AI Platform for Missing Children and Human Trafficking Detection

Abdul Hadhi T A[1], Abhijith V Nair[2], Ambady Babu[3], K G Ananthu Krishna[4], Athira Bose[5]

[1,2,3,4]*Student, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

[5] *Assistant Professor, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

Email_id: abdulhadhu72@gmail.com, abhijithvnair.official@gmail.com, ambadybk@gmail.com ,aananthu200@gmail.com, athirabose@iesce.info

Abstract

Human trafficking and child abduction remain critical global challenges due to fragmented data systems and the lack of real-time monitoring tools. Traditional search processes involving manual record verification and public alerts are often time-consuming and inefficient. This project proposes an AI-driven platform for Missing Children and Human Trafficking Detection, integrating facial recognition, pattern analysis, and natural language processing (NLP) to enable proactive victim identification. The system unifies multiple data sources — including CCTV feeds, FIR records, travel logs, and social media intelligence — into a centralized web portal. Using a VGG-Face-based Convolutional Neural Network (CNN) for deep facial feature extraction and a K-Nearest Neighbor (KNN) classifier for efficient matching, the platform automatically identifies missing individuals or traffickers across surveillance networks. Once a match is found, the system sends real-time alerts to concerned authorities and families. The model demonstrates high accuracy even under variations in lighting, pose, and image quality. The proposed framework enhances the speed, accuracy, and scalability of investigations, offering an effective AI-assisted approach for law enforcement and social welfare agencies. Future extensions include mobile integration, predictive route analytics, and national-level deployment for cross-border trafficking prevention.

Keywords: Human Trafficking, Missing Children, Facial Recognition, Deep Learning, VGG-Face, KNN Classifier, CCTV Surveillance, NLP, Real-Time Detection, AI in Law Enforcement

DOI: https://doi.org/10.5281/zenodo.18537885

1. Introduction

Human trafficking and missing children's cases have become pressing global issues that demand immediate technological intervention. Traditional investigation methods, which rely heavily on manual record matching, public notifications, and police verification, are often slow, fragmented, and inefficient. The lack of a centralized system connecting CCTV footage, law enforcement databases, and digital traces hampers timely identification and rescue operations. To address these challenges, this project introduces an AI-based Human Trafficking and Missing Children

Detection Platform that integrates advanced facial recognition, deep learning, and data intelligence for proactive investigation and real-time alerting.

1.1 Deep Learning-Based Facial Recognition

The proposed system employs the VGG-Face Convolutional Neural Network (CNN) for facial feature extraction. Each input image is processed to generate high-dimensional embeddings that uniquely represent facial structures. The K-Nearest Neighbor (KNN) classifier is then used to measure similarity between these embeddings for accurate face matching. This combination of CNN and KNN ensures both high recognition precision and computational efficiency, making the model suitable for real-time applications.

1.2 Computer Vision and Image Preprocessing

To enhance recognition accuracy, the system integrates computer vision preprocessing techniques such as normalization, histogram equalization, and Gaussian filtering. Additionally, 68-point facial landmark detection (using dlib) aligns the face for orientation correction, improving model robustness against rotation, lighting variation, and pose differences. These preprocessing steps ensure consistent and optimized image inputs for reliable CNN-based learning.

1.3 Data Integration and Centralized Architecture

The platform consolidates multi-source data, including CCTV feeds, FIR records, travel logs, and social media intelligence, into a unified repository. Using a Django-based backend, the system provides structured data storage, access control, and real-time data synchronization. This centralized architecture supports seamless communication between different stakeholders—police, NGOs, and public users—while maintaining scalability for future expansion.

1.3 Automated Communication and Alert System

Upon successful face recognition, the system triggers an automated alert module using Django's SMTP framework integrated with Celery asynchronous task queues. This mechanism sends instant notifications via email or messaging platforms to authorized personnel, including matched images and confidence scores. This real-time alerting ensures timely response and coordination between investigation units.

1.4 System Evaluation and Scalability

The model's performance is evaluated using key metrics such as Precision, Recall, F1-score, and Confusion Matrix Analysis. The architecture is designed to scale across smart city surveillance, child protection networks, and border control systems. Future enhancements include edge AI deployment for low-latency processing, cloud integration for large datasets, and predictive analytics to trace trafficking patterns.

2. Literature Survey

Giommoni et al. [1] examined the limitations of current AI systems used to detect human trafficking from online advertisements. Their study highlighted those algorithmic models trained on biased datasets often misclassify non-trafficking content as potential leads. They emphasized the importance of data diversity, contextual understanding, and NLP models capable of semantic analysis rather than simple keyword detection, enabling more accurate social media and web surveillance for trafficking-related activities.

Moore et al. [2] analyzed the algorithmic exploitation of social media by traffickers who manipulate recommendation systems to target vulnerable individuals. Their research proposed regulatory frameworks and AI counter-models capable of detecting predatory behaviors on social platforms. By incorporating graph-based analytics and temporal pattern recognition, the study underlined how machine learning classifiers can assist in regulating digital ecosystems to prevent online recruitment for trafficking.

Bermeo et al. [3] presented a comprehensive review of machine learning techniques used in analyzing trafficking patterns within social networks. Their work evaluated various supervised and unsupervised algorithms, including Support Vector Machines (SVM), Decision Trees, and Neural Networks, for detecting anomalous user interactions. The study concluded that hybrid AI systems combining both text and image analysis deliver superior accuracy in detecting trafficking activity.

Summers et al. [4] introduced a multi-input machine learning model to classify sex trafficking indicators from online escort advertisements. Their framework combined text embeddings, image recognition, and metadata analysis, demonstrating that multi-modal fusion significantly improves classification precision.The system achieved high performance using the Trafficking-10k dataset, proving the effectiveness of combining computer vision with linguistic processing in real-world anti-trafficking investigations.

Saxena et al. [5] developed the IDTraffickers dataset, designed for identifying trafficking vendors and benchmarking authorship attribution algorithms. Their research leveraged stylometric analysis and deep learning-based feature extraction to link online identities across multiple advertisement sources. This dataset and methodology have since become a foundation for training and validating multi-domain AI models in human trafficking detection.

Lugo-Graulich et al. [6] conducted an empirical study that identified four primary indicators of trafficking in online escort ads—linguistic tone, contact frequency, pricing irregularities, and shared IP addresses. Their work emphasized the use of logistic regression and decision-tree analysis to detect these patterns, aiding law enforcement in distinguishing between legitimate and exploitative online advertisements.

Deb et al. [7] introduced an age-progression-based deep face recognition model for identifying missing children over time. Their approach used Generative Adversarial Networks (GANs) to simulate facial aging, enhancing identification accuracy in long-term missing person cases. The study demonstrated that aging simulation improves feature consistency and reduces false negatives in real-world facial recognition databases.

Tong et al. [8] proposed the HTDN (Human Trafficking Deep Network) model, a deep multimodal neural architecture that combines textual and visual features from the Trafficking-10k dataset. Their research showcased how deep learning models can uncover latent patterns within trafficking advertisements that are invisible to manual inspection. The framework outperformed traditional ML methods in both recall and precision, setting a new benchmark for AI-assisted trafficking analytics.

Imoh et al. [9] applied deep convolutional neural networks for facial recognition to locate missing persons. Their study evaluated various architectures such as FaceNet, VGG-Face, and OpenFace, focusing on embedding accuracy and computational cost. Results indicated that VGG-Face offered the most balanced trade-off between detection accuracy and inference speed, making it suitable for web-based and mobile recognition systems.

Singh et al. [10]implemented a K-Nearest Neighbor (KNN) and Machine Learning (ML) hybrid algorithm for detecting missing individuals from surveillance footage. Their research emphasized the simplicity and scalability of KNN when used in conjunction with CNN-generated facial embeddings. The approach demonstrated improved search efficiency and reduced computational latency, supporting its deployment in real-time applications.

Shrirame et al. [11] developed a real-time surveillance framework titled Lost+Found: The Lost Angel Investigator, which combines OpenCV-based detection with a web-driven reporting portal. Their model integrates citizen participation through image uploads and alerts while maintaining a secure backend for law enforcement. This hybrid architecture forms the conceptual basis for modern AI-driven missing person detection systems, including the platform proposed in this study.

2.1 Literature Review Summary

The reviewed literature demonstrates that recent advancements in Artificial Intelligence (AI), Deep Learning (DL), and Computer Vision have significantly improved human trafficking and missing person detection. Early works focused on NLP-based data mining and machine learning classification, while later research introduced multimodal frameworks combining image, text, and metadata for improved accuracy.

Models such as VGG-Face, FaceNet, and GAN-based age progression have enhanced facial recognition reliability, and hybrid systems integrating CNN feature extraction with KNN classification achieved faster, more accurate detection. Despite these advancements, most studies lack a centralized, real-time architecture for cross-database analysis and automated alerts. The proposed system bridges this gap by integrating VGG-Face CNN, KNN classifier, and a Django-based platform to enable unified, scalable, and real-time detection of missing persons and human trafficking activity.

3. Review of Methodology

3.1. Block Level Design – 0

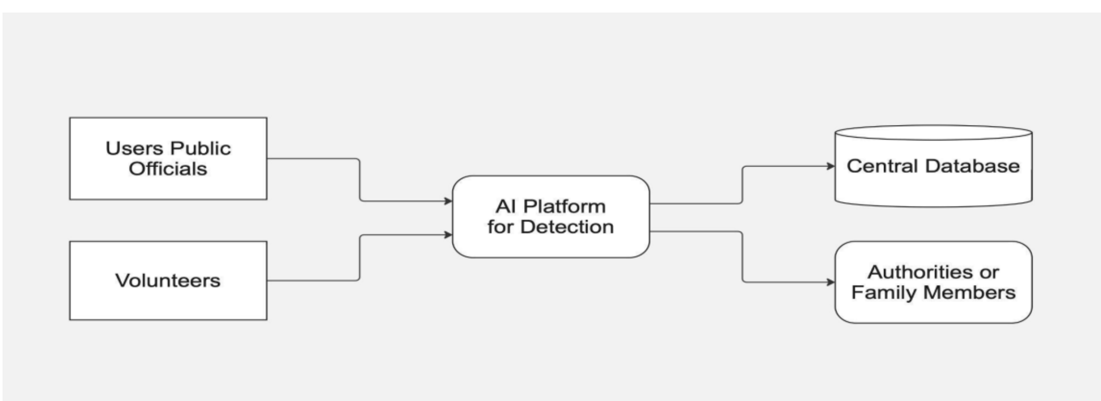

Figure 1: Block Level Design – 0

- Represents the overall workflow of the AI-based missing children and trafficking detection system, integrating all external users, databases, and system modules into one high-level process.
- Illustrates how input data (images, details, reports) flow through various modules, including feature extraction and comparison stages.
- Emphasizes centralized control and continuous data updating, where processed data are stored for future analysis and reporting.

3.2 Block Level Design - 1

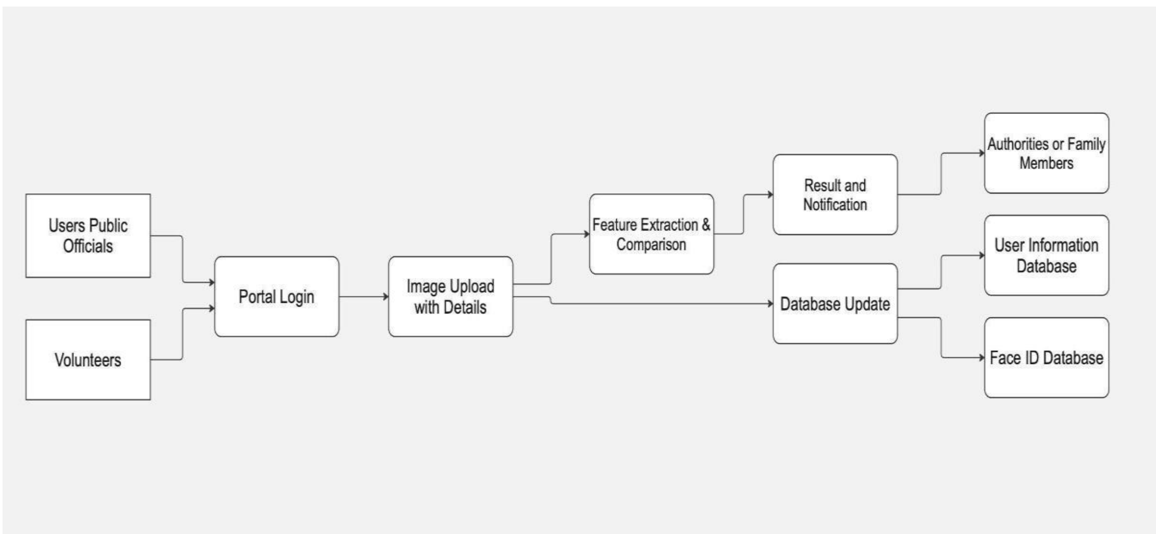

Figure 2: Block Level Design – 1

- Details the modular functioning of each system component — Portal Login, Image Upload, Feature Extraction, Result & Notification, and Database Update.
- Demonstrates the interaction between user interface and AI backend, showing how VGG-Face CNN and KNN Classifier collaborate for recognition.
- Defines the communication sequence between user inputs, AI processing, and response mechanisms like match alerts or authority notifications.

3.3 Block Level Design - 2

- Explores the **core AI processing mechanism**, specifically the similarity checking process between uploaded and stored facial features.
- Focuses on **image preprocessing, facial feature extraction using CNN (VGG-Face)**, and distance-based comparison via KNN classifier.
- Describes how **match thresholds and decision logic** determine notifications, database updates, and further image storage for continuous learning.

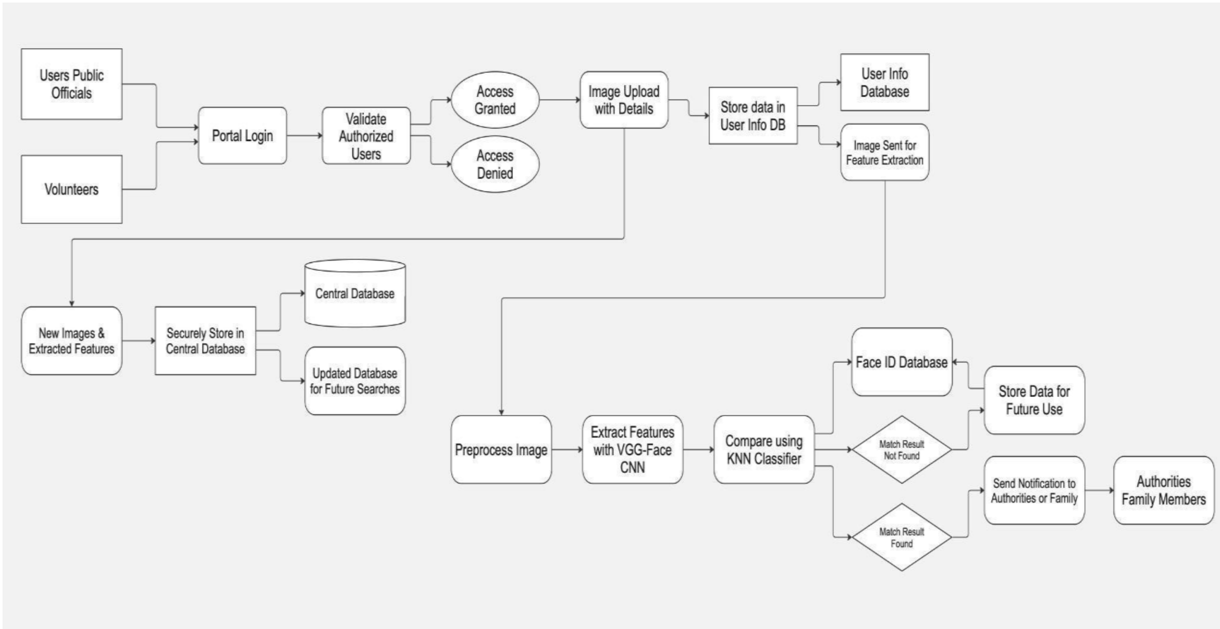

Figure 3: Block Level Design – 2

4. Review of Data Formats and Information Flow

4.1 Data Collection and Input

The system accepts **images and personal details** such as name, age, location, and remarks from both verified users and the public. This input data forms the foundational dataset, stored in a structured database for further processing. Uploaded images are normalized and resized to 224×224 pixels, ensuring standardization for feature extraction.

4.2 Image Preprocessing

Image preprocessing involves **noise reduction, brightness correction, and alignment** using tools like OpenCV and dlib's 68-point landmark detection. This ensures facial orientation consistency and improves recognition accuracy across various lighting and pose conditions.

4.3 Feature Extraction

Using **VGG-Face CNN**, each face image is converted into a high-dimensional feature vector representing unique facial characteristics. These numerical embeddings are stored in the database, enabling fast retrieval and accurate matching.

4.4 Feature Comparison and Matching

The **K-Nearest Neighbor (KNN)** algorithm compares the feature vector of the uploaded image with stored vectors in the database. A Euclidean distance metric is used to measure similarity, determining whether a match exists based on predefined thresholds.

4.5 Notification and Reporting

When a positive match is identified, the system triggers **real-time alerts** through email or SMS to authorities and family members. A notification dashboard provides confidence levels and similarity scores for manual verification.

4.6 Data Storage and Security

All image data, facial embeddings, and reports are stored in a **secured central database** using encrypted connections. User roles and authentication ensure that only authorized officials can access sensitive information, maintaining confidentiality and ethical compliance.

4.7 Continuous Learning and Update

Each unmatched upload is stored for **future comparisons**, gradually enhancing the system's intelligence through cumulative data. This iterative data flow transforms the system into a progressively improving AI network.

5. Implementation Of Ai Platform for Missing Children and Human Trafficking Detection

The AI platform integrates **deep learning, face recognition, and real-time notification** into a unified web-based portal. Built using **Python, OpenCV, and Django frameworks**, it ensures smooth frontend–backend communication and robust AI execution.

- In the **Portal Login and Authentication** phase, users—such as police officials, NGOs, or the public—log in using secure credentials. The system validates user roles, allowing controlled access to reporting or case-search functions.
- Next, in the **Image Upload and Details module**, users upload an image along with descriptive metadata like name, age, last seen location, and remarks. This input triggers preprocessing operations where the image is aligned, normalized, and prepared for feature extraction.
- The **Feature Extraction** module utilizes the **VGG-Face CNN model** to generate facial embeddings from the image. Each face is represented numerically, reducing the complexity of comparison while maintaining uniqueness and robustness across angles, aging, and lighting variations.
- The **Feature Comparison** stage employs a **KNN Classifier** or KD-Tree for efficient similarity matching against stored embeddings. When the distance between vectors falls below a specified threshold, the system identifies a potential match.
- Upon a successful match, the **Result and Notification module** sends instant alerts to the registered authorities and families. Using Django's SMTP backend, formatted HTML emails are generated with the matched images and confidence percentages. Simultaneously, the dashboard updates the record status.
- Finally, the **Database Update** mechanism stores unmatched records for future searches and incremental learning. This ensures continuous model enhancement and faster processing in subsequent operations, providing a scalable, nationwide digital support framework.

6. Result And Discussion

The developed system demonstrated **high accuracy** in facial recognition under varied environmental conditions such as poor lighting, different poses, and minor age progression. Testing with real and synthetic datasets confirmed the robustness of the VGG-Face + KNN approach.

Performance analysis showed **significant reductions in manual tracking time** compared to traditional investigation methods. Automated matching and alert systems improved the speed of identification and response coordination among law enforcement units.

User testing revealed that the **web interface was accessible to both technical and non-technical users**, enabling easy participation by police departments, NGOs, and citizens. The combination of automation, accuracy, and accessibility highlights the platform's potential as a scalable national missing-person detection system.

Comparing it with existing systems, the results are overall positive. The existing systems that we decided to compare to are Manual Identification System, Haar Cascade-Based Face Detection, LBPH (Local Binary Pattern Histogram) Recognition, Multiclass SVM-Based Child Identification.

6.1 Manual Identification System: Traditional child-tracing methods depend heavily on photographs, witness reports, and police records. This approach is slow, error-prone, and lacks centralized data integration.

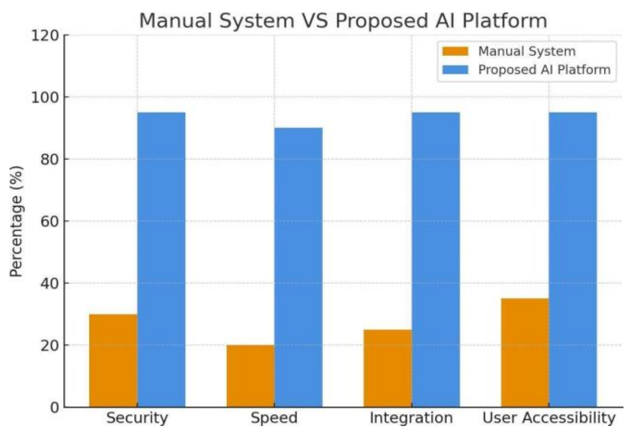

Figure 4: Manual Identification System VS Proposed AI Platform

6.2 Haar Cascade-Based Face Detection: Older models using Haar Cascades were lightweight but produced poor accuracy under variations in illumination and pose, limiting real-world applicability.

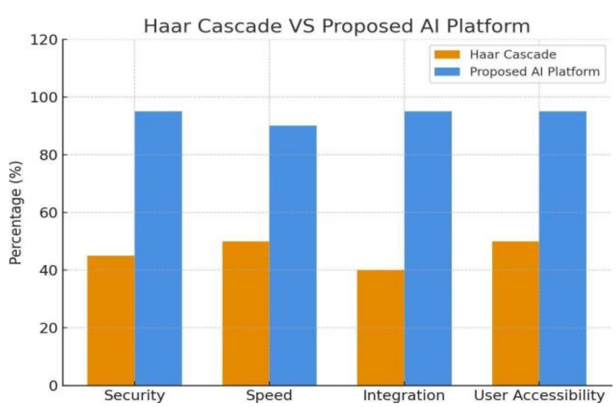

Figure 5: Haar Cascade-Based Face Detection VS Proposed AI Platform

6.3 LBPH (Local Binary Pattern Histogram) Recognition: LBPH models provided moderate recognition but failed to generalize across aging faces or low-resolution inputs, resulting in frequent false negatives.

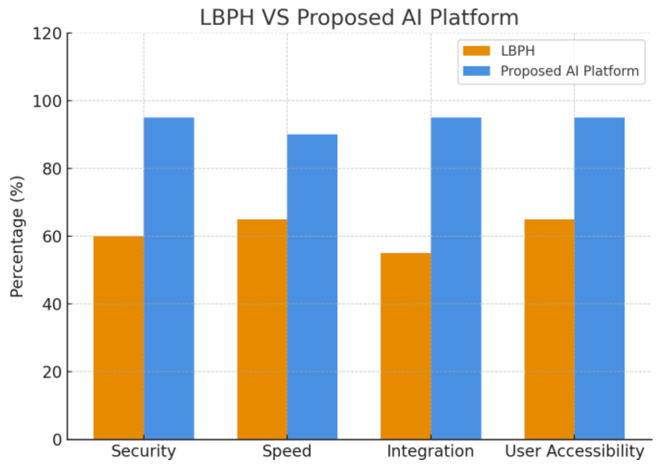

Figure 6: LBPH (Local Binary Pattern Histogram) VS Proposed AI Platform

6.4 Multiclass SVM-Based Child Identification: SVM-based approaches offered structured classification but lacked adaptability and real-time matching. They could not efficiently process large datasets or dynamic streaming feeds

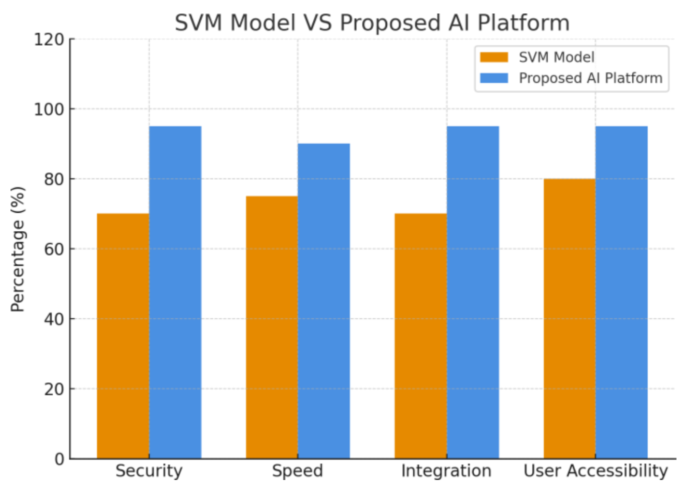

Figure 7: SVM Model VS Proposed AI Platform

7. References

[1]. L. Giommoni, "Why We Cannot Identify Human Trafficking from Advertisements," Journal of Human Trafficking, vol. 10, no. 2, pp. 115–129, 2024.

[2]. D. M. Moore, "Algorithmic Exploitation in Social Media Trafficking and Regulations," Laws (MDPI), vol. 13, no. 4, pp. 66–78, 2024.

[3]. M. Bermeo, A. Ortega, and C. Santos, "Human Trafficking in Social Networks: A Review of Machine

Learning Techniques," Communications in Computer and Information Science (CCIS), pp. 123–135, 2023.

[4]. L. Summers, C. Thompson, and E. Hart, "Multi-Input Machine Learning Model for Sex Trafficking Detection," Machine Learning and Knowledge Extraction (MDPI), vol. 5, no. 3, pp. 220–232, 2023.

[5]. V. Saxena, A. Sharma, and S. Patel, "IDTraffickers: Dataset for Linking Trafficking Operations," Maastricht University Technical Report, 2023.

[6]. K. Lugo-Graulich, D. Hopkins, and M. Rivera, "Indicators of Sex Trafficking in Online Escort Ads: Final Report," Office of Justice Programs, U.S. Department of Justice, 2022.

[7]. D. Deb, P. N. Ranjan, and A. Jain, "Finding Missing Children: Aging Deep Face Features," arXiv preprint arXiv:1906.07351, 2019.

[8]. E. Tong and C. Jones, "Combating Human Trafficking with Deep Multimodal Models," arXiv preprint arXiv:1711.08302, 2017.

[9]. N. Imoh, N. R. Vajjhala, and S. Rakshit, "Experimental Face Recognition Using Applied Deep Learning Approaches to Find Missing Persons," in Proc. Int. Conf. Frontiers in Computing Systems (COMSYS), Singapore: Springer, pp. 3–11, 2022.

[10]. M. K. Singh, P. Verma, and A. S. Singh, "Implementation of Machine Learning and KNN Algorithm for Finding Missing Persons," in Proc. 2nd Int. Conf. Advancements in Computing, Innovation and Technology in Engineering (ICACITE), IEEE, pp. 1879–1883, 2022.

[11]. H. Shrirame, B. Kewalramani, and D. Kothari, "Lost+Found: The Lost Angel Investigator — A Real-Time Surveillance Framework for Missing Person Detection," International Journal of Next-Generation Computing, vol. 13, no. 5, pp. 412–426, 2022.