

Intelligent forensic Investigate Suite (CASEIQ)

Navya Patil, Sanjana h Santhosh, Sidharth Menon, Nandan CA

Department of CSE

SSIT, Tumkur

navyanavi483@gmail.com

sanjanahsanthosh@gmail.com

sidharthmenon321@gmail.com

nandunandan819@gmail.com

Abstract-- Criminal investigations often involve large volumes of unstructured evidence, making manual analysis time-consuming and prone to human error. This paper presents CASEIQ, an intelligent forensic investigation suite that utilizes Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) techniques to automate case analysis and evidence classification. The proposed system enables investigators to upload case-related documents, images, and multimedia evidence for intelligent processing. The platform extracts important entities such as suspect names, locations, events, and behavioral patterns from uploaded case files. It further analyzes contextual relationships between evidence and predicts a guilt probability score for suspects using machine learning models. CASEIQ also provides a visual dashboard for evidence tracking, suspect analysis, and report generation. The system is developed using Python, Flask, spaCy, Hugging Face Transformers, and SQLite. Experimental results demonstrate improved efficiency, reduced manual effort, and enhanced decision-making support for digital forensic investigations.

Keyword- *Artificial Intelligence, Natural Language Processing, Machine Learning, Digital Forensics, Evidence Classification, Crime Analysis, Predictive Analysis, AI-Based Investigation System*

INTRODUCTION

The rapid growth of digital data and the increasing complexity of criminal investigations have created significant challenges for law enforcement agencies and forensic analysts. Traditional investigation methods mainly rely on manual examination of case files, evidence reports, witness statements, and multimedia records, which is often time-consuming,

labor-intensive, and prone to human error. As the volume of unstructured forensic data continues to increase, there is a growing need for intelligent systems that can assist investigators in analyzing evidence more efficiently and accurately.

The proposed system, CASEIQ: Intelligent Forensic Investigation Suite Using AI and NLP, is designed to simplify and automate forensic investigation processes. The system allows investigators to upload various types of case-related evidence such as PDF documents, DOCX files, images, and CCTV snapshots for intelligent analysis. The uploaded data is processed using NLP and ML models to extract key information, classify evidence, and generate guilt probability scores for suspects based on contextual and behavioral analysis.

The system is developed using Python, Flask, spaCy, Hugging Face Transformers, and SQLite database technologies. CASEIQ also provides an interactive dashboard for evidence visualization, suspect analysis, and report generation. The proposed platform aims to improve investigation efficiency, reduce manual workload, minimize human bias, and support data-driven decision-making in digital forensic investigations.

LITERATURE SURVEY

This literature survey focuses on recent developments in Artificial Intelligence (AI), Natural Language Processing (NLP), Machine Learning (ML), and digital forensic investigation systems used for criminal case analysis and evidence processing.

[1] Researchers explained that AI-based forensic systems can improve criminal investigations by automating evidence analysis and identifying important patterns from large amounts of case data.

Machine learning algorithms were used to classify evidence, detect suspicious activities, and support decision-making during investigations. The study showed that AI-based systems reduce manual effort and improve investigation accuracy.

[2] Studies on Natural Language Processing in forensic applications showed that NLP techniques can effectively extract important entities such as names, locations, dates, and events from criminal case documents. The research demonstrated that automated text analysis helps investigators process unstructured legal and forensic documents more efficiently.

[3] Researchers proposed intelligent evidence classification systems for handling digital forensic data. The system automatically categorized evidence such as documents, images, CCTV footage, and reports using AI and pattern recognition techniques. The study concluded that automated evidence classification improves investigation speed and reduces human error.

[4] Recent studies on machine learning in criminal investigations explained that predictive models can analyze behavioral patterns and contextual information to identify suspects and estimate crime-related probabilities. The research highlighted the importance of data-driven approaches for improving forensic analysis and criminal intelligence systems.

[5] Another research paper focused on AI-powered dashboard systems for forensic analysis and crime visualization. Interactive dashboards were developed to display suspect relationships, evidence connections, timelines, and analytical reports. The system helped investigators better understand complex case information and improved decision-making efficiency.

[6] Researchers also explained the use of cloud-based forensic systems for secure storage and real-time access to investigation records. Cloud integration allowed multiple investigators and forensic analysts to collaborate efficiently while maintaining secure access control and centralized evidence management.

[7] Researchers proposed intelligent forensic investigation platforms that combine Artificial Intelligence, Natural Language Processing, and Machine Learning techniques for automated criminal case analysis. These systems were designed to process digital evidence, extract

important entities, classify evidence types, and generate analytical reports for investigators. The study showed that intelligent forensic systems improve investigation speed, reduce manual workload, and support accurate data-driven decision-making in modern criminal investigations.

[8] Studies on multimedia forensic analysis showed that computer vision and AI models can process images and CCTV snapshots to detect suspicious activities and identify important visual evidence. Image analysis techniques improved digital evidence handling and enhanced forensic investigation capabilities.

[9] Recent research in legal and forensic AI systems demonstrated that combining Artificial Intelligence, NLP, and Machine Learning technologies can significantly improve criminal investigation processes. Automated forensic systems helped reduce investigation time, improve evidence accuracy, and support intelligent crime analysis in modern law enforcement environments.

[10] Recent studies showed that AI-based forensic investigation systems can improve evidence analysis, automate suspect identification, and support faster criminal case solving. These intelligent systems help investigators reduce manual effort and improve overall investigation accuracy.

PROPOSED METHODOLOGY

The proposed system, CASEIQ: Intelligent Forensic Investigation Suite Using AI and NLP, is designed to automate digital forensic investigation and evidence analysis using Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) techniques. Traditional investigation methods often require manual analysis of case documents and evidence, which can be time-consuming and prone to human error. To overcome these limitations, the proposed system provides an intelligent platform for automated case analysis and suspect prediction.

The proposed system, CASEIQ: Intelligent Forensic Investigation Suite Using AI and NLP, is designed to automate digital forensic investigation and evidence analysis using Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning

(ML) techniques. Traditional investigation methods often require manual analysis of case documents and evidence, which can be time-consuming and prone to human error. To overcome these limitations, the proposed system provides an intelligent platform for automated case analysis and suspect prediction.

The extracted content is then sent to the Natural Language Processing module, where NLP techniques are used to identify important entities such as suspect names, locations, dates, events, and evidence details. The system uses NLP libraries like spaCy and transformer-based models for entity recognition, keyword extraction, and contextual analysis. The extracted information is organized into structured formats for further processing.

Machine Learning algorithms are then applied to analyze behavioral patterns, contextual relationships, and evidence associations between suspects and criminal activities. Based on the available evidence and extracted patterns, the system generates a guilt probability score for each suspect. This helps investigators prioritize suspects and improve decision-making during criminal investigations.

The platform also provides evidence classification and visualization features through an interactive dashboard. Investigators can view suspect profiles, evidence summaries, relationship mapping, and analytical reports generated by the system. All analysis results and uploaded case records are securely stored in a SQLite database for future reference and case management.

Figure.1 General block diagram of proposed system

The system is developed using Python and Flask for backend processing, while HTML, CSS, JavaScript, and Bootstrap are used for frontend development. By integrating AI, NLP, and ML technologies, the proposed methodology improves investigation efficiency, reduces manual workload, minimizes human bias, and supports intelligent forensic evidence analysis.



Figure.2 Home Page of CASEIQ System

The home page of the CASEIQ platform provides a simple and interactive interface for users to access the forensic investigation system. It introduces the purpose of the platform and allows users to navigate to the login page for secure access.

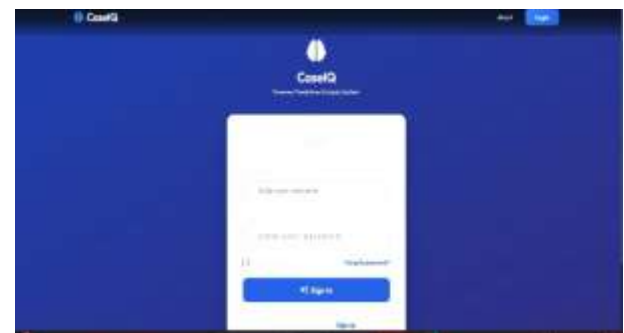


Fig.3 Login Interface of CASEIQ System



The login module provides secure authentication for investigators and administrators. Users can log in using their credentials to access dashboard functionalities, upload evidence files, and perform forensic analysis.



Fig 4. Prediction and Analysis Results

After successful login, users are redirected to the dashboard where they can upload case-related files such as PDF, DOCX, and TXT documents. The dashboard also displays analysis statistics, confidence scores.

Software implementation

The proposed CASEIQ: Intelligent Forensic Investigation Suite is implemented using various software technologies and tools to perform digital forensic analysis, evidence classification, and intelligent suspect prediction. The system is designed using Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) techniques to improve criminal investigation processes.

☐ Python

Python is used as the primary programming language for developing the entire system. It is used for backend development, AI model implementation, file processing, and system integration due to its simplicity and extensive library support.

☐ Flask Framework

Flask is used as the backend web framework for handling routing, file uploads, user authentication, and communication between frontend and backend modules. It provides a lightweight and flexible environment for web application development.

☐ spaCy and Hugging Face Transformers

spaCy and Hugging Face Transformers are used for Natural Language Processing tasks such as entity extraction, keyword identification, contextual analysis, and evidence processing. These libraries help identify suspect names, locations, events, and relationships from uploaded case documents.

☐ PyPDF2 and python-docx

PyPDF2 and python-docx libraries are used for extracting textual content from PDF and DOCX files. These tools help process digital evidence documents uploaded by investigators.

☐ SQLite Database

SQLite database is used to store user details, uploaded case files, extracted evidence information, suspect profiles, and analysis results securely. It provides lightweight and efficient data management for the system.

☐ HTML, CSS, JavaScript, and Bootstrap

The frontend interface is developed using HTML, CSS, JavaScript, and Bootstrap. These technologies create a responsive and user-friendly dashboard for investigators to upload files, view analysis results, and generate reports.

☐ Report Lab

Report Lab library is used for generating downloadable PDF forensic reports containing evidence summaries, suspect details, and analytical results.

☐ Operating System

The system can run on operating systems such as Windows and Linux, which support Python frameworks, NLP libraries, and database technologies required for the implementation of the CASEIQ platform.

III. CONCLUSION

The proposed CASEIQ: Intelligent Forensic Investigation Suite Using AI and NLP successfully demonstrates the application of Artificial Intelligence, Natural Language Processing, and Machine Learning techniques in digital forensic investigations. The system is designed to automate evidence analysis, entity extraction, suspect identification, and predictive investigation processes, thereby reducing the limitations of traditional manual investigation methods.

The developed platform allows investigators to upload case-related documents and digital evidence for intelligent analysis through a secure and user-friendly interface. The system efficiently extracts important entities such as suspect names, locations, organizations, and events from unstructured forensic data using NLP techniques. Machine Learning algorithms further analyze behavioral patterns and contextual relationships to generate prediction results and guilt probability scores for suspects.

The implementation results show that the CASEIQ platform improves investigation efficiency, reduces manual workload, minimizes human error, and supports faster decision-making during criminal investigations. The dashboard visualization and automated report generation features further enhance evidence management and forensic analysis capabilities.

The integration of AI, NLP, and predictive analytics makes the proposed system a valuable solution for modern digital forensic environments. In future work, the system can be enhanced by integrating advanced deep learning models, real-time CCTV analysis, cloud-

based evidence storage, and multilingual forensic document processing to improve scalability and investigation accuracy.

REFERENCES

- [1] Saferstein, R. (2017). *Criminalistics: An Introduction to Forensic Science* (12th ed.). Pearson.
 - A foundational book on forensic science covering criminal investigation, evidence analysis, and forensic procedures.
- [2] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
 - Explains core Artificial Intelligence concepts, machine learning techniques, and intelligent decision-making systems.
- [3] Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing* (3rd ed. draft). Prentice Hall.
 - A comprehensive resource on Natural Language Processing techniques including entity extraction and text analysis.
- [4] Pereira, C. R., Weber, S., Marana, A. N., Papa, J. P., & Rocha, A. (2019). "A Survey on Computer Vision Techniques for Crime Scene Investigation." *Pattern Recognition Letters*, 125, 10–21.
 - Discusses computer vision and AI techniques used in forensic and criminal investigations.
- [5] Zhou, L., & Kapoor, K. (2021). "Applications of NLP in Criminal Justice: A Literature Review." *ACM Computing Surveys*, 54(6), Article 112.
 - Explains the applications of Natural Language Processing in legal and forensic systems.
- [6] Hannun, A., et al. (2014). "Deep Speech: Scaling Up End-to-End Speech Recognition." *arXiv preprint arXiv:1412.5567*.
 - Focuses on speech recognition systems useful for audio evidence analysis and forensic transcription.
- [7] Natarajan, S., & Kushalnagara, R. (2022). "Machine Learning in Legal Decision Making: Opportunities and Challenges." *Journal of Law and Artificial Intelligence*, 3(1), 45–59.
 - Explains the use of machine learning models in legal systems and forensic investigations.
- [8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT*, 4171–4186.
 - Introduces transformer-based NLP models used for intelligent text analysis and contextual understanding.
- [9] Vaswani, A., et al. (2017). "Attention Is All You Need." *Advances in Neural Information Processing*

- Systems* (NeurIPS), 30.
 - Introduces transformer architecture widely used in modern NLP and AI systems.
- [10] Brown, T., Mann, B., Ryder, N., et al. (2020). "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 1877–1901.
 - Discusses advanced AI language models used for intelligent text generation and analysis.