

Article

VERA: A Privacy-Preserving Framework for Deep Learning Data Collection and Object Detection in Private Settings

Manuel H. Jimenez ¹, Onur Toker ² and Luis G. Jaimes ^{1,*} ¹ Department of Computer Science, Florida Polytechnic University, Lakeland, FL 33805, USA; roswellman86@gmail.com² Department of Electrical and Computer Engineering, Florida Polytechnic University, Lakeland, FL 33805, USA; otoker@floridapoly.edu

* Correspondence: ljames@floridapoly.edu

Abstract

This paper introduces VERA (Vision Expert Real Analysis), a privacy-supporting cyber-physical framework designed for real-time data collection and visual analysis in healthcare environments. VERA limits exposure to identifiable RGB content by ensuring that annotators interact only with non-identifiable edge-based representations, while original images remain encrypted at rest using AES-CFB, with integrity verification performed before in-memory decryption. The system integrates edge-based obfuscation, secure annotation, in-memory decryption, and dynamic data augmentation to train YOLO-based person detection models without compromising patient privacy. Experimental results on a curated COCO subset show that VERA enables effective person detection, improving mean Average Precision (mAP) from an intentionally minimal baseline of 0.61 percent to 99.94 percent after full training and augmentation. This baseline is used solely to illustrate the contribution of the secure data preparation pipeline and is not intended to represent a fully optimized YOLO configuration. The results demonstrate that privacy-supportive workflows can maintain strong model performance while aligning with data protection practices common in regulated environments. Although this work focuses on person detection as a foundational stage, the VERA architecture is designed to support future extensions toward privacy-preserving Human Activity Recognition (HAR) tasks in clinical and assisted-living settings.

Keywords: health informatics; computer vision; activity recognition

1. Introduction

The adoption of artificial intelligence (AI) and computer vision technologies in healthcare has enabled significant advancements in real-time diagnostics, patient monitoring, and proactive care delivery. Human Activity Recognition (HAR) systems are increasingly used in healthcare to monitor patients' behaviors, detect abnormal events, and support clinical decision-making. Through the integration of computer vision and deep learning techniques, modern HAR solutions enable continuous observation of patient motion patterns in hospital rooms, intensive care units (ICUs), and elderly-care facilities. These developments are particularly valuable in sensitive environments where timely detection of patient behavior and medical conditions can lead to improved outcomes and reduced risks.

However, these capabilities introduce major ethical and technical challenges related to patient privacy. Hospital environments contain sensitive visual information, including patients, staff, and medical equipment that must be handled under strict confidentiality



Academic Editors: Luigi Bibbò,
J. Artur Serrano and Roger Narayan

Received: 31 October 2025

Revised: 26 January 2026

Accepted: 11 February 2026

Published: 23 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

rules defined by frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) [1] in the United States and the General Data Protection Regulation (GDPR) [2] in Europe. The use of visual data, especially facial and bodily imagery, raises concerns about unauthorized exposure, re-identification, and compliance with regulatory standards. In practice, healthcare institutions often face a trade-off between system effectiveness and strict data privacy requirements.

Existing privacy-preserving approaches for visual data include blurring, pixelation, and color masking. These techniques conceal identity but frequently degrade structural information essential for action understanding. More advanced methods, such as adversarial privacy protection [3] and differential privacy [4], offer stronger protection but generally require considerable computational resources or introduce noise that reduces accuracy. There remains a need for an efficient solution that preserves geometric structure relevant to HAR while concealing identifiable content and maintaining compatibility with real-time clinical workflows.

To address this need, we propose VERA (Vision Expert Real Analysis), a privacy-preserving data collection and activity recognition framework designed specifically for medical and clinical environments. VERA transforms raw RGB inputs into edge-based representations that retain spatial structure while removing color and texture. These edge-only representations are encrypted, labeled, and used to train a deep neural network for person detection and activity inference. The framework employs a dual-mode architecture that separates secure Development workflows from real-time Production inference. Central elements of VERA include edge-based obfuscation, AES in Cipher Feedback (CFB) mode for image encryption [5], and dynamic augmentation for robust YOLO-based training.

1.1. Contributions

The main contributions of this work include:

1. A dual-mode architecture enabling real-time, privacy-aware patient monitoring with clear separation between secure training and production inference.
2. An edge-based obfuscation strategy that preserves structural cues while eliminating identifiable visual features.
3. A secure data pipeline integrating Canny edge detection, encrypted RGB storage, and annotation performed exclusively on non-identifiable images.
4. A privacy-aware training methodology based on controlled decryption within memory, dynamic augmentation, and secure data loading.
5. A customized YOLOv5 model trained on obfuscated data achieving high detection accuracy ($\text{mAP} \approx 99.94\%$) for person detection.
6. A complete evaluation demonstrating that VERA maintains detection performance comparable to traditional RGB-based training while improving privacy protection.

1.2. Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews prior work in privacy-preserving AI and medical image security. Section 3 introduces the VERA system architecture, including both development and production modes. Section 4 describes preprocessing, edge-based obfuscation, encryption, labeling, and model training. Section 5 presents the experimental evaluation. Section 6 outlines potential applications and deployment scenarios. Section 7 discusses the main limitations of the current framework. Finally, Section 8 concludes the paper and highlights future research directions.

2. Related Work

Computer vision systems designed for healthcare, surveillance, and assisted living environments must balance detection accuracy with strict privacy constraints. A wide range of methods have been proposed to support privacy-aware visual processing, including image transformation, de-identification, encryption, adversarial perturbation, and differential privacy. This section reviews key contributions in these areas and situates the VERA framework within the broader landscape of privacy-preserving visual analytics.

Early approaches to privacy in medical imaging focused primarily on text-based anonymization and metadata removal from DICOM headers. More recent analyses demonstrate the limitations of such approaches, showing that visual data itself can contain identifiable cues even when metadata is removed. Works such as [6,7] emphasize the need for privacy methods that act directly on image content to ensure safety in the era of deep learning.

2.1. Blurring and Pixelation Methods

Traditional image masking methods remove fine-grained facial details by applying blurring, pixelation, or black-box occlusion to regions of interest. These methods are simple and lightweight, but they often degrade the geometric and structural detail needed for downstream tasks such as pose estimation or activity recognition. Moreover, recent studies suggest that deep learning models can sometimes recover identity cues even from blurred content, raising concerns about residual identifiability.

2.2. Advanced De-Identification Techniques

More advanced de-identification approaches include segmentation-based masking, synthetic face replacement, and perturbation-based obfuscation. For example, the Digital Mask technique proposed in [8] removes identity features while preserving diagnostic structures in medical images. While this is effective for modality-specific tasks such as endoscopy or ophthalmology imaging, the method can be computationally demanding and is not designed for real-time, general-purpose scene analysis.

2.3. Adversarial and Generative Privacy Methods

Generative and adversarial methods have also been explored for image privacy protection. Instead of removing content, these techniques introduce structured perturbations that prevent deep learning models from recovering identity cues. Tong et al. [3] propose an adversarial perturbation generative network that modifies images in a way that protects identity while preserving utility for downstream tasks. These methods offer strong privacy guarantees but often introduce visual artifacts or require significant training resources, making real-time deployment challenging.

2.4. Differential Privacy in Computer Vision

Differential privacy provides a formal mathematical framework for limiting the disclosure risk of individual data samples by injecting calibrated noise into computations [4]. Although widely used in statistical data analysis, its application to computer vision remains limited because injecting noise directly into images or deep features often degrades performance significantly. For tasks that require detailed spatial reasoning, such as person detection or pose estimation, differential privacy noise can reduce accuracy to impractical levels.

2.5. Encrypted and Privacy-Supporting Data Pipelines

Beyond transformation-based anonymization, several works have explored secure data pipelines using encryption or controlled-access architectures. For example, Xu et al. [9] propose secure integrity verification for healthcare cyber-physical systems, highlighting

the importance of encrypted storage and secure access channels. These approaches complement privacy by ensuring confidentiality and preventing unauthorized access but do not inherently remove identifiable visual content.

2.6. Comparison with Privacy-Preserving Methods

The VERA framework differs from the above approaches by combining edge-based obfuscation with encrypted RGB storage in a unified data collection and training pipeline. Edge representations remove texture and color information, reducing the exposure of identifiable features, while preserving geometric structure needed for annotation. Unlike blurring or pixelation, edges retain the contours necessary for manual labeling. Unlike Digital Mask and adversarial perturbation methods, VERA avoids computationally heavy generative models. Unlike differential privacy, VERA does not inject random noise into the data and therefore avoids the associated performance degradation. Table 1 provides a qualitative comparison of representative privacy-preserving methods discussed in this section.

Table 1. Qualitative comparison of privacy-preserving methods.

Method	Identity Removal	Annotation Suitability	Compute Cost
Blurring or Pixelation	Low to Medium	Low	Very Low
Digital Mask [8]	High	Medium	High
Adversarial Perturbation [3]	High	Medium	High
Differential Privacy [4]	High (theoretical)	Low	Medium
VERA (Edge + Encryption)	Medium	High	Low

2.7. Ethical and Responsible AI in Healthcare Vision Systems

Beyond technical privacy mechanisms, the deployment of AI systems in healthcare environments raises broader ethical considerations, including transparency, accountability, fairness, data governance, and human oversight. Several frameworks have been proposed to guide the responsible development of AI systems. The European Commission's High-Level Expert Group on AI defines Trustworthy AI by requiring privacy and data governance, technical robustness, and human agency [10]. Similarly, the IEEE Ethically Aligned Design guidelines and the NIST AI Risk Management Framework emphasize the importance of minimizing harm, ensuring traceability, and maintaining human control in safety-critical applications.

In healthcare specifically, ethical AI discussions highlight the risks of re-identification, data misuse, and loss of patient autonomy when visual monitoring systems are deployed. Scholars have argued that privacy-preserving technical design should be integrated with ethical-by-design approaches rather than treated as an afterthought [11]. Within this context, VERA contributes a privacy-supporting architecture that operationalizes key ethical AI principles by limiting exposure to identifiable data, enforcing controlled access to sensitive content, and enabling monitoring without direct visual disclosure.

In summary, VERA offers a practical privacy-supporting pipeline that avoids the computational overhead of generative or differentially private methods while preventing annotators and storage systems from accessing identifiable RGB data. This balance of privacy, practicality, and geometric fidelity positions VERA as a useful framework for real-time person detection in environments subject to strict privacy constraints.

3. System Architecture

The VERA framework is designed to operate in development and production modes to support both secure model training and real-time inference. The system architecture balances privacy preservation with detection performance by integrating modular compo-

nents for data acquisition, encryption, image obfuscation, annotation, augmentation, and deep learning inference.

The system is designed as a modular, privacy-supporting workflow that integrates edge-based obfuscation, encrypted RGB storage, and secure model training. Its purpose is to enable real-time and offline deep learning workflows while minimizing exposure to identifiable visual content.

3.1. Overview of the VERA Pipeline

The VERA framework is an end-to-end pipeline composed of two complementary modules: (i) a Production (Real-Time) module for live visual monitoring in secure environments such as hospitals or clinics, and (ii) a Development (Offline) module for model training, fine-tuning, and performance evaluation. Together, these components support privacy-oriented workflows while maintaining functional accuracy for human detection and activity recognition tasks.

The term “privacy-supporting” is used to distinguish VERA from strict anonymization systems. Edge detection reduces exposure of identifiable visual content but does not provide formal anonymization guarantees. Instead, VERA minimizes disclosure risk by restricting annotators to edge-based views and decrypting original images only within protected memory during training.

The framework is divided into two primary operational modes:

- Development Mode (Non-Real-Time): Supports secure model training by handling encrypted RGB images, manual annotations on edge-detected images, and dynamic image augmentation.
- Production Mode (Real-Time): Enables real-time image acquisition, inference using a pre-trained model, and secure on-device processing without compromising visual privacy.

The VERA workflow can be summarized as follows:

1. Data Acquisition: RGB images are captured from cameras or sensors within a private environment. No raw video data is stored; all subsequent processing occurs in-memory to reduce disclosure risks.
2. Edge-Based Obfuscation: Each captured image undergoes Canny edge detection to produce an abstract, texture-free representation that preserves geometric contours while concealing identity-revealing features.
3. Encryption and Secure Storage: Original RGB images are encrypted using the Advanced Encryption Standard (AES) in Cipher Feedback (CFB) mode. Encryption keys and initialization vectors (IVs) are stored separately under multiple security layers.
4. Manual Labeling and Annotation: Annotators label only the edge-based images using bounding boxes. These label files, combined with securely decrypted RGB images, form the dataset used for YOLO-based model training.
5. Data Augmentation and Loading: A dynamic augmentation module applies random geometric and photometric transformations to increase dataset variability.
6. Neural Network Training: A YOLOv5 model is trained using the augmented and securely decrypted data. Decryption occurs exclusively in memory.
7. Inference and Real-Time Deployment: The trained model performs real-time detection on edge images or RGB inputs, providing bounding boxes and class labels without revealing sensitive information during annotation.

VERA is designed for deployment on a range of hardware platforms. While compatible with lightweight devices such as Raspberry Pi and NVIDIA Jetson boards, the present work does not include hardware benchmarking. These evaluations are planned for future work.

3.2. Development Mode (Non-Real-Time)

The Development Mode in VERA focuses on securely preparing and training deep learning models using privacy-preserved data. Figure 1 shows the main components of this stage.

Input RGB images are first acquired from a dataset such as COCO [12] or from institutional sources. They are processed through two parallel paths:

- **Edge Detection Path:** A Canny edge detector transforms the RGB image into a contour representation. These images are used for annotation, preventing exposure to identifiable visual content.
- **Encryption Path:** The original RGB image is encrypted using AES-CFB and stored securely with its associated symmetric key and initialization vector.

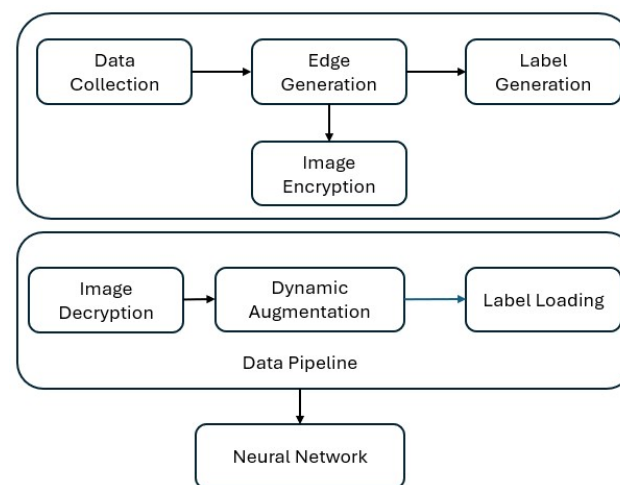


Figure 1. VERA development pipeline. All images displayed in this figure originate from the COCO public dataset. COCO manages image rights and permissions, and no protected health information or clinical imagery is used.

All examples shown in this section originate solely from the COCO dataset, ensuring that no private or clinical content is included in the manuscript.

Annotation is carried out using tools such as Labellmg [13]. Label files are stored separately from encrypted RGB images to maintain a strong separation between annotation data and identifiable content.

Dynamic image augmentation is applied in memory to decrypted RGB data, improving model robustness.

The lower module of Figure 1 illustrates the synchronized loading pipeline:

1. Secure in-memory decryption of RGB images
2. Augmentation with geometric and photometric transformations
3. Loading the corresponding annotation files

3.3. Production Mode (Real-Time)

Figure 2 illustrates VERA operating in production mode, supporting deployment in clinical and other privacy-sensitive environments.

- **Service 1: Inference mode** (Figure 2, top). Incoming frames are processed by the YOLOv5 model for real-time person detection.
- **Service 2: Secure data capture** (Figure 2, bottom). Incoming RGB images are transformed using edge-based processing for visualization and encrypted prior to storage.

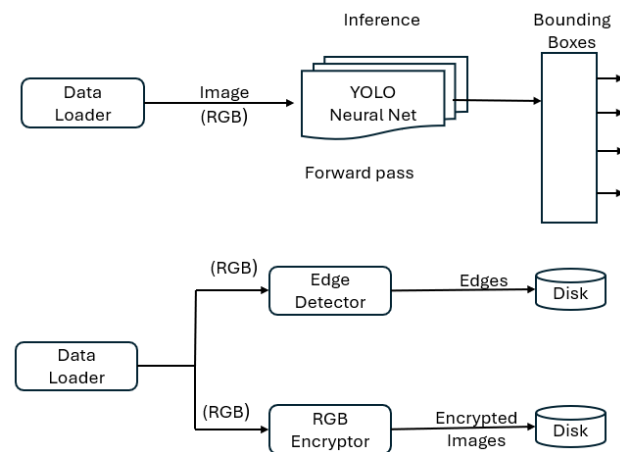


Figure 2. VERA production pipeline. All images shown originate from the COCO dataset, which contains only publicly released images and no protected clinical content.

All inference and data capture operations run entirely on local hardware to prevent unauthorized data transmission.

3.4. Security and Privacy Layer

VERA supports privacy-focused workflows by reducing exposure to identifiable content, encrypting RGB images at rest, enforcing integrity verification, and performing decryption exclusively in memory, and separating annotation views from original imagery. These combined practices help align the pipeline with privacy principles used in regulated environments.

3.4.1. Image Encryption Using AES-CFB

AES-CFB is used to encrypt RGB images.

Images are decrypted only within protected memory during training.

3.4.2. Edge Detection for Visual Anonymization

Canny edge detection removes color and texture while preserving geometric features needed for localization.

Although edge detection reduces exposure to identifiable content, it does not guarantee complete anonymity. VERA is therefore described as a privacy-supporting workflow rather than an anonymization algorithm.

3.4.3. Label Security and Linking Mechanism

Encrypted images are stored as files with the extension .enc. while labels are stored separately using YOLO-compatible format.

3.4.4. Key Management, Integrity, and Auditability

While AES-CFB provides confidentiality for RGB images stored at rest, additional mechanisms are required to ensure data integrity and authenticity. To address this, VERA complements symmetric encryption with message authentication, allowing detection of unauthorized modification of encrypted content prior to decryption.

Integrity and Authenticity Verification

For each encrypted RGB image, VERA associates a message authentication code (MAC) computed over the ciphertext and selected non-sensitive metadata. In practice, this can be realized using an HMAC constructed with a dedicated authentication key. Before any in-memory decryption is performed during training or data loading, the integrity tag is verified. If verification fails, the image is discarded and no plaintext is produced. This

prevents silent tampering with encrypted image blocks and ensures that only authenticated data are processed by the learning pipeline.

Key Separation and Secure Storage

VERA enforces logical separation between encryption keys and stored image data. Symmetric encryption keys and authentication keys are not colocated with encrypted RGB files. In institutional deployments, keys may be protected using standard security mechanisms such as hardware security modules (HSMs) or key management services (KMS), with access restricted to authorized components of the training or ingestion workflow. This separation follows the principle of least privilege and limits the impact of potential storage compromise.

Key Rotation and Lifecycle Management

To reduce long-term exposure risk, VERA supports periodic key rotation policies. Encryption and authentication keys can be rotated on a scheduled basis (e.g., per dataset batch or time interval), with previously used keys retained only as needed for controlled reprocessing or reproducibility. Retired keys can be revoked according to institutional retention and security policies.

Auditing and Accountability

VERA also supports auditing of cryptographic operations by recording non-sensitive events such as encryption, integrity verification, and decryption attempts. Audit logs include timestamps and operation outcomes but never expose plaintext data. This capability improves traceability and accountability in regulated environments without increasing privacy risk.

Alignment with Security Best Practices

Although VERA is presented as a research framework, the combination of confidentiality, integrity verification, key separation, rotation, and auditing aligns with widely adopted cybersecurity best practices in healthcare-oriented systems. These design choices are consistent with key management lifecycle recommendations and information security management principles commonly referenced in standards-based environments.

3.5. What Are the Novel Features of VERA?

The proposed framework, referred to as VERA, is designed to operate in two distinct modes: production (real-time inference) and development (offline training). This dual-mode operation enables VERA to maximize privacy without real-time mode detection performance or speed degradation. The only potential impact is limited to offline training speed, but not to the quality of the training process, and this is simply because increased use of protected system memory during the training process.

Unlike strict anonymization-based approaches, which permanently remove or obscure sensitive information and may degrade model accuracy, VERA employs a two-tier system architecture that selectively controls data accessibility. During the development phase, which corresponds to offline training, sensitive raw data are never directly exposed to human operators. Instead, the data labeling process relies on edge-detected image representations, ensuring that annotators interact only with privacy-preserving data formats rather than original images.

All data stored on persistent disk drives remain encrypted at rest, and decryption is strictly confined to protected memory regions. During offline training, decrypted images are accessed solely within these secure memory environments, and all intermediate training artifacts are retained in protected memory for the duration of the training process. While this approach introduces additional computational overhead—potentially slowing training

due to encryption, decryption, and secure memory management—it preserves the full informational content of the data for model optimization. This is probably the most important advantage compared to strict anonymization-based approaches.

In contrast, during the production phase, which corresponds to real-time inference, the inference is limited to edge devices only without strong any privacy sensitive data. Since inference is computationally less demanding compared to training, a lower cost and simpler edge computing type devices will be sufficient. As a result, VERA enables unhindered real-time processing and superior runtime detection performance compared to strict anonymization systems, which often sacrifice inference accuracy or speed to maintain privacy guarantees.

In summary, although the integration of encryption mechanisms and protected memory management increases overall system complexity and may reduce efficiency during offline training, VERA achieves a favorable trade-off by delivering high-performance, real-time inference without compromising data privacy, thereby outperforming traditional anonymization-based frameworks in production environments.

4. Image Processing and Model Training

This section presents the core components of the VERA pipeline used to transform raw images into training-ready data for deep learning. We describe the privacy-supporting image processing techniques, the secure labeling strategy, and the complete workflow for training person-detection models using encrypted RGB images and edge-based annotations.

4.1. Preprocessing and Edge Detection

VERA uses Canny edge detection as its primary obfuscation technique. The goal is to remove identifying features such as skin tone, facial texture, and clothing details, while retaining structural information required for annotation and detection.

Edge detection provides an abstracted view of the scene that reduces exposure to personally identifiable information while preserving geometric detail essential for bounding-box annotation.

4.1.1. Principle of Edge Detection

Edge detection locates regions in an image where the intensity gradient exhibits sharp discontinuities. If $I(x, y)$ represents the intensity at pixel (x, y) , the gradient magnitude $G(x, y)$ and direction $\theta(x, y)$ are defined as:

$$G(x, y) = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1}\left(\frac{\partial I / \partial y}{\partial I / \partial x}\right) \quad (2)$$

Regions with large gradient magnitude correspond to edges in the image. The Canny operator [14] was selected due to its robustness to noise and precise localization capability.

4.1.2. Canny Edge Detection Procedure

The Canny operator consists of the following stages:

1. Gaussian Smoothing: The input image is convolved with a Gaussian filter G_σ to reduce high-frequency noise:

$$I_s(x, y) = G_\sigma * I(x, y) \quad (3)$$

2. Gradient Computation: Sobel operators estimate horizontal and vertical derivatives.

3. Non-Maximum Suppression: Only local maxima in gradient magnitude are preserved, producing thin edges.
4. Hysteresis Thresholding: Two thresholds (T_{low} , T_{high}) classify strong and weak edges. Weak edges connected to strong edges are preserved.

The resultant binary edge map $E(x, y)$ is:

$$E(x, y) = \begin{cases} 1, & \text{if } G(x, y) \geq T_{low} \text{ and connected to a strong edge} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

4.2. Encryption and Secure Labeling

While edge-detected images are suitable for manual annotation (Figure 3), original RGB images must be encrypted to maintain confidentiality. Encryption uses the AES algorithm in Cipher Feedback (CFB) mode, which supports partial data streaming and is well suited for continuous camera feeds.

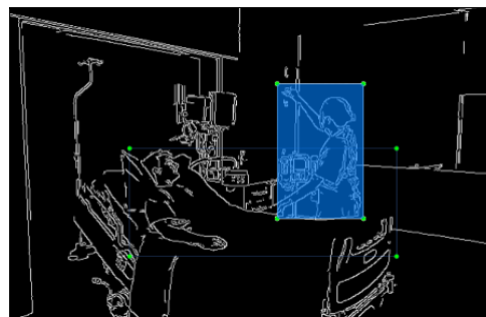


Figure 3. Secure labeling on edge-based images. All images displayed in this example originate from the COCO dataset, which contains publicly released images and no protected clinical information.

Edge-based annotation ensures that annotators do not access identifiable RGB content, supporting privacy-preserving labeling workflows.

4.2.1. Symmetric Encryption Scheme

Let P_i denote the i -th plaintext block, C_i the corresponding ciphertext block, K the secret key, and E_K the AES encryption function. AES-CFB encryption is defined as:

$$C_i = P_i \oplus E_K(C_{i-1}), \quad C_0 = IV \quad (5)$$

Decryption follows:

$$P_i = C_i \oplus E_K(C_{i-1}) \quad (6)$$

This stream-oriented mode supports block-wise encryption suitable for real-time image capture.

4.2.2. Implementation Details

Each RGB image is encrypted with a unique 256-bit key and a 128-bit initialization vector (IV). Encrypted files use the .enc extension, while the original RGB images are securely deleted or moved to protected storage. The following file structure is used:

- image_XXXX.jpg.enc: Encrypted RGB image file.
- image_XXXX.txt: Corresponding annotation file.
- image_XXXX.png: Edge-based annotation visualization.

Annotations follow the YOLO format:

```
<class> <x_center> <y_center> <width> <height>
```

This organization preserves a strict separation between encrypted RGB data and non-sensitive annotation data.

4.3. YOLO-Based Model Training

The YOLOv5 [15] model was selected due to its real-time performance and widespread adoption in embedded systems. Figure 4 illustrates the training process, which consists the following steps:

1. The encrypted RGB image is decrypted securely in memory.
2. The corresponding annotation file is loaded.
3. Optional augmentation is applied (Section 4.4).
4. The processed image and labels are passed to the YOLO network.

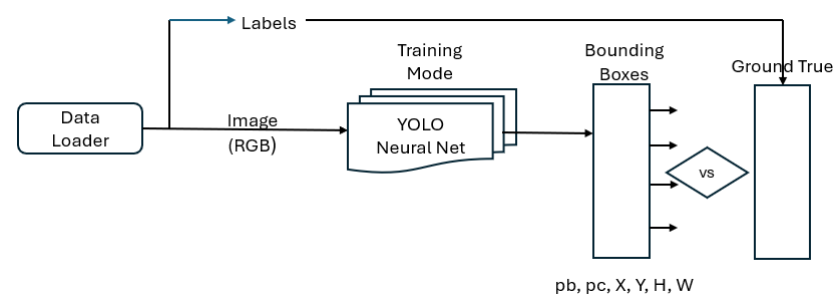


Figure 4. YOLO training workflow using encrypted-decrypted RGB images and edge-based annotations. All sample images originate from the COCO dataset.

During training, decrypted images exist only as temporary arrays in memory and are erased immediately after processing.

This design ensures that original RGB content is never written to disk in unencrypted form, aligning with privacy-supporting requirements in sensitive environments.

4.4. Dynamic Image Augmentation

To improve model generalization, VERA applies dynamic augmentation techniques during training. These include:

- Horizontal flipping
- Random cropping and scaling
- Rotation and affine transformations
- Brightness and contrast variation

Bounding boxes are recalculated after each transformation. Augmentation occurs after in-memory decryption and before network training.

Dynamic augmentation increases dataset variability and improves robustness by exposing the model to diverse geometric and photometric conditions.

5. Experimental Evaluation

This section presents a comparative analysis of two YOLO-based training configurations: a Baseline model trained on raw COCO images and a VERA model trained using encrypted RGB images, edge-based annotation, and secure data loading. Figures 5 and 6 illustrate the evolution of mean average precision (mAP) and loss during training for both configurations.

All images used in the experiments originate exclusively from the COCO dataset [12], a publicly released dataset that does not contain protected clinical or patient information.

5.1. Dataset

A subset of the COCO 2014 dataset was constructed focusing exclusively on the person class, which is particularly relevant for healthcare applications such as patient activity monitoring and fall risk assessment. Table 2 presents the details of the final subset and its subsequent division into training and testing sets.

Table 2. Dataset composition and experimental split.

Description	Details
Original dataset	COCO 2014 (over 40,000 labeled person images).
Curated subset	4000 images selected for experimentation.
Train/test split	3000 images for training (75%), 1000 for testing (25%).

Each image in the curated dataset underwent the following steps:

1. Encryption using AES in CFB mode.
2. Generation of an edge-detected version for annotation.
3. Association with YOLO-style annotation files produced either from COCO metadata or manual bounding-box labeling on edge images.

The dataset was filtered to remove ambiguous cases or views with limited visibility to strengthen annotation quality.

Using edge-based annotation ensures that annotators interact only with obfuscated representations, supporting privacy-preserving data handling without altering the underlying detection task.

5.2. Metrics and Setup

Table 3 summarizes the metrics used for quantitative evaluation, which follow standard object-detection methodology.

Table 3. Performance metrics used for model evaluation.

Metric	Description
Precision	Percentage of correct positive detections.
Recall	Percentage of actual positives correctly identified.
F1-score	Harmonic mean of precision and recall.
mAP@0.5	Mean Average Precision at 0.5 IoU threshold.
Average IoU	Spatial overlap between predictions and ground truth.
Inference latency	Forward-pass processing time during inference.

Training was performed under the configuration summarized in Table 4.

Table 4. Experimental setup and training configuration.

Parameter	Specification
Hardware	NVIDIA Tesla T4 (16 GB VRAM), 32 GB RAM, 4 vCPUs.
Framework	YOLOv5 (version 6.2) implemented in PyTorch version 1.13.1.
Batch size	16.
Epochs	Up to 1000 with early stopping.
Augmentation	Enabled (see Section 4.4).
Secure data loading	On-the-fly decryption and label loading.

5.3. Results and Analysis

The experimental comparison centers on two pipelines:

- **Baseline Model:** Direct training with raw RGB images and COCO annotations.
- **VERA Model:** Secure training using encrypted RGB images, in-memory decryption, and edge-based annotation.

5.3.1. Baseline Results

Table 5 shows the results of the training on raw RGB images without additional preprocessing (baseline). The model achieved only an mAP of 0.61% after 10 epochs, with high loss and an imbalance between true and false detections as shown in Figure 5.

Table 5. mAP calculation results for the Baseline model.

Metric	Value
Precision	0.08
Recall	0.26
F1-score	0.13
True Positives	22
False Positives	238
False Negatives	63
Average IoU	6.46%
mAP@0.50	0.006089 (0.61%)

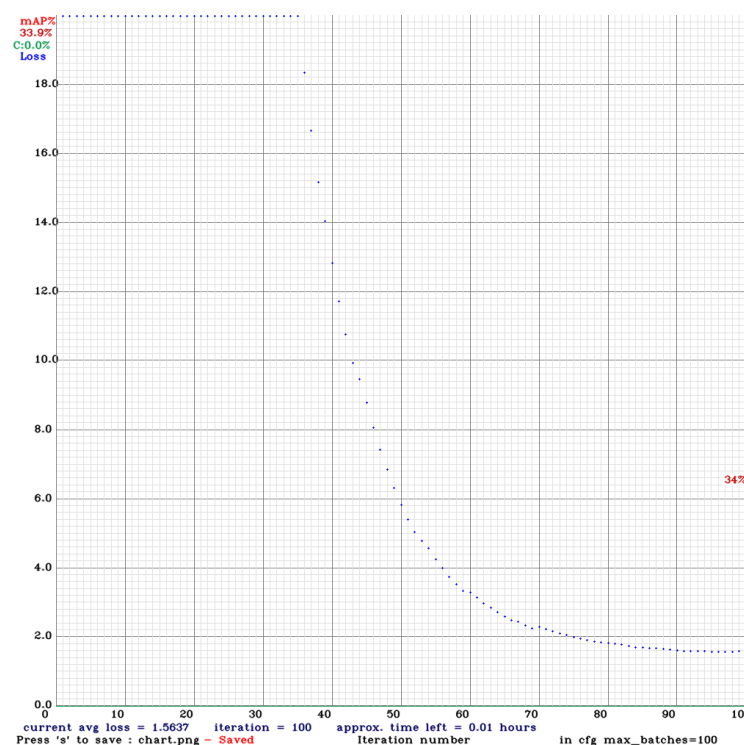


Figure 5. Baseline model: mAP versus loss. The mAP@0.50 value is too low to be shown on the plot. The red 34% line is included only as a reference value to reconcile the 0–100% percentage scale with the actual 0–18 range shown on the y-axis. All images presented are sourced from the COCO dataset.

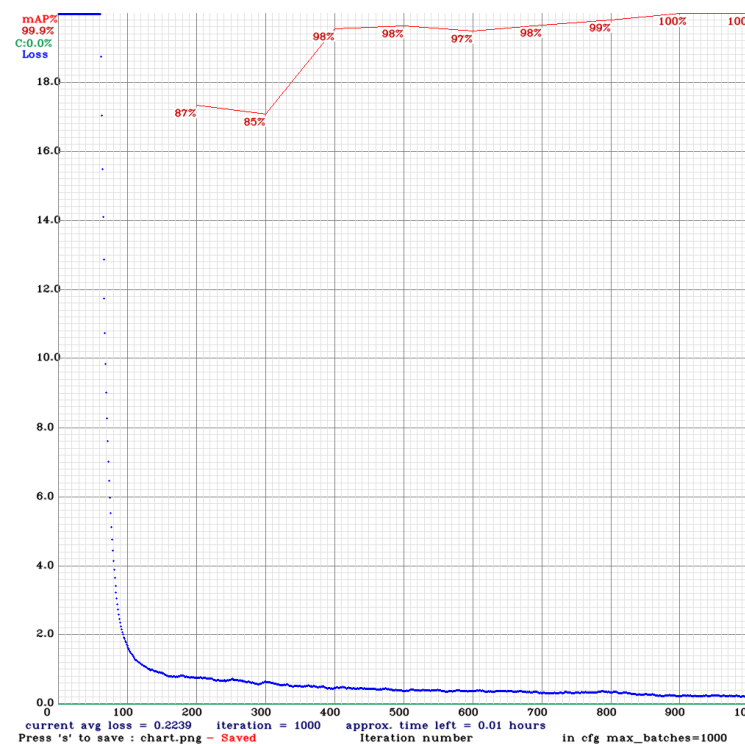
These results highlight the importance of a more structured and augmented training pipeline when working with limited data or unbalanced annotations.

5.3.2. Privacy-Preserved Training Results

Table 6 shows the metrics evaluated after training using the VERA methodology, which led to substantial improvements over the baseline. After fine-tuning for up to 1000 epochs with augmentation and secure data loading, the model converged to an mAP of 99.94%, with near-perfect precision and recall as shown in Figure 6

Table 6. mAP calculation results for the Fine-Tuned VERA model.

Metric	Value
Precision	0.99
Recall	1.00
F1-score	0.99
True Positives	85
False Positives	1
False Negatives	0
Average IoU	79.45%
mAP@0.50	0.999412 (99.94%)
Inference latency	18 ms per frame (GPU-based)

**Figure 6.** Fine-tuned model: mAP (red) versus loss over training epochs (blue). All images originate from the COCO dataset.

The improvement reflects the benefits of structured preprocessing, high-quality annotation, dynamic augmentation, and sufficient training duration rather than an inherent performance gain from privacy-preserving steps. The encryption–decryption pipeline does not degrade the training signal, demonstrating that secure handling of RGB data can coexist with high detection accuracy.

5.3.3. Qualitative Results

The VERA-trained model demonstrated accurate and stable person detection under varied visual conditions:

- partial occlusions
- variations in lighting
- multiple subjects per frame
- different viewpoints and subject poses

During visualization, bounding boxes were placed on edge-detected output images to maintain privacy in potential deployment environments.

5.3.4. Discussion of Results

The experimental results indicate that privacy-supporting training using encrypted RGB images and edge-based annotations can achieve performance comparable to traditional pipelines when combined with careful preprocessing and augmentation.

The substantial difference between the baseline and fine-tuned results is attributed primarily to improvements in training procedure (augmentation, number of epochs, and data quality) rather than to privacy mechanisms themselves. This distinction clarifies that VERA enables secure data handling without impairing model performance, but does not inherently improve accuracy beyond what a well-structured pipeline would achieve.

The average IoU of 79.45% supports the suitability of the model for downstream applications in person detection and monitoring workflows.

6. Applications and Use Cases

The VERA framework is designed to meet the stringent privacy, real-time processing, and accuracy requirements of clinical environments. Its modular architecture makes it suitable for a wide range of healthcare applications in which visual data are sensitive yet essential for decision-making.

VERA supports privacy-oriented operation by restricting annotation workflows to edge-based representations and ensuring that encrypted RGB images are never exposed during inference or monitoring. This enables visual analytics in environments where raw image disclosure is not permitted.

6.1. Real-Time Patient Movement Monitoring

One of the primary use cases for VERA is continuous observation of patients in sensitive environments such as hospitals, intensive care units (ICUs), and eldercare facilities. The system detects the presence and location of individuals in real time using person detection applied to privacy-preserved inputs.

VERA enables continuous visual monitoring without revealing identifiable image content, which helps support compliance with privacy regulations such as HIPAA and GDPR.

By visualizing detections on edge-based images rather than RGB frames, caregivers retain situational awareness while minimizing exposure to sensitive visual information.

6.2. Alert Systems for High-Risk Events

VERA can be integrated with alert or notification systems to support early detection of high-risk events, including:

- Falls: Detection of unusual posture or sudden transitions that may indicate a fall.
- movements: Alerts when mobility-restricted patients attempt to leave beds or designated safe areas.

These automatic alerts can improve patient safety and reduce the need for continuous physical supervision, particularly in settings with limited staffing.

6.3. Pressure Ulcer and Bedsore Prevention

Prolonged immobility increases the risk of pressure ulcers for patients undergoing extended treatments or with reduced mobility. VERA can help monitor patient posture over time by detecting how long a subject remains in a static position.

Future extensions may incorporate:

- Temporal analysis of posture sequences
- Heatmap-like visual summaries of patient positioning
- Automated reminders for caregiver-assisted repositioning

Integration with electronic health record (EHR) systems could further assist caregivers by documenting repositioning intervals and supporting preventive care routines.

6.4. Fall Detection and Prevention

Falls are a major health concern for elderly and mobility-limited populations. Approximately 1.6 million adults in the United States require emergency treatment annually for fall-related injuries. Early detection and prevention can significantly reduce these risks.

VERA can be extended with temporal models such as LSTMs or 3D CNNs to detect and analyze motion patterns leading up to a fall. Indicators may include:

- Abrupt or unstable body movements
- Loss of balance indicators
- Post-fall immobility prompting immediate response

Such extensions would require additional temporal datasets and would operate on edge-based motion streams to preserve visual privacy while enabling robust activity recognition.

6.5. Deployment in Hospitals and Nursing Homes

VERA's privacy-supporting design and lightweight computation requirements allow deployment across multiple healthcare environments, which may include:

- Nursing homes: Passive monitoring without exposing sensitive imagery.
- Hospitals: Assistance with real-time risk assessment and patient movement oversight.
- Clinics and home care: Remote or assisted monitoring for elderly individuals or patients requiring supervision.

VERA operates locally without storing or transmitting raw RGB data, which supports deployment in bandwidth-limited or distributed environments.

The framework can run on a variety of hardware platforms, including low-cost or embedded devices, provided that they support the required inference load.

Supported platforms include:

- Raspberry Pi for cost-effective installations
- NVIDIA Jetson Nano or Jetson Xavier for enhanced edge-based compute
- Standard workstations for centralized monitoring hubs

These deployment options allow VERA to scale from small clinics to large multi-room facilities while maintaining consistent privacy protections.

7. Limitations

Although the VERA framework demonstrates that privacy-aware pipelines can be effectively integrated into person detection workflows, several limitations should be acknowledged. First, the current evaluation focuses exclusively on static person detection rather than full Human Activity Recognition (HAR). Temporal models for fall detection or posture analysis are not yet implemented, and their integration remains future work. Second, all experiments rely on the COCO dataset rather than real hospital imagery. While COCO provides high annotation quality and diverse scenes, clinical environments present different lighting conditions, occlusions, and patient movements that must be validated in future studies. Third, although edge detection reduces identifiable visual information, it does not guarantee formal anonymity. No empirical analysis of re-identification risk was performed, and VERA should therefore be understood as supporting privacy compliance, not providing absolute de-identification. Fourth, the baseline model in the experimental section is intentionally minimal and does not represent a fully optimized YOLO configuration. A stronger baseline and cross-dataset evaluation will be incorporated in future

extensions. Finally, although the architecture is compatible with resource-constrained hardware such as Raspberry Pi or NVIDIA Jetson boards, no benchmarking results are reported. Quantitative measurements of latency, throughput, and encryption overhead on edge devices will be explored in subsequent versions of the framework.

8. Conclusions and Future Work

In this work, we presented VERA (Vision Expert Real Analysis), a privacy-supporting framework for real-time image-based data processing in healthcare environments. VERA enables secure training and deployment of deep learning models through encrypted image handling, edge-based annotation, and a modular architecture that supports both development and production modes.

The framework is designed to minimize exposure to identifiable visual content by restricting annotation workflows to edge-based representations and performing all decryption exclusively in memory. These design elements support the use of computer vision in sensitive clinical settings while reducing the risk of disclosing protected visual information.

Experimental results showed that the fine-tuned VERA model achieved a mean average precision (mAP) of 99.94%. This stands in contrast to the much lower baseline performance of 0.61%.

The improvement is attributed to structured preprocessing, consistent annotation quality, dynamic augmentation, and extended training rather than to the privacy mechanisms themselves. These findings demonstrate that privacy-supporting pipelines can be integrated into object detection workflows without degrading model performance.

The key contributions of VERA include:

1. Dual-mode architecture separating development and production workflows.
2. Edge-based obfuscation that preserves structural cues while reducing identifiable content.
3. AES-CFB encryption for secure storage and controlled access to RGB data.
4. Privacy-supporting annotation via non-identifiable edge images.
5. Dynamic image augmentation to improve model robustness.
6. Real-time inference capabilities suitable for edge devices.

VERA's modular structure enables flexible deployment across hospitals, nursing homes, private clinics, and home-care environments. It supports a range of hardware platforms, from embedded devices to workstation-class systems.

Author Contributions: Conceptualization, M.H.J., O.T. and L.G.J.; methodology, M.H.J. and L.G.J.; software, M.H.J.; validation, M.H.J., O.T. and L.G.J.; formal analysis, M.H.J.; investigation, M.H.J.; resources, L.G.J.; data curation, M.H.J.; writing—original draft preparation, M.H.J.; writing—review and editing, M.H.J., O.T. and L.G.J.; visualization, M.H.J.; supervision, O.T. and L.G.J.; project administration, L.G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable for studies not involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors acknowledge the support of Florida Polytechnic University in conducting this research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

VERA	Vision Expert Real Analysis
AI	Artificial Intelligence
HAR	Human Activity Recognition
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation
PII	Personally Identifiable Information
AES	Advanced Encryption Standard
CFB	Cipher Feedback Mode
YOLO	You Only Look Once
mAP	mean Average Precision
IoU	Intersection over Union
ICU	Intensive Care Unit
COCO	Common Objects in Context
RGB	Red Green Blue
IV	Initialization Vector
EHR	Electronic Health Record
DICOM	Digital Imaging and Communications in Medicine
DM	Digital Mask
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
GAN	Generative Adversarial Network

References

1. Dwyer, S.J., III; Weaver, A.C.; Hughes, K.K. Health insurance portability and accountability act. *Secur. Issues Digit. Med Enterp.* **2004**, *72*, 9–18.
2. Protection, F.D. General data protection regulation (GDPR). In *Intersoft Consulting, Accessed in October*; Woody's Express Parcels: Stornoway, UK, 2018; Volume 24.
3. Tong, C.; Zhang, M.; Lang, C.; Zheng, Z. An image privacy protection algorithm based on adversarial perturbation generative networks. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2021**, *17*, 1–14.
4. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Theory of Cryptography Conference*, New York, NY, USA, 4–7 March 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
5. Asaad, R.R.; Abdulrahman, S.M.; Hani, A.A. Advanced Encryption Standard Enhancement with Output Feedback Block Mode Operation. *Acad. J. Nawroz Univ.* **2017**, *6*, 1–10.
6. Lotan, E.; Tschider, C.; Sodickson, D.K.; Caplan, A.L.; Bruno, M.; Zhang, B.; Lui, Y.W. Medical imaging and privacy in the era of artificial intelligence: myth, fallacy, and the future. *J. Am. Coll. Radiol. JACR* **2020**, *17*, 1159.
7. Abouelmehdi, K.; Beni-Hessane, A.; Khaloufi, H. Big healthcare data: Preserving security and privacy. *J. Big Data* **2018**, *5*, 1–18.
8. Yang, Y.; Lyu, J.; Wang, R.; Wen, Q.; Zhao, L.; Chen, W.; Bi, S.; Meng, J.; Mao, K.; Xiao, Y.; et al. A digital mask to safeguard patient privacy. *Nat. Med.* **2022**, *28*, 1883–1892.
9. Xu, J.; Wei, L.; Wu, W.; Wang, A.; Zhang, Y.; Zhou, F. Privacy-preserving data integrity verification by using lightweight streaming authenticated data structures for healthcare cyber-physical system. *Future Gener. Comput. Syst.* **2020**, *108*, 1287–1296.
10. Smuha, N. A. The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International* **2019**, *20*, 97–106.
11. Shahriari, K.; Shahriari, M. IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In *Proceedings of the 2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, Toronto, ON, Canada, 20–21 July 2017; pp. 197–201.
12. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
13. Tariq, M.; Choi, K. YOLO11-Driven Deep Learning Approach for Enhanced Detection and Visualization of Wrist Fractures in X-Ray Images. *Mathematics* **2025**, *13*, 1419.

14. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, PAMI-8, 679–698.
15. Jocher, G.; Stoken, A.; Borovec, J.; Chaurasia, A.; Changyu, L.; Hogan, A.; Hajek, J.; Diaconu, L.; Kwon, Y.; Defretin, Y.; et al. ultralytics/yolov5: V5. 0-YOLOv5-P6 1280 models, AWS, Supervise. ly and YouTube integrations. *Zenodo* 2021. <https://doi.org/10.5281/zenodo.4679653>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.