

Patchhealer: Counterfactual Image Segment Transplants with chest X-ray domain check

Hakan Lane¹

hlane@uni-mainz.de

Michal Valko²

misko.valko007@gmail.com

Veda Sahaja Bandi³

sankarvedasahaja@gmail.com

Nandini Lokesh Reddy³

lreddynandini@gmail.com

Stefan Kramer¹

www.datamining.informatik.uni-mainz.de

¹ Department of Computer Sciences

Data Mining Group

Johannes Gutenberg University

Staudingerweg 9, 55130, Mainz, Germany

² Theological Institute

Catholic University in Ruzomberok

Hrabovecka cesta 1A, 03401,

Ruzomberok, Slovakia

³ Department of Data Science

University of Massachusetts Dartmouth

285 Old Westport Rd, 02747, North

Dartmouth, United States of America

Abstract

Counterfactual (CF) techniques extend XAI by shifting from a path- to a goal-focused paradigm, deriving what changes are required to produce a shift in the predicted class. We present a novel method for CF image generation inspired by human organ transplants. The CF generation procedure copies segments of pixels from the source (of the opposite class) to the target until the desired class is reached, based on an efficientnet classification network. The method is combined without of distribution detection with a Convolutional Neural Network (CNN) trained on the original training set together with fake images contaminated by patches from other cardiovascular scanning types. The methods were employed on a set of chest X-rays labelled either as Cardiomegaly or Healthy. The out of domain detection reached accuracies of 90% or higher and rejected other medical scans better than purely white patches.

1 Introduction

Machine learning for image-based prediction has become an integral tool in the diagnosis of many diseases, such as heart disease, breast cancer, diabetes, and liver disease [1]. However, explaining these predictions and gaining trust and acceptance of end users such as doctors and patients remains daunting. Multiple approaches including GRAD-CAM heatmaps [2], Layer-Relevance Propagation (LRP) [3] and so-called DeepLift methods [4] have been explored with varying degrees of success. The presence of out of domain (OOD) samples poses significant challenge for accurate predictions, as the contamination of only a few images from another domain may yield distinctly poorer performance [5].

In recent years, counterfactual (CF) techniques have attracted increased levels of attention and have been applied to many medical fields in the task of disease classification and diagnosis. It rests on the concept of explaining a prediction by demonstrating how an object from a different class appears when it is shifted across the decision boundary [18]. Within cardiovascular imaging, Thigarajan *et al.* [58] used a CF method to explain and visualise detection of anomalies on chest X-ray images.

This study introduces a new method to produce counterfactual images via copying segments by transferring image segments within the same domain but featuring different pathologies. The method was augmented by an out of domain. Our research questions are a) Does the rejection success depend on what domain the intrusion comes from? b) Does the rejection success depend on what domain the network is trained on? c) Does the rejection success depend on whether the network is trained on images with a diagnosis or from healthy subjects?

2 Previous Work

Cardiomegaly X-ray Prediction. Within image-based classification, convolutional neural networks (CNNs) have, to date, achieved accuracies of 90% or higher [10, 27, 42], reflective of their potential as clinical expert systems [29, 30]. For chest X-ray datasets, a number of deep learning methods have been employed, producing accuracies between 70 and 80% [9]. Many of these approaches use hybrid models integrating CNN with other architectures, such as U-Net [39].

Counterfactuals. The counterfactual approach has demonstrated efficacy in satisfying the information needs of end-users, and in clarifying medical disease diagnoses [22]. Although user-studies remain limited, preliminary surveys indicate promising results. For example, Delaney *et al.* [5] have shown that users presented with counterfactual images were more likely to correctly classify images across different domains. Frameworks for producing a variety of workable counterfactual explanations have been created in different contexts [24, 31]. One such framework by Nagesh *et al.* [26] adopted a variational autoencoder-based methodology that yielded counterfactuals that outperformed previous approaches, both qualitatively and quantitatively. Additionally, other studies have reported that exposure to counterfactual explanations increases participants' trust in the models. Individuals who viewed counterfactual images reported reduced levels of stress and frustration compared to the control group [22]. Both users and system designers benefit from the counterfactual explanation, which facilitates improved model comprehension and debugging. Furthermore, counterfactual explanations perform robustly in traditional quantitative assessment [26]. Using the counterfactual maps as a framework, Oh *et al.* [28] developed an attention-based feature refining model to improve the diagnostic model's generalisation. Recently, Tan *et al.* [36] proposed two different assessment criteria to evaluate better utilisation of counterfactual explanations.

Image Patch Translation. Various imaging modalities and processing techniques have been used to enhance medical education [6, 13]. Patch-based medical image segmentation techniques have also been applied with Quantum Tensor Networks [32] and latent aspect models for contextual classification [23]. One assessment of a proposed segmentation scheme was conducted using 20 cardiac MRIs and 20 CT scans [44]. Barnes *et al.* [0] introduced a novel approach to image modification with application in medical retargeting and

completion of medical images. Patch-based image synthesis and inpainting approaches offer potential for improving the synthesis of medical images [21, 82].

Superpixel Segmentation. Superpixel image segmentation refers to an image processing technique that groups similar pixels into larger perceptually meaningful regions known as superpixels [85]. Superpixels are typically more uniform and compact than individual pixels, facilitating easier analysis and processing [86]. This segmentation method is commonly used in tasks such as object detection, image classification, and image editing [87]. The segmentation of a chest X-ray is shown in Figure 1.

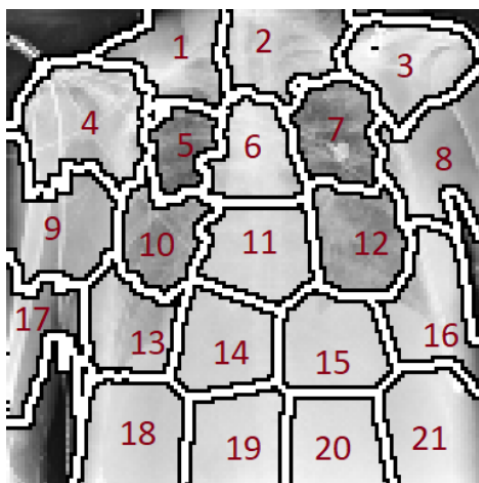


Figure 1: Segmentation via superpixels of chest X-ray

Out-Domain Detection Out-of-domain detection (OOD) in the medical imaging field is crucial for ensuring the reliability and accuracy of diagnostic models. This process involves identifying images or data that significantly differ from the training dataset, which can occur due to variations in imaging protocols, patient demographics, or disease presentations [88]. Techniques such as deep learning-based anomaly detection and domain adaptation strategies have been employed to enhance the robustness of medical imaging systems against out-of-domain samples. For instance, methods leveraging convolutional neural networks (CNNs) have shown promise in distinguishing between in-domain and out-of-domain images by learning feature representations that capture the underlying data distribution [89]. Additionally, the use of generative adversarial networks (GANs) has been explored to synthesise in-domain-like images from out-of-domain samples, thereby improving model performance [90]. Addressing out-of-domain detection is essential for maintaining the clinical applicability of AI systems in diverse medical settings, as it directly impacts patient safety and diagnostic accuracy.

3 Method

Image Patch Translation. A synthetic image is generated through this patch-based image-to-image transfer by copying a set of patches from the counterfactual image to the original image. The translated patches are selected using a modified version of the ViCE algorithm [10] that we have adapted for images. Each image was segmented into a number of patches via the superpixels function [15]. Unlike conventional approaches that use smaller patches (e.g. 2x2 or 10x10 pixels method) to create CF images, our method reflects anatomical significance of chest area. It allows larger and more realistic transplants that better match medical pathology, such as cardiomegaly, where large section of heart are involved.

Segment Transfer. Given two patches, one from the original image, and one from the counterfactual image, we adapt the morphmix algorithm from the MiMICRI system [10] to transfer the selected patch from the counterfactual image into the original image. Morphmix first aligns the counterfactual patch centroid to the centroid of the original image patch. The algorithm then copies pixels from the counterfactual into the original using a flood-fill starting from the centroid pixel. In cases where the counterfactual image patch is smaller than the original image patch, we estimate missing pixel values by interpolating between corresponding pixels in the counterfactual and original image.

Crossing Decision Boundary. The algorithm iteratively replaces patches from the original with copies of the corresponding patch from the counterfactual image. The iterations continue until either a) the predicted class for X becomes the class of Y or b) a user-defined threshold of maximum translated fraction has been reached. By default, the threshold for the number of patches that can be replaced is the total number of patches in X (i.e., the entire image is replaced). This threshold can be adjusted by users to limit the extent of amount of changes made to the image before the procedure is deemed to be pruned. In the case study, EfficientNet [16], a family of CNNs [16] designed for high efficiency and effectiveness in image classification, was used to distinguish between CM and Healthy. Utilising a compound scaling method, it uniformly scales the network’s depth, width, and resolution, optimising performance and resource use. An example of a successful transplant in one step is displayed in Figure 2

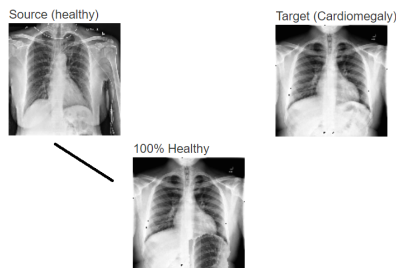


Figure 2: Example successful translation

Pseudocode. Based on the class prediction operator $predictClass(X)$, a domain operator $inDomain(X)$ and the functions $nextSegment$, $mostAlignedSegment$ and $transferPatch$, the

pseudocode for the algorithm is presented as Algorithm 1.

Algorithm 1 Image counterfactual algorithm

Require: X target image, Y source

$Class(X) \leftarrow predictClass(X)$

$count \leftarrow 0$

while $Class(X) \neq Class(Y) \wedge count \leq nsegments(X)$ **do**

$i \leftarrow nextSegment(X)$

$j \leftarrow mostAlignedSegment(X, Y, i)$

$X' \leftarrow transferPatch(X, i, Y, j)$

if $inDomain(X') \equiv TRUE$ **then**

\triangleright Check if segment is valid

$X \leftarrow X'$

$Class(X) \leftarrow predictClass(X)$

$count \leftarrow count + 1$

end if

end while

Out of Distribution detection The classifier used for our patch-based counterfactual explanation method is a convolutional neural network with the architecture in Table 1.

Table 1: Specification of the CNN

Layer	Type	Size
	(Conv2D)	128 x 128 x 32
	(Activation)	128 x 128 x 32
	(Conv2D)	128 x 128 x 16
	(LeakyReLU)	128 x 128 x 16
	(Batch Normalisation)	128 x 128 x 16
	(Max Pooling)	64 x 64 x 16
	(Dropout)	64 x 64 x 16
	(Flatten)	1 x 65536
	(Dense)	1 x 100
	(Activation)	1 x 100
	(Dropout)	1 x 100
	(Dense)	1 x 2
	(Activation)	1 x 2

Note: Conv2D= Two-dimensional convolutional neural network with a 3x3 pixel kernel, LeakyReLU= Rectified Linear Unit with a positive gradient while not active.

The network classifies images into the classes CXR (Chest X-ray) and *unreal* (fake). The input is a grayscale 128 x 128 pixel matrix (range 0-1) and the output a two valued binary class probability vector. The R packages superpixels, OpenImageR, and magick were used for image processing, keras3 and Mimicri for the translation section and shiny for the user interface ¹

¹<https://nightingale.zdv.uni-mainz.de:3838/cmgame>

Dataset The presented case study uses the chest X-ray data set from the National Institute of Health (NIH), United States [4], with over 100,000 scanned images with annotations performed by doctors. A subset with 5552 images in total with only the cardiomegaly and healthy labels were used in this study. To create the desired out of domain sample, the images were distorted, with blank areas or patches from other imaging. The intruding segments were chosen from the following public image sets: a) Tufts Medical Echocardiogram DataSet (Echos) [5], b) Completely white patches, c) Chest Computer Tomography (CT) from Lincoln University College, Omega Hospitals [6]. The network was trained with cardiomegaly and healthy images, a sample of the distortion images is shown in Figure 3

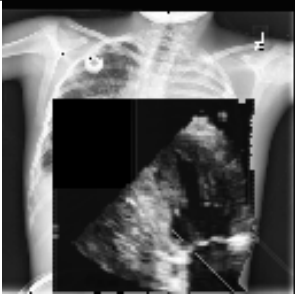

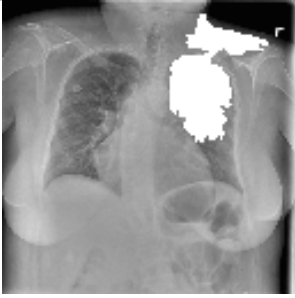
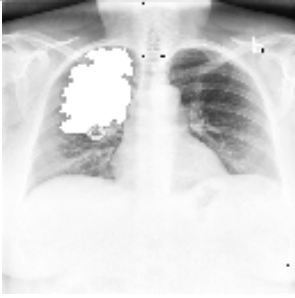
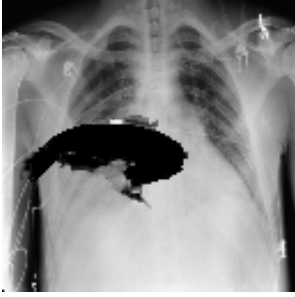
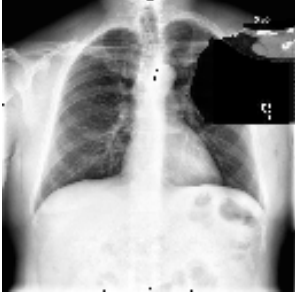
	Cardiomegaly	Healthy
Echocardiography		
White		
CT		

Figure 3: A 2x3 table with images and headers

4 Results

4.1 Image similarities

The closeness between a randomly selected image of each type as measured via cosine similarity is presented in Table 2.

Table 2: Cosine similarities sample

Source	OCM	OH	ECM	EH	WCM	WH	CCM	CH
OCM	-	0.86	0.69	0.83	0.88	0.85	0.66	0.81
OH	-	-	0.75	0.93	0.94	0.96	0.80	0.92

Note: CM:Cardiomegaly; H:Healthy; O:Original; E:Echos; W:White; C:Computer Tomography The white distortion leads to smaller dissimilarity between images compared to a contamination from the other scanning domains.

4.2 Domain detection.

The generic model from Table 1 was trained on training data in three different configurations: a) Separate models for each source of distortion (echo, white, CT); b) Separate models per training class (CM/Healthy); c) One model where all training classes and distortions are included.

For each image in the test set, segment size was chosen randomly within the size vector (6,9,12,15,18,21,24,27), the testing image segmented with superpixels, a target segment to transfer was chosen randomly and the corresponding segment in the source (distorted) image was found via the bounding box of the assigned patch. The test set accuracies (in percent) of correctly classified images (in/out of domain) for each combination of intrusion type and pathology are presented in Table 3, Table 4, and Table 5, respectively. The numbers are based on a test sample of 319-320 images of each cell (Label+Distortion).

Table 3: Accuracy when training on same domain

Source	Cardiomegaly	Healthy
Echo	91.2	92.9
White	83.5	85.1
CT	97.7	99.4

Table 4: Accuracy when training same label

Source	Cardiomegaly	Healthy
Echo	97.1	94.7
White	69.4	80.4
CT	97.6	96.4

Overall, the network is rejecting intrusions from other medical images better than the purely white patches, and performs worse when the space of distortions becomes less homogeneous.

Table 5: Accuracy when training all

Source	Cardiomegaly	Healthy
Echo	80.6	79.3
White	78.9	78.6
CT	80.6	80.4

4.3 User tests

A user interface has been developed with the graphics embedded as portrayed in Figure 4. It provides the "player" of the gamified Shiny app options to select a patch for translation in the source and which segment it should replace. Following the algorithm, the system will notify the user when the decision boundary has been crossed, i.e. the CM image has been converted into a healthy prediction.

CardioCNN: Transplant to Healthy

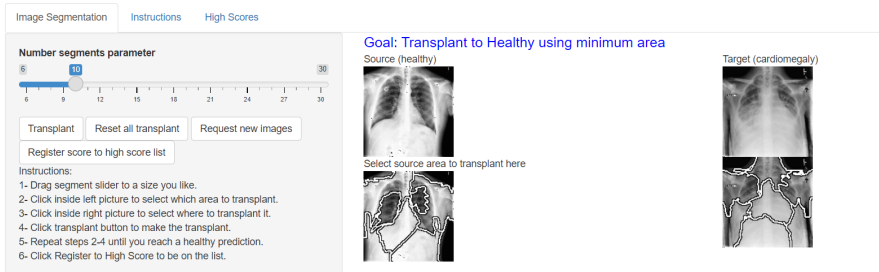


Figure 4: Example successful translation

Early testing showed a high rate of success, with an average of 30 percent of the pixels being transplanted at first try.

5 Discussion

The classification reached very balanced results with an accuracy close to 80%, reflecting its effectiveness in identifying relevant instances while minimising false positives.

The transplant results demonstrated the potential of this approach to effectively transfer content between classes in the same domain, enabling the generation of images that depict medical imaging outcomes from opposing categories while retaining a realistic visual appearance. Overall, the algorithm achieved succession rate approximately 85%, requiring the translation of roughly three fifths of the target image to achieve the desired class shift. The proposed approach offers several advantages compared to existing methods. It operates directly on raw image patches, preserving the spatial relationships and contextual information within the scene. This allows for more natural and realistic translations, compared to methods reliant on hand-crafted features or pre-trained segmentation models.

The domain check method was trained on patchwise intrusions similar to the reached maintains a check of the method's integrity against attempts to use services with "faked" data, such as combining images from different scan techniques. It thus guards against synthetic results with a level of accuracy comparable to state of the art from challenges on

medical images [45], and considerably higher than detection in many other domains [49]. It is apparent that the detection of "fake" images is less accurate when the out of domain contamination is more similar to the original. This issue resembles the "phishing" technique of making the intrusion resemble a normal occurrence as close as possible [47].

Despite these promising findings, several avenues for future research remain. One direction involves developing more sophisticated patch alignment techniques to further enhance translation accuracy and consistency. The domain check technique should be trained with more sophisticated data resembling expected user intrusion tactics. Additionally, investigating the impact of patch extraction strategies and incorporating semantic information into the patch translation process could further improve the quality and realism.

6 Conclusion

In conclusion, the present study has demonstrated the potential of image-to-image patch translation for generating semifactual images of a different class. This was done through an architecture with a neural network for image classification, identification of the nearest counterfactual from a data set with two classes, superpixel image segmentation into patches, selection of the matching patches in the opposing image, a morphing algorithm matching segment geometries for translating patches between images of different classes, and a convolutional model for rejecting attempts to translate images from other domains.

The tests indicate that the detection of hostile images displayed a high dependence on the degree of similarity between the original and the fake, as the accuracy was much lower when they were closer to each other, illustrating the security problem of detecting insider threats and phishing.

The method could be used as a tool demonstrating the effect of transplants and as an XAI gamified technique to illustrate image classification. While the proposed method for generating counterfactual images shows promising results in terms of accuracy and interpretability, further work is necessary to evaluate their effectiveness in real-world scenarios. Future research should focus on refining patch alignment methods, incorporating semantic information, and investigating the application of this approach to more complex image transformations.

Acknowledgments The National Institutes of Health, Bethesda, MD, United States, has provided the chest X-rays with labels based on pathology. The method was developed with the support of the curAIHeart project funded by the German Federal Ministry of Education and Research "Clusters of Future" program with program code 03ZU120212. Dr. Himanshu Verma, CogniCode IT Solutions, Gwalior, Madhya Pradesh, India, has curated the data from the NIH and provided a balanced dataset via the Kaggle² web site.

References

- [1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012. doi: 10.1109/TPAMI.2012.120.

²www.kaggle.com

- [2] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transaction on Graphic*, 28(3):619–629, 2009. doi: 10.1145/3596711.3596777.
- [3] Haralabos Bougias, Eleni Georgiadou, Christina Malamateniou, and Nikolaos Stogianos. Identifying cardiomegaly in chest x-rays: a cross-sectional study of evaluation and comparison between different transfer learning methods. *Acta Radiologica*, 62(12):1601–1609, 2020. doi: 10.1177/0284185120973630.
- [4] Hongkun Chen, Jie Cao, and Mingdi Yi. Out of distribution detection for medical images. In *International Conference on Computer Vision, Application, and Algorithm (CVAA 2022)*, volume 12613, pages 95–102. SPIE, 2023.
- [5] Eoin Delaney, Arjun Pakrashi, Derek Greene, and Mark T Keane. Counterfactual explanations for misclassified images: How human and machine explanations differ. *Artificial Intelligence*, 324:103995, 2023. doi: 10.1016/j.artint.2023.103995.
- [6] Parvati Dev. Imaging and visualization in medical education. *IEEE Computer Graphics and Applications*, 19(3):21–31, 1999. doi: 10.1109/38.761545.
- [7] Bhanu Prakash Doppala. Chest-ct images, 2020. URL <https://data.mendeley.com/datasets/w4mv8ypr3/1>. original document from Bhanu Prakash Doppala.
- [8] Marion Dörrich, Markus Hecht, Rainer Fietkau, Arndt Hartmann, Heinrich Iro, Antoniu-Oreste Gostian, Markus Eckstein, and Andreas M Kist. Explainable convolutional neural networks for assessing head and neck cancer histopathology. *Diagnostic Pathology*, 18(1):121, 2023. doi: 10.1186/s13000-023-01407-8.
- [9] Meherwar Fatima and Maruf Pasha. Survey of machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications*, 9(01):1–16, 2017. doi: 10.4236/jilsa.2017.91001.
- [10] Yunendah Nur Fu’adah and Ki Moo Lim. Classification of atrial fibrillation and congestive heart failure using convolutional neural network with electrocardiogram. *Electronics*, 11(15):2456, 2022. doi: 10.3390/electronics11152456.
- [11] Oscar Gomez, Steffen Holter, Jun Yuan, and Enrico Bertini. Vice: Visual counterfactual explanations for machine learning models. In *Proceedings of the 25th international conference on intelligent user interfaces*, pages 531–535, 2020. doi: 10.1145/3377325.3377536.
- [12] Grace Guo, Lifu Deng, Animesh Tandon, Alex Endert, and Bum Chul Kwon. Mimicri: Towards domain-centered counterfactual explanations of cardiovascular image classification models. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1861–1874, 2024. doi: 10.1145/3630106.3659011.
- [13] Michael Haubner, Christian Krapichler, Andreas Losch, Karl Hans Englmeier, and Wilhelm Van Eimeren. Virtual reality in medicine-computer graphics and interaction techniques. *IEEE Transactions on Information Technology in Biomedicine*, 1(1):61–72, 1997. doi: 10.1109/4233.594047.

- [14] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. 2016. doi: 10.48550/ARXIV.1610.02136.
- [15] Zhe Huang, Gary Long, Benjamin Wessler, and Michael C. Hughes. A new semi-supervised learning benchmark for classifying view and diagnosing aortic stenosis from echocardiograms. In *Proceedings of the 6th Machine Learning for Healthcare Conference (MLHC)*, 2021. URL https://tmed.cs.tufts.edu/papers/HuangEtAl_MLHC_2021.pdf.
- [16] Sakshi Indolia, Anil Kumar Goswami, Surya Prakesh Mishra, and Pooja Asopa. Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science International Conference on Computational Intelligence and Data Science*, 132:679–688, 2018. doi: <https://doi.org/10.1016/j.procs.2018.05.069>.
- [17] Ankit Kumar Jain and B. B. Gupta. Phishing detection: Analysis of visual similarity based approaches. *Security and Communication Networks*, 2017:1–20, 2017. ISSN 1939-0122. doi: 10.1155/2017/5421046. URL <http://dx.doi.org/10.1155/2017/5421046>.
- [18] Eoin M Kenny and Mark T Keane. On generating plausible counterfactual and semi-factual explanations for deep learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11575–11585, 2021. doi: 10.1609/aaai.v35i13.17377.
- [19] Byung Chun Kim, Byungro Kim, and Yoonsuk Hyun. Investigation of out-of-distribution detection across various models and training methodologies. *Neural Networks*, 175:106288, 2024. doi: <https://doi.org/10.1016/j.neunet.2024.106288>.
- [20] Daniel Krakowczyk, David Robert Reich, Paul Prasse, Sebastian Lopuschkin, Lena Ann Jäger, and Tobias Scheffer. Selection of xai methods matters: Evaluation of feature attribution methods for oculomotoric biometric identification. In *Gaze Meets Machine Learning Workshop*, pages 66–97, 2023. URL <https://proceedings.mlr.press/v210/krakowczyk23a.html>.
- [21] Shiguang Liu and Jingting Wu. Fast patch-based image hybrids synthesis. In *2011 12th International Conference on Computer-Aided Design and Computer Graphics*, pages 191–197, 2011. doi: 10.1109/CAD/Graphics.2011.83.
- [22] Silvan Mertes, Tobias Huber, Katharina Weitz, Alexander Heimerl, and Elisabeth André. Ganterfactual—counterfactual explanations for medical non-experts using generative adversarial learning. *Frontiers in artificial intelligence*, 5:825565, 2022. doi: 10.3389/frai.2022.825565.
- [23] Florent Monay, Pedro Quelhas, Jean-Marc Odobez, and Daniel Gatica-Perez. Contextual classification of image patches with latent aspect models. *EURASIP Journal on Image and Video Processing*, 2009:1–20, 2009. doi: 10.1155/2009/602920.
- [24] Ramaravind Kommiya Mothilal, Amit Sharma, and Chenhao Tan. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 607–617, 2020. doi: 10.1145/3351095.3372850.

- [25] Lampros Mouselimis. *SuperpixelImageSegmentation: Image Segmentation using Superpixels, Affinity Propagation and Kmeans Clustering*, 2022. URL <https://CRAN.R-project.org/package=SuperpixelImageSegmentation>. R package version 1.0.5.
- [26] Supriya Nagesh, Nina Mishra, Yonatan Naamad, James M Rehg, Mehul A Shah, and Alexei Wagner. Explaining a machine learning decision to physicians via counterfactuals. In *Conference on Health, Inference, and Learning*, pages 556–577, 2023. URL <https://proceedings.mlr.press/v209/nagesh23a.html>.
- [27] Jeffrey J Nirshchl, Andrew Janowczyk, Eliot G Peyster, Renee Frank, Kenneth B Margulies, Michael D Feldman, and Anant Madabhushi. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of h&e tissue. *PLoS one*, 13(4):e0192726, 2018. doi: 10.1371/journal.pone.0192726.
- [28] Kwanseok Oh, Jee Seok Yoon, and Heung-II Suk. Learn-explain-reinforce: counterfactual reasoning and its guidance to reinforce an alzheimer’s disease diagnosis model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4843–4857, 2022. doi: 10.1109/TPAMI.2022.3197845.
- [29] Alexandre Tomazati Oliveira and Euripedes Guilherme de Oliveira Nobrega. A novel arrhythmia classification method based on convolutional neural networks interpretation of electrocardiogram images. In *2019 IEEE International Conference on Industrial Technology (ICIT)*, pages 841–846, 2019. doi: 10.1109/ICIT.2019.8755177.
- [30] Andreas Østvik, Erik Smistad, Svein Arne Aase, Bjørn Olav Haugen, and Lasse Lovstakken. Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. *Ultrasound in medicine & biology*, 45(2):374–384, 2019. doi: 10.1016/j.ultrasmedbio.2018.07.024.
- [31] Martin Pawelczyk, Klaus Broelemann, and Gjergji Kasneci. Learning model-agnostic counterfactual explanations for tabular data. In *Proceedings of the web conference 2020*, pages 3126–3132, 2020. doi: 10.1145/3366423.3380087.
- [32] Tijana Ružić and Aleksandra Pižurica. Context-aware patch-based image inpainting using markov random field modeling. *IEEE transactions on image processing*, 24(1): 444–456, 2014. doi: 10.1109/TIP.2014.2372479.
- [33] Clemens Seibold, Anna Hilsman, and Peter Eisert. Focused lrp: Explainable ai for face morphing attack detection. In *2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*, page 88–96, 2021. doi: 10.1109/wacvw52041.2021.00014.
- [34] Raghavendra Selvan, Erik B Dam, Søren Alexander Flensburg, and Jens Petersen. Patch-based medical image segmentation using matrix product state tensor networks. *arXiv preprint arXiv:2109.07138*, 2021. URL <https://arxiv.org/abs/2109.07138>.
- [35] Subhashree Subudhi, Ram Narayan Patro, Pradyut Kumar Biswal, and Fabio Dell’Acqua. A survey on superpixel segmentation as a preprocessing step in hyper-spectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5015–5035, 2021. doi: 10.1109/JSTARS.2021.3076005.

- [36] Juntao Tan, Shuyuan Xu, Yingqiang Ge, Yunqi Li, Xu Chen, and Yongfeng Zhang. Counterfactual explainable recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 1784–1793, 2021. doi: 10.1145/3459637.3482420.
- [37] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 6105–6114, 2019. URL <https://proceedings.mlr.press/v97/tan19a.html>.
- [38] Jayaraman J. Thiagarajan, Kowshik Thopalli, Deepta Rajan, and Pavan Turaga. Training calibration-based counterfactual explainers for deep learning models in medical image analysis. *Scientific Reports*, 12(1):597, 2022. doi: 10.1038/s41598-021-04529-5.
- [39] Daiju Ueda, Toshimasa Matsumoto, Shoichi Ehara, Akira Yamamoto, Shannon L Walston, Asahiro Ito, Taro Shimono, Masatsugu Shiba, Tohru Takeshita, Daiju Fukuda, and Yukio Miki. Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. *The Lancet Digital Health*, 5(8):e525–e533, 2023. doi: 10.1016/s2589-7500(23)00107-3.
- [40] Francesco Verdoja and Marco Grangetto. Fast superpixel-based hierarchical approach to image segmentation. In *Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part I 18*, pages 364–374, 2015. doi: 10.1007/978-3-319-23231-7_33.
- [41] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017. doi: 10.1109/CVPR.2017.369.
- [42] Farah Yasmin, Syed Muhammad Ismail Shah, Aisha Naeem, Syed Muhammad Shujaiddin, Adina Jabeen, Sana Kazmi, Sarush Ahmed Siddiqui, Pankaj Kumar, Shiza Salman, Syed Adeel Hassan, et al. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. *Reviews in cardiovascular medicine*, 22(4):1095–1113, 2021. doi: 10.31083/j.rcm2204121.
- [43] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017. URL <https://arxiv.org/abs/1703.10593>.
- [44] Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016. doi: 10.1016/j.media.2016.02.006.
- [45] David Zimmerer, Peter M. Full, Fabian Isensee, Paul Jager, Tim Adler, Jens Petersen, Gregor Kohler, Tobias Ross, Annika Reinke, Antanas Kascenas, Bjorn Sand Jensen, Alison Q. O’Neil, Jeremy Tan, Benjamin Hou, James Batten, Huaqi Qiu, Bernhard Kainz, Nina Shvetsova, Irina Fedulova, Dmitry V. Dylov, Baolun Yu, Jianyang Zhai, Jingtao Hu, Runxuan Si, Sihang Zhou, Siqi Wang, Xinyang Li, Xuerun Chen, Yang

Zhao, Sergio Naval Marimont, Giacomo Tarroni, Victor Saase, Lena Maier-Hein, and Klaus Maier-Hein. Mood 2020: A public benchmark for out-of-distribution detection and localization on medical images. *IEEE Transactions on Medical Imaging*, 41(10): 2728–2738, 2022. doi: 10.1109/tmi.2022.3170077.