Journal of Digital Imaging (2023) 36:1941–1953
https://doi.org/10.1007/s10278-023-00850-9

Patient Identification Based on Deep Metric Learning for Preventing Human Errors in Follow-up X-Ray Examinations

Yasuyuki Ueda[1] · **Junji Morishita**[2]

Received: 18 January 2023 / Revised: 8 May 2023 / Accepted: 10 May 2023 / Published online: 12 June 2023

Abstract
Biological fingerprints extracted from clinical images can be used for patient identity verification to determine misfiled clinical images in picture archiving and communication systems. However, such methods have not been incorporated into clinical use, and their performance can degrade with variability in the clinical images. Deep learning can be used to improve the performance of these methods. A novel method is proposed to automatically identify individuals among examined patients using posteroanterior (PA) and anteroposterior (AP) chest X-ray images. The proposed method uses deep metric learning based on a deep convolutional neural network (DCNN) to overcome the extreme classification requirements for patient validation and identification. It was trained on the NIH chest X-ray dataset (ChestX-ray8) in three steps: preprocessing, DCNN feature extraction with an EfficientNetV2-S backbone, and classification with deep metric learning. The proposed method was evaluated using two public datasets and two clinical chest X-ray image datasets containing data from patients undergoing screening and hospital care. A 1280-dimensional feature extractor pretrained for 300 epochs performed the best with an area under the receiver operating characteristic curve of 0.9894, an equal error rate of 0.0269, and a top-1 accuracy of 0.839 on the PadChest dataset containing both PA and AP view positions. The findings of this study provide considerable insights into the development of automated patient identification to reduce the possibility of medical malpractice due to human errors.

Keywords Chest X-ray image · Patient identification · Patient verification · Biometrics · Deep metric learning

Introduction

Patient misidentification due to human error remains a major concern in the healthcare industry [1–8]. According to a previous study [3], the average rate of misfiled cases for radiography at a hospital is 0.075%.

A previous paper [5] reported that "a 2006 study [6] reviewing several adverse events databases in the United States estimated that between 1,300 and 2,700 adverse events relating to wrong side or site, wrong procedure, and wrong patient occurred annually. It identified radiology as the second most common hospital department (after the operating room) to perform wrong-site procedures." Although the primary risk of patient misidentification in clinical X-ray examinations is radiation exposure of the wrong patient, it can lead to serious incidents or accidents, such as operating of the wrong patient [6].

Risk management in clinics and hospitals employs a standard process for patient verification because confirming patient-specific information (such as the full name and date of birth) is critical [9]. Although using two patient identifiers can improve the reliability of patient identification, such a process is not the most effective for preventing human errors, resulting in misfiling and misidentification [10, 11]. To minimize the risk of human error and improve healthcare workflow and utility, clinics and hospitals may use biometrics, such as fingerprints and face recognition, for routine examinations. However, the use of these methods can increase the workload of healthcare providers.

Digital transformation in healthcare can be realized using various technologies to streamline operations for healthcare providers and ensure patient safety. Several types of clinical images, such as chest X-ray or scout images, exhibit

✉ Yasuyuki Ueda
ueda@sahs.med.osaka-u.ac.jp

[1] Department of Medical Physics and Engineering, Area of Medical Imaging Technology and Science, Graduate School of Medicine, Division of Health Sciences, Osaka University, Osaka, Japan

[2] Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan

considerable potential for use in biometric patient verification [1–3, 12–27] and forensic personal identification [28–35]. The potential risk of patient registration errors owing to multiple factors, including human errors, can be reduced by consolidating clinical information and using it for patient identity verification. Furthermore, such a technology has the potential to identify the personal information of a patient without the patient verbally stating their name or date of birth. Several biometric methods based on clinical images have been proposed to achieve satisfactory patient verification and identification [1–3, 12–24].

Deep learning-based technologies have been applied to several tasks using clinical X-ray images [36–39] of the chest [12, 28, 36, 40–46], including lesion detection and estimation of individual attributes, such as gender and age. Some issues need to be addressed for deep learning-based biometric systems to be utilized in clinics or hospitals. In hospitals and clinics, a biometric system can identify a large number of patients during the second examination without retraining the system for each new patient. Several methods have been reported for the classification of chest X-ray images used in clinics or hospitals [12]; however, such methods are yet to be used in actual clinical applications.

The performance of such biometric techniques applied to clinics and hospitals can degrade owing to variabilities in the clinical chest X-ray images, including variations in posteroanterior (PA) and anteroposterior (AP) view positions, image processing, and patient conditions (pre- and post-operative or trauma) [13, 14]. In particular, regarding view positions, healthy subjects are examined in the standing and PA view positions in routine chest X-ray imaging. However, depending on the patient's condition (such as inability to stand or postmortem), the examination is performed in the spine or in the sitting AP view position. Previous studies have revealed that classification between PA and AP view positions is possible in chest X-ray images [36]; this also implies that there is a domain-shift problem between PA and AP view positions. Therefore, differences in the positions of PA and AP views, including differences in the patient's condition and imaging geometry, can have a significant impact on their biometric performance. Although some studies have addressed this problem [28], the domain-shift problem caused by differences in view position has not been resolved, and concerns regarding its clinical application are yet to be addressed. Biometric systems for clinics and hospitals are expected to not only have adequate biometric performance based on clinical images [47–49] but also exhibit robustness under clinical image variabilities; therefore, such a biometric system, which exhibits good performance under varied view positions, will potentially be effective for use in clinics and hospitals.

The objective of the study was to (a) develop a deep learning-based feature extractor and classifier using the similarity index to perform patient identity verification during chest X-ray examination and (b) quantitatively evaluate the outcomes of the proposed method on clinical chest X-ray datasets with clinical variabilities, such as different view positions.

Materials and Methods

Outline of the Proposed Method

The proposed deep learning metric-based patient verification and identification system is illustrated in Fig. 1. The system comprises three units: (i) image acquisition and preprocessing, (ii) feature extraction, and (iii) identification. The first unit acquires an image with a routine chest X-ray examination and performs image postprocessing. Healthcare providers are not required to perform any additional work in biometric patient verification. In the second unit, the feature extractor trained by the proposed model extracts the features that identify an individual from chest X-ray images. In the final unit, the similarity index of the extracted features between clinical chest X-ray images and stored clinical datasets is used to determine whether the patients are the same.

The proposed method was evaluated using four test datasets different from the dataset used for training and validation. Two datasets were public datasets, whereas the other two were clinical chest X-ray image datasets.

Datasets

The experiment was a retrospective, observational study and was approved by the Institutional Review Boards at Osaka University (Approval Number: 19422–6). Therefore, informed consent was not mandatory. All procedures performed in this study conformed to the guidelines and ethics set by the Helsinki Declaration.

Table 1 summarizes the patient demographics in the five datasets used for training, validating, and testing the DCNN used in the proposed method. The ChestX-Ray8 dataset [50], used for training and validation, included 57,452 chest X-ray images from 3,245 patients with at least eight chest X-ray images per patient that were examined on both PA and AP view positions. All 57,452 chest X-ray images were used in either training or validating subsets, but not both. Of the 57,452 chest X-ray images of 3,245 patients, 50,962 images were used for training and 6,490 images for validating the trained model. To enable mini-batch learning within the computation limitations for training and validation, the image pixel spacing of all the images was resampled to 1.0×1.0 mm^2 (276 to 495 and 261 to 495 pixels for rows and columns, respectively) using bicubic interpolation and by cropping the central 320×320 pixels.

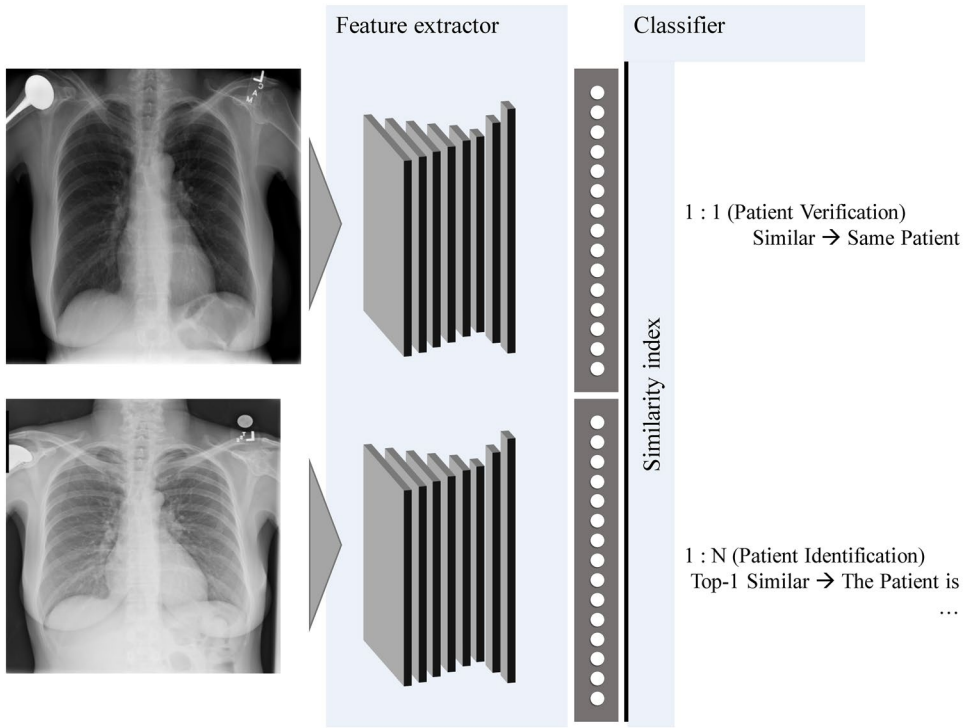

Fig. 1 Proposed system for patient verification and identification

Two public datasets and two clinical datasets were used for testing. The first public dataset, PadChest [51], included 19,713 chest X-ray images from 6376 patients with at least two PA or AP chest X-ray images per patient. The image pixel spacing of all images was resampled to 1.0×1.0 mm^2 (151 to 431 and 158 to 431 pixels for rows and columns, respectively) using bicubic interpolation. Furthermore, the bit depth was rescaled linearly down to 8 bits. The second public dataset, CheXpert [52], included 32,972 chest X-ray images from 10,224 patients. We used the CheXpert-V1.0-Small dataset in which the image pixel size and bit depth of all images were 512×512 pixels and 8 bits, respectively.

The first clinical dataset named "Morishita Laboratory (ML)" comprised 81,109 chest X-ray images from 31,366 patients who underwent chest X-ray screening examinations between March 1986 and October 2004 in Iwate, Japan. Images of patients in the ML dataset were acquired using a computed radiography system (Fujifilm, Tokyo, Japan) under a matrix size of 1760×1760 pixels, pixel spacing of 0.2×0.2 mm^2, and 10-bit greyscale. The second clinical dataset named "Yamaguchi University (YU)" comprised 25,547 chest X-ray images from 12,517 patients obtained by random sampling from patients aged 20 years or older who underwent routine chest X-ray examinations between October 2015 and March 2017 at Yamaguchi University Hospital, Japan. All the images in the YU dataset were acquired using a flat-panel detector system (AeroDR; Konica Minolta Inc., Tokyo, Japan) or a computed radiography system (REGIUS MODEL 110; Konica Minolta Inc., Tokyo, Japan) with a matrix size of 2430×1994 pixels, pixel spacing of 0.175×0.175 mm^2, and 12-bit greyscale. The data were processed using the diagnostic imaging workstation (model CS-7 version 1.20–1.30; Konica Minolta Inc., Tokyo, Japan). In addition to the public dataset, we resampled all the images in the two clinical datasets to images with a pixel spacing of 1.0×1.0 mm^2 using bicubic interpolation. Furthermore, the bit depth was rescaled linearly down to 8 bits.

DCNN Learning

Outline of DCNN Learning by the Proposed Model

The EfficientNetV2-S backbone network with metric learning was used to train the DCNN for classifying individual examined patients based on chest X-ray images. Deep learning was performed using a computer with a GeForce 3090 Ti (NVIDIA, Santa Clara, Calif) graphics processing unit, Core i9-10900X 3.70 GHz (Intel, Corp., Santa Clara, Calif) central processing unit, and 128 GB of random-access memory.

Python 3.10.8 and PyTorch 1.13.0 + cu117 were used to perform DCNN training, validation, and testing using the element-wise adaptive sharpness-aware minimization optimizer [53] (base optimizer, SGD with a momentum of 0.8, and a weight decay of 0.0005) with a neighborhood size, rho, of 2.0, batch size of 60, smooth cross-entropy loss function with a label smoothing of 0.1, and cosine-decay learning rate scheduler (initial–last: 0.02–0.0004) without warmup. During

Table 1 Datasets used for DCNN training, validation, and testing

Public dataset	ChestXray8	CheXpert	PadChest	
Number of patients	3,245	10,224	6,376	Patients
Number of images	57,452	32,972	19,713	Images
Gender				Patients
Male	1,892	5,272	3,594	
Female	1,353	4,952	2,780	
Others	0	0	2	
View position				Images
PA	26,018	6	7,867	
AP	31,434	32,966	11,846	
Age categories				Patients
Neonate, infant (0 years)	0		152	
Young child (1–4 years)	0		222	
Older child (5–10 years)	0		85	
Adolescent (11–19 years)	0		62	
Adult (20 years or more)	3,245		5,855	
Examination interval categories (from baseline to follow-up X-ray examination dates)				Image–pairs
Under 1 year			10,559	
From 1 to 5 years			2,241	
Over 5 years			537	

Clinical dataset	ML	YU	
Number of images	81,109	25,547	images
Number of patients	31,366	12,517	patients
Examined baseline X-rays	12,790	8,417	
Examined baseline and follow-up X-rays	18,576	4,100	
Age (at the baseline examination)			years
Mean ± SD	55 ± 15	64 ± 16	
Range	15–95	20–100	
Gender			patients
Male	13,845	6,653	
Female	17,521	5,864	
View position			images
PA	81,109	23,953	
AP	0	1,594	
Subsets of clinical dataset (with at least two examinations per patient)			
Number of images	68,319	17,130	images
Number of patients	18,576	4,100	patients
Pair of view positions (baseline and follow-up)			pairs
PA–PA	49,743	11,487	
AP–PA	0	613	
PA–AP	0	306	
AP–AP	0	624	
Examination intervals between the baseline and follow-up X-ray examination dates			days
Mean ± SD	1,051 ± 635	100 ± 115	
Range	0–2,932	0–489	

ML Morishita Laboratory, *YU* Yamaguchi University, *PA* Posteroanterior, *AP* Anteroposterior, *SD* Standard Deviation, *PA–PA* subset for PA view at both examinations, *AP–PA* subset for AP view at baseline and PA view at follow–up examinations, *PA–AP* subset for PA view at baseline and AP view at follow–up examinations, *AP–AP* subset for AP view for both examinations

each training epoch, each image was augmented to correspond to the variabilities in clinical chest X-ray images as follows: random rotation between − 10° and 10° and random perspective transformation with a probability of 0.1 and distortion scale of 0.25 with bicubic resampling. The training time of the proposed method was approximately 70 h for 300 epochs of learning.

Feature Extractor Used by the Proposed Model

The EfficientNetV2-S backbone network with metric learning was used to train the DCNN for classifying individual examined patients based on chest X-ray images. EfficientNet uses mobile inverted bottleneck convolution (MBConv), similar to MobileNetV2 [54] and MnasNet [55]. A compound scaling method that uniformly scales each dimension with a fixed set of scaling coefficients is incorporated in the system [56]. In EfficientNetV2, fused-MBConv layers replace the 3 × 3 depth-wise convolution (conv3 × 3), and the 1 × 1 boosted convolution (conv1 × 1) is expanded in MBConv in the original EfficientNet with a normal 3 × 3 convolution to improve training performance [57]. The output values of the feature extractor trained using the proposed method were used as features to characterize individuals.

Classifier Used by the Proposed Model

Conventional image classifiers cannot achieve high generalizability even with deep learning and require numerous training samples of medical images and a long training time, which degrades the performance of the DCNN. By contrast, deep metric learning is a technique for improving the accuracy of classifiers [58]. Deep metric learning can be used to develop a feature extractor with high generalizability using fewer classes and images [58].

Figure 2 displays the classifier used by the proposed method that converts the features in the feature extractor output into the prediction of the 3245 classes of patients for the training subset. The classifier network comprises a nonlinear fully connected (FC) layer with metric learning layer (Fig. 2). The nonlinear FC layers comprise the following sequential layers: a linear layer, dropout layer, rectified linear unit (ReLU) activation function, and linear layer. Linear function applies a linear transformation to input matrices. ReLU is an activation function that renders the network nonlinear and fits complex data. Dropout is a regularization method for reducing overfitting in neural networks during training. Furthermore, we introduced AdaCos [59], which is a deep metric learning technique, to control dimensionality reduction complexity while preserving the individual characteristics of each patient. AdaCos [59] can automatically determine hyperparameters and successfully perform deep metric learning without additional tuning steps.

Similarity Index by the Proposed Model

The cosine similarity between the features in the follow-up and baseline chest X-ray images was used as the similarity index. The cosine similarity ranged from − 1 to 1 and was used for distinguishing same or distinct matched patient pairs. Values closer to 1 present the same patient pair. The cosine similarity is considered superior to the distance measurement in face verification [60].

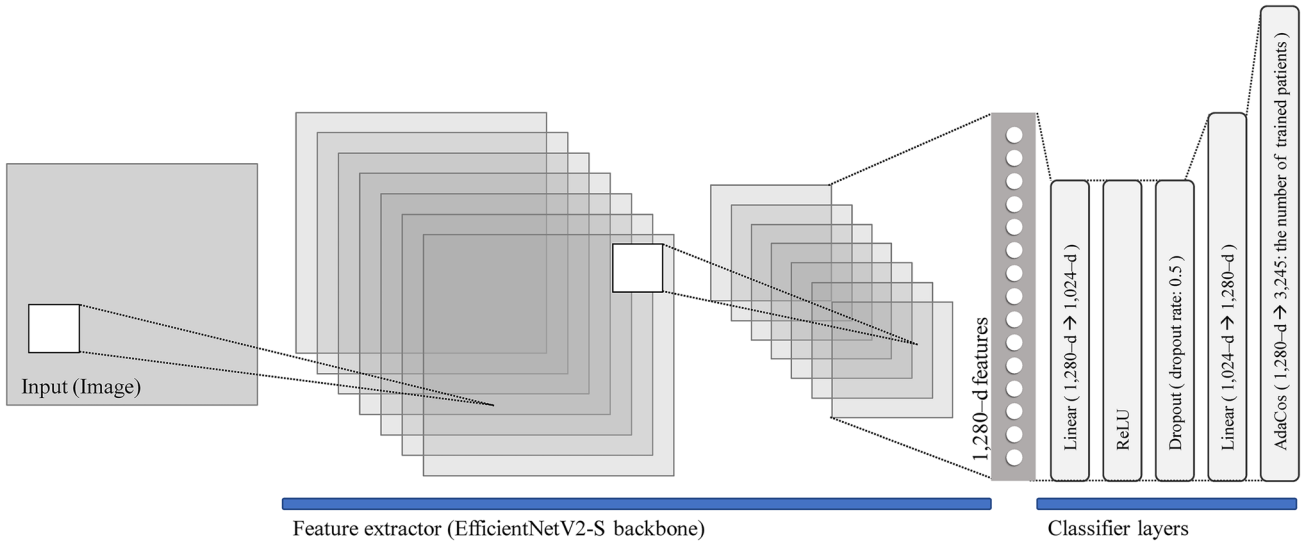

Fig. 2 Overview of the DCNN used in the proposed method. EfficientNetV2-S backbone is used as the feature extractor, and the classifier network consists of a linear layer, ReLU activation function, dropout layer, linear layer, and AdaCos [59] in that order. ReLU: rectified linear unit

Hyperparameter Evaluation of the Proposed Model

Several hyperparameter settings of the proposed method were evaluated in terms of each top-1 accuracy on the PadChest dataset divided into the following four subsets in terms of view position combinations at the baseline and each follow-up examination: PA view at both examinations (PA–PA), AP view at the baseline and PA view at follow-up examinations (AP–PA), PA view at the baseline and AP view at follow-up examinations (PA–AP), and AP view for both examinations (AP–AP).

In this study, we evaluated the following hyperparameters: number of epochs and number of features in the feature extractor output. The number of features was varied by further applying a linear transformation to the output from the feature extractor.

Verification Performance Evaluation of the Proposed Method

The image with the oldest examination date in the PadChest, YU, and ML datasets was used as the baseline image, and the other images were used as follow-up images. To evaluate the verification performance, this study used all image pairs from the same patient and the image pairs from different patients by random sampling: PadChest (6376 and 6376 of respective same and different patient pairs), YU (4100 and 4100 of respective same and different patient pairs), and ML (18,576 and 18,576 of respective same and different patient pairs) datasets. For the CheXpert dataset, we used the image-pair list, of which 8243 were of same patient pairs and 8243 of different patient pairs, used in a previous study [12].

The verification performance of the proposed method was evaluated in terms of the area under the receiver operating characteristic (ROC) curve (AUC) and equal error rate (EER). The ROC curves were statistically compared using the unpaired DeLong's test [61].

Closed-Set Identification Performance Evaluation of the Proposed Method

The image with the oldest examination date in the PadChest, YU, and ML datasets was used as the baseline image, and the other images were used as the follow-up image by employing all combinations for closed-set identification performance evaluation: PadChest (baseline 6,376 images of 6376 patients and follow-up 13,337 images of 6376 patients), YU (baseline 12,517 images of 12,517 patients and follow-up 13,030 images of 4100 patients), and ML (baseline 31,366 images of 31,366 patients and follow-up 49,743 images of 18,576 patients) datasets.

The closed-set identification performance of the proposed method was evaluated in terms of top-1 and top-2 accuracies. Top-2 accuracy refers to the accuracy when any of the pairs within the top-two higher cosine similarities in all patient comparisons are the same patient pairs. The top-1 accuracy on each subset by view position combinations was compared statistically using the two proportion *Z*-test.

Results

Figure 3 shows the top-1 accuracy transition of the hyperparameter variabilities on the PadChest dataset and four subsets, which are PA–PA, PA–AP, AP–PA, and AP–AP, using the proposed method. As can be observed in Fig. 3a, the performance of the proposed method improved with the increase in the number of epochs to 300. As shown in Fig. 3b, the top-1 accuracy transition remains flat regardless of the number of features.

Figure 4 shows the top-1 accuracy transition of (a) patient age in the baseline examination and (b) examination interval between the variabilities in the baseline and follow-up categories on the PadChest dataset using the proposed method with the result of the two proportion *Z*-test. In Fig. 4a, it can be observed that the top-1 accuracy increased until the age category reached adolescent. The top-1 accuracy of the adult category was significantly lower than that of adolescent ($p < 0.05$). Furthermore, in Fig. 4b, it is seen that the top-1 accuracy with an examination interval of less than one year was not significantly different from those with examination intervals from one to five years and greater than five years.

Table 2 summarizes the verification performance of the proposed method in terms of AUC and EER values on two public and two clinical datasets with 300 epochs and 1,280 features. Significant differences were observed between each different dataset ($p < 0.01$). The AUC values of the proposed method on the CheXpert (AUC = 0.9943) and ML datasets (AUC = 0.9999) were superior to those of the methods proposed by Packhäuser et al. on the CheXpert dataset (AUC = 0.9870) [12], Morishita et al. on the ML dataset (AUC = 0.993) [16], and Shimizu et al. on the ML dataset (AUC = 0.994) [17].

Table 3 summarizes the closed-set identification performance of the proposed method in terms of top-1 and top-2 accuracies on a public and two clinical datasets with 300 epochs and 1280 features. The closed-set identification performance of follow-up examinations with different view positions (PA–AP or AP–PA) was significantly lower than that with a same view position (PA–PA or AP–AP).

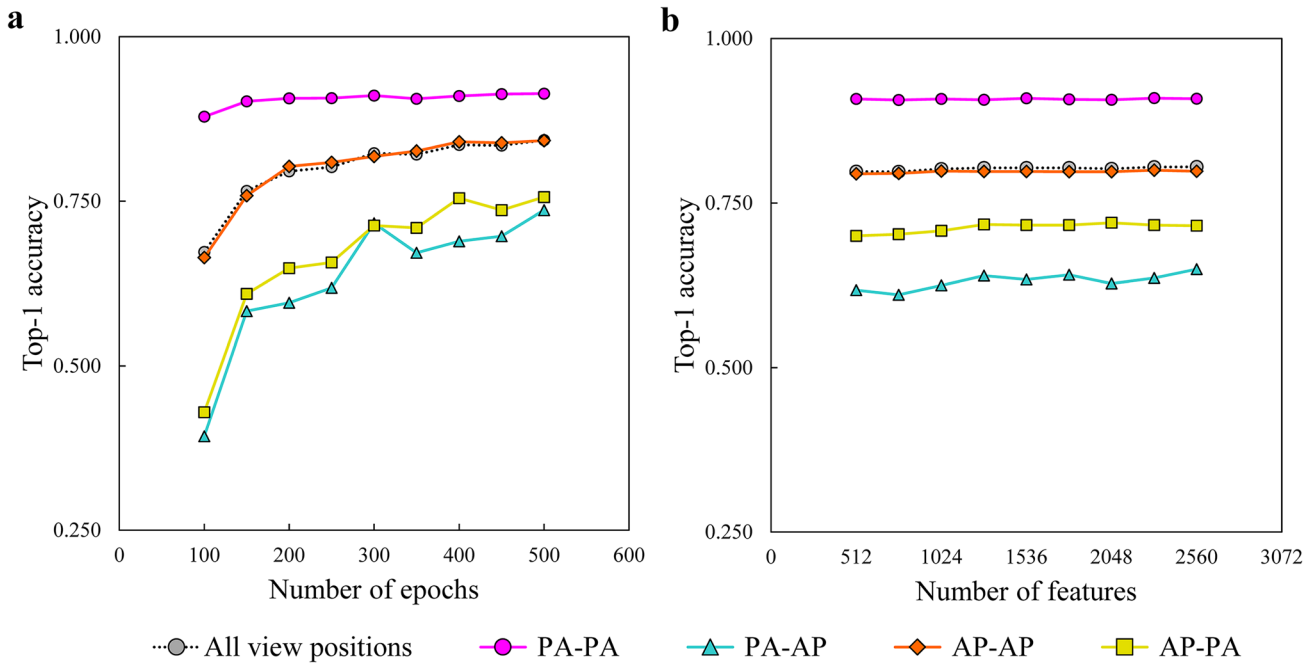

Fig. 3 Hyperparameter transition in the proposed method on the PadChest dataset. Two hyperparameters, number of epochs, and number of features in the feature extractor output variabilities show **a** the relationship between the number of epochs and top-1 accuracy, **b** the relationship between the number of features and top-1 accuracy. PA–PA subset for PA view at both examinations, PA–AP subset for PA view at baseline and AP view at follow-up examinations, AP–AP subset for AP view for both examinations, AP–PA subset for AP view at baseline and PA view at follow-up examinations

Furthermore, although the top-2 accuracy increased to 0.984, the top-1 accuracy showed a sharp decrease to 0.749 on the YU dataset.

Figure 5 shows an important area visualization that identifies the patient using the model trained based on the proposed method by Grad-CAM [62]. Grad-CAM generates heat maps of characteristic features and superimposes them on the original X-ray image to visualize the important regions that affect patient verification.

Discussion

In this study, a DCNN feature extractor and classifier based on the similarity index was used to verify the identity of examined patients using clinical chest X-ray images. The proposed method is available for use not only from the second examination onwards without additional training but also extracts robust features regarding different view positions. The number of features output by the EfficientNetV2-S backbone is used as a hyperparameter in the proposed method. We set the number of features to 1280 and the number of epochs to 300. The results revealed that this evaluation method achieved high performance not only on public datasets containing a large amount data and different view positions but also on clinical datasets that include clinical image variabilities (Tables 2–3). Packhäuser et al. [12] achieved an AUC value of 0.9870 on the CheXpert dataset. Furthermore, Morishita et al. [16] and Shimizu et al. [17] achieved the highest AUC values of 0.993 and 0.994, respectively, on the ML dataset. In comparison, the proposed method achieved AUC values of 0.9999 and 0.9943 on the ML and CheXpert datasets, respectively, which are both higher than those achieved by the previous methods [12, 16, 17].

In the comparison (PA–AP and AP–PA subsets in the PadChest dataset) between the combinations of different view positions shown in Fig. 3a, the top-1 accuracy improved when the number of epochs was increased to 300. The results in the PA–PA and AP–AP comparisons, that is, combinations of the same view position, were sufficiently accurate even with 200 epochs, and increasing the number of epochs did not improve the accuracy. Therefore, the optimal number of epochs depends on the type of view positions present in the dataset. These results suggest that underfitting may occur in feature extraction at different view positions up to 300 epochs. More than 300 epochs are required to extract robust features for different view positions. We assume that, depending on the model, when the number of epochs exceeds 300 during training, techniques such as regularization and data augmentation are needed to suppress overfitting in the proposed method.

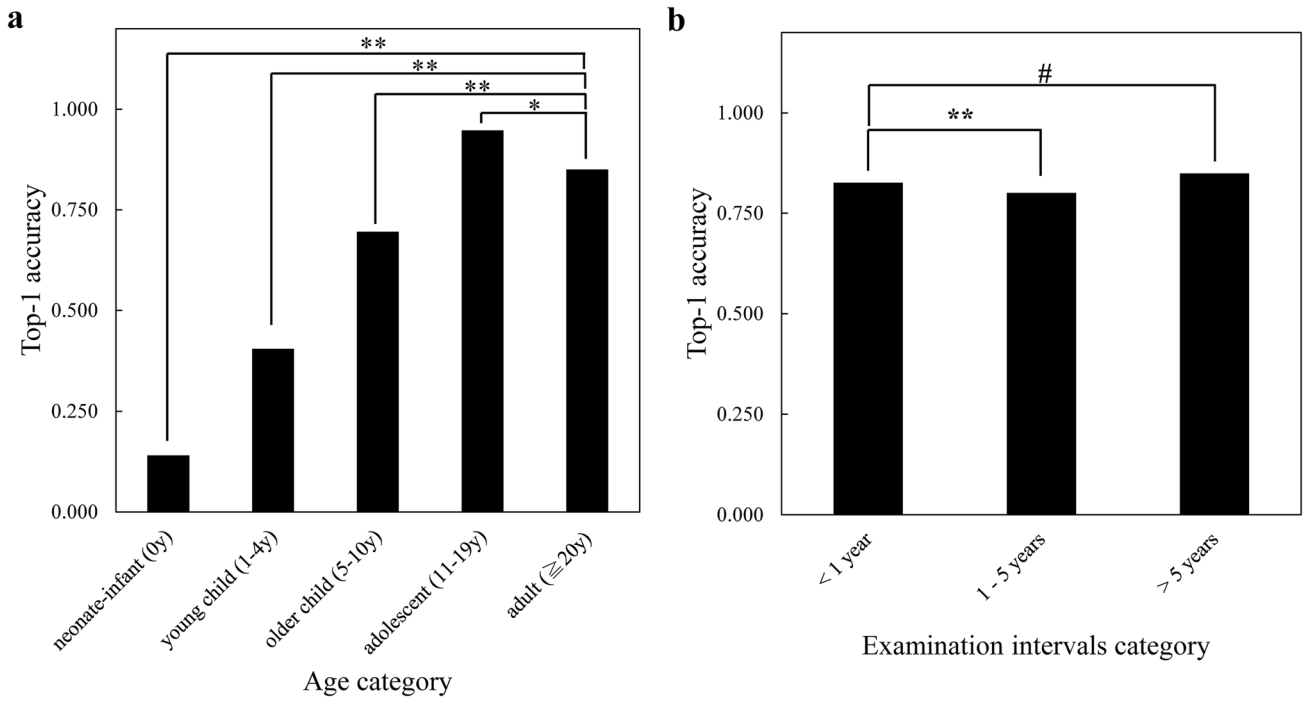

Fig. 4 Top-1 accuracy performance using the proposed method on the PadChest dataset. Two categories, age and examination intervals, variabilities show **a** the relationship between the age category subsets and top-1 accuracy, **b** the relationship between the examination intervals category subsets and top-1 accuracy. Neonate-infant: subset under the age of 0 years; young child: subset aged between 1 and 4 years; older child: subset aged between 5 and 10 years; adolescent: subset aged between 11 and 19 years; adult: subset aged 20 years or more. < 1 year: subset in which the examination interval is less than 1 year, 1–5 years: subset in which the examination interval is between 1 and 5 years, > 5 years: subset in which the examination interval is 5 years or more. Asterisks: significant difference (two proportion *Z*-test; $*p<0.05$ and $**p<0.01$). Pound: no significant difference (two proportion *Z*-test; $\#p>0.05$)

An increase in the interval between two examinations negligibly affected the top-1 accuracy of the proposed method, as shown in Fig. 4b; therefore, the examination intervals are expected to not be problematic for clinical applications of this method.

Areas around the devices implanted in the three patient examples shown in Fig. 5 are hardly recognized as important areas for feature extraction. Although we are concerned that the devices might be recognized as characteristic features, we believe that the proposed method has the potential to be applied to patient verification and identification from chest X-ray images even if high-contrast objects, such as accessories (removable) or metal implants (not removable), are present in the imaging area. Furthermore, changes in lung condition due to pneumonia or slightly oblique patient positions do not affect the performance of the proposed method in extracting patient-characteristic features.

This study has certain limitations. First, the results suggest that patient identification using chest X-ray images has a domain-shift problem in different view positions, including the reproducibility of the positioning. The performance of the proposed method on the ML dataset was consistently superior to that on the YU, PadChest, and CheXpert datasets in terms of AUC, EER, and top-1 and top-2 accuracies,

Table 2 Verification performance of the proposed method on two public and two clinical datasets

Dataset	EER	AUC	p value		
			vs. CheXpert	vs. ML	vs. YU
PadChest	2.70×10^{-2}	0.9899	2.02×10^{-7}	2.20×10^{-16}	2.20×10^{-16}
CheXpert	2.96×10^{-2}	0.9943	–	2.20×10^{-16}	2.98×10^{-6}
ML	5.65×10^{-4}	0.9999	–	–	9.10×10^{-4}
YU	5.12×10^{-3}	0.9980	–	–	–

p value calculated for AUC values by the unpaired DeLong test. *ML* Morishita Laboratory, *YU* Yamaguchi University, *EER* Equal Error Rate, *AUC* Area under the receiver operating characteristic curve

Table 3 Identification performance of the proposed method on a public and two clinical datasets

Dataset	Subset	Number of patients	Top-1 accuracy	*p* value	Top-2 accuracy
PadChest	PA–PA	3,584	0.910	1.42×10^{-68}	
	PA–AP	1,412	0.717		
	AP–AP	7,184	0.818	1.01×10^{-16}	
	AP–PA	1,157	0.713		
	All view positions	13,337	0.823		0.875
ML	PA–PA	49,743	0.993		0.999
YU	PA–PA	11,487	0.738	0.742	
	PA–AP	613	0.731		
	AP–AP	624	0.923	1.23×10^{-9}	
	AP–PA	306	0.781		
	All view positions	13,030	0.747		0.981

p value calculated for top-1 accuracy by the two proportion *Z*-test. *ML* Morishita Laboratory, *YU* Yamaguchi University, *PA–PA* subset for PA view at both examinations, *AP–PA* subset for AP view at baseline and PA view at follow-up examinations, *PA–AP* subset for PA view at baseline and AP view at follow-up examinations, *AP–AP* subset for AP view for both examinations

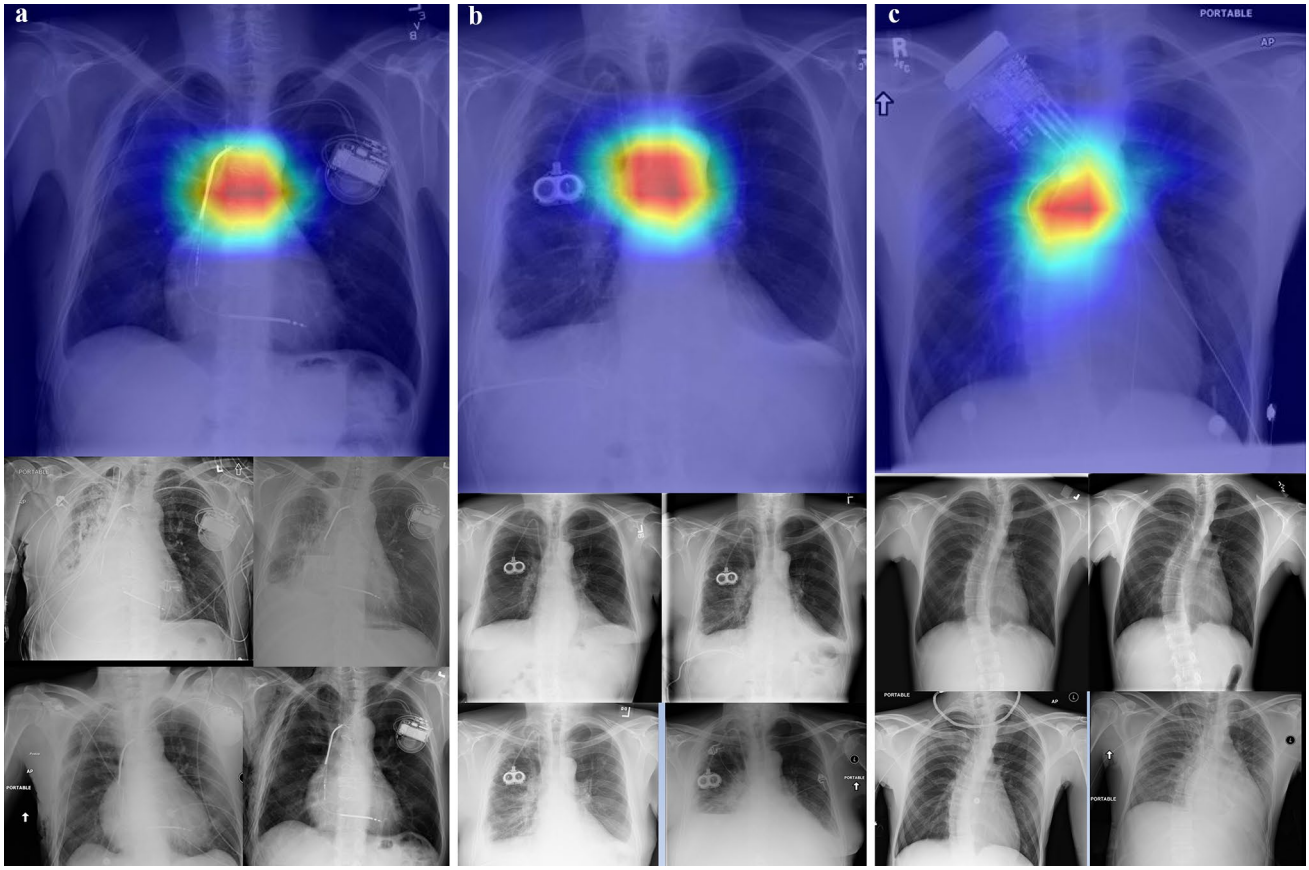

Fig. 5 Visualization of an important area that identifies the patient with the trained model of the proposed method using Grad-CAM. Example images of three patients (**a**–**c**) with five images each, one image from the validation dataset and four images from the training dataset. **a**–**b** Example image of patients with an implanted device (**a** pacemaker and **b** CV port). **c** Example image of a patient with temporary devices (some devices are in the training image and the other is in the validation image). The validation images overlayed the Grad-CAM [62] map to visualize the parts of an image that are important for patient identification using the proposed method. The patients in the first and second rows are different

likely because all patients in the ML dataset underwent the chest X-ray imaging in a standing or sitting PA view position. In terms of the reproducibility of positioning, the ML dataset, which does not include the AP view position, is comprehensively superior to the YU, PadChest, and CheXpert datasets. Furthermore, results revealed a degradation in the top-1 accuracy performance between different view positions (AP–PA or PA–AP) compared with images from the same view position (PA–PA or AP–AP). We demonstrated that one of the solutions to this problem is to increase the number of epochs and suppress underfitting. Although the performance on the YU, PadChest, and CheXpert datasets was not as effective as that on the ML dataset, the metrics showed reasonably good results. However, there is a statistically significant difference between the top-1 accuracies for the same and different view positions. We expect further accuracy improvements using applications such as domain adaptation techniques [12, 28].

Second, in Fig. 4a, significant differences can be observed in the performance of the proposed method in terms of the top-1 accuracies for each patient age category. The top-1 accuracies were significantly lower for the neonate-infant, young child, and older child categories compared with the adult category. We concluded that the performance degradation was because the model underfitted to the changes in the patient's body over the growth period, which could be attributed to no images of patients aged under 20 in the training dataset. Further research is required on efficient learning with small datasets.

Third, although the top-2 accuracy increased to 0.984, the top-1 accuracy showed a sharp decrease to 0.749 on the YU dataset. At the hospital that provided the YU dataset, a new patient identifier (ID) is issued when a patient visits for the first time; however, in some cases, an already registered patient with a patient ID was assigned a new patient ID. In this case, two patient IDs were assigned to the same patient. In this study, the patient ID registered in the DICOM header was treated as one patient with one ID. Therefore, if the patient ID recorded in the DICOM header is different, it will be regarded as the image of a different patient, even if it is the same patient. However, if the similarity is high because it is the same patient, the top-1 accuracy still degrades. Furthermore, the YU dataset contains some inadequate X-ray images resulting from errors that occur while reading and transferring the images. However, the patient ID is correctly recorded in the header of these images. Although it is the same patient on the ID recorded in the DICOM images, the results show that the similarity index is low between inadequate and other adequate images. A record that has been assigned two IDs is managed in the hospital information system (HIS) or radiology information system (RIS). It is, therefore, desirable to use the patient IDs recorded in HIS/RIS information, not the patient IDs recorded in DICOM images, for lists used in biometric systems in clinical applications. Moreover, we believe that measures to disable inadequate images presented above are desirable.

In addition, in this study, each clinical dataset was obtained using X-ray imaging devices with the same model and similar acquisition parameters. Different X-ray imaging devices have different image contrast in chest X-ray images; therefore, the performance achieved in this study may not be reproducible. To evaluate the biometric performance on a variety of image contrasts, a dataset consisting of clinical X-ray images from several clinics or hospitals using common patient identifier is required.

The proposed method can confirm that the patient who underwent chest X-ray examination and the patient to be examined are the same. Moreover, the proposed method is available for checking the misfiling patient identifier on the picture archiving and communication system. We thereby expect that the proposed method will be useful for safety systems that could prevent these patient misidentification problems and will be able to contribute to the reduction of the risk of incorrectly identifying patient due to human error.

Conclusions

Potential risks that can lead to medical malpractice, such as patient misidentification in clinical examinations and misfiling of images stored in picture archiving and communication systems, are still inherent in clinical work. In this study, a patient verification and identification method was proposed using a deep feature extractor with robustness under different view positions from images obtained by clinical chest X-ray examination. The performance of the proposed method was verified not only on two public datasets but also on two clinical image datasets, namely ML and YU. Clinical chest X-ray examinations that utilize cutting-edge technology (such as the technique presented in this study) provide a key solution to prevent patient misidentification resulting from human errors.

Acknowledgements The authors thank Dr. Shohei Kudomi at Yamaguchi University Hospital, Japan, for his valuable insights on improving our manuscript; Drs. Masatoshi Yamane and Takuya Uehara at Yamaguchi University Hospital, Japan, for organizing the Yamaguchi University Dataset; Drs. Yongsu Yoon at Dongseo University, Republic of Korea, and Yoichiro Shimizu at Yamaguchi University Hospital, Japan, for organizing the Morishita Lab. Dataset; and Editage (http://www.editage.com) for providing English language editing.

Author Contribution All authors contributed to the study conception and design. Material preparation and data collection were performed by Yasuyuki Ueda and Junji Morishita, and analysis was performed by Yasuyuki Ueda. The first draft of the manuscript was written by Yasuyuki Ueda, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by JSPS KAKENHI [grant numbers: JP18K15590, 19K08122 and JP21K15827]. Open access funding provided by Osaka University.

Data Availability All chest X-ray images in this study were compiled from three public and two clinical datasets. ChestX-Ray8 published by Wang X et al. [50] is available in the ChestXray-NIHCC repository at https://nihcc.app.box.com/v/ChestXray-NIHCC/. PadChest published by Bustos A et al. [51] is available in the PadChest repository at https://bimcv.cipf.es/bimcv-projects/padchest/. CheXpert published by Irvin J et al. [52]. is available in the CheXpert repository at https://stanfordmlgroup.github.io/competitions/chexpert/. Two clinical datasets, ML and YU, are owned by the Morishita Laboratory in Kyushu University, Japan, and Yamaguchi University Hospital, Japan, respectively. ML and YU cannot be made publicly available owing to patient privacy, its proprietary nature, and ethical concerns.

Code Availability The code necessary to reproduce the results achieved by the proposed method on public datasets can be found at https://github.com/d83yk/dml-cxr-patient-identification.

Declarations

Ethics Approval All procedures in this study were in conformance with the ethical standards of the Institutional Review Board at each authors' affiliated institutions and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent to Participate Written informed consent was not required for this study because of the retrospective nature of the study.

Consent for Publication Written consent of the study participants or their legal guardians was not required for publishing the data of the individuals in the manuscript because of the retrospective nature of this study.

Competing Interests The authors declare no competing interests.

References

1. Morishita J, Ueda Y: New solutions for automated image recognition and identification: challenges to radiologic technology and forensic pathology. Radiol Phys Technol 14:123-133, 2021. https://doi.org/10.1007/s12194-021-00611-9.
2. Morishita J, Katsuragawa S, Kondo K, Doi K: An automated patient recognition method based on an image-matching technique using previous chest radiographs in the picture archiving and communication system environment. Med Phys 28(6):1093–1097, 2001. https://doi.org/10.1118/1.1373403
3. Morishita J, Watanabe H, Katsuragawa S, et al.: Investigation of misfiled cases in the PACS environment and a solution to prevent filing errors for chest radiographs. Acad Radiol 12(1):97-103, 2005. https://doi.org/10.1016/j.acra.2004.11.008.
4. Bittle MJ, Charache P, Wassilchalk DM: Registration-associated patient misidentification in an academic medical center: causes and corrections. Jt Comm J Qual Patient Saf 33(1):25-33, 2007. https://doi.org/10.1016/S1553-7250(07)33004-3
5. Danaher LA, Howells J, Holmes P, Scally P: Is it possible to eliminate patient identification errors in medical imaging? J Am Coll Radiol 8(8):568-574, 2011. https://doi.org/10.1016/j.jacr.2011.02.021.
6. Seiden SC, Barach P. Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: Are they preventable? Arch Surg 141(9):931-939, 2006. https://doi.org/10.1001/archsurg.141.9.931
7. Henneman PL, Fisher DL, Henneman EA, Pham TA, Campbell MM, Nathanson BH: Patient identification errors are common in a simulated setting. Ann Emerg Med 55(6):503-509, 2010. https://doi.org/10.1016/j.annemergmed.2009.11.017.
8. Schulmeister L: Patient misidentification in oncology care. Clin J Oncol Nurs 12(3):495-498, 2008. https://doi.org/10.1188/08.CJON.495-498.
9. The Joint Commission: National Patient Safety Goals Effective July 2020 for the Hospital Program. 2020. Available at https://www.jointcommission.org/-/media/tjc/documents/standards/national-patient-safety-goals/2020/npsg_chapter_hap_jul2020.pdf. Accessed Sep 28, 2021.
10. Emergency Care Research Institute: Patient identification errors. 2016. Available at https://www.ecri.org/Resources/HIT/Patient%20ID/Patient_Identification_Evidence_Based_Literature_final.pdf. Accessed Sep 28, 2021.
11. Healthcare Financial Management Association. 2016. The Value of Precise Patient Identification. Accessed Sep 28, 2021. https://www.imprivata.com/resources/whitepapers/hfma-educational-report-value-precise-patient-identification.
12. Packhäuser K, Gündel S, Münster N, et al.: Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data. Sci Rep 12:14851, 2022. https://doi.org/10.1038/s41598-022-19045-3
13. Ueda Y, Morishita J, Kudomi S, Ueda K: Usefulness of biological fingerprint in magnetic resonance imaging for patient verification. Med Biol Eng Comput 54:1341-1351, 2016. https://doi.org/10.1007/s11517-015-1380-x.
14. Ueda Y, Morishita J, Hongyo T: Biological fingerprint using scout computed tomographic images for positive patient identification. Med Phys 46:4600-4609, 2019. https://doi.org/10.1002/mp.13779.
15. Ueda Y, Morishita J, Kudomi S: Biological fingerprint for patient verification using trunk scout views at various scan ranges in computed tomography. Radiol Phys Technol 15: 398-408, 2022. https://doi.org/10.1007/s12194-022-00682-2
16. Morishita J, Katsuragawa S, Sasaki Y, Doi K: Potential usefulness of biological fingerprints in chest radiographs for automated patient recognition and identification. Acad Radiol 11:309–315, 2004. https://doi.org/10.1016/s1076-6332(03)00655-x.
17. Shimizu Y, Matsunobu Y, Morishita J: Evaluation of the usefulness of modified biological fingerprints in chest radiographs for patient recognition and identification. Radiol Phys Technol 9:240-244, 2016. https://doi.org/10.1007/s12194-016-0355-4.
18. Shimizu Y, Morishita J: Development of a method of automated extraction of biological fingerprints from chest radiographs as preprocessing of patient recognition and identification. Radiol Phys Technol 10:376-381, 2017. https://doi.org/10.1007/s12194-017-0400-y.
19. Kao EF, Lin WC, Jaw TS, Liu GC, Wu JS, Lee CN: Automated patient identity recognition by analysis of chest radiograph features. Acad Radiol 20:1024-1031, 2013. https://doi.org/10.1016/j.acra.2013.04.006.
20. Shamir L, Ling S, Rahimi S, Ferrucci L, Goldberg IG: Biometric identification using knee X-rays. Int J Biom 1:365-370, 2009. https://doi.org/10.1504/IJBM.2009.024279.

21. Lamb JM, Agazaryan N, Low DA: Automated patient identification and localization error detection using 2-dimensional to 3-dimensional registration of kilovoltage x-ray setup images. Int J Radiat Oncol Biol Phys 87:390-393, 2013. https://doi.org/10.1016/j.ijrobp.2013.05.021.
22. Silverstein E, Snyder M: Implementation of facial recognition with Microsoft Kinect v2 sensor for patient verification. Med Phys 44:2391-2399. 2017. https://doi.org/10.1002/mp.12241
23. Wiant DB, Verchick Q, Gates P, et al.: A novel method for radiotherapy patient identification using surface imaging. J Appl Clin Med Phys 17:271-278, 2016. https://doi.org/10.1120/jacmp.v17i2.6066
24. Parks CL, Monson KL: Automated facial recognition of computed tomography-derived facial images: patient privacy implications. J Digit Imaging 30:204-214, 2017. https://doi.org/10.1007/s10278-016-9932-7
25. Koike-Akino T, Mahajan R, Marks TK, et al.: High-accuracy user identification using EEG biometrics. Conf Proc IEEE Eng Med Biol Soc 2016: 854-858, 2016.
26. Toge R, Morishita J, Sasaki Y, Doi K: Computerized image-searching method for finding correct patients for misfiled chest radiographs in a PACS server by use of biological fingerprints. Radiol Phys Technol 6:437–443, 2013. https://doi.org/10.1007/s12194-013-0221-6
27. Sakai Y, Takahashi K, Shimizu Y, Ishibashi E, Kato T, Morishita J: Clinical application of biological fingerprints extracted from averaged chest radiographs and template-matching technique for preventing left-right flipping mistakes in chest radiography. Radiol Phys Technol 12:216-223, 2019. https://doi.org/10.1007/s12194-019-00504-y
28. Nguyen K, Nguyen HH, Tiulpin A: AdaTriplet: Adaptive gradient triplet loss with automatic margin learning for forensic medical image matching. In: Wang L, Dou Q, Fletcher PT, Speidel S, Li S (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol 13438, 725–735, 2022. https://doi.org/10.1007/978-3-031-16452-1_69
29. Ciaffi R, Gibelli D, Cattaneo C: Forensic radiology and personal identification of unidentified bodies: a review. Radiol Med 116:960-968, 2011. https://doi.org/10.1007/s11547-011-0677-6
30. Clemente MA, La Tegola L, Mattera M, Guglielmi G: Forensic radiology: An update. J Belg Soc Radiol 101(Suppl 2):21, 2017. https://doi.org/10.5334/jbr-btr.1420
31. Matsunobu Y, Morishita J, Usumoto Y, Okumura M, Ikeda N: Bone comparison identification method based on chest computed tomography imaging. Leg Med (Tokyo) 29:1-5, 2017. https://doi.org/10.1016/j.legalmed.2017.08.002
32. Wada Y, Morishita J, Yoon Y, Okumura M, Ikeda N: A simple method for the automatic classification of body parts and detection of implanted metal using postmortem computed tomography scout view. Radiol Phys Technol 13:378-384, 2020. https://doi.org/10.1007/s12194-020-00581-4
33. Kawazoe Y, Morishita J, Matsunobu Y, Okumura M, Shin S, Usumoto Y, Ikeda N: A simple method for semi-automatic readjustment for positioning in post-mortem head computed tomography imaging. J Forensic Radiol Imaging 16:57–64, 2019. https://doi.org/10.1016/j.jofri.2019.01.004
34. Krishan K, Chatterjee PM, Kanchan T, Kaur S, Baryah N, Singh RK: A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. Forensic Sci Int 261:165.e1-165.e1658, 2016. https://doi.org/10.1016/j.forsciint.2016.02.007
35. Tsubaki S, Morishita J, Usumoto Y, et al.: Sex determination based on a thoracic vertebra and ribs evaluation using clinical chest radiography. Leg Med (Tokyo) 27:19-24, 2017. https://doi.org/10.1016/j.legalmed.2017.06.003
36. Kim TK, Yi PH, Wei J et al.: Deep learning method for automated classification of anteroposterior and posteroanterior chest radiographs. J Digit Imaging 32:925-930, 2019. https://doi.org/10.1007/s10278-019-00208-0
37. Bae J, Yu S, Oh J, Kim TH, Chung JH, Byun H, Yoon MS, Ahn C, Lee DK: External validation of deep learning algorithm for detecting and visualizing femoral neck fracture including displaced and non-displaced fracture on plain X-ray. J Digit Imaging. 34:1099-1109, 2021. https://doi.org/10.1007/s10278-021-00499-2
38. Suzuki T, Maki S, Yamazaki T, Wakita H, Toguchi Y, Horii M, Yamauchi T, Kawamura K, Aramomi M, Sugiyama H, Matsuura Y, Yamashita T, Orita S, Ohtori S: Detecting distal radial fractures from wrist radiographs using a deep convolutional neural network with an accuracy comparable to hand orthopedic surgeons. J Digit Imaging. 35:39-46, 2022. https://doi.org/10.1007/s10278-021-00519-1
39. Liu F, Gao L, Wan J, Lyu ZL, Huang YY, Liu C, Han M: Recognition of digital dental X-ray images using a convolutional neural network. J Digit Imaging. 36:73-79, 2023. https://doi.org/10.1007/s10278-022-00694-9.
40. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K: Deep learning for chest X-ray analysis: A survey. Med Image Anal. 72:102125, 2021. https://doi.org/10.1016/j.media.2021.102125
41. Agrawal T, Choudhary P: Segmentation and classification on chest radiography: a systematic survey. Vis Comput. 39:875-913, 2023. https://doi.org/10.1007/s00371-021-02352-7
42. Baltruschat IM, Nickisch H, Grass M et al.: Comparison of deep learning approaches for multi-label chest X-ray classification. Sci Rep 9: 6381, 2019. https://doi.org/10.1038/s41598-019-42294-8
43. Ieki H, Ito K, Saji M et al.: Deep learning-based age estimation from chest X-rays indicates cardiovascular prognosis. Commun Med (Lond) 2:159, 2022. https://doi.org/10.1038/s43856-022-00220-6
44. Gichoya JW, Banerjee I, Bhimireddy AR et al.: AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health. 4:e406-e414, 2022. https://doi.org/10.1016/S2589-7500(22)00063-2.
45. Yang M, Tanaka H, Ishida T: Performance improvement in multi-label thoracic abnormality classification of chest X-rays with noisy labels. Int J Comput Assist Radiol Surg 18(1):181-189, 2022. https://doi.org/10.1007/s11548-022-02684-2.
46. Kawakubo M, Waki H, Shirasaka T, Kojima T, Mikayama R, Hamasaki H, Akamine H, Kato T, Baba S, Ushiro S, Ishigami K: A deep learning model based on fusion images of chest radiography and X-ray sponge images supports human visual characteristics of retained surgical items detection. Int J Comput Assist Radiol Surg 2022. https://doi.org/10.1007/s11548-022-02816-8.
47. Wu C, Wu H, Lei S, Li X, Tong H: Deep learning in automatic fingerprint identification. 2021 IEEE 6th International Conference on Smart Cloud. 111–116, 2021. https://doi.org/10.1109/SmartCloud52277.2021.00027
48. Finizola J, Targino J, Teodoro F, Lima C: Comparative study between deep face, autoencoder and traditional machine learning techniques aiming at biometric facial recognition. 2019 International Joint Conference on Neural Networks (IJCNN) 1–8, 2022. https://doi.org/10.1109/IJCNN.2019.8852273
49. Priyadharshini AR, Arivazhagan S, Arun M: A deep learning approach for person identification using ear biometrics. Appl Intell (Dordr). 51:2161–2172, 2021. https://doi.org/10.1007/s10489-020-01995-8
50. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM: ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 3462–3471, 2017. https://doi.org/10.1109/CVPR.2017.369
51. Bustos A, Pertusa A, Salinas JM, de la Iglesia-Vayá M: PadChest: A large chest x-ray image dataset with multi-label annotated

reports. Med Image Anal 66, 101797, 2020. https://doi.org/10.1016/j.media.2020.101797
52. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghgoo B, Ball R, Shpanskaya K, Seekins J. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence 33:590-597, 2019. https://doi.org/10.1609/aaai.v33i01.3301590
53. Kwon J, Kim J, Park H, Choi IK: Asam: Adaptive sharpness-aware minimization for scale-invariant learning of deep neural networks. In Proceedings of the 38th International Conference on Machine Learning. PMLR. 2021. https://proceedings.mlr.press/v139/kwon21b.html.
54. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC.: Mobilenetv2: Inverted residuals and linear bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 4510–4520, 2018. https://doi.org/10.1109/CVPR.2018.00474
55. Tan M, Chen B, Pang R, et al.: MnasNet: Platform-aware neural architecture search for mobile. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2815–2823, 2019. https://doi.org/10.1109/CVPR.2019.00293
56. Tan M, Le Q: EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019. 6105–6114, 2019.
57. Tan M, Le Q: EfficientNetV2: Smaller models and faster training. Proceedings of the 38th International Conference on Machine Learning, PMLR 139:10096–10106, 2021.
58. Musgrave K, Belongie S, Lim SN: A metric learning reality check. In: Vedaldi A, Bischof H, Brox T, Frahm JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, 12370. 681–699, Springer, Cham. 2020. https://doi.org/10.1007/978-3-030-58595-2_41
59. Zhang X, Zhao R, Qiao Y, Wang X, Li H: AdaCos: Adaptively scaling cosine logits for effectively learning deep face representations. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 10815–10824, 2019. https://doi.org/10.1109/CVPR.2019.01108
60. Nguyen HV, Bai L: Cosine similarity metric learning for face verification. In: Kimmel R, Klette R, Sugimoto A. (eds) Computer Vision – ACCV 2010. ACCV 2010. Lecture Notes in Computer Science, vol 6493. 709–720, https://doi.org/10.1007/978-3-642-19309-5_55
61. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 44:837-845, 1988. https://doi.org/10.2307/2531595
62. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128:336-359, 2020. https://doi.org/10.1007/s11263-019-01228-7