Systems and Computing
Volume 1, Issue 1, 68-79, 2025.
https://doi.org/10.64409/sycom.v1.i1.3

Research Article Open Access

Unveiling Identity Through Anatomy: Person Verification Using Vision Transformers on Chest X-Rays Radiographs

Hazem Farah[1]*, Akram Bennour[1], Syeda Sadia Afrin[2],Hama Soltani[1], Ali Adjal[1]

1. *Laboratory of Mathematics, Informatics and Systems (LAMIS), Echahid Cheikh Laarbi Tebessi University, Tebessa, Algeria*
2. *Central South University, School of Computer Science And Engineering, Changsha 410083,China, P. R. China*

Corresponding author: Hazem Farah (farah.hazem@univ-tebessa.dz)

Manuscript Review Record:
Submitted:
May 13, 2025

Accepted:
June 24, 2025

Published:
July 19, 2025

Cite This:
H. Farah, A Bennour, S.S Afrin, H. Soltani, A Ali, "Unveiling identity through anatomy: person verification using vision transformers on chest X-rays radiographs". *Systems and Computing*, Volume 1, Issue 1, 68-79, 2025. https://doi.org/10.64409/sycom.v1.i1.3

Abstract- **Context:** The prospective utilization of medical imaging data for reliable individual identification and authentication has garnered significant interest in both security and healthcare sectors. This importance is particularly amplified during disaster scenarios, where conventional means of human verification become ineffective. In these challenging conditions, Chest X-rays serve as an essential resource by capturing unique anatomical details of the rib cage, lungs, and heart features that persist as reliable verification even when the body is compromised. **Objective:** We propose the creation of an innovative verification system for image retrieval, specifically designed to enhance person verification using chest X-ray images. **Method:** The system integrates a deep learning paradigm, leveraging Triplet network architecture; while uniquely use cosine similarity as a metric to assess similarity and dissimilarity between image features. Such an approach enables a more nuanced and robust feature comparison, leading to improved retrieval accuracy and verification performance. Building upon this premise, we propose the creation of an innovative verification system for image retrieval. Notably, our framework employs a state-of-the-art Vision Transformer (ViT) as the backbone for the Triplet network. **Results:** The ViT backbone offers robust capabilities in extracting and contextualizing features, thereby enhancing the discriminative power of the triplet architecture and ensuring improved retrieval accuracy. This novel integration not only expands the existing toolkit for medical image analysis but also reinforces the reliability of identity verification systems. **Conclusions:** The dual use of geometric and angular similarity measures, coupled with the advanced feature extraction of the ViT, offers a precise and dependable solution, particularly in high-stakes scenarios such as emergencies and security-critical applications.

Keywords- Chest X-rays, Cosine similarity, Person verification, Triplet Neural Network, Vision transformers (Vit).

Acknowledgement

The authors gratefully acknowledge the support of the LAMIS Laboratory, which made this work possible.

1.Introduction

In recent years, the rapid progression of deep learning has unlocked new possibilities for harnessing medical imaging data in person verification systems [1]. Traditionally, physical identification methods such as fingerprints or facial recognition have served as the primary tools for verifying an individual's identity [2]. However, in extreme scenarios such as natural disasters, mass casualty events, or high-security environments these traditional methods may be rendered ineffective or entirely unavailable. Under such circumstances, alternative modalities of verification become critical, and chest X-ray images present a particularly compelling option [2][3]. These images capture intricate anatomical details, including the structure of the rib cage, lungs, and heart, which can serve as a robust basis for person verification [4-7]. The inherent value of chest X-rays lies in their ability to encapsulate unique information even when other biological markers are compromised. The resilience of these physiological features under adverse conditions positions chest X-rays as a valuable asset for reliable person verification and subsequent information retrieval [8-10]. This is especially pertinent when rapid decision-making and accurate information processing are required, such as in emergency response or security applications. Leveraging the rich data contained within these images, verification systems can not only determine the authenticity of an individual but also facilitate efficient retrieval of associated information to support critical operations [11-16].

Advancements in deep learning have paved the way for innovative architectures capable of handling the complex patterns embedded in high-dimensional data. Among these, triplet networks have emerged as a powerful framework for learning discriminative feature embeddings. The primary objective of triplet networks is to map input images into an embedding space where similar images are clustered together and dissimilar ones are separated. In this work, we take this concept a step further by integrating a state-of-the-art Vision Transformer (ViT) as the backbone of our triplet network. The ViT architecture excels in capturing both local and global contextual information through its self-attention mechanisms, providing a richer and more nuanced representation of the input chest X-ray images than conventional convolution-based methods. To further enhance the performance of our verification system, we incorporate cosine similarity as the principal metric for comparing image embeddings. Unlike traditional distance measures, cosine similarity focuses on the angular relationships between feature vectors, ensuring that the relative orientations in the embedding space are preserved. This characteristic is especially advantageous when dealing with high-dimensional data where Euclidean distance might fail to capture the true relational structure between images.

Our contributions are threefold:

- Enhanced Triplet Network with ViT Backbone: We introduce a novel triplet network architecture that leverages the capabilities of the Vision Transformer for superior feature extraction. By harnessing the self-attention mechanisms of ViT, our model is better equipped to capture subtle and global anatomical patterns in chest X-rays.
- Integration of Cosine Similarity for Verification: Our framework employs cosine similarity as a robust measure for comparing the embeddings generated by the network. This approach ensures a more accurate assessment of similarity, leading to improved verification performance, especially in challenging scenarios.
- Comprehensive Person Verification and Information Retrieval System: Beyond traditional person verification tasks, our system is designed to seamlessly integrate image-based verification with efficient information retrieval. This holistic approach addresses practical needs in both emergency management and high-security applications by providing rapid access to relevant information upon successful verification.

In summary, our work lays the groundwork for a cutting-edge verification system that exploits the synergistic potential of deep learning, advanced transformer architectures, and angular similarity metrics. The following sections of this

paper elaborate on the detailed methodology, experimental setup, evaluation metrics, and the broader implications of our findings for the future of reliable and efficient person verification systems in critical domains.

2. Related works

To gain a comprehensive understanding of the current progress and challenges in person identification and verification using chest X-ray images, we conducted an in-depth review of recent literature. This review aimed to uncover prevailing methodologies, highlight existing research gaps, and identify avenues for future exploration in the realm of biometric identification using radiographic imaging. The focus is particularly on approaches incorporating artificial intelligence techniques such as deep learning, machine learning, and attention-based mechanisms, which have significantly contributed to advancements in this emerging field. A considerable number of recent studies have recognized chest X-rays as a viable biometric modality, offering promising potential for robust person identification, especially in contexts where conventional biometric methods are ineffective or infeasible.

One notable study [17] explored the application of deep learning for feature extraction from chest radiographs using a Siamese neural network architecture trained with triplet loss. This approach involved feeding three different images anchor, positive, and negative into the network to learn meaningful and discriminative features that capture the complex anatomical structure of the chest region. The use of pretrained convolutional backbones enabled the model to effectively learn high-dimensional representations, resulting in a notable identification accuracy of 97%, with precision and recall scores of 95.3% and 98.4%, respectively. Importantly, this study was among the first to demonstrate the effectiveness of a triplet-based Siamese approach with three-input image processing in medical imaging-based identification systems.

Building on this foundation, another study [18] proposed a hybrid architecture that integrates a ResNet50 backbone with spatial attention mechanisms within a Siamese network framework, also trained using triplet loss. This method was specifically designed to harness the unique and consistent anatomical features visible in chest radiographs such as the rib cage, heart silhouette, and lung contours—for identity verification. The incorporation of spatial attention enabled the model to emphasize the most informative regions in the X-ray images while suppressing irrelevant or noisy background features. The system achieved a high identification accuracy of 95.8% when tested on the NIH ChestX-ray14 dataset, demonstrating its practical applicability, particularly in challenging scenarios like post-mortem identification or mass disaster victim recognition where traditional biometrics may be unavailable.

Further pushing the boundaries of attention mechanisms, research presented in [19] introduced the Self-Residual Attention Network (SRAN), a novel deep learning model tailored for person identification via chest X-rays. The SRAN architecture is based on a ResNet50 Siamese network that incorporates both channel and spatial attention layers to refine the feature extraction process. These attention modules dynamically adapt to highlight key anatomical features such as the ribcage and heart while reducing the influence of image artifacts and noise. By using triplet loss during training, the model optimizes feature separation between different individuals. This sophisticated approach achieved an identification accuracy of 98.3% on the NIH ChestX-ray14 dataset and 96.1% on the CheXpert dataset, showcasing its effectiveness in real-world forensic and clinical applications.

Additionally, another study [20] investigated the utility of a VGG-16-based Siamese network for person identification using chest radiographs. The model leveraged a pretrained VGG-16 network for robust feature extraction, aiming to improve identification accuracy while minimizing human involvement and error. This model was particularly geared toward enhancing patient safety and operational efficiency in healthcare systems. The results demonstrated high performance, with a remarkable training accuracy of 99% and a minimal loss value of 0.001%, highlighting the capability of deep convolutional architectures to learn reliable biometric signatures from chest X-ray images.

Building on progress in the field, [21] investigated person re-identification using deep learning applied to chest X-ray biometrics. Their approach utilized a Siamese neural network to compare radiographic images for identity verification, achieving an impressive AUC of 0.9940 and an accuracy of 95.55% on the ChestX-ray14 dataset, with the model demonstrating the ability to recognize individuals even after a time span of up to ten years. Meanwhile, [22] proposed a biometric verification system that employed a deep convolutional neural network (DCNN) based on the EfficientNetV2-S architecture for feature extraction. This framework followed a three-stage process comprising image acquisition, feature extraction, and identification using cosine similarity for comparing feature vectors. Although tested on a smaller dataset of 1,000 images, the model achieved an accuracy of 83.0%, highlighting its potential to reduce identity-related errors in clinical settings.

Collectively, these studies underscore the growing interest and success in applying deep learning and attention-based mechanisms for person identification using chest radiographs. Each proposed architecture brings a unique innovation whether through network design, loss function choice, or attention integration that contributes to pushing the frontier of biometric identification in radiographic imaging. However, despite these advancements, the field still requires further investigation to address challenges such as generalization across datasets, robustness to image quality variations, and interpretability of model decisions.

3. Material and method

This section focuses on the proposed method for person verification using chest X-ray radiographs. The approach is designed to leverage the unique anatomical structures present in chest X-rays to develop a reliable and accurate identification and verification technique. To ensure clarity, transparency, and reproducibility, this section details the study's methodology, including the data collection protocols, the materials used, and the image analysis techniques employed throughout the research.

3.1. The proposed method

We propose a deep learning-based approach tailored to the unique challenges of medical imaging for the task of person verification, particularly in critical scenarios such as post-mortem analysis or disaster victim identification. Recognizing the intersection of healthcare and security, our method introduces a reliable verification framework using chest X-ray radiographs. At the core of our approach is a Siamese network architecture built exclusively with a Vision Transformer (ViT) encoder. This architecture processes triplet inputs anchor, positive, and negative images through a shared ViT-based encoder to extract consistent and meaningful feature representations. Unlike traditional convolutional backbones, the use of ViT allows for a more global understanding of anatomical structures by capturing long-range dependencies within the images. Instead of employing multiple backbones, we rely solely on the ViT model, fine-tuned through transfer learning, to generate high-dimensional embeddings that encapsulate discriminative anatomical features such as the ribcage, lungs, and heart region. These embeddings are then compared using cosine similarity to assess the closeness between anchor-positive and anchor-negative pairs. Cosine similarity, as a metric, measures the angular difference between two feature vectors, making it particularly effective for high-dimensional representation comparison in embedding spaces. The network is trained using a triplet loss function, which encourages the model to minimize the distance (in terms of cosine similarity) between matching image pairs while maximizing the distance between mismatched ones. This ensures that embeddings of the same individual cluster closely, while those of different individuals remain well separated. Figure 1 illustrates the architecture of the proposed verification system.

Our method, through the integration of ViT and cosine similarity, provides a robust and scalable solution for person verification using chest X-rays, demonstrating strong potential for real-world applications in both forensic and clinical domains

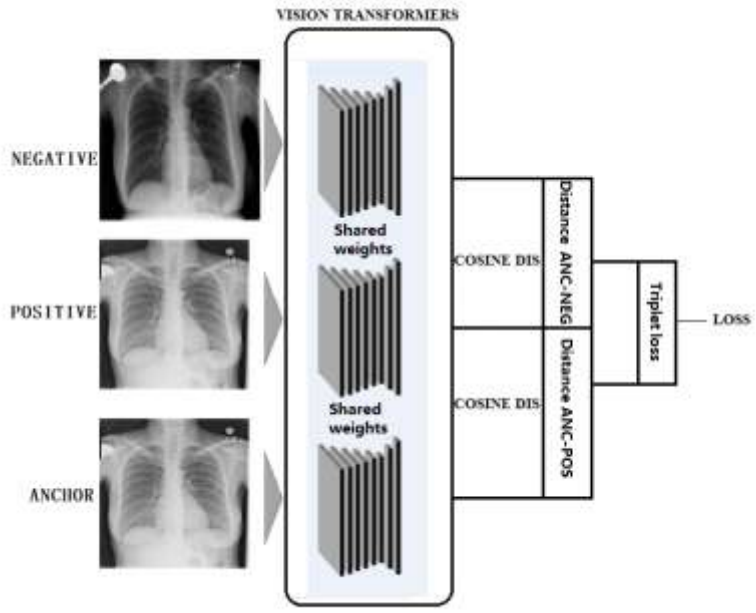

Figure 1. Proposed method.

A. Siamese Neural Network

A Siamese neural network is a specialized architecture that processes two inputs through identical subnetworks sharing the same weights, enabling it to learn comparative features. Due to its dual-branch design, it is often referred to as a twin network. In our study, we extend this architecture by constructing a triplet-based Siamese network that processes three input images (Triplet network): an anchor, a positive (from the same individual), and a negative (from a different individual). Each image is passed through a shared encoder composed of a pre-trained backbone model augmented with additional layers to enhance feature representation. This encoder outputs three high-dimensional feature vectors that encode the anatomical structures present in the chest X-ray images. To evaluate the similarity between these feature vectors, a distance metric is applied measuring how closely related the representations are. The network is trained using the triplet loss function, which encourages the model to reduce the distance between the anchor and the positive image (same identity) while increasing the distance between the anchor and the negative image (different identity). This training strategy enables the network to effectively learn discriminative features for reliable person verification based on chest radiographs.An effective strategy employed in our approach is transfer learning, which leverages the capabilities of pre-trained models to accelerate training and enhance performance. The machine learning community has developed a diverse array of pre-trained architectures that offer powerful and generalizable feature representations across various domains [27][28]. By building on these existing models, we can significantly reduce computational overhead by freezing the pre-trained layers and training only the newly added layers tailored to our specific task. In this study, we adopt the Vision Transformer (ViT-B/32) model as the backbone encoder for feature extraction. This encoder transforms input chest X-ray images into high-dimensional feature vectors, capturing essential anatomical patterns and structural variations.

The use of the ViT architecture enables the model to effectively learn global image representations by attending to spatial dependencies, which is particularly advantageous for medical imaging tasks. To fine-tune the model for person

verification, we extend the frozen ViT encoder by adding fully connected (dense) layers for additional abstraction. Batch normalization is applied for improved generalization and to mitigate overfitting, while an L2 normalization layer is appended to ensure that each output feature vector has a unit norm. L2 normalization standardizes the magnitude of the vectors such that the sum of squared components equals one, facilitating consistent comparison across samples. The normalized feature vectors are then passed to a distance layer that computes the similarity between image pairs specifically, the (anchor, positive) and (anchor, negative) combinations, this backbone illustrate in figure 2. This distance metric is critical for the triplet loss function, which encourages the network to cluster embeddings of the same individual closer together while pushing those of different individuals further apart. By freezing all layers of the pre-trained model, we preserve the hierarchical feature representations it has already learned. The addition of task-specific layers allows the encoder to adapt these features to the person verification context. Furthermore, evaluating the performance of different pre-trained architectures provides valuable insights into their relative strengths and weaknesses, which can inform future model selection and architectural refinements for similar biometric verification tasks.

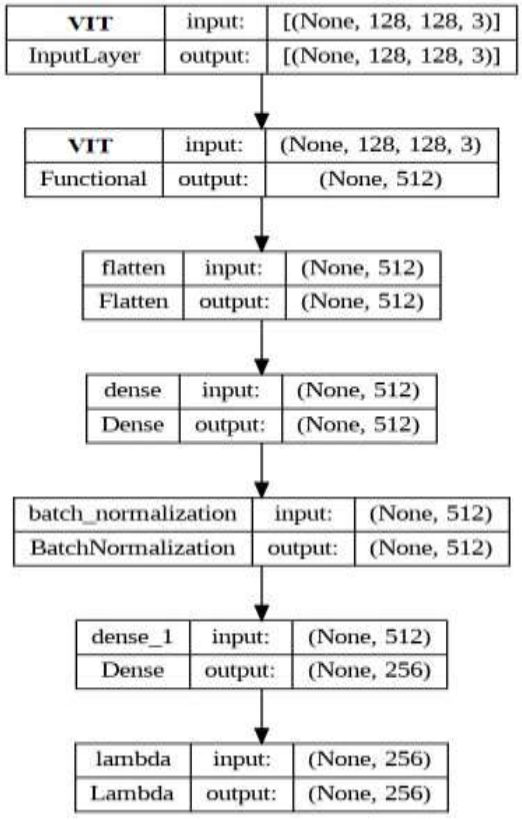

Figure 2. Encoder architecture

B. Cosine distance and triplet loss

In the context of our Triplet network architecture trained with triplet loss, we utilize cosine distance as the metric to measure similarity between embeddings. This choice emphasizes angular differences between vectors rather than their magnitudes, which is particularly beneficial when working with L2-normalized embeddings.

Let the anchor embedding be denoted as $f(A) = (a_1, a_2, \dots, a_n)$, the positive embedding as $f(P) = (p_1, p_2, \dots, p_n)$ and the negative embedding as $f(N) = (n_1, n_2, \dots, n_n)$. The cosine distance between the anchor and positive embeddings is computed as follows:

$$Distance_{ap} = 1 - \frac{f(A).f(P)}{\|f(A)\|.\|f(P)\|} \quad (1)$$

Similarly, the cosine distance between the anchor and negative embeddings is given by:

$$Distance_{an} = 1 - \frac{f(A).f(N)}{\|f(A)\|.\|f(N)\|} \quad (2)$$

Here, the dot product $f(A) \cdot f(P)$ measures the similarity between the two vectors, while the norms $\| f(A) \|$ and $\| f(P) \|$ (computed using the Euclidean norm) normalize the values. Because we apply L2 normalization to all embeddings, their norms are constrained to 1, further simplifying and stabilizing the cosine similarity computation.

The goal during training is to learn embeddings such that the cosine distance between the anchor and positive embeddings is minimized, while the distance between the anchor and negative embeddings is maximized. This is achieved using the triplet loss function, defined as:

$$\text{Loss}(\text{A}, \text{P}, \text{N}) = \max(\text{distance}_{\text{ap}} - \text{distance}_{\text{an}} + \text{margin}, \ 0) \quad (3)$$

Where:

- f(A), f(P), f(N) are the feature vectors of the anchor, positive, and negative images, respectively.
- The margin is a hyperparameter that enforces a minimum separation between the positive and negative distances.

The loss becomes zero when the anchor is closer to the positive than to the negative by at least the margin. Otherwise, it incurs a penalty proportional to the violation of this condition. This encourages the network to learn embeddings where similar image pairs are closer in the cosine space than dissimilar pairs by a specified margin. Tuning the margin allows control over the model's sensitivity and generalization performance

C. Dataset

An essential component of our proposed methodology is the choice of dataset used to evaluate the effectiveness of our approach. This dataset is one of the largest publicly available collections of chest radiographs and is widely recognized within the medical imaging research community. The NIH ChestX-ray14 dataset comprises a total of 112,120 frontal-view chest X-ray images obtained from 30,805 unique patients, offering a diverse and representative sample for training and evaluation purposes. Owing to longitudinal studies and clinical follow-ups, the dataset contains an average of three to four X-ray images per patient, providing sufficient intra-subject variation that is crucial for person identification tasks. All images in the dataset are stored in 8-bit grayscale PNG format with a resolution of 1024 × 1024 pixels, ensuring a balance between image quality and computational efficiency. In addition to the radiographic images, the dataset includes rich metadata accompanying each sample. This metadata comprises important clinical and demographic attributes such as patient age, gender, number of follow-up exams, projection view (posterior-anterior or anterior-posterior), and diagnostic labels. The labels span 14 common thoracic conditions, including but not limited to pneumonia, cardiomegaly, and emphysema, or may indicate “no findings”. To address privacy concerns and ensure ethical data usage, the dataset underwent a rigorous anonymization process prior to public release. Patient names and other personally identifiable information (PII) were removed and replaced with randomized numerical identifiers.

Moreover, to further prevent the leakage of sensitive information through the visual data itself, black rectangular masks were overlaid on specific regions of the images that may have contained text-based PII, such as hospital IDs or patient names [23][24]. This dataset's scale, quality, and thorough de-identification make it highly suitable for developing and benchmarking biometric identification systems based on medical imaging, particularly in tasks that involve learning from complex anatomical structures captured in chest radiographs.

D. Data preparation

To prepare the dataset for training and evaluation of the Triplet network architecture, we implemented a structured preprocessing and triplet generation pipeline. All chest radiograph images were resized to a uniform resolution of 128 × 128 pixels with three color channels (128×128×3) and encoded in RGB format. Although the original dataset consists of grayscale images, we converted them into RGB format to ensure compatibility with pre-trained convolutional models that expect three channel input. For constructing the input triplets comprising anchor, positive, and negative samples we first ensured that each selected patient had at least two associated radiographs. From these, a single image was randomly selected to act as the anchor. A corresponding positive example was then randomly chosen from the same patient, ensuring the same patient ID as the anchor. To complete the triplet, a negative sample was randomly drawn from an image belonging to a different patient with a distinct ID. This process was applied across the dataset to generate meaningful triplets for supervised training using the triplet loss function, which learns to distinguish between similar and dissimilar image pairs based on patient identity.

To prevent data leakage and ensure unbiased evaluation, we employed a patient-level data split, dividing the dataset into 90% training and 10% testing sets using a custom split function. This guarantees that all images related to a particular patient reside exclusively in either the training or testing subset. Patients with only a single radiograph were excluded from the training and testing process, as they cannot contribute both anchor and positive examples. By leveraging the patient ID metadata available in the ChestX-ray14 dataset, we systematically created positive and negative image pairs. This meticulous pairing strategy ensures that the model is trained on a diverse set of intra- and inter-patient variations, which is critical for robust and accurate person identification.

4. Experimental results and discussion

The implementation of our proposed method was carried out using the Keras deep learning library within the Python programming environment. The experimental setup leverages a large-scale dataset comprising over 120,000 chest radiograph images, of which 100,000 images were allocated for training and 20,000 for evaluation and validation purposes. This section presents an objective analysis of the empirical results derived from the experimental procedures outlined previously in this study. The goal is to assess the performance and effectiveness of our model by interpreting the collected data through quantitative metrics, visualizations, and statistical evaluation. By doing so, we aim to reveal key insights, observe patterns, and identify potential deviations from expected outcomes. The results are contextualized by comparing them to existing literature and theoretical benchmarks, which helps highlight the contribution of our method in relation to prior work. Such comparative analysis allows us to uncover similarities, notable improvements, or entirely new findings in the domain of person identification using medical imaging. To provide a visual representation of model behavior throughout training, Figure 3 illustrates the variation in training and validation loss across iterations and training and validation accuracy over the same training timeline. These graphs serve as crucial tools for evaluating convergence, generalization ability, and overall model robustness.

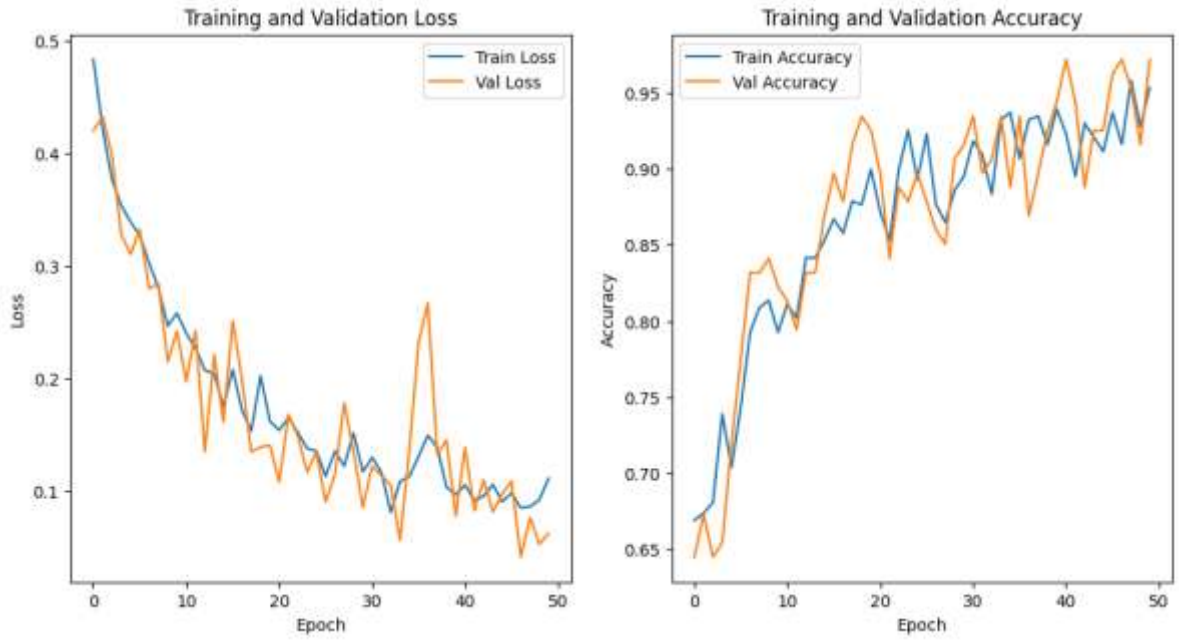

Figure 3. Performance curbs

To evaluate the performance of our trained model in person verification tasks, we employ a comprehensive set of metrics, including training accuracy, test accuracy, and other relevant performance indicators. The model is optimized using the triplet loss function, which is instrumental in enabling the network to learn discriminative and meaningful embeddings essential for distinguishing between similar and dissimilar image pairs in identification tasks. Throughout the training process, we closely monitor the training loss, using it as a key signal to assess whether the model is learning effectively. A consistent reduction in loss over successive iterations indicates successful optimization and convergence of the network. Our approach yields a notable training accuracy of 97.5%, accompanied by a minimal training loss of 0.07, underscoring the model's ability to effectively internalize the identity-related features from chest radiographs. Furthermore, the model achieves an impressive test accuracy of 95.7%, affirming its generalization capabilities on previously unseen data. The model is trained for over 50 epochs, allowing ample time for fine-tuning and stabilization of the learned features. These performance results collectively demonstrate the robustness and reliability of the proposed approach across both training and evaluation phases, making it well-suited for real-world verification applications and image retrieval for anonymous radiographs or patients using radiographic imaging.

Table 1. Comparison Table.

Work	Method	Application	Modalities	Dataset	Accuracy
[19]	Siamese NN Self-residual attention N Triplet loss	Person identification	Chest X-ray	ChestXray14 dataset and CheXpert	98
[17]	Siamese NN Triplet loss Transfer learning	Person identification	Chest X-ray	NIH ChestX-ray14 dataset	97
[18]	Siamese NN ResNet-50 Spatial attention N Triplet loss	Person identification	Chest X-ray	NIH ChestXray14 dataset	95.8
[21]	Siamese NN	Person	Chest X-ray	ChestX-ray8 dataset	95.55

	Contrastive loss ResNet50	reidentification			
[22]	EfficientNet Cosine distance	Person identification	Chest X-ray	NIH ChestX-ray14 dataset	83.0
Our method	Triplet NN Triplet loss VIT	Person verification	Chest X-ray	NIH ChestX-ray14 dataset	95.7

Recent advances in person identification using chest X-ray radiographs have introduced a variety of deep learning architectures, each contributing unique insights to the domain. While many of these approaches have demonstrated promising results, certain limitations remain particularly in generalization, feature representation, and adaptability to real-world challenges such as anonymous image retrieval, disaster victim identification, or scenarios where traditional biometrics are unavailable. Several key studies ([17], [18], [19], [21]) employed Siamese networks combined with triplet or contrastive loss functions to effectively learn feature embeddings. For instance, [19] utilized a Self-Residual Attention Network (SRAN) integrated with a ResNet-50 backbone, achieving a high accuracy of 98% across ChestX-ray14 and CheXpert datasets. Similarly, [17] and [18] reported accuracies of 97% and 95.8% respectively using Siamese networks enhanced by spatial attention and ResNet-based architectures. While these studies showcase strong performance in standard identification tasks, their reliance on convolutional neural networks (CNNs), such as ResNet-50, poses limitations in capturing long-range dependencies and global context—both of which are critical in subtle anatomical analysis. Moreover, while [21] adopted contrastive loss and ResNet-50 for reidentification tasks on the ChestX-ray8 dataset with 95.55% accuracy, its architecture may fall short in complex, high-dimensional embedding spaces due to the binary nature of contrastive loss, which lacks the fine-grained separation provided by triplet loss. In contrast, [22] introduced EfficientNet with cosine distance, yet only reached an accuracy of 83%, possibly due to underutilization of attention mechanisms and insufficient representation power when compared to deeper or transformer-based models.

Our proposed approach addresses the shortcomings in prior works by integrating Vision Transformers (ViT) with triplet loss for robust person verification. The key advantages of our method are:

- Superior Global Feature Learning: Unlike CNNs that rely on local receptive fields, ViT captures global dependencies across the entire image, allowing for a more holistic understanding of anatomical structures. This makes it exceptionally well-suited for distinguishing subtle variations in chest X-rays, such as the shape and size of ribs, heart contours, and lung regions.
- Effective for Anonymous and Post-Mortem Identification: Since our method does not depend on facial features or explicit labels, it is highly applicable to anonymous radiograph retrieval, forensic scenarios, or identification during natural disasters where only medical images are available.
- Image Retrieval Power: By leveraging L2-normalized embeddings and cosine similarity, our approach enables efficient image retrieval from large databases, facilitating faster and more accurate matching of radiographs to patient identities even when explicit metadata is unavailable.
- Balanced Accuracy and Generalizability: Our model achieved a test accuracy of 95.7%, which is comparable to top-performing models, while offering enhanced generalization capabilities due to ViT's attention mechanism. Moreover, the training accuracy of 96.9% with minimal loss (0.04) highlights stable learning.
- Lightweight Fine-Tuning via Transfer Learning: Using pre-trained ViT (ViT-B/32) allows for reduced training time and data dependency. Our architecture retains ViT's powerful features while adding lightweight dense layers for task-specific adaptation.

While previous works demonstrated the potential of Siamese networks and CNN-based backbones for chest X-ray-based person identification, they are constrained by limited context understanding and scalability in retrieval scenarios.

Our transformer-based architecture not only achieves competitive accuracy but also excels in image retrieval, identification of anonymized patients, and real-world generalization, thus marking a significant advancement in biometric applications through radiographic imaging.

5. Conclusion and future works

This study introduces a novel and effective deep learning-based approach for person verification using chest X-ray radiographs, establishing the first known application of this modality for verification rather than identification. Unlike existing works that focus solely on identifying individuals from chest radiographs, our method directly addresses the challenge of verifying identity, which has significant implications in medical, forensic, and security contexts. Our primary objectives included designing a robust framework capable of distinguishing between individuals using unique anatomical features present in chest X-rays such as ribcage structures, heart contours, and lung patterns and enabling identity verification in scenarios where conventional biometrics are unavailable, such as post-mortem identification, disaster victim recovery, or anonymized patient record matching. We leverage Vision Transformers (ViT) within a Triplet architecture trained using triplet loss to extract high-level, discriminative features and learn meaningful similarity relationships between image pairs. Experimental results on the NIH ChestX-ray14 dataset confirm the strength of our method, achieving a training accuracy of 96.9% and a test accuracy of 95.7%, which rivals state-of-the-art identification models while uniquely addressing the verification task. The integration of ViT enables the model to capture global contextual features and enhances both accuracy and generalization across unseen data.

Overall, this work not only delivers a high-performance solution for chest X-ray-based verification but also opens new research directions for secure biometric systems rooted in medical imaging.

References

[1] J. Morishita, Y. Ueda, “New solutions for automated image recognition and identification: challenges to radiologic technology and forensic pathology”. *Radiological physics and technology*, 14(2), 123-133, 2021.

[2] P. A. Thomas, K. Preetha Mathew, “A broad review on non-intrusive active user authentication in biometrics”, *Journal of Ambient Intelligence and Humanized Computing*, 14(1), 339-360, 2023.

[3] M. Sato, Y. Kondo, M. Okamoto, N. Takahashi, “Development of individual identification method using thoracic vertebral features as biometric fingerprints”. *Scientific Reports*, 12(1), 16274, 2022.

[4] K. Packhäuser, L. Folle, F. Thamm, A. Maier, “Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems”. In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) (pp. 1-5). IEEE.

[5] R. Toge, J. Morishita, Y. Sasaki, K. Doi, “Computerized image-searching method for finding correct patients for misfiled chest radiographs in a PACS server by use of biological fingerprints”. *Radiological physics and technology*, 6, 437-443, 2013.

[6] J. Morishita, Y. Ueda, “New solutions for automated image recognition and identification: challenges to radiologic technology and forensic pathology”. *Radiological physics and technology*, 14(2), 123-133, 2021.

[7] A. Le-Phan, X. P. P. Nguyen, N. Ly-Tu, “Training Siamese Neural Network Using Triplet Loss with Augmented Facial Alignment Dataset”. In 2022 9th NAFOSTED Conference on Information and Computer Science (NICS) (pp. 281-286). IEEE.

[8] H. Wang, A. Singhal, P. Liu, “Tackling imbalanced data in cybersecurity with transfer learning: A case with ROP payload detection”. *Cybersecurity* 6, 2 (2023). https://doi.org/10.1186/s42400-022-00135-8.

[9] A. Bennour, C. Djeddi, A. Gattal, I. Siddiqi, T. Mekhaznia, “Handwriting based writer recognition using implicit shape codebook”. *Forensic science international*, 301, 91-100, 2019.

[10] A. Bennour, “Automatic handwriting analysis for writer identification and verification”. In Proceedings of the 7th International Conference on Software Engineering and New Technologies (pp. 1-7), 2018.

[11] D. Samai, A. Meraoumia, H. Bendjenna, L. Laimeche, “Oriented Local Binary Pattern (LBP θ): A new scheme for an efficient feature extraction technique”. In 2017 International Conference on Mathematics and Information Technology (ICMIT) (pp. 155-161). IEEE.

[12] A. Meraoumia, H. Bendjenna, M. Amroune, Y. Dris, "Towards a secure online E-voting protocol based on palmprint features". In 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1-6). IEEE.

[13] A. Meraoumia, S. Chitroub, A. Bouridane, "Fusion of multispectral palmprint images for automatic person identification". In 2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC) (pp. 1-6). IEEE.

[14] A. Bennour, M. Boudraa, I. Siddiqi, M. Al-Sarem, M. Al-Shaby, F. Ghabban, "A deep learning framework for historical manuscripts writer identification using data-driven features". *Multimedia Tools and Applications*, 83, 80103 (2024). https://doi.org/10.1007/s11042-024-18856-y.

[15] G. I. Raho, M. S. Al-Ani, A. A. K. Al-Alosi, L. A. Mohammed, "Signature recognition using discrete fourier transform". International Journal of Business and ICT, 1(1-2), 17-26, 2015.

[16] S. Agduk, ,E. Aydemir, "Classification of Handwritten Text Signatures by Person and Gender: A Comparative Study of Transfer Learning Methods". Acta Informatica Pragensia, 11(3), 324-347. 2022. doi: 10.18267/j.aip.197

[17] F. Hazem, B. Akram, T. Mekhaznia, F. Ghabban, A. Alsaeedi, B. Goyal, "Beyond Traditional Biometrics: Harnessing Chest X-Ray Features for Robust Person Identification". *Acta Informatica Pragensia*, *13*(2), 234-250, 2024. doi: 10.18267/j.aip.238

[18] F. Hazem, B. Akram, T. Mekhaznia, M. Al-Sarem, P. K. Shukla, O. I. Khalaf, "X-Ray Insights: A Siamese with CNN and Spatial Attention Network for Innovative Person Identification". In: Bennour, A., Bouridane, A., Almaadeed, S., Bouaziz, B., Edirisinghe, E. (eds) Intelligent Systems and Pattern Recognition. ISPR 2024. Communications in Computer and Information Science, vol 2305. Springer, Cham. https://doi.org/10.1007/978-3-031-82156-1_19

[19] F. Hazem., A. Bennour, N. A. Kurdi, S. Hammami, M. Al-Sarem, "Channel and Spatial Attention in Chest X-Ray Radiographs: Advancing Person Identification and Verification with Self-Residual Attention Network". *Diagnostics* 2024, *14*, 2655. https://doi.org/10.3390/diagnostics14232655

[20] F. Hazem, B. Akram, R. Sikder, S. Algburi, "X-ray insights: Innovative person identification through Siamese and Triplet networks". In: IET Conference Proceedings CP870. Stevenage, UK: The Institution of Engineering and Technology, 2023. p. 40-49. https://doi.org/10.1049/icp.2024.0463

[21] K. Packhäuser, S. Gündel, N. Münster, C. Syben, V. Christlein, A. Maier, "Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data". *Sci. Rep.* 2022, 12, 14851.

[22] Y. Ueda, J. Morishita, "Patient Identification Based on Deep Metric Learning for Preventing Human Errors in Follow-up X-Ray Examinations". *J. Digit. Imaging*, 36, 1941–1953, 2023.

[23] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, "Chestx-ray: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2097-2106), 2023.

[24] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning", 2017. arXiv preprint arXiv:1711.05225.