Title

MISSING PERSON AND WANTED CRIMINAL DATABASE SYSTEM

Author

JAMILLA ISSAH

Co-Author

MR. RICHARD PHIRI

Issue November 2025

Certificate AR2025ER5I27

ABSTRACT

The absence of a centralized system for reporting and tracking missing persons and wanted criminals poses significant challenges for law enforcement and the public. This web-based application addresses these challenges by enabling both police and citizens to report missing persons with details such as photos and last known locations. Law enforcement can also flag wanted criminals, allowing citizens to submit anonymous tips that aid investigations. Email notifications keep users informed about new cases and important developments, ensuring timely dissemination of information. By leveraging technology for real-time reporting and tracking, this system enhances public safety, streamlines police efforts, and increases the chances of reuniting missing individuals with their families. A critical innovation is the integration of a machine learning module using the Random Forest algorithm, which analyzes past cases and predicts potential connections between reports. This intelligent feature assists officers in identifying trends or recurring patterns, improving investigation accuracy. Overall, the system contributes to enhancing public safety, improving law enforcement efficiency, and establishing a digital foundation for crime prevention and social security in Malawi.

KEYWORDS: Missing persons, Wanted criminals, Database system, Machine learning, Random Forest, Web-based application, Email notifications

INTRODUCTION

Background of Study

The rapid increase in reported cases of missing persons and wanted criminals has become a pressing issue for both government and society. In Malawi and other developing countries, the process of locating missing individuals or apprehending fugitives is typically slowed by inefficient manual record-keeping systems and lack of coordination between police departments and the public. These traditional methods, including posters, social media posts, and local announcements, often fail to reach a wide audience or provide timely updates. The inefficiency results in delayed responses, reduced recovery rates, and increased risk to vulnerable individuals.

This research introduces a centralized web-based platform designed to bridge the communication gap between law enforcement agencies and the public. The proposed system allows citizens to report missing persons and wanted criminals through an online portal, upload photographs, and provide last-known details. The police can then verify, classify, and update reports in real time, making the process transparent and faster.

In addition, the inclusion of email notifications ensures that relevant authorities and families receive immediate updates on reported cases.

A critical innovation of this project is the integration of a machine learning module using the Random Forest algorithm, which analyzes past cases and predicts potential connections between reports. This intelligent feature assists officers in identifying trends or recurring patterns, improving investigation accuracy. Overall, the system contributes to enhancing public safety, improving law enforcement efficiency, and establishing a digital foundation for crime prevention and social security in Malawi.

Objectives

The primary goal of this research is to develop and evaluate a Missing Person and Wanted Criminal Database System capable of accurately reporting, tracking, and predicting connections in cases using machine learning techniques. The study is guided by the following specific objectives:

1. To promote public engagement and safety in reporting cases: This objective emphasizes the social and community impact of the system. Many cases in Malawi remain unresolved due to lack of public involvement and centralized reporting. By integrating these into a digital platform, the project helps safeguard public safety and ensures that citizens and authorities can engage with transparent tools without barriers, promoting accountability and inclusion. Moreover, it supports authorities and families in documenting and resolving cases for safety and heritage purposes.
2. To enhance accessibility and user interaction through digital technology: This objective focuses on usability and inclusiveness. The project aims to build a platform that enables citizens, especially those in remote areas, to interact easily with law enforcement systems through web access. Instead of manual reporting or checks, users can submit details digitally in real time. This improves access to information and services, especially in rural or underserved communities. It also benefits individuals with limited mobility by offering a more natural and convenient way of reporting through digital interaction.
3. To design an efficient and adaptable ML model for real-time case prediction: The goal is to build a robust ML model that performs case recognition and prediction quickly and accurately, even in varying data environments or with different inputs. The model should also be adaptable, meaning it can be trained or fine-tuned to recognize additional

patterns or new cases as more data becomes available. Real-time processing capability is essential to ensure smooth and instant matching during investigations, public services, or digital communication. Ultimately, this will make the system reliable, efficient, and scalable for broader applications across regions and case types.

LITERATURE REVIEW

A literature review serves as a critical analysis of existing scholarly works relevant to a particular topic or research question. It provides a comprehensive overview of the current state of knowledge, identifies gaps, and synthesizes key findings to inform further research. By examining and synthesizing existing literature, researchers gain insights, contextualize their own work, and contribute to the advancement of knowledge in the field.

Overview of Research Studies

The study of missing persons and criminal tracking has been widely explored in academic and law enforcement research. Several technological advancements, including online databases and predictive analytics, have been implemented to improve case resolution.

Miethe, T. D., Wetherell, K. T. E., & Regoeczi, W. C. (2023). Missing Persons: A National Survey Approach Assessing the Predictors of Case Outcomes. This study examines the key factors influencing the outcomes of missing person cases, focusing on demographic characteristics and the effectiveness of search efforts. The findings highlight the importance of coordinated law enforcement strategies and public engagement in locating missing individuals.

Phoenix, J., & Francis, B. J. (2023). Police Risk Assessment and Case Outcomes in Missing Person Investigations. The research investigates how law enforcement classifies missing persons as high-risk and examines the impact of age, sex, and situational factors on case resolutions. It underscores the importance of early intervention strategies in improving case outcomes.

Richards, Y., McClish, M., & Keatley, D. (2023). Finding the Missing: Using Statement Analysis to Assist in Missing Persons Cases. This paper explores how linguistic analysis techniques can be used to distinguish truthful statements from deceptive ones in missing person cases. It highlights the potential of forensic linguistics in aiding investigative processes.

Papesh, M. H., Cash, D. K., Pinto, J. D. G., & Lomba, S. V. (2024). Spotting Missing or Wanted People: Racial Biases in Prospective Person Memory. The study analyzes racial biases in recognizing missing and wanted

individuals, revealing disparities in identification accuracy across different racial groups. The findings suggest the need for bias-aware training programs in law enforcement.

McKie, R. (2024). The Nameless Dead: Scientists Hunt for Identities of Thousands Who Tried to Reach Europe. This article discusses the challenges of identifying deceased migrants who attempted to reach Europe, highlighting new scientific methods for forensic identification. It explores the role of DNA matching, biometric databases, and international collaboration in resolving unidentified cases.

METHODOLOGY AND TOOLS

The system was developed using the Agile Software Development methodology, which focuses on iterative design, user feedback, and continuous improvement. The project was divided into several development phases, including requirements gathering, system design, implementation, testing, and deployment. In this context, the key challenge addressed is the limited availability of centralized systems for reporting missing persons and wanted criminals, especially within low-resource settings like Malawi.

The Agile framework was suitable as it integrates both scientific rigor and practical innovation, enabling a structured yet adaptable process for designing and evaluating a machine learning-based database application. The methodology followed three major phases: system design, system development, and system evaluation.

Each phase was guided by the agile methodology, which supports iterative development, rapid prototyping, user feedback, and continuous system improvement. Agile divides the development cycle into short, manageable sprints, ensuring that user input and real-world testing inform every iteration of the system.

System Design Phase

The design phase began with the identification of both functional and non-functional requirements. Data collection involved interviews and observation with police officers and community representatives to understand the shortcomings of current systems, case diversity, pattern variations, and user expectations.

The system’s architecture was then conceptualized, focusing on modularity, scalability, and adaptability to multiple case types. The design process emphasized:

- A data acquisition module to capture and preprocess input;
- An ML-based recognition engine for case-specific prediction and modeling; and

- A dashboard interface for real-time notification output.

The design also included database structures for storing user data, case samples, and prediction results. Furthermore, experts and local stakeholders were engaged to help identify pattern variations and indicators commonly used in target setups. This ensured practical and technical relevance of the system.

System Development Phase

The development phase involved implementing the designed architecture into a functional prototype. Development was conducted in Agile sprints, where each sprint targeted specific components such as data preprocessing, prediction modeling, ML alert generation, and graphical user interface (GUI) creation.

Key tools used in this phase included PHP, MySQL, and scikit-learn for model training and case recognition, while Bootstrap was used for frontend management. The report-to-prediction pipeline integrated a pre-trained model fine-tuned using locally collected case datasets.

Each sprint ended with functional testing, where developers and experts validated the accuracy of recognition and execution output. Feedback from users was incorporated before moving to the next sprint, promoting an iterative and user-centered design process.

System Evaluation Phase

In the evaluation phase, the prototype system was tested in a controlled environment involving participants familiar with case scenarios. A pilot test was conducted, where participants submitted various reports, and outputs were analyzed for accuracy, latency, and reliability.

Evaluation metrics included:

- Prediction Error Rate for recognition accuracy;
- Response Time for real-time processing; and
- User Satisfaction Scores through post-test surveys.

The system achieved high accuracy with stable performance in real-time operation. Ethical considerations such as informed consent, data anonymization, and participant privacy were strictly maintained throughout the testing process.

Justification for Agile Methodology

The Agile methodology was adopted due to its adaptability, focus on user collaboration, and iterative improvement cycle. Unlike traditional

waterfall models, Agile allowed developers to respond rapidly to challenges, such as differences in case patterns or data clarity.

Frequent feedback from users, experts, and technical evaluators ensured that modifications were implemented promptly without disrupting the entire workflow. This approach minimized development risks, improved system usability, and enhanced stakeholder engagement—all of which are critical for a technology-sensitive system that evolves with user interaction and case diversity.

Development Tools

The implementation of the Missing Person and Wanted Criminal Database System required a combination of programming languages, frameworks, and tools to enable robust backend processing, secure data handling.

System Architecture

The development of the Missing Person and Wanted Criminal Database System relied on a combination of backend, frontend, and auxiliary tools to ensure efficiency, accuracy, and scalability. The backend tools used were PHP and MySQL, which handled the system's server-side logic, user authentication, and secure data management. The system adopts a three-tier architecture consisting of the presentation layer, application layer, and data layer. The frontend tools included HTML, CSS, JavaScript, and Bootstrap, which were used to design a responsive and user-friendly interface, enabling users to submit or view cases and outputs seamlessly across multiple devices. For machine learning and model development, Python with scikit-learn served as the core, supporting data preprocessing, model training, and feature extraction. In addition, tools like session control were utilized for secure authentication, including token-based access and normalization. Together, these tools created a robust and integrated technological foundation that ensured the system's ability to accurately manage cases while maintaining efficiency, data security, and user accessibility.

Data Collection and Preprocessing Data Sources

The Missing Person and Wanted Criminal Database System relies heavily on the quality and diversity of its data. Datasets were collected from a combination of open-source repositories and locally recorded samples to ensure high prediction accuracy and relevance. The project adopted a multi-layered approach to curate, preprocess, and validate all datasets used to train and test the ML model.

Open-Source Datasets: Foundational data for the system was obtained from publicly available corpora such as INTERPOL records and historical case examples. These datasets provided general examples across different patterns and conditions. To ensure the system could recognize local variations, additional samples were recorded from stakeholders in community settings.

Locally Curated and Expert-Reviewed Data

To ensure accuracy, locally recorded files were reviewed by experts and users. The reviewers verified consistency, action accuracy, and distinctions for setups with limited forms. Their contributions were essential in eliminating errors and ensuring that the system reflected authentic patterns.

This collaborative process helped enhance the model's reliability, appropriateness, and depth.

Data Cleaning and Noise Filtering

Before training, all collected data underwent rigorous preprocessing. Samples were cleaned using tools to remove noise, normalize, and ensure clear input. The recordings were segmented into smaller, uniform clips for easier processing.

Details were standardized to remove filler words, repeated sounds, and inconsistent spellings. This improved data quality and helped the ML engine learn precise relationships between input and action.

Language and Tone Filtering

Because cases often express meaning through context and details, classification tools were integrated during preprocessing. Each sample was labeled based on variation (e.g., high-risk, recurring). This step was critical in helping the model distinguish between semantically similar cases that vary by context.

Localization and Language Support

Considering diversity, the system was designed to handle multilingual input, focusing on English and related variations. Surveys and interviews were conducted to collect expressions, slang, and region-specific details. This localization ensured that the model was contextually accurate, sensitive, and adaptable to local communication styles.

Testing and Evaluation Study Design

To evaluate system performance, a pilot study was conducted involving participants who were users of various scenarios. Participants submitted short reports, which were then processed by the system. The resulting actions

were compared against manually verified references. Feedback was collected through surveys to assess usability, speed, and accuracy of the system.

Types of Testing Performed

- Usability Testing: Examined how intuitive and accessible the application interface was. Participants evaluated ease of submitting reports, playback, and output viewing.
- Functional Testing: Verified whether the main functions including input, saving, output generation, and export worked correctly under different use conditions.
- Accuracy Testing: Measured how precisely the system matched cases. Results were compared with ground truth to determine the Prediction Error Rate and Execution Accuracy.
- Performance and Reliability Testing: Evaluated response time, stability, and system performance under varied workloads.
- Security and Data Handling Testing: Ensured data confidentiality through encryption, secure authentication, and anonymization of files, in compliance with ethical standards.

Evaluation Metrics

The project evaluated several performance indicators:

- Prediction Accuracy: Degree of precision between recognized and actual cases.
- Ease of Use: Simplicity and navigability of the interface.
- Processing Speed: Time taken to convert input to output.
- System Reliability: Uptime, crash rates, and response times.
- Relevance: Effectiveness in handling local expressions.

Ethical Considerations

All participants gave informed consent prior to testing. The project adhered to strict data protection and ethical research standards. No personal identifiers were stored; all samples were anonymized and encrypted. Participants were fully informed that their data would be used solely for research and system improvement purposes.

RESULTS

The results of the Missing Person and Wanted Criminal Database System were evaluated and analyzed based on three key dimensions: system performance, user experience, and technological impact. These dimensions provide a holistic

understanding of the system's effectiveness, usability, and contribution to ML advancement.

System Performance

The first dimension focused on the technical accuracy and efficiency of the ML model. The system was tested using multiple case samples collected from diverse users differing in scenarios and contexts. The ML model demonstrated high prediction accuracy, which significantly improved after additional training and noise reduction techniques.

Processing time was found to be efficient, allowing near real-time execution. Furthermore, the system successfully handled variations and differences, achieving reliable results even in low-quality inputs. The integration of algorithms such as Random Forest contributed to higher accuracy in continuous recognition. These findings affirm that the system performs effectively in real-world environments and can be optimized further through additional expansion.

User Experience

Evaluated usability, accessibility, and user satisfaction. Field testing was conducted with participants from local communities, officers, and families. Feedback revealed that users found the system intuitive, responsive, and user-friendly, especially those with limited technical skills.

The web-based interaction allowed users to communicate naturally without switching to manual processes. This created a sense of inclusion among users. Additionally, the visual output was clear, and execution accuracy built trust in the system's capability. The system's interface also enabled users to toggle between modes, promoting flexibility. Overall, users rated the system as helpful and easy to use, confirming its practical value in community and safety settings.

Technological and Societal Impact

This focused on examining the broader technological relevance and social contribution of the project. The introduction of ML-powered prediction for cases represents a major step toward inclusivity and digital equity. The system not only bridges the communication gap between technology and users but also preserves safety through digital means.

From a technological standpoint, the project demonstrated the feasibility of low-resource ML development—a challenge often faced by underrepresented setups. The successful implementation using limited resources proves that transfer learning and modeling can overcome constraints. Furthermore, the system has potential applications in governance, investigations, and public services, where

management of cases can enhance data accessibility and community engagement.

In essence, the project's impact goes beyond technology—it empowers populations to interact, report, and resolve digitally, contributing to sustainable digital transformation.

Discussion

The findings underscore the potential of ML-based database systems in promoting inclusivity and accessibility. High usability and accuracy scores demonstrate that with proper preparation and tuning, cases can be effectively digitized. Compared with conventional setups, this model performed better in recognizing patterns and expressions, making it more relatable to users.

Nevertheless, some limitations such as challenges in limited internet environments and mixed inputs highlight the need for more diverse training and improvements. The system represents a significant step toward bridging the gap between technology and users.

CONCLUSION

The Missing Person and Wanted Criminal Database System provides a modern and efficient digital solution for addressing one of the most persistent societal problems — locating missing persons and tracking wanted criminals. By leveraging web technologies and machine learning, the system offers a centralized platform that connects citizens with law enforcement agencies in real time. It promotes transparency, accountability, and timely communication, thereby improving the rate of successful investigations.

The project successfully met its objectives of developing a secure, user-friendly, and intelligent platform. It introduced innovations such as automated notifications, role-based access control, and data-driven decision support. The integration of the Random Forest algorithm proved highly beneficial in identifying potential matches, accelerating investigations.

For future improvements, the system can be enhanced with SMS notifications, mobile app integration, and AI-powered facial recognition to strengthen identification accuracy. Furthermore, incorporating a data analytics dashboard could provide visual insights into case trends, helping police agencies allocate resources strategically.

In conclusion, this project demonstrates how digital transformation can enhance public safety and support law enforcement efforts in Malawi and beyond. The implementation of such systems could revolutionize how cases are managed, ensuring that every missing person and every wanted criminal is effectively traced through technology-driven intelligence.

REFERENCES

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Elsevier.

Hargittai, E. (2005). Survey Measures of Web-Oriented Digital Literacy. Social Science Computer Review, 23(3), 371–379.

INTERPOL. (2023). Yellow and Red Notices for Missing and Wanted Persons. Retrieved from: https://www.interpol.int

Laudon, K. C., & Laudon, J. P. (2020). Management Information Systems: Managing the Digital Firm (16th ed.). Pearson Education.

McKie, R. (2024). The Nameless Dead: Scientists Hunt for Identities of Thousands Who Tried to Reach Europe. The Guardian Science.

Miethe, T. D., Wetherell, K. T. E., & Regoeczi, W. C. (2023). Missing Persons: A National Survey Approach Assessing the Predictors of Case Outcomes. Journal of Criminal Justice.

National Crime Agency. (2022). Missing Persons Data Report 2021-2022. Retrieved from: https://www.nca.gov.uk

Papesh, M. H., Cash, D. K., Pinto, J. D. G., & Lomba, S. V. (2024). Spotting Missing or Wanted People: Racial Biases in Prospective Person Memory. Psychology, Crime & Law.

Phoenix, J., & Francis, B. J. (2023). Police Risk Assessment and Case Outcomes in Missing Person Investigations. Policing and Society.

Richards, Y., McClish, M., & Keatley, D. (2023). Finding the Missing: Using Statement Analysis to Assist in Missing Persons Cases. Forensic Linguistics Review.

Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.

World Health Organization. (2022). Guidelines on Personal Data Protection in Health and Safety Systems. WHO Publications.