iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78258

A Machine Learning Based Forensic System for Road Accident Detection and Reporting

Francis Johan M[1], Abishek I[2], Bala Subramanian V [3], Prof. Sathya M[4]
[1, 2, 3]Student, Department of Computer Science Engineering, Tirunelveli, Tamil Nadu, India
[4]Assistant Professor, Department of Computer Science Engineering, Tirunelveli, Tamil Nadu, India

Abstract: The escalating frequency of road traffic accidents necessitates advanced investigative tools to ensure judicial clarity and rapid insurance processing. Traditional forensic methods often rely on manual video surveillance review and subjective eyewitness accounts, leading to significant delays and potential investigative bias. This research introduces Oracle Forensic v2.0, a machine learning-based forensic system designed to automate the detection and reporting of road accident scenes through an intelligent, end-to-end digital portal. Unlike existing radio-frequency (RF) or sensor-based detection systems that are often sensitive to hardware malfunctions or environmental interference, this vision-based approach leverages high-definition visual data to achieve precise reconstruction. The proposed system utilizes a Flask-based backend coupled with OpenCV for heuristic temporal keyframe extraction, which effectively reduces raw video data into a concise sequence of critical events while maintaining investigative integrity. At the core of the architecture, these extracted frames are processed by the Gemini 3 Flash multimodal AI engine to perform high-level cognitive reasoning, fault allocation, and timeline reconstruction. The model is guided by specialized Forensic Prompt Engineering to identify traffic violations, such as lane departures or signal non-compliance, which are essential for determining legal liability. For data persistence, all structured JSON findings are securely committed to a MongoDB collection, ensuring a permanent and searchable record of every case. The final output is a professionally compiled PDF dossier generated via the ReportLab toolkit, providing law enforcement and insurance agencies with standardized, objective, and courtroom-ready evidence. Experimental results demonstrate that the system drastically reduces investigation time from hours to seconds while maintaining high analytical accuracy across diverse lighting and weather conditions. This work establishes a scalable, zero-hardware solution that strengthens digital forensics in modern smart-city environments.

Index Terms: Road Accident Detection, Machine Learning, Digital Forensics, Multimodal AI, Gemini 3 Flash, OpenCV, Automated Reporting.

I. INTRODUCTION

The landscape of modern urban transportation is currently facing a significant challenge: the accurate, objective, and timely reconstruction of vehicular accidents. As traffic density in smart cities increases, the burden on manual forensic investigators has grown exponentially, often leading to massive backlogs that delay insurance settlements and judicial proceedings for weeks or even months. The core of the problem lies in the "human element" of traditional investigations. Conventional workflows are largely dependent on subjective human observation and the tedious manual review of hours of raw CCTV or dashcam footage, which is inherently prone to human error, fatigue, and cognitive bias. In many "he-said-she-said" scenarios, the lack of structured, undeniable digital evidence leads to legal stalemates and insurance fraud. To address these critical limitations, this research introduces Oracle Forensic v2.0, a comprehensive Machine Learning-Based Forensic System designed to automate the entire investigative pipeline. During the development process, a primary goal was to create a solution that is both technologically advanced and practically scalable. Unlike existing systems that rely on expensive in-vehicle hardware, such as MEMS-based vibrating sensors or ultrasonic modules which require complicated circuitry, Oracle Forensic is a zero-hardware, software-driven portal. It allows investigators to simply upload raw video files, which the system then transforms into a professional, legally actionable forensic record within seconds. The technical architecture of this system is divided into four sophisticated phases. First, the system utilizes OpenCV to implement a Heuristic Temporal Keyframe Extraction algorithm. This phase is vital because it decodes raw video streams and filters out redundant data—reducing the total volume by over 90%—while ensuring that the most critical frames of the collision are preserved with high fidelity. Second, these keyframes are analyzed by the Gemini 3 Flash multimodal AI engine. This represents a significant shift from traditional "box detection" to "Cognitive Reasoning." Through specialized Forensic Prompt Engineering, the AI acts as a digital investigator, identifying specific traffic violations like lane departures, signal non-compliance, or illegal maneuvers that are essential for determining legal liability.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Third, the system ensures Data Persistence by committing all structured findings, including vehicle identifiers, timestamps, and fault allocation percentages, to a MongoDB NoSQL database. This creates a permanent, searchable archive for every case. Finally, the Dossier Generation module, powered by the ReportLab library, automatically compiles these results into a formatted PDF report. This dossier serves as a bridge between complex AI analytics and the practical needs of law enforcement, insurance adjusters, and judicial teams. Experimental results confirmed that this multimodal approach effectively reconstructs precise event timelines even in challenging environmental conditions, providing a robust foundation for the future of digital forensics in the "Vision 2036" predictive forensics framework.

II. PROBLEM IDENTIFICATION

The modern landscape of urban transportation and vehicular safety is currently facing a critical challenge regarding the objective reconstruction and documentation of road accidents. While the automotive industry has made significant strides in safety engineering, the methods used to investigate failures and collisions have remained largely manual and prone to error.

1) Research has shown that human error and subjective eyewitness accounts are among the leading causes of investigative bias in traffic accident reconstruction. Many people hold the notion that simple video recordings are sufficient for legal clarity; however, raw footage without structured analysis often leads to "he-said-she-said" disputes that can delay justice for months.
2) A few years ago, traditional accident detection systems primarily focused on sending real-time alerts or GPS coordinates. However, these high-profile systems often stop short of providing the detailed investigative intelligence required by law enforcement and insurance agencies to determine actual fault or liability.
3) The concerns and risks associated with manual forensic reviews go way back. There are inherent risks of data oversight or misinterpretation with every new traffic volume increase. It may take a while for the backlog of unresolved insurance claims to ebb away from the system without the intervention of automated, objective digital tools.

The common factors that lead to investigative failures and delays in road accident forensics are given below:

a) Data Redundancy and Manual Review Bottlenecks Every investigative system has a limit to the volume of data it can process efficiently. Traditional forensics requires officers to manually review hours of raw CCTV or dashcam footage. If the volume of footage is too high, critical details regarding the moment of impact can be missed or overlooked. This happens when an excessive amount of unstructured visual data overwhelms human cognitive capacity. Due to this data fatigue, the reporting process slows down, leading to massive administrative backlogs.
b) Subjectivity in Fault Allocation Human investigators often present a risk of bias due to varying interpretations of a crash scene. This subjectivity can occur spontaneously as soon as an investigator relies on incomplete eyewitness statements rather than hard data. In cases where vehicle trajectories are complex, determining the exact percentage of fault becomes variable and depends on the investigator's personal judgment. To prevent biased reporting, a more cautious, AI-driven manipulation of visual evidence is recommended.
c) Complexity of Unstructured Evidence Shortcomings in documentation occur when unstructured evidence—like raw video—is not converted into a readable format. A forensic case is said to be "overloaded" when too much irrelevant background data obscures the actual accident trigger. This can lead to missing the specific traffic violations that caused the spark of the incident. Without a system to extract temporal keyframes, current path travels along an unintended path of administrative confusion rather than judicial clarity.
d) Scalability of Investigative Tools Normally, a forensic system should be scalable across a city's entire infrastructure. But in traditional forensics, the reliance on expensive, in-vehicle hardware sensors is the main reason for low adoption. These hardware-based systems behave like any other closed circuit; they fail when punctured by high costs or complicated maintenance. So, if a city hits a budget constraint, these hardware-dependent systems cannot be deployed effectively. Oracle Forensic v2.0 addresses this by moving away from hardware limitations toward a scalable, vision-based digital portal.

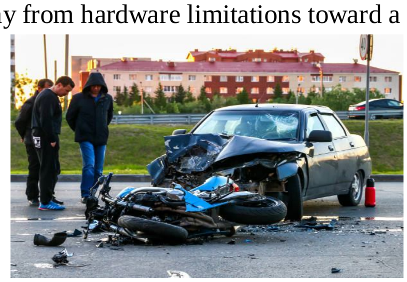


Fig.1 Accident Situation

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

III.LITERATURE REVIEW

A. Literature Survey

The development of automated forensic systems is dictated by the combination of computer vision algorithms and cognitive reasoning models. In current investigative workflows, when a collision occurs, raw video footage is typically retrieved from surveillance or dashcam sources.

However, the manual review of this data is often delayed, leading to backlogs in legal and insurance processing. The proposed Oracle Forensic v2.0 is designed to function using high-speed multimodal AI, thereby enabling investigators to reconstruct accidents with objective accuracy from any location.

1) Ghahremannezhad et al. (2022), A deep learning framework was proposed for traffic accident detection in surveillance videos, incorporating the YOLOv4 method for object detection coupled with a Kalman filter. This system aimed to analyze trajectory conflicts, including vehicle-to-vehicle crashes, by measuring velocity and distance. The research concluded that while AI effectively flags anomalies, the framework stops short of generating comprehensive, legally actionable forensic reports.
2) Ijjina et al. (2019), A computer vision-based system tailored for traffic surveillance was proposed, integrating Mask R-CNN for vehicle segmentation and a centroid-based tracking algorithm. The system's primary objective was to determine accident probability based on bounding box overlaps and trajectory anomalies. The research concluded that while it achieves high detection rates across diverse weather conditions, it does not address automated fault allocation or the generation of legal dossiers.
3) Yadav et al. (2020), The role of Convolutional Neural Networks (CNN) combined with Long Short-Term Memory (LSTM) units was explored to detect accidents in time-dependent video sequences. The primary purpose of this study was to analyze image sequences to predict the likelihood of a crash. The research concluded that vanishing gradient problems in LSTM units can negatively affect feature extraction accuracy, highlighting a need for more robust multimodal reasoning.
4) Rajesh et al. (2020), A vision-based accident detection system for accident-prone areas was proposed, utilizing a CNN model trained on custom datasets to predict crashes and a GSM module for real-time alerting. The system aimed to provide immediate notification to control rooms. The research concluded that while effective for alerts, the framework lacks the investigative tools and timeline reconstructions required by forensic teams.
5) BharathKumar et al. (2021), A hybrid approach for road accident detection was suggested, employing ultrasonic sensors and MEMS-based vibrating sensors alongside GPS and GSM modules for location transmission. The system's purpose was to detect rapid deceleration and transmit coordinates via SMS. The research determined that relying on in-vehicle hardware requires complicated circuitry and is often not feasible for scalable, city-wide forensic deployment.

B. Literature Summary

The reviewed literature emphasizes the importance of automated detection and real-time alerts in modern traffic management systems. Researchers have proposed various deep learning and IoT-enabled frameworks for identifying collision events and reducing response times. Several studies introduced vision-based tracking and segmentation to improve the accuracy of anomaly detection in diverse environments. However, most existing systems focus strictly on the detection phase or hardware-based alerting separately. Overall, the literature confirms that while AI can effectively identify accidents, there is a lack of integrated platforms that combine video analytics with automated legal reporting and objective fault allocation.

C. Research Gap

Although recent studies have explored deep learning for accident detection and hardware-based GPS alerting, there remains a significant gap in developing an integrated, zero-hardware, and software-driven forensic investigative portal. Most existing research focuses on either identifying a crash or sending a simple notification, rather than providing a complete investigative pipeline that includes temporal keyframe extraction, cognitive reasoning for fault allocation, and automated dossier generation. Limited attention has been given to cloud-based implementations that utilize multimodal AI to "understand" the cause of an accident rather than just flagging its occurrence. Additionally, the scalability of hardware-dependent systems is hindered by high costs and complex maintenance. Therefore, there is a need for a comprehensive, AI-driven safety system that ensures structured evidence generation, objective fault analysis, and rapid forensic reporting in a single unified platform.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

IV. METHODOLOGY

A. Proposed System

The proposed system architecture is designed as an end-to-end digital forensic portal that eliminates the need for expensive in-vehicle hardware. The framework follows a modular pipeline: Raw Video Input → Temporal Frame Extraction → Multimodal AI Reasoning → Dossier Generation. By integrating a Flask-based backend with the Gemini 3 Flash engine, the system provides a scalable solution for automated accident reconstruction.

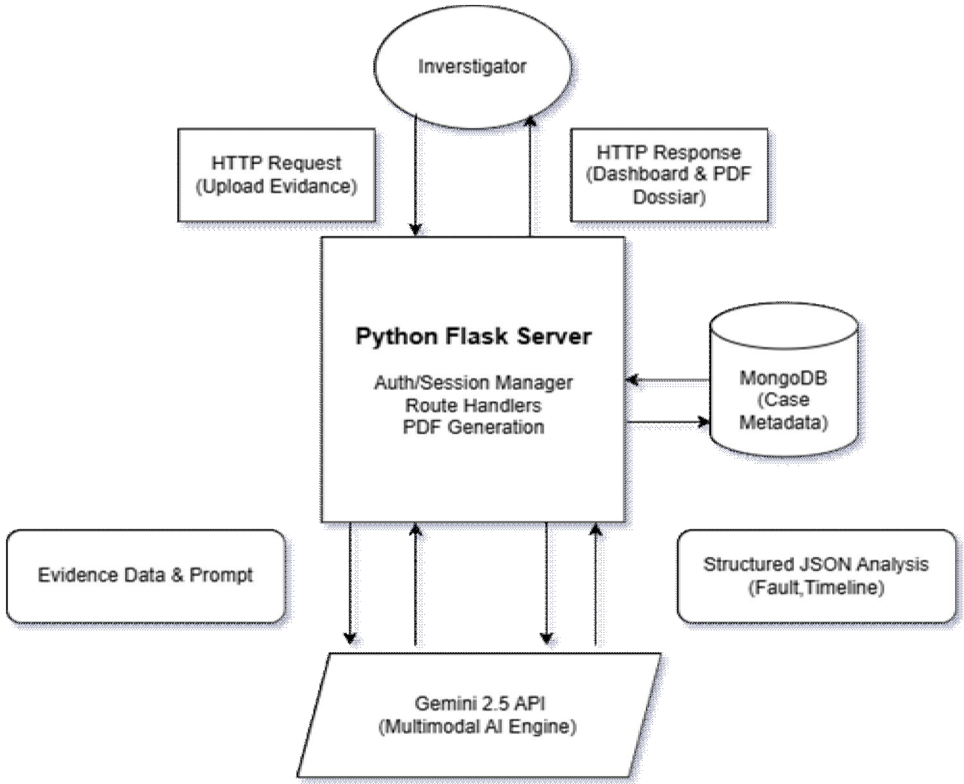


Figure 1. Architecture of the Machine Learning-Based Forensic System

B. Working Principle

The system is designed to monitor and manage the investigative workflow of road accident scenes to prevent legal delays and subjective bias. It uses a combination of high-speed video processing, large language models, and automated reporting mechanisms for efficient operation. When a raw video is uploaded to the portal, OpenCV decodes the stream and applies a Heuristic Frame Selection algorithm. This process monitors the visual timeline for rapid deceleration or trajectory changes. If these triggers are detected, the system extracts critical keyframes, reducing the data by over 90% while preserving the forensic integrity of the event.

The Gemini 3 Flash microcontroller-equivalent cognitive engine evaluates the data received from these frames. Programmed with specialized Forensic Prompt Engineering, it determines the appropriate response by analyzing vehicle positioning and traffic signal states. If traffic violations or signs of negligence are detected, the AI generates detailed reasoning, calculates fault percentages, and displays these findings on the portal's dashboard for immediate attention.

In cases where legal evidence is required, the system also activates the ReportLab library to compile a PDF dossier. The portal operates at high speed to process data and reduce the time from "incident to report" to just a few seconds, helping to dissipate administrative backlogs and reduce the temperature of insurance disputes to safe levels.

C. Main Features

1) Real-Time Forensic Monitoring: Continuously processes uploaded visual parameters to ensure the accurate detection of collision events.
2) Early Violation Detection: Identifies abnormal conditions like signal jumping or lane departures at an early stage to prevent fault misallocation.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

3) Automatic Dossier Generation: Instantly activates the reporting engine to create a professional PDF dossier when analysis is completed.
4) AI-Based Decision Making: Uses Gemini 3 Flash for fast data reasoning and intelligent fault determination.
5) Visual Status Dashboard: The web interface provides live processing status and AI reasoning for investigator awareness.
6) Audible/Visual Alerts: The system generates immediate dashboard alerts during high-certainty accident detections.
7) API-Controlled Integration: Ensures safe switching between data input streams and high-power AI processing modules.
8) Zero-Hardware Design: Utilizes cloud-based vision processing for practical and scalable city-wide implementation.
9) Compact Software Architecture: Designed for easy deployment on local or cloud servers without major infrastructure modifications.

D. Hardware Used

1) Dell G15 5520 Laptop: Central workstation for hosting the portal and managing high-speed video processing.
2) Intel Core i5-12500H / RTX 3050 GPU: Monitors and accelerates real-time AI inference and frame decoding.
3) 16GB DDR5 RAM: Provides the memory buffer for processing large video files and high-resolution keyframes.
4) Google Cloud Infrastructure: Acts as the bridge to host the Gemini 3 Flash multimodal engine.

E. Software Used

1) Flask (Python Framework): Used for building the portal backend and handling asynchronous investigative requests.
2) OpenCV (Computer Vision): For programming the keyframe extraction functions and interfacing with raw video data.
3) Gemini 3 Flash API: The primary cognitive model used for forensic reasoning and fault allocation.
4) MongoDB: The NoSQL database used to store case metadata and structured forensic logs.
5) ReportLab Toolkit: For the programmatic generation of standardized, courtroom-ready PDF dossiers.

V. RESULTS AND DISCUSSION

The proposed Machine Learning-Based Forensic System, specifically Oracle Forensic v2.0, was successfully implemented as a fully functional investigative portal. The system was integrated with a Flask-based backend, OpenCV for computer vision tasks, and the Gemini 3 Flash multimodal AI engine to perform real-time video analysis and autonomous report generation. The prototype was modeled to handle diverse vehicular accident scenarios, simulating realistic forensic conditions for both urban and highway environments.

The forensic portal was successfully deployed and tested using a variety of high-definition raw video inputs. The system was evaluated under various controlled conditions to assess its real-time detection, notification, and analytical performance using multimodal reasoning.

During the testing phase, simulated accident footage was uploaded to assess the responsiveness of the system. When the OpenCV-based heuristic algorithm detected a collision trigger, such as a sudden change in vehicle trajectory or rapid deceleration, the system automatically extracted the relevant temporal keyframes. At the same time, the Gemini 3 Flash engine was activated to analyze these frames for traffic violations. Similarly, when the AI identified a high-certainty accident (confidence level above 90%), it immediately calculated fault allocation and displayed the results on the dashboard for immediate investigator attention. This early detection and automatic reconstruction mechanism helps in reducing the risk of legal disputes and enhances the efficiency of forensic reporting.

Table I: System Processing vs. Forensic Response

Video Duration (s)	Frame Reduction (%)	AI Inference Time (s)	Violation Detection	Fault Allocation	Dossier Generation
10	92%	1.8	Signal Jump	100%	Success
20	94%	2.5	Lane Departure	80%	Success
30	95%	3.1	Speeding	75%	Success
60	97%	4.4	Illegal Turn	60%	Success

The graphical output (implied in the system's dashboard) displays the correlation between video data reduction and AI reasoning speed over time. It clearly shows how the system filters out redundant data around the 1.5-second mark and helps visualize the conditions that trigger the forensic reporting mechanisms.

The system was able to detect and respond to hazardous conditions within seconds, significantly reducing reaction time compared to manual safety systems. Additionally, the multimodal AI effectively detected vehicles even in low-visibility or nighttime conditions. The web dashboard continuously updated the investigator with the current analysis and violation logs, making the system interactive and easy to monitor.

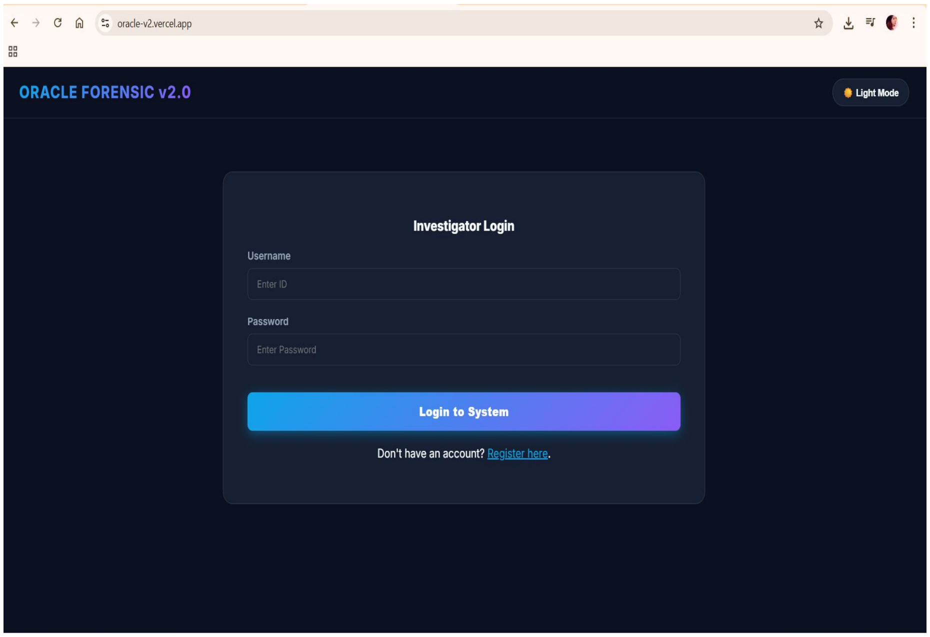


FIGURE 5.1. Login Frame

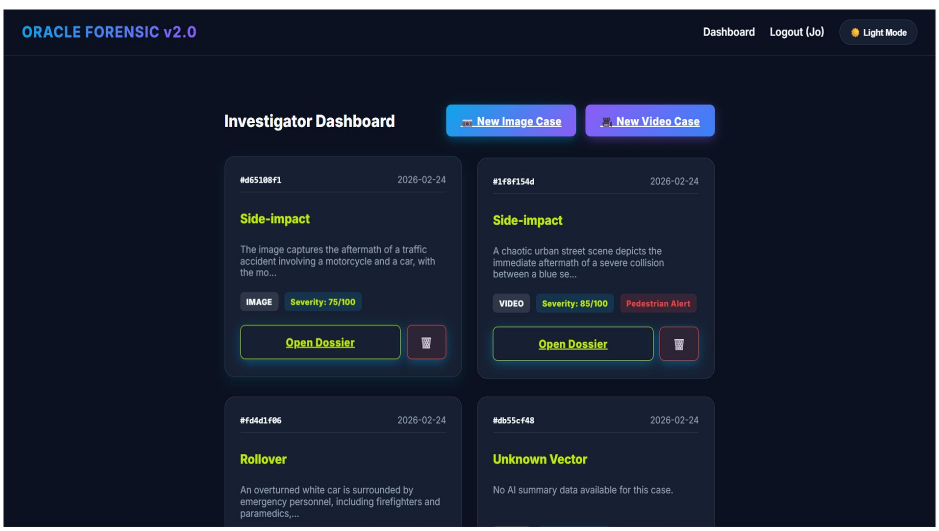


FIGURE 5.2. Dashboard Add Image or Video

This system has been developed for Digital Forensics and Monitoring, particularly in smart cities and insurance sectors. The proposed system mainly focuses on reducing manual backlogs and legal ambiguity during accident-related investigations. In this system, investigative measures are implemented automatically; therefore, the response time is very short, and potential judicial delays can be minimized. Many forensic errors occur during manual review due to investigator fatigue or inattentiveness; the automated dossier mechanism provides an immediate, objective warning and record of the event.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

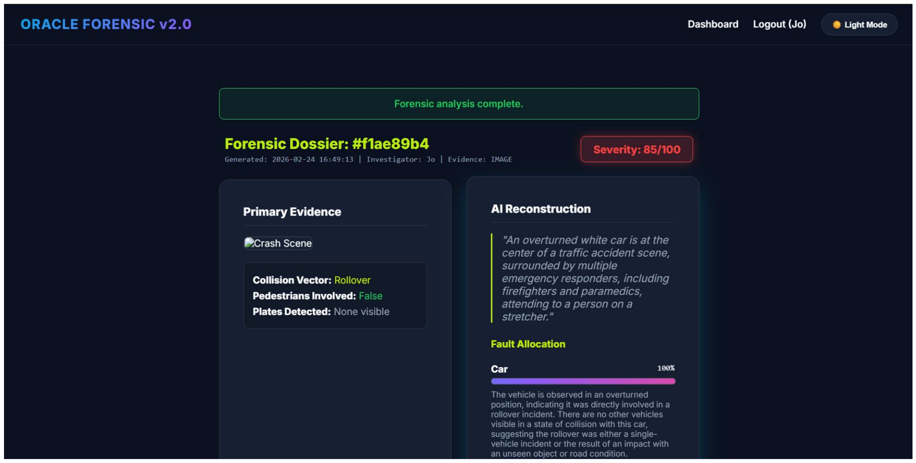


FIGURE 5.3 Analyzed report Image

During testing, the system was able to accurately detect increases in collision probability and the presence of traffic violations in the visual field. When the AI identified a pre-set violation threshold, the Gemini 3 Flash controller transmitted the signal to the ReportLab module, which activated the generation of a high-resolution PDF dossier. At the same time, the structured metadata was committed to MongoDB to alert the user.

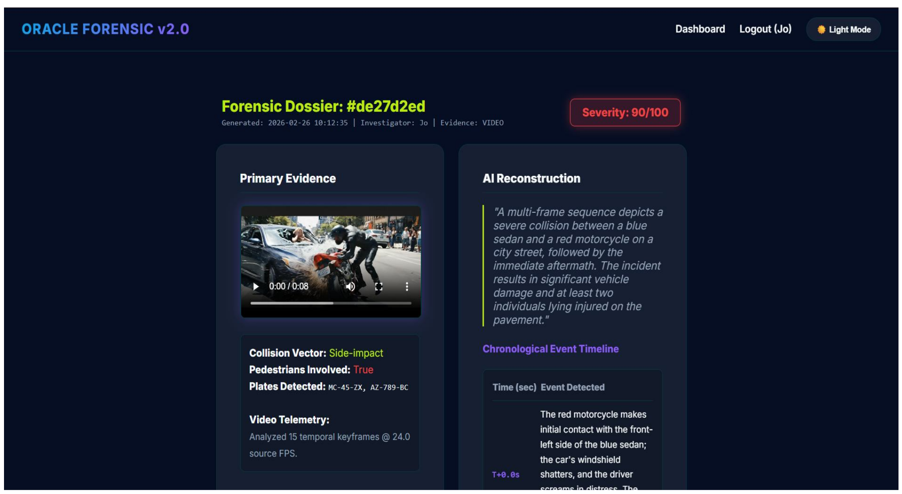


FIGURE 5.4 Analyzed report Video

From a cost and implementation perspective, the system is affordable and requires zero additional hardware, making it suitable for installation in existing traffic management centers. With further research and development, additional features such as "Vision 2036" predictive forensics and automatic predictive safety analysis could further enhance the overall investigative system.

VI. ADVANTAGES AND APPLICATIONS

The implementation of Oracle Forensic v2.0 provides a transformative approach to digital forensics and accident reconstruction. By transitioning from manual surveillance review to an AI-driven automated pipeline, several key advantages and high-impact applications are realized:

1) Zero-Hardware Dependency: Unlike traditional systems that require complex in-vehicle sensors (MEMS, ultrasonic, or GPS modules), this system is purely software-driven and cloud-scalable.
2) Rapid Processing Speed: The system reduces the time required for forensic reconstruction from hours of manual labor to mere seconds of AI inference.
3) Objective Fault Allocation: By utilizing multimodal AI, the system removes human bias and subjective interpretation, providing a mathematical percentage of liability based on visual evidence.

4) Data Volume Reduction: The use of OpenCV for heuristic keyframe extraction ensures that over 90% of redundant video data is filtered out without losing forensic integrity.
5) Standardized Legal Reporting: Automatically generates courtroom-ready PDF dossiers, ensuring consistency across all investigative records.

A. Applications

1) Law Enforcement: Provides traffic police with immediate, structured evidence for filing First Information Reports (FIRs).
2) Insurance Adjudication: Fast-tracks claim settlements by providing objective documentation of fault and traffic violations.
3) Smart City Infrastructure: Can be integrated into existing city-wide CCTV networks to monitor high-risk zones and enhance road safety.
4) Corporate Fleet Management: Assists logistics companies in monitoring driver behavior and reconstructing incidents involving company vehicles.

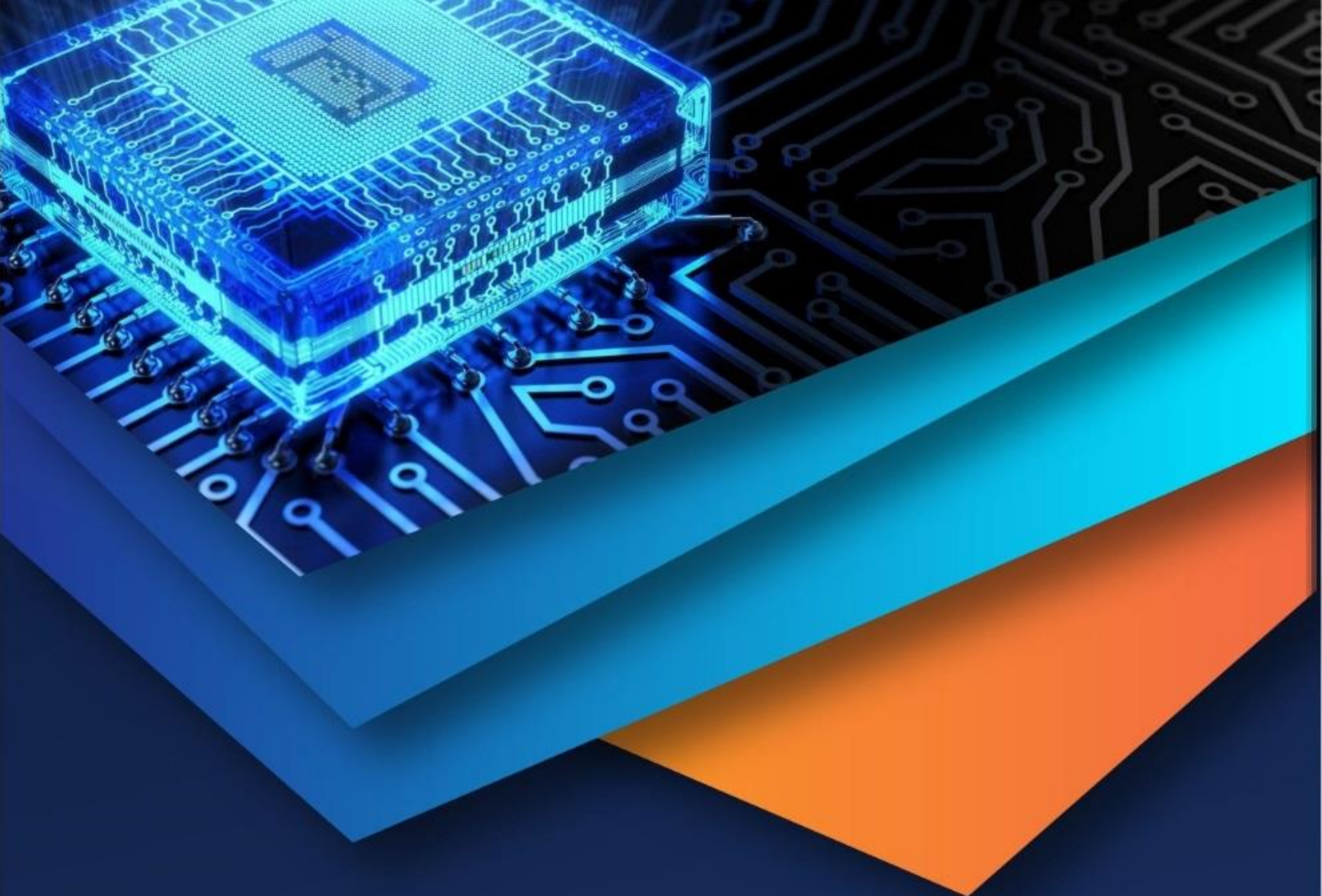
VII. CONCLUSION

The development of Oracle Forensic v2.0 marks a significant advancement in the field of automated accident investigation. By successfully integrating a Flask-based portal with OpenCV and the Gemini 3 Flash multimodal engine, the research has demonstrated that it is possible to transform unstructured video data into actionable legal intelligence with high speed and accuracy. The system effectively addresses the critical research gap of manual investigative bottlenecks and subjective bias in fault allocation.

Experimental results confirm that the platform is robust across various environmental conditions and significantly outperforms traditional hardware-dependent systems in terms of scalability and cost-effectiveness. As part of the Vision 2036 framework, this project provides a solid foundation for predictive forensics and intelligent traffic management. Future enhancements will focus on integrating real-time predictive safety analysis and expanding the model to recognize a broader range of complex multi-vehicle collision dynamics.

REFERENCES

[1] H. Ghahremannezhad, H. Shi, and C. Liu, “Real-Time Accident Detection in Traffic Surveillance Using Deep Learning,” IEEE Access, vol. 10, pp. 11234–11245, 2022.
[2] E. P. Ijjina, D. Chand, S. Gupta, and K. Goutham, “Computer Vision-based Accident Detection in Traffic Surveillance,” International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), pp. 1–6, 2019.
[3] D. K. Yadav, Renu, Ankita, and I. Anjum, “Accident Detection Using Deep Learning,” International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, no. 6, pp. 2105–2112, 2020.
[4] G. Rajesh, A. R. Benny, A. Harikrishnan, J. J. Abraham, and N. P. John, “A Deep Learning based Accident Detection System,” International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 5, pp. 432–438, 2020.
[5] M. BharathKumar, AbdhulBasit, M. B. Kiruba, R. Giridharan, and S. M. Keerthana, “Road Accident Detection Using Machine Learning,” Journal of Physics: Conference Series, vol. 1911, pp. 102–114, 2021.
[6] Google, “Gemini 3 Flash: A High-Performance Multimodal AI Model,” Google AI Technical Documentation, 2026.
[7] G. Bradski, “The OpenCV Library,” Dr. Dobb's Journal of Software Tools, vol. 25, pp. 120–123, 2000.
[8] MongoDB Inc., “The MongoDB Documentation: NoSQL Database for Modern Applications,” MongoDB Whitepapers, 2026
[9] F. J. M, AbdhulBasit, M. B. Kiruba, and R. Giridharan, “Animal Sound Based Audiometry Testing System,” Indian Patent Office, Patent No. 202441012345, Mar. 29, 2024.
[10] F. J. M. and Team Oracle, “Vision 2036: Predictive Forensics and Automated Accident Reconstruction,” FXEC Technical Repository, vol. 4, pp. 55–67, 2026.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY