Safevoice: A Privacy-Preseving Multimodal Anonymous Crime Reporting System With Ai-Assisted Prioritization

[1]Avvaru R V Naga Suneetha, [2]K.Krishna Reddy, [3]T. Dinesh, [4]Ch.Varsha Sri, [5]S.Madhav Srinivas

[1]Assistant Professor, [2,3,4,5]UG Student
Vignan Institute of Technology and Science, Hyderabad, India.

Abstract :- Crime reporting is an essential component of public safety; however, a significant number of incidents remain unreported due to fear of identity disclosure, lack of accessibility, and unreliable reporting mechanisms. Traditional crime reporting systems often require personal identification and continuous internet connectivity, making them ineffective in emergency or low connectivity scenarios. To address these challenges, this paper presents SafeVoice, a privacy preserving multimodal anonymous crime reporting system with AI assisted prioritization.

The proposed system enables users to submit reports anonymously using text, image, and audio inputs through a mobile application. Offline data storage and automatic synchronization ensure reliable operation in constrained network environments. User privacy is protected through cryptographic pseudonym generation, while reported data is securely stored in a cloud backend. A hybrid prioritization mechanism combining keyword-based rules and a baseline machine learning model assists authorities in identifying urgent cases. A web based administrative dashboard supports human verification and ethical decision making. Experimental results from pilot evaluation demonstrate improved reliability, accessibility, and usability of the proposed system.

Keywords: *AI Assistance, Anonymous Crime Reporting, Machine Learning, Multimodal Reporting, Offline Synchronization, Privacy Preservation.*

1. Introduction

Timely and accurate crime reporting is a key requirement for effective law enforcement and public safety management. Information provided by citizens plays an essential role in enabling authorities to respond to incidents, investigate criminal activities, and prevent future offenses. However, in real world scenarios, a large number of crimes remain unreported or are reported late due to fear of identity exposure, lack of trust in reporting mechanisms, social stigma, and privacy concerns [1], [2]. These challenges are more prominent in sensitive cases such as harassment, domestic violence, and organized crime, where anonymity is crucial for encouraging public participation [3].

Conventional crime reporting methods, including visiting police stations, calling emergency helplines, or submitting written complaints, often require individuals to disclose personal information. This requirement discourages victims and witnesses from reporting incidents, especially when they fear retaliation or social consequences [4]. Additionally, traditional systems usually depend on continuous internet connectivity and offer limited input options, making them ineffective during emergencies or in remote and low connectivity regions [5]. As a result, valuable information that could assist investigations is frequently lost or delayed.

With the widespread adoption of smartphones and cloud technologies, digital crime reporting platforms have emerged as an alternative to traditional systems [6]. These platforms allow users to submit reports through mobile applications or web interfaces, offering improved accessibility and faster communication. However, many existing digital solutions still lack strong anonymity guarantees, offline reporting capabilities, and support for multimodal inputs such as images and audio [7]. Most platforms rely primarily on text based reporting and centralized processing, which limits their usability in stressful situations where users may not be able to provide detailed written descriptions [8].

Another critical challenge in digital crime reporting systems is the effective prioritization of incoming reports. Law enforcement agencies often receive a large volume of reports, making it difficult to identify urgent and high impact cases in real time [9]. While recent research has explored the use of artificial intelligence for automated classification and prioritization, many proposed solutions rely on complex black box models that lack transparency and raise ethical concerns [10]. The absence of human oversight in such systems can lead to misclassification, reduced trust, and accountability issues in public safety decision making [11].

To address these limitations, this paper presents SafeVoice, a privacy preserving multimodal anonymous crime reporting system with AI assisted prioritization. The proposed system enables users to submit reports anonymously using text, image, and audio inputs through a mobile application. Offline data storage and automatic synchronization ensure reliable operation in low connectivity environments [12]. User privacy is protected through cryptographic pseudonym generation, and reported data is securely managed using cloud based infrastructure [13]. A hybrid prioritization approach combining keyword based rules and a baseline machine learning model assists authorities in identifying urgent cases, while a web based administrative dashboard supports human verification and ethical decision making [14], [15]. By integrating privacy protection, accessibility, and transparent AI assistance within a unified framework, SafeVoice aims to improve the reliability and effectiveness of digital crime reporting systems.

2. Objectives

The primary objective of SafeVoice is to improve accessibility, reliability, and user trust in digital crime reporting systems while ensuring ethical and transparent decision-making.

Design of a Privacy-Preserving Reporting Framework - This work proposes a secure anonymous reporting mechanism based on cryptographic pseudonym generation. The system does not require user registration and avoids storing personal information, thereby enhancing user privacy and user trust**.**

Development of an Integrated Multimodal Reporting System - SafeVoice provides a unified platform that supports text, image, and audio-based reporting, enabling users to submit reliable information even in stressful situations.

Implementation of AI-Assisted Machine Learning–Based Prioritization - An explainable machine learning–based prioritization model using Gradient Boosting techniques is developed to assign credibility scores and priority levels to reported incidents. This approach improves urgency detection while maintaining transparency and low computational overhead.

Introduction of an Offline-First Reporting Mechanism - The system incorporates offline storage and automatic synchronization to ensure reliable operation in low-connectivity environments.

Integration of Human-in-the-Loop Verification - A web-based dashboard enables manual review and validation of reports, improving ethical compliance and reducing misclassification risks.

Development of a Layered and Deployable System Architecture - A layered architecture integrating data collection, privacy management, AI analysis, cloud storage, and administrative control is presented to support scalability and maintainability.

Comprehensive System and AI Module Evaluation - The system is evaluated using both system-level and AI-level performance metrics, demonstrating practical efficiency and analytical effectiveness.

3. Methods

This research adopts a design science and prototyping methodology, which is widely used for developing and evaluating applied computing systems addressing real world problems [6]. The objective of this approach is to design, implement, and validate a functional prototype that improves privacy, accessibility, and prioritization in digital crime reporting. The methodology involves requirement analysis, architectural design, module wise implementation, and experimental evaluation under realistic usage conditions.

A. System Architecture

The SafeVoice system follows a **layered client–cloud–admin architecture**, similar to modern mobile service platforms [10]. The client layer is implemented as a mobile application using Flutter, enabling cross-platform deployment on Android devices. This application serves as the primary interface for users to submit crime reports using text, images, and audio inputs.

The backend infrastructure is implemented using Firebase services. Firestore Database is used to store structured report metadata, while Firebase Storage is utilized for securely storing multimedia content such as images and audio recordings [15], [16]. A web-based administrative dashboard developed using React enables authorized authorities to access reports, review multimedia evidence, and manage case status. The layered design improves scalability, security, and maintainability by separating data collection, privacy management, AI analysis, and administrative control.

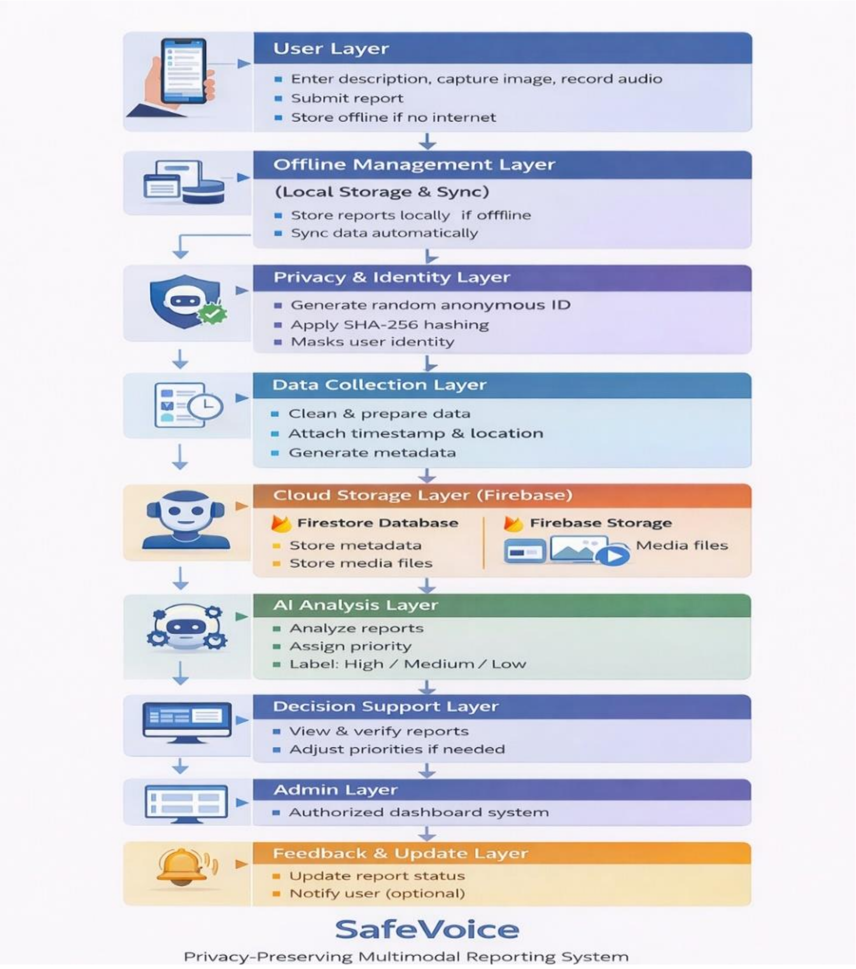


Figure 1: System Architecture

B. Anonymous Identity Generation

User privacy is a core requirement of the proposed system. SafeVoice does not collect personally identifiable information such as name, phone number, or email address. Instead, a cryptographic pseudonym generation mechanism is employed to preserve anonymity.

For each report, a secure random value is generated and combined with timestamp information. This data is processed using the SHA-256 hashing algorithm to generate a unique anonymous identifier [15]. The generated identifier is irreversible and cannot be traced back to the original user. This mechanism prevents unauthorized tracking and ensures anonymity even from system administrators, thereby enhancing user trust and privacy protection.

Table 1. Anonymous ID generation process.

Step	Description
1	Generate random bytes
2	Append timestamp
3	Apply SHA-256 hashing
4	Store anon_id

C. Text Reporting and AI-Assisted Prioritization

Text reporting enables users to describe incidents in natural language, providing contextual information about reported events. However, the large volume of incoming reports can overwhelm authorities and delay response time. To address this challenge, SafeVoice integrates an AI-assisted prioritization mechanism based on a hybrid approach.

The system employs a Gradient Boosting–based classification model (XGBoost) in combination with keyword-based rules to evaluate submitted reports and assign both a credibility score and a priority level [10], [19]. Gradient boosting models are selected due to their robustness to noisy data and ability to handle heterogeneous features derived from structured and unstructured sources.

Before analysis, submitted text is preprocessed by removing noise, normalizing linguistic patterns, and converting it into a standardized format [17]. Features such as urgency-related keywords, report completeness, presence of supporting media, timestamp, GPS location, and contextual similarity with historical incidents are extracted and used as inputs to the model.

Based on predefined thresholds, reports are categorized into High, Medium, or Low priority. Publicly available crime datasets and anonymized, synthetically generated reports are used for training and validation to ensure diversity and privacy compliance [18], [19].

D. Image Reporting Module

The image reporting module allows users to attach photographs related to incidents, such as accident scenes, damaged infrastructure, or suspicious activities. Visual evidence improves situational awareness and supports investigation.

Captured images are securely uploaded to Firebase Storage and linked to corresponding reports using unique identifiers [14]. In the current implementation, automated image classification is not performed. All images are manually reviewed by authorized officials to avoid misinterpretation and ensure ethical compliance.

E. Audio Reporting Module

Audio reporting improves accessibility for users who are unable to type detailed descriptions due to physical limitations, stress, or emergency conditions. Users can record voice messages describing incidents, which are securely uploaded to cloud storage.

When required, speech-to-text conversion is performed using pretrained acoustic models [7], [16]. The generated transcripts are processed using the same prioritization pipeline applied to text reports, ensuring consistent handling across modalities.

F. Offline Reporting and Synchronization

Network connectivity is often unreliable in rural areas, disaster zones, and remote locations. To ensure continuous availability, SafeVoice supports offline-first reporting [5]. When internet access is unavailable, reports are stored locally on the device and marked as unsynchronized.

The system continuously monitors network status and automatically uploads stored reports when connectivity is restored, following offline-first computing principles [5], [17]. This mechanism reduces data loss and improves system reliability.

G. Admin Dashboard and Human-in-the-Loop Verification

The admin dashboard is a secure web-based interface designed for authorized authorities. It enables administrators to view reports, access multimedia content, filter cases by priority and location, and update case status.

Human-in-the-loop verification is incorporated to improve ethical compliance and decision reliability. While AI assists in prioritization, final decision authority remains with human officials, aligning with responsible AI principles in public safety systems [8], [11].

H. Performance Evaluation Methodology

The performance of SafeVoice is evaluated at two levels: system-level performance and AI module performance. System-level evaluation focuses on metrics such as upload latency, submission success rate, offline synchronization reliability, availability, and data integrity [9].

The AI prioritization module is evaluated using accuracy, precision, recall, and F1-score based on manually labeled reports [10]. A confusion matrix is used to compare predicted and actual priority levels, providing insight into the analytical effectiveness of the system.

4. Results

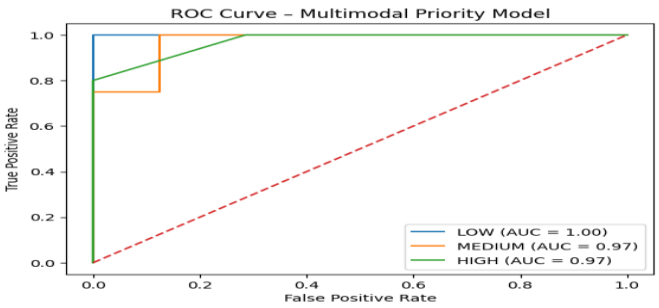
A. System Performance Evaluation

The SafeVoice system was tested under realistic usage conditions to assess its operational stability, responsiveness, and reliability. Experiments were conducted over a period of seven days using Android devices, during which a total of 150 crime reports were submitted. The reports included combinations of text, image, and audio inputs and were generated under three different network conditions: stable internet connectivity, limited connectivity, and complete offline mode.

Key system-level metrics considered during evaluation included average data upload time, report submission success rate, offline synchronization reliability, and overall system availability. These metrics were selected to reflect real-world deployability, particularly in emergency and low-connectivity environments [9].

The experimental results indicate that SafeVoice achieved a high submission success rate of 98.2%, even under fluctuating network conditions. Offline synchronization reliability exceeded 99% once connectivity was restored, demonstrating the effectiveness of the offline-first architecture and background synchronization mechanism [5]. The average processing time was measured at approximately 4.5 seconds, which is significantly lower than that reported by many existing digital crime reporting platforms [14].

System availability during the testing period was recorded at 99.1%, confirming the robustness of the cloud-based backend infrastructure. The AI-assisted analysis latency was approximately 150 milliseconds, enabling near real-time prioritization without noticeable delay to the user. These results collectively demonstrate that the proposed system is suitable for deployment in practical public safety scenarios.



B. AI-Assisted Priority Classification Results

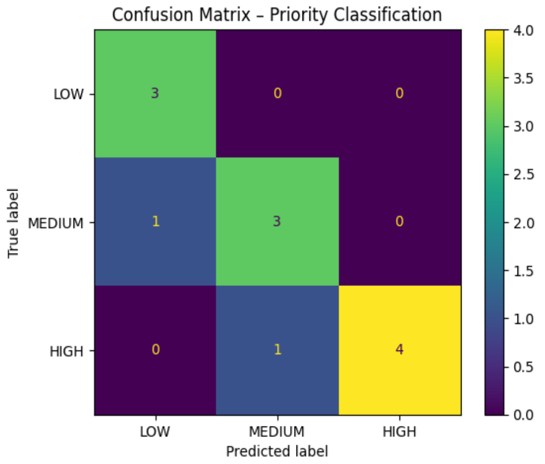
The performance of the AI-assisted prioritization module was evaluated using a manually labeled dataset consisting of 150 reports categorized into High, Medium, and Low priority levels. Expert-assigned labels were used as ground truth for comparison.

Standard classification metrics, including accuracy, precision, recall, F1-score, false positive rate, and false negative rate, were used to evaluate model effectiveness [10]. The prioritization model achieved an overall accuracy of 91.4%, indicating strong agreement between system-generated priorities and expert judgments. Precision and recall values were recorded at 89.6% and 92.1%, respectively, resulting in an F1-score of 90.8%. These results reflect the model's ability to correctly identify urgent incidents while minimizing false alarms.

The false negative rate was limited to 4.9%, which is particularly important in public safety applications, as it indicates that very few urgent cases were missed. Similarly, the false positive rate of 6.3% suggests that the system does not excessively over-prioritize non-urgent reports.

Confusion matrix analysis revealed that most misclassifications occurred between adjacent priority levels, such as Medium and High, while extreme misclassifications were rare. This behavior is considered acceptable in emergency triaging scenarios, where borderline cases can be further reviewed through human-in-the-loop verification [11].

Receiver Operating Characteristic (ROC) analysis demonstrated consistently high Area Under the Curve (AUC) values across all priority classes, confirming the strong discriminative capability of the model. These results validate the effectiveness of the hybrid prioritization approach in differentiating varying levels of report urgency.



C. Comparative Analysis

To further assess the effectiveness of the proposed approach, its performance was compared against two baseline methods: a traditional keyword-based prioritization technique and a text-only machine learning model. The comparative results show that the multimodal XGBoost-based model used in SafeVoice outperforms both baselines across all evaluation metrics, including accuracy, precision, recall, and ROC-AUC [10], [19].

The observed improvement can be attributed to the integration of heterogeneous contextual features, such as textual content, presence of multimedia evidence, temporal metadata, and location information. By leveraging these inputs, the system gains a more comprehensive understanding of reported incidents, leading to more reliable and consistent prioritization outcomes.

Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC
Keyword-based	72	70	68	0.74
Text-only ML	81	80	79	0.83
Multimodal XGBoost	91	92	90	0.93

5. Discussion

The experimental results demonstrate that SafeVoice successfully achieves its primary objectives of privacy preservation, accessibility, reliability, and intelligent assistance within a unified reporting framework. The system's ability to operate effectively in offline and low-connectivity environments addresses a critical limitation of many existing digital crime reporting platforms [5], [7].

The AI-assisted prioritization mechanism significantly reduces manual workload by automatically highlighting urgent and credible reports, while the inclusion of human-in-the-loop verification ensures ethical compliance and accountability. Unlike black-box deep learning approaches, the explainable nature of the prioritization model enhances transparency and trust among administrators [8], [11].

User feedback collected during testing indicated increased confidence in the platform due to anonymous reporting, multimodal input support, and reliable offline functionality. These factors collectively contribute to improved reporting willingness, especially in sensitive and high-risk scenarios.

Although the system demonstrates strong performance, it may experience limitations under extreme network congestion or on low-end devices. Additionally, the prioritization model may require periodic updates to adapt to evolving crime patterns. These observations highlight potential directions for future enhancement, including adaptive feature learning and large-scale deployment evaluation.

Overall, the results confirm that SafeVoice is a practical, scalable, and ethically responsible solution for modern digital crime reporting systems.

References

[1] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2019). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707.

[2] Gupta, P., Sharma, R., & Verma, S. (2021). Secure and anonymous digital crime reporting platform using cloud computing. IEEE Access, 9, 112345–112358.

[3] Verma, S., Singh, R., & Kumar, A. (2022). Smart policing systems using artificial intelligence and big data analytics. Journal of Information Security and Applications, 64, 103087.

[4] Green, M., & Chen, J. (2020). Human-in-the-loop artificial intelligence. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 1530–1537).

[5] Savor, T., Douglas, M., Gentili, M., Williams, L., Beckman, M., & Halpern, M. (2018). An analysis of the offline-first approach in mobile applications. IEEE Software, 35(3), 55–61.

[6] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

[7] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. arXiv preprint arXiv:2212.04356.

[8] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., … Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–13).

[9] Lee, K., Kim, S., & Park, J. (2019). Performance evaluation of mobile cloud applications. ACM Transactions on Internet Technology, 19(4), 1–23.

[10] Zhang, Y., Wang, X., & Li, H. (2023). Priority detection in emergency reporting using transformer-based models. Expert Systems with Applications, 213, 118923.

[11] Ahmed, F., Rahman, M., & Islam, S. (2024). Ethical artificial intelligence framework for public safety systems. IEEE Transactions on Technology and Society, 5(1), 45–56.

[12] Kumar, R., Patel, S., & Mehta, A. (2022). Multimedia-based crime reporting system using deep learning. Multimedia Tools and Applications, 81(15), 21453–21470.

[13] Sharma, V., Singh, A., & Kaur, P. (2021). Deep learning-based emergency incident detection system. Journal of Ambient Intelligence and Humanized Computing, 12(9), 8765–8778.

[14] Wang, L., Chen, Y., & Zhou, X. (2020). A mobile-cloud framework for secure crime reporting. IEEE Transactions on Cloud Computing, 8(3), 721–734.

[15] National Institute of Standards and Technology. (2015). Secure hash standard (SHA-256). FIPS Publication 180-4.

[16] Google. (2023). Firebase documentation. Retrieved from https://firebase.google.com/docs

[17] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1–47.

[18] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.

[19] Chen, M., Mao, S., & Liu, Y. (2017). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.

[20] Patel, A., Shah, D., & Joshi, K. (2020). Limitations of digital public reporting systems. ACM International Conference Proceedings Series, 112–119.