VISUAL CORRESPONDENCE-BASED EXPLANATIONS IMPROVE CONVOLUTIONAL NEURAL NETWORKS FOR CLASIFICATIONOF MAMMOGRAMS

Manuscript Info

Manuscript History
Received: xxxxxxxxxxxxxxx
Final Accepted: xxxxxxxxxxxx
Published: xxxxxxxxxxxxxxx

Key words:-
Visual Correspondence, Mammograms, EMD-Corr, Convolutional Beural Networks.

Abstract

Explainable Artificial Intelligence (XAI) classifications are increasingly important in convolutional neural networks for the classification of mammograms. In this paper, we propose a novel architecture of Visual Correspondence Based Explanations that improve Convolutional Neural Networks (CNNs) for Classification of Mammograms of self-interpretable image classifiers that first explain, and then predict by harnessing the visual correspondences between a query breast cancer X-Ray image and exemplars. In the evaluation of our proposed models, the k-nearest neighbor (kNN) classifier improves upon ResNet-18 on Breast Cancer datasets. This means during testing, kNN outperforms ResNet-18 model (e.g. +2.28 for Accuracy, +1.45 for Sensitivity, and +0.98 for Specificity). Our visual correspondence model (EMD-Corr) improves kNN and ResNet 18 where EMD-Corr slightly underperforms kNN model (e.g. -0.09 for Accuracy, -1.02 for Sensitivity, and -0.48 for Specificity) and EMD-Corr outperforms the ResNet 18 (e.g. +2.17 for Accuracy, +0.44 for Sensitivity, and +.09 for Specificity). On the other hand, EMD-Corr model can improve the localization of abnormal areas of breast cancer by showing patch locations of cancer regions in Breast Cancer images in comparison with the doctor's annotation.

Introduction:-

Explainable Artificial Intelligence (XAI) focuses on the term “explainable” or “interpretable,” referring to models that can produce outputs understandable by humans. An “explainable” model is one that inherently provides insights into its decision-making process. Currently, one of the core techniques in XAI involves comparing input images with training-set prototypes, serving as the foundation for many practical applications by G. Casalino et al. (2024), Shin, Y.-M. et al. (2022). In recent years, additional research on XAI for localizing medical images has emerged, alongside studies that demonstrate how XAI improves medical prediction accuracy by Agarwal, C. and Nguyen, A. (2020), Alqaraawi, A. et al. (2020), Phuong, N. H.el al. (2022). Several review articles have compared different XAI applications and methodologies in healthcare, highlighting both opportunities and challenges of Karim, M. R. et al (2021), Kim, K. H. et al. (2021). Beyond existing healthcare-focused methods, new XAI approaches aim to enhance model accuracy. Notably, visual correspondence-based explanations have been proposed to improve both the robustness of AI models and the collaborative accuracy of human-AI teams by Salahuddin, Z. et al., (2022), Shen, Y.et al. (2021).

In recent years, the study of patch-wise Explanation methods presented in Taesiri, M. R. et al. (2022) introduced a novel Explainable Artificial Intelligence (XAI) approach called Visual Correspondence-Based Explanations. This method is designed to enhance both the robustness of AI models and the accuracy of human-AI collaboration. The key idea is that in tasks such as face identification or prototype-based classification (e.g., birds), comparing images at the patch level—rather than as whole images - can improve performance. Specifically, by re-ranking the results of a k-Nearest Neighbors (kNN) search based on the similarity between small patches of the input and training images,

classification on out-of-distribution (OOD) data improves significantly. Rather than computing similarity between all possible patch pairs (e.g., 49×49 = 2,401 combinations as done in earlier methods like in G. Casalino et al. (2024)), the new models - EMD-Corr and CHM-Corr - only use the top 5 most relevant patch pairs, making the approach more efficient while still effective. These classifiers base their decisions on a set of support image-patch pairs which also serve as an explanation to users. Fig. 1 in a Visual Corrby Taesiri, M. R. et al. (2022) shows this problem with one positive example and one negative example.

Figure1:- Operating at the image-level visual similarity, kNN incorrectly labels the input toaster due to theadversarial patch (a). EMD-Corr instead ignores the adversarial patch and only uses the head and neck patches of the hen to make decisions (b).

(a) Explanations for kNN's toaster decision (top) and EMD-NN (bottom) (b) Explanations for EMD-Corr's hen decision

In fact, cancer is one of the most dangerous diseases in Vietnam. One of the useful ways is to develop a system for classifying X-Ray images into three classes: Normal, Benign, and Malignant. The authors have applied the ResNet 18 trained on 15,040 Vietnamese mammograms collected by the radiologists from the Vietnam National Cancer Hospital to classify X-Ray images into three classes above by Phuong N. H.(2024). This paper describes how to apply the Visual Correspondence Based Explanations of by Taesiri, M. R. et al. (2022) to improve CNNs for the Classification of Mammograms.

To summarize, the main contributions of this paper are:

- kNN explanations. From the top-20 nearest neighbors (as k = 20 in our kNN), we show the first five breast cancer X-Ray images that are from the predicted class. Note that our kNN explanations consist of five support images for each decision of kNN as opposed to the post-hoc nearest examples, which do not reflect a classifier's decisions.
- Calculating the Interaction of Union (IoU) of five patches with the localization of prototype breast cancer X-Ray images by doctors.

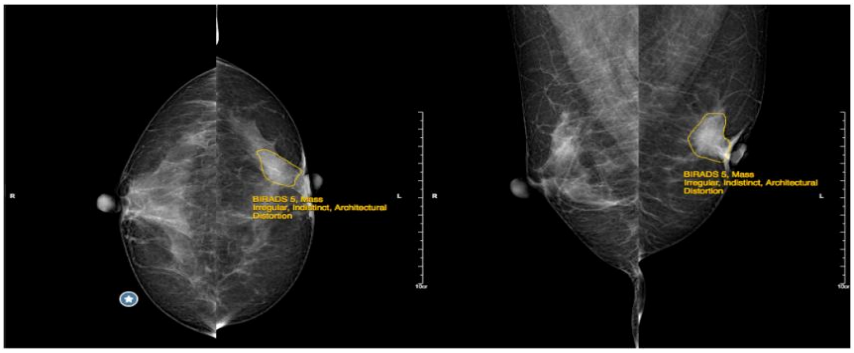
The rest of the paper consists of the following sections: Section 2 presents a proposed method. Section 3 presents experiments. Section 4 shows the evaluation. Section 5 gives some conclusions.

Proposed Methods:-

Datasets

The data set of breast cancer images in this study was built with the support of the radiologists from the Vietnam National Cancer Hospital, Hanoi, Vietnam. Mammography data is captured through the hospital's machine system and then stored in the central database of the PACS (Picture Archiving and Communication Systems) system. Radiologists access and label mammograms through a web-based client program connected to the PACS system. After selecting a mammogram from the mammography database, radiologists will label this mammogram as Normal, Benign, or Malignant based on their medical expertise. In this case, if the doctors observe the abnormal image areas, they will evaluate the suspected image areas from different viewing angles and identify abnormal areas with specific boundaries. This is also the basis for radiologists to have a general conclusion about the patient's condition to record in the medical record. To perform the labeling, the software provides a tool to mark the contour of the abnormal area and specifically annotate the information about the abnormal area. Each abnormal region would be annotated as a malignant tumor or a benign tumor (see Fig. 2).

Figure2:- Localization of malignant areas of R-CC, L-CC, R-MLO, L-MLO images (left to right)



If an image has no abnormal region, it would be marked as the normal state. Among the dataset, 7,939 images have the normal state, 4,055 images contain benign regions, 3,169 images contain malignant regions, and 123 images contain both benign and malignant regions, which are described in Table 1.

Table 1:- Label distribution in the dataset

Type	Number of images
Normal	7,939
Benign	4,055
Malignant	3,169
Benign and malignant	123
All	15,040

The dataset includes 15,040 digital screening mammography images. Each image corresponds to one of four standard views used in screening mammography: R-CC (right craniocaudal), L-CC (left craniocaudal), R-MLO (right mediolateral oblique), L-MLO (left mediolateral oblique). The type and views of images in the dataset are described in Table 2.

Table 2:- The type and views of images in the dataset

Type	L-CC	L-MLO	R-CC	R-MLO
Normal	1,710	2,168	1,805	2,256
Benign	889	1,183	849	1,134
Malignant	682	956	615	916
Benign and malignant	27	42	21	33

In this study, we divided mammograms into three groups: malignant, normal, and benign, and distributed them into the training set, validation set, and test set (Table 3) as in Phuong N. H. et al. (2024).

Table 3:- Details of the distribution in the training set, validation set, and test set

Category	Malignant	Normal	Benign	Total
Training set	2,534	6,349	3,145	12,028
Validation set	316	794	393	1,503
Test set	319	796	394	1,509
Total	3,169	7,939	3,932	15,040

Classifiers

We harness the same ResNet-18 Denselayer256 backbone (where a ResNet-18 model is used to extract features, and the output of this model is passed through a dense layer with 256 units, producing a 256-dimensional feature vector) as the main feature extractor for all main classifiers. Therefore, to test the effectiveness of our models, we compare them with (1) a ResNet-18 classifier; and (2) a kNN classifier that uses the same pretrained DenseLayer256 features. We report the evaluation of all classifiers in Table 4.

ResNet-18

For experiments on dataset with 15,040 breast cancer X-Ray images from the Vietnam National Cancer Hospital, Hanoi, Vietnam and retrain only the last 256-output classification layer (right after avgpool) to create a competitive, baseline ResNet-18 as in Fig. 2. The output of the ResNet 18 is the output of the Dense 256 layer which is a vector with 256 values.

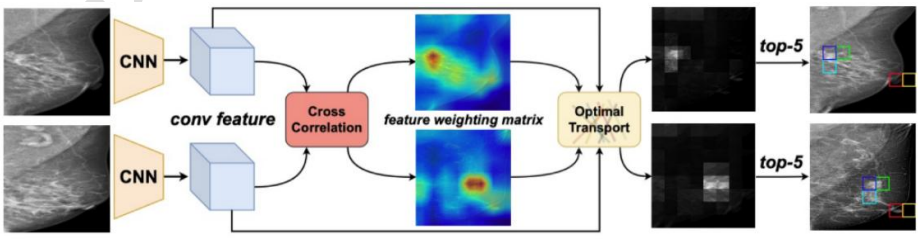
kNN

We realize kNN with the input as the output of the Dense layer 256 of ResNet 18. So, the input of the classifier kNN is a vector of 256 features of ResNet-18. That is, given a query image Q, we sort all training-set images Gi based on their distance D(Q, Gi), which is the cosine distance between the two corresponding image features f(Q) and f(Gi) ∈ R256 where f(.) outputs the avgpool feature of layer256 of ResNet-18. The predicted label of Q is the dominant class among the top-k nearest neighbors. We choose k = 20 as it performs the best among the tested values of k ∈{10, 20, 50, 100}. The formula for the calculation of the cosine distance between Q and Gi:

$$D(Q, G_i) = 1 - \frac{\langle f(Q), f(G_i) \rangle}{\|f(Q)\| \|f(G_i)\|} (1)$$

Where, f(Q) is a feature vector calculated by the input Q, f(Gi) is a feature vector calculated by the training-set image G_i, $\langle \cdot \rangle$ is the dot product, and $\|\cdot\|$ is the L_2 norm operator. Based on the received results, we rank and choose a subset of the nearest images to the input images. The final result is calculated by majority statistics. In our case, we take k=14. The result of the classification of kNN is one of three classes: Malignant, Benign, or Normal.

Figure 3:EMD-Corr re-ranks kNN's top-50 candidates using the patch-wise similarity between the query and each candidate over the top-5 pairs of patches that are the most important and the most similar (i.e. highest EMD flows in EMD-Corr).



EMD-Corr

As kNN compares breast cancer X-Ray images using only image-level features, it lacks the capability of paying attention to fine details in images. Therefore, we propose EMD-Corr, a visual correspondence-based classifier that (1) re-ranks the top-N (here, N = 50) candidates of kNN using their Earth Mover's Distance (EMD) with the query in a patch embedding space (see Fig. 3) and (2), similarly to kNN, takes the dominant class among the re-ranked top-20 as the predicted label. First, using layer 256 features (8×8×512) (which is the output of backbone ResNet 18 and through layer AvgPool with size 8x8x512), we divide an image into 49 patches, whose embeddings are ∈

R^{512}.Second, for interpretability, in re-ranking, we only use patch-wise EMD instead of a 3-linear combination of image-level cosine distance and patch-level EMD as in [13], which makes it more opaque how a specific image patch contributes to re-ranking. Third, while the original deep patch-wise EMD in Zhao, W.(2021) between two images is defined as the sum over the weighted cosine distances of all 49 × 49 patch pairs, we only use L = 5 pairs as explanations and therefore, only sum over the corresponding 5 flow weights returned by Sinkhorn optimization by Cuturi, M. (2013). We find N = 50 to perform the best among N ∈ 50, 100, 200. We choose L = 5, which is also the most common in nearest-neighbor visualizations by Krizhevsky, A. et al. (2012), Li, H. et al. (2022)..

Calculation of Intersection over Union (IoU) of the result with the doctor's breast cancer X-Ray image localization

Intersection over Union is a popular metric to measure localization accuracy and compute localization errors in object detection models. It calculates the amount of overlap between two bounding boxes - a predicted bounding box and a ground truth bounding box. In detail:

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \qquad (2)$$

In our study, IoU is used for comparison of two binary regions where one region is a region of a breast cancer X-Ray image localized by medical doctors and another one is a region consisting of rectangles resulting by EMD-Corr classifier. For breast cancer X-Ray images localized by medical doctors, we assume that the image background is black and the localization region is white. In a similar way, for the output images, the image background is black and the rectangular regions are white. The result of IoU is the intersection of the white regions of two images.

Result of EMD-Corr shows the label as Malignant or Benign or Normal together with 5 rectangles in the breast cancer X-Ray image. We choose the resulting images given by EMD-Corr and these images will be compared with the breast cancer X-Ray images localized by medical doctors in the training dataset. The evaluation result consists of the data content and the image content. The data content is calculated by the measure of IoU. For each breast cancer X-Ray image resulting from EMD-Corr which issues a rectangular then we use this rectangular evaluate IoU with localization regions by medical doctors. The average of IoU over all rectangles is the result of IoU. On the other side, the image content will be a visual image of rectangular locations together with the image of localization areas of doctors in order one can see the intersections.

Evaluation:-

kNN classifiers improve upon ResNet-18 on Breast Cancer datasets

One of the important goals of training convolutional neural networks is to enable the networks to generalize to variations of the object of interest based on the known examples provided through the images. To further explore this issue for training a convolutional neural network model for the problem of breast cancer detection from mammograms, we compare the results of the kNN model with the Resnet-18 architecture-based classification model on the mammogram image dataset from the Vietnam National Cancer Hospital.

Table 4:- Evaluation results of ResNet-18. kNN, and EMD-Corr

Methods	Average Accuracy	Average Sensitivity	Average Specificity
ResNet-18	80.16	68.39	83.58
kNN	**82.42** (+2.28)	**69.84** (+1.45)	**84.56** (+0.98)
EMD-Corr	82.33 (-0.09)	68.83(-1.02)	84.48 (-0.48)

In Table 4, Average Accuracy, Average Sensitivity, and Average Specificity are the average measures of Malignant, Benign, and Normal. From the evaluation results of the mammogram dataset, kNN has slightly better performance than the Resnet-18 architecture-based classifier. Specifically, the kNN model achieves 82.42 in terms of the accuracy score, 69.84 in terms of the sensitivity score, and 84.56 in terms of the specificity score while the Resnet-18 architecture-based classifier achieves 80.16 in terms of the accuracy score, 68.39 in terms of the sensitivity score, and 83.58 in terms of the specificity score. This means during testing, kNN outperforms ResNet-18 model (e.g. +2.28 for Accuracy, +1.45 for Sensitivity, and +0.98 for Specificity). These results are generated on the basis of both cases using the same backbone. Intuitively, reusing data samples in the training set to decide labels for test cases can benefit from the appearance of out-of-distribution situations in the Vietnam National Cancer Hospital mammogram dataset based on the computation results of the Resnet-18 backbone architecture.

Visual correspondence-based explanations improve kNNand ResNet 18

Some recent studies in Hai Phan, A. Nguyen. (2022), Zhang, C. et al. (2020) have focused on the problem of re-ranking the selected candidate samples of kNN. Accordingly, re-ranking based on applying patch-level similarity assessment between query samples and samples in the training set has shown the ability to improve the quality of image classification. In addition, the patch-level approach is also suitable for semantics in some specific types of image objects, especially with the classification of mammograms to determine the possibility of breast cancer. In this case, the important information is only in a few small image regions with malignant tumors, while the majority of other image regions are healthy. In the case study in Taesiri, M. R.el al. (2022), the authors have shown significant improvements in the candidate samples from kNN based on patch-level similarity assessment on each pair of images for re-ranking. Our research is also inspired by this work and we have evaluated similarity based on the top 5 matching patch pairs instead of the entire 64x64 patch pairs. In the experiment, we re-ranked 50 candidates as the samples of the kNN model’s processing results. The results will be taken according to the majority label in the group has size 14 of the top candidates. The metrics are also calculated similarly to the kNN case. In general, the metrics are slightly lower than the kNN case. This means during testing, EMD-Corr underperforms kNN model (e.g. -0.09 for Accuracy, -1.02 for Sensitivity, and -0.48 for Specificity). On the other side, EMD-Corr outperforms the ResNet 18 (e.g. +2.17 for Accuracy, +0.44 for Sensitivity, and +.09 for Specificity) (see Table 4).In terms of perception, in some cases on mammograms, sometimes small image areas with characteristic lines and contrast levels, such as the nipple area, may have a more dominant effect than the tumor area. Moreover, in images without tumors, semantically, characteristic local locations will not appear. Thus in these cases, it becomes more difficult to evaluate using a small set of the most similar patch pairs.

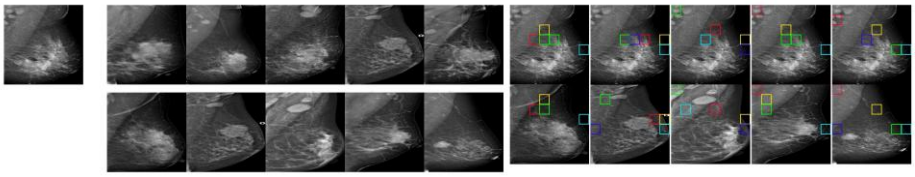
Visual correspondence-based explanations improve the localization of abnormal areas of breast cancer.

For the medical profession to diagnose breast cancer from mammography in the medical profession, it is important to identify the area of the image with malignant tumors on the mammography image. This is the basis for performing a confirmation assessment of the disease condition. Therefore, a task posed when building a mammogram database for breast cancer diagnosis is to label tumor areas on mammograms. When building the dataset, Vietnam National Cancer Hospital radiologists performed tumor contour localization on mammograms to clearly record the boundaries of tumors on mammograms. With the method implemented in this paper, a computational result is to give the most suitable patch pairs when considering the query image and an image in the training data set. The meaning of the patch pairs implies that these two images are similar in the locations of the corresponding patch pairs. Thus, in terms of perception, when considering images with the same cancer, the images have all areas with malignant tumors.

To clarify the above statement, we will compare the image region that the doctor has labeled with the image region corresponding to the determined location of the found patch pairs. If the regions that the doctor has labeled, that is, the malignant regions, have similar locations with the regions corresponding to the calculated patch pairs, it can be understood that the method has evaluated the query image and the image in the training data set as the same because

of the malignant regions in both images. The calculation result of breast cancer diagnosis on mammograms can be explained in the sense that it uses the information in the malignant image region. However, comparing the doctor's area and the image area corresponding to the patch location will have certain limitations. First, the patches are square areas, so these results will be difficult to match the detailed boundaries of the tumor areas, which are zoned by the radiologist. The boundaries have alevel of detail at the pixel level, while the patches will be larger, in this case 64x64 pixels. Second, tumors have different sizes. That means that tumor areas have different areas on different mammograms, while each patch has a fixed area, in this case, a 64x64 square area. In the test cases, the area of the patches will have specific limits, for example, here we will choose 5 pairs of patches (see Fig. 4).

Figure 4:-Explanation of a malignant sample



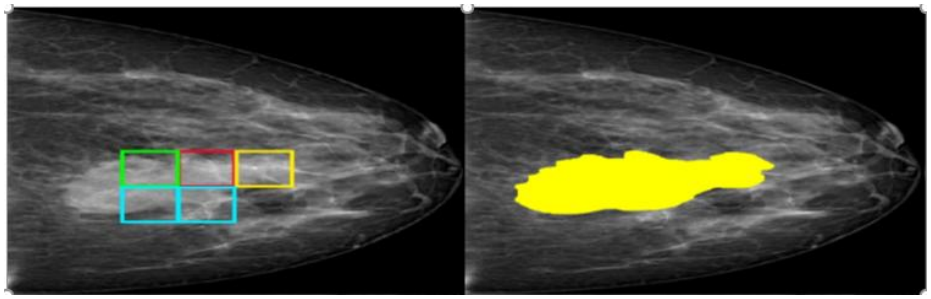
The resulting patch locations are also similar to the labels assigned in the segmentation problem. Here we directly use the image regions bounded by patches as the prediction results as in the case of the segmentation problem. Therefore, we will compare the doctor-labeled region and the image region corresponding to the patch locations based on the application of the segmentation problem's measures. In this study, we use the IoU measure, which is a famous measure that has been applied in many studies of image segmentation problems. In this experiment, we conducted on a dataset of mammograms of cancer cases. Specifically, the training set will include 2696 images and the test set will include 299 images. Each image in the test set willbe calculated by the EMD-Corr method. Thus, in the similarity evaluation step based on using a set of the most suitable patch pairs, we will have the results of the pairs in which each pair includes the query image and the image in the training set. For each query image that includes labeled regions, itsdata corresponds to the segmentation result form. Each case of the image pair will have different patch pairs.

Table 5. Evaluation of localization results

Measures	1st closest sample	2nd closest sample	3rd closest sample	2 closest samples	3 closest samples
Average of IoU	16.71	16.69	17.43	16.70	16.94
Tumor finding rate	79.60	80.60	80.60	83.61	85.62

Table 5 describes the comparison values between the regions labeled by the doctor and the regions determined from the calculated patch pairs. The first row is the average IoU value between the query image results. The second row is the rate of finding the tumor region in the query image. The tumor region is found if the image region resulting from the patch pairs overlaps with the tumor region labeled by the doctor. Accordingly, the values are calculated according to the cases in the columns. In the first column, the values are calculated with the closest sample calculated by EMD-Corr. In the second column, the values are calculated with the second closest sample calculated by EMD-Corr. In the third column, the values are calculated with the third closest sample calculated by EMD-Corr. Inthe fourth column, the values are calculated with the 2 closest samples calculated by EMD-Corr. Inthe fifth column, the values are calculatedwith the 3 closest samples calculated by EMD-Corr. We see that the average IoU values usually fluctuate between 16.69 and 17.43.

Figure5: Comparison between patch locations (left) versus doctor's annotation (right)



This is also expected from the hypothesis due to the different characteristics of the doctor’s tumor labeling region and the result region generated from the most matched patch pairs. More importantly, this is also reflected in the relatively high tumor finding rate in most cases. Although only 5 patch pairs were identified, the tumor finding rate in the last 3 samples alone was 85.62. We see samples of the comparison between patch locations versus the doctor’s annotation in Fig. 5.

Conclusions:-

In this paper, we propose a novel architecture of Visual Correspondence Based Explanations to improve Convolutional Neural Networks for Classification of Mammograms of self-interpretable image classifiers that first explain, and then predict by harnessing the visual correspondences between a query breast cancer X-Ray image and exemplars. We compare the results of the kNN model with the Resnet-18 architecture-based classification model on the mammogram image dataset from the Vietnam National Cancer Hospital and the result shows that kNN classifiers improve upon ResNet-18 on the Breast Cancer dataset. Our proposed model EMD-Corr is a Visual correspondence-based explanation that improveskNN robustness further and EMD-Corrcan improve the localization of abnormal areas of breast cancer by showing patch locations of cancer regions in Breast Cancer images in comparison with the doctor's annotation.

Acknowledgement

This study is in part granted by the project “Research on XAI and applications in Healthcare”, 2024-2025, Thang Long University, Vietnam.Thanks to the doctors of the Vietnam National Cancer Hospital for providingmammograms for this research by the project project KC-4.0-18/19-25.

References

1. G. Casalino, G. Castellano, K. Kaczmarek-Majer and G. Zaza,(2024) "Prototype-Based Explanations to Improve Understanding of Unsupervised Datasets," 2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Yokohama, Japan, pp. 1-8.
2. Shin, Y.-M., Kim, S.-W., Yoon, E.-B., & Shin, W.-Y. (2022).Prototype-Based Explanations for Graph Neural Networks (Student Abstract). Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 13047-13048.
3. Agarwal, C. and Nguyen, A. (2020). Explaining image classifiers by removing input features using generative models. In Proceedings of the Asian Conference on Computer Vision.
4. Alqaraawi, A., Schuessler, M., Weis, P., Costanza, E., and Berthouze, N. (2020). Evaluating saliency map explanations for convolutional neural networks: a user study. In Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 275–285.
5. Phuong, N. H., Toan, H. M., Linh, L. T., Cuong, N. N., and Hanh, B. M. (2022). Weakly supervised localization of the abnormal regions in breast cancer x-ray images using patches classification. In Biomedical and Other Applications of Soft Computing, pp. 203–212. Springer.

6. Karim, M. R., Jiao, J., Doehmen, T., Cochez, M., Beyan, O., Rebholz-Schuhmann, D., and Decker, S. (2021). Deepkneeexplainer: explainable knee osteoarthritis diagnosis from radiographs and magnetic resonance imaging. IEEE Access, 9:39757–39780.
7. Kim, K. H., Koo, H.-W., Lee, B.-J., Yoon, S.-W., and Sohn, M.-J. (2021). Cerebral hemorrhage detection and localization with medical imaging for cerebrovascular disease diagnosis and treatment using explainable deep learning. Journal of the Korean Physical Society, 79(3):321–327.
8. Salahuddin, Z., Woodruff, H. C., Chatterjee, A., and Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. Computers in biology and medicine, 140: 105111.
9. Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., Heacock, L., Kim, S. G., Moy, L., Cho, K., et al. (2021). An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. Medical image analysis, 68:101908.
10. Taesiri, M. R., Nguyen, G., and Nguyen, A. (2022). Visual correspondence-based explanations improve ai robustness and human-AI team accuracy. Advances in Neural Information Processing Systems, 35:34287–34301.
11. G. Casalino, G. Castellano, K. Kaczmarek-Majer and G. Zaza, (2024). "Prototype-Based Explanations to Improve Understanding of Unsupervised Datasets," 2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Yokohama, Japan, pp. 1-8.
12. Phuong N. H., Nguyen D. D., Nguyen V. D., Ha M. T., Nguyen K. D., and Dao V. T. (2024). A Classification System of Mammograms Based on Convolutional Neural Networks, In Nguyen Hoang Phuong, Nguyen ThịHuyen Chau, VladikKreinovich (eds.), Machine Learning and Other Soft Computing Techniques: Biomedical and Related Applications, Springer, pp. 149-158
13. Zhao, W., Rao, Y., Wang, Z., Lu, J., and Zhou, J. (2021).Towards interpretable deep metric learning with structural matching. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9887–9896.
14. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
15. Li, H., Wang, Y., Wu, A., Wei, H., and Qu, H. (2022). Structure-aware visualization retrieval. In CHI Conference on Human Factors in Computing Systems, pp. 1–14.
16. Hai Phan, A. Nguyen. (2022). Deepface-emd: Re-ranking using patch-wise earth mover's distance improves out-of-distribution face identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
17. Zhang, C., Cai, Y., Lin, G., and Shen, C. (2020). Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12203–12213.
18. Cuturi, M. (2013).Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems, 26.