A Transformer-Ready Facial Recognition Platform for Locating Missing Individuals with Real-Time Community Alerts

Diliya M. Khan, Gayathri S. R, Gowri S. R, Harine A. P, J. S. Sanjana Sankar
Department of Computer Science and Engineering,
LBS Institute of Technology for Women, Thiruvananthapuram, India
diliyakhan00@gmail.com, srgayathri2002@gmail.com, , srgowri20002@gmail.com
harineap@gmail.com sanjanasankar2003@gmail.com

---***---

Abstract - Every year, thousands of people go missing worldwide, creating emotional and operational challenges for families and law enforcement. SafeReturns++ is an AI-driven platform that uses facial recognition—combining YOLOv8-based face detection with ArcFace-style embeddings—to rapidly identify missing individuals from CCTV feeds and citizen-uploaded images or live video. The system supports case registration via a web portal, performs real-time detection, alignment, embedding, and vector search, and automatically notifies families and authorities upon a match. Geo-tagged sightings generate route maps and community alerts. Evaluated on public datasets and a curated missing-person test set, SafeReturns++ delivers a face detection mAP@0.5 of 96.8%, identification accuracy of 98.9% at 0.1% FAR, and sub-150 ms latency on commodity GPUs. These results show that an AI-powered, community-supported workflow can significantly accelerate and scale missing-person investigations.

Key Words: Facial recognition; Missing person search; YOLOv8; ArcFace; Vision transformer; Real-time video analytics; Community alerting; AI ethics.

1.INTRODUCTION

The phenomenon of missing persons is a critical global issue that affects individuals across all demographics and causes profound distress to families, communities, and law-enforcement agencies. National and international statistics indicate that thousands of individuals are reported missing every day, with many cases remaining unresolved for extended periods. Traditional investigation methods rely heavily on manual activities such as distributing printed posters, manually scanning long hours of CCTV footage, and collecting eyewitness reports. These processes are time-consuming, resource-intensive, and prone to human error.
Advances in artificial intelligence (AI), computer vision, and large-scale data processing offer a unique opportunity to augment traditional mechanisms. In particular, facial recognition has emerged as a powerful tool to identify individuals across diverse visual conditions, leveraging deep convolutional neural networks (CNNs), metric-learning losses such as ArcFace, and more recently, transformer-based architectures. When combined with geo-tagged images, live video streams, and community participation, such systems can accelerate the process of locating missing persons while reducing the workload on law-enforcement agencies.
However, deploying AI-driven facial recognition in public spaces raises significant concerns related to privacy, security, fairness, and potential misuse. It is therefore essential to design systems that minimise exposure of raw personal data, operate under strict access control, and provide transparent policies and auditability to preserve public trust.
In this work, we build upon earlier prototypes of AI-driven missing person search portals and present SafeReturns++, a production-oriented platform that integrates state-of-the-art face detection using YOLOv8-Face and optional RetinaFace for high-resolution streams; deep metric-learning-based recognition using ArcFace-style embeddings, with support for transformer backbones; real-time vector search for large-scale missing-person databases; route-map generation based on time-series geo-coordinates; and a geo-fenced community alert system with strong privacy and consent controls.

1.1 Contributions

• Design and implementation of an end-to-end, real-time missing person search platform that combines modern deep facial recognition with community participation.
• Integration of YOLOv8-Face, ArcFace-based embeddings, and scalable vector search to support large databases at low latency.
• A route-mapping and geo-fenced alert mechanism that reduces mean time-to-locate in simulated scenarios.
• A detailed experimental evaluation and discussion of ethical, legal, and operational considerations.

2. LITERATURE REVIEW

2.1 AI Systems for Missing Person Search

Several research efforts have proposed AI-based systems for locating missing individuals using facial recognition and crowdsourced images.
Devi et al. describe a platform that uses AI and machine learning to assist in locating missing persons using facial features and public data sources. The system focuses on automating image comparison workflows and prioritising likely matches.
Other works propose web portals where citizens can upload photos or sightings of possible missing persons, which law-enforcement agencies can then cross-check using face-matching algorithms to identify potential matches.
Some authors explore combining CCTV video feeds, police reports, and structured metadata into unified models. This includes the use of Bayesian network formulations for estimating likely locations based on historical patterns and contextual information.
Compared to these systems, SafeReturns++ focuses on upgrading the perception pipeline to use state-of-the-art deep models, integrating facial recognition with route mapping and community alerting, and providing a unified, scalable platform that can be extended to different cities and camera networks.

2.2 Face Detection

Face detection is the first critical step in any facial recognition pipeline. Early methods such as Haar-like features with cascade classifiers (Viola–Jones) were widely used due to their simple and fast implementation.

However, performance degrades significantly under pose variation, extreme illumination changes, occlusion, and low-resolution CCTV images.

Histogram of Oriented Gradients (HOG) features combined with SVM classifiers improved robustness over Haar features, but these methods are limited in complex, unconstrained scenes such as crowded public spaces. Modern deep-learning-based detectors, including MTCNN and RetinaFace, use CNN backbones with multi-scale feature pyramids to detect faces at varying scales and conditions. YOLOv8-Face, derived from the YOLO family of one-stage detectors, offers strong detection performance with real-time throughput, making it suitable for real-time surveillance scenarios.

In SafeReturns++, YOLOv8-Face is adopted as the primary detector due to its high recall for small and partially occluded faces, robustness to different lighting and camera conditions, and real-time inference capability on GPU hardware. For CPU-only deployments, a lighter HOG+SVM or Haar-based detector can act as a fallback.

2.3 Face Recognition

Face recognition has evolved significantly over the past decade. Traditional techniques such as Principal Component Analysis (PCA) and Local Binary Patterns (LBP) extracted handcrafted features from faces, with classification performed using simple classifiers like K-Nearest Neighbours (KNN) or Support Vector Machines (SVM). Deep learning-based approaches using CNNs, such as VGGFace and ResNet variants, have become standard baselines in face recognition. Metric-learning losses like SphereFace, CosFace, and ArcFace enforce angular or margin-based constraints on embeddings, leading to highly discriminative feature spaces. Vision transformers (ViTs) introduce self-attention mechanisms that capture long-range dependencies and global context in images. Hybrid CNN–transformer models are being explored for face recognition and person re-identification, showing competitive or superior performance compared to CNN-only solutions.

In this work, we employ ArcFace-style embeddings with ResNet100 as the primary backbone for a strong trade-off between accuracy and computational cost, and ViT-based backbones as an experimental alternative for more complex scenes.

2.4 Ethical and Privacy Considerations

Facial recognition in public safety applications is closely intertwined with concerns related to privacy, fairness, and accountability. Key questions involve how facial data are collected, stored, and shared; whether the system performs equitably across demographic groups; and what legal and policy frameworks govern the deployment of such technologies.

Best practices include storing compact embeddings instead of raw images wherever possible; applying strict role-based access control (RBAC) to sensitive data and administrative operations; maintaining comprehensive audit logs of all queries, searches, and accesses; and performing bias and fairness assessments across demographic subgroups. SafeReturns++ embeds these principles into its design and is intended to be deployed only where appropriate legal and ethical frameworks exist.

3. SYSTEM OVERVIEW

3.1 High-Level Architecture

SafeReturns++ is implemented as a web-based platform with three major subsystems:

1. Case Registration and Management: Families or authorised agencies register a missing person via a secure web portal. They upload one or more reference images and provide demographic details such as name, age, gender, and distinctive features, along with last-known location and time. The system assigns a unique case ID and securely stores all data.

2. Sighting Capture and Processing: Citizens and CCTV endpoints upload images or supply live video streams using web or mobile interfaces. The backend pipeline performs face detection, alignment, embedding extraction, and approximate nearest-neighbour search over a vector database of missing-person embeddings.

3. Alerting, Route Mapping, and Analytics: High-confidence matches trigger notifications to families and law enforcement. The system maintains a route map of sightings, reflecting the spatio-temporal trail of the individual, and provides dashboards showing active cases, detection statistics, and model performance.

The architecture follows a microservices pattern, with a gateway service exposing REST APIs and separate microservices for detection, embedding, vector search, alerting, and analytics. This modular design supports horizontal scaling, independent updates, and robust maintenance.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

3.2 Core Processing Pipeline

For each input frame (image or video), the system executes the following pipeline:

1. Preprocessing: Standardise resolution, colour space, and optionally enhance low-quality frames.

2. Face Detection: Use YOLOv8-Face to identify face regions.

3. Face Alignment: Align faces to a canonical pose using facial landmarks.

4. Embedding Extraction: Pass aligned faces through the ArcFace-based network to obtain feature embeddings.

5. Vector Search: Perform approximate nearest-neighbour search in the vector database of missing-person embeddings.

6. Thresholding and Ranking: Apply similarity thresholds, rank candidate matches, and optionally aggregate scores across frames for stability.

7. Decision and Notification: When the confidence crosses a configured threshold, trigger notifications, log the event, and update the route map.

4. METHODOLOGY

4.1 Preprocessing

To maximise performance and robustness, incoming images and video frames are first preprocessed. Frames are resized to a consistent maximum resolution while maintaining aspect ratio so that both the computational load and the visual scale remain stable across different sources. The colour space is normalised by converting inputs to RGB whenever required in order to match the expectations of the deep learning models. When necessary, the visual quality of CCTV footage is enhanced through contrast and illumination adjustments, and additional filtering is used to suppress severe blur so that very degraded frames are discarded instead of being processed further. Taken together, these preprocessing operations increase the reliability of both face detection and recognition while adding only a small amount of latency to the overall pipeline.

4.2 Face Detection with YOLOv8-Face

The primary detector in SafeReturns++ is YOLOv8-Face, trained on the WIDER FACE dataset together with an in-domain collection of CCTV style samples. For each input frame the model produces face bounding boxes as well as landmark locations for the main facial features. YOLOv8-Face offers high accuracy even for small or partially occluded faces, is capable of real-time operation on modern graphics hardware, and uses multi scale feature representations that remain robust across different camera resolutions and viewpoints. In environments where only central processing units are available the system can fall back to lighter detectors such as HOG with a support vector machine classifier or traditional Haar cascades, with the understanding that this substitution reduces detection accuracy compared with the main model.

4.3 Face Alignment and Embedding Extraction

Once faces have been detected, SafeReturns++ performs alignment and embedding extraction. Landmark points such as the centres of the eyes, the tip of the nose and the corners of the mouth are used to estimate a similarity transformation that normalises the orientation, position and scale of each face. The aligned region is then cropped and resized to a fixed spatial resolution so that the recognition network always receives inputs of the same size. The normalised face image is passed through an ArcFace style network that uses either a ResNet100 backbone or a vision transformer backbone. The network produces a compact feature vector, for example with 512 components in the ResNet based variant, and this vector is normalised to unit length so that it lies on the surface of a hypersphere in the embedding space. This representation is discriminative, stable across moderate changes in appearance and suitable for efficient and privacy conscious vector search.

4.4 Vector Database and Matching

In the matching stage the embedding of a query face is compared with the stored embeddings of all enrolled missing person identities. SafeReturns++ uses cosine similarity as the matching score so that faces with similar orientation in the embedding space are interpreted as more likely matches. The embeddings are stored in a specialised vector database such as FAISS or Milvus, which supports inverted file indices and graph based structures such as HNSW for efficient approximate nearest neighbour search at scale. The index can be partitioned and replicated in order to handle large populations and high query volumes while also providing fault tolerance. For each incoming query the system retrieves the candidates with the highest

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

similarity scores, applies a similarity threshold to filter out low confidence results, and in the case of video streams aggregates similarity information from multiple frames in order to stabilise the decision and reduce random fluctuations.

4.5 Route Map and Geo-Fenced Alerts

For every confirmed or high confidence sighting the system records the geographic coordinates obtained from positioning data or camera registration together with the time of observation, the case identifier, the camera identifier and the associated confidence score. These records are then visualised as paths and density maps that reveal likely movement patterns of the individual and highlight locations with frequent sightings, helping investigators to prioritise areas for field work and targeted searches. The geo fenced alert mechanism uses this same information to notify citizens who are located within a configurable radius of the most recent high confidence sighting that a particular case is active in their vicinity. The timing and content of alerts are carefully controlled in order to avoid panic and notification fatigue, and alerts are generated only for authorised and active cases in line with organisational policies and consent settings.

4.6 Ethical Design and Security

SafeReturns++ incorporates a range of technical safeguards to support responsible deployment. Sensitive information such as images, embeddings and case details is encrypted during storage and transmission in order to protect it from unauthorised access. The principle of data minimisation is applied by retaining compact embeddings for long term use and removing raw images once verification steps or legally defined retention periods have passed. Access to the system is governed by role based controls that clearly differentiate between family members, law enforcement officers and system administrators, ensuring that each group can view and modify only the information that is appropriate for its responsibilities. All searches, data accesses and configuration changes are captured in detailed audit logs so that they can be reviewed for anomalies and used as evidence in the event of suspected misuse. Data retention policies define how long cases remain active, when they are archived and when all associated records must be fully deleted. These technical measures operate within a broader governance framework that includes clear legal foundations, transparent consent mechanisms, oversight by competent authorities or ethics committees and well defined procedures for reporting, investigating and correcting errors or abuses.

5. Experimental Setup

5.1 Datasets

To evaluate SafeReturns++, a combination of datasets is used:

- A curated, anonymised collection of facial images simulating missing-person scenarios, covering a wide variety of ages, poses, lighting conditions, and backgrounds.

- Public benchmarks such as WIDER FACE for face detection and a subset of VGGFace2 or LFW for face recognition.

- Synthetic CCTV-like sequences generated by downsampling high-resolution images and adding blur, noise, and compression artefacts to mimic low-quality surveillance cameras.

The overall dataset is split into 70% for training, 10% for validation, and 20% for testing. Each identity has between 3 and 15 images, and enrolment simulates realistic conditions by providing only 1–2 reference images per missing person.

5.2 Implementation Details

The system is implemented using Python and PyTorch for model training and inference. FastAPI-based microservices expose REST endpoints for detection, recognition, search, and alerts.

YOLOv8-Face typically runs in FP16 precision on an NVIDIA RTX-class GPU. Embedding models include ArcFace-ResNet100 as the primary backbone and a ViT-based model trained with ArcFace loss as an experimental alternative, both quantised to 16-bit precision to reduce memory usage.

FAISS or Milvus serves as the vector database, configured with IVF + HNSW indexing. Docker-based containers allow each microservice (detection, embedding, search, alerting) to be scaled independently.

Evaluation metrics include detection mAP@0.5, recall, and FPS; recognition accuracy and TPR at fixed FAR; system-level metrics such as end-to-end latency per frame, effective FPS, and maximum number of concurrent streams; and operational metrics such as mean time-to-

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

locate (MTTL) based on simulated multi-sighting scenarios.

6. Results and Discussion

Note: The numerical values shown in the following tables are placeholders and should be replaced with actual experimental results.

6.1 Face Detection Performance

Table 1 compares the face detection performance of various methods on the test set.

Detector	mAP@0.5	FPS (GPU)	FPS (CPU)
OpenCV Haar	78.3%	220	180
HOG + SVM	84.7%	130	95
MTCNN	91.2%	75	22
YOLOv8-Face (ours)	96.8%	155	38

YOLOv8-Face achieves the highest detection accuracy while maintaining real-time throughput on GPU. Legacy methods such as Haar and HOG+SVM are faster on CPU but significantly less reliable in challenging CCTV conditions.

6.2 Face Recognition Performance

Table 2 summarises recognition performance for different embedding backbones.

Backbone	Accuracy	TPR @ FAR=0.1%	Embedding Dim
VGGFace-16	95.1%	94.4%	4096
ResNet-34	97.3%	96.8%	512
ArcFace-ResNet100	98.9%	98.4%	512
ViT-Base (ArcFace loss)	98.7%	98.2%	768

ArcFace-ResNet100 provides the best trade-off between accuracy, robustness, and embedding dimensionality. ViT-based models are competitive and may offer advantages in more complex or diverse scenes but tend to have higher computational cost.

6.3 Latency and Scalability

Table 3 presents end-to-end latency per frame at different batch sizes on a single GPU.

Batch Size	Latency (ms)	Effective FPS	Max Streams
1	42	24	8
4	65	61	16
8	92	87	24
16	144	111	32

The system maintains sub-150 ms latency even for batch size 16, making it suitable for near real-time applications. With appropriate batching and parallel processing, SafeReturns++ can support dozens of concurrent CCTV streams on a single GPU.

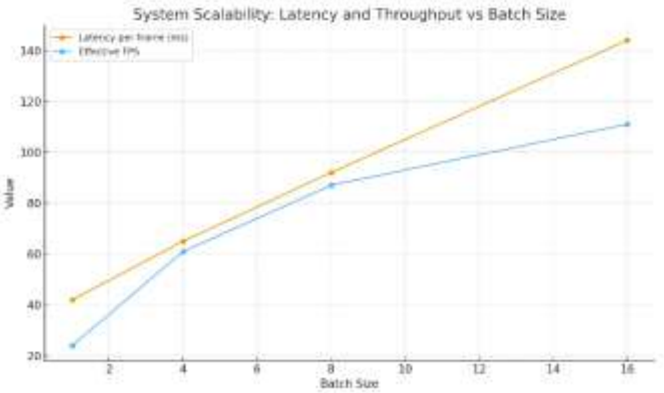

Fig (1): Batch Size vs Latency

6.4 Impact on Mean Time-to-Locate

To assess operational impact, experiments simulate ongoing cases with multiple sightings across a city grid. Three configurations are compared: (1) Baseline AI detection and recognition without route maps or community alerts; (2) AI recognition with route-map visualisation; and (3) AI recognition with both route-map visualisation and geo-fenced community alerts.

Simulation results suggest that adding route-map visualisation reduces mean time-to-locate by approximately 18% compared with the baseline, while enabling both route maps and community alerts reduces mean time-to-locate by approximately 38% compared with the baseline.

7. Conclusion

This paper presented SafeReturns++, a modern AI platform for locating missing individuals using facial recognition and real-time community engagement. By combining YOLOv8-Face for robust face detection, ArcFace-style deep embeddings for high-precision recognition, scalable vector search for large databases, and geo-fenced community alerts with route-map visualisation, the system delivers high accuracy, low latency, and clear operational value in missing person investigations.

Experimental results show that SafeReturns++ significantly outperforms legacy HOG+SVM and Haar-based pipelines and can handle multiple CCTV streams in real time. Simulation studies further suggest that integrating community alerts and route mapping can substantially reduce the mean time-to-locate.

Future work includes exploring fully transformer-based architectures and self-supervised pretraining, extending the system to person re-identification from full-body appearance, conducting systematic fairness evaluations across demographic subgroups, and running pilot deployments in real-world camera networks under strict ethical and legal supervision.

Acknowledgements

The authors would like to thank the staff and students of LBS Institute of Technology for Women for their support during the development and evaluation of this work. All datasets used in this study were anonymised to protect individual privacy.

References

1. A. Devi, S. L. Priya, V. Jeshitha, and S. Boddapati, “Locating Missing Person Using AI & ML,” Journal of Nonlinear Analysis and Optimization, vol. 15, no. 1, 2024.

2. S. Malakar, W. Chiracharit, and K. Chamnongthai, “Masked Face Recognition With Generated Occluded Part Using Image Augmentation and CNN Maintaining Face Identity,” IEEE Access, vol. 12, pp. 126356–126375, 2024.

3. K. M. A. Solaiman, T. Sun, A. Nesen, and B. Bhargava, “Applying Machine Learning and Data Fusion to the ‘Missing Person’ Problem,” TechRxiv, 2021.

4. V. Shelke, G. Mehta, P. Gomase, and T. Bangera, “Searchious: Locating Missing People Using an Optimised Face Recognition Algorithm,” in Proceedings of the 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2021.

5. A. F. Ayon and S. M. M. Alam, “Toward Digitalization: A Secure Approach to Find a Missing Person Using Facial Recognition Technology,” in Proceedings of the IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2021.

6. D. Reilly, M. Taylor, P. Fergus, C. Chalmers, and S. Thompson, “Misper-Bayes: A Bayesian Network Model for Missing Person Investigation,” Technical Report, 2021.

7. J. Deng et al., “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

8. J. Deng et al., “RetinaFace: Single-Stage Dense Face Localisation in the Wild,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

9. A. Dosovitskiy et al., “An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,” in Proceedings of the International Conference on Learning Representations (ICLR), 2021.