PROJECT SYNOPSIS (AIDS 451)

On

MULTIMODAL FUSION & ALERT SYSTEM

Submitted for partial fulfilment of award of the degree of

Bachelor of Technology

In

Artificial Intelligence and Machine Learning

Submitted by

Taniya Dixit – 35418011922

Under the Guidance of

Guide Name : Dr. Upasana Joshi

Designation: HOD (AI)

Affiliated to GGSIP University, New Delhi
Approved by AICTE & Council of Architecture

Department of Artificial Intelligence

DELHI TECHNICAL CAMPUS, GREATER NOIDA

(Affiliated Guru Gobind Singh Indraprastha University, New Delhi)

Session 2025-2026 (ODD SEM)

CONTENTS

ABSTRACT

Road accidents caused by driver distraction, drowsiness, and unsafe driving conditions are a growing concern worldwide. The project *“Multimodal Fusion & Rule-Based Alert System”* aims to enhance driver safety by integrating data from multiple sources—**vision, audio, and sensors**—to provide real-time alerts.

The **vision module** uses cameras with **YOLO object detection** to identify vehicles, pedestrians, traffic signs, and monitor driver attention. The **audio module** captures environmental sounds such as horns, sirens, and yawning, classifying them to detect potential risks. The **sensor module** collects vehicle speed, GPS location, steering grip, and environmental conditions like rain via OBD-II or embedded sensors.

All data are combined in a **fusion engine** to create a unified representation of the driving environment. A **rule-based system** then evaluates the fused data using IF–THEN logic to detect hazards such as collisions, drowsiness, or traffic violations. Alerts are delivered in **real-time** through audio, visual, and haptic feedback to warn the driver, while all events are logged for analysis.

This system demonstrates a **practical, modular, and scalable approach** to real-time driver assistance, reducing accident risk and enabling future improvements such as deep learning integration, cloud monitoring, and autonomous vehicle adaptation.

Keywords: *Multimodal Fusion* (combining vision, audio, and sensor data), *Rule-Based Alert System* (IF–THEN logic for safety alerts), *YOLO Object Detection* (detecting vehicles, pedestrians, traffic signs), *Driver Drowsiness Monitoring* (detecting tired drivers), *Real-Time Alerts* (instant audio, visual, haptic warnings), *Traffic Violation Detection* (identifying unsafe driving and signal violations).

1.) INTRODUCTION

Road safety has become a critical concern due to the increasing number of accidents caused by driver distraction, drowsiness, and environmental factors. Traditional safety systems often rely on a single source of information, such as cameras or sensors, which limits their effectiveness in detecting complex driving risks. To address this limitation, multimodal systems combine data from multiple sources—vision, audio, and sensors—to provide a more comprehensive understanding of the driving environment and driver behavior.

In this project, a Multimodal Fusion & Rule-Based Alert System is designed to enhance vehicle safety by integrating these diverse inputs. The vision module captures visual information from cameras to detect vehicles, pedestrians, traffic signs, and driver attention. The audio module processes environmental sounds such as horns, sirens, or yawning to detect potential risks. The sensor module gathers data on vehicle speed, GPS location, steering, and environmental conditions like rain.

The system employs a fusion engine to combine these inputs into a structured format, followed by a rule-based alerting mechanism that applies IF–THEN logic to identify hazardous situations. When a risk is detected, the system provides real-time alerts through audio, visual, and haptic feedback to warn the driver and reduce the likelihood of accidents.

This project demonstrates a practical approach to real-time driver assistance, offering proactive warnings and the potential for future upgrades in smart vehicles and autonomous systems.

2) PROBLEM STATEMENT

For building a multimodal driver alert system from scratch, we face several challenges. Currently, most vehicle safety systems rely on a **single type of data**, such as cameras or sensors, which limits their ability to accurately detect risky situations. To design a robust system, we need to integrate **vision, audio, and sensor data**, but this raises multiple questions about **data collection, fusion, and alerting**.

For example, the vision module may detect pedestrians or vehicles, but it may fail in poor lighting or occlusion. The audio module can detect horns or sirens, but background noise can interfere. Sensor data like speed, GPS, or environmental conditions may vary across vehicles. Moreover, combining these heterogeneous data sources requires a **fusion strategy** that maintains accuracy and supports real-time processing.

Q. How can the system detect hazards accurately when one of the modalities fails?
Q. How should the data from different modules be fused to make reliable decisions?
Q. How can we define rules that trigger real-time alerts effectively for collisions, drowsiness, and traffic violations?
Q. How can the system ensure low latency so that alerts reach the driver immediately?

3) OBJECTIVES

The main objective of this project is to develop a Multimodal Fusion & Rule-Based Alert System that can enhance driver safety by integrating data from multiple sources and providing real-time alerts. To achieve this, the project focuses on several sub-goals:

- **Data Collection:** Collect and organize data from vision (camera), audio (microphone), and sensor (OBD-II, GPS, environmental sensors) modules to capture comprehensive information about the vehicle and its surroundings.

- **Data Preprocessing:** Process the collected data to remove noise, normalize sensor readings, and extract relevant features for accurate analysis.

- **Multimodal Fusion:** Combine information from all modules into a unified format to create a holistic representation of the driving environment and driver behavior.

- **Rule-Based Alert System:** Design and implement IF–THEN rules to detect hazards such as collisions, driver drowsiness, traffic violations, and unsafe driving conditions.

- **Real-Time Alerts:** Generate audio, visual, and haptic feedback to warn the driver promptly and prevent accidents.

- **Logging & Analysis:** Maintain a **system log** of events for performance evaluation, future improvements, and possible integration with fleet management or smart vehicle systems.

4) FEASIBILITY STUDY

The feasibility study evaluates the practicality of developing the **Multimodal Fusion & Rule-Based Alert System** in terms of technical, operational, and economic aspects.

1. Technical Feasibility

- The project uses **Python**, **YOLO (v8/v11)** for object detection, **OpenCV** for video processing, and **Librosa/PyAudio** for audio analysis, all of which are well-supported and easy to implement on standard CPUs.
- Data fusion and rule-based alerting are computationally lightweight, making real-time implementation feasible on **CPU-based systems** without requiring high-end GPUs.
- Sensor integration (OBD-II, GPS, environmental sensors) is compatible with standard vehicles, and software libraries like **python-OBD** enable easy interfacing.

2. Operational Feasibility

- The system provides **real-time alerts** through audio, visual, and haptic feedback, ensuring driver safety.
- Modular design allows easy **upgrade or replacement** of any module (vision, audio, or sensor).
- Training and testing can be conducted with publicly available datasets (COCO, KITTI, UrbanSound8K) and small-scale custom datasets.

3. Economic Feasibility

- Minimal hardware cost: standard webcam, microphone, OBD-II adapter, and basic sensors.
- No expensive GPU required for CPU-focused implementation.
- Software tools are **open-source** or freely available, reducing licensing costs.

5) NEED AND SIGNIFICANCE

Road safety is a major global concern, as the number of accidents caused by driver distraction, drowsiness, and unsafe driving conditions continues to rise. Despite advances in vehicle safety systems, most existing solutions rely on a single type of data, such as cameras or sensors, which limits their ability to detect complex hazards accurately. For instance, a camera-based system may fail in poor lighting, while a sensor-only system may not detect pedestrians or sudden environmental changes. This highlights the need for a comprehensive, real-time system that can integrate multiple sources of information and provide timely alerts to prevent accidents.

The Multimodal Fusion & Rule-Based Alert System addresses this need by combining vision, audio, and sensor data to create a holistic understanding of the driving environment and driver behavior. By employing a fusion engine and rule-based alerting mechanism, the system can detect hazards such as collisions, driver drowsiness, and traffic violations accurately and provide real-time feedback through audio, visual, and haptic signals.

The significance of this project lies in its potential to enhance driver safety, reduce accidents, and provide a scalable solution for modern vehicles. Its modular design allows future integration with smart vehicles, fleet management systems, and autonomous driving technologies. Additionally, the system logs events for performance evaluation and research, making it a valuable tool for improving intelligent transportation systems. Moreover, using CPU-based processing and readily available sensors makes the system economically feasible for widespread adoption, enabling practical implementation without high-cost hardware.

In conclusion, the project not only fulfills a critical need for proactive driver assistance but also offers a significant contribution to road safety, vehicle intelligence, and accident prevention, making it a highly relevant and impactful solution in the field of intelligent transportation systems.

6) METHODOLOGY

The methodology for the Multimodal Fusion & Rule-Based Alert System involves a approach that integrates vision, audio, and sensor data to generate real-time alerts for drivers. The project is divided into the following steps:

1. Data Acquisition

- Vision Module**:** Capture images and video using cameras to detect vehicles, pedestrians, traffic signs, and driver attention.
- Audio Module: Record environmental sounds such as horns, sirens, and yawning through microphones.
- Sensor Module: Collect vehicle and environmental data, including speed, GPS coordinates, steering grip, and rain detection, using OBD-II adapters or other sensors.

2. Data Preprocessing

- Remove noise and irrelevant information from vision, audio, and sensor data.
- Normalize sensor readings and extract relevant features (e.g., bounding boxes for objects, MFCC features for audio, speed thresholds).
- Annotate custom datasets using tools like CVAT for images and Audacity for audio.

3. Multimodal Data Fusion

- Combine heterogeneous data from all modules into a unified format for analysis.
- Use timestamp synchronization to align events across vision, audio, and sensor streams.
- Generate a **holistic representation** of the driving environment and driver behavior.

4. Rule-Based Alert System

- Define **IF–THEN rules** to detect potential hazards, such as:
  - IF driver eyes closed > 2 seconds → trigger drowsiness alert
  - IF vehicle speed > limit AND proximity < safe distance → collision alert
  - IF horn detected nearby → attention alert
- Evaluate fused data against these rules to generate **real-time alerts**.

5. Alert Generation

- Provide **audio alerts** through speakers.
- Show **visual alerts** on the dashboard or HUD.

- Use **haptic feedback** via vibration motors for critical warnings.

6. Logging and Analysis

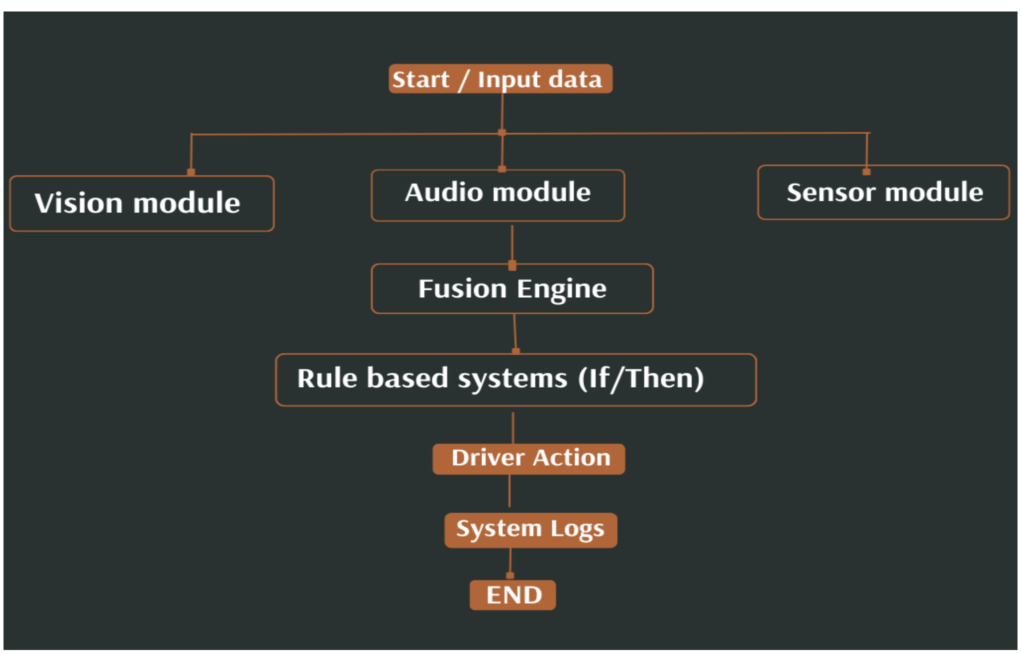
- Record all events and alerts for performance evaluation.
- Use logged data to refine rules, improve accuracy, and support research in intelligent transportation.

7) INTENDED USERS

The Multimodal Fusion & Rule-Based Alert System is designed for a wide range of users who require enhanced driving safety and real-time hazard detection:

- Individual Drivers: Personal vehicle owners seeking proactive safety alerts to prevent accidents.
- Commercial Vehicle Drivers: Taxi, bus, and truck drivers who operate for long hours and are at risk of fatigue or drowsiness.
- Fleet Management Companies: Organizations that monitor multiple vehicles and want real-time alerts to reduce accidents and improve driver behavior.
- Automobile Manufacturers: For integration into Advanced Driver Assistance Systems (ADAS) in new vehicles.
- Researchers and Developers: For studying multimodal fusion, driver behavior monitoring, and intelligent transportation systems.

8) DIAGRAM : FLOWCHART



GANTT CHART

phase	Week 1-2	Week 3-4	Week 5-6	Week 7-8	Week 9-10
Requirement analysis					
Dataset collection					
Pre processing					
Model development (yolo					
Rule based system design					
Integration and fusion					
Testing and evaluation					

9) ABBREVIATIONS AND ACRONYMS

Abbreviation / Acronym	Full Form / Meaning
ADAS	Advanced Driver Assistance Systems
OBD-II	On-Board Diagnostics, Second Generation
GPS	Global Positioning System

Abbreviation / Acronym	Full Form / Meaning
YOLO	You Only Look Once (Object Detection Model)
CVAT	Computer Vision Annotation Tool
MFCC	Mel-Frequency Cepstral Coefficients (Audio Feature)
CPU	Central Processing Unit
HUD	Head-Up Display
ADAS	Advanced Driver Assistance System
API	Application Programming Interface
FPS	Frames Per Second
ML	Machine Learning
DL	Deep Learning

10) LITERATURE REVIEW

The rise in road accidents and unsafe driving behavior has prompted extensive research on driver assistance systems and real-time alert mechanisms. Traditional approaches often rely on a single modality, such as vision-based or sensor-based systems, which limits their effectiveness in complex driving scenarios.

- **Vision-Based Systems**

Several studies have used **object detection models** like YOLO (You Only Look Once) for vehicle and pedestrian detection. Bochkovskiy et al. (2020) demonstrated YOLOv4's efficiency in real-time object detection, providing high accuracy with low computational cost. Vision-based systems are effective in detecting road obstacles, traffic signs, and driver attention, but performance can degrade under poor lighting or occlusion conditions.

- **Audio-Based Systems**

Audio cues such as horns, sirens, and driver sounds (yawning, speech) have been used to detect risky situations. UrbanSound8K dataset (Salamon & Bello, 2014) is widely used for environmental sound classification. Audio systems complement vision-based modules by providing **non-visual cues**, but they face challenges from background noise and overlapping sounds.

- **Sensor-Based Systems**

Vehicle sensors (speed, GPS, steering grip) and environmental sensors are used to monitor driving conditions. Integration with OBD-II allows real-time data logging and detection of unsafe driving practices. Sensor-based systems are robust but cannot detect visual hazards like pedestrians or traffic signs.

- **Multimodal Fusion Approaches**

Recent research emphasizes **combining multiple modalities** to improve accuracy. Fusion of vision, audio, and sensor data enables a comprehensive understanding of the driving environment and driver behavior. Multimodal systems can leverage the strengths of each module, mitigating the weaknesses of single-modality systems.

- **Rule-Based Alert Systems**

Rule-based systems use **IF–THEN logic** to generate alerts when certain thresholds are crossed (e.g., drowsiness, overspeeding, close proximity to obstacles). Redmon & Farhadi (2018) and Ultralytics YOLOv8 demonstrate real-time object detection, which can be integrated with rule-based logic for immediate hazard detection.

Gaps Identified

- Single-modality systems are prone to failure under adverse conditions.
- Existing multimodal systems are often GPU-dependent, limiting low-cost CPU implementations.
- Integration of real-time fusion, rule-based alerts, and low-cost hardware remains a challenge.

11) SOFTWARE REQUIREMENTS

Front End Purpose: User interface for displaying alerts, logs, and system status.

Tools/Technologies:

- Tkinter / PyQt: For GUI in Python to show visual alerts and dashboard.
- HTML, CSS, JavaScript (optional): For web-based dashboards if needed.
- Matplotlib / Seaborn: For plotting sensor data or alert statistics.

Back End Purpose: Handles data processing, multimodal fusion, rule evaluation, and alert generation.

Tools/Technologies:

- Python: Main programming language for integrating modules.

- OpenCV: For vision processing and object detection (YOLO).
- PyTorch : For running object detection models.
- Librosa / PyAudio: For audio feature extraction and sound classification.
- Python-OBD / GPS libraries: For sensor data acquisition.
- CSV: For storing logs and event records.
- Threading / Multiprocessing: To enable real-time processing.

OTHER Tools:

- VS Code / Jupyter Notebook: For development and testing.
- Git: Version control for code management.

12) FUTURE SCOPE AND APPLICATIONS

Future Scope

- Integration with Advanced Driver Assistance Systems (ADAS): The system can be expanded to work alongside lane departure warnings, adaptive cruise control, and automatic braking.
- Autonomous Vehicles: Multimodal fusion can enhance decision-making for self-driving cars, improving obstacle detection and driver safety.
- Cloud-Based Fleet Monitoring: Real-time data from multiple vehicles can be analyzed in the cloud for predictive maintenance and safety analytics.
- Deep Learning Enhancements: Future versions can incorporate deep learning models for predictive hazard detection, improving accuracy and adaptability.
- Wearable and IoT Integration: Integration with smartwatches, wearables, or vehicle IoT devices to monitor driver vitals (heart rate, fatigue) for more proactive safety.

Applications

- Personal Vehicles: Real-time alerts to prevent accidents and improve driver awareness.
- Commercial Vehicles & Fleet Management: Monitor multiple drivers, reduce accident risks, and ensure regulatory compliance.
- Public Transport Safety: Buses and taxis can use the system to monitor driver attention and road conditions.
- Smart City & Traffic Management: Data from multiple vehicles can be used for traffic flow optimization, accident analysis, and road safety planning.

- Insurance & Risk Management: Logged data can support insurance claims, driver scoring, and personalized insurance plans.

CONCLUSION

The proposed Multimodal Fusion & Rule-Based Alert System combines vision, audio, and sensor data to provide a reliable and real-time safety mechanism for drivers. By integrating multiple modalities, the system overcomes the limitations of single-source monitoring and ensures higher accuracy in detecting hazards such as drowsiness, distractions, overspeeding, and collision risks.

The use of a rule-based alerting mechanism ensures quick, interpretable, and low-cost implementation, making it suitable for CPU-based systems without requiring expensive hardware. This makes the solution practical for personal vehicles, commercial fleets, and public transport systems.

Ultimately, the project contributes to reducing road accidents, enhancing driver awareness, and paving the way for intelligent transportation systems. It also opens future opportunities for integration with ADAS, autonomous driving technologies, and smart city infrastructure.

REFERENCES

- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection.* arXiv:2004.10934.
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement.* arXiv:1804.02767.
- Ultralytics (2023). *YOLOv8: State-of-the-art Real-Time Object Detection.* Available at: https://docs.ultralytics.com
- Salamon, J., & Bello, J. P. (2014). *UrbanSound8K: A Dataset for Urban Sound Research.* Proceedings of the 22nd ACM International Conference on Multimedia.
- National Highway Traffic Safety Administration (NHTSA). (2021). *Traffic Safety Facts: Overview of Motor Vehicle Crashes.* U.S. Department of Transportation.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.
- Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). *DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving.* IEEE International Conference on Computer Vision (ICCV).
- Sivaraman, S., & Trivedi, M. M. (2013). *Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis.* IEEE Transactions on Intelligent Transportation Systems.
- Abdi, H., & Williams, L. J. (2010). *Principal Component Analysis.* Wiley Interdisciplinary Reviews: Computational Statistics.