IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

SACHET: A Crowdsourced Missing Child Alert System with Predictive Location Analytics

Prof. Moushmee Kuri[1], Harshil Buch[2], Vijay Bhilare[3], Dadasaheb Bhosure[4], Manasi Gaur[5]
Assistant Professor, CSE, MIT ADT University Pune, India[1]
Students, CSE, MIT ADT University, Pune, India[2-5]

Abstract: *Child safety remains one of the most pressing humanitarian concerns in India. According to the National Crime Records Bureau (NCRB) Crime in India 2022 report, 83,350 children were reported missing in a single year, with 62,946 of them being girls — approximately 75 per cent of total missing cases. Even though we have applications like TrackChild and Khoya-Paya , due to many factors , it still fails , due to late reporting, fragmented communication and less awareness between people .*

This paper presents SACHET (System for Alert, Community Help, and Emergency Tracking),covering the gap by enabling rapid crowdsourced reporting of missing children , real-time sighting documentation with automated multi-channel broadcast alerts via Telegram , Radio, Television etc. Sachet is a full stack web application built on Python Flask framework. The system uses a LightGBM ensemble model which is trained on demographic, temporal, and spatial case features to generate the estimation using probability-weighted location. It iteratively refines these predictions as community sighting data accumulate, using the Haversine great-circle distance formula for spatial triangulation.

SACHET further incorporates risk-zone analytics that cluster historical case data to produce heatmaps of geographic hotspots, enabling proactive policing. Media persistence is handled via Cloudinary, while case data is stored in PostgreSQL for production reliability. Administrative access is restricted by role-based authentication, and the system supports dual deployment on serverless (Vercel) and container-based (Render) environments with Redis-backed ML result caching.

The architecture, algorithmic design, empirical problem context, and comparative analysis with existing systems — including India's TrackChild and the United States AMBER Alert — are discussed in detail. The findings suggest that SACHET's crowdsourced-predictive hybrid model can substantially reduce the mean time to recovery for missing children by enabling faster, geographically targeted search efforts.

Keywords: Missing Child Alert System, Predictive Analytics, LightGBM, Geospatial Intelligence, Crowdsourcing, Flask, Child Safety, India, NCRB, Haversine Formula, Telegram Broadcast

I. INTRODUCTION

When a child disappears, the first starting hours are very crucial, families whose child goes missing are statistically the most critical for survival and safe recovery. Research by the Federal Bureau of Investigation in the United States found that in abduction-murder cases, approximately 74 per cent of child victims are killed within the first three hours of their disappearance (Hanfland et al., 1997; Griffin et al., 2021). This figure underscores why rapid, community-wide alert mechanisms are not merely desirable but essential.

Studies show that in India the rate of missing children is escalating. The NCRB's Crime in India 2022 report recorded 83,350 missing children in that year alone — a 7.5 percent increase over 2021 and representing a near-1,000 percent increase in reported cases over the fourteen-year period from 2008 to 2022 (The Print, 2024; The News Minute, 2024). As of the end of 2022, 47,313 children remained untraced, of whom 71.4 per cent — roughly 33,798 — were girls. West Bengal and Madhya Pradesh consistently account for the largest proportion of missing children nationally, with West Bengal alone reporting 12,455 missing children in 2022 (Dainik Jagran, 2026). The link between missing girls and human trafficking, child labour, and early forced marriage is well-documented but

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 18, April 2026 Impact Factor: 8.2

systematically under-prosecuted, with only 424 trafficking cases formally registered in 2022 despite the scale of disappearances [1] "Missing children and trafficking concerns in India," The News Minute, 2024.

Government-initiated platforms — primarily the Ministry of Women and Child Development's TrackChild portal and its citizen-facing Khoya-Paya module — have provided a structured mechanism for recording missing and found children since 2012. The portal recorded 139,443,896 website hits by December 2017, demonstrating citizen awareness (PIB, 2018). However, these portals mainly focus on data-recording systems rather than real-time alert mechanisms; they do not give notifications to community members in a child's vicinity.

Sachet fills the gap and addresses these gaps with a muilti-pronged approach: immediate public broadcast of missing-child details at the moment of case registration; a community sighting portal that geotags reported observations and appends them to the case record; ML-powered location prediction that synthesises case demographics, abduction context, and sighting history into a probability-weighted coordinate estimate; and administrative analytics that identify and visualise high-risk zones for proactive law-enforcement focus.

The paper is made as follows. Section 2 reviews the relevant literature on alert systems, predictive analytics, and crowdsourcing in child safety. Section 3 describes the problem statement and motivation. Section 4 details the system architecture and design. Section 5 explains the machine learning methodology. Section 6 outlines the key features and implementation. Section 7 presents the technology stack. Section 8 discusses deployment and infrastructure. Section 9 compares SACHET with existing systems. Section 10 addresses limitations and future directions. Section 11 concludes the paper.

II. LITERATURE REVIEW

A. The AMBER Alert Model and Its Limitations

The America's Missing: Broadcast Emergency Response (AMBER) Alert system, established in 1996 because of a girl named Amber who was abducted in her neighbourhood, represents the most widely studied child-abduction broadcast mechanism globally. As of December 2025, the system has contributed to the safe recovery of 1,292 children through 81 state-level plans in the United States (Office of Justice Programs, 2025).

Even though many people support Amber Alert, researchers are still in doubt. Griffin et al. (2007) conducted the first systematic empirical analysis of 333 publicised AMBER Alert successes and found that the system rarely operates within the critical three-hour window. A subsequent analysis by the same research team of 448 alerts concluded that while over 25 per cent of alerts were associated with safe recoveries, there was little evidence of lifesaving impact in a statistically meaningful sense: the primary predictor of safe recovery was the abductor's relationship to the child, not whether an alert was issued (Griffin & Wiecko, 2016). Critically, fewer than one-third of AMBER Alerts between 2017 and 2021 were issued within three hours of the initial missing-child report, a delay that is especially damaging given the mortality statistics cited above (BAMFI, 2023).

The system can be improved, if instead of waiting for people to give location, smart tools can be used like prediction, in this way Amber Alert would work much faster than they are now. Because the current system takes much time due to the rules and criteria that need to be followed before the alert.

B. India's Existing Frameworks: TrackChild, Khoya-Paya, and Operation Muskan

India's primary portal and solution to missing children is the TrackChild portal, a Ministry of Women and Child Development initiative implemented in partnership with the Ministry of Home Affairs, state governments, railways, Child Welfare Committees, and Juvenile Justice Boards. The portal maintains a centralised database of missing children and those in Child Care Institutions, enabling record-matching. The citizen-facing Khoya-Paya module allows the public to report sightings or missing cases online (PIB, 2018; India.gov.in, 2024).

Complementing these digital tools are operational programmes including Operation Muskan, a periodic police campaign to trace missing children from shelter homes and train stations, the Childline 1098 helpline (operational 24/7), and Anti-Human Trafficking Units (AHTUs) funded by the MHA across all districts (Government of India, 2024). The Supreme Court in September 2025 directed the central government to create a national unified missing-persons portal, acknowledging ongoing coordination failures (Dainik Jagran, 2026).

Sources say despite this infrastructure, recovery rates remain uneven. Kerala achieves an 86 percent recovery rate for missing children, while West Bengal — with the highest absolute volume — recovers only around 52 percent. Some cases are still unsolved , which shows there are serious problems in the current system.

C. Machine Learning in Missing-Persons and Geospatial Search

The application of machine learning to geospatial prediction and location estimation has matured considerably over the past decade. Boersma et al. (2019) demonstrated that GPS tracker data combined with unsupervised clustering and supervised classification could predict future locations of wandering dementia patients with sufficient accuracy to support targeted search operations. Their multi-stage approach — identifying routine patterns before modelling anomalous trajectories — is methodologically relevant to the child-recovery domain, where understanding the likely movement corridor of an abductor from a given urban or rural starting point is a key search parameter.

The LightGBM (Light Gradient Boosting Machine) framework, introduced by Microsoft Research (Ke et al., 2017), has emerged as a leading algorithm for tabular classification and regression tasks requiring rapid inference, particularly where datasets are moderate in size and training must be repeated as new data accumulates. In the context of location prediction, a study published in SN Computer Science used ensemble LightGBM to predict next points of interest for users on a university campus, achieving 95.11 per cent accuracy on historical movement data — demonstrating the feasibility of the gradient-boosting approach for spatial prediction on behavioural tabular features (SN Computer Science, 2023).

For calculating distances between geospatial coordinates, the Haversine formula — derived from spherical trigonometry — provides a computationally efficient and numerically stable method for computing great-circle distances on the Earth's surface. It is expressed as:

$$a = \sin^2(\Delta\varphi/2) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2(\Delta\lambda/2)$$

$$d = 2R \cdot \arcsin(\sqrt{a})$$

where φ represents latitude, λ represents longitude, R is the Earth's mean radius (6,371 km), and d is the great-circle distance (Wikipedia, Haversine Formula; Movable Type Scripts, 2024). SACHET uses this formula within its sighting-refinement module to weight and triangulate sighting reports relative to the initial last-known location of the missing child.

D. Crowdsourcing in Emergency Management

Crowdsourcing — the practice of mobilising distributed public participation for information gathering — has proven effective in disaster response and crisis management. A systematic review of AI-enhanced crowdsourcing for disaster management (PMC, 2024) found that AI systems capable of processing real-time social media data can deliver timely alerts, coordinate emergency responses, and increase community engagement. Key enabling conditions include low-friction reporting interfaces, multi-channel broadcast mechanisms, and feedback loops that reward community participants with case-status updates.

In the child safety context, crowdsourcing presents both opportunity and risk: community-submitted sightings can rapidly narrow a search area, but unverified reports carry the risk of false positives that divert search resources. SACHET addresses this by treating community sightings as probabilistic inputs to a refined ML prediction rather than definitive location assertions, preserving investigative efficiency.

III. PROBLEM STATEMENT AND MOTIVATION

India's missing-children crisis is characterised by three structural failures that compound one another:

Reporting Latency: The process of formally registering a missing child report, attaching photographs, and disseminating alerts to field officers and community members typically takes several hours — often exceeding the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 18, April 2026 Impact Factor: 8.2

three-hour window within which the risk of irreversible harm is highest. First Information Reports (FIRs) and portal submissions are sequential, manual processes that impose delays at every step.

Fragmented Communication: Alert dissemination currently relies on police communication channels, bulletin boards, and periodic social media posts by concerned families. There is no mechanism for instantly reaching all community members in the last-known-location vicinity of a missing child.

Absence of Predictive Search Assistance: Search operations are conducted based on the last-known location and informal experiential knowledge of investigating officers. No structured machine-learning tool currently integrates demographic, temporal, and spatial case features to generate quantitative probability estimates for where a missing child may be found, leaving search coordinators without decision-support tools.

SACHET is motivated in order to address all three gaps in a single, deployable, open-source system that requires no specialised hardware and is accessible via standard web browsers on any device — reducing the barrier between both citizens and law-enforcement personnel in resource-constrained environments.

IV. SYSTEM ARCHITECTURE AND DESIGN

A. Architectural Overview

SACHET follows a Model-View-Controller (MVC) architectural pattern, implemented within the Python Flask micro-framework. The application is divided into three principal layers:

Presentation Layer: Bootstrap 5 responsive HTML templates rendered by Flask's Jinja2 templating engine. Leaflet.js handles interactive map rendering with colour-coded case and sighting markers.

Application Layer: Flask route handlers implement business logic, including case submission, sighting recording, ML prediction invocation, alert dispatch, and role-based access control.

Data Layer: Flask-SQLAlchemy ORM with PostgreSQL as the production database backend, configured with connection pooling (pool_recycle: 300 seconds, pool_pre_ping) for resilience. SQLite is used as a development fallback.

Media assets (photographs and audio recordings of missing children) are stored on Cloudinary's cloud storage service rather than on the application server, ensuring persistence across deployments and providing CDN-accelerated delivery. ML model artefacts (serialised LightGBM pipelines in joblib format) are loaded at application startup and maintained in application memory.

B. Database Schema

The core data model consists of three primary entities:

Missing Child: Stores report id (unique UUID-based identifier), name, age, gender, last seen location, last seen at, last seen ng, description, photo filename, audio filename, emergency contact, date reported, status, abduction time (hour of day as float), abductor relation, region type, population density, and missing date. The

ML-derived fields ml prediction, ml refined, and ml status are stored as serialised JSON to enable cache persistence.

Sighting: Records report id (foreign key to Missing Child), location, latitude, longitude, description, sighting time, and verifier name for each community-submitted sighting.

Risk Zone: Aggregates historical case data into geographic grid cells, storing centroid coordinates, case ids, and a computed risk score for use in analytics dashboards.

C. Request Flow

When a missing-child case is submitted through the web portal, the following sequence executes:

1. Form data and media files (photo, audio) are validated server-side and it is uploaded.
2. A new Missing Child database record is created and committed to PostgreSQL.
3. A rich-format Telegram alert (including photo thumbnail if available) is broadcast to the configured Telegram channel via the Bot API.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

4. Risk zone analysis is triggered asynchronously to update geographic hotspot data.
5. The ML prediction module is queued for the new case, generating initial location estimates based on case input features.

When a community sighting is submitted, steps 3-5 repeat, and the ML refinement module is triggered to update the location prediction with the new sighting data point.

V. MACHINE LEARNING METHODOLOGY

A. Model Architecture

SACHET's predictive analytics engine uses three LightGBM models housed in a scikit-learn pipeline and serialised as joblib files:

Clf risk: A binary classifier trained to predict the risk category of a case (high-risk vs. standard-risk) based on case input features.

Clf recovered: A binary classifier trained to predict the probability that a child will be recovered, incorporating time-elapsed features and regional historical recovery rates.

Reg recovery (implicit in pipeline): A regression component estimating the latitude and longitude of the most probable recovery location.

LightGBM was selected for this task for several reasons documented in the literature: its histogram-based split-finding algorithm provides significant speed advantages over XGBoost on moderate-sized datasets; it natively handles categorical features and missing values, reducing preprocessing overhead; and it supports leaf-wise tree growth that tends to achieve better accuracy on tasks where the target distribution has long tails — as is the case with geographic coordinates (Ke et al., 2017; SN Computer Science, 2023).

B. Feature Engineering

Case input features passed to the ML pipeline are constructed from the MissingChild record via the build case input from child function. These features include:

Temporal features: abduction time (hour of day as float, e.g., 14.5 for 2:30 PM), computed hours since missing.

Demographic features: age, gender.

Contextual features: abductor relation (stranger, family member, acquaintance), region type (Urban, Semi-Urban, Rural), population density.

Geospatial features: last seen at, and distance to nearest city centre (computed via the Haversine formula against a pre-loaded dictionary of major Indian city centre coordinates).

The distance-to-city-centre feature is computed at case-input construction time for local (non-serverless) deployments. For serverless environments, external ML service endpoints receive the serialised case input and return prediction results, enabling separation of the compute-intensive model inference from the web-serving layer.

C. Location Prediction and Refinement

The initial location prediction (predict initial case) produces a dictionary containing predicted latitude, predicted longitude, and associated confidence and risk scores. This constitutes the first-pass search centroid recommended to field officers.

As community sightings are submitted, the refine location with sightings function is invoked. This function builds a list of sighting dictionaries — each containing latitude, longitude, and hours since (computed from sighting time relative to the current UTC time) — and passes them alongside the initial prediction to a refinement algorithm.

The refinement algorithm applies a weighted centroid calculation, where each sighting's weight is inversely proportional to both its distance from the initial predicted location (computed via Haversine) and the elapsed time since the sighting was reported. The intuition is that recent, geographically proximate sightings are more informative

DOI: 10.48175/IJARSCT-34398

than older, distal ones. The output is a refined (rlat, rlon) coordinate pair that updates the ml refined field of the case record.

Location names for both initial and refined predictions are resolved through a reverse-geocoding lookup (get location name from coordinates), with a fallback to a nearest-known-location algorithm that matches coordinates against labelled points from the case record itself — ensuring meaningful location descriptions even when external geocoding APIs are unavailable.

D. ML Result Caching

ML inference is computationally non-trivial in serverless environments with ephemeral memory. SACHET implements a two-tier caching strategy:

Redis Cache (Primary): When a Redis instance is available (via REDIS_URL environment variable), ML predictions are stored in Redis with a configurable TTL (default: 86,400 seconds / 24 hours) using keys of the form ml case cache:{report id}. Results are JSON-serialised.

In-Memory LRU Cache (Secondary): An in-process dictionary (ML_CASE_CACHE) with a maximum of 500 entries and LRU eviction serves as a fallback when Redis is unavailable, providing sub-millisecond lookup for recently computed predictions within a single process lifetime.

Database Persistence: ML predictions are also serialised to JSON and stored in the MissingChild database record, allowing them to survive process restarts and be rehydrated into the in-memory cache on first request.

Cache invalidation is triggered automatically when new sightings are added to a case, ensuring that stale predictions do not mislead search teams.

E. Risk Zone Analysis

Beyond individual case prediction, SACHET sums up geographic risk zones using the analyze risk zones function. This function partitions the spatial extent of historical cases into a grid using a bucket-key function (_risk bucket) that rounds latitude and longitude to one decimal place — creating approximately 10x10 km cells at Indian latitudes. All the cases in groups are compiled together ,then of that group risk is calculated . The risk depends on, how recent the cases are, how many cases are there, how serious they are.

Risk zones with scores above 70 are flagged as HIGH RISK and highlighted on the analytics dashboard for law-enforcement administrators. Zones scoring 40-70 are flagged as MEDIUM RISK and recommended for monitoring. This layered risk communication model enables a triage approach to resource allocation in multi-case scenarios.

VI. KEY FEATURES AND IMPLEMENTATION

A. Missing Child Reporting Portal

The case submission form captures comprehensive case information: child name, age, gender, last-seen location (with a Leaflet.js map for coordinate pinning), physical description, photograph upload, optional audio recording, emergency contact information, and ML-relevant contextual fields (time of disappearance, abductor relationship classification, region type, population density estimate, and exact date of disappearance).

Photo and audio files are uploaded to Cloudinary asynchronously. The application supports JPEG, PNG, and GIF formats for photographs and common audio formats for voice recordings. The maximum file size is configured at 16MB per upload. Cloudinary-hosted media URLs are stored in the database rather than local file paths, enabling the application to remain stateless across deployments.

B. Community Sighting System

Any member can open and navigate to a specific case's detail page and submit a sighting report. Sighting submissions require: a location description, latitude and longitude (auto-populated by the browser's Geolocation API with map confirmation), a text description of the observed child and their apparent condition, and an optional photograph. The submitter's name is recorded.

Upon sighting submission, a Telegram alert is broadcast to the channel, including sighting details and the geographic distance from the child's last known location (computed via the Haversine formula). This dual-alert mechanism — for both case registration and sightings — ensures that subscribers receive actionable updates throughout the search lifecycle rather than only at initial reporting.

C. Real-Time Alert System

SACHET's alert infrastructure evolved from an initial Twilio SMS implementation to a Telegram-based broadcast mechanism, driven by cost and scalability considerations. Telegram Bot API allows unlimited message delivery to arbitrarily large subscriber channels without per-message fees. The system supports rich alerts, including:

Text alerts with case details, last seen location, emergency contact, and a case-specific hyperlink for sighting submission.

Photo alerts where the child's photograph is sent as an inline Telegram image with the alert caption.

Location-segmented alert routing: the select numbers for location function parses the free-text location field and matches it against a pre-configured state-to-phone-number mapping, enabling the system to notify location-specific contacts (e.g., local police or NGO coordinators) rather than broadcasting universally.

The architecture also provides hooks for Discord webhook integration, enabling parallel notifications to Discord-based law-enforcement or NGO coordination servers.

D. Interactive Geospatial Visualisation

Each case detail page includes a Leaflet.js interactive map with three marker layers: the child's last known location (red marker), community sighting locations (orange markers with sighting descriptions in pop-ups), and the ML-predicted recovery location (blue marker with confidence annotations). The analytics dashboard for administrators renders a map with colored areas , where each area is colored based on how risky it is. This helps admin to quickly find most dangerous spot.

E. Administrative Portal

There is special admin portal accessible at /admin/login provides law-enforcement administrators and case managers with tools to update case statuses (active, found, closed), delete cases if needed, review all sightings, and access the analytics dashboard. Admin access is restricted by brute-force-hardened authentication (five-attempt lockout with 15-minute cooling period, configurable via environment variables). A separate police login portal provides authorised police personnel with the ability to submit official missing-child cases without requiring full admin privileges.

F. Security Architecture

Security controls embedded in the application include: environment-variable-based credential management (no default secrets in production), IP-tracked failed-login counters with time-expiring lockouts, route-guarding decorators that return HTTP 404 rather than 401 to mask admin endpoints from unauthorised enumeration, Flask-Login session management with secure cookie configuration, and Werkzeug password-hashing for credential storage. Cloudinary API credentials and Telegram bot tokens are never exposed to the client layer.

VII. TECHNOLOGY STACK

The following table summarises the technology stack employed in SACHET:

TABLE I TECHNOLOGY STACK

Features	Technology Used	Description
Backend Framework	Python Flask	Framework of the Web application
ORM & Database	Flask-SQLAlchemy + PostgreSQL	Storage
ML Framework	LightGBM + scikit-learn Pipeline	To find the risk heat map
Model Serialisation	joblib (.joblib files)	Pipeline channel
Frontend UI	Bootstrap 5 + FontAwesome	User Interface
Geospatial Maps	Leaflet.js	Map for interaction
Cloud Media Storage	Cloudinary	Cloud Storage
Alert Broadcast	Telegram Bot API	Alert message channel
Cache Layer	Redis + In-Memory LRU	ML for predicting Cache
Authentication	Flask-Login + Werkzeug	Security restrictions on access by users
Image Processing	Pillow (PIL)	Image Validation before uploading
Containerisation	Docker + Procfile	Deployment of containers
Geospatial Data	GADM India Shapefiles	Visualization of risk zones

VIII. DEPLOYMENT AND INFRASTRUCTURE

A. Multi-Environment Strategy

SACHET is architected for two cloud infrastructures, working hand-in-hand:

Render (Container-Based): This is our main cloud setup, in which all services will run. It utilizes a managed PostgreSQL database provided by Render. All machine learning dependencies like LightGBM, scikit-learn, numpy, and shapely are available via requirements.txt in the Build stage. Procfile starts up the application using gunicorn. As the entire stack is present here, the inference of ML is done locally.

Vercel (Serverless / Edge): This is the secondary deployment using Vercel Cloud, optimized for performance and access. Since serverless function sizes are tightly bound, the heavy ML dependencies aren' t part of the infrastructure. This particular setup makes calls to an external ML service (ML_SERVICE_URL) to get its results; in most cases, this refers to the Render deployment. The API index.py initializes the Flask web app using Vercel's ASGI wrapper, allowing the frontend to be as lightweight as possible yet capable of running full ML.

In summary, this solution is a blend of sorts, somewhat between serverless and the traditional monolithic backend, which enables the exact codebase to work in two clouds with minimal configuration (VERCEL and RENDER environment variables).

B. Configuration Management

All passwords, tokens, and other sensitive data as well as all other deployment variables are handled via environment variables, which are parsed by the Config class in config.py via an environmental variable resolution function (_read env) that standardises any extra quotation marks around the variables, which is common when copy-pasting from cloud service dashboards. The app handles both full URL connections for PostgreSQL and individual variables for building a connection string (host, user, password, database, port).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

IX. COMPARATIVE ANALYSIS WITH EXISTING SYSTEMS

The following table compares SACHET against India's primary existing missing-child platform and the internationally benchmarked AMBER Alert system:

TABLE II: COMPARISON WITH EXISTING TECHNOLOGIES

Feature	Proposed System (SACHET)	Existing System (India Missing Child Portal)	AMBER Alert System (International)
Alert Mechanism	Telegram broadcast (instant)	None (data portal only)	Emergency broadcast; wireless alerts
Community Sightings	Real-time with geotagging	Basic web form	Phone tips to law enforcement
Predictive Location ML	LightGBM + Haversine refinement	None	None
Risk Zone Analytics	Automated hotspot mapping	None	None
Deployment Time	Sub-minute (automated)	Manual FIR required	Multi-hour (criteria gating)
Access Barrier	Open public portal	Account registration	Law enforcement gated
Media Handling	Photo + audio (Cloudinary)	Photo upload	Text description only
Admin Dashboard	Full case management + analytics	Government officials only	Law enforcement backend
Open Source	Yes (MIT License)	No	No

The comparative analysis highlights SACHET's primary differentiators: the integration of machine learning for location prediction (absent from all comparable systems at the time of writing), automated real-time broadcast at the moment of case submission, and the absence of gatekeeping requirements that delay alert issuance in the AMBER Alert model. SACHET's open-source nature under the MIT License further enables adaptation by NGOs, state governments, and research institutions without licensing barriers.

X. LIMITATIONS AND FUTURE WORK

A. Current Limitations

Some constraints of SACHET's current design that need to be taken into account are listed below:

Limited Training Data: The LightGBM models use limited or artificially generated datasets of missing children cases. The quality of predictions depends on the size and representativeness of case history data available; however, such data cannot be made public for ML model training in India. For the purpose of deployment, an agreement on using the data needs to be reached with relevant authorities.

Fake Sightings: Unverified community sightings may lead to inefficient searches. In future iterations of SACHET, a reputation system for community sightings should be considered or a law enforcement officer confirmation status added.

Telegram Broadcast Limits: Effectiveness of broadcasts depends on the number of Telegram subscribers. Due to inherent limitations, SACHET will be less effective in reaching rural communities where few people use Telegram. An ability to send alerts via SMS should be re-implemented.

Reverse Geocoding Limitations: Availability of an external geocoding API is essential for translating predicted coordinates to familiar locations. Without it, SACHET users see just numbers.

DOI: 10.48175/IJARSCT-34398

B. Future Directions

Below are some of the technology advancements:

Facial Recognition API: The application of facial recognition software (e.g., FaceNet or DeepFace) through a neural-network algorithm for the automatic matching of images shared by the community and the registered faces of missing children.

WhatsApp Business API: Given the large number of WhatsApp users in the country (as many as 560 million as of 2024), the use of the WhatsApp Business API would help in improving outreach capabilities, particularly within the semi-urban and rural areas.

Government Portal APIs: API-based integration with government portals such as TrackChild and Mission Vatsalya to allow for simultaneous registration of cases from SACHET into the government database without the need to duplicate efforts.

Mobile Application Development: Mobile applications to be developed on both iOS and Android platforms with push notification and offline viewing capabilities.

Federated Learning: Federated learning can be enabled in the case of implementation of SACHET in multiple locations while avoiding centralized case data collection.

Temporal Decay Function: The inclusion of temporal decay in the location prediction model owing to the realization that over time the potential radius of search increases with the movements of the abducted children.

XI. CONCLUSION

SACHET is thus an important step forward in expanding the technological tools needed to cope with missing child crises in India. Through integrating three innovative technologies that work together – crowdsourced sight spotting from the community to aggregate sightings, automatic broadcasting via Telegram to spread alerts, and predictive analytics using LightGBM technology to guide searches – the system works to fill the gap left by both India's current government websites and alert systems used around the world.

Indeed, the scale of the challenge requires immediate action. As noted, India saw 83,000 cases of child disappearance in 2022, of whom only 22,687 were found at the end of the year – leaving 47,313 missing, 71% of whom were girls. The toll of human life caused by insufficient analysis and technological capabilities can hardly be calculated; each hour wasted searching for a missing child is one where the odds of success get lower and lower.

Open-source nature, the system's multi-cloud architecture, and its modularity help tailor it to fit the practical and financial constraints of NGOs, state police forces, and child welfare organizations in India. Collaboration with the Ministry of Women and Child Development and the crime records offices of states will help leverage real-world data from cases for training purposes and the use of APIs for government portals, turning SACHET from a mere prototype into a national lifeline.

Child safety is not merely a law-enforcement challenge — it is a community responsibility. SACHET operationalises that responsibility through technology.

XII. ACKNOWLEDGEMENTS

The authors acknowledge the contributions of the open-source communities behind Python Flask, LightGBM, Leaflet.js, and Bootstrap. They also acknowledge the data reporting work of India's National Crime Records Bureau (NCRB), whose annual Crime in India reports provide the empirical foundation for quantifying the scale of the missing-children crisis. The GADM administrative boundary shapefiles used for geographic risk visualisation are made available by the Database of Global Administrative Areas under their respective academic licences.

REFERENCES

[1] National Crime Records Bureau (NCRB). (2022). Crime in India 2022 (Book 1). Ministry of Home Affairs, Government of India. Retrieved from https://www.ncrb.gov.in

[2] The Print. (2024, January 3). More than 47,000 children are missing in India, 71% are girls, according to NCRB data. Retrieved from https://theprint.in
[3] The News Minute. (2024, February 6). The gone girls of India: CRY report on NCRB data on missing girls says situation grim. Retrieved from https://www.thenewsminute.com
[4] Dainik Jagran. (2026, March 8). 33,577 Missing Children, 7 States With Zero Cases: India's Child Safety Crisis in Black and White. Retrieved from https://english.dainikjagranmpcg.com
[5] Down to Earth. (2025, May 25). World Missing Children's Day 2025: No growth is inclusive with our daughters still missing. Retrieved from https://www.downtoearth.org.in
[6] Government of India, Ministry of Home Affairs. (2024, February 6). Lok Sabha Unstarred Question No. 467: Missing Children Data and Technology. Parliamentary Answer.
[7] Press Information Bureau, Government of India. (2018). Online Tracking System/Portals for Missing Children. Retrieved from https://www.pib.gov.in
[8] Press Information Bureau, Government of India. (2019). Drive to Rescue Missing Children. Retrieved from https://www.pib.gov.in
[9] Office of Justice Programs, U.S. Department of Justice. (2025). AMBER Alert Statistics. Retrieved from https://amberalert.ojp.gov/statistics
[10] Griffin, T., Miller, M. K., Hoppe, J., Rebideaux, A., & Hammack, R. (2007). An empirical examination of AMBER Alert 'successes'. Criminal Justice Policy Review, 18(4), 418-436.
[11] Griffin, T., & Wiecko, F. (2016). Does AMBER Alert 'save lives'? An empirical analysis and critical implications. Journal of Crime and Justice, 39(4). https://doi.org/10.1080/0735648X.2014.1003577
[12] Griffin, T., Williams, J. H., Miller, M. K., & Wooldredge, J. (2021). AMBER Alert Effectiveness Reexamined. Criminology and Criminal Justice (published on ResearchGate). Retrieved from https://www.researchgate.net
[13] Black and Missing Foundation Inc. (BAMFI). (2023). Do Amber Alerts work? Data shows how often they help bring missing kids home. Retrieved from https://www.blackandmissinginc.com
[14] Hanfland, K. A., Keppel, R. D., & Weis, J. G. (1997). Case management for a missing children homicide investigation. U.S. Department of Justice / Attorney General of Washington.
[15] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.
[16] Pandey, M., & Litoriya, R. (2023). LightGBM: Next Point of Interest Location Prediction Using Ensemble Machine Learning. SN Computer Science, 4, Article 739. https://doi.org/10.1007/s42979-023-02254-6
[17] Boersma, P., van der Waal, M., & Schuurman, I. (2019). Location prediction using GPS trackers: Can machine learning help locate the missing people with dementia? ScienceDirect. https://doi.org/10.1016/j.smate.2018.11.003
[18] Maria, E., Budiman, E., Taruk, M., et al. (2020). Measure distance by locating nearest public facilities using Haversine and Euclidean methods. Journal of Physics: Conference Series, 1450, 012080. IOP Publishing.
[19] Inman, J. (1835). Navigation and Nautical Astronomy for the Use of British Seamen (3rd ed.). London: W. Woodward, C. & J. Rivington. [Original haversine formula derivation]
[20] PMC / Frontiers in Public Health. (2024). AI-enhanced crowdsourcing for disaster management: Strengthening community resilience through social media. PubMed Central. Retrieved from https://pmc.ncbi.nlm.nih.gov
[21] Bloomquist, J. (2018). An Analysis of the Unintended Effects of the AMBER Alert System. Honours Thesis, Bemidji State University. Retrieved from https://www.bemidjistate.edu
[22] Sciammarella, S. (2021). AMBER Alert System: An Analysis on how Access to Technology and Data-Driven Approaches Could Improve Effectiveness. Master's Project, San Jose State University. SJSU ScholarWorks.
[23] Wikipedia. (2024). Haversine Formula. Wikimedia Foundation. Retrieved from https://en.wikipedia.org/wiki/Haversine_formula
[24] Movable Type Scripts. (2024). Calculate distance and bearing between two Latitude/Longitude points using haversine formula. Retrieved from https://www.movable-type.co.uk/scripts/latlong.html

[25] Mission Vatsalya / WCD Ministry. (2024). TrackChild / Khoya-Paya Portal. Retrieved from https://missionvatsalya.wcd.gov.in and https://trackthemissingchild.gov.in
[26] Grinberg, M. (2024). A Year In Review: Flask in 2024. Retrieved from https://blog.miguelgrinberg.com
[27] Cloudinary Inc. (2024). Cloudinary Developer Documentation. Retrieved from https://cloudinary.com/documentation
[28] Telegram. (2024). Bot API Documentation. Retrieved from https://core.telegram.org/bots/api
[29] NCPCR / Ministry of WCD. (2024). GHAR — Go Home and Re-Unite portal. Retrieved from https://services.india.gov.in