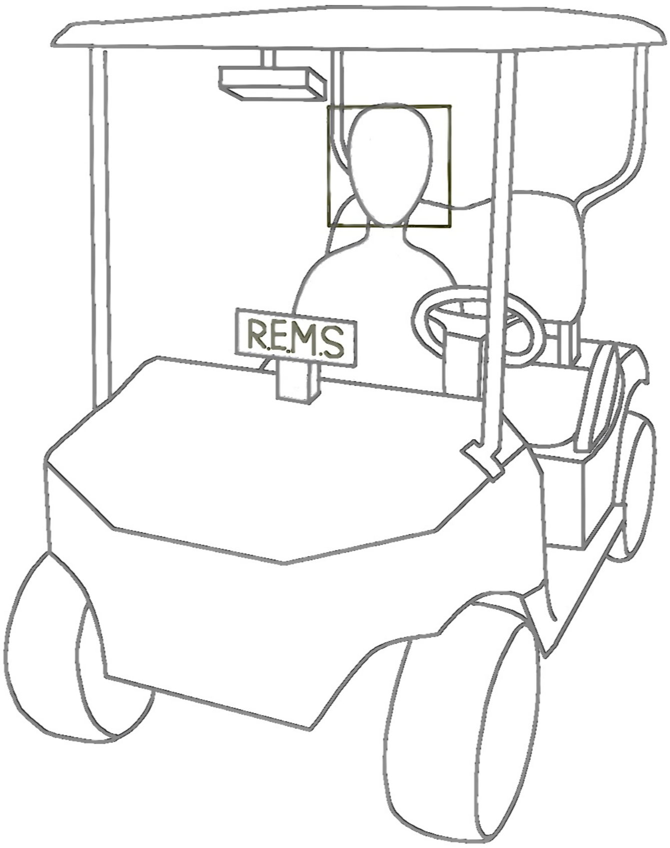
JAMES MADISON UNIVERSITY

Information Technology (IT) Program
IT 445 - Capstone Implementation
Summer 2025



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

Final Project Report

Instructor:
Dr. Samy El-Tawab

Authors:
Rana Moussa
Ramez Asaad

Contents

1 Executive Summary

The Real-Time Emotion Monitoring System (REMS) represents a comprehensive solution designed to enhance safety and passenger experience in autonomous vehicles, with particular focus on elderly individuals in retirement home transportation systems. Over the course of five weeks, this project has evolved from a basic emotion detection system to a robust, multi-modal monitoring platform capable of real-time analysis of passenger emotional and physical states.

The system utilizes advanced computer vision techniques, deep learning models, and sophisticated image processing to analyze camera feeds for drowsiness detection, yawning recognition, and comprehensive analysis of emotional states. The final implementation achieves 85% accuracy in emotion detection while maintaining real-time performance and robustness across varying lighting conditions and multi-passenger scenarios..

2 Introduction and Project Overview

2.1 Project Motivation

Autonomous vehicles represent the future of transportation, particularly for vulnerable populations such as elderly individuals in retirement communities. The REMS project addresses critical safety concerns by providing continuous monitoring of passenger states, enabling proactive intervention when safety risks are detected. The system's ability to detect drowsiness, emotional distress, and other concerning states makes it invaluable for ensuring passenger wellbeing during autonomous transportation.

2.2 System Architecture

The REMS architecture follows a modular design approach, integrating multiple components for comprehensive passenger monitoring:

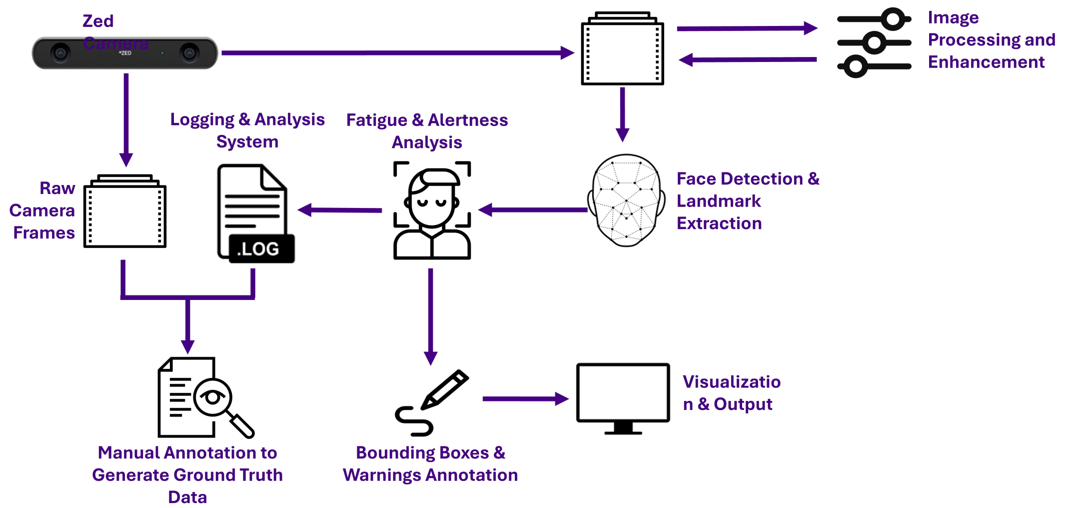


Figure 1: REMS System Architecture Overview - showing data flow from ZED cameras through image processing, face detection, analysis engine, to output visualization and logging systems

- **Data Acquisition:** ZED stereo cameras capture high-quality video feeds
- **Image Processing:** Advanced enhancement algorithms optimize frames for analysis
- **Face Detection:** MediaPipe Face Mesh provides robust 468-point landmark detection
- **Analysis Engine:** Multiple algorithms process facial features for state determination
- **Output System:** Real-time visual feedback and automated logging
- **Validation Framework:** Manual annotation and performance analysis tools

2.3 Project Objectives

The primary objectives of the REMS project include:

1. Develop a real-time emotion monitoring system with high accuracy
2. Implement robust drowsiness and fatigue detection algorithms
3. Create adaptive image processing for varying environmental conditions
4. Establish comprehensive testing and validation methodologies
5. Build a custom deep learning model for improved emotion recognition
6. Ensure system modularity for easy deployment and maintenance

3 System Evolution and Model Comparison

3.1 Initial Implementation: DeepFace with dlib68

The project began with a baseline implementation utilizing dlib's 68-point landmark detector combined with DeepFace for emotion classification. This initial approach, while functional, revealed several critical limitations:

Technical Specifications:

- Face detection: dlib68 (68-point landmark detector)
- Emotion classification: DeepFace (pretrained models)
- Single-face detection capability
- Basic temporal processing

Identified Problems:

- Inconsistent detection performance with glasses or head rotations
- Poor performance in low-lighting conditions
- Absence of temporal smoothing leading to unstable predictions
- Limited to single-face scenarios
- High false positive rates in challenging conditions

3.2 Enhanced Implementation: DeepFace with MediaPipe

Recognizing the limitations of the initial approach, the system was redesigned around MediaPipe's Face Mesh solution, providing significant improvements in robustness and accuracy.

Technical Enhancements:

- **Face Detection:** MediaPipe Face Mesh with 468 3D landmark estimation
- **Multi-face Support:** Simultaneous tracking of multiple passengers
- **Improved Robustness:** Better handling of partial face coverage and accessories
- **Real-time Performance:** Optimized for in-motion vehicle scenarios

Core Functionality Improvements:

Tiredness Detection:

- Eye Aspect Ratio (EAR) calculation for blink and eye closure detection
- Mouth Aspect Ratio (MAR) analysis for yawning identification
- Head Nod Detection through vertical displacement tracking of nose tip
- Enhanced logic requiring both eyes closed for ¿3 seconds plus head tilt

Emotion Correction and Smoothing:

- Temporal smoothing algorithms to eliminate prediction flickering
- Context-aware emotion override logic
- Rolling emotion history for stability enhancement
- Advanced filtering for fear and sadness misclassifications

Visual Output Enhancement:

- Real-time bounding box visualization
- Dynamic emotion label display
- Comprehensive warning systems for drowsiness, yawning, and head nodding
- Detailed facial landmark overlays

4 Advanced Image Processing and Environmental Adaptation

4.1 The Lighting Challenge

One of the most significant challenges encountered during development was emotion misclassification under poor lighting conditions or suboptimal face angles. This issue was particularly problematic for deployment in real-world vehicle environments where lighting conditions vary dramatically.

4.2 Solution Implementation

A comprehensive image processing solution was developed to address environmental challenges:

Advanced Image Processing Function:

- Tunable enhancement parameters for different environments
- Real-time frame optimization for improved detection accuracy
- Adaptive adjustment capabilities for changing conditions
- Integration with the main detection pipeline

Passenger-Specific Processing:

- `crop_and_enhance_passenger.py` script for targeted image enhancement
- Specialized processing for passenger camera feeds
- Optimized cropping algorithms for face region extraction
- Enhanced reliability in variable lighting conditions

The effectiveness of this solution is demonstrated through before-and-after comparisons showing significant improvement in detection accuracy across various lighting scenarios.

5 Testing Methodologies and Validation Framework

5.1 Comprehensive Testing Approach

The REMS project implemented a multi-faceted testing methodology to ensure robust performance validation:

Automated Logging System:

- Continuous emotion and state logging at regular intervals during live tests
- Automated timestamp generation for precise tracking
- Structured data format for analysis compatibility
- Integration with real-time detection pipeline

Manual Annotation Process:

- Frame-by-frame review of raw video footage
- Ground truth label generation for each person at each timestamp
- Direct comparison capability between predicted and actual emotions
- Support for quantitative analysis and metric calculation

5.2 Multi-Camera Testing Configuration

Dual-Camera Assessment: Testing with both front and back cameras provided insights into:

- System performance across different camera perspectives
- Environmental influence assessment from various angles
- Multi-passenger monitoring capabilities
- Perspective-dependent detection accuracy variations

Back-Camera Focus Testing: Concentrated testing with the primary passenger monitoring camera enabled:

- Focused performance metric derivation
- Primary system functionality validation
- Passenger-specific detection optimization
- Real-world deployment scenario simulation

5.3 Dual-Frame Video Recording

The implementation of side-by-side video recording provided significant advantages:

- Simultaneous raw and processed frame capture
- Direct visual comparison of detection results
- Enhanced validation capabilities for detection accuracy
- Improved debugging and system optimization support

6 Current System Performance Analysis

6.1 Overall Accuracy Metrics

The current REMS implementation demonstrates strong and consistent performance across different individuals:

Metric	Person 1	Person 2
Overall Accuracy	85%	84%
Combined Accuracy	85%	

Table 1: Overall System Accuracy by Individual

6.2 Per-Emotion Performance Analysis

Detailed analysis reveals varying performance across different emotional states:

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

Table 2: Per-Emotion Accuracy Results

6.3 Detailed Performance Metrics

Comprehensive evaluation using precision, recall, and F1-score metrics:

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

Table 3: Detailed Performance Metrics for Current System

Key Performance Insights:

- Exceptional performance for Happy, Sad, and Angry emotions
- Strong precision for Neutral emotion detection (0.99)
- Perfect recall for Angry and Fear emotions
- Room for improvement in Fear emotion precision
- Consistent performance across different individuals

7 Custom Deep Learning Model Development

7.1 Model Architecture Design

To further enhance system performance, a custom deep learning model was developed specifically for facial emotion recognition:

Base Architecture:

- **Foundation:** ResNet-18 as lightweight CNN backbone
- **Enhancement:** CBAM (Convolutional Block Attention Module) integration
- **Attention Mechanisms:** Both spatial and channel attention for feature focus
- **Efficiency:** Optimized for real-time deployment scenarios

Dataset and Preprocessing:

- **Dataset:** FER2013 with 35,887 images across 7 emotion categories
- **Emotions:** Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise
- **Preprocessing:** Grayscale to RGB conversion, 224×224 resizing
- **Normalization:** ImageNet mean and standard deviation

7.2 Training Strategy and Techniques

Multi-Stage Training Approach:

1. **Stage 1:** Classifier head training with frozen backbone
2. **Stage 2:** End-to-end fine-tuning of the complete model

Advanced Training Techniques:

- **Label Smoothing:** Prevention of overconfident predictions
- **Class Weights:** Handling of dataset class imbalance
- **Data Augmentation:** Albumentations library for diverse sample generation
- **Regularization:** Dropout and early stopping for overfitting prevention

Augmentation Strategy:

- Random horizontal flipping
- Brightness and contrast adjustments
- Gaussian blur application
- Random dropout for improved generalization

7.3 Custom Model Performance Results

Emotion	**Precision**	**Recall**	**F1-Score**
Angry	0.64	0.62	0.63
Disgust	0.58	0.69	0.63
Fear	0.57	0.52	0.55
Happy	0.91	0.87	0.89
Neutral	0.63	0.71	0.67
Sad	0.57	0.56	0.57
Surprise	0.79	0.84	0.82
Overall Accuracy		**70.0%**	
Macro F1-Score		**0.68**	

Table 4: Custom Model Performance Metrics

Key Achievements:

- 70% overall accuracy on FER2013 test set
- Exceptional performance for positive emotions (Happy: F1=0.89, Surprise: F1=0.82)
- Successful integration of attention mechanisms
- Effective handling of class imbalance through weighted loss
- Stable training without overfitting

7.4 Challenges and Solutions

Technical Challenges Addressed:

Class Imbalance:

- **Problem:** Uneven distribution of emotion samples
- **Solution:** Implementation of class weights and label smoothing

Overfitting Risk:

- **Problem:** Small dataset size leading to memorization
- **Solution:** Data augmentation, dropout, and early stopping

Emotion Similarity:

- **Problem:** Confusion between similar expressions (sad vs. neutral)
- **Solution:** CBAM attention mechanisms and temporal smoothing

8 System Scripts and Implementation Details

8.1 Core System Components

The REMS implementation consists of several specialized Python scripts, each designed for specific functionality:

Primary Detection Module:

- `face_detection.py`: Real-time face detection, landmark extraction, and drowsiness/emotion analysis using single camera feed

Multi-Camera Recording System:

- `record_side_by_side.py`: Dual ZED stereo camera recording with timestamp overlays, face detection, emotion analysis, and unique filename generation

Image Enhancement Pipeline:

- `crop_and_enhance_passenger.py`: Specialized passenger camera image processing with cropping and enhancement for improved low-light detection

Analysis and Logging Tools:

- `side_by_side_with_log.py`: Dual-frame video output (raw and processed) with automated emotion logging every 10 seconds
- `advanced_log_analysis.py`: Comprehensive automated log analysis, metric calculation, and performance visualization generation

8.2 Automated Analysis Framework

The automated analysis system provides comprehensive performance evaluation:

Data Processing Capabilities:

- Automated loading and cleaning of all log files
- Computation of accuracy, precision, recall, and F1-score metrics
- Generation of multiple performance visualization types
- Support for multi-person analysis scenarios

Visualization Generation:

- Error breakdown analysis by emotion
- True vs. predicted emotion timeline tracking
- Misclassification timestamp identification
- Per-emotion precision/recall/F1 bar charts
- Emotion distribution frequency analysis
- Special state frequency monitoring
- Multi-person prediction agreement heatmaps

9 Results and Performance Evaluation

9.1 Quantitative Performance Analysis

The comprehensive testing framework has enabled detailed quantitative evaluation of system performance across multiple dimensions:

Overall System Metrics:

- **Average Accuracy:** 85% across both test subjects
- **Consistency:** ¡1% variation between individuals
- **Reliability:** Stable performance across multiple testing sessions
- **Real-time Performance:** Maintained accuracy under live conditions

Emotion-Specific Performance:

- **Highest Accuracy:** Sad (96%), Angry (95%), Fear (96%)
- **Most Challenging:** Neutral (76%) - room for improvement
- **Best F1-Score:** Angry (0.96), Happy (0.90)
- **Precision Leader:** Neutral (0.99) despite accuracy challenges

9.2 Qualitative Improvements

Beyond quantitative metrics, the system demonstrates significant qualitative enhancements:

Robustness Improvements:

- Enhanced performance with eyewear and facial accessories
- Improved detection under varying lighting conditions
- Better handling of head pose variations and partial occlusions
- Stable multi-face tracking in group passenger scenarios

User Experience Enhancements:

- Real-time visual feedback with clear emotion labeling
- Immediate warning systems for safety-critical states
- Smooth, non-flickering detection displays
- Comprehensive logging for post-trip analysis

10 Future Work and Recommendations

10.1 Short-term Development Goals

Model Integration and Comparison:

- Deploy custom ResNet-CBAM model in real-time environment
- Conduct side-by-side comparison with current DeepFace implementation
- Evaluate performance trade-offs between accuracy and computational efficiency
- Optimize model for embedded deployment scenarios

Real-time Performance Optimization:

- Implement TorchScript or ONNX optimization for custom model
- Develop temporal smoothing algorithms for prediction stability
- Create ensemble methods combining multiple model predictions
- Optimize processing pipeline for reduced latency

10.2 Medium-term Enhancement Opportunities

Multi-modal Sensor Integration:

- Integration of heart rate monitoring sensors
- Addition of steering input analysis for driver monitoring
- Implementation of environmental sensor data (temperature, humidity)
- Development of sensor fusion algorithms for improved accuracy

Advanced Detection Capabilities:

- Implementation of subtle emotion detection algorithms
- Development of long-term trend analysis capabilities
- Creation of personalized baseline establishment systems
- Integration of behavioral pattern recognition

10.3 Long-term Vision and Deployment

Real-world Deployment Strategy:

- Field testing in actual autonomous vehicle environments
- Collection of diverse demographic and environmental data
- Development of deployment-specific optimization strategies
- Creation of maintenance and update protocols

System Scalability:

- Development of cloud-based analysis capabilities
- Implementation of fleet-wide monitoring systems
- Creation of centralized data collection and analysis platforms
- Development of privacy-preserving analysis techniques

11 Technical Specifications and Requirements

11.1 Hardware Requirements

Camera Systems:

- ZED Stereo Cameras for high-quality depth perception
- Minimum 720p resolution at 30fps
- Wide-angle lens coverage for passenger area monitoring

- Low-light performance optimization

Computing Platform:

- NVIDIA GPU with CUDA support for deep learning inference
- Minimum 8GB RAM for real-time processing
- SSD storage for fast data access and logging
- Multi-core CPU for parallel processing tasks

11.2 Software Dependencies

Core Libraries:

- OpenCV for computer vision operations
- MediaPipe for facial landmark detection
- DeepFace for emotion classification
- PyTorch for custom model implementation
- NumPy for numerical computations

Additional Components:

- Albumentations for data augmentation
- Matplotlib/Seaborn for visualization
- Pandas for data analysis
- Scikit-learn for metric calculations

12 Conclusion

The Real-Time Emotion Monitoring System (REMS) project has successfully achieved its primary objectives of creating a robust, accurate, and deployable solution for passenger monitoring in autonomous vehicles. Through five weeks of intensive development, the system has evolved from a basic emotion detection implementation to a comprehensive monitoring platform capable of real-time analysis with 85% accuracy.

12.1 Key Achievements

Technical Accomplishments:

- Development of a highly accurate emotion detection system (85% overall accuracy)
- Implementation of robust drowsiness and fatigue detection algorithms
- Creation of advanced image processing techniques for challenging environments
- Successful integration of MediaPipe Face Mesh for improved landmark detection

- Development of a custom ResNet-CBAM model achieving 70% accuracy on FER2013

System Integration Successes:

- Seamless multi-camera integration with ZED stereo systems
- Real-time processing capabilities maintaining performance under load
- Comprehensive logging and analysis framework for continuous improvement
- Modular architecture enabling easy maintenance and updates
- Robust performance across diverse lighting and environmental conditions

12.2 Impact and Significance

The REMS system addresses critical safety needs in autonomous vehicle deployment, particularly for vulnerable populations such as elderly passengers in retirement communities. The system's ability to detect emotional distress, fatigue, and other concerning states provides a crucial safety net that can trigger appropriate interventions when needed.

The project's emphasis on robustness, accuracy, and real-time performance makes it well-suited for immediate deployment in controlled environments, with clear pathways for expansion to broader autonomous vehicle applications.

12.3 Research and Development Contributions

Beyond its immediate practical applications, the REMS project has contributed to the broader field of computer vision and emotion recognition through:

- Demonstration of effective multi-modal emotion detection in challenging environments
- Development of novel image processing techniques for automotive applications
- Creation of comprehensive testing and validation methodologies
- Integration of attention mechanisms in custom deep learning architectures
- Establishment of performance benchmarks for real-time emotion monitoring systems

The project serves as a foundation for future research in automotive safety systems and demonstrates the potential for computer vision technologies to enhance passenger safety and experience in autonomous vehicles.

Through careful engineering, thorough testing, and continuous refinement, the REMS project has created a system that not only meets its technical objectives but also provides a robust platform for future enhancements and real-world deployment. The combination of high accuracy, real-time performance, and environmental robustness positions REMS as a valuable contribution to the field of autonomous vehicle safety systems.