A The INTRPRT guideline: A case study

To contextualize the *INTRPRT guideline*, we present case studies to demonstrate the envisioned use. Because none of the surveyed papers take into account all aspects mentioned in the *INTRPRT guideline*, we include three published papers that are representative of different parts of the *INTRPRT guideline*. Case 1 (C1): Xie et al. [1] focused on the formative user research stage to determine physicians' needs when exploring and understanding an Artificial Intelligence (AI)-generated analysis report in the context of chest X-rays diagnosis. Through surveys, a co-design process, and user evaluations, different features of the explanations that physicians expect when they interact with the system were identified. Case 2 (C2): Motivated by the fact that the radiologists characterize breast masses according to Breast Imaging Reporting and Data System (BI-RADS) (public evidence), Kim et al. [2] directly encoded the BI-RADS characteristics of breast masses to build a deep learning model for mass classification. Case 3 (C3): Sabol et al. [3] created an explainable system for colorectal cancer diagnosis from histopathological images by providing human-friendly explanations for a certain prediction. The validation of the system included the assessment of transparency through a human factors evaluation with 14 pathologists who interacted with the system through a graphical user interface.

We identify which aspects within the three cases follow the *INTRPRT guideline*:

G1: Specify the clinical scenario, constraints, requirements, and end users.
In C1, Machine Learning (ML) designers determined the requirements in current clinical practice by conducting paired surveys on physicians and radiologists, where the latter provide a diagnosis and the former have to interpret the results. The target users (referring physicians) were identified within the context in which the system will be used.

G2: Justify the choice of transparency and determine the level of evidence.
In C1, to inform the design of AI-enabled chest X-ray analysis, iteratively developed evidence was generated by involving potential end users of the system through interactions with low and high-fidelity prototypes. The design choices to build the prototypes were based on the initial needs identified with a survey of explanations between referring physicians and radiologists.
In C2, the acBI-RADS standard serves as public evidence because it is widely accepted and applied among clinicians for mass classification.

G3: Clarify how the model follows the justification of transparency.
In C2, the public evidence, namely BI-RADS characteristics of breast masses, are encoded as features and concatenated with deep encoded image features for the final tissue classification. As a result, BI-RADS characteristics are explicitly extracted and have direct impact in the final decision making.

G4: Determine how to communicate with end users.
In C1, a graphical interface was implemented as a high-fidelity prototype. The interface presented one patient's case at a time, displaying the Chest X-ray (CXR) image, the significant observations generated from an AI model (as textual labels), and eight explanations features that were manually generated, such as highlighting evidence towards a specific diagnosis in the image, probabilities for each possible conclusion or comparisons with previous patients' cases.
C3 created a graphical user interface that showed the original image of the Whole Slide Image (WSI) and the corresponding label map with a color code for different tissue types. Pathologists can examine an arbitrary area of the WSI by clicking on the desired area. Subsequently, the outcomes of the system were displayed, including the prediction result with a semantical explanation. Besides, a visualization of the training image most similar to the current one as well as training images with other tissue types are presented as references and support for the prediction.

G5: Report task performance of the ML systems.
In C2, quantitative results of the method with/without implementing BI-RADS characteristics are presented on a public mammogram database. Both accuracy and Area under the ROC Curve (AUC) are used as evaluation metrics of the model itself.
C3 reported the accuracy of the classifier in a class-balanced dataset of tissue slides.

G6: Assess correctness and human factors of system transparency.
C3 involved 14 pathologists to evaluate four human factors of the system: usefulness, level of detail, reliability, and experience quality. Pathologists were first asked to examine 20 arbitrary areas in a

graphical user interface and evaluate the prediction outcome. At the end of the experiment session, every participant was asked to fill out a questionnaire based on their perception of the system.

Lastly, we present examples on where and how authors in the three cases could integrate the use of the *INTRPRT guideline* in their articles:

C1: At the beginning of the Method section: "We were primarily concerned on the user-centered iterative design of a proof-of-concept system prototype of an AI-based medical image analysis tool with different explanation features to assist physicians. The outcome of the user-centered empirical formative research approach includes recommendations and insights that may benefit future development of explainable algorithms for medical image analysis."

C2: In the Method section: "We followed the widely accepted BI-RADS standard for breast mass assessment as a readily interpretable backbone as a basis for our deep network." In the Results section: "The verification of the system included a task performance comparison against a model without the interpretable BI-RADS component on a public database. Additional qualitative visualizations showed relevant areas where the model exploited more information."

C3: At the end of the Introduction section: "We developed and discussed the mathematical structure of a classifier that provides different outputs to explain the plausibility of the decision. A quantitative evaluation of the system's performance was conducted on a public database. We also report the system's acceptability assessed through a user study with 14 pathologists."

B Search strategy

We use the following search term for PubMed and EMBASE to screen titles, abstracts, and keywords of all available records:

("interpretable" OR "explainable" OR "interpretability" OR "explainability" OR "interpretation" OR "explanation" OR "interpreting" OR "explaining" OR "interpret" OR "explain") AND ("artificial intelligence" OR "deep learning" OR "machine learning" OR "neural network") AND ("image" OR "imaging") AND ("healthcare" OR "health care" OR "clinical" OR "medical")

In addition to the above search terms, Compendex offers "controlled terms" to better locate desired articles. We first filter all records with the following search term for titles, abstracts, and keywords:

(interpret* OR explain* OR "explanation")

Then we use "controlled terms" to further filter all the remaining records:

("artificial intelligence" OR "neural networks" OR "machine learning" OR "deep learning" OR "deep neural networks" OR "convolutional neural networks" OR "learning systems" OR "supervised learning" OR "network architecture") AND ("medical imaging" OR "medical image processing" OR "medical computing" OR "tissue")

For screening in all three databases, we also exclude articles with ("survey" OR "review") in the tile and ("workshop") in the title and abstract.

C Details of screening and full-text review

The initial search resulted in 2508 records, and after removal of duplicates, 1731 unique studies were included for screening. During screening, 1514 articles were excluded because they 1) were not transparent methods (n=947); 2) were not imaging ML methods (n=422); 3) only had simple visualization explanations (n=129) and 4) were not for medical problems (n=19). We found unsubstantiated claims around transparency with only simple visualizations such as Class Activation Maps (CAMs) widely occur in ML methods or using existing transparent methods. As a result, we excluded them and focused on articles with other transparent ML methods. The remaining 217 articles were included in full-text review. In total, 149 records were further excluded because they 1) used exactly the same transparency mechanism as previously proposed work for natural image problems (n=48); 2) were not transparent methods (n=43); 3) were not long articles and therefore could not be analyzed with the required detail (n=24); 4) were not imaging ML methods (n=21); 5) were not for medical problems (n=5); 6) did not have available full text (n=5); 7) were repeated works (n=3). The criterion for long

articles we applied was single-column articles longer or equal to 8 pages or double-column articles longer or equal to 6 pages, excluding reference pages.

D Data extraction

Table 1 describes in detail our data extraction approach for the studies included in this review. Tables 2 and 3 present the details of each study included in the review in the design preparation and implementation, respectively.

Table 1: Data extraction strategy related to the six themes.

Theme	Item	Description
Incorporation	Clinician engineering team	Any clinical stakeholders part of the study team and author list
	Formative user research	Any technique to understand the target population
Target	End users	Users of the systems
Prior	Justification of transparency	Description of the choice of transparency
	Prior type	Computer vision / clinical knowledge prior
Task	Inputs & outputs	Task inputs & outputs
	Task difficulty	Routine / super-human tasks
Interpretability	Technical mechanism	Explicit transparency technique
	Transparency type	Interpretable (provides its own explanations) / Explainable (needs post-hoc explanations)
Reporting	Metrics	Task performance evaluation and transparency assessment metrics
	Transparency performance	Performance against comparable baseline models
	Incorporation performance	Performance of human-AI incorporation against AI alone
	Human subjects	Number of end users involved in evaluation

Table 2: A summary of transparency design preparation details of each study reviewed. The definition of each column is the same as in Table 1.

Study ID	Author team	End users	Task type	Formative user research	Task difficulty	Clear justification of transparency	Prior type
[4]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[5]	Yes	Decision providers	Prediction	None	Routine	No	Computer vision
[6]	Yes	Decision providers	Prediction	None	Routine	No	Clinical knowledge
[7]	No	Not specified	Segmentation	None	Routine	Yes	Computer vision
[8]	No	Not specified	Prediction	None	Routine	Yes	Computer vision
[9]	Yes	Decision providers	Prediction	None	Super-human	Yes	Clinical knowledge
[10]	Yes	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge

[11]	Yes	Not specified	Prediction	None	Routine	No	Computer vision
[12]	No	Decision providers	Prediction	None	Routine	Yes	Computer vision
[13]	Yes	Decision providers	Segmentation	None	Routine	No	Clinical knowledge
[14]	Cannot tell	Not specified	Prediction	None	Cannot tell	Yes	Computer vision
[14]	Yes	Not specified	Segmentation	None	Routine	Yes	Computer vision
[15]	Yes	Decision providers	Image grouping	None	Routine	Yes	Clinical knowledge
[16]	No	Not specified	Prediction	None	Routine	Yes	Computer vision
[17]	No	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[18]	Yes	Not specified	Prediction	None	Routine	No	Computer vision
[19]	Yes	Decision providers	Prediction	None	Super-human	Yes	Computer vision
[20]	Yes	Decision providers	Prediction	None	Routine	Yes	Computer vision
[21]	Yes	Not specified	prediction	None	Cannot tell	Yes	Computer vision
[22]	Yes	Not specified	Segmentation	None	Routine	Yes	Computer vision
[23]	No	Decision providers	Prediction	None	Routine	Yes	Computer vision
[24]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[25]	Yes	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[26]	No	Not specified	Prediction	None	Routine	Yes	Computer vision
[27]	Yes	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[28]	Cannot tell	Not specified	Prediction	None	Routine	Yes	Clinical knowledge
[29]	No	Not specified	Prediction	None	Cannot tell	Yes	Clinical knowledge
[30]	No	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[31]	No	Not specified	Predictions	None	Routine	No	Clinical knowledge
[32]	Yes	Not specified	Prediction	None	Routine	Yes	Clinical knowledge
[33]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[34]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[35]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[36]	Yes	Decision providers	Segmentation	None	Routine	Yes	Computer vision
[37]	Yes	Not specified	Prediction	None	Super-human	Yes	Computer vision
[38]	No	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[2]	No	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[39]	Yes	Not specified	prediction	None	Routine	Yes	Clinical knowledge

[3]	Yes	Decision providers	Prediction	None	Routine	Yes	Computer vision
[40]	No	Not specified	Segmentation	None	Routine	No	Computer vision
[41]	No	Not specified	prediction	None	Routine	No	Computer vision
[42]	No	Not specified	Prediction	None	Routine	Yes	Clinical knowledge
[43]	No	Not specified	Prediction	None	Routine	Yes	Clinical knowledge
[44]	No	Not specified	prediction	None	Routine	Yes	Computer vision
[45]	Yes	Not specified	Segmentation	None	Routine	Yes	Clinical knowledge
[46]	Yes	Decision providers	Prediction	None	Routine	Yes	Computer vision
[47]	Yes	Not specified	Segmentation	None	Routine	Yes	Computer vision
[48]	No	Decision providers	prediction	None	Routine	Yes	Computer vision
[49]	No	Decision providers	Prediction	None	Routine	No	Computer vision
[50]	No	Decision providers	Prediction	None	Super-human	Yes	Clinical knowledge
[51]	Yes	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[52]	Yes	Decision providers	Prediction	None	Routine	No	Computer vision
[53]	Yes	Decision providers	prediction	None	Routine	No	Clinical knowledge
[54]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[55]	Yes	Not specified	Segmentation	None	Routine	Yes	Computer vision
[56]	Yes	Decision providers	Super-resolution	None	Super-human	Yes	Computer vision
[57]	Yes	Not specified	Prediction	None	Routine	Yes	Clinical knowledge
[58]	No	Decision providers	Prediction	None	Routine	Yes	Computer vision
[59]	No	Decision providers	prediction	None	Routine	Yes	Computer vision
[60]	Yes	Decision providers	Prediction	None	Routine	No	Clinical knowledge
[61]	Yes	Decision providers	Prediction	None	Routine	No	Computer vision
[62]	Yes	Decision providers	Prediction	None	Cannot tell	No	Clinical knowledge
[63]	Yes	Not specified	Prediction	None	Routine	Yes	Computer vision
[64]	Yes	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[65]	Yes	Decision providers	Prediction	None	Routine	No	Computer vision
[66]	No	Decision providers	Prediction	None	Routine	Yes	Clinical knowledge
[67]	No	Not specified	Prediction	None	Routine	Yes	Computer vision
[68]	No	Not specified	Prediction	None	Routine	Yes	Clinical knowledge

Table 3: A summary of transparency design implementation and validation details of each study reviewed. The definition of each column is the same as in Table 1.

Study ID	Transparency type	Technical mechanism	Transparency assessment	Transparency performance	Incorporation performance	Human subjects
[4]	Interpretable	Attention	Spearman coefficients	Better	None	None
[5]	Interpretable	Attention	None	Better	None	None
[6]	Interpretable	Clustering	None	Better	None	None
[7]	Explainable	activation maximization	None	None	None	None
[8]	Explainable	Input perturbation	Visualization	None	None	None
[9]	Interpretable	Automatic extraction of clinically important features	Similarity	None	None	None
[10]	Interpretable	Automatic extraction of clinically important features	Visualization	None	None	None
[11]	Interpretable	Attention	None	Better	None	None
[12]	Explainable	Saliency maps	Level of trust	None	None	8
[13]	Interpretable	Hand-crafted feature computation	Confusion matrix	None	None	None
[14]	Interpretable	Network architecture pruning	Kappa, R^2 matrix	Better	None	None
[69]	Interpretable	Attention	Visualization	Better	None	None
[15]	Interpretable	Clustering	None	Better	Better	1
[16]	Interpretable	Attention	None	None	None	None
[17]	Interpretable	Decoded by simple transparent models	None	Comparable	None	None
[18]	Interpretable	Latent variable evolution	None	None	None	None
[19]	Interpretable	Uncertainty estimation / confidence calibration	None	Better	None	None
[20]	Interpretable	Ranking of features	C-index, visualization	Better	None	None
[21]	Interpretable	Attention	Pearson correlation, visualization	Better	None	None
[22]	Interpretable	Attention	None	Better	None	None
[23]	Interpretable	Attention	None	Better	None	None
[24]	Explainable	Input corruption	Visualization	None	None	None
[25]	Interpretable	Decoded by simple transparent models	T-SNE visualization	Better	None	None
[26]	Explainable	Visualization	Visualization	Better	None	None
[27]	Interpretable	Automatic extraction of clinically important features	None	None	None	None

[28]	Interpretable	Network structure modification	None	Better	None	None
[29]	Interpretable	Causal inference	Causal relationships, visualization	None	None	None
[30]	Interpretable	Network structure modification	Visualization	Better	None	None
[31]	Interpretable	Automatic extraction of clinically important features	None	Better	None	None
[32]	Interpretable	Automatic extraction of clinically important features	None	None	None	None
[33]	Interpretable	Attention	C-index	Better	None	None
[34]	Explainable	Class activation mapping	Visualization	Better	None	None
[35]	Explainable	Decoded by simple transparent models	Correctness, completeness and compactness	None	None	None
[36]	Interpretable	Clustering	None	None	None	None
[37]	Interpretable	Decoded by simple transparent models	Faithfulness, relevance scores	None	None	None
[38]	Interpretable	Hand-crafted feature computation	None	Better	None	None
[2]	Interpretable	Automatic extraction of clinically important features	Visualization	Better	None	None
[39]	Interpretable	Relation analysis	None	Better	None	None
[3]	Interpretable	Uncertainty estimation / confidence calibration	Certainty rate & error, User perception	Comparable	None	14
[40]	Explainable	Attention	Deletion metric	None	None	None
[41]	Interpretable	Clustering	Visualization	Better	None	None
[42]	Interpretable	Attention with domain knowledge	None	Better	None	None
[43]	Interpretable	Decoded by simple transparent models	None	Better	None	None
[44]	Interpretable	Representative feature extraction	None	Better	None	None
[45]	Interpretable	Prior knowledge into latent space	Agreement	Better	None	2
[46]	Interpretable	Image retrieval	Visualization	Better	None	None

[47]	Explainable	Decoded by simple transparent models	Visualization	Comparable	None	None
[48]	Interpretable	Perturbation analysis	Visualization	Better	None	None
[49]	Explainable	Feature importance analysis	Visualizations	None	None	None
[50]	Interpretable	Network structure modification	Visualization	Comparable	None	None
[51]	Interpretable	Relation analysis	R^2 matrix	None	None	None
[52]	Interpretable	Automatic extraction of clinically important features	Kappa	Better	None	None
[53]	Interpretable	Causal inference	None	Comparable	None	None
[54]	Interpretable	Attention	R^2 matrix, visualization	Better	None	None
[55]	Interpretable	Attention	Visualization	Better	None	None
[56]	Interpretable	Uncertainty estimation / confidence calibration	Reliability	Better	None	None
[57]	Interpretable	Causal inference	None	None	None	None
[58]	Explainable	Perturbation analysis	None	None	None	None
[59]	Interpretable	Decoded by simple transparent models	Visualization	None	None	None
[60]	Interpretable	Automatic extraction of clinically important features	None	None	None	None
[61]	Explainable	Visualization	Visualization	None	None	None
[62]	Explainable	Relation analysis	Visualization	None	None	None
[63]	Interpretable	Attention	Visualization	Better	None	None
[64]	Interpretable	Decoded by simple transparent models	Visualization	Better	None	None
[65]	Interpretable	Attention	Visualization	Better	None	None
[66]	Interpretable	Concept's importance analysis	R^2 matrix	None	None	None
[67]	Explainable	Backpropagation guidance	Visualization, Kendall's Tau metric	None	None	None
[68]	Explainable	Perturbation analysis	Shapley values	None	None	None

References

[1] Xie Y, Chen M, Kao D, Gao G, Chen X. CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; 2020. p. 1-13.

[2] Kim ST, Lee JH, Lee H, Ro YM. Visually interpretable deep network for diagnosis of breast masses on mammograms. Physics in Medicine and Biology. 2018;63(23):235025. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L628170713&from=exporthttp://dx.doi.org/10.1088/1361-6560/aaef0a.

[3] Sabol P, Sinčák P, Hartono P, Kočan P, Benetinová Z, Blichárová A, et al. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. Journal of Biomedical Informatics. 2020;109. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2007460563&from=exporthttp://dx.doi.org/10.1016/j.jbi.2020.103523.

[4] Abdel Magid S, Jang WD, Schapiro D, Wei D, Tompkin J, Sorger PK, et al. Channel Embedding for Informative Protein Identification from Highly Multiplexed Images. 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2020, October 4, 2020 - October 8, 2020. 2020;12265 LNCS(PG - 3-13):3-13. Available from: http://dx.doi.org/10.1007/978-3-030-59722-1_1NS-.

[5] Afshar P, Naderkhani F, Oikonomou A, Rafiee MJ, Mohammadi A, Plataniotis KN. MIXCAPS: A capsule network-based mixture of experts for lung nodule malignancy prediction. Pattern Recognition. 2021;116(PG -). Available from: http://dx.doi.org/10.1016/j.patcog.2021.107942NS-.

[6] Codella NCF, Lin CC, Halpern A, Hind M, Feris R, Smith JR. Collaborative human-AI (CHAI): Evidence-based interpretable melanoma classification in dermoscopic images. 1st International Workshop on Machine Learning in Clinical Neuroimaging, MLCN 2018, 1st International Workshop on Deep Learning Fails, DLF 2018, and 1st International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC. 2018;11038 LNCS(PG - 97-105):97-105. Available from: http://dx.doi.org/10.1007/978-3-030-02628-8_11NS-.

[7] Couteaux V, Nempont O, Pizaine G, Bloch I. Towards interpretability of segmentation networks by analyzing deepDreams. 2nd International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2019, and the 9th International Workshop on Multimodal Learning for Clinical Decision Support, ML-CDS 2019, held in conjunction with the 22nd Interna. 2019;11797 LNCS(PG - 56-63):56-63. Available from: http://dx.doi.org/10.1007/978-3-030-33850-3_7NS-.

[8] de Sousa IP, Vellasco MMBR, da Silva EC. Approximate Explanations for Classification of Histopathology Patches. Workshops of the 20th Joint European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD, September 14, 2020 - September 18, 2020. 2020;1323(PG - 517-526):517-26. Available from: http://dx.doi.org/10.1007/978-3-030-65965-3_35NS-.

[9] Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nature Communications. 2021;12(1). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2010776995&from=exporthttp://dx.doi.org/10.1038/s41467-021-21896-9.

[10] Dong Y, Wan J, Wang X, Xue J, Zou J, He H, et al. A Polarization-Imaging-Based Machine Learning Framework for Quantitative Pathological Diagnosis of Cervical Precancerous Lesions. IEEE Transactions on Medical Imaging. 2021. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L635538309&from=exporthttp://dx.doi.org/10.1109/TMI.2021.3097200.

[11] Fan M, Chakraborti T, Chang EIC, Xu Y, Rittscher J. Microscopic Fine-Grained Instance Classification Through Deep Attention. 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2020, October 4, 2020 - October 8, 2020. 2020;12265 LNCS(PG - 490-499):490-9. Available from: http://dx.doi.org/10.1007/978-3-030-59722-1_47NS-.

[12] Folke T, Yang SCH, Anderson S, Shafto P. Explainable AI for medical imaging: Explaining pneumothorax diagnoses with Bayesian teaching. Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III 2021, April 12, 2021 - April 16, 2021. 2021;11746(PG - The Society of Photo-Optical Instrumentation Engineers (SPIE)):The Society of Photo-Optical Instrumentation Engin. Available from: http://dx.doi.org/10.1117/12.2585967NS-.

[13] Giannini V, Rosati S, Regge D, Balestra G. Texture features and artificial neural networks: A way to improve the specificity of a CAD system for multiparametric MR prostate cancer. 14th Mediterranean Conference on Medical and Biological Engineering and Computing, MEDICON 2016, March 31, 2016 - April 2, 2016. 2016;57(PG - 296-301):296-301. Available from: http://dx.doi.org/10.1007/978-3-319-32703-7_59NS-.

[14] Graziani M, Lompech T, Muller H, Depeursinge A, Andrearczyk V. Interpretable CNN Pruning for Preserving Scale-Covariant Features in Medical Imaging. 3rd International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2020, the 2nd International Workshop on Medical Image Learning with Less Labels and Imperfect Data, MIL3ID 2020, and the 5th International Workshop o. 2020;12446 LNCS(PG - 23-32):23-32. Available from: http://dx.doi.org/10.1007/978-3-030-61166-8_3NS-.

[15] Guo X, Yu Q, Li R, Alm CO, Calvelli C, Shi P, et al. Intelligent medical image grouping through interactive learning. International Journal of Data Science and Analytics. 2016;2(3-4 PG - 95-105):95-105. Available from: http://dx.doi.org/10.1007/s41060-016-0021-2NS-.

[16] An F, Li X, Ma X. Medical Image Classification Algorithm Based on Visual Attention Mechanism-MCNN. Oxidative Medicine and Cellular Longevity. 2021;2021. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2011217895&from=exporthttp://dx.doi.org/10.1155/2021/6280690.

[17] Barata C, Celebi ME, Marques JS. Explainable skin lesion diagnosis using taxonomies. Pattern Recognition. 2021;110(PG -). Available from: http://dx.doi.org/10.1016/j.patcog.2020.107413NS-.

[18] Biffi C, Oktay O, Tarroni G, Bai W, De Marvao A, Doumou G, et al. Learning interpretable anatomical features through deep generative models: Application to cardiac remodeling. 21st International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2018, September 16, 2018 - September 20, 2018. 2018;11071 LNCS(PG - 464-471):464-71. Available from: http://dx.doi.org/10.1007/978-3-030-00934-2_52NS-.

[19] Carneiro G, Zorron Cheng Tao Pu L, Singh R, Burt A. Deep learning uncertainty and confidence calibration for the five-class polyp classification from colonoscopy. Medical Image Analysis. 2020;62. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2005189093&from=exporthttp://dx.doi.org/10.1016/j.media.2020.101653.

[20] Hao J, Kosaraju SC, Tsaku NZ, Song DH, Kang M. PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2020;25:355-66. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L630059597&from=export.

[21] He S, Pereira D, David Perez J, Gollub RL, Murphy SN, Prabhu S, et al. Multi-channel attention-fusion neural network for brain age estimation: Accuracy, generality, and interpretation with 16,705 healthy MRIs across lifespan. Medical Image Analysis. 2021;72. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2012117928&from=exporthttp://dx.doi.org/10.1016/j.media.2021.102091.

[22] Hou B, Kang G, Xu X, Hu C. Cross Attention Densely Connected Networks for Multiple Sclerosis Lesion Segmentation. 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, November 18, 2019 - November 21, 2019. 2019;(PG - 2356-2361):2356-61. Available from: http://dx.doi.org/10.1109/BIBM47256.2019.8983149NS-.

[23] Huang Y, Chung ACS. Evidence localization for pathology images using weakly supervised learning. 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2019, October 13, 2019 - October 17, 2019. 2019;11764 LNCS(PG - 613-621):613-21. Available from: http://dx.doi.org/10.1007/978-3-030-32239-7_68NS-.

[24] Li X, Dvornek NC, Zhuang J, Ventola P, Duncan JS. Brain biomarker interpretation in ASD using deep learning and fMRI. 21st International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2018, September 16, 2018 - September 20, 2018. 2018;11072 LNCS(PG - 206-214):206-14. Available from: http://dx.doi.org/10.1007/978-3-030-00931-1_24NS-.

[25] Li Y, Chen J, Xue P, Tang C, Chang J, Chu C, et al. Computer-Aided Cervical Cancer Diagnosis Using Time-Lapsed Colposcopic Images. IEEE Transactions on Medical Imaging. 2020;39(11):3403-15. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L631763414&from=exporthttp://dx.doi.org/10.1109/TMI.2020.2994778.

[26] Liao W, Zou B, Zhao R, Chen Y, He Z, Zhou M. Clinical Interpretable Deep Learning Model for Glaucoma Diagnosis. IEEE Journal of Biomedical and Health Informatics. 2020;24(5):1405-12. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L631748506&from=exporthttp://dx.doi.org/10.1109/JBHI.2019.2949075.

[27] Lin Y, Wei L, Han SX, Aberle DR, Hsu W. EDICNet: An end-to-end detection and interpretable malignancy classification network for pulmonary nodules in computed tomography. Medical Imaging 2020: Computer-Aided Diagnosis, February 16, 2020 - February 19, 2020. 2020;11314(PG - The Society of Photo-Optical Instrumentation Engineers (SPIE)):The Society of Photo-Optical Instrumentation Engin. Available from: http://dx.doi.org/10.1117/12.2551220NS-.

[28] Liu J, Wang W, Guan T, Zhao N, Han X, Li Z. Ultrasound Liver Fibrosis Diagnosis Using Multi-indicator Guided Deep Neural Networks. 10th International Workshop on Machine Learning in Medical Imaging, MLMI 2019 held in conjunction with the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2019, October 13, 2019 - October 13, 2019. 2019;11861 LNCS(PG - 230-237):230-7. Available from: http://dx.doi.org/10.1007/978-3-030-32692-0_27NS-.

[29] Liu Y, Li Z, Ge Q, Lin N, Xiong M. Deep Feature Selection and Causal Analysis of Alzheimer's Disease. Frontiers in Neuroscience. 2019;13. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L629992085&from=exporthttp://dx.doi.org/10.3389/fnins.2019.01198.

[30] Liu Y, Zhang F, Chen C, Wang S, Wang Y, Yu Y. Act Like a Radiologist: Towards Reliable Multi-view Correspondence Reasoning for Mammogram Mass Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021;(PG -). Available from: http://dx.doi.org/10.1109/TPAMI.2021.3085783NS-.

[31] Loveymi S, Dezfoulian MH, Mansoorizadeh M. Generate Structured Radiology Report from CT Images Using Image Annotation Techniques: Preliminary Results with Liver CT. Journal of Digital Imaging. 2020;33(2 PG - 375-390):375-90. Available from: http://dx.doi.org/10.1007/s10278-019-00298-wNS-.

[32] MacCormick IJC, Williams BM, Zheng Y, Li K, Al-Bander B, Czanner S, et al. Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. PLoS ONE. 2019;14(1). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L625837308&from=exporthttp://dx.doi.org/10.1371/journal.pone.0209409.

[33] Morvan L, Nanni C, Michaud AV, Jamet B, Bailly C, Bodet-Milin C, et al. Learned Deep Radiomics for Survival Analysis with Attention. 3rd International Workshop on Predictive Intelligence in Medicine, PRIME 2020, held in conjunction with the Medical Image Computing and Computer Assisted Intervention, MICCAI 2020, October 8, 2020 - October 8, 2020. 2020;12329 LNCS(PG - 35-45):35-45. Available from: http://dx.doi.org/10.1007/978-3-030-59354-4_4NS-.

[34] Shinde S, Chougule T, Saini J, Ingalhalikar M. HR-CAM: Precise localization of pathology using multi-level learning in CNNS. 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2019, October 13, 2019 - October 17, 2019. 2019;11767 LNCS(PG - 298-306):298-306. Available from: http://dx.doi.org/10.1007/978-3-030-32251-9_33NS-.

[35] Silva W, Fernandes K, Cardoso MJ, Cardoso JS. Towards complementary explanations using deep neural networks. 1st International Workshop on Machine Learning in Clinical Neuroimaging, MLCN 2018, 1st International Workshop on Deep Learning Fails, DLF 2018, and 1st International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC. 2018;11038 LNCS(PG - 133-140):133-40. Available from: http://dx.doi.org/10.1007/978-3-030-02628-8_15NS-.

[36] Janik A, Dodd J, Ifrim G, Sankaran K, Curran K. Interpretability of a deep learning model in the application of cardiac MRI segmentation with an ACDC challenge dataset. Medical Imaging 2021: Image Processing, February 15, 2021 - February 19, 2021. 2021;11596(PG - The Society of Photo-Optical Instrumentation Engineers (SPIE)):The Society of Photo-Optical Instrumentation Engin. Available from: http://dx.doi.org/10.1117/12.2582227NS-.

[37] Khaleel M, Tavanapong W, Wong J, Oh J, De Groen P. Hierarchical visual concept interpretation for medical image classification. 34th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2021, June 7, 2021 - June 9, 2021. 2021;2021-June(PG - 25-30):25-30. Available from: http://dx.doi.org/10.1109/CBMS52027.2021.00012NS-.

[38] Kim ST, Lee H, Kim HG, Ro YM. ICADx: Interpretable computer aided diagnosis of breast masses. Medical Imaging 2018: Computer-Aided Diagnosis, February 12, 2018 - February 15, 2018. 2018;10575(PG - DECTRIS Ltd.; The Society of Photo-Optical Instrumentation Engineers (SPIE)):DECTRIS Ltd.; The Society of Photo-Optical Instrum. Available from: http://dx.doi.org/10.1117/12.2293570NS-.

[39] Kunapuli G, Varghese BA, Ganapathy P, Desai B, Cen S, Aron M, et al. A Decision-Support Tool for Renal Mass Classification. Journal of Digital Imaging. 2018;31(6):929-39. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L625181034&from=exporthttp://dx.doi.org/10.1007/s10278-018-0100-0.

[40] Saleem H, Shahid AR, Raza B. Visual interpretability in 3D brain tumor segmentation network. Computers in Biology and Medicine. 2021;133. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2011734982&from=exporthttp://dx.doi.org/10.1016/j.compbiomed.2021.104410.

[41] Sari CT, Gunduz-Demir C. Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images. IEEE Transactions on Medical Imaging. 2019;38(5 PG - 1139-1149):1139-49. Available from: http://dx.doi.org/10.1109/TMI.2018.2879369NS-.

[42] Shahamat H, Saniee Abadeh M. Brain MRI analysis using a deep learning based evolutionary approach. Neural Networks. 2020;126:218-34. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2005472077&from=exporthttp://dx.doi.org/10.1016/j.neunet.2020.03.017.

[43] Shen T, Wang J, Gou C, Wang FY. Hierarchical Fused Model with Deep Learning and Type-2 Fuzzy Learning for Breast Cancer Diagnosis. IEEE Transactions on Fuzzy Systems. 2020;28(12 PG - 3204-3218):3204-18. Available from: http://dx.doi.org/10.1109/TFUZZ.2020.3013681NS-.

[44] Li J, Shi H, Hwang KS. An explainable ensemble feedforward method with Gaussian convolutional filter. Knowledge-Based Systems. 2021;225(PG -). Available from: http://dx.doi.org/10.1016/j.knosys.2021.107103NS-.

[45] Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation. IEEE Transactions on Medical Imaging. 2018;37(2 PG - 384-395):384-95. Available from: http://dx.doi.org/10.1109/TMI.2017.2743464NS-.

[46] Peng T, Boxberg M, Weichert W, Navab N, Marr C. Multi-task learning of a deep K-nearest neighbour network for histopathological image classification and retrieval. 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2019, October 13, 2019 - October 17, 2019. 2019;11764 LNCS(PG - 676-684):676-84. Available from: http://dx.doi.org/10.1007/978-3-030-32239-7_75NS-.

[47] Pereira S, Meier R, McKinley R, Wiest R, Alves V, Silva CA, et al. Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation. Medical Image Analysis. 2018;44:228-44. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L619966103&from=exporthttp://dx.doi.org/10.1016/j.media.2017.12.009.

[48] Uzunova H, Ehrhardt J, Kepp T, Handels H. Interpretable explanations of black box classifiers applied on medical images by meaningful perturbations using variational autoencoders. Medical Imaging 2019: Image Processing, February 19, 2019 - February 21, 2019. 2019;10949(PG - The Society of Photo-Optical Instrumentation Engineers (SPIE)):The Society of Photo-Optical Instrumentation Engin. Available from: http://dx.doi.org/10.1117/12.2511964NS-.

[49] Pirovano A, Heuberger H, Berlemont S, Ladjal S, Bloch I. Improving Interpretability for Computer-Aided Diagnosis Tools on Whole Slide Imaging with Multiple Instance Learning and Gradient-Based Explanations. 3rd International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2020, the 2nd International Workshop on Medical Image Learning with Less Labels and Imperfect Data, MIL3ID 2020, and the 5th International Workshop o. 2020;12446 LNCS(PG - 43-53):43-53. Available from: http://dx.doi.org/10.1007/978-3-030-61166-8_5NS-.

[50] Puyol-Antón E, Chen C, Clough JR, Ruijsink B, Sidhu BS, Gould J, et al. Interpretable Deep Models for Cardiac Resynchronisation Therapy Response Prediction. Med Image Comput Comput Assist Interv. 2020;2020(PG - 284-293):284-93. Available from: NS-.

[51] Puyol-Anton E, Ruijsink B, Clough JR, Oksuz I, Rueckert D, Razavi R, et al. Assessing the Impact of Blood Pressure on Cardiac Function Using Interpretable Biomarkers and Variational Autoencoders. 10th International Workshop on Statistical Atlases and Computational Models of the Heart, STACOM 2019, held in conjunction with the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2019, October 13, 2019. 2020;12009 LNCS(PG - 22-30):22-30. Available from: http://dx.doi.org/10.1007/978-3-030-39074-7_3NS-.

[52] Quellec G, Al Hajj H, Lamard M, Conze PH, Massin P, Cochener B. ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis. Medical Image Analysis. 2021;72. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2012995582&from=exporthttp://dx.doi.org/10.1016/j.media.2021.102118.

[53] Ren H, Wong AB, Lian W, Cheng W, Zhang Y, He J, et al. Interpretable Pneumonia Detection by Combining Deep Learning and Explainable Models with Multisource Data. Ieee Access. 2021;9(PG - 95872-95883):95872-83. Available from: http://dx.doi.org/10.1109/ACCESS.2021.3090215NS-.

[54] Singla S, Gong M, Ravanbakhsh S, Sciurba F, Poczos B, Batmanghelich KN. Subject2Vec: Generative-Discriminative Approach from a Set of Image Patches to a Vector. Med Image Comput Comput Assist Interv. 2018;11070(PG - 502-510):502-10. Available from: NS-.

[55] Sun J, Darbehani F, Zaidi M, Wang B. SAUNet: Shape Attentive U-Net for Interpretable Medical Image Segmentation. 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2020, October 4, 2020 - October 8, 2020. 2020;12264 LNCS(PG - 797-806):797-806. Available from: http://dx.doi.org/10.1007/978-3-030-59719-1_77NS-.

[56] Tanno R, Worrall DE, Kaden E, Ghosh A, Grussu F, Bizzi A, et al. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. NeuroImage. 2021;225. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2008373754&from=exporthttp://dx.doi.org/10.1016/j.neuroimage.2020.117366.

[57] Velikova M, Lucas PJF, Samulski M, Karssemeijer N. On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. Artificial Intelligence in Medicine. 2013;57(1):73-86. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L52426996&from=exporthttp://dx.doi.org/10.1016/j.artmed.2012.12.004.

[58] Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific Reports. 2021;11(1):3254. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L634212207&from=exporthttp://dx.doi.org/10.1038/s41598-020-74399-w.

[59] Verma A, Shukla P, Abhishek, Verma S. An interpretable SVM based model for cancer prediction in mammograms. 1st International Conference on Communication, Networks and Computing, CNC 2018, March 22, 2018 - March 24, 2018. 2019;839(PG - 443-451):443-51. Available from: http://dx.doi.org/10.1007/978-981-13-2372-0_39NS-.

[60] Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. European Radiology. 2019;29(7):3348-57. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L627809141&from=exporthttp://dx.doi.org/10.1007/s00330-019-06214-8.

[61] Wang K, Patel BK, Wang L, Wu T, Zheng B, Li J. A dual-mode deep transfer learning (D2TL) system for breast cancer detection using contrast enhanced digital mammograms. IISE Transactions on Healthcare Systems Engineering. 2019;9(4):357-70. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L628355909&from=exporthttp://dx.doi.org/10.1080/24725579.2019.1628133.

[62] Wongvibulsin S, Wu KC, Zeger SL. Improving Clinical Translation of Machine Learning Approaches Through Clinician-Tailored Visual Displays of Black Box Algorithms: Development and Validation. JMIR Med Inform. 2020;8(6 PG - e15791):e15791. Available from: NS-.

[63] Xu X, Guan Y, Li J, Ma Z, Zhang L, Li L. Automatic glaucoma detection based on transfer induced attention network. Biomedical Engineering Online. 2021;20(1):39. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L634874614&from=exporthttp://dx.doi.org/10.1186/s12938-021-00877-5.

[64] Yan K, Peng Y, Sandfort V, Bagheri M, Lu Z, Summers RM. Holistic and comprehensive annotation of clinically significant findings on diverse CT images: Learning from radiology reports and label ontology. 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, June 16, 2019 - June 20, 2019. 2019;2019-June(PG - 8515-8524):8515-24. Available from: http://dx.doi.org/10.1109/CVPR.2019.00872NS-.

[65] Yang H, Kim JY, Kim H, Adhikari SP. Guided Soft Attention Network for Classification of Breast Cancer Histopathology Images. IEEE Transactions on Medical Imaging. 2020;39(5 PG - 1306-1315):1306-15. Available from: http://dx.doi.org/10.1109/TMI.2019.2948026NS-.

[66] Yeche H, Harrison J, Berthier T. UBS: A dimension-agnostic metric for concept vector interpretability applied to radiomics. 2nd International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2019, and the 9th International Workshop on Multimodal Learning for Clinical Decision Support, ML-CDS 2019, held in conjunction with the 22nd Interna. 2019;11797 LNCS(PG - 12-20):12-20. Available from: http://dx.doi.org/10.1007/978-3-030-33850-3_2NS-.

[67] Zhao G, Zhou B, Wang K, Jiang R, Xu M. Respond-CAM: Analyzing deep models for 3D imaging data by visualizations. 21st International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2018, September 16, 2018 - September 20, 2018. 2018;11070 LNCS(PG - 485-492):485-92. Available from: http://dx.doi.org/10.1007/978-3-030-00928-1_55NS-.

[68] Zhu P, Ogino M. Guideline-based additive explanation for computer-aided diagnosis of lung nodules. 2nd International Workshop on Interpretability of Machine Intelligence in Medical Image Computing, iMIMIC 2019, and the 9th International Workshop on Multimodal Learning for Clinical Decision Support, ML-CDS 2019, held in conjunction with the 22nd Interna. 2019;11797 LNCS(PG - 39-47):39-47. Available from: http://dx.doi.org/10.1007/978-3-030-33850-3_5NS-.

[69] Gu R, Wang G, Song T, Huang R, Aertsen M, Deprest J, et al. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. IEEE Transactions on Medical Imaging. 2021;40(2):699-711. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L634109018&from=exporthttp://dx.doi.org/10.1109/TMI.2020.3035253.