PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Acta Informatica Pragensia
2024, Volume 13, Issue 2, pp. 234–250
Special Issue: Future Trends of Machine Intelligence in Science and Industry
https://doi.org/10.18267/j.aip.238

Article Open Access

Beyond Traditional Biometrics: Harnessing Chest X-Ray Features for Robust Person Identification

Farah Hazem [1], Bennour Akram [1], Tahar Mekhaznia [1], Fahad Ghabban [2], Abdullah Alsaeedi [2], Bhawna Goyal [3]

[1] Laboratory of Mathematics, Informatics and Systems, Larbi Tebessi University – Tebessa, Algeria
[2] College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia
[3] Department of Electronics and Communication Engineering, Chandigarh University, Punjab, India

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

Abstract

Person identification through chest X-ray radiographs stands as a vanguard in both healthcare and biometrical security domains. In contrast to traditional biometric modalities, such as facial recognition, fingerprints and iris scans, the research orientation towards chest X-ray recognition has been spurred by its remarkable recognition rates. Capturing the intricate anatomical nuances of an individual's rib cage, lungs and heart, chest X-ray images emerge as a focal point for identification, even in scenarios where the human body is entirely damaged. Concerning the field of deep learning, a paradigm is exemplified in contemporary generations, with promising outcomes in classification and image similarity challenges. However, the training of convolutional neural networks (CNNs) requires copious labelled data and is time-consuming. In this study, we delve into the rich repository of the NIH ChestX-ray14 dataset, comprising 112,120 frontal-view chest radiographs from 30,805 unique patients. Our methodology is nuanced, employing the potency of Siamese neural networks and the triplet loss in conjunction with refined CNN models for feature extraction. The Siamese networks facilitate robust image similarity comparison, while the triplet loss optimizes the embedding space, mitigating intra-class variations and amplifying inter-class distances. A meticulous examination of our experimental results reveals profound insights into our model performance. Noteworthy is the remarkable accuracy achieved by the VGG-19 model, standing at an impressive 97%. This achievement is underpinned by a well-balanced precision of 95.3% and an outstanding recall of 98.4%. Surpassing other CNN models utilized in our research and outshining existing state-of-the-art models, our approach establishes itself as a vanguard in the pursuit of person identification through chest X-ray images.

Keywords

Deep learning; Convolutional neural network; CNN; Siamese neural network; Triplet loss.

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

Special Issue Editors: Hakim Bendjenna, Larbi Tebessi University – Tebessa, Algeria
Lawrence Chung, University of Texas at Dallas, USA
Abdallah Meraoumia, Larbi Tebessi University – Tebessa, Algeria

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

Copyright: © 2024 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

In recent years, the identification of individuals both before and after death has emerged as a pivotal area of research within the medical field (Morishita and Ueda, 2021). Various biological prints, such as fingerprints, facial recognition, DNA and eye prints, have been employed for human identification, while newer modalities such as handwriting recognition or palmprints have also shown promise (Bennour et al., 2019; Bennour, 2018; Samai et al., 2017; Meraoumia et al., 2018; Meraoumia et al., 2011; Bennour et al., 2024). However, the deterioration of certain soft tissues, as observed in natural disasters or aircraft crashes, renders many bodily biometrics impractical for post-mortem identification, and traditional means such as visual recognition or behavioural biometrics become unviable (Raho et al., 2015; Thomas & Preetha Mathew, 2023; Sato et al., 2022).

In scenarios where the human body undergoes severe damage, disfigurement or hindering of conventional identification methods, the utilization of chest X-ray images emerges as an alternative strategy. Use of artificial intelligence, particularly deep learning methods, has proven instrumental in person identification using chest X-ray radiography data (Packhäuser et al., 2023; Toge et al., 2013; Morishita et al., 2021). While conventional approaches, including convolutional neural networks (CNNs) and basic machine learning techniques, have shown promise, they often fall short in capturing nuanced details essential for precise identification.

This research introduces an innovative method that significantly enhances person identification using chest X-ray radiographs, employing the Siamese and triplet loss, powerful deep learning architectures. The primary objective is to develop an innovative approach that enhances post-mortem identification using chest X-ray radiographs. This involves employing deep learning methods, specifically Siamese and triplet networks, to extract highly discriminative features from X-ray images. This approach contributes to the state of the art by optimizing Siamese networks for image similarity comparison and triplet loss for an enhanced embedding space, minimizing intra-class variations and maximizing inter-class distances (Le-Phan et al., 2022). The integration of these architectures, coupled with the utilization of transfer learning, involving several pre-trained models such as VGG-19, VGG-16, ResNet-50, ResNet-101 and DenseNet as feature extractors in the encoder, results in a robust model capable of discerning individuals with improved discrimination.

The deliberate choice to employ transfer learning from different pre-trained models is particularly pertinent when dealing with unbalanced datasets where negative pairs significantly outnumber positive ones (Wang et al., 2023; Agduk et al., 2022). This strategic decision not only addresses issues with contrastive loss research but also showcases a deep understanding of the unique challenges posed by the dataset. By incorporating three inputs (anchor, positive and negative samples), our model effectively harnesses both positive and negative connections, leading to more efficient and robust learning. The comparative analysis of feature extraction using various pre-trained models allows identifying the one that exhibits the best performance in person identification.

This paper makes the following contributions:

- It introduces a novel approach using the Siamese network, based on its ability to learn highly discriminative features by comparing and contrasting pairs of images.
- It incorporates triplet loss to further enhance the discriminative power of the proposed approach, forcing the network to learn a feature space where the dissimilarity between different individuals is maximized.
- It successfully explores the use of previously pre-trained CNNs for different tasks on large benchmarks for person identification using chest X-ray radiographies.
- It makes an in-depth performance evaluation of five prominent pre-trained convolutional neural network (CNN) architectures, namely VGG-16, VGG-19, DenseNet, ResNet-50 and ResNet-101,

employed independently as backbones of Siamese neural network architectures for extracting features.

This paper aims to provide critical insights that advance our comprehension of the subject matter. Starting with an exploration of foundational concepts, our inquiry delves into existing research to provide a comprehensive overview in Section 2. Section 3 elucidates the meticulously devised structure underpinning our study, serving as a solid foundation for the experiments detailed in the following section. In Section 4, a thorough analysis of findings unveils valuable information contributing to the broader discourse in the field. Lastly, Section 5 consolidates the amassed data, underscores their significance and proposes potential avenues for future research, including recommendations.

2 Literature Review

By reviewing latest relevant literature, we can identify the most recent approaches, knowledge gaps and areas that require further research. In this context, we provide an overview of the research and literature related to person identification using chest X-ray images and other studies that discuss identification using Siamese networks and triplet techniques. Our interest leans towards artificial intelligence approaches, such as machine learning and deep learning, to apply particular architectures. We also examine the effects of pre-training and transfer learning. We discuss the most recent research in the area of person identification, which is assessed based on a number of key factors.

Multiple studies explore the use of chest X-ray radiographs for individual identification. Wang et al. (2023) proposed a stochastic modelling approach, grounded in image processing technologies for identifying human bodies post mortem through radiographs and CT scans. The methodology involves building a database by comparing images from a CT scan of an unidentified body with chest X-ray images. The primary processing steps encompass feature extraction, a matching process and ranking. Feature extraction entails deriving feature quantities from chest X-ray images, while the matching process involves comparing produced data with the database content. Subsequently, the ranking process arranges similar chest radiographs. The proposed approach to human identification utilizing chest X-ray images incorporates a decision-making mechanism, two-level feature extraction procedures for both database and query X-ray images and the computation of similarity measures. In the initial stage of feature extraction for each X-ray image in the database, the bag of words (BoW) and histogram of orientations of gradients (HOG) features are applied to chest radiographs. The second stage utilizes the BoW classifier to locate similar features between the images used and the database content. Ultimately, the ranking is determined based on the Euclidean distance between the final scores, yielding an accuracy ranging from 44.4% to 63.0%.

Cho et al. (2018) created a method for identifying a deceased individual that involves four primary steps, namely pre-processing, feature extraction, boundary extraction, and similarity calculation and ranking. The first step addresses the issue of low contrast in radiographic images. The second is implemented to achieve more sharply defined rib contours within radiographs and CT scan images using various methods. The authors utilized the contrast-limited adaptive histogram equalization (CLAHE) as a pre-processing tool, enhancing local contrast in images and achieving a more distinct ribcage contour. Within this step, the two-dimensional discrete Fourier transform (DFT) technique is employed. The process involves utilizing these attributes to assess the similarity between two radiograph images. The Euclidean distance method is then employed to compute similarity measures between the feature vectors of the query image and those in the database. Ultimately, the accuracy of this method is reported to be 74.07%.

Packhäuser et al. (2022) presented a technique focused on person re-identification using deep learning, specifically using the biometric properties of medical chest X-ray data. They employed Siamese neural networks (SNN) for both identification and re-identification tasks. In the identification scenario, the SNN is trained to compare two chest radiographs and predict whether they belong to the same patient. For re-identification, the SNN is utilized to compare a given patient's chest radiograph with a chest radiograph dataset, retrieving the most similar images. The study demonstrates a high performance in both scenarios, underscoring the potential of deep learning techniques for patient matching in medical image datasets. Using the ChestX-ray14 dataset, containing 100,000 chest X-ray images from over 30,000 unique patients, the system achieved remarkable results, with an AUC of 0.9940 and a classification accuracy of 95.55%. Notably, the system proved its ability of identifying the same person even ten or more years after the initial scan.

Ueda & Morishita (2023) proposed a biometric patient verification method utilizing chest X-ray images and a deep convolutional neural network (DCNN) with an EfficientNetV2-S backbone for feature extraction. The method involves three steps: image acquisition and pre-processing, feature extraction, and identification. The feature extraction step utilizes the DCNN to extract features from chest X-ray images, subsequently calculating a similarity index between clinical chest X-ray images and stored clinical datasets in order to identify a given patient. The cosine similarity formula is employed to compute the similarity between features in follow-up and baseline chest X-ray images. This formula measures the cosine of the angle between two non-zero vectors in an inner product space. The proposed method, evaluated on a dataset of 1,000 chest X-ray images, achieved an accuracy of 83.0%. The authors suggested that implementing this method could enhance patient safety and healthcare workflow by minimizing the risk of human error in patient identification.

In another context, several research studies have explored the use of Siamese networks, triplets or combined models for identification and verification. Thapar et al. (2019) proposed a deep learning-based framework for palm vein detection termed PVSNet using a Siamese network. The Siamese network is trained with triplet loss and a convolutional encoder-decoder network (CED), which make up the framework. In order to overcome the difficulties involved in creating a deep learning network for biometric verification with a small number of training samples, the authors also suggested a novel strategy. The method uses adaptive margin-based hard negative mining in conjunction with generative domain-specific features. The efficacy of the suggested method was demonstrated with experimental findings on three distinct palm vein databases. The paper showed that the proposed PVSNet framework achieves a state-of-the-art performance on three different palm vein databases. The limitations that occur in similar literature were critically examined in the paper, with particular attention paid to the clear deficiencies regarding their ability to withstand the effects of outdoor illumination and image quality degradation. Interestingly, the studies lack robustness, as evidenced by their irregular ROI segmentation, which makes them more prone to errors in feature extraction. Moreover, difficulties occur in precisely aligning obscure image portions, which may be false positives or false negatives, and in the availability of large-scale, precisely labelled datasets needed for deep learning models to be trained efficiently. Consequently, the framework achieved an EER of 3.71% on the most challenging CASIA dataset, which was significantly better than the similar works. The ROC curves and AUC values also demonstrated an effectiveness of the proposed framework in distinguishing between genuine and impostor palm vein images. The paper also presented a comparative study of the performance of PVSNet and FaceNet, a popular deep learning-based face recognition framework, using a zero-shot learning protocol. The results showed that PVSNet outperforms FaceNet in terms of EER and CRR (correct recognition rate).

Kumar et al. (2023a) introduced a two-pronged approach to facial detection, demonstrating the implementation of two distinct models. The first makes use of a convolutional neural network (CNN) to extract key points, subsequently employs a k-nearest neighbours (KNN) algorithm for classification. The second approach employs a Siamese network, a technique dedicated for image similarity. Emphasizing the critical phases of data collection and training, the paper offered a detailed exposition of both methodologies. Through meticulous performance evaluation and clear results, the study provided valuable insights into the comparative efficacy of these approaches, underscoring their respective strengths and potential applications in the ever-evolving landscape of facial detection technology.

Lai & Lam (2021) presented a cross-resolution triplet loss function and a deep Siamese network-based low-resolution face recognition (LRFR) technique. Matching low-resolution query faces with high-resolution gallery faces was the primary issue. By extracting discriminative features based on various resolutions and reducing the domain gap between high-resolution and low-resolution deep features, the suggested approach tackled this problem. In order to extract discriminative features across various resolutions, a deep Siamese network was trained as part of the proposed low-resolution facial recognition technique. The K-way face recognition network G is responsible for obtaining the deep features from both high-resolution and low-resolution face images. The shared classifier receives the HR and LR deep features, after which the classification loss is calculated using the additive margin softmax loss (AM-softmax). Additionally, the network is trained using a new loss function, the cross-resolution triplet loss, to further narrow the domain gap between the high-resolution and low-resolution deep features across various resolutions. For using this technique, the model showed a high performance: on the LFW dataset, it achieved a verification rate of 94.8% and an AUC of 0.982 at a false acceptance rate of 0.1%.

Kumar et al. (2023b) proposed a method involving training a convolutional neural network (CNN) model with a comprehensive dataset featuring diverse individuals, encompassing various facial expressions and lighting conditions. The unique aspect of one-shot learning lies in the capacity of the model to glean meaningful insights from a single input image. In this context, the Siamese neural network was employed to ascertain facial similarity, enabling the model to recognize an individual based on a solitary image. This innovative approach holds promise for overcoming the challenges posed by small datasets. It achieved 91% in terms of accuracy.

Wu et al. (2021) created a multi-scale deep supervision with an attention feature deep learning model for person re-identification (PReID). ResNet-50, the main model component, was utilized to extract several hierarchical deep features from the input image. Four classification losses and a ranked triplet loss were used to train the model. The middle-level feature learning incorporates a reverse attention module to reduce the problem of feature information loss due to conventional attention processes. In addition, the softmax scores are generated by the attention block. The network was further trained using a multi-scale feature learning layer under deep supervision. The suggested modules were used during the training phase. The proposed model showed promising results. As an example, on the Market-1501 dataset, the proposed model achieved a mAP of 89.0% and accuracy of 95.5%. The effectiveness of the multi-scale deep supervision block was also verified, as introducing it improved the performance of the model by increasing the mAP and rank-1 accuracy respectively.

This literature review covers different methods of person identification using biometric data and chest X-ray images, with a focus on triplet and Siamese networks. Approaches to feature extraction, similarity calculation and deep learning architectures such as convolutional neural networks (CNNs) and Siamese neural networks (SNNs) have all been examined in a number of studies. In tasks involving patient identification, these methods show promise, obtaining high accuracy rates between 44.4% to 95.55%. Significant research gaps and obstacles still exist, nevertheless, such as the need for robustness against outdoor illumination and image quality degradation and the absence of large-scale, accurately labelled datasets necessary for effective deep learning model training. Furthermore, there is ongoing exploration of Siamese networks and triplets in other biometric identification contexts, such as palm vein detection and facial recognition, highlighting the versatility of these techniques. Addressing these gaps and challenges constitutes the primary aim of the present study, which seeks to contribute novel methodologies and insights to enhance the accuracy and robustness of person identification systems, particularly in medical contexts, utilizing chest X-ray images.

3 Material and Proposed Method

The proposed approach to person identification using chest X-ray radiographs is the main topic of this section. By utilizing the distinct anatomical characteristics found in chest X-ray radiographs, we propose to create a reliable and precise technique for person identification. This section provides the methodology for person identification including study materials, data gathering procedures and image analysis processes.

3.1 Dataset

A key component of our methodology is the dataset. We used the NIH ChestX-ray14 dataset, which is the same as the one used by Ueda & Morishita (2023) with the best accuracy performance, to assess the efficacy of our suggested strategy. The NIH ChestX-ray14 dataset collection, which includes 112,120 frontal-view chest radiographs from 30,805 different patients, is one of the largest publicly accessible databases of chest X-ray radiographs in the scientific world. Because of follow-up scans, the image collection provides three to four photographs of each patient on average. The initial radiographs that were acquired were made available as 8-bit, 1024×1024 pixel PNG images. Metadata related to every photograph in the dataset are accessible. The patient's age and gender, the quantity of follow-up images acquired, the projection perspective (posterior-posterior or posterior-anterior) used to acquire the radiograph and the underlying illness patterns (no findings or a combination of up to 14 common thoracic diseases) are all included in the additional data.

According to the publisher, the dataset was put through a robust screening procedure to remove any personally identifiable information before it was made public. Consequently, patient names were replaced with numerical IDs. Furthermore, personal data in the image domain itself were made incomprehensible by overlaying black boxes over the pertinent image regions (Rajpurkar et al., 2017; Wang et al., 2017).

3.2 Proposed method

We propose a strategy that deliberately combines deep learning technology, specifically designed for identification of complex medical imaging and addressing the critical task of reliably identifying individuals in the aftermath of disasters or after death. Understanding the crucial nexus between security and healthcare, our methodology incorporates a meticulously developed solution for person identification using chest X-ray radiographs. Firstly, we employ a Siamese neural network for its effectiveness in comparing input pairs with identical weights, aiming to create discriminative representations of individuals' chests by encoding anchor, positive and negative images into feature vectors.

Thus, we suggest using a common encoder in our Siamese network architecture to handle the anchor, positive and negative inputs using similar weights as showed in Figure 2. This shared encoder uses transfer learning, involving several pre-trained models such as VGG-19, VGG-16, ResNet-50, ResNet-101 and DenseNet as feature extractors, resulting in a robust model capable of discerning individuals with improved discrimination. Our method uses a strong encoder built on a set of layers able to deal with feature extraction and normalization. This Siamese network is designed to process sets of anchor, positive and negative samples, computing the distances between anchor-positive and anchor-negative pairs through a custom distance layer as showed in Figure 1. We employ Euclidean distance and triplet loss metrics to quantify similarity and dissimilarity between images, further reinforcing learning objectives and ensuring robustness across diverse scenarios. The distance formula commonly used in Siamese networks for comparing two feature vectors $f(A)$ and $f(B)$ is the Euclidean distance. The Euclidean distance d between two vectors $x=(x1,x2,\dots,xn)$ and $y=(y1,y2,\dots,yn)$ in an n-dimensional space is given by Equation (1).

$$d(x,y)=\sum i=1n(xi-yi)2 \qquad (1)$$

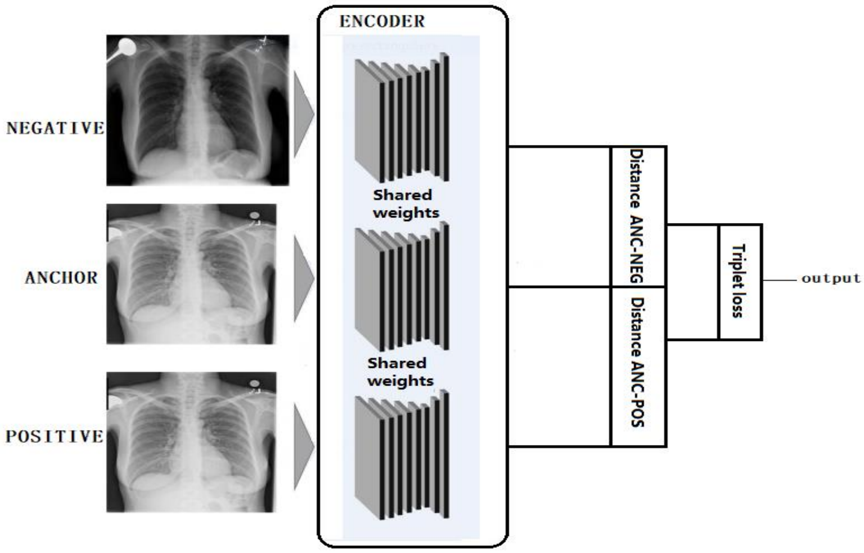

Figure 1. *Proposed method.*

3.2.1 Siamese neural network

Given two inputs with similar weights, the Siamese neural network compares their output vectors. Consequently, a twin neural network is another term for it. Initially, it was used for signature verification and facial recognition (De Rosa & Papa, 2022). The neural network receives two input images that are employed in the network architecture. In our experiment, we build a Siamese network that encodes three input images (anchor, positive and negative) to their feature vectors using an encoder that contains a pre-trained model with some added layers. The model produces three vectors that depict chests. The distance between these vectors is produced by using the distance function to calculate the similarity between the three feature maps. Using the triplet loss function, the network learns to place greater distance between dissimilar images and less distance between identical images.

3.2.2 Transfer learning

Creating a chest X-ray identification system from scratch requires a significant investment of time and money. Utilizing transfer learning from trained models is an additional method. A wide range of pre-trained models and ideas are presented by the current machine learning community (Guo et al., 2020; Cho et al., 2023). We built our own model based on pre-trained models. This can be accomplished by training only recently added layers, which need the least amount of time and resources to train, and freezing existing model layers. There are several such models, such as VGG-19, VGG-16, ResNet-50, ResNet-101 and DenseNet. In our work, we use encoder for converting the input images into their feature vectors. We use pre-trained models to extract features; our goal is to evaluate each pre-trained model separately within the same encoder architecture for the same task.

The use of transfer learning can significantly reduce the training time and size of the dataset. The model is connected to fully connected (dense) layers and the last layer normalizes the data using L2 normalization (a method that modifies the dataset values in a way that in each row the sum of the squares will always be up to 1). By utilizing those pre-trained models, we harness their ability to capture intricate hierarchical features from images and achieve the best performance. To ensure the preservation of these learned features, all layers of the pre-trained models are set as non-trainable. Subsequently, we construct a tailored encoder model by extending the pre-trained architecture. This includes the addition of dense layers for further abstraction, batch normalization for regularization and a final L2 normalization layer for feature vector normalization. Those features are passed to a distance layer, which computes the distance between anchor-positive and anchor-negative pairs. By analysing the performance of each pre-trained model to gain insights into their strengths and weaknesses for this task. This understanding can guide future model selection or improvements.

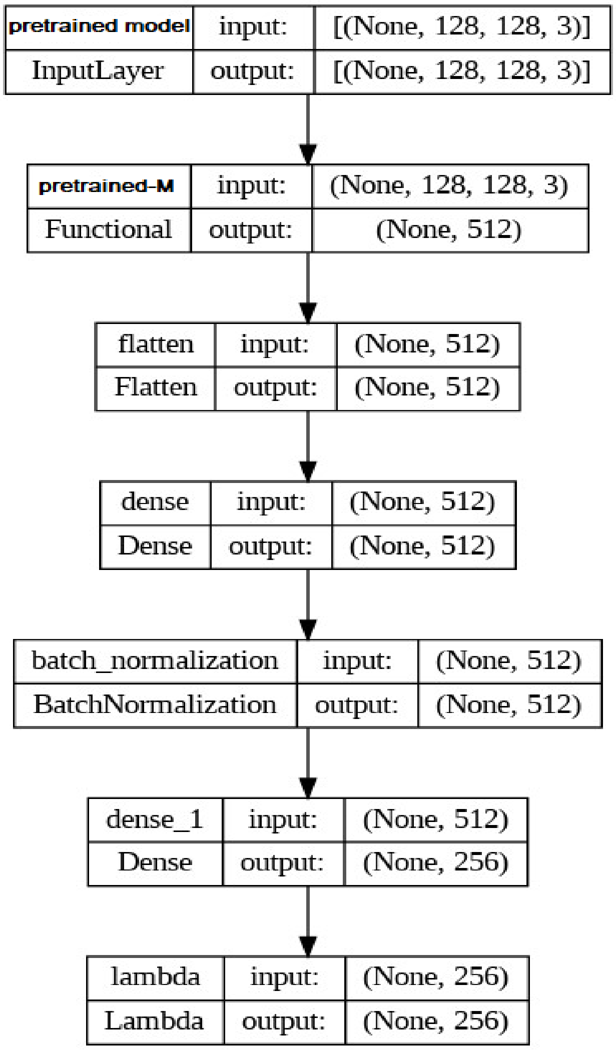

Figure 2. *Encoder.*

3.2.3 Euclidean distance and triplet loss

In the context of Siamese networks with triplet loss in our case, the distance formula between the anchor *f(A)* and positive *f(P)* embeddings, as well as between the anchor *f(A)* and negative *f(N)* embeddings, can be expressed in this way. Let: $f(A) = (a1,a2,...,an)$ be the anchor embedding, $f(P) = (p1,p2,...,pn)$ be the positive embedding and $f(N) = (n1,n2,...,nn)$ be the negative embedding. The Euclidean distance between the anchor *f(A)* and positive *f(P)* embeddings is given by Equation (2). Similarly, the Euclidean distance between the anchor *f(A)* and negative *f(N)* embeddings is given by Equation (3).

$$distance\_ap = \sum i = 1n(ai - pi)2 \tag{2}$$

$$distance\_an = \sum i = 1n(ai - ni)2 \tag{3}$$

The goal is to learn embeddings in such a way that the distance between the anchor and positive embeddings is minimized, while the distance between the anchor and negative embeddings is maximized using triplet loss. This loss function is mathematically expressed as follows:

The triplet loss for a triplet *(A,P,N)*, where *A* is the anchor image, *P* is the positive image and *N* is the negative image, is defined by Equation (4):

$$loss(A, P, N) = max(ap\_distance - an\_distance + margin, 0.0 \quad (4)$$

Where:

- ‖·‖‖·‖ represents the Euclidean norm.
- *f(A)*, *f(P)* and *f(N)* are the embeddings of the anchor, positive and negative images, respectively.
- *margin* is a hyperparameter that specifies the minimum margin or separation between the distances of the anchor and positive embeddings *f(A)−f(P)∥2)* and the anchor and negative embeddings *f(A)−f(N)∥2)*.

The goal during training is to minimize this triplet loss. The loss is zero when the distances between the anchor and positive embeddings are smaller than the distances between the anchor and negative embeddings by at least the specified margin. Otherwise, the loss becomes the positive difference between the squared distances (Guo et al., 2020).

This formulation encourages the model to learn embeddings in which positive pairs are closer together than negative pairs by at least the margin. Adjusting the margin allows us to control the trade-off between embedding separation and model robustness.

3.3 Data preparation

We preprocess images to be of the size (128,128,3) and read them encoded in RGB. We choose a random image to serve as the anchor for one patient, with the requirement that each patient have at least two examples. Then, we generate positive examples for each patient that have the same patient ID with the anchor image, and we choose random negative examples for each patient that do not. We read the RGB-encoded images and preprocess them to the size of 128x128x3. In order to evaluate and train Siamese network architectures using triplets, we pick an anchor image and create positive pairings (the same patient ID) and negative pairings (different patient IDs) from the images in the ChestX-ray14 dataset. We employ our split data function, which splits the dataset into 90% training data and 10% testing data, to guarantee that images from a single patient only appear in the training or testing set. We build the real image pairs for every subset, excluding patients who only have one example of a radiograph by using the patient ID label from the dataset to create positive and negative pairs.

4 Experimental Results and Discussion

The findings may be compared with existing literature or theoretical expectations to highlight similarities, differences or novel discoveries. In this section, we examine the empirical data that came from the experimental protocols described earlier in the study. Presenting the results objectively is the main purpose, and this is frequently accomplished by using tables, figures and statistical analysis. Our proposed technique is implemented using the Keras library in Python. A total of over 120,000 images are present in the dataset used, of which 100,000 images were used for training and 20,000 for evaluation. The confusion matrix and the performance metrics of the proposed method are presented in Table 1. The graph for evaluation loss against iterations is presented in Figure 3. The graph for test accuracy against iterations is presented in Figure 4.

4.1 Performance evaluation

To evaluate the performance of our proposed person identification system, we employ various assessment metrics tailored to assess the accuracy of identification tasks. The confusion matrix is one of the important measures used to assess classification metrics. For this purpose, we use it to provide a comprehensive imagine of the performance of the classification model in distinguishing between dissimilar and similar images. The matrix reveals that the model accurately predicted dissimilar images (true negatives) and similar images (true positives). However, there were instances that dissimilar images were erroneously classified as similar (false positives) and similar images were mistakenly categorized as dissimilar (false negatives). Precision denotes the model accuracy in predicting similar images among its positive predictions. Recall indicates the ability of the model to identify all actual similar images. The F1 score, a balanced metric between precision and recall, encapsulates the model effectiveness in handling both dissimilar and similar image classifications (Powers, 2020).

$$\text{Precision [P]} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall [R]} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \textbf{(7)}$$

$$\text{F1 score} = 2\,\frac{P*R}{P+R} \quad (8)$$

4.2 Results and comparison of pre-trained models

To evaluate the identification performance of our trained model, several metrics are computed, namely training loss, test accuracy and other metrics. The identification model is trained using the triplet loss function to enhance its ability to learn effective representations for tasks such as identification, image similarity and other tasks where understanding the similarity or dissimilarity between inputs is important. We evaluate our model with training loss to ensure that our model is learning. To prove that our loss is decreasing over time, Figure 3 shows the training loss curve. In the proposed model, we obtain a training accuracy of 99.9% and loss of 0.004%; in testing, we get an accuracy of 97% using the pre-trained model VGG-19. That is proof that our proposed method has a very powerful performance in learning and testing in terms of accuracy and loss and in other metrics, as shown in Table 1. We trained our model for over 50 epochs.

Table 1. *Confusion matrix and performance metrics.*

Pre-trained model	TN	TP	FP	FN	Recall	Precision	F1 score	Accuracy
VGG-16	49.63	46.97	0.7	3.03	93.94	98.5	96.1	96.4
VGG-19	**49.67**	**47.24**	**2.33**	**0.76**	**98.4**	**95.3**	**96.8**	**97**
ResNet-50	43.89	46.91	3.09	6.11	88.4	93.8	91.01	90.8
ResNet-101	48.64	46.98	3.02	1.36	97.1	93.9	95.5	94.9
DenseNet	45.92	42.78	7.22	4.08	91.2	85.5	88.3	88.7

In this comparative analysis of pre-trained models for feature extraction in our encoder, our experimental results provide valuable insights into their respective performances. VGG-16 exhibited an impressive accuracy of 96.4%, showcasing a high precision of 98.5% and a robust recall of 93.94%. VGG-19 demonstrated an even higher accuracy of 97%, driven by a well-balanced precision of 95.3% and a remarkable recall of 98.4%. ResNet-50, while achieving an accuracy of 90.8%, displayed a slightly lower precision of 93.8% but maintained a commendable recall of 88.4%. ResNet-101 emerged as a strong contender with a balanced accuracy of 95.6%, combining a high recall of 97.1% and a precision of 93.9%. DenseNet, although showing a lower accuracy of 88.7%, delivered a balanced precision of 85.5% and a recall of 91.2%. The best evaluation model is VGG-19, with the highest overall accuracy, precision and recall. Its impressive performance makes it the optimal choice for feature extraction in our experimental setup, emphasizing its efficacy in accurately identifying positive instances while minimizing false positives and false negatives.

The comprehensive evaluation presented here guides the selection of the most suitable pre-trained model for the given task, setting the stage for improved outcomes in related applications. In conclusion, while all models demonstrate strong performance overall, there are nuances in their abilities to balance precision, recall and accuracy. VGG-16 and ResNet-101 stand out for their high precision and recall rates, making them suitable for tasks where minimizing both false positives and false negatives is crucial. ResNet-50, although exhibiting slightly lower recall, still maintains high accuracy, indicating its reliability in making correct classifications. VGG-19 and DenseNet, while performing well, may require additional attention to minimize false positives and negatives, respectively. These insights can inform the selection and optimization of pre-trained models based on specific research objectives and application requirements.

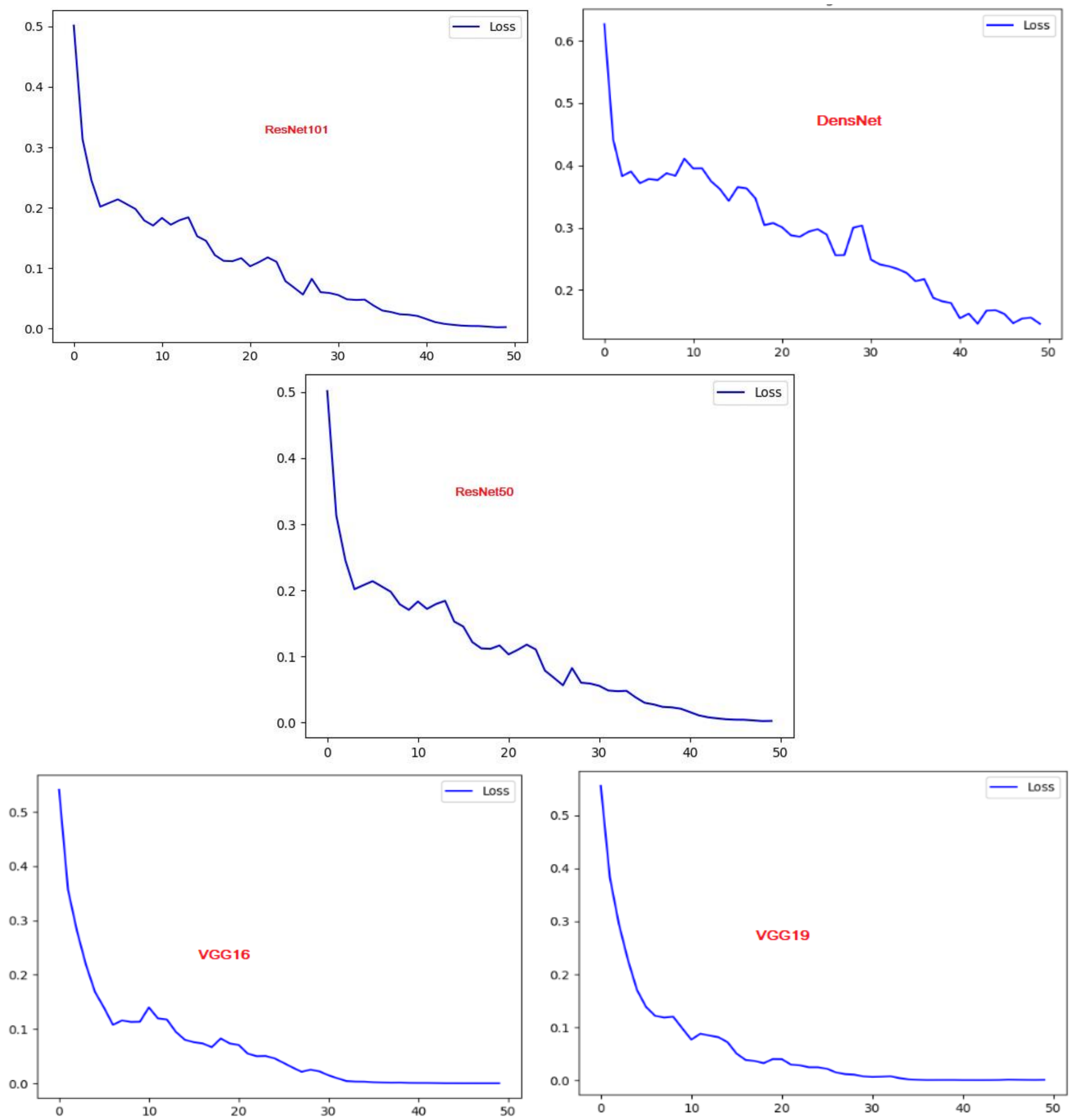

Figure 3. *Training losses of different pre-trained models over epochs.*

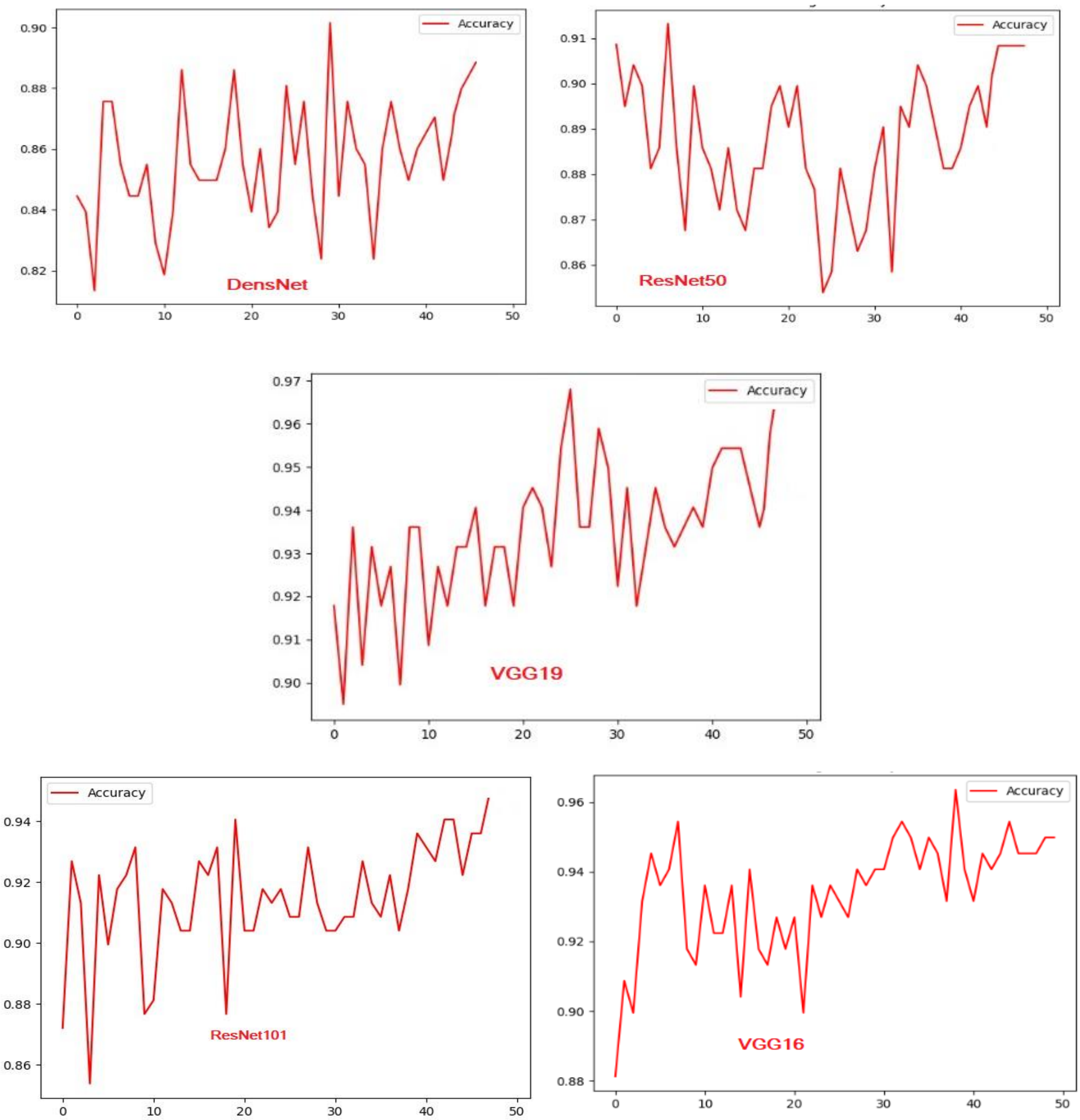

Figure 4. *Testing accuracy of different pre-trained models over epochs.*

4.3 Comparison with similar works

Person identification using chest X-ray images is a difficult task. Convolutional neural networks (CNNs), and in particular, deep learning models are particularly good at automatically extracting characteristics from data. Deep learning models are able to automatically extract important features from chest X-ray radiographs without the need for handmade feature engineering or annotation, which is necessary in classical machine learning approaches. Thus, such machine learning approaches could have more difficulty extracting unique features in chest X-ray images, as proposed by Ishigami et al. (2017) and Cho et al. (2018). In their studies, they complained about blurring problems in the images, resulting in obtaining a small quantity of features of poor quality. They obtained low model performance in terms of accuracy. The first experiment achieved 63% and the second one achieved 74%.

Packhäuser et al. (2022) used a Siamese network with two sub-networks that computed the difference between two images. In training, they used a contrastive loss function. In our case, where the dataset is imbalanced, with a significantly higher number of negative pairs compared to positive pairs, contrastive loss may struggle to effectively learn discriminative features. In such cases, triplet loss, which inherently addresses imbalanced scenarios by considering an anchor, a positive and a negative example, will perform better. The study obtained the best performance in terms of accuracy, but we have outperformed it in terms of accuracy. The researchers used only two inputs(positive pair), which could be a weak point. By considering the negative sample in addition to the anchor and positive samples, the network is encouraged to learn a more discriminative feature space. This helps in creating larger inter-class separations and smaller intra-class variations. The use of three inputs in a Siamese network with triplet loss facilitates more effective and robust learning by incorporating information about both positive and negative relationships. Additionally, we find that the use of pre-trained models as features extractors is more useful than the Siamese feature extractor. That is why we obtained a very powerful model that uses VGG-19 as the feature extractor that has a testing accuracy of 97% on the same dataset.

For facial detection, Kumar et al. (2023) employed a combination of CNN for key point extraction, KNN and a Siamese network with triplet loss. However, accuracy metrics for their method are not available. In contrast, our model excels with a 97% accuracy in person identification, outperforming Lai & Lam (2021), Kumar et al. (2023), Wu et al. (2021) and all of the state-of-the-art models. This clarification endorses the superiority of our model, particularly in the context of person identification using chest X-ray images, where it consistently outperforms the state-of-the-art methods in accuracy. In Table 2, we provide a comparison of works that have used chest X-ray for identification. The metrics shown in Table 1 clearly show how much better our suggested method is in identifying people from chest X-ray images. With an impressive 97% accuracy rate, our model performs better than earlier research, which struggled with problems including feature quality degradation and blurring. Using a Siamese network with triplet loss strategically is essential, particularly when dealing with an unbalanced dataset where negative pairs greatly outnumber positive pairs. This design decision not only addresses drawbacks found in research using contrastive loss, but it also shows a sophisticated comprehension of the particular difficulties presented by our dataset. We are able to utilize both positive and negative associations by combining three inputs (anchor, positive and negative samples), which promotes more robust and efficient learning. Furthermore, using transfer learning for feature extraction, specially VGG-19, proves to be a better option than other feature extraction techniques, greatly enhancing the model capacity to extract and utilize pertinent features.

***Table 2.** Comparison with state-of-the-art works.*

Work	Application	Method	Dataset	Accuracy
Ishigami et al. (2017)	Person identification using chest X-ray	HOG, BOW classifier, Euclidean distance	Collected from data stored in Miyazaki University School of Medicine database	44.4 to 63
Cho et al. (2018)	Person identification (using chest X-ray	Clahi, two-dimensional discrete Fourier transform (DFT)	Collected from data stored in Miyazaki University School of Medicine database	74.07
Packhäuser et al. (2022)	Person re-identification using chest X-ray	Siamese NN, contrastive loss,ResNet-50	ChestX-ray8 dataset	95.55
Ueda and Morishita (2023)	Person identification using chest X-ray	EfficientNet, cosine distance	ChestX-ray14 dataset	83
Thapar et al. (2019)	Palm vein detection	Siamese network with triplet loss CED, adaptive margin-based hard negative mining, generative domain-specific features	------	------

Work	Application	Method	Dataset	Accuracy
Kumar et al. (2023)	Facial detection	CNN for key point extraction, KNN for classification, Siamese network and triplet loss	------	------
Lai & Lam (2021)	Face recognition	Triplet loss, deep Siamese network, K-way face recognition network	LFW dataset	94.8
Kumar et al. (2023)	Face recognition	Deep Siamese network, triplet loss	------	91
Wu et al. (2021)	Face recognition	Attention feature learning, ResNet-50, triplet loss	Market-1501 dataset	95.5
Presented solution	**Person identification using chest X-ray**	**Siamese NN, triplet loss, VGG-19**	**ChestXray14 dataset**	**97**

The utilization of a Siamese neural network with triplet loss in our approach has proven its effectiveness in achieving a remarkably high accuracy. The robustness of the Siamese architecture lies in its ability to learn highly discriminative features by comparing and contrasting pairs of images. The incorporation of triplet loss further enhances this discriminative power, compelling the network to learn a feature space where the dissimilarity between different individuals is maximized. The choice of pre-trained models as the backbone of our Siamese network contributes significantly to the superior performance observed. VGG-19 is renowned for its deep and homogeneous structure, allowing it to capture intricate patterns and features within the input data. The hierarchical feature representation of the network, facilitated by its stacked convolutional layers, proves crucial in discerning subtle differences in chest X-ray images for effective person identification. The comparative analysis with existing methods highlights the clear superiority of our proposed approach. Notably, our accuracy of 97% on the ChestX-ray14 dataset outperforms other state-of-the-art methods, including HOG, BoW, Clahi, Siamese networks with contrastive loss and ResNet 50, as well as the pertinent models VGG-16, VGG-19, ResNet-50, ResNet-101 and DenseNet. The high accuracy achieved by our model holds significant implications for various real-world applications. In the realm of medical diagnostics, accurate person identification through chest X-ray images can streamline patient record management and enhance the efficiency of healthcare systems. Furthermore, in security systems and border control, our methodology could be integral in establishing robust identity verification protocols, particularly in scenarios where traditional identification methods may be impractical or insufficient.

5 Conclusion

In summary, our work introduced an innovative approach to person identification by utilizing distinctive anatomical features present in chest X-ray radiographs. The methodology, built upon Siamese and triplet networks, employs various pre-trained models (VGG-16, VGG-19, ResNet-50, ResNet-101, DenseNet) individually as feature extractors. This departure from conventional practices offers a fresh perspective on integrating deep learning technologies at the intersection of healthcare and security. Emphasizing the unique characteristics of anatomical elements such as the heart, lungs and rib cage, our study underscores the potential for accurate identification, even in situations where harm to the human body is prevalent, as is often the case in emergency scenarios. The strategic adoption of Siamese and triplet networks showcases the versatility of these deep learning architectures in capturing intricate correlations within chest X-ray images.

Importantly, our research compared the performance of each pre-trained model, demonstrating nuanced variations in feature extraction capabilities. This comprehensive analysis reveals that VGG-19 outperforms the others, achieving superior training and testing accuracy. This signifies a great advancement in the field and highlights the efficacy of our approach in learning and representing complex relationships within medical images.

This research not only contributes to the evolution of person identification techniques but also holds promise for broader applications in healthcare and security domains. The intersection of anatomical specificity and deep learning sophistication, combined with a careful comparison of pre-trained models, presents a holistic and robust methodology. In conclusion, our results not only showcase the superior performance of our proposed approach in person identification using chest X-ray images, but also provide valuable insights into the comparative effectiveness of different pre-trained models. This work positions our methodology as a noteworthy advancement, paving the way for more accurate and reliable person identification from medical imaging data.

A limitation of our study is the presence of class imbalance within the dataset, potentially biasing model performance. Future research could address this by implementing advanced data augmentation techniques and specialized loss functions to mitigate imbalance effects, as well as investigating the scalability and computational efficiency of the proposed approach in large-scale deployment scenarios.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Acknowledgments: This work would not have been possible without the support of the LAMIS Laboratory, Larbi Tebessi University – Tebessa and the authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of education in Saudi Arabia.

Funding: This work was funded by the Deputyship for Research & Innovation, Ministry of education in in Saudi Arabia, Grant No: 445-9-902.

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: F.H.: Writing – Original draft preparation, Methodology, Software. B.A.: Supervision, Conceptualization. T.M.: Data curation, Writing – Original draft preparation. F.G.: Visualization, Investigation. A.A.: Data curation, Writing – Original draft preparation. B.G.: Software, Validation, Writing – Reviewing and Editing.

Data Availability: The authors used previously published dataset, which is available in National Institutes of Health repository at https://nihcc.app.box.com/v/ChestXray-NIHCC.

References

Agduk, S., & Aydemir, E. (2022). Classification of handwritten text signatures by person and gender: A Comparative Study of transfer learning methods. *Acta Informatica Pragensia*, 11(3), 324–347. https://doi.org/10.18267/j.aip.197

Bennour, A. (2018). Automatic handwriting analysis for writer identification and verification. In *Proceedings of the 7th International Conference on Software Engineering and New Technologies* (pp. 1–7). ACM. https://doi.org/10.1145/3330089.3330129

Bennour, A., Boudraa, M., Siddiqi, I., Al-Sarem, M., Al-Shaby, M., & Ghabban, F. (2024). A deep learning framework for historical manuscripts writer identification using data-driven features. *Multimedia Tools and Applications*, (in press). https://doi.org/10.1007/s11042-024-18187-y

Bennour, A., Djeddi, C., Gattal, A., Siddiqi, I., & Mekhaznia, T. (2019). Handwriting based writer recognition using implicit shape codebook. *Forensic science international*, 301, 91–100. https://doi.org/10.1016/j.forsciint.2019.05.014

Bhimireddy, A. R., Burns, J. L., Purkayastha, S., & Gichoya, J. W. (2022). Few-Shot Transfer Learning to improve Chest X-Ray pathology detection using limited triplets. *arXiv preprint arXiv:2204.07824*. https://doi.org/10.48550/arXiv.2204.07824

Cho, H., Zin, T. T., Shinkawa, N., & Nishii, R. (2018). Post-Mortem human identification using chest X-Ray and CT scan images. *Soft Computing*, 23(2), 51–57. https://doi.org/10.24466/ijbschs.23.2_51

Cho, K., Kim, K. D., Nam, Y., Jeong, J., Kim, J., Choi, C., Lee, S., Lee, J. S., Woo, S., Hong, G., Seo, J. B., & Kim, N. (2023). CHESS: Chest X-Ray Pre-trained model via self-supervised contrastive learning. *Journal of Digital Imaging*, 36(3), 902–910. https://doi.org/10.1007/s10278-023-00782-4

De Rosa, G. H., & Papa, J. P. (2022). Learning to weight similarity measures with Siamese networks: a case study on optimum-path forest. In *Optimum-Path Forest* (pp. 155-173). Academic Press. https://doi.org/10.1016/B978-0-12-822688-9.00015-3

Guo, J., Wang, H., Cheng, Z., Zhang, X., & Yan, D. M. (2020). Learning local shape descriptors for computing non-rigid dense correspondence. *Computational Visual Media*, 6, 95-112. https://doi.org/10.1007/s41095-020-0163-y

Ishigami, R., Zin, T. T., Shinkawa, N., & Nishii, R. (2017). Human identification using X-Ray image matching. In *Proceedings of the International Multi Conference of Engineers and Computer Scientists* (pp. 1–4). IAENG.

Kumar, C. R., Saranya, N., Priyadharshini, M., & Gilchrist, D. (2023a). Face recognition using CNN and siamese network. *Measurement: Sensors*, 27, 100800. https://doi.org/10.1016/j.measen.2023.100800

Kumar, K. V., Teja, K. A., Bhargav, R. T., Satpute, V., Naveen, C., & Kamble, V. (2023b). One-shot face recognition. In *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS),* (pp. 1-6). IEEE. https://doi.org/10.1109/PCEMS58491.2023.10136112

Le-Phan, A., Nguyen, X. P. P., & Ly-Tu, N. (2022). 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. https://doi.org/10.1109/NICS56915.2022.10013393

Lai, S. C., & Lam, K. M. (2021). Deep siamese network for low-resolution face recognition. In *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC),* (pp. 1444–1449). IEEE.

Meraoumia, A., Bendjenna, H., Amroune, M., & Dris, Y. (2018). 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. https://doi.org/10.1109/PAIS.2018.8598520

Meraoumia, A., Chitroub, S., & Bouridane, A. (2011). Fusion of multispectral palmprint images for automatic person identification. In *2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC),* (pp. 1–6). IEEE. https://doi.org/10.1109/SIECPC.2011.5876898

Morishita, J., & Ueda, Y. (2021). New solutions for automated image recognition and identification: challenges to radiologic technology and forensic pathology. *Radiological physics and technology*, 14(2), 123-133. https://doi.org/10.1007/s12194-021-00611-9

Packhäuser, K., Folle, L., Thamm, F., & Maier, A. (2023). 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. https://doi.org/10.1109/ISBI53787.2023.10230346

Packhäuser, K., Gündel, S., Münster, N., Syben, C., Christlein, V., & Maier, A. (2022). Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data. *Scientific Reports*, 12(1), 14851. https://doi.org/10.1038/s41598-022-19045-3

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. https://doi.org/10.48550/arXiv.2010.16061

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

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*. https://doi.org/10.48550/arXiv.1711.05225

Samai, D., Meraoumia, A., Bendjenna, H., & Laimeche, L. (2017). 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. https://doi.org/10.1109/MATHIT.2017.8259710

Sato, M., Kondo, Y., Okamoto, M., & Takahashi, N. (2022). Development of individual identification method using thoracic vertebral features as biometric fingerprints. *Scientific Reports*, 12(1). https://doi.org/10.1038/s41598-022-20748-w

Thapar, D., Jaswal, G., Nigam, A., & Kanhangad, V. (2019). PVSNet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features. In *2019 IEEE 5th international conference on identity, security, and behavior analysis (ISBA),* (pp. 1–8). IEEE. https://doi.org/10.1109/ISBA.2019.8778623

Thomas, P. A., & Preetha Mathew, K. (2023). A broad review on non-intrusive active user authentication in biometrics. *Journal of Ambient Intelligence and Humanized Computing*, 14(1), 339–360. https://doi.org/10.1007/s12652-021-03301-x

Toge, R., Morishita, J., Sasaki, Y., & Doi, K. (2013). 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. https://doi.org/10.1007/s12194-013-0221-6

Ueda, Y., & Morishita, J. (2023). Patient identification based on deep metric learning for preventing human errors in follow-up X-ray examinations. *Journal of Digital Imaging*, 36(5), 1941–1953. https://doi.org/10.1007/s10278-023-00850-9

Wang, H., Singhal, A., & Liu, P. (2023). Tackling imbalanced data in cybersecurity with transfer learning: A case with ROP payload detection. *Cybersecurity*, 6(1), 2. https://doi.org/10.1186/s42400-022-00135-8

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-Ray8: 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). IEEE. https://doi.org/10.1109/CVPR.2017.369

Wu, D., Wang, C., Wu, Y., Wang, Q., & Huang, D. (2021). Attention Deep model with Multi-Scale Deep Supervision for Person Re-Identification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(1), 70–78. https://doi.org/10.1109/tetci.2020.3034606

Editorial record: The article has been peer-reviewed. First submission received on 5 February 2024. Revisions received on 12 April 2024 and 24 May 2024. Accepted for publication on 2 June 2024. The editors coordinating the peer-review of this manuscript were Hakim Bendjenna, Lawrence Chung, Abdallah Meraoumia, and Zdenek Smutny. The editor in charge of approving this manuscript for publication was Zdenek Smutny.

Special Issue: Future Trends of Machine Intelligence in Science and Industry. Selected papers from the National Conference on Artificial Intelligence: From Theory to Practice (NCAI'2023).

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951