# Re-Identification in Chest Radiographs After Anatomical Masking

**Kiyan Yaghnam**[1*]

*Received September 23, 2026*
*Accepted October 26, 2026*
*Electronic access November 30, 2026*

Medical imaging is often treated as anonymous once all patient-identifying metadata has been removed. This assumption does not always hold. A chest radiograph can be matched to other images of the same patient. One possible response is to mask the part of the image that gives the patient away. Whether that method truly works depends on how much anatomical information a model actually needs. We trained two re-identification models on ChestX-ray14 and evaluated them on 1,400 held-out patients, using anteroposterior (AP) queries against a posteroanterior (PA) gallery of reference images. We masked ten anatomical regions at four levels of severity, pairing each ablation with a control that placed the same mask at a random location. We find that masking any one of the ten tested regions does not prevent patient matching: masking all pixels within one lung field drops ResNet-50 precision@1 from 0.946 to 0.582, yet the correct patient remains among the top 20 of 4,611 gallery images for 87% of queries. With cumulative masking, the mean precision@1 for ResNet-50 with three seeds decreased from 0.9386 for the full image to 0.0204 when the two lung fields were masked and 0.0019 when all ten regions were masked. Only masking the lung fields and mediastinum produced consistently larger drops in retrieval performance than placing the same masks at random positions across both tested models. Medical images masked in a single region should be tested against a recognition model and should not automatically be assumed safe.

**Keywords:** Chest radiographs, patient re-identification, anatomical masking, medical image privacy, shape-matched ablation.

## Introduction

Medical imaging is generally considered to be anonymous after removing the patient's name and record number. But the removal of this identifiable metadata does not mean that the image is non-identifiable. For example, it has been shown that when given two chest X-rays, a deep learning algorithm can tell if they belong to the same person with an area under the curve greater than 0.99, even if the two images are captured over a decade apart[1]. Macpherson et al.[2] carried out patient matching and database search using more than one million radiographs in their internal dataset. This raises a challenge for sharing of radiographs because removing common identifiers does not eliminate the possibility of image-based matching.

One potential solution is to mask out the structures that the recognition algorithm looks at. However, it then becomes difficult to determine whether the importance of a certain region depends on where it is located or simply on how much space it takes up compared to other regions. For instance, masking out the lung field removes much more image information compared to the clavicle. These effects cannot be separated without a matched control.

Three questions guide our evaluation. Can re-identification survive with full masking of any single anatomical area? Does masking an area in its current position impair recognition more than relocating that particular mask to another spot? Are the findings consistent across both architectures and the training seeds we use? Our primary assessment utilizes AP radiographs as queries and PA radiographs as gallery images, where the images come from patients unseen during training. Retrieval is measured only if the matching patient is in the gallery. Rejection of queries whose patient is absent is not tested.

### Contributions

Anatomical masking is assessed using AP queries, a fixed PA gallery, and patients not seen in the training set. We also conducted a region-by-region ablation study over ten segmented structures at four severity levels, each paired with a shape-matched control (the identical mask moved to a random position) that separates location from amount and shape. Within the tested models and dataset, re-identification remained possible after complete masking of each anatomical region individually. Removing the most detrimental region tested still leaves the correct patient in the top 20 of 4,611 candidates for 87% of queries. Masking the lung fields and mediastinum caused larger drops in retrieval performance than placing the same masks at random locations under both architectures. The other tested regions did not consistently show this pattern.

[1] *ACS Amman; Amman, Jordan*
[*] *Corresponding author:* yaghnamkiyan@gmail.com

© The National High School Journal of Science 2026

## Related work

### Patient re-identification across acquisitions

Removing identifying metadata from chest radiographs does not necessarily make them anonymous. Packhäuser et al. (2022)[1] showed that radiographs can be used to determine whether two images belong to the same patient. Macpherson et al.[2] showed that chest X-ray re-identification also works as a retrieval problem and measured how accuracy falls as the search pool grows.

Radiographic identification has a longer history in forensic research, where skeletal landmarks have been compared to establish identity[3]. Ueda and Morishita[4] evaluated PA–PA, AP–AP, PA–AP, and AP–PA comparisons using PadChest. Kim et al.[5] tested patient verification across internal and external datasets and compared performance with radiologists. Sony et al.[6] examined the usefulness of anatomical regions such as the vertebrae and clavicles for forensic identification.

Gichoya et al.[7] used image degradation and anatomical occlusion to investigate race prediction, rather than patient matching, and did not compare occlusion with an identical mask moved elsewhere. Re-identification risks also extend beyond chest radiographs: Ganz et al.[8] demonstrated patient matching from histopathology slides. Enomoto et al.[9] identified patients from dental panoramic radiographs despite treatment-related changes.

Our study examines what happens when anatomical regions are removed from AP query images matched against a fixed PA gallery. Each removal is compared with the identical mask placed at another location, keeping its shape and pixel count unchanged. This follows perturbation-based occlusion methods[10,11] rather than gradient-based heatmaps[12]. The comparison helps measure how retrieval depends on where information is removed, although it cannot fully separate anatomical content from spatial position or acquisition-related signals that persist across views.

### Privacy-preserving transformations of medical images

Stripping identifiers from a medical image does not make it anonymous. The HIPAA Safe Harbor guidance[13] specifies categories of identifying information to remove, but compliance with an identifier list does not, by itself, measure the biometric information remaining in image pixels. Brain imaging has the longest record on this problem. Faces reconstructed from research brain MRI scans can be matched to photographs by commercial face-recognition software, even after every identifying label has been removed[14], and the same is true for brain PET, where reconstructed faces matched photographs at 32–42% depending on the tracer[15]. de Sitter et al.[16] and Buimer et al.[17] looked at the other side of the trade-off and found that removing facial features can change neuroimaging measurements and downstream processing. All of this work concerns defacing brain images. None of it asks how much individual anatomical regions contribute to retrieval in chest radiographs.

A second line of work tries to hide patient identity while keeping images clinically useful. Packhäuser et al. (2023)[18] built a deformation-based anonymization model for chest radiographs, and follow-up work[19] used latent diffusion to generate synthetic images instead. Montenegro and Cardoso[20] replaced identity features, Heinrich and Hansen[21] obscured identity by mixing patients through a neural network, and Li et al. (2025)[22] applied generative anonymization to MIMIC-CXR.

These methods transform the image. These studies evaluate privacy-preserving transformations, rather than the controlled anatomical-location ablation used here. They also show why a masked image should not be equated with a demonstrably anonymous one: a transformation can preserve diagnostically useful information and still leave some identifying signal. Dar et al.[23] found that latent diffusion models memorized and replicated training images. In their experiments on chest radiographs, 32.6% of the training images were memorized, and 54.5% of generated images were copies. This is a distinct privacy issue from the residual identifying information explored in our masking experiments. Molchanova et al.[24] evaluated the trade-off between MRI refacing, re-identification risk and brain-morphometry consistency.

### Recognition and retrieval with unseen identities

Biometric retrieval systems convert images into feature embeddings and rank potential matches by similarity. This allows new identities to be added without retraining the model. Matching becomes harder when viewpoints change or parts of an image are obscured. He et al.[25] studied camera and viewpoint differences. Wang et al. (2022)[26] examined matching under occlusion. Other work explored semantic-controlled pretraining[27], anatomy-based chest-radiograph retrieval[28], and part-aware matching for occluded people[29].

A separate challenge is determining whether a person is in the gallery at all. Dong et al.[30] studied open-set face identification and protected face templates. Wang et al. (2026)[31] investigated rejecting unknown identities using gallery-distance distributions. AdaFace[32] examined how image quality affects identity embeddings.

This distinction matters for our study. An unseen-identity evaluation holds test people out of training. Full open-set identification additionally permits a query with no correct identity in the gallery and measures whether the system rejects it. Our test patients were unseen during training, but every AP query had a matching patient in the PA gallery. We therefore evaluate ranked retrieval, not the rejection of unknown


© The National High School Journal of Science 2026

identities. Table 1 compares earlier studies by data source, cross-view evaluation, use of anatomical region masks, mask translation controls, and testing of unseen subject identities in training. Within the published studies compared in Table 1, we found no chest-radiograph experiment that pairs a complete anatomical-region mask with an exact translated copy as the retrieval control. What previous work does not establish in the retrieval setting is whether masking a particular region hurts patient matching more than removing the same amount and shape of image content elsewhere.

## Methods

### Data

We used ChestX-ray14[33] with 112,120 frontal chest X-rays of 30,805 patients. The associated file Data_Entry_2017_v2020.csv contains an anonymized "Patient ID", "Patient Age", "Patient Gender" and "View Position", which is either posteroanterior (PA, taken from the back) or anteroposterior (AP, taken from the front). PA is taken standing for a sharper image, while AP is used when a patient cannot stand. Because the heart sits farther from the detector on an AP view, it appears larger.

Patients having fewer than two images were excluded from our dataset because retrieval requires at least one image to perform search, and another one to identify. Also, only patients who had both PA and AP chest X-rays were included, because our cross-view setup described below requires both types of images. For each patient we capped the number of images at ten, selected with a fixed random seed. The distribution of the number of images per patient is highly unbalanced (median $= 6$, max $= 173$). Thus, to have a fair setup, where no patient will have an advantage over others, we limited the maximum number of images, since a patient with 173 gallery images has far more chances to be found than one with two. Finally, our dataset consists of 6,936 patients and 44,044 images, separated by patients into 4,836 for training, 700 for validation, and 1,400 for testing (see Appendix S1).

The age and sex characteristics of the training, validation, and test datasets were similar (Table S3). In comparison to the entire dataset of ChestX-ray14, our selected subset of patients was slightly older (47.25 vs. 45.53 years) and contained a larger fraction of male patients (57.44% vs. 53.98%). Such a disparity is consistent with our criteria for choosing at least two images from each patient and two different acquisition views. Nonetheless, it should be noted that the selected subset may be non-representative of the entire dataset. Capped patients were similar in demographics to uncapped patients ($n = 1{,}857$ vs. $n = 5{,}079$). The average age of capped patients was 47.01 years while the average age of uncapped patients was 47.33 years. The proportion of males was 59.45% among capped patients and 56.70% among uncapped patients. These small differences suggest that the cap did not substantially alter the cohort's age and sex composition, although other sources of selection bias remain possible.

Since all images of each patient were put into the same split, no patient was present in both training and test. Had this not been done, the model could have memorized the patients during training, giving us higher numbers without any actual ability to generalize. No patient in the test data is part of the training data, and the model has no output unit for any of them. We held out 1,400 patients for testing. As reference, Macpherson et al.[2] describe another ChestX-ray14 test subset of 1,424 patients (24,223 images) from the repository described by Packhäuser et al. (2022)[1]. These subsets cannot be seen as equivalent: the patient selection, gallery creation, and view-pairing protocols differ.

Images were converted to grayscale, resized to $320 \times 320$, saved in JPEG at quality 90 and loaded at size $224 \times 224$. Lossy compression is covered in the Limitations section of the Discussion.

The patient identifiers in the dataset are anonymous. We used these identifiers to check whether two images were of the same patient, not to identify anyone by name.

### Re-identification Model

We treat re-identification as a metric learning task. Our model maps each X-ray to a vector embedding, such that vectors from the same person are similar, and vectors from different people are dissimilar. For identification, the vector of the person to identify is calculated and matched to the closest vector from the archive. During inference, the model generates embeddings, not class labels for patients, which enables us to match the images of previously unseen patients.

Our primary model architecture is ResNet-50[34], pre-trained on ImageNet. Since ImageNet models are pre-trained on colored images, we modified the initial convolutional layer to use only one channel, the weights of which were defined as a sum of the weights of its three channels. Thus, the pre-trained filters were kept instead of being discarded. The classification layer was substituted with a linear transformation to a 512-dimensional space followed by batch normalization. We also repeated the entire experiment with DenseNet-121[35], adapted the same way.

The training used the ArcFace loss[36] with scale $s = 30$ and margin $m = 0.30$. ArcFace forces classes to be separated by a certain angle, which results in tighter clusters compared to a regular classification loss and is consistent with our cosine-similarity testing. Appendix S6 compares ArcFace and batch-hard triplet loss and shows a high correlation in regional $\Delta$ values ($r = 0.939$) with the same effect direction for seven out of ten regions. Training took place over 30 epochs with AdamW

© The National High School Journal of Science 2026

**Table 1** Study-design comparison of radiographic identification, privacy transformation and biometric retrieval. Cross-view means a fixed query view differs from the gallery view; region mask means a defined image area is removed or occluded; translated control means the identical mask is moved elsewhere; unseen IDs means test identities were absent from training. Yes — the study includes that feature. No — the study does not include it. NR — not reported or cannot be determined. N/A — not applicable to that study. The translated-control column identifies the methodological gap.

| Study | Data | Cross-view | Region mask | Translated control | Unseen IDs |
|---|---|---|---|---|---|
| **Radiographic identity** | | | | | |
| Packhäuser 2022 | ChestX-ray14 | No | No | No | Yes |
| Macpherson 2023 | Multi-site CXR | NR | No | No | Yes |
| Ueda 2023 | PadChest, CheXpert | Yes | No | No | Yes |
| Kim 2025 | Multi-site CXR | NR | No | No | Yes |
| Sony 2025 | NIH, MSUFAL | N/A | Yes | No | NR |
| Stephan 2011 | Forensic CXR | N/A | No | No | N/A |
| **Privacy transformation** | | | | | |
| Gichoya 2022 | Multi-site CXR | N/A | Yes | No | N/A |
| Packhäuser 2023 | ChestX-ray14 | No | No | No | NR |
| Li 2025 | MIMIC-CXR | N/A | No | No | NR |
| **Biometric retrieval** | | | | | |
| He 2021 | ReID benchmarks | Yes | No | No | Yes |
| Wang 2026 | ReID benchmarks | Yes | No | No | Yes |
| Wang 2022 | Occluded ReID | Yes | Yes | No | Yes |
| Present study | ChestX-ray14 | Yes | Yes | Yes | Yes |

(learning rate $3 \times 10^{-4}$, weight decay $10^{-4}$), batch size 64 and input image size of 224 $\times$ 224 normalized to $[-1, 1]$. The full list of hyperparameters and augmentation settings can be found in Appendix S2.

Horizontal flipping was not applied as augmentation since the chest does not possess left-right symmetry (with the heart being located on the left side). Thus, flipping would promote learning to ignore one of the sources of identification.

Three different random seeds were used for each architecture while keeping all other parameters equal. All experiments were run on a single NVIDIA A100 GPU.

### Retrieval Evaluation

Retrieval operates as follows. The query image is compared to a gallery of possible images, which is ranked in order of cosine similarity relative to the query. Our primary measure is precision@1: the proportion of queries for which the single nearest image in the gallery has the correct identity. Top-k metrics and mAP@R are in Appendix S4. In our reporting, we always include the size of the gallery since it is much harder to find an individual out of 4,611 candidates compared to 100[2].

The primary restriction we impose is that of closed gallery retrieval without having any identities seen during training: the entire set of 4,499 test AP images as queries against a gallery of 4,611 PA images from the 1,400 test patients. Since the query and the gallery images are never consistent in terms of the view, an accurate match cannot rest solely on view-specific information. Portable AP and standing PA X-rays differ systematically in geometry, exposure and equipment. This is a stronger constraint than retrieval that does not control for view, as in previous research.

We only have one small change to make in order to achieve proper same-view matching. In such cases, the query image is from the gallery and therefore the best match for the query image would obviously be the query image itself. We disregard such matches. However, there is a problem with this approach. In some instances, patients have only a single image which belongs to the same view as the query image, and by disregarding the self-match, we will not have any proper match at all for the query image. Thus, same-view results are used for those patients only if they have at least two images in the same view. Cross-view results are unaffected, because every patient in our set has both views.

### Anatomical Ablation

The regions were defined by the ChestX-Det PSPNet segmentation model available in the torchxrayvision package[37]. It generates 14 masks for each X-ray image. From these 14 we chose both clavicles, both scapulae, spine, both lung fields, heart, mediastinum, and diaphragm. The remaining four were deemed too small to be erased in a meaningful way or too anatomically ambiguous (see Appendix S3). These masks were computed once per test image and cached.

Pixels erased from the image become mid-gray 128. We


© The National High School Journal of Science 2026

did not use black because it sits at the extreme of the brightness range and creates a sharp artificial edge, which the model could learn instead of registering the missing information. Using mid-gray lessens this risk.

Severity refers to the proportion of a region that is removed, at 0.25, 0.50, 0.75, and 1.00, measured from the top of the region downward. One solid block is used instead of removing scattered pixels because the latter would retain the outline of the region and let the model fill in the gaps, testing robustness to noise instead of information loss.

Since we cut from the top downwards, severity is a proportion of a region's height and not its area, because regions have different widths as we move downwards. Therefore, the precise number of pixels removed in each condition is recorded. The control is matched on the mask itself and hence has an equal pixel count.

### Shape-matched random control

Each erasure is accompanied by a control that erases the identical mask in a randomly selected location. We determine the bounding box of the anatomical mask, pick a random top-left corner that ensures the entire box is within the $320 \times 320$ image, and apply this same mask to that location. The erasures are done on the query images at their stored $320 \times 320$ resolution before being down-scaled to $224 \times 224$ for the model. The gallery images are not masked. The control location is selected independently of the anatomical masks and may intersect them. Thus, the control removes exactly the same number of pixels in the same shape but at a different location and over different image content.

Our previous protocol relied on a square of the same area as the mask instead of a re-positioned mask. The square matches the region's area only approximately and not its shape, and the shift to the precise mask changed the signs of five out of the ten effects, so we report the shape-matched control throughout.

The location of the control is re-calculated separately for each of the 4,499 query images rather than one calculation for each region, so that each value of $\Delta$ represents an average of 4,499 independent placements. Each seed uses independent random sequences.

Our primary quantity is

$$\Delta = \mathrm{p@1}_{\text{anatomical}} - \mathrm{p@1}_{\text{shape-matched}} \tag{1}$$

A negative value for $\Delta$ indicates that masking the region in place hurt retrieval more than masking the same shape at a randomly chosen location. A positive $\Delta$ means it hurt less. The anatomical masks are in row (a) of Figure 1, with their shape-matched controls in row (b), across the four severity levels.

The type of control used in this study was unconfined random placement (Anywhere), where the mask was placed inside the $320 \times 320$ image. In addition to this type of control, four different control types were considered. Thorax confined the mask within the bounding box of the segmented anatomy, Same half confined the mask on the same side of the region centroid, and Band confined the mask center within $\pm 40$ pixels from the region centroid in the vertical direction. Low overlap attempted to constrain the anatomical overlap under 20% but achieved overlaps of 38.3% and 39.0% for the left and right lung fields, respectively. Only 2.0% of Same half placements of the diaphragm satisfied the constraint. The remaining placements were unconstrained.

The cumulative masking was done with ResNet-50 using three seeds in descending order of precision@1 loss. No modifications were made to the PA gallery. The control was performed using the mask of the combined regions moved to another location within the image. However, the larger the mask was, the fewer places there were to move it to. Furthermore, we tried all 45 possible pairings of the regions, taking the interaction to be the difference between the precision@1 loss for the combined regions and the sum of the losses of the individual regions. For two lungs, we calculated an independence approximation by multiplication of the individual precision@1 scores divided by the whole baseline. We also experimented with ResNet-50 (seed 0) with a cap of five images per patient without any other changes in our training and ablation protocol, on the original test set as well as on the separately capped test set (Appendix S10).

### Statistical analysis

We report $\Delta$ as the mean and standard deviation across three seeds.

The reason for the uncertainty calculation being a bootstrap sampling procedure of patients and not individual queries is that out of our total of 4,499 queries, there are only 1,400 patients with 3.21 images per patient on average. What this means is that queries from the same patients cannot be treated as independent queries since the same anatomy and body shape are measured from these images and, more importantly, since the images are captured from the same device, if one image from a patient matches, then probably the others also match. Treating these as independent queries would make our confidence intervals narrower than they should be. We draw 1,400 patients with replacement, include every query from each drawn patient, and repeat this 20,000 times.

We do multiple comparisons between ten regions and their respective controls. Therefore, we need to make adjustments to our results. We give Bonferroni-corrected 99.5% confidence intervals aimed at achieving a family-wise error rate of 5% over the ten tests. Regions that have an adjusted confidence interval not containing zero will be considered different from the controls. The uncorrected 95% confidence intervals

© The National High School Journal of Science 2026

are given beside the adjusted confidence intervals in Tables S6 and S7 (see Appendix S5). The confidence intervals obtained via the percentile method suffer from loss of accuracy when calculating extreme quantiles. Therefore, the 99.5% intervals should be considered conservative bounds rather than exact coverage. The number of resamples increased from 4,000 to 20,000. This caused changes of less than 0.001 in the ends of confidence intervals without changing any conclusions.

Patient clustering rather than query clustering broadens the unadjusted 95% confidence intervals by 1.01–1.26×, with more widening for the largest effects, as expected when the variance among patients outweighs the within-patient variance. The interval broadening is less than 3.21 queries per patient would lead one to expect, implying that repeated images of the same individual contain less redundant information as retrieval queries than one would think. We report only patient-clustered intervals because the patient is the appropriate unit of analysis.

Bootstrapped confidence intervals and seed-level standard deviations measure different kinds of variability and are thus reported separately. The bootstrapped confidence intervals for complete regional ablation are derived from the per-query results of one particular model (seed 0). They quantify the variability due to sampling of test patients, not variability due to training new models or redrawing the random control masks. Thus the bootstrapped confidence interval may be narrower than the seed-level standard deviation without any inconsistency. We retained only the per-query results of seed 0, and thus did not calculate the bootstrapped confidence intervals for the other seeds.

Seed-level t-tests are not provided because with three seeds they have only two degrees of freedom and are not informative.

Across the evaluation sessions, we noted a difference of 0.0011 for the precision@1 metric and 0.0016 for $\Delta$. This occurred due to cuDNN algorithm selection. $\pm 0.002$ was chosen as an approximate resolution and 0.005 as a threshold for determining the direction of the effect. Segmentation produced empty masks in about 1.2% of cases, including some left- and right-clavicle masks. These queries did not change in either condition, possibly attenuating $\Delta$ towards zero.

## Results

### Baseline re-identification

The precision@1 for same-view retrieval of the PA images in a gallery of 4,242 (ResNet-50, seed 0) amounts to 0.9934, which is comparable to the 0.976 achieved by Packhäuser et al. (2022)[1] at the same image size. The patient set, gallery and training setup of these authors differ from ours. The most stringent cross-view setting achieves 0.9500 precision@1 (AP queries in a PA gallery of 4,611) for 1,400 previously unseen patients. Large cross-view accuracy implies that the identity information is not destroyed by large differences in view position, so that the identity signature is unlikely to be dominated by view-specific artifacts. It does not rule out features that persist across different views, such as those associated with a certain scanning site, scanner or positioning protocol. The results for all four view combinations can be found in Table S5 (Appendix S4).

### Reported baselines

The 0.9500 baseline is computed by scoring the saved image files directly. The ablation experiments read images along a different resize path and give a seed-0 baseline of 0.9458 for ResNet-50; all of the ablation results, $\Delta$ values and intervals in this paper use that resize path. Using three different seeds, the ablation-path baseline is 0.929–0.946 for ResNet-50 and 0.955–0.966 for DenseNet-121.

### Patient matching persists after single-region masking

Masking out an entire lung field, the worst case among all single-region conditions tested, reduces the precision@1 from 0.9458 to 0.5817 when compared against a gallery of 4,611 images (ResNet-50, seed 0). The gallery consists of 4,611 images from 1,400 patients, giving roughly 3.1 correct matches per query, meaning the guessing baseline is approximately $7 \times 10^{-4}$. Even with roughly 15% of the image erased, the model is still able to find the correct patient more than 800 times better than guessing. Masking out any individual bone region yields precision@1 greater than 0.78.

Accuracy decreases gradually; it is not abrupt (Figure 2). No sudden decrease in retrieval was observed for any of the four severity levels tested.

Ranked retrieval is much more robust than the results on precision@1 imply. Following the removal of a whole lung field, the correct patient remains among the top 20 of 4,611 candidates in 87.0% of queries (left lung) and 87.5% (right lung). For all other regions, top-20 retrieval remains above 94%: 94.6% after the removal of the spine, 95.9% after the mediastinum, and above 98% for either clavicle (Table 2). In the absence of any masking, top-20 retrieval is 98.6%. Thus, removal of a whole lung field incurs a cost of 11.7 percentage points at rank 20 but 36 percentage points at rank 1. Figure 3 depicts the effects of masking on retrieval performance across ranks.

At complete ablation, the median rank value was 1 for all regions when using ResNet-50 (seed 0), even for either lung field. The mean rank value went up from 4.93 when using intact images to 33.12 when using left lung masks and to 30.71 when using right lung masks. This suggests that masking sub-


© The National High School Journal of Science 2026

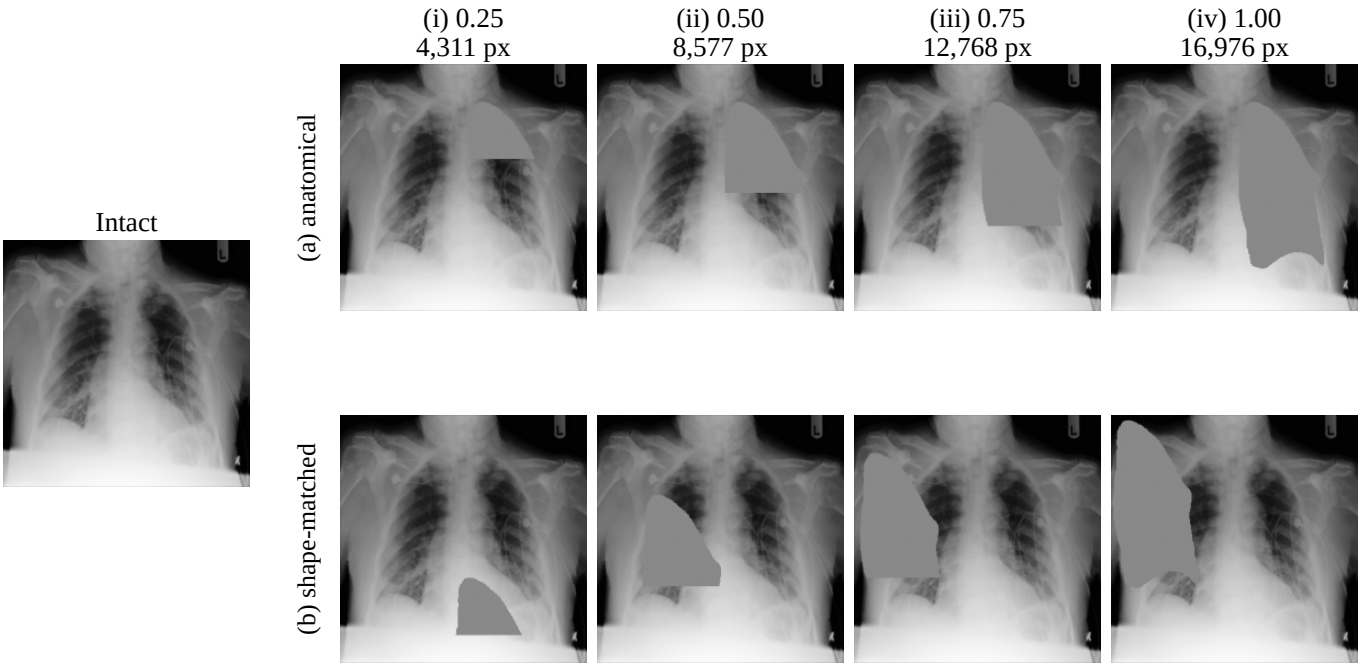



**Figure. 1** Graded masking of the left lung field. The intact image is shown on the left. (a) Anatomical ablation and (b) shape-matched controls. Columns (i)–(iv) show masking at 0.25, 0.50, 0.75, and 1.00 severity, respectively.

stantially worsened retrieval for some queries while leaving the typical match at rank 1.

### Lung-field and mediastinum masking results in larger drops than the matched control

Only three areas showed significantly larger reduction in retrieval performance compared to their shape-matched controls in both architectures. At seed 0, these three areas are the left lung field ($\Delta = -0.079$, Bonferroni-corrected 99.5% CI $[-0.106, -0.052]$ for ResNet-50 and $-0.134$, $[-0.161, -0.108]$ for DenseNet-121), right lung field ($-0.091$, $[-0.119, -0.064]$ and $-0.025$, $[-0.0484, -0.0004]$), and mediastinum ($-0.045$, $[-0.062, -0.028]$ and $-0.034$, $[-0.050, -0.019]$).

There were two regions which cost significantly less under both architectures. Removal of the heart region cost less than random placement of the equivalent mask ($+0.052$, $[+0.035, +0.069]$ and $+0.056$, $[+0.042, +0.070]$), and similarly, removal of the diaphragm region cost less than random placement of the equivalent mask ($+0.284$, $[+0.260, +0.307]$ and $+0.245$, $[+0.224, +0.266]$). In both cases, the removal in its specific location was less costly than random placement.

Neither clavicle nor either scapula shows significance under both architectures post-correction, but the spine region shows significance under DenseNet-121 only, in the positive direction.

**Full intervals are in Appendix S5**

### Positional effects dominate for peripheral structures

The diaphragm provides the largest effect in the study and in the reverse direction of the lungs: $\Delta = +0.284$ when using ResNet-50 and $+0.245$ for DenseNet-121 (seed 0). Removing the entire diaphragm brings the precision@1 to 0.921, while removing a region of the same shape somewhere else reduces it down to about 0.64.

### Location of the control mask and anatomical coverage

The random placement of control masks frequently included other anatomical regions apart from the background space. In particular, controls deleted 75.7%–87.1% as much anatomical structure as their corresponding regional masks (Table 3). To investigate the importance of control placement, we examined five different rules: anywhere in the image, inside the thorax, same lateral half, along a horizontal strip at the same height, or low anatomical overlap (Table 4).

Applying the low-overlap rule to larger masks proved to be quite challenging. While our initial goal was to have no more than 20% anatomical overlap, in reality, there was 38.3% overlap on the left lung, 39.0% overlap on the right lung, and 52.5% overlap on the diaphragm (seed 0). It simply was not possible to manipulate such large masks without overlapping them on any other structure (see Figure 4). As a result, this

© The National High School Journal of Science 2026

**Table 2** Complete-region top-k retrieval (seed 0). For both architectures, the results were calculated using AP queries to a 4,611-image PA gallery. Each row fully masks the designated region. The mean rank is given only for ResNet-50 since ranks for individual queries were stored for this network only.

| Region ablated | ResNet-50 p@1 | ResNet-50 top-5 | ResNet-50 top-20 | ResNet-50 mean rank | DenseNet-121 p@1 | DenseNet-121 top-5 | DenseNet-121 top-20 |
|---|---|---|---|---|---|---|---|
| None (intact) | 0.9458 | 0.9747 | 0.9862 | 4.93 | 0.9549 | 0.9780 | 0.9884 |
| Left lung field | 0.5817 | 0.7571 | 0.8695 | 33.12 | 0.5655 | 0.7426 | 0.8742 |
| Right lung field | 0.5824 | 0.7597 | 0.8753 | 30.71 | 0.6993 | 0.8266 | 0.9109 |
| Spine | 0.7884 | 0.8949 | 0.9464 | 13.68 | 0.8651 | 0.9364 | 0.9671 |
| Right scapula | 0.8195 | 0.9126 | 0.9595 | 7.95 | 0.8760 | 0.9400 | 0.9720 |
| Left scapula | 0.8331 | 0.9206 | 0.9653 | 8.12 | 0.8680 | 0.9360 | 0.9664 |
| Mediastinum | 0.8360 | 0.9155 | 0.9589 | 10.26 | 0.8722 | 0.9344 | 0.9667 |
| Heart | 0.8791 | 0.9391 | 0.9707 | 7.72 | 0.9204 | 0.9600 | 0.9787 |
| Diaphragmatic surface | 0.9211 | 0.9647 | 0.9827 | 5.97 | 0.9400 | 0.9684 | 0.9829 |
| Right clavicle | 0.9244 | 0.9673 | 0.9831 | 5.26 | 0.9467 | 0.9742 | 0.9871 |
| Left clavicle | 0.9315 | 0.9664 | 0.9831 | 5.07 | 0.9462 | 0.9751 | 0.9862 |

rule could not provide a clean background control for those regions.

Effects that were distinguishable from numerical drift maintained their direction under the first four placement rules, while under the low-overlap rule, only five regions exhibited such behavior. In addition, the spatial analysis presented in Appendix S8 indicates that masking led to greater retrieval losses in the upper central thorax region rather than around the edge of any particular body part (Figure S2). This suggests the importance of mask placement, although this experiment is not capable of dissociating this aspect from which body part is removed.

### Regional effects largely replicate across architectures

DenseNet-121 has a higher unablated baseline than ResNet-50 (0.955–0.966 against 0.929–0.946 across seeds).

Out of ten regions, nine have the same sign regardless of which architecture is used, and all five regions that are significant under both architectures after correction have the same sign in both: both lung fields and the mediastinum are negative, and the heart and diaphragm positive. Largest and smallest effects show up in the same regions: the diaphragm is the largest and the clavicles and spine the smallest (Table 5, Figure 5).

Sizes do differ. On average across seeds, the right lung field is $-0.075$ under ResNet-50 but $-0.031$ under DenseNet-121, and the left lung field is $-0.077$ versus $-0.117$, so the two models weigh the two lungs differently but agree on direction. The only sign disagreement is the spine: at seed 0, negative and not significant under ResNet-50 ($-0.017$, $[-0.037, 0.002]$), significantly positive under DenseNet-121 ($+0.022$, $[+0.006, +0.038]$).

### Cumulative multi-region ablation and pairwise interactions

For ResNet-50, the mean precision@1 score with three seeds declined from 0.9386 in the intact images to 0.5475 for the masking of the left lung field and 0.0204 for the masking of both lung fields. For the masking of all ten regions, the mean precision@1 score fell to 0.0019 (Figure 6). The change in retrieval from a displacement of a $40 \times 40$ pixel block in the image with seed 0 is shown as well (Appendix S8). At a precision@1 score of 0.0019, the performance was about 2.7 times what would be expected by chance at 0.0007. Thus, the retrieval was close to chance, but it does not prove complete anonymization.

The pairwise interaction was defined as the difference between the combined loss in precision@1 score and the sum of the individual losses in precision@1 score. On average, the 45 pairwise interactions resulted in an interaction value of $+0.0812$, and only seven of them were sub-additive. All four same-side interactions between clavicle/scapula and lung fields resulted in a negative interaction (between $-0.0831$ and $-0.0029$), whereas all four opposite-side interactions yielded positive interaction values (between $+0.1131$ and $+0.3465$). Appendix S9 provides all 45 pairwise interaction values and shows the regional interaction patterns (Figure S3).

Because the additivity model is only a descriptive one for precision@1, an independence approximation for the two lung fields was considered as well, with an expectation of precision@1 of 0.32 versus the observed precision@1 of 0.020. The laterality pattern could be caused by an overlap of


© The National High School Journal of Science 2026

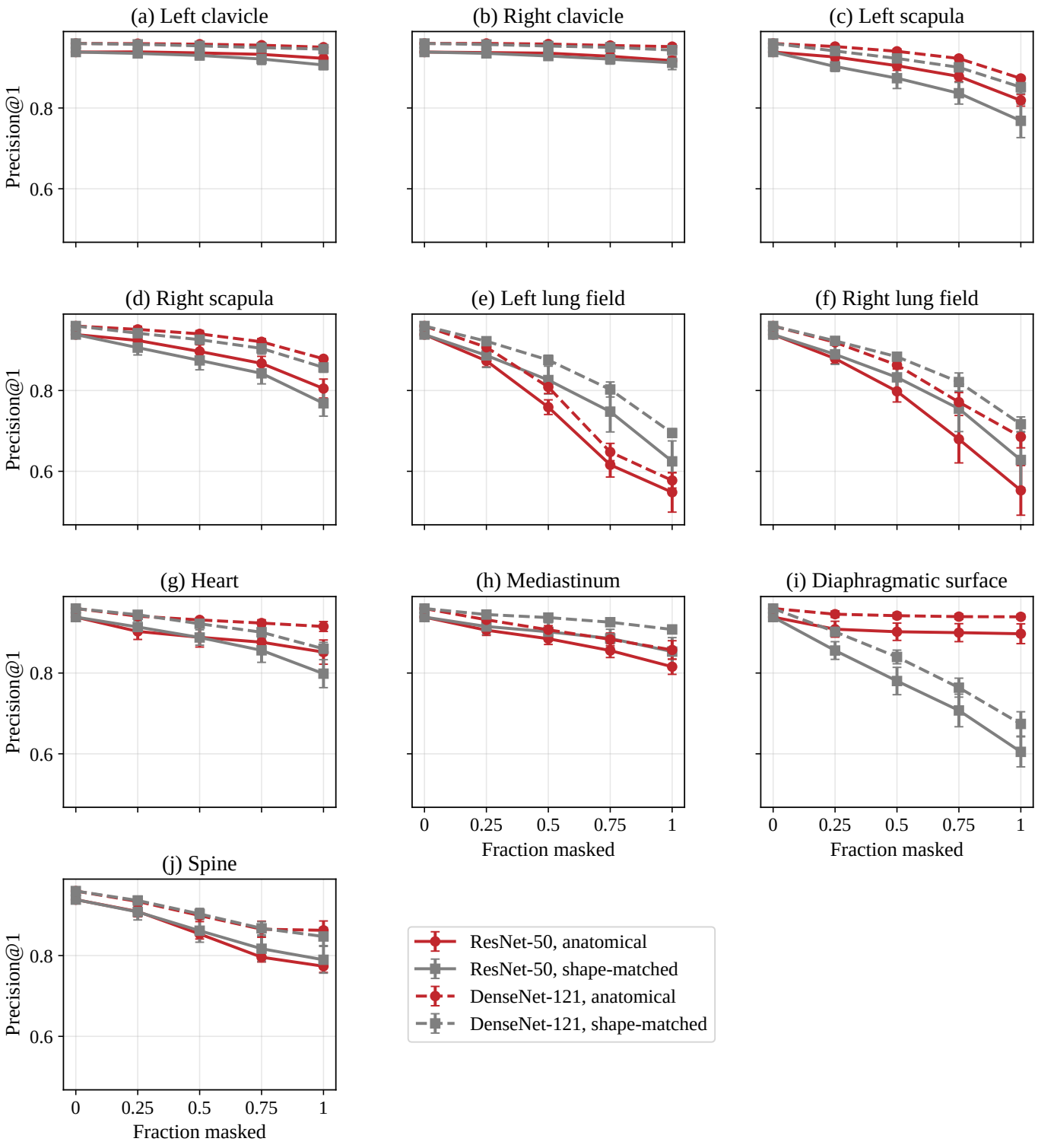



**Figure. 2** Precision@1 with respect to the masked fraction of each anatomical region and its shape-matched control. Circles represent anatomically masked regions and squares represent shape-matched controls; solid lines correspond to ResNet-50, and dashed lines correspond to DenseNet-121. The error bars correspond to the standard deviation for three different random seeds. At full masking, lung fields and mediastinum suffer from higher retrieval losses than their shape-matched controls under both architectures.

© The National High School Journal of Science 2026

**Table 3** Anatomy stripped out by translated control masks. Columns indicate the original mask size, the number of control-mask pixels covering the segmented anatomy, and the proportion (Share) for each region, averaged across 4,499 queries. Control masks strip out 75.7% to 87.1% as much segmented anatomy as the anatomical masks, rather than falling mostly on background.

| Region | Region mask | On anatomy | Share |
|---|---|---|---|
| Left lung field | 15,770 | 13,516 | 0.857 |
| Right lung field | 14,847 | 12,818 | 0.863 |
| Mediastinum | 5,061 | 4,108 | 0.812 |
| Spine | 9,331 | 7,065 | 0.757 |
| Right clavicle | 1,354 | 1,123 | 0.830 |
| Left clavicle | 1,459 | 1,207 | 0.828 |
| Right scapula | 6,477 | 5,642 | 0.871 |
| Left scapula | 6,818 | 5,879 | 0.862 |
| Heart | 7,993 | 6,924 | 0.866 |
| Diaphragmatic surface | 16,242 | 13,829 | 0.851 |

**Table 4** Sensitivity of regional effects to control positioning. The numbers indicate the average Δ over three seeds with full ablation. Anywhere: no restriction of placement in the image. Thorax: within the bounding box of the segmented-anatomy mask. Same half: in the same half (left or right) as the region's center. Band: control-mask center placed within ±40 pixels of the region's center vertically. Low overlap: least amount of overlap with the segmented anatomy. Numbers smaller than 0.005 in absolute value are undetermined (n.d.). † Direction of Δ does not change across any of the rules with a definite effect. ‡ Only 2.0% of placements satisfied the diaphragm same-half rule, hence this cannot be considered an independent check.

| Region | Anywhere | Thorax | Same half | Band | Low overlap |
|---|---|---|---|---|---|
| Left lung field† | –0.0771 | –0.0481 | –0.1285 | –0.0450 | –0.2949 |
| Right lung field† | –0.0758 | –0.0432 | –0.1297 | –0.0367 | –0.2881 |
| Mediastinum† | –0.0379 | –0.0244 | –0.0336 | –0.0105 | –0.1075 |
| Spine† | –0.0156 | n.d. | –0.0153 | –0.0148 | –0.1543 |
| Right clavicle | +0.0054 | +0.0113 | n.d. | +0.0087 | –0.0135 |
| Left clavicle | +0.0167 | +0.0196 | +0.0093 | +0.0133 | –0.0079 |
| Right scapula | +0.0360 | +0.0540 | +0.0052 | +0.0472 | –0.0755 |
| Left scapula | +0.0507 | +0.0724 | +0.0156 | +0.0588 | –0.0543 |
| Heart | +0.0530 | +0.0695 | +0.0237 | +0.0242 | –0.0362 |
| Diaphragmatic surface† | +0.2919 | +0.3245 | +0.2918‡ | +0.0405 | +0.1420 |

anatomy in projectional radiography, but this is not confirmed experimentally.

## Discussion

Patient matching remained possible after masking each of the ten anatomical regions we tested. Erasing an entire lung field, the most detrimental single-region condition we tested, drops precision@1 from 0.946 to 0.582, yet the correct patient is still in the top 20 of 4,611 candidates for 87% of queries. Our results suggest that masking a single anatomical region should not, by itself, be treated as proof that an image is anonymous.

The signal is unevenly distributed, but not in the way a landmark-based view would predict. Only masking the lung fields and mediastinum caused significantly larger retrieval losses than their shape-matched controls under both architectures. The clavicles and scapulae did not show a consistently significant difference from their controls across both models, the spine behaved differently between models, and masking the heart hurt retrieval less than its control. This is worth pausing on, because forensic radiographic comparison leans heavily on clavicle and vertebral shape[3], so we expected the


© The National High School Journal of Science 2026

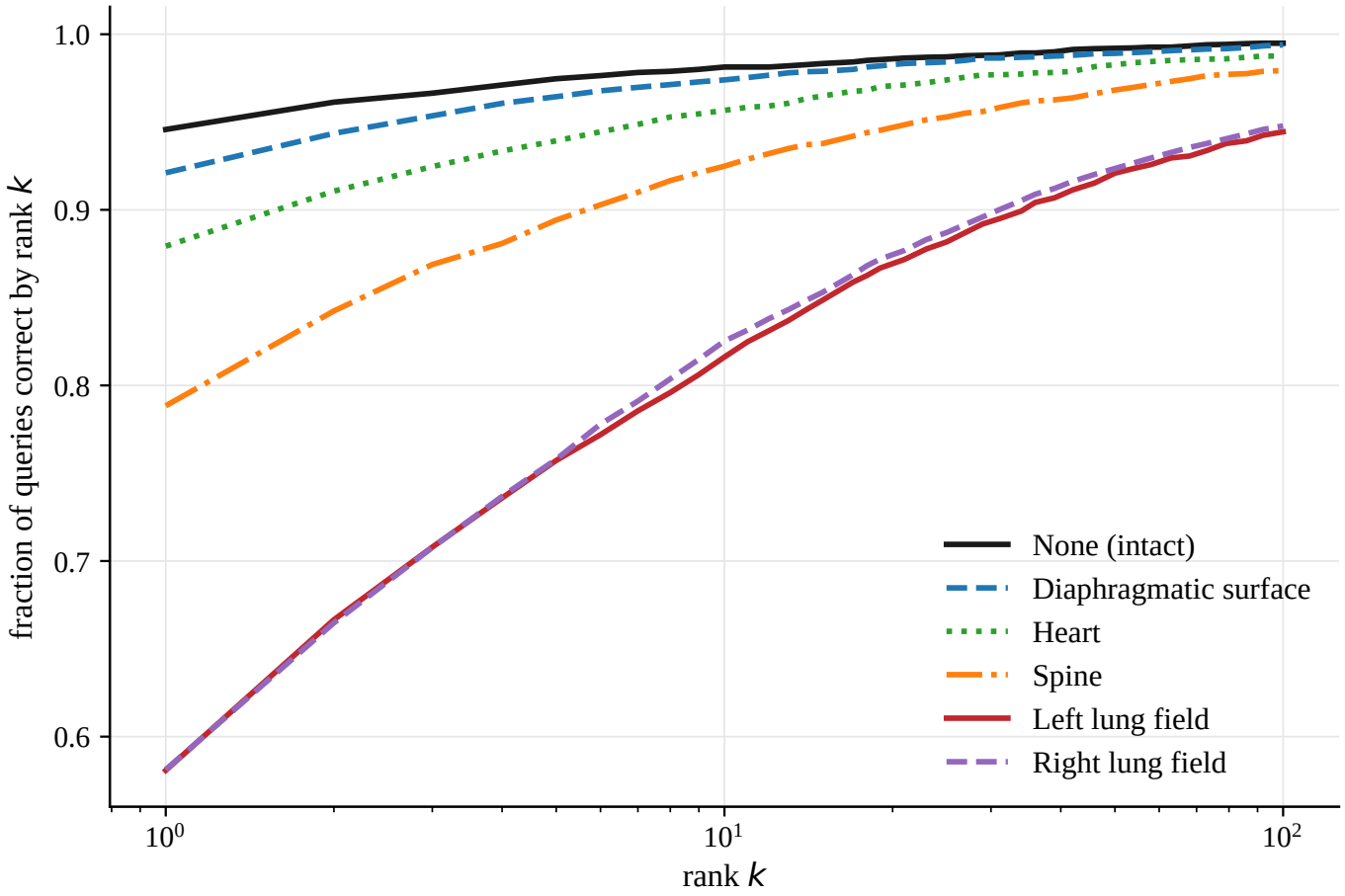



**Figure. 3** Rank-based retrieval following complete masking (ResNet-50, seed 0). Rank k is on the horizontal axis, and the fraction of queries with a correct match by rank k is provided on the vertical axis. Changes beyond the top-1 decision are illustrated with the curves. The "intact" case and five representative regional masks are included for illustrative purposes.

**Table 5** Δ (Equation 1) of precision@1 for complete ablation of the region compared to shape-matched controls, mean ± SD from three random seeds per architecture. Positive values denote increased precision@1 with the masking of the region compared to the placement of the same shape at other locations, whereas negative values denote reduced precision@1. Out of ten regions considered, only the spine fails to satisfy the uniformity of signs between architectures.

| Region | ResNet-50 | DenseNet-121 | Sign agrees |
|---|---|---|---|
| Left lung field | $-0.077 \pm 0.002$ | $-0.117 \pm 0.016$ | yes |
| Right lung field | $-0.075 \pm 0.014$ | $-0.031 \pm 0.009$ | yes |
| Mediastinum | $-0.038 \pm 0.019$ | $-0.051 \pm 0.017$ | yes |
| Spine | $-0.016 \pm 0.022$ | $+0.015 \pm 0.006$ | no |
| Right clavicle | $+0.005 \pm 0.008$ | $+0.009 \pm 0.003$ | yes |
| Left clavicle | $+0.016 \pm 0.001$ | $+0.005 \pm 0.002$ | yes |
| Right scapula | $+0.037 \pm 0.016$ | $+0.021 \pm 0.012$ | yes |
| Left scapula | $+0.051 \pm 0.028$ | $+0.022 \pm 0.015$ | yes |
| Heart | $+0.053 \pm 0.006$ | $+0.055 \pm 0.004$ | yes |
| Diaphragmatic surface | $+0.292 \pm 0.013$ | $+0.265 \pm 0.025$ | yes |

© The National High School Journal of Science 2026

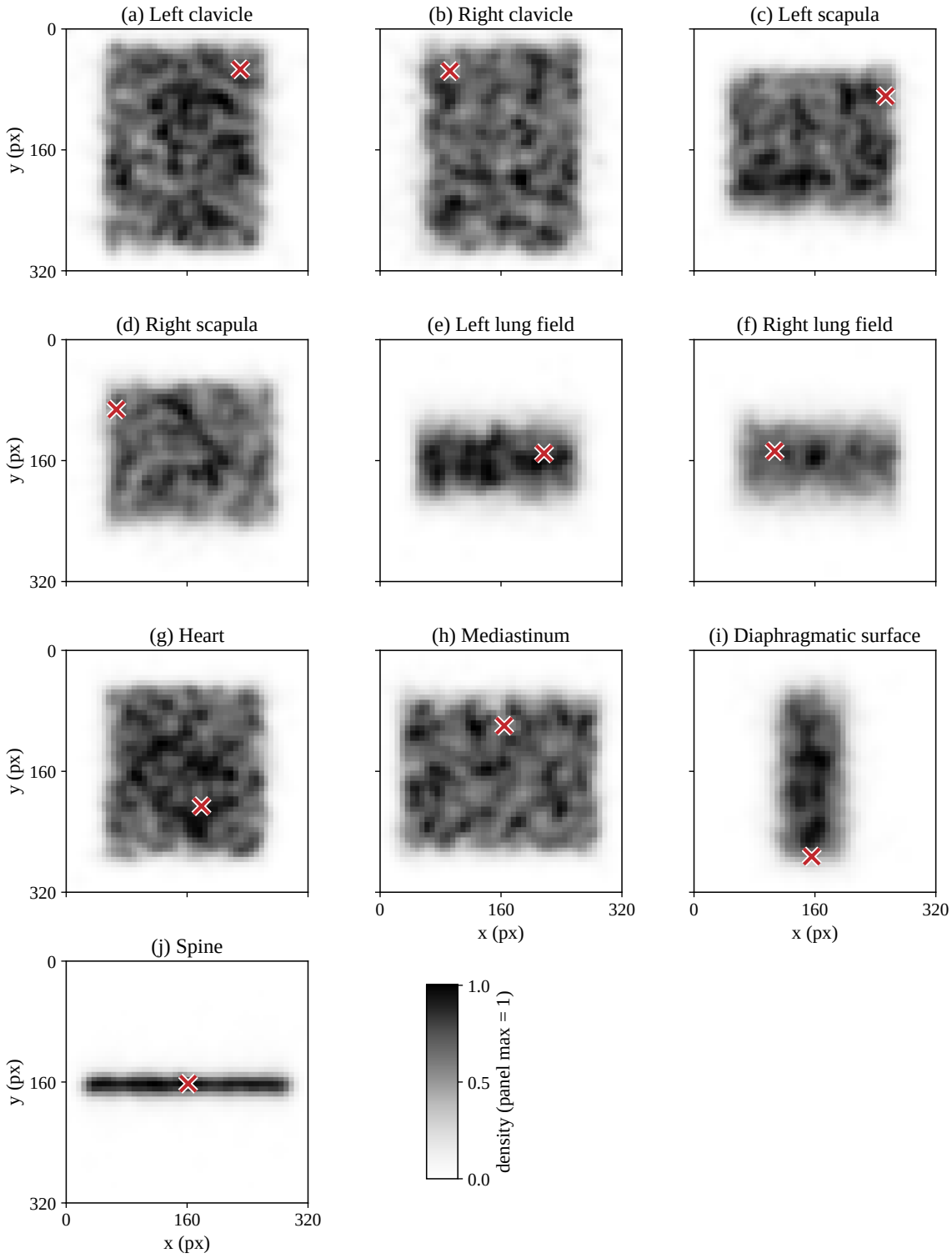



**Figure. 4** Distribution of locations of the control masks. Panels show the smoothed density of placement based on 4,499 queries; the red cross denotes the mean location of the anatomical mask to illustrate how the locations of the control masks diverge from the location of the original area. Pixels within images define identical coordinate systems for all panels; individual normalizations of shade are done per panel, so darkness should not be compared between regions.


© The National High School Journal of Science 2026

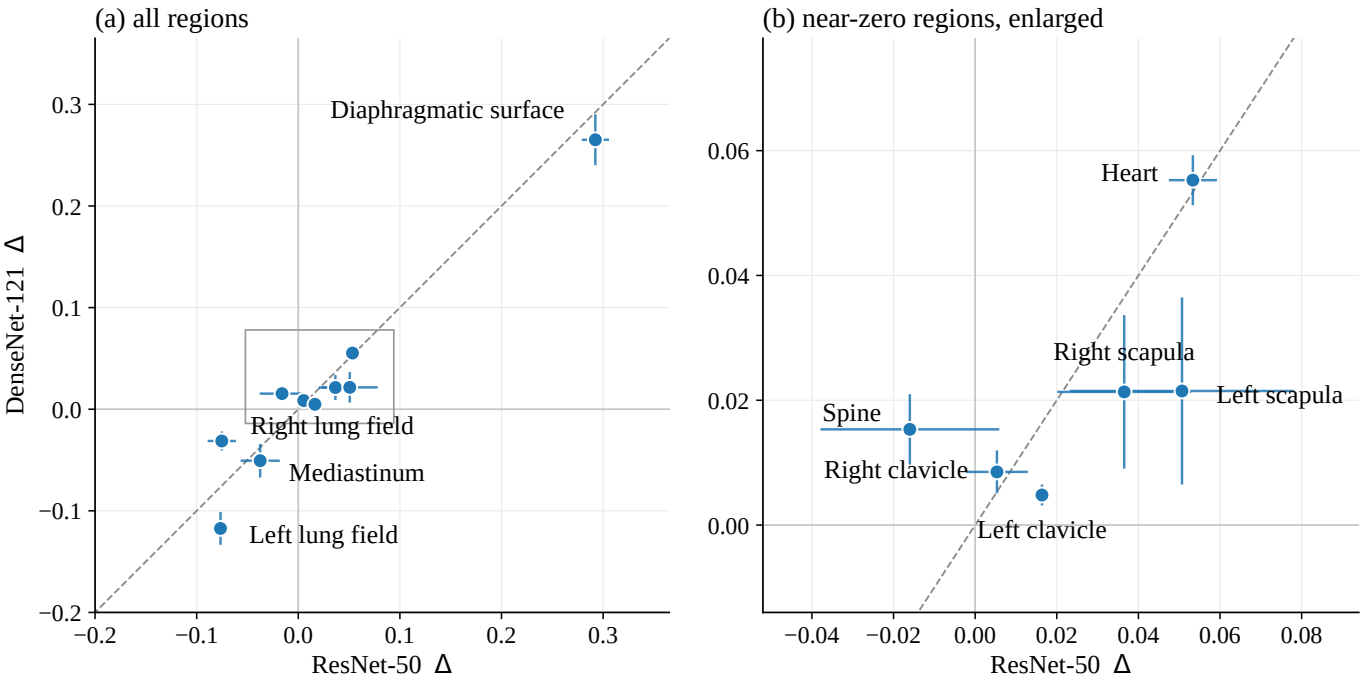



**Figure. 5** Δ at complete ablation for ResNet-50 versus DenseNet-121, where the standard deviation has been calculated using three random seeds, and the dashed line represents perfect match. While panel (a) shows all regions, panel (b) represents the close-to-zero values marked in the gray box in panel (a). From the ten regions, there are nine regions that show consistent direction of influence, and the maximum and minimum influences fall in the same regions (the diaphragm largest, the clavicles and spine smallest). Only the spine region is inconsistent in direction.

clavicles and spine to matter most for re-identification. Instead, once mask shape and size are controlled for, masking the lung fields and mediastinum causes the largest additional drops in retrieval. That does not mean lung tissue itself is uniquely identifying, though. The lung-field masks also cover the ribs that lie over the lungs, so soft tissue and bone are mixed together in that result.

The regional pattern is similar across the two models, but not identical. Nine of ten regions have the same sign under both architectures, and all five regions significant under both agree. The exceptions are informative: the two models weight the left and right lungs differently, and the spine is negative under ResNet-50 but significantly positive under DenseNet-121. The agreement suggests the broad pattern is not a quirk of one network. The disagreements suggest that testing against a single model may not be enough to validate anonymization. A transformation that hides the features one model uses may leave the features another model uses intact. Validation should therefore use several recognition architectures rather than treating one model's failure as proof that an image is anonymous.

We did not test any real de-identification tool. Filling a region with uniform gray removes its original pixel values but leaves the mask boundary visible, whereas other methods blur, pixelate, crop, or synthesize[19]. Our results show that patient matching can persist after the single-region masks we tested; they do not measure what any particular tool does achieve.

## Limitations

The computational variability makes it difficult to interpret effects near zero. The empty segmentation masks could have reduced the effects of smaller masks. Single seed comparison with five images per patient cap resulted in preservation of regional effect direction but a reversal of the left lung effect (Appendix S10). Lung-field masks include the ribs lying over the lungs, so the lung-field effect cannot be attributed to soft tissue rather than bone. Separating their contributions would require a method that distinguishes bone from soft tissue, which our experiment does not provide. All results come from one dataset collected at one institution. We have not tested whether they generalize to other scanners, protocols, or patient populations. We tested two convolutional architectures; transformer-based models have different inductive biases and may weight regions differently, and the disagreement we already see between two CNNs suggests this should be tested rather than assumed. Severity is a fraction of a region's

© The National High School Journal of Science 2026

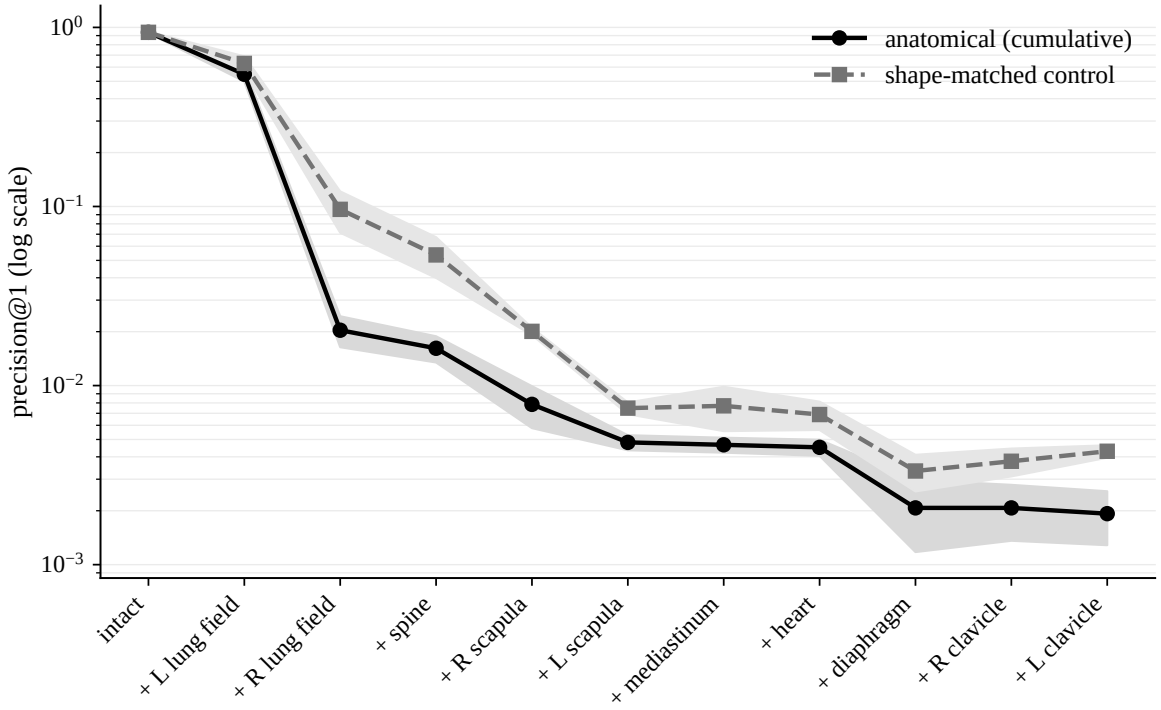



**Figure. 6** Cumulative multi-region ablation (ResNet-50). Regions are considered in the order of their diminishing individual damage to the model's performance, whereas the control shifts the same combined mask randomly. The plots show the mean for three different seeds, and shaded bands are one standard deviation. Precision@1 is shown on a logarithmic scale. Starting from about 0.002, the tail is constant within the measurement resolution (one correct answer out of 4,499 queries changes precision@1 by about 0.0002), so small fluctuations are meaningless.

height, not its area, because the cut is made along the vertical axis. Controls are matched on the mask itself so this does not affect $\Delta$, but the severity axis in Figure 2 is not linear in area. Images were stored as JPEG at quality 90; a small comparison against lossless inputs found identical retrieval (Appendix S7), but the sample is small and the test correspondingly weak. We erase only the query and leave the gallery intact, which matches the scenario we care about but leaves the case where both sides are masked untested. Finally, the shape-matched control changes location while holding shape and pixel count fixed, so a positive $\Delta$ means masking that region hurts retrieval less than masking the same shape at the sampled random positions. It does not separate position from anatomy: a structure that sits at the edge because of where it is in the body, like the diaphragm, cannot be tested in the center while still being that structure. Our claims are therefore about the identity signal available to these models at a region's location and extent, not about what the anatomy itself contains.

### Closing thought

Masking any one of the ten regions we tested did not prevent patient matching, and the regions traditionally treated as identifying landmarks did not produce the largest drops relative to shape-matched controls in our tests. Together, this argues for judging de-identification by the whole released image rather than by a checklist of structures to hide.

## Acknowledgments

The author thanks Laurin Rupp for extensive comments on drafts of this manuscript, including critiques of the methods and suggestions for revision. Claude (Anthropic) was used to rebuild tables and to help with LaTeX formatting. It was not used to draft the manuscript.

## References

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

2 M. S. Macpherson, C. E. Hutchinson, C. Horst, V. Goh, G. Montana, Patient reidentification from chest radiographs: an interpretable deep metric learning approach and its applications. Radiology: Artificial Intelligence. Vol. 5, pg. e230019, 2023., https://doi.org/10.1148/ryai.230019.


© The National High School Journal of Science 2026

3 C. N. Stephan, A. P. Winburn, A. F. Christensen, A. J. Tyrrell, Skeletal identification by radiographic comparison: blind tests of a morphoscopic method using antemortem chest radiographs. Journal of Forensic Sciences. Vol. 56, pg. 320–332, 2011., https://doi.org/10.1111/j.1556-4029.2010.01673.x.

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

5 K. Kim, K. Cho, Y. Eo, J. Kim, J. Yun, Y. Ahn, J. B. Seo, G.-S. Hong, N. Kim, Screening patient misidentification errors using a deep learning model of chest radiography: a seven reader study. Journal of Imaging Informatics in Medicine. Vol. 38, pg. 694–702, 2025., https://doi.org/10.1007/s10278-024-01245-0.

6 R. Sony, C. V. Isaac, A. VanBaarle, C. J. Devota, T. Fenton, J. T. Hefner, A. Ross, Automatic comparative chest radiography using deep neural networks. IEEE Access. Vol. 13, pg. 4398–4410, 2025., https://doi.org/10.1109/ACCESS.2025.3525579.

7 J. W. Gichoya, I. Banerjee, A. R. Bhimireddy, J. L. Burns, L. A. Celi, L.-C. Chen, R. Correa, N. Dullerud, M. Ghassemi, S.-C. Huang, P.-C. Kuo, M. P. Lungren, L. J. Palmer, B. J. Price, S. Purkayastha, A. T. Pyrros, L. Oakden-Rayner, C. Okechukwu, L. Seyyed-Kalantari, H. Trivedi, R. Wang, Z. Zaiman, H. Zhang, AI recognition of patient race in medical imaging: a modelling study. The Lancet Digital Health. Vol. 4, pg. e406–e414, 2022., https://doi.org/10.1016/S2589-7500(22)00063-2.

8 J. Ganz, J. Ammeling, S. Jabari, K. Breininger, M. Aubreville, Re-identification from histopathology images. Medical Image Analysis. Vol. 99, article 103335, 2025., https://doi.org/10.1016/j.media.2024.103335.

9 A. Enomoto, A.-D. Lee, M. Sukedai, T. Shimoide, R. Katada, K. Sugimoto, H. Matsumoto, Automatic identification of individuals using deep learning method on panoramic radiographs. Journal of Dental Sciences. Vol. 18, pg. 696–701, 2023., https://doi.org/10.1016/j.jds.2022.10.021.

10 M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks. European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science. Vol. 8689, pg. 818–833, 2014., https://doi.org/10.1007/978-3-319-10590-1_53.

11 R. C. Fong, A. Vedaldi, Interpretable explanations of black boxes by meaningful perturbation. ICCV, 2017., https://doi.org/10.1109/ICCV.2017.371.

12 R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization. ICCV, 2017., https://doi.org/10.1109/ICCV.2017.74.

13 U.S. Department of Health and Human Services, Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. 2012., https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

14 C. G. Schwarz, W. K. Kremers, T. M. Therneau, R. R. Sharp, J. L. Gunter, P. Vemuri, A. Arani, A. J. Spychalla, K. Kantarci, D. S. Knopman, R. C. Petersen, C. R. Jack Jr., Identification of anonymous MRI research participants with face-recognition software. New England Journal of Medicine. Vol. 381, pg. 1684–1686, 2019., https://doi.org/10.1056/NEJMc1908881.

15 C. G. Schwarz, W. K. Kremers, V. J. Lowe, M. Savvides, J. L. Gunter, M. L. Senjem, P. Vemuri, K. Kantarci, D. S. Knopman, R. C. Petersen, C. R. Jack Jr., Alzheimer's Disease Neuroimaging Initiative, Face recognition from research brain PET: an unexpected PET problem. NeuroImage. Vol. 258, pg. 119357, 2022., https://doi.org/10.1016/j.neuroimage.2022.119357.

16 A. de Sitter, M. Visser, I. Brouwer, K. S. Cover, R. A. van Schijndel, R. S. Eijgelaar, D. M. J. Müller, S. Ropele, L. Kappos, Á. Rovira, M. Filippi, C. Enzinger, J. Frederiksen, O. Ciccarelli, C. R. G. Guttmann, M. P. Wattjes, M. G. Witte, P. C. de Witt Hamer, F. Barkhof, H. Vrenken, MAGNIMS Study Group and Alzheimer's Disease Neuroimaging Initiative, Facing privacy in neuroimaging: removing facial features degrades performance of image analysis methods. European Radiology. Vol. 30, pg. 1062–1074, 2020., https://doi.org/10.1007/s00330-019-06459-3.

17 E. E. L. Buimer, H. G. Schnack, Y. Caspi, N. E. M. van Haren, M. Milchenko, P. Pas, H. E. Hulshoff Pol, R. M. Brouwer, De-identification procedures for magnetic resonance images and the impact on structural brain measures at different ages. Human Brain Mapping. Vol. 42, pg. 3643–3655, 2021., https://doi.org/10.1002/hbm.25459.

18 K. Packhäuser, S. Gündel, F. Thamm, F. Denzinger, A. Maier, Deep learning-based anonymization of chest radiographs: a utility-preserving measure for patient privacy. Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science. Vol. 14222, pg. 262–272, 2023., https://doi.org/10.1007/978-3-031-43898-1_26.

19 K. Packhäuser, L. Folle, F. Thamm, A. Maier, Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pg. 1–5, 2023., https://doi.org/10.1109/ISBI53787.2023.10230346.

20 H. Montenegro, J. S. Cardoso, Anonymizing medical case-based explanations through disentanglement. Medical Image Analysis. Vol. 95, pg. 103209, 2024., https://doi.org/10.1016/j.media.2024.103209.

21 M. P. Heinrich, L. Hansen, Implicit neural obfuscation for privacy preserving medical image sharing. Proceedings of Machine Learning Research (MIDL). Vol. 250, pg. 596–609, 2024., https://proceedings.mlr.press/v250/heinrich24a.html.

22 H. Li, N. Ayache, H. Delingette, Generative medical image anonymization based on latent code projection and optimization. 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pg. 1–4, 2025., https://doi.org/10.1109/ISBI60581.2025.10981125.

23 S. U. H. Dar, M. Seyfarth, I. Ayx, T. Papavassiliu, S. O. Schoenberg, R. M. Siepmann, F. C. Laqua, J. Kahmann, N. Frey, B. Baeßler, S. Foersch, D. Truhn, J. N. Kather, S. Engelhardt, Unconditional latent diffusion models memorize patient imaging data. Nature Biomedical Engineering. Vol. 10, pg. 458–472, 2026., https://doi.org/10.1038/s41551-025-01468-8.

24 N. Molchanova, B. Marechal, J.-P. Thiran, T. Kober, T. Huelnhagen, J. Richiardi, Fast refacing of MR images with a generative neural network lowers re-identification risk and preserves volumetric consistency. Human Brain Mapping. Vol. 45, article e26721, 2024., https://doi.org/10.1002/hbm.26721.

25 S. He, H. Luo, P. Wang, F. Wang, H. Li, W. Jiang, TransReID: transformer-based object re-identification. Proceedings of the IEEE/CVF International Conference on Computer Vision. pg. 15013–15022, 2021., https://openaccess.thecvf.com/content/ICCV2021/html/He_TransReID_Transformer-Based_Object_Re-Identification_ICCV_2021_paper.html.

26 Z. Wang, F. Zhu, S. Tang, R. Zhao, L. He, J. Song, Feature erasing and diffusion network for occluded person re-identification. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pg. 4754–4763, 2022., https://doi.org/10.1109/CVPR52688.2022.00471.

27 W. Chen, X. Xu, J. Jia, H. Luo, Y. Wang, F. Wang, R. Jin, X. Sun, Beyond appearance: a semantic controllable self-supervised learning framework for human-centric visual tasks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pg. 15050–15061, 2023., https://openaccess.thecvf.com/content/CVPR2023/html/Chen_Beyond_Appearance_A_Semantic_Controllable_Self-Supervised_Learning_Framework_for_Human


© The National High School Journal of Science 2026

-Centric_CVPR_2023_paper.html.
28 N. Akash R J, A. Tadanki, J. Sivaswamy, CheXtriev: anatomy-centered representation for case-based retrieval of chest radiographs. Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science. Vol. 15001, pg. 461–471, 2024., https://doi.org/10.1007/978-3-031-72378-0_43.
29 Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, F. Wu, Diverse part discovery: occluded person re-identification with part-aware transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pg. 2898–2907, 2021., https://openaccess.thecvf.com/content/CVPR2021/html/Li_Diverse_Part_Discovery_Occluded_Person_Re-Identification_With_Part-Aware_Transformer_CVPR_2021_paper.html.
30 X. Dong, S. Kim, Z. Jin, J. Y. Hwang, S. Cho, A. B. J. Teoh, Open-set face identification with index-of-max hashing by learning. Pattern Recognition. Vol. 103, pg. 107277, 2020., https://doi.org/10.1016/j.patcog.2020.107277.
31 J. Wang, J. Cao, C. Song, Gallery-distance profiles for open-set person re-identification with frozen encoders in camera networks. Sensors. Vol. 26, pg. 5479, 2026., https://doi.org/10.3390/s26175479.
32 M. Kim, A. K. Jain, X. Liu, AdaFace: quality adaptive margin for face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pg. 18750–18759, 2022., https://doi.org/10.1109/CVPR52688.2022.01819.
33 X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. CVPR, 2017., https://doi.org/10.1109/CVPR.2017.369.
34 K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. CVPR, 2016., https://doi.org/10.1109/CVPR.2016.90.
35 G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks. CVPR, 2017., https://doi.org/10.1109/CVPR.2017.243.
36 J. Deng, J. Guo, N. Xue, S. Zafeiriou, ArcFace: additive angular margin loss for deep face recognition. CVPR, 2019., https://doi.org/10.1109/CVPR.2019.00482.
37 J. P. Cohen, J. D. Viviano, P. Bertin, P. Morrison, P. Torabian, M. Guarrera, M. P. Lungren, A. Chaudhari, R. Brooks, M. Hashir, H. Bertrand, TorchXRayVision: a library of chest X-ray datasets and models. Proceedings of the 5th International Conference on Medical Imaging with Deep Learning. Vol. 172, pg. 231–249, 2022., https://proceedings.mlr.press/v172/cohen22a.html.

## Supplementary Information

The online version contains supplementary material available at https://nhsjs.com/?p=63440


© The National High School Journal of Science 2026