Available online at **www.joac.info**

ISSN: 2278-1862

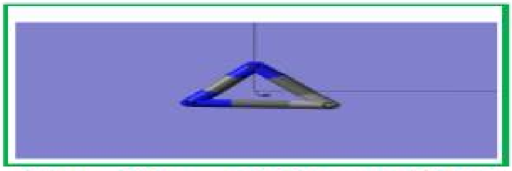
Journal of Applicable Chemistry

2023, 12 (4): 553-596

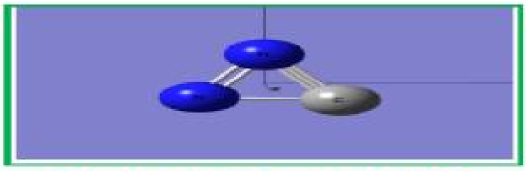
(International Peer Reviewed Journal)

New Chemistry News

$^{-}N=C=N^{-}$



New News of Chem (NNC)



ChemNewsNew (CNN)

CNN-55--Fit (Figure Image TableScript...)Bases (Bfit) Part 3.xAI. Architectures . Medicine – Part-2

Information Source	sciencedirect.com;	
S. Narasinga Rao M D **Associate Professor,** **Dept. of General Medicine,** **Government medical college,** **government general hospital,** **Srikakulam, AP, India**	**K. SomasekharaRao, Ph D** **Dept. of Chemistry,** **Acharya Nagarjuna Univ.,** **Dr. M.R.Appa Rao Campus,** **Nuzvid-521 201, India**	**R. Sambasiva Rao, Ph D** **Dept. of Chemistry,** **Andhra University,** **Visakhapatnam 530 003,** **India**
snrnaveen007@gmail.com (+91 9848136704)	sr_kaza1947@yahoo.com (+91 98 48 94 26 18)	rsr.chem@gmail.com (+91 99 85 86 01 82)

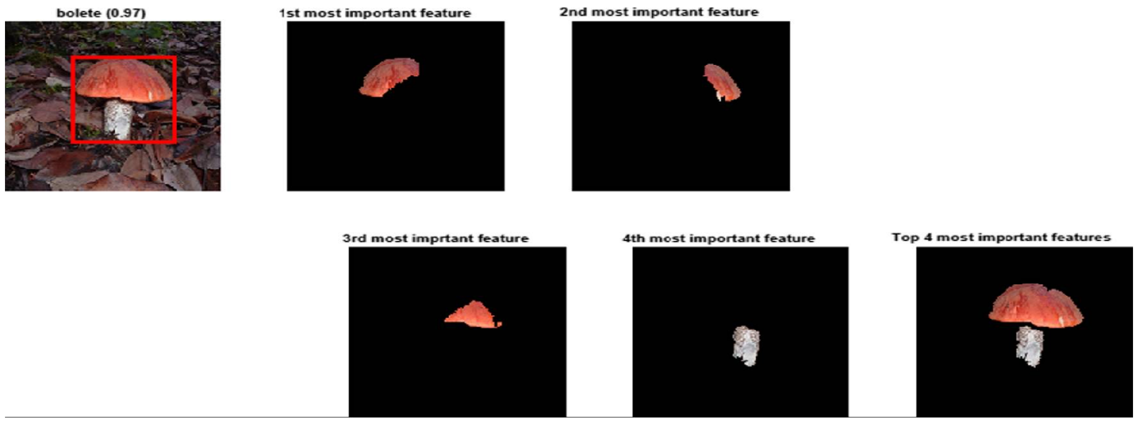
Conspectus: The evolution of scientific Artificial Intelligence (AI) (save consciousness) implementable on computers since mid-nineteen fifties and emergence of eXplainable AI (xAI) during last one decade is briefly described. The extensive application in most of critical research disciplines brought this approach to the fore-front-of state-of-knowledge projects including medicine, defence, industry, commerce etc. The recent trend setting xAI-probes resulted in laying confidence in AI-embedded methods/products/materials. The value and acceptability increased probing more into trust-worthy and responsible AI to comply with stake-holder's expectations. This indispensable upward trend of more and more AI-integrated technologies is in harmony with safety and security oflife forms and environment. Here, some typical case studies shedding light on benefits of xAI methods in Medical diagnosis and health care are incorporated.

Keywords: AI; evolution (1950-to-2023-Future); eXplainable AI (xAI); interpretable/ Responsible/ Trustworthy AI; DARPA and NSF targets; Industry (Health, Defence, research requirements)-Deep architectures; CNN; Capsule Nets; ALEX; TRANSFORMERS; Machine Learning; Deep Learning—Supervised/unsupervised/Self-supervised data; Figures; Images, Tables, Scripts, Numerical values;

LIME plot

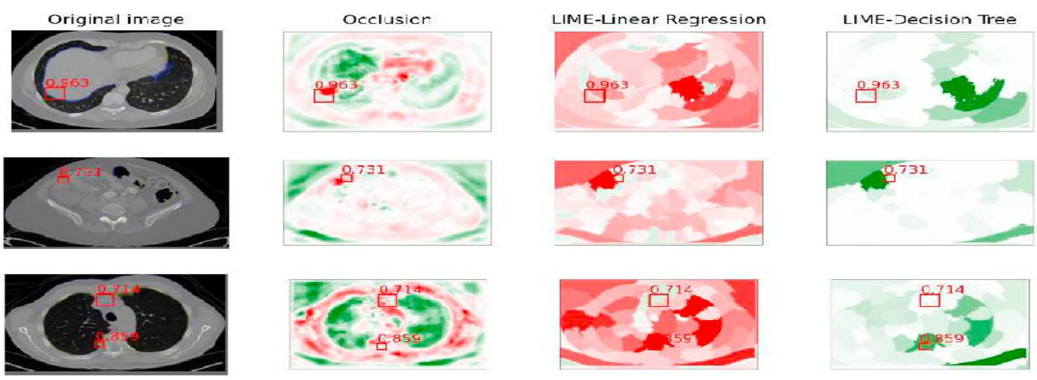
xAI.Med. 2022-58

LIME maps

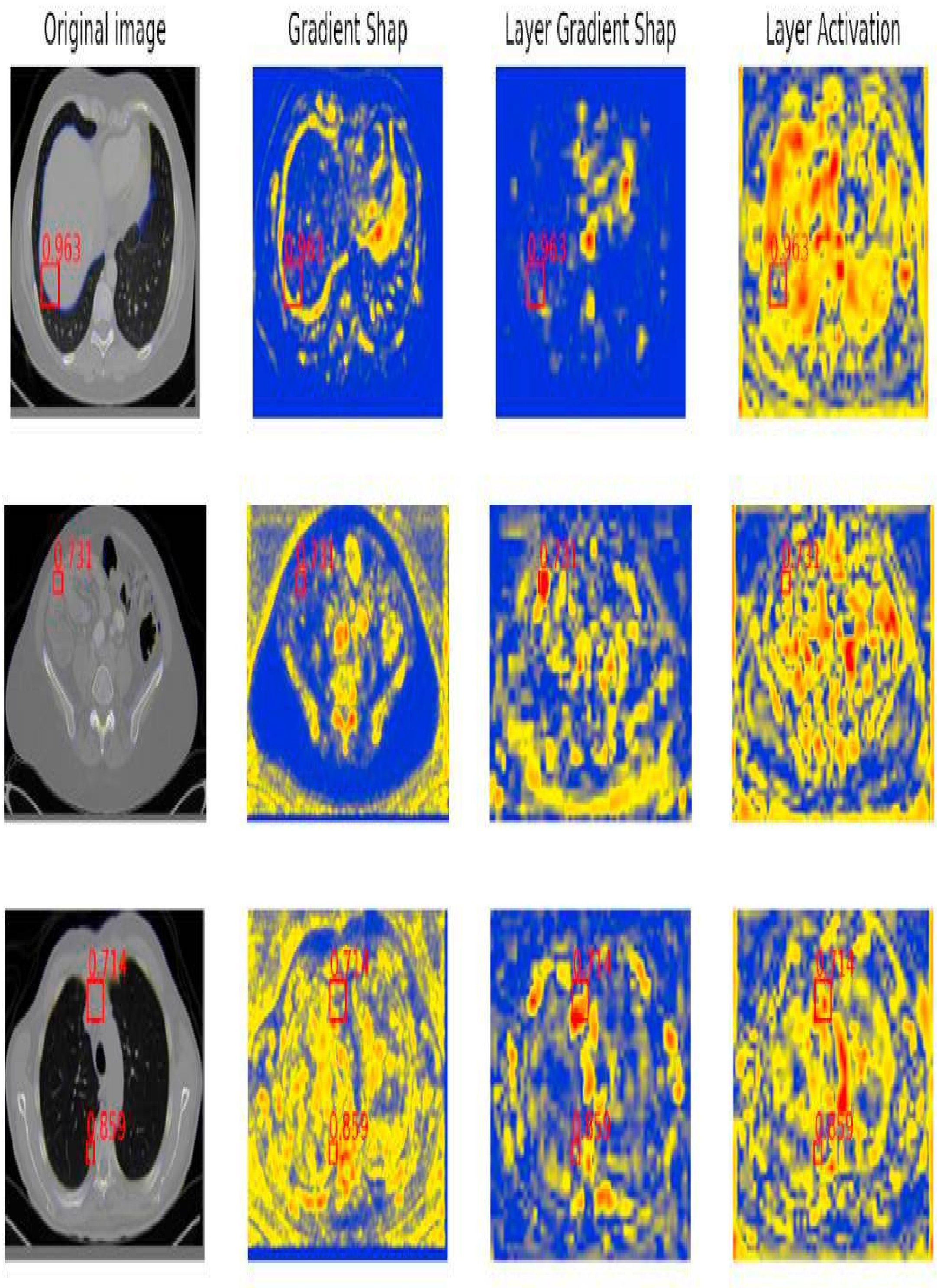


xAI.Med. 2022-59

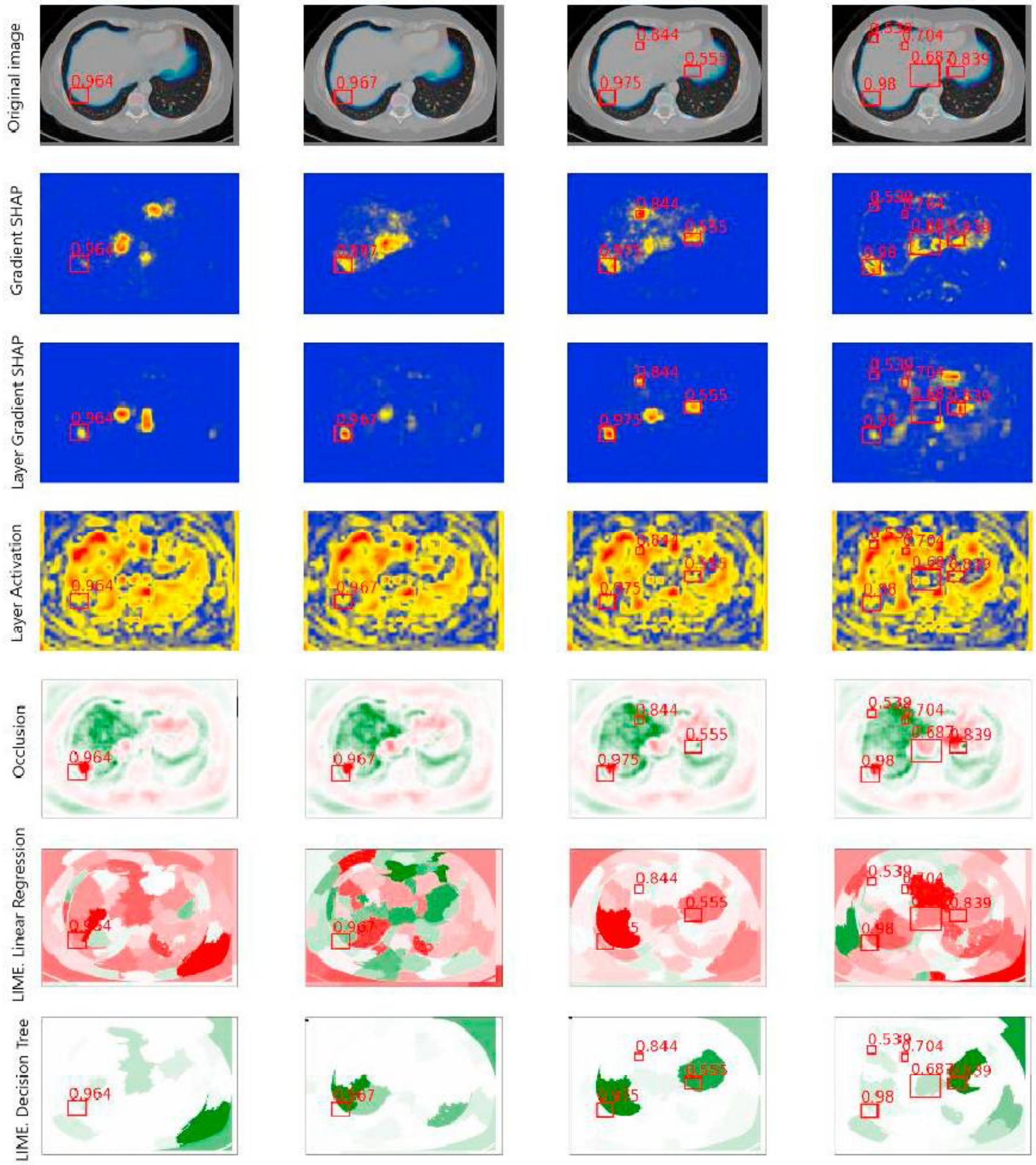
LIME -- Linear Regression and Decision Tree



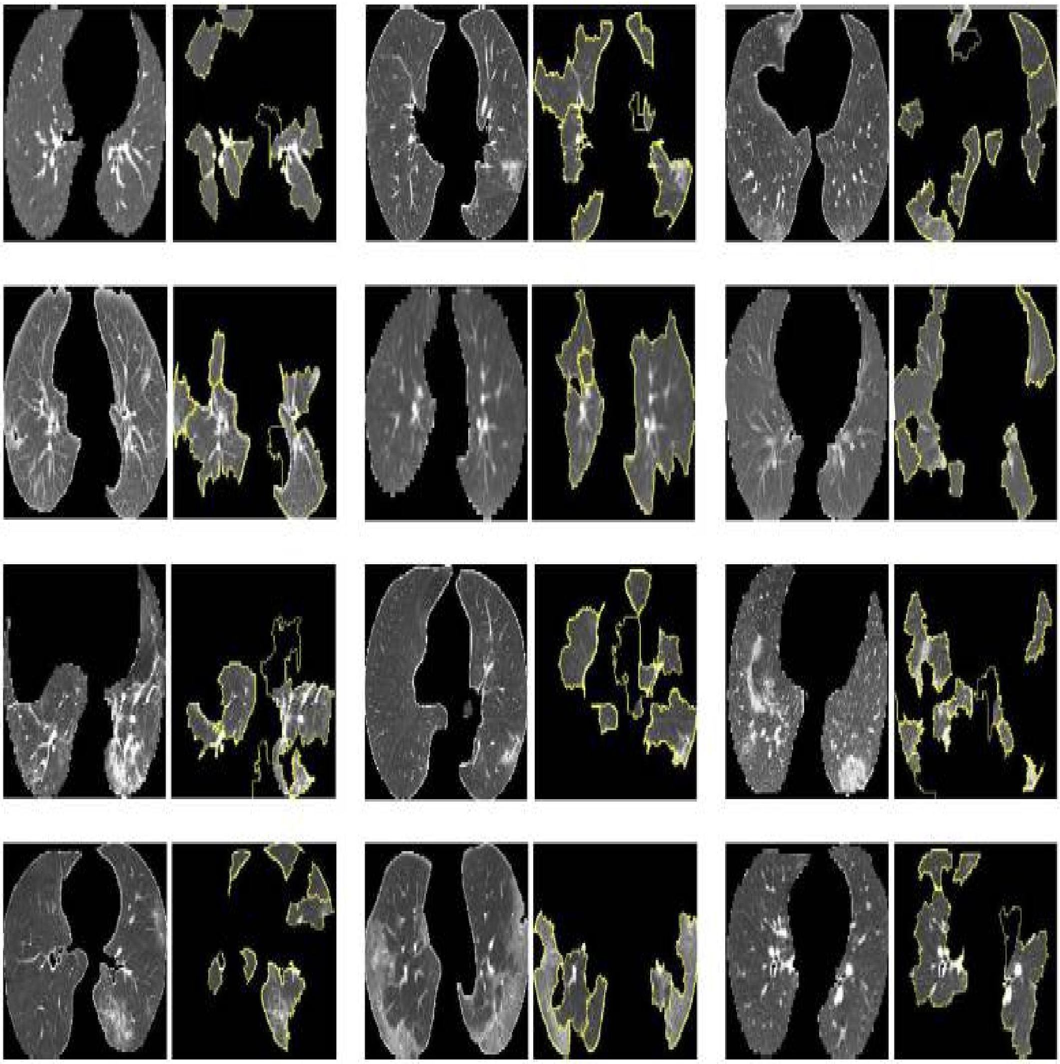
Shap –Layer activation



. Consistency evaluation of axial slices of original images

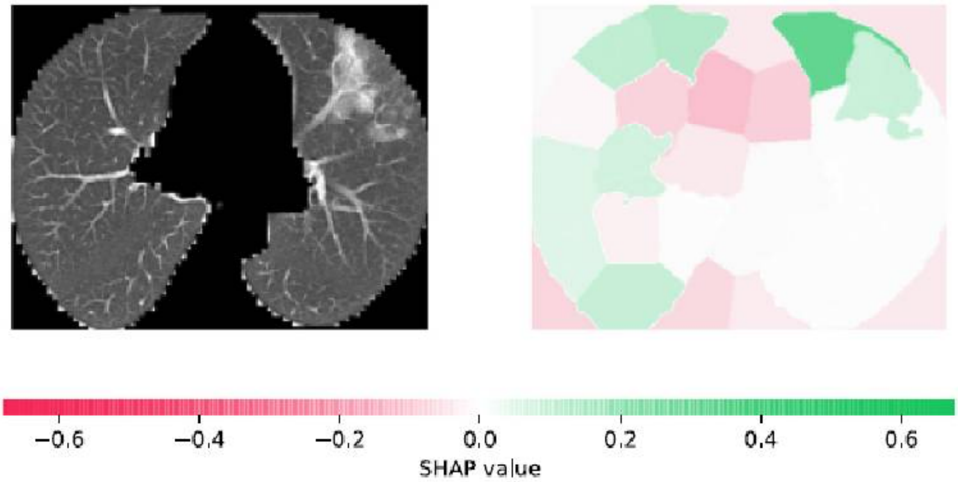


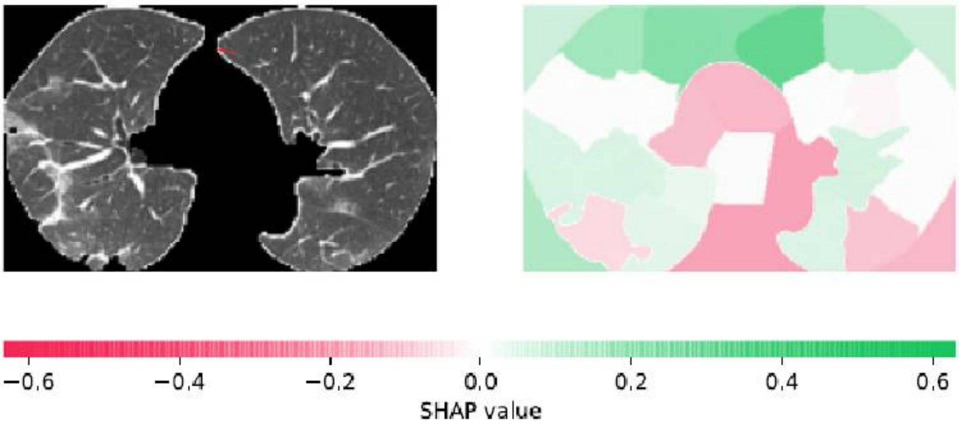
Visualisation of the super-pixels which are
positively contributed to the predictions via the LIME method



xAI.Med. | 2022-62

SHAP values for differentsuper-pixels of the sampled images by Kernel SHAP method





- ✓ Super-pixel with positive SHAP value: positive impact to the positive prediction
- ✓ Negative value : super-pixel contributes to the negative prediction

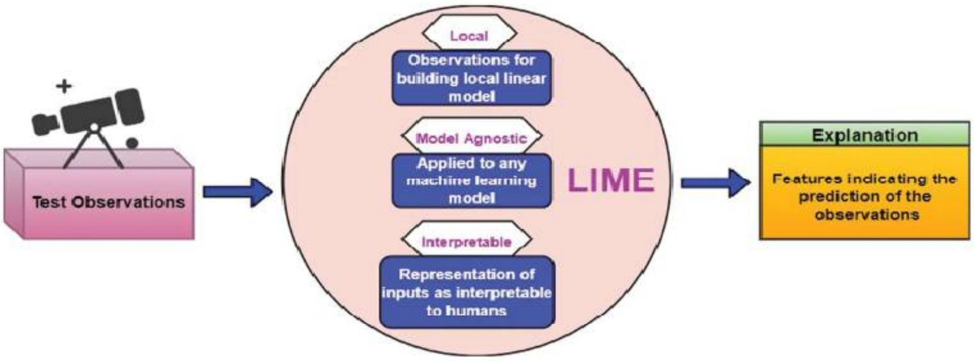
xAI.Med. | 2022-209

Larger bar, the larger is its contribution

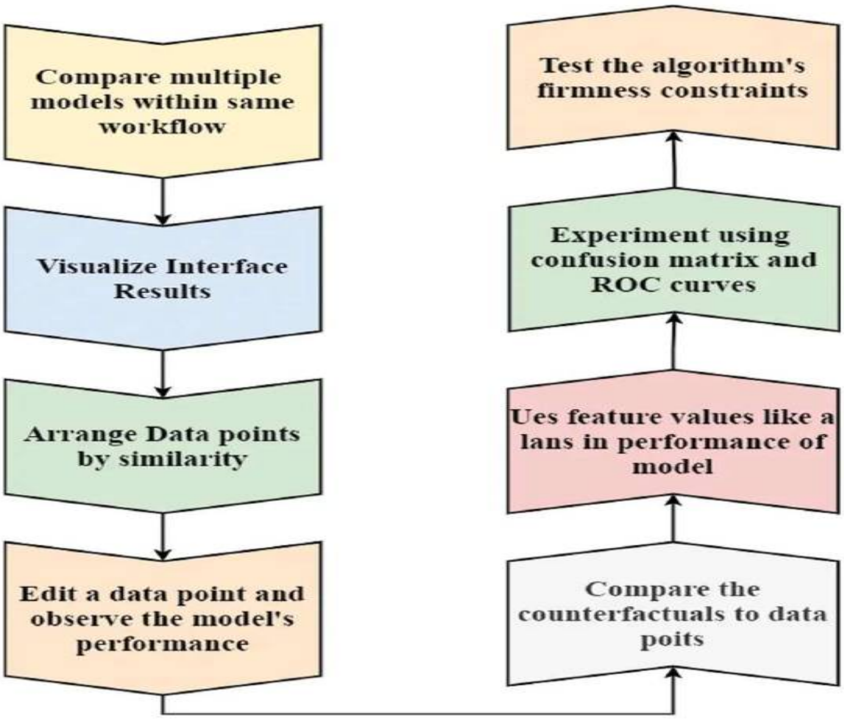
- The deep purple bar indicates the final prediction for a chosen patient. The plot also shows the contributions of each variable to the final prognosis.
- Green bars indicate positive contributions, meaning that the value of a certain variable implies an increase in prediction
- Red bars suggest a negative contribution
- ✓ Shapley values indicate positive and negative contributions to the final prediction

xAI.Med. 2022-205

LIME

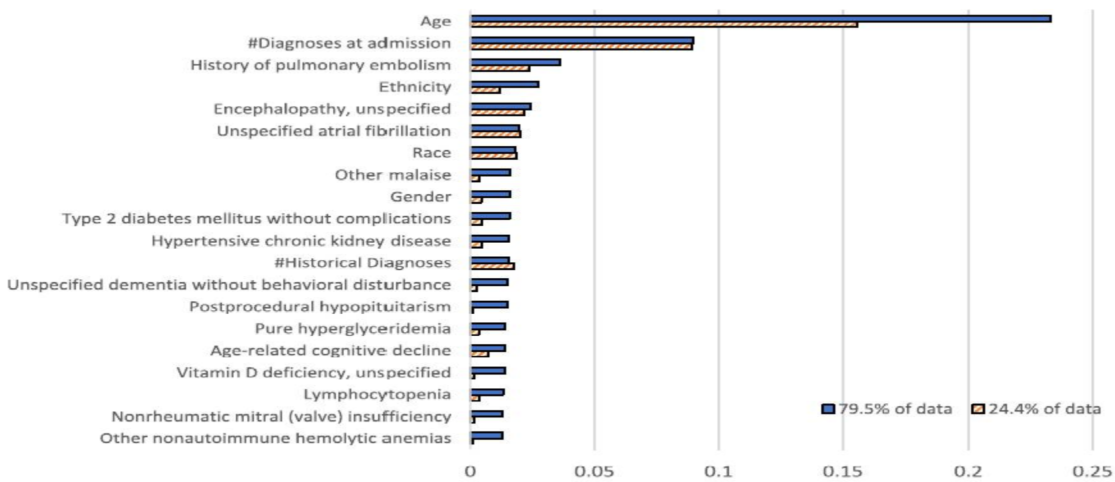


What-if-Tool



https://towardsdatascience.com/using-what-if-tool-to-investigate-machine-learning-models-913c7d4118f)

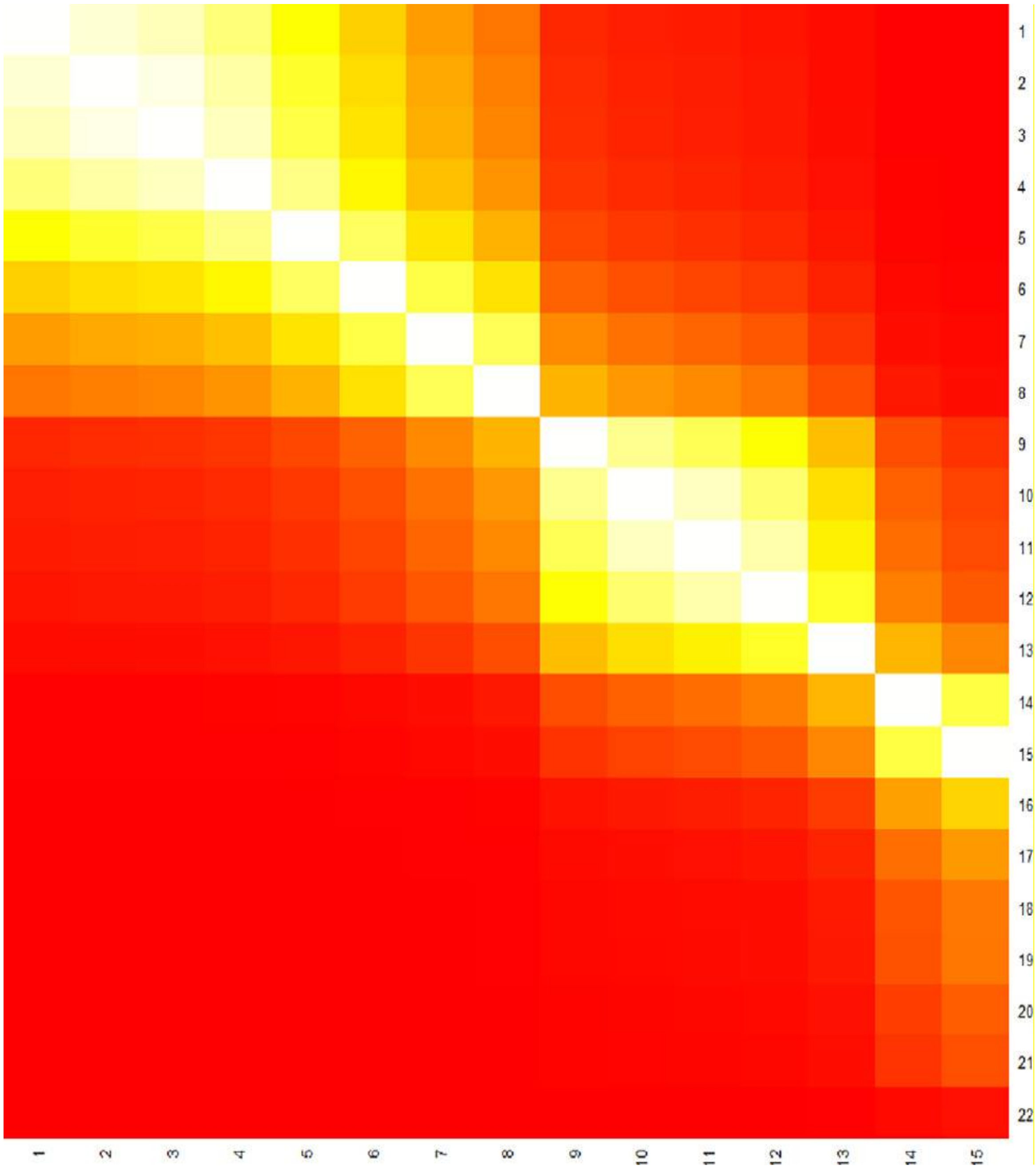
SHAP scores



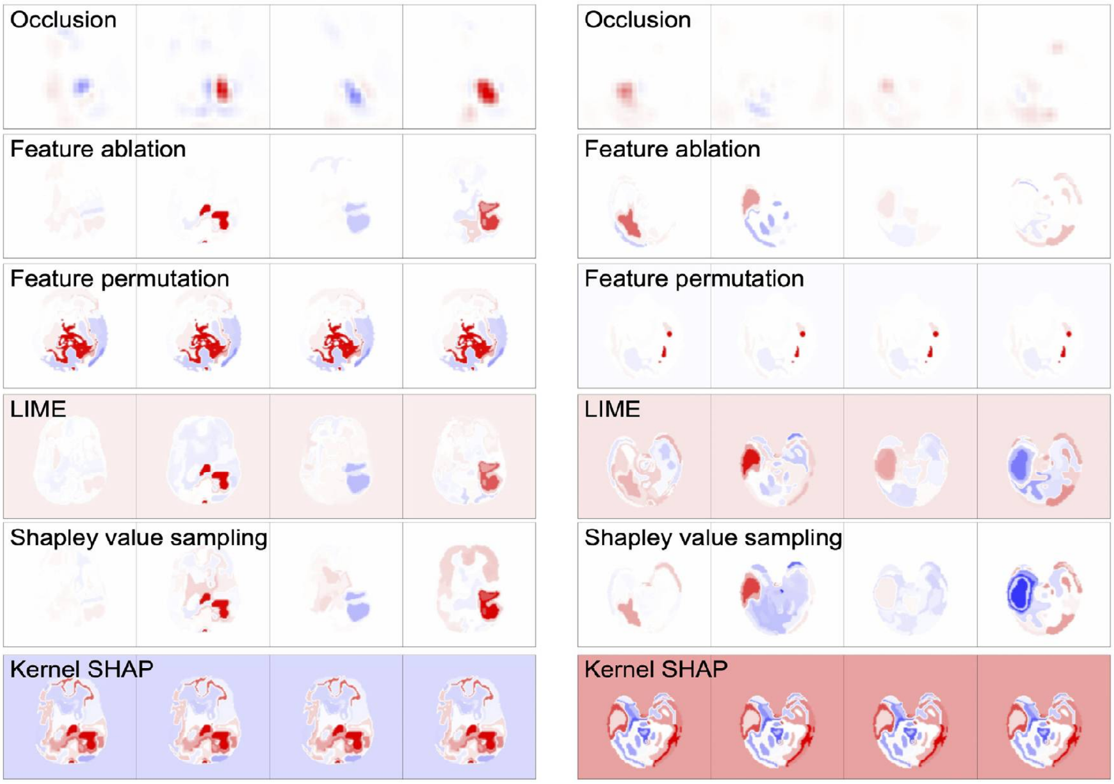
Heat maps

xAI.Med. | 2023-45

Heat map for p-values of the Friedman test



Heatmap explanations generated from perturbation-based explanation methods



Gradient-based explanation methods (Ex: Guided BackProp, DeepLift)

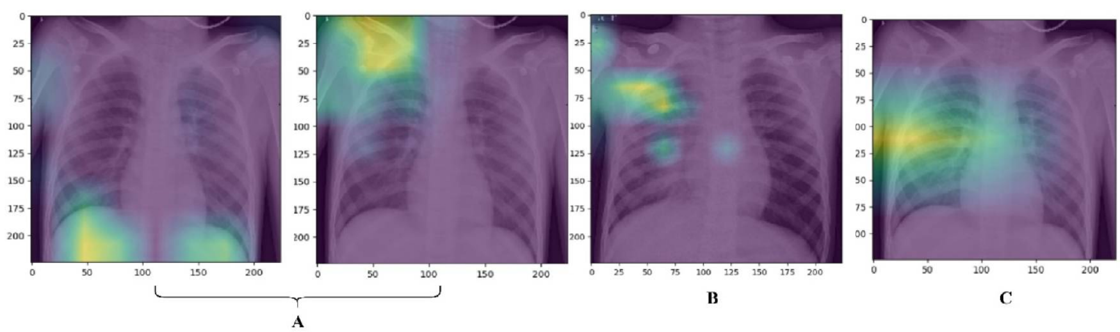
- ✓ Utilize the gradient signal to estimate the feature importance for model prediction

Perturbation-based methods (Ex: occlusion, LIME, kernel SHAP)

- ✓ Utilize the input-output sampling pairs to estimate the feature importance

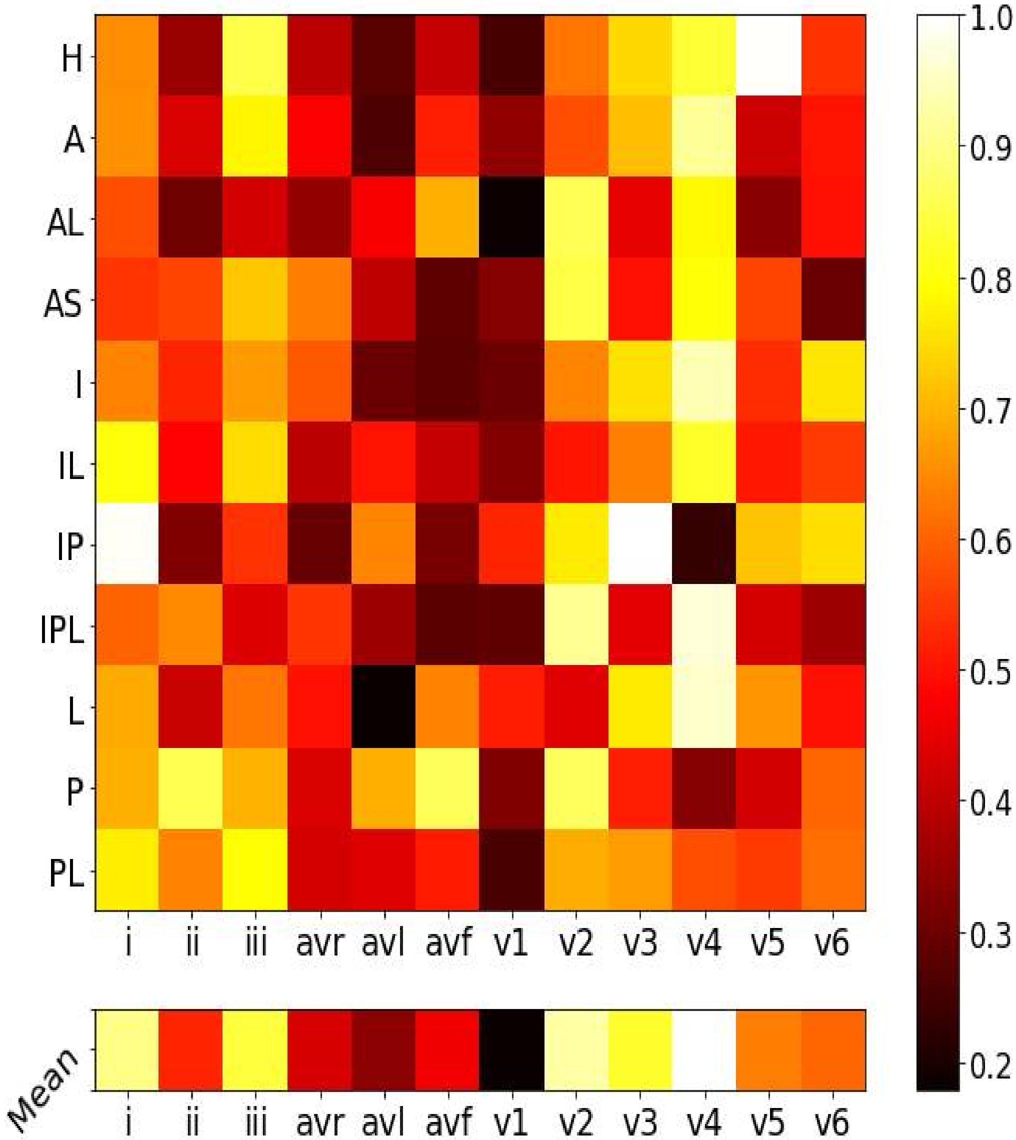
xAI.Med. 2022-07

Visual explainable heat maps (i.e., saliency maps) of the chest X-ray pneumonia image



- ✓ Heat maps of the pre-train DenseNet201and VGG16 models
- ✓ Heat map of the ensemble A deep learning model
- ✓ (C) Saliency map for proposed hybrid deep learning framework

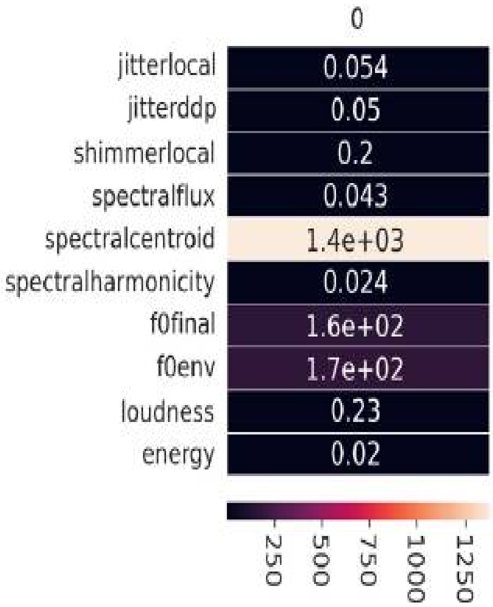
Heat maps of individual lead activations



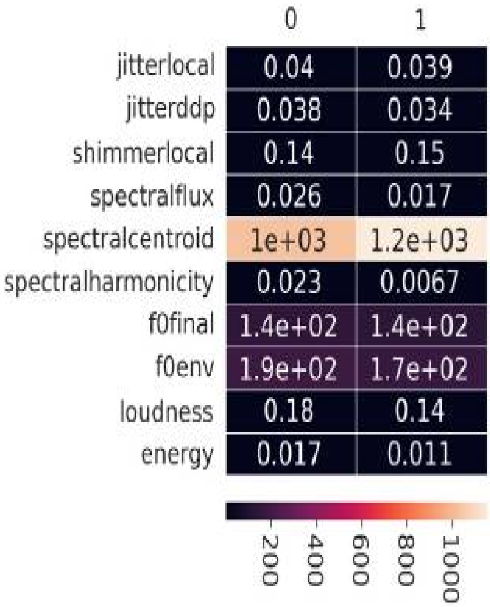
DenseNet and CNN models.

Anterior, A; anterior lateral, AL; anterior septal, AS; healthy, H;
inferior, I; inferior lateral, IL; inferior posterior, IP; inferior posterior lateral, IPL;
lateral, L; posterior, P; posterior lateral, PL

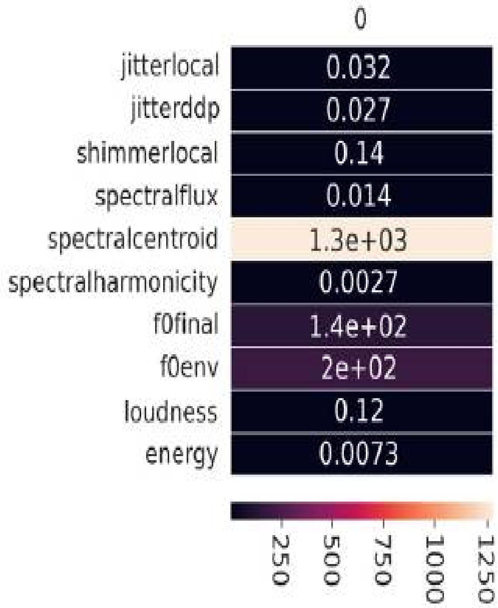
Heatmap visualization



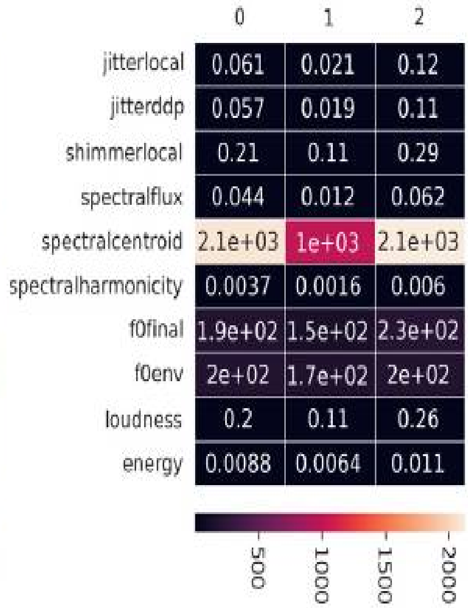
(a) chunk#4



(b) chunk#5



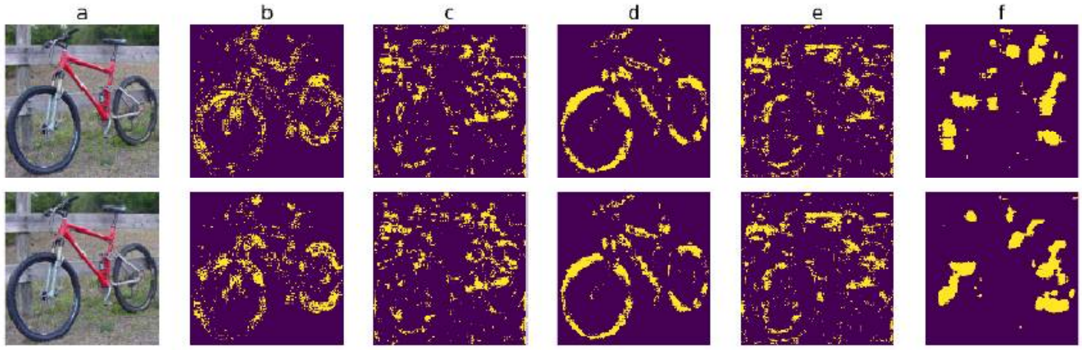
(c) chunk#6



(d) chunk#8

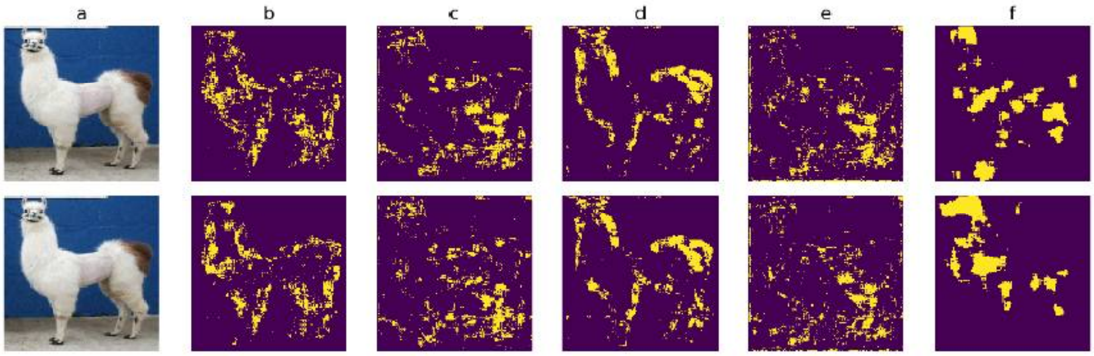
xAI.Med. | 2022-201

Heatmaps

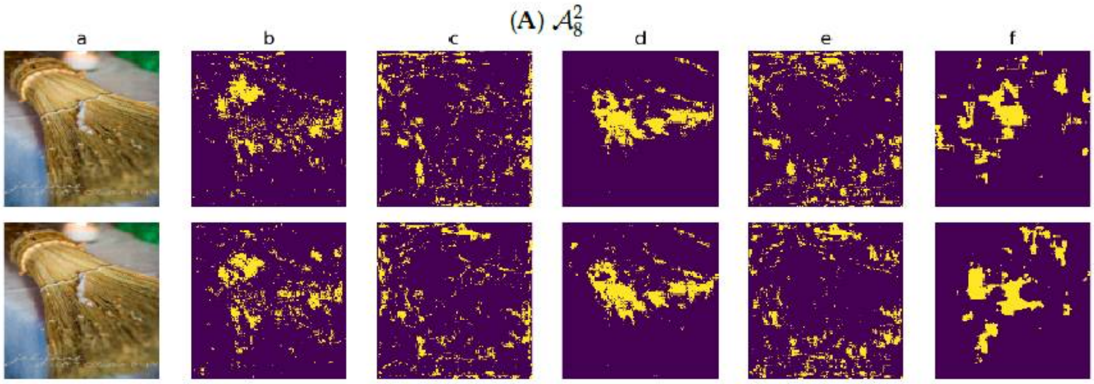


xAI.Med. | 2022-

Heatmaps



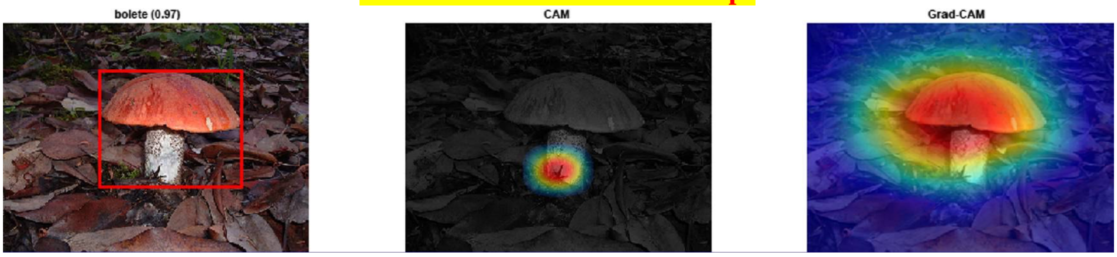
(A) $\mathcal{A}_8^2$



Heatmaps Grad-CAM

xAI.Med.		2022-58

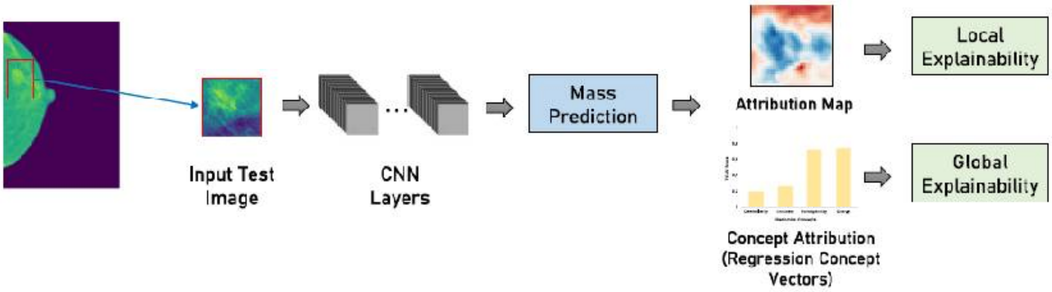
CAM and Grad-CAM heatmaps



Attribution map

xAI.Med.	Mass detection in mammograms	2022-68

Post-hoc interpretability



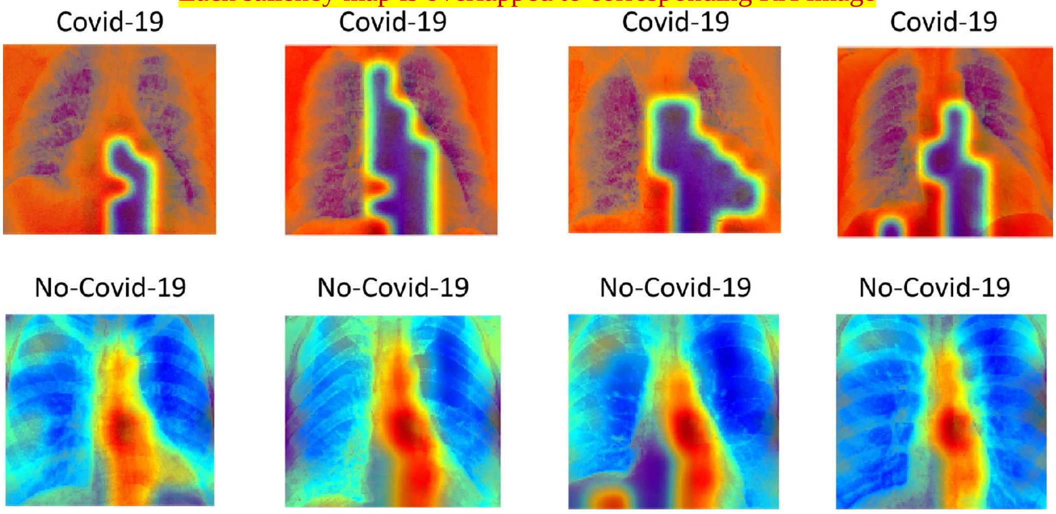
Attribution map

- Explains each test image by highlighting regions of image
- Model Provides global interpretability by quantifying influence of radiomics features on prediction for each class.

Saliency maps

xAI.Med. | 2022-

Each saliency map is overlapped to corresponding RX image

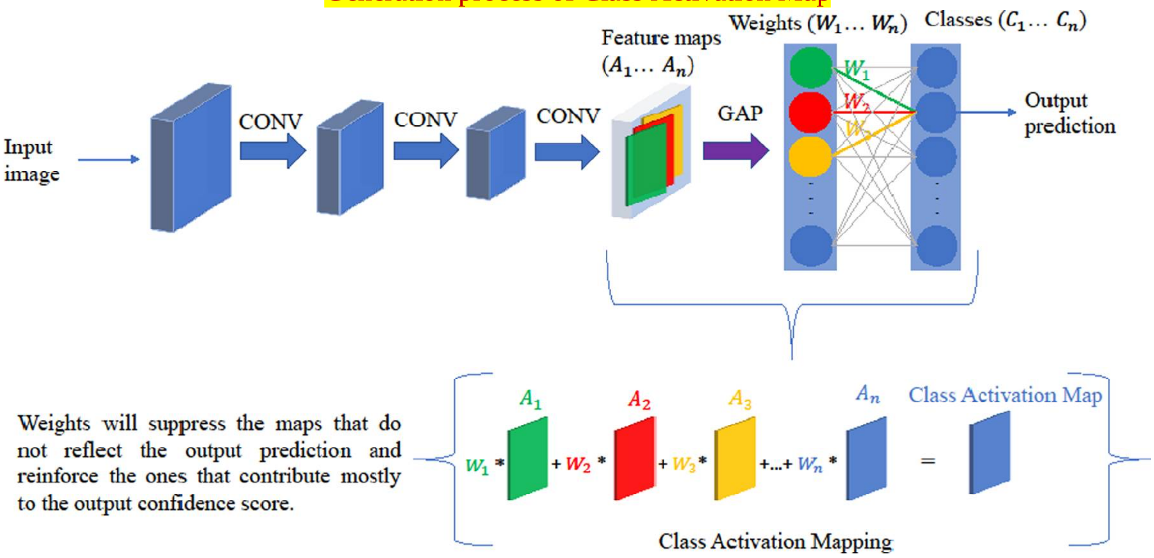


- Coloration of pixels in the saliency map: blue (low relevance) ; red (high relevance)

Class Activation maps

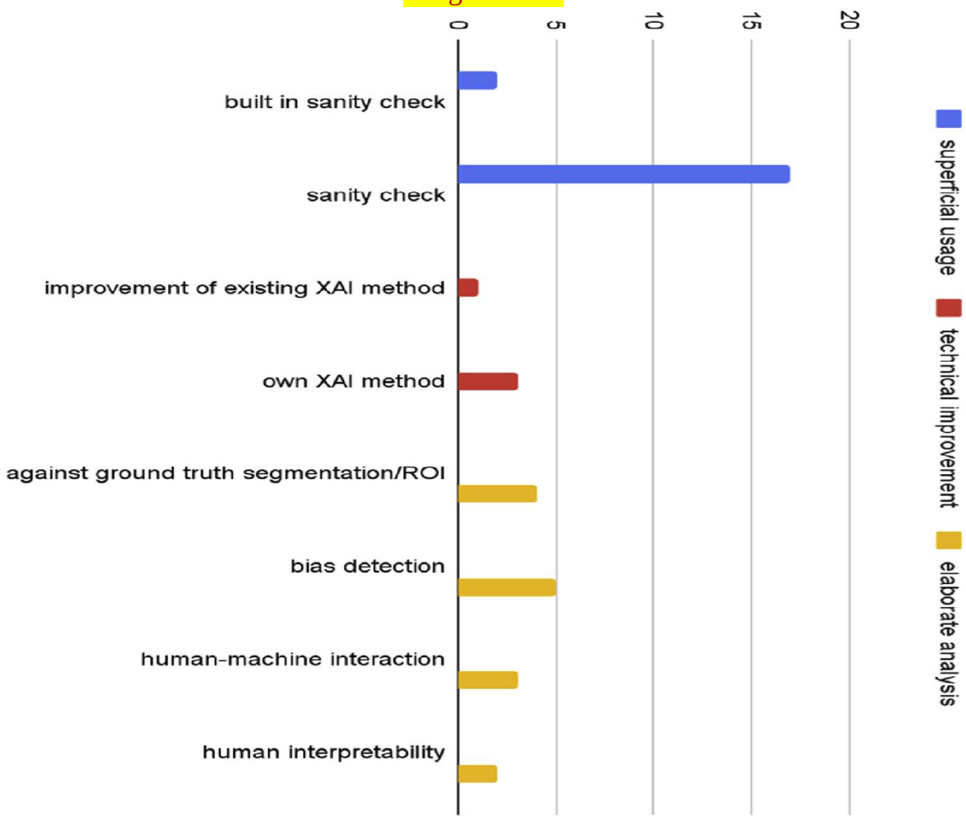
xAI.Med. | 2022-58

Generation process of Class Activation Map



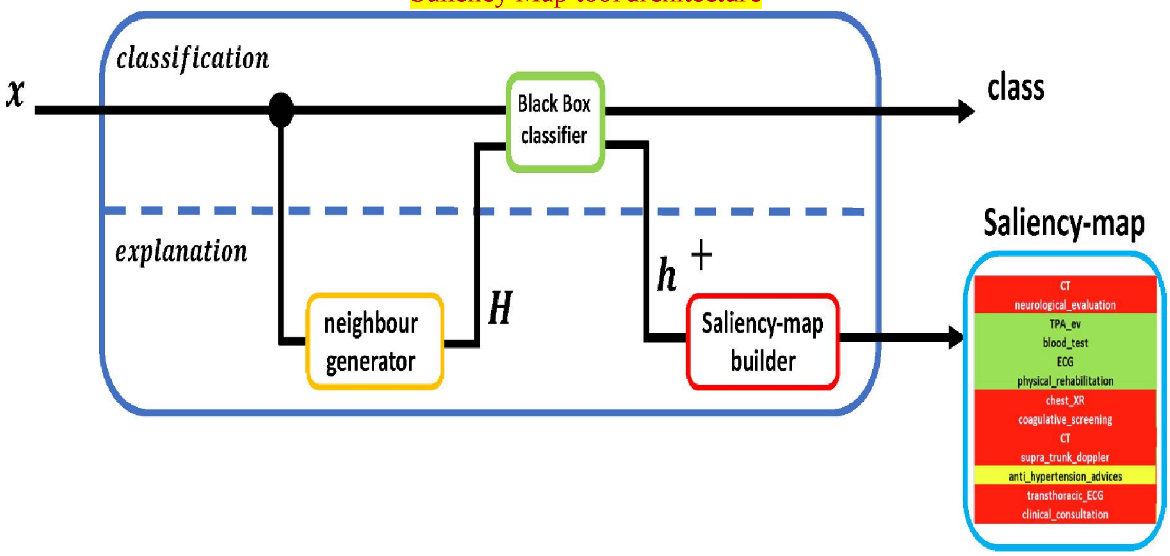
xAI.Med. 2022-51

Usage of XAI



xAI.Med. 2022-72

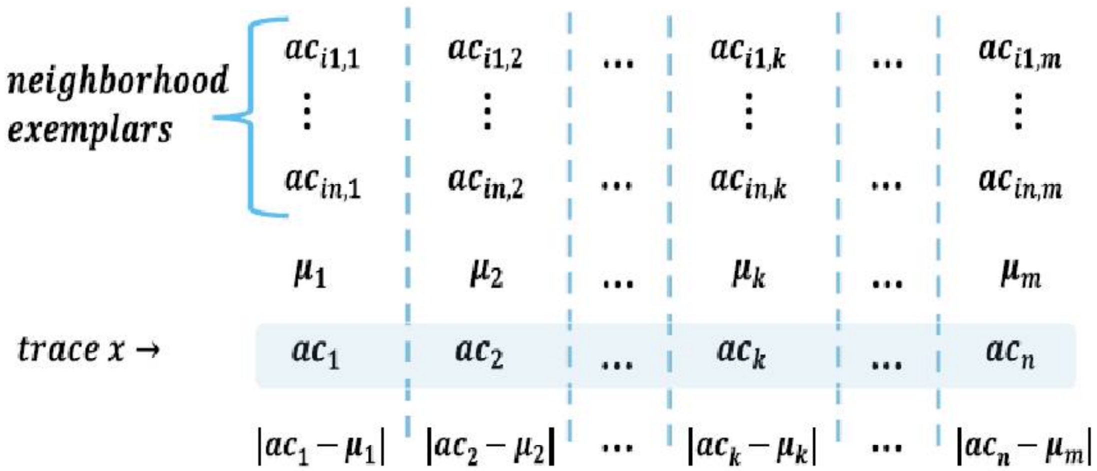
Saliency Map tool architecture



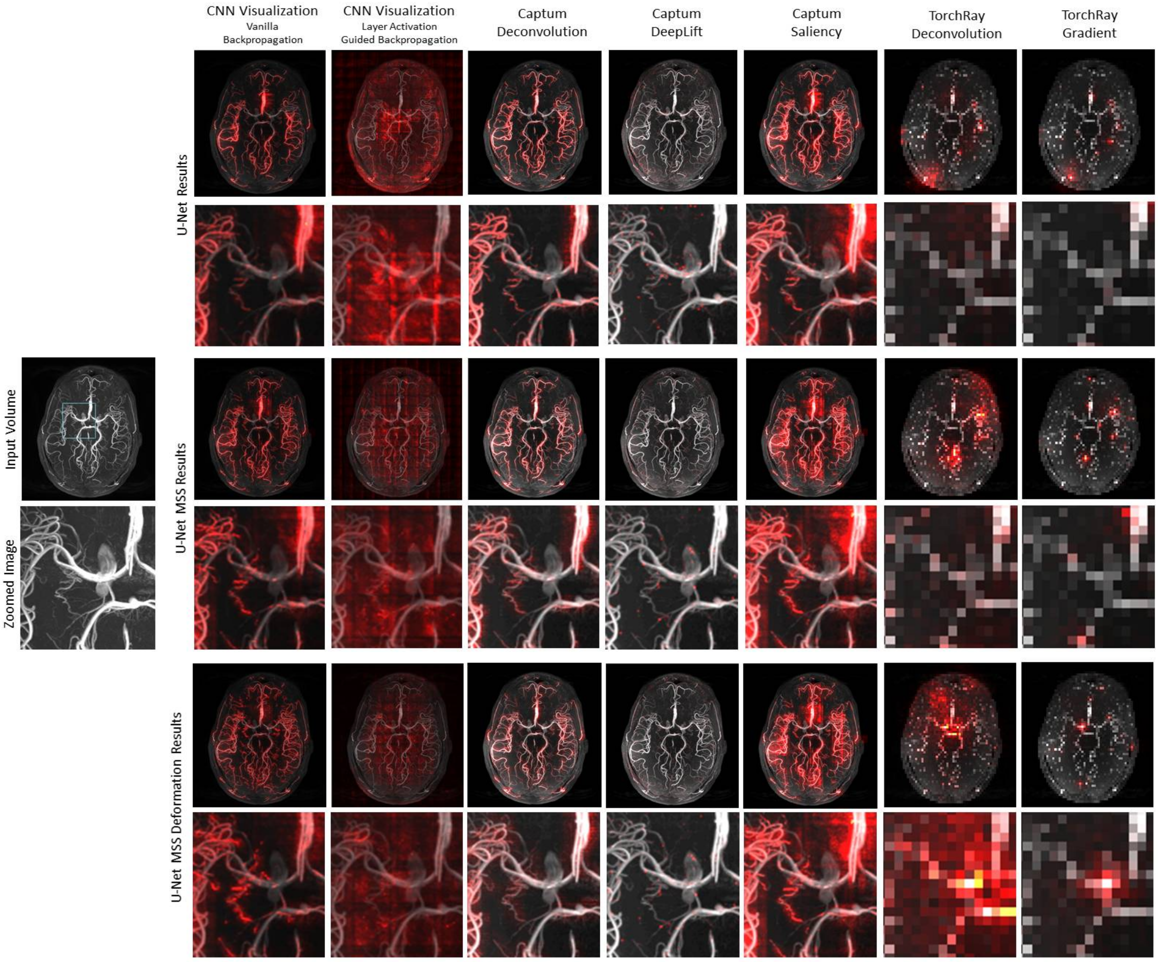
Saliency maps of a simple patient trace	Complex patient trace of saliency maps
CT	CT
neurological_evaluation	neurological_evaluation
TPA	ECG
blood_test	supra_trunk_doppler
ECG	transcranial_doppler
physical_rehabilitation	neurosurgery_consultation
chest_XR	blood_test
coagulative_screening	TPA
CT	ECG
supra_trunk_doppler	antiaggregants
anti_hypertension_advices	chest_XR
transthoracic_ECG	CT
clinical_consultation	venous_sinus_thrombosis
	physiatrist_consultation
	physical_rehabilitation
	transthoracic_ECG
	anti_hypertension_therapy
	...

Saliency maps of a simple patient trace | Complex patient trace of saliency maps

Trace saliency map construction procedure



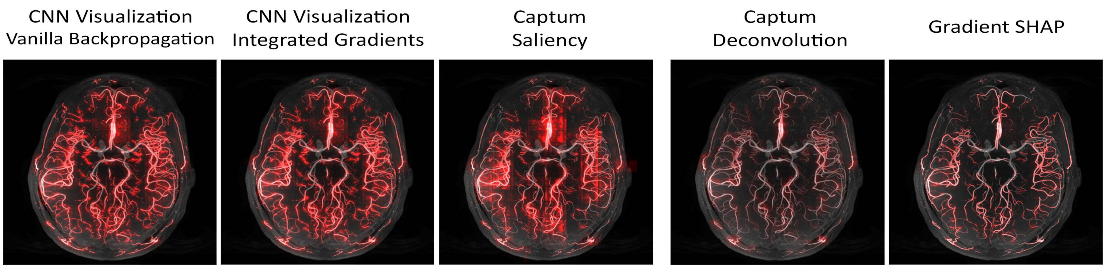
xAI.Med. 2022-211



Higher the intensity of red, higher is concentration of focus of the network

! Looking at the regions where the network did not focus, it can be understood which parts of the segmentation prediction might be wrong

xAI.Med. 2022-211

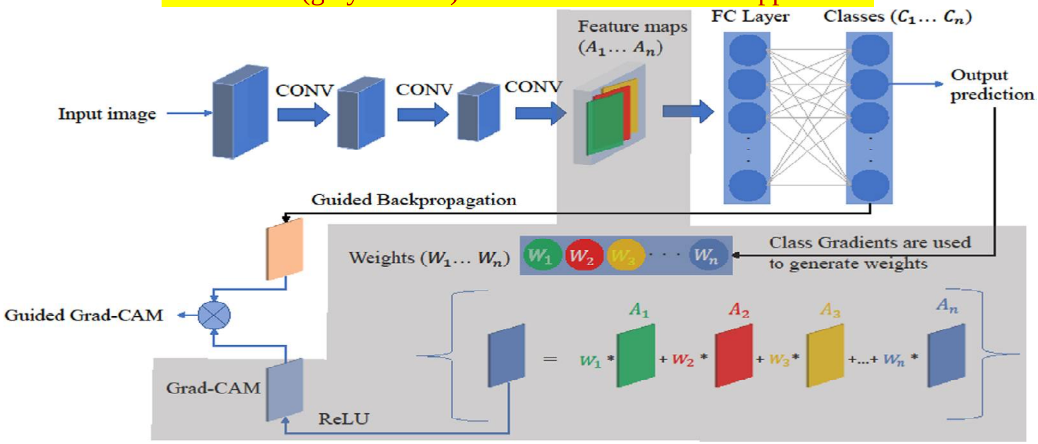


- Saliency show similar focus areas: on the anterior, posterior, right, and left regions of the brain
- Captum Deconvolution and Gradient SHAP emphasise specifically on cerebral artery itself, not the whole area

CAM	
EigenCAM	CAMs Aconv
Grad- CAM	Guided Grad-CAM
LayerGradCAM	

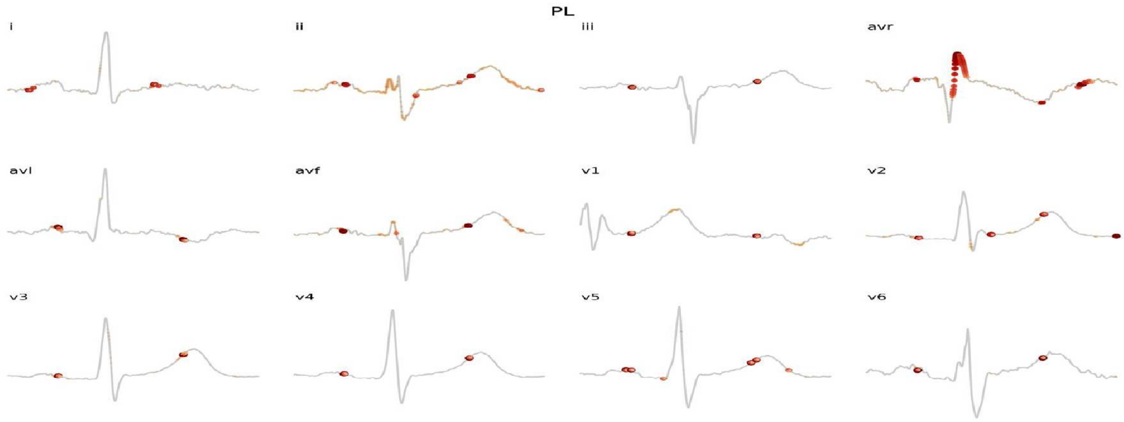
xAI.Med.		2022-58

Grad-CAM (grey-shaded) and Guided Grad-CAM approaches

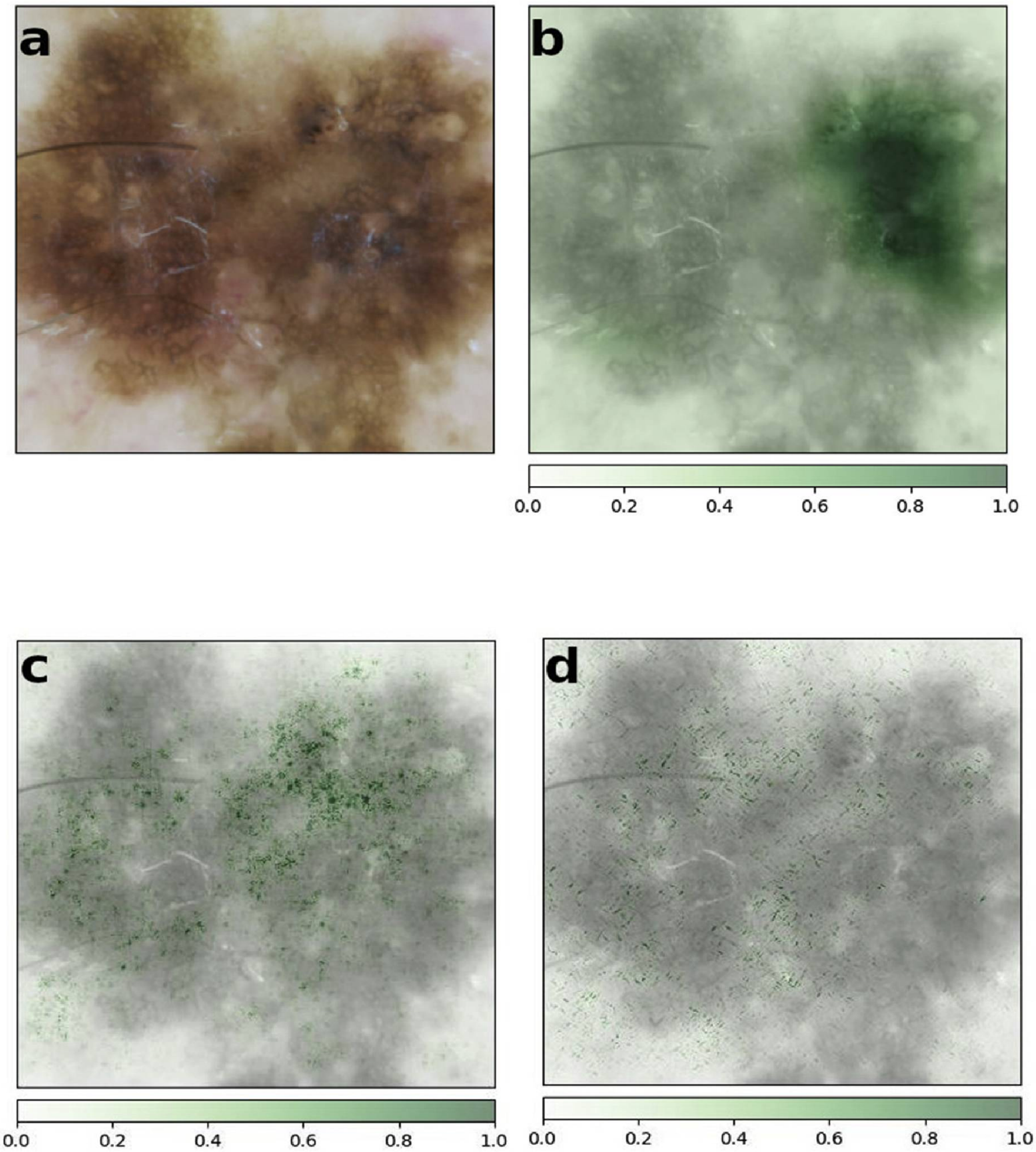


xAI.Med.		2022-32

Activation maps of each myocardial infarction class by Grad-CAM method



Exemplary XAI visualisations,

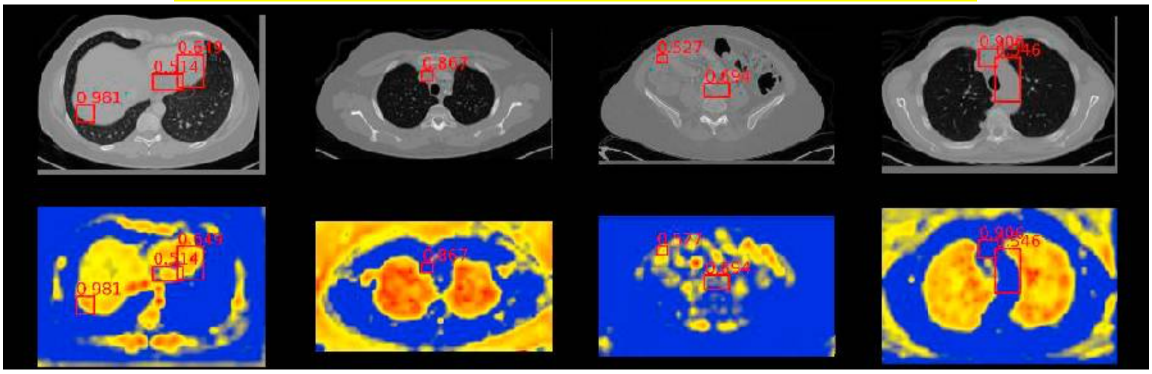


- a: Dermoscopic image of a melanoma from the ISIC archive and corresponding explanations for class "melanoma" created with
- b: GradCAM
- c: Backpropagation
- d: Integrated Gradients

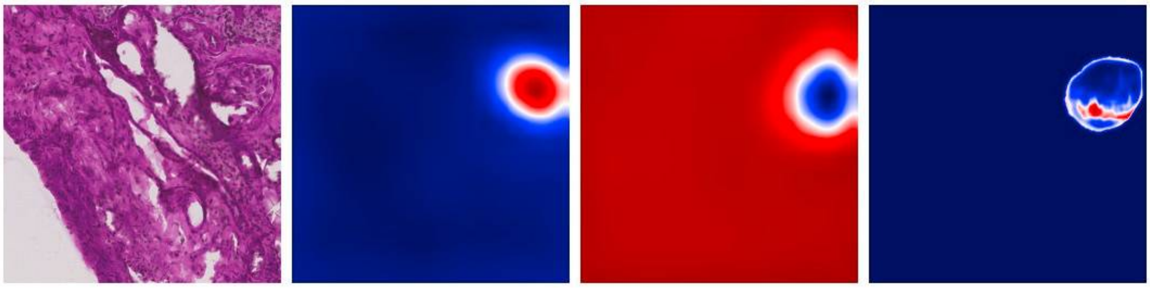
xAI.Med. 2022-59

Original images (first row)

Representations of their attribution maps for EigenCAM (second row)



xAI.Med. 2022-60

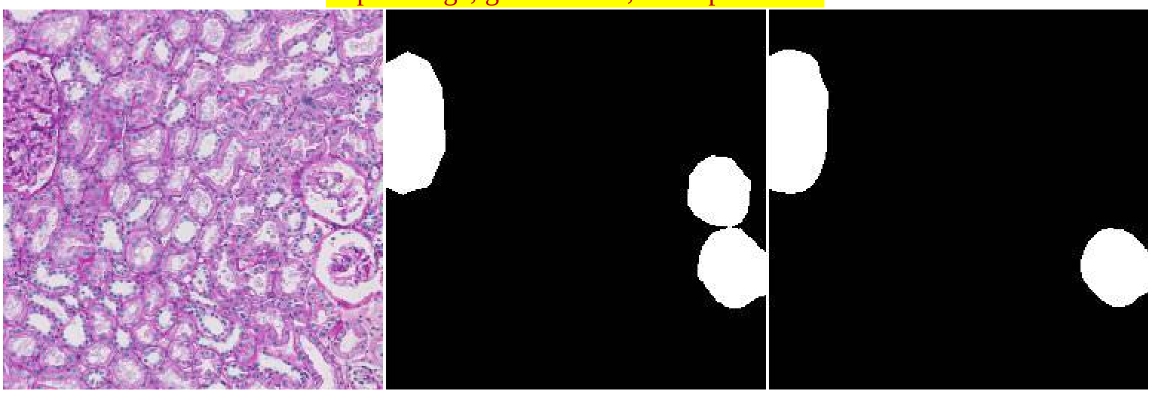


a) b) c) d)

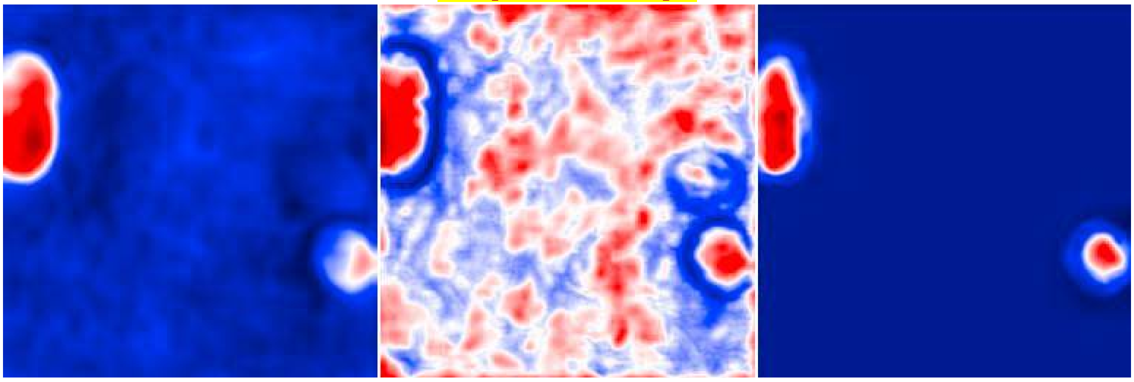
- ✓ Inputimage,a)
- ✓ LayerGradCAM,b)
- ✓ LayerActivation,c)
- ✓ LayerGradientSHAPoftheeslectedlayer (d)

xAI.Med. 2022-60

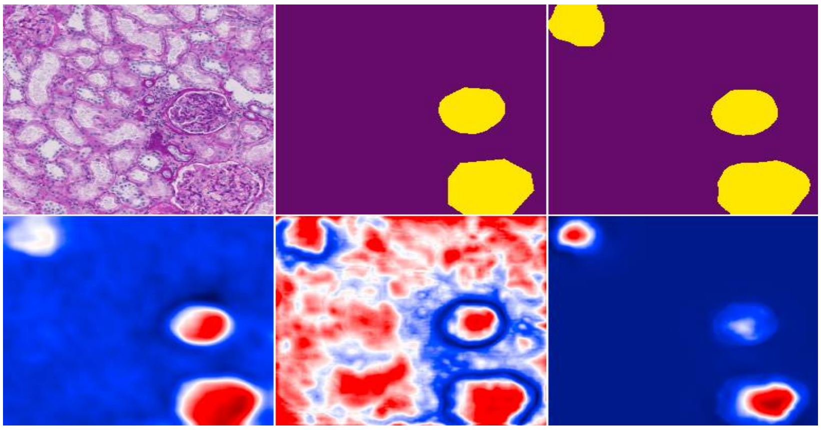
Input image, ground truth, mask prediction



Interpretation example



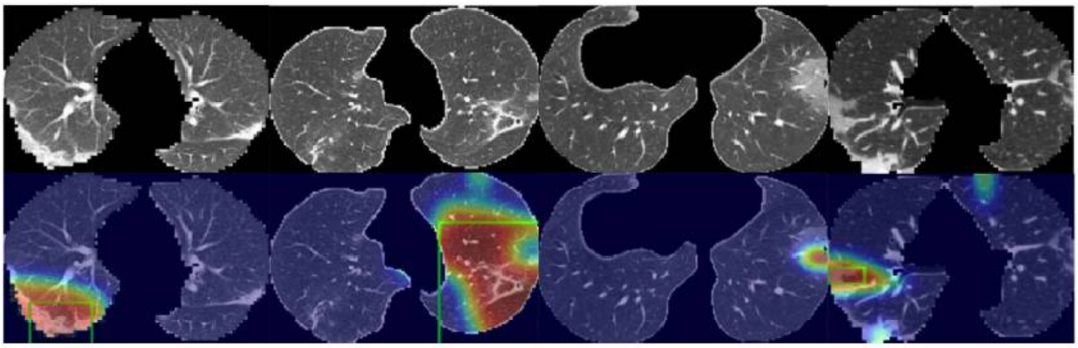
✓ Layer GradCAM, Layer Activation, Layer Gradient SHAP



First and second row:

✓ Crop, ground truth mask, prediction, Layer GradCAM, Layer Activation and Layer Gradient SHAP

COVID-19 positive patients

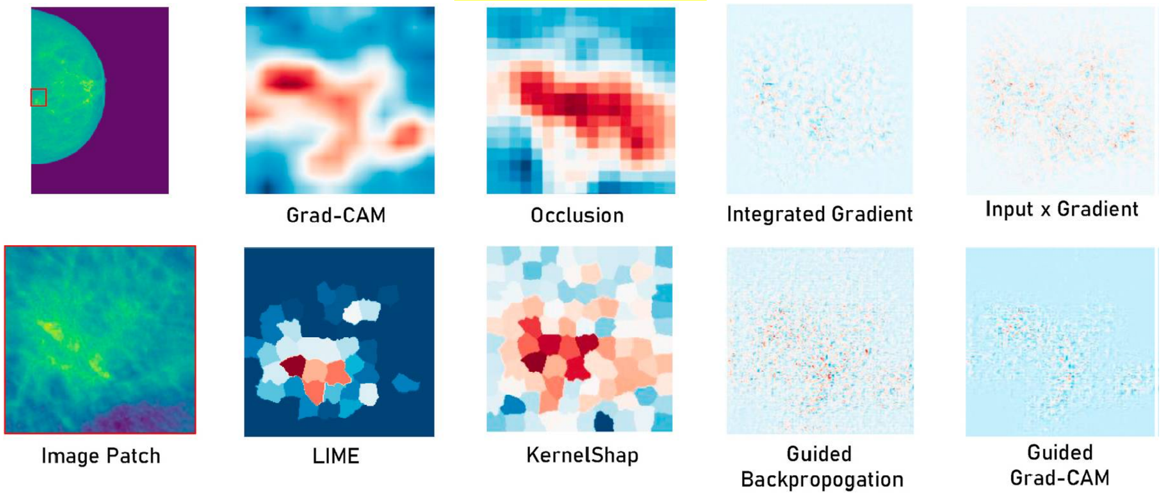


✓ First row : original CT-scan image slices

✓ Second row : heatmaps of CAMs $Aconv$ with bounding boxes confined to the infected areas

xAI.Med. 2022-

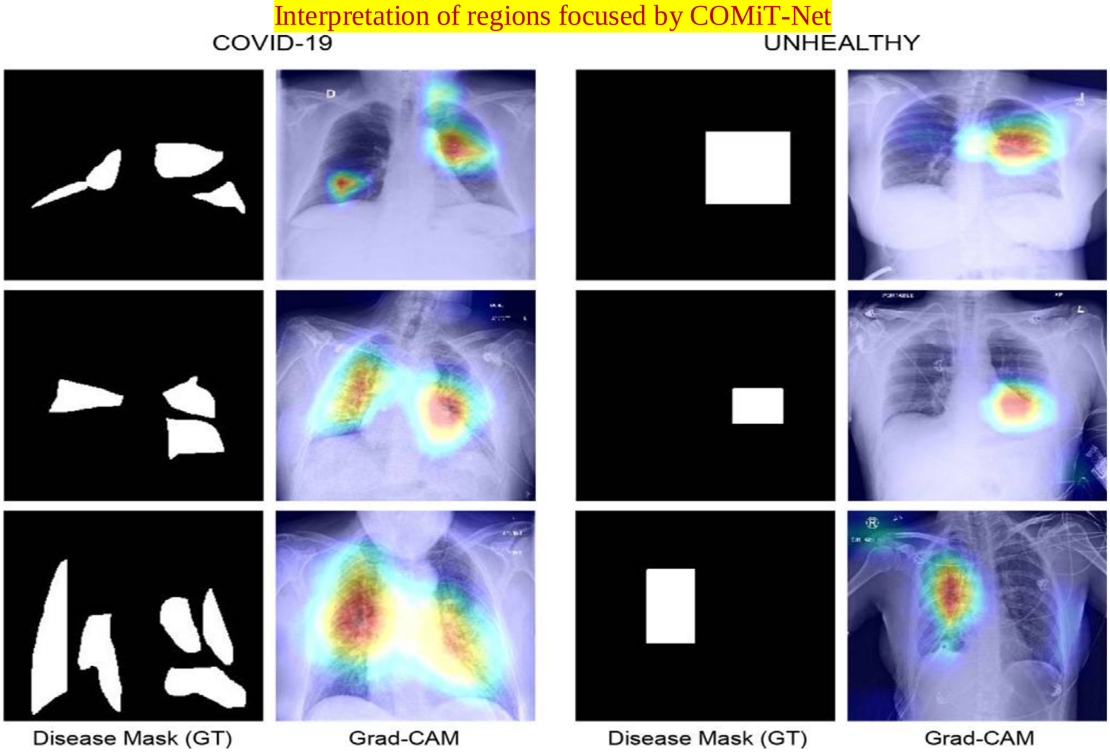
Attribution maps
Breast mass in an image patch extracted from a digitalmammogram



- ✓ First column : image patch extracted from the mammogram that is provided to DL model
- ✓ Red color corresponds to important regions
- ✓ Blue color corresponds to irrelevant image regions for classification

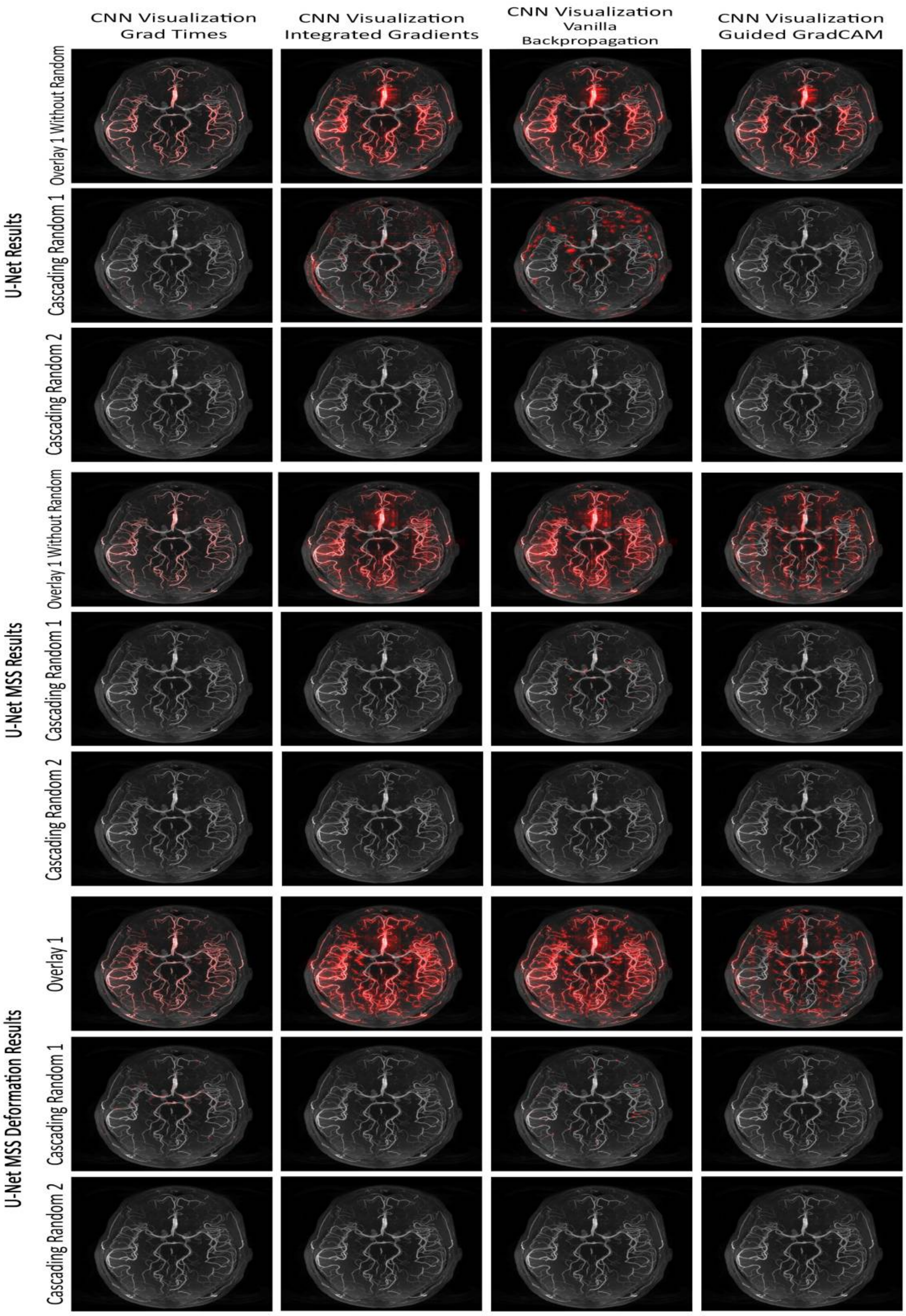
xAI.Med. 2022-75

Interpretation of regions focused by COMiT-Net

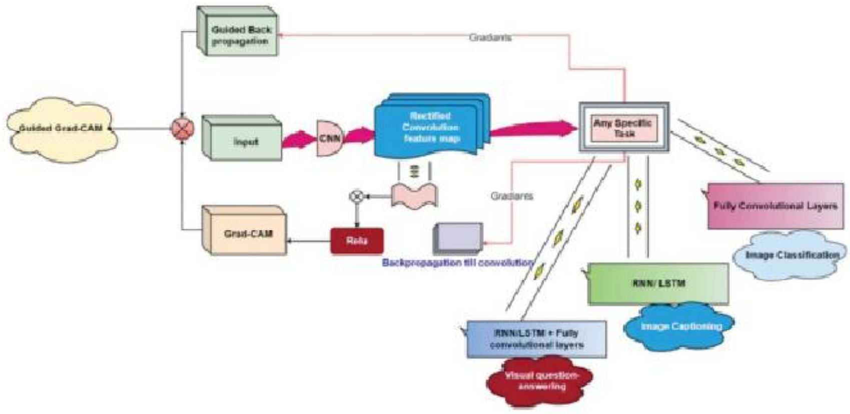


- ✓ Lung annotation helps comit-Net to focus on unhealthy regions

Outputs


CNN Visualization
Grad Times
CNN Visualization
Integrated Gradients
CNN Visualization
Vanilla
Backpropagation
CNN Visualization
Guided GradCAM
U-Net Results
Overlay 1 Without Random
Cascading Random 1
Cascading Random 2
U-Net MSS Results
Overlay 1 Without Random
Cascading Random 1
Cascading Random 2
U-Net MSS Deformation Results
Overlay 1
Cascading Random 1
Cascading Random 2

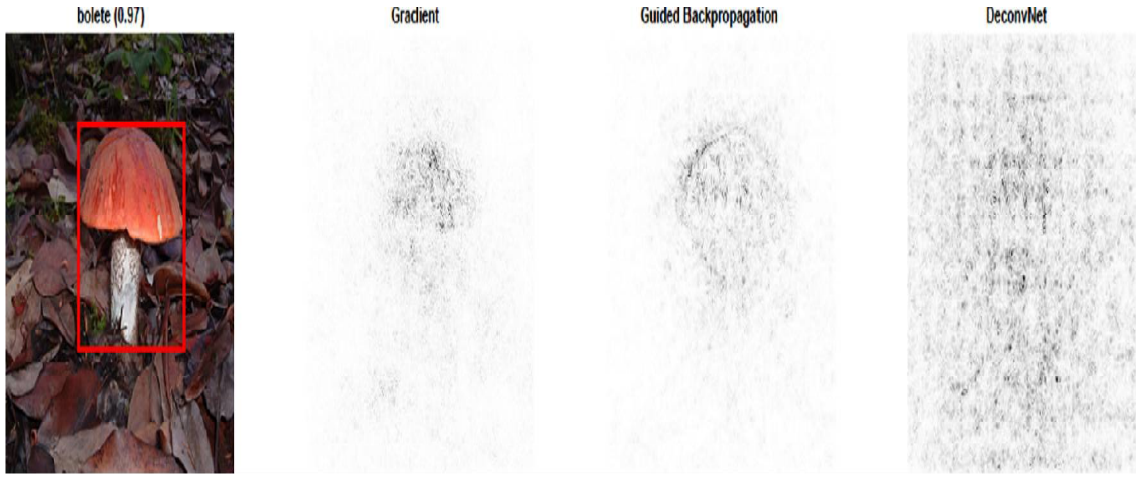
xAI.Med.	2022-205



✓ Grad CAM

xAI.Med.	2022-58

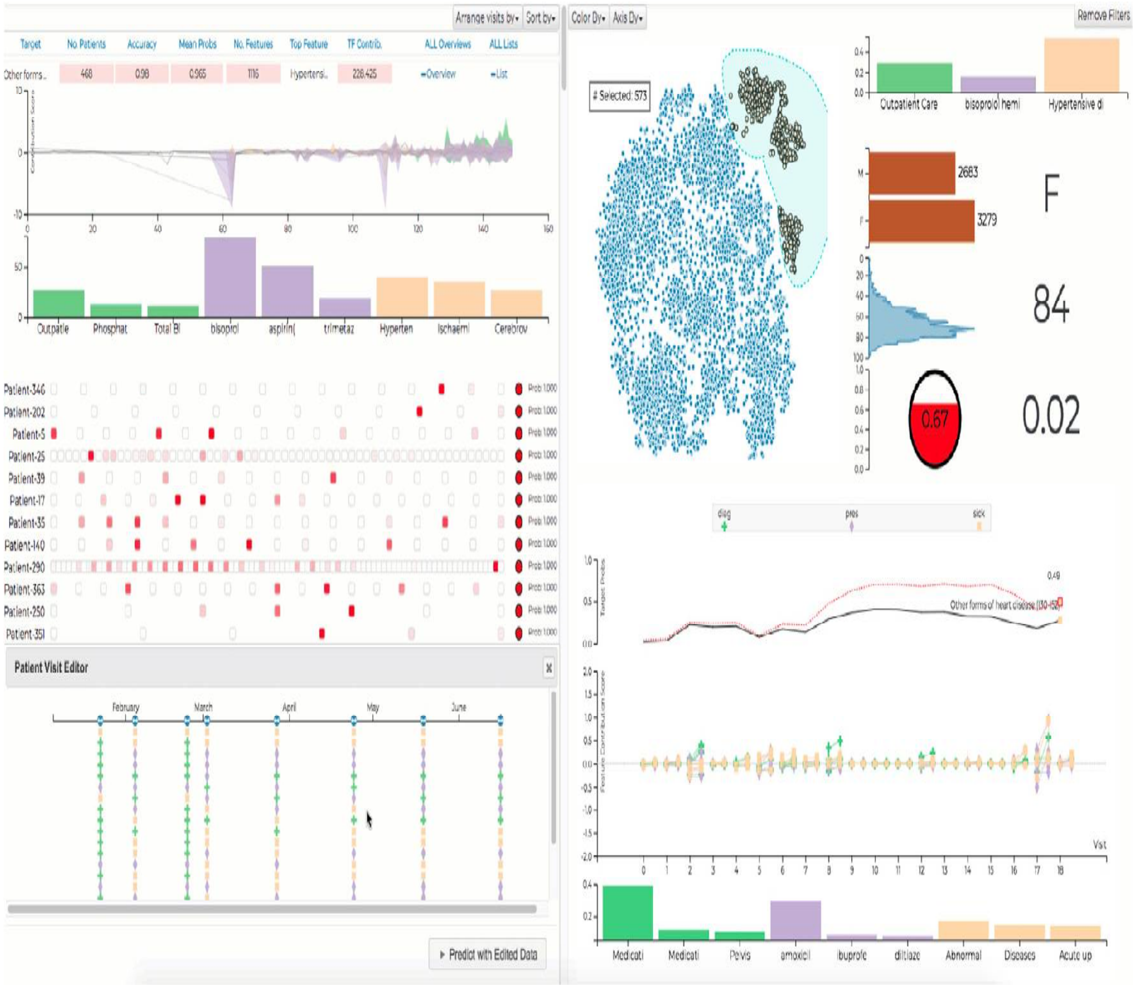
Examples of gradient-based methods



tSNE plot

t-distributed stochastic neighbor embedding

what-if analysis tools



[Script$$

(Left) RETAINVis83 RNN ''RETAIN'' model

- Shows contribution to the overall outcome of patient visits through feature contribution score, representing drugs (violet), diagnosis (yellow), or physiological markers (green) for each visit

(Bottom) Patient list

- Shows individual patients in a row of rectangles. In the patient list, users can select a patient of interest to view details, shown below, and edit patients to conduct a what-if analysis.
-

(Right top) Dimensionality reductiontechniques like t-SNE

! t-distributed stochastic neighbor embedding result in the blue scatterplot to gain an overview and then build patient cohorts using lasso selection tools and take a look at the distribution for demographic information like biological sex, age, and risk prediction scores (red circle).

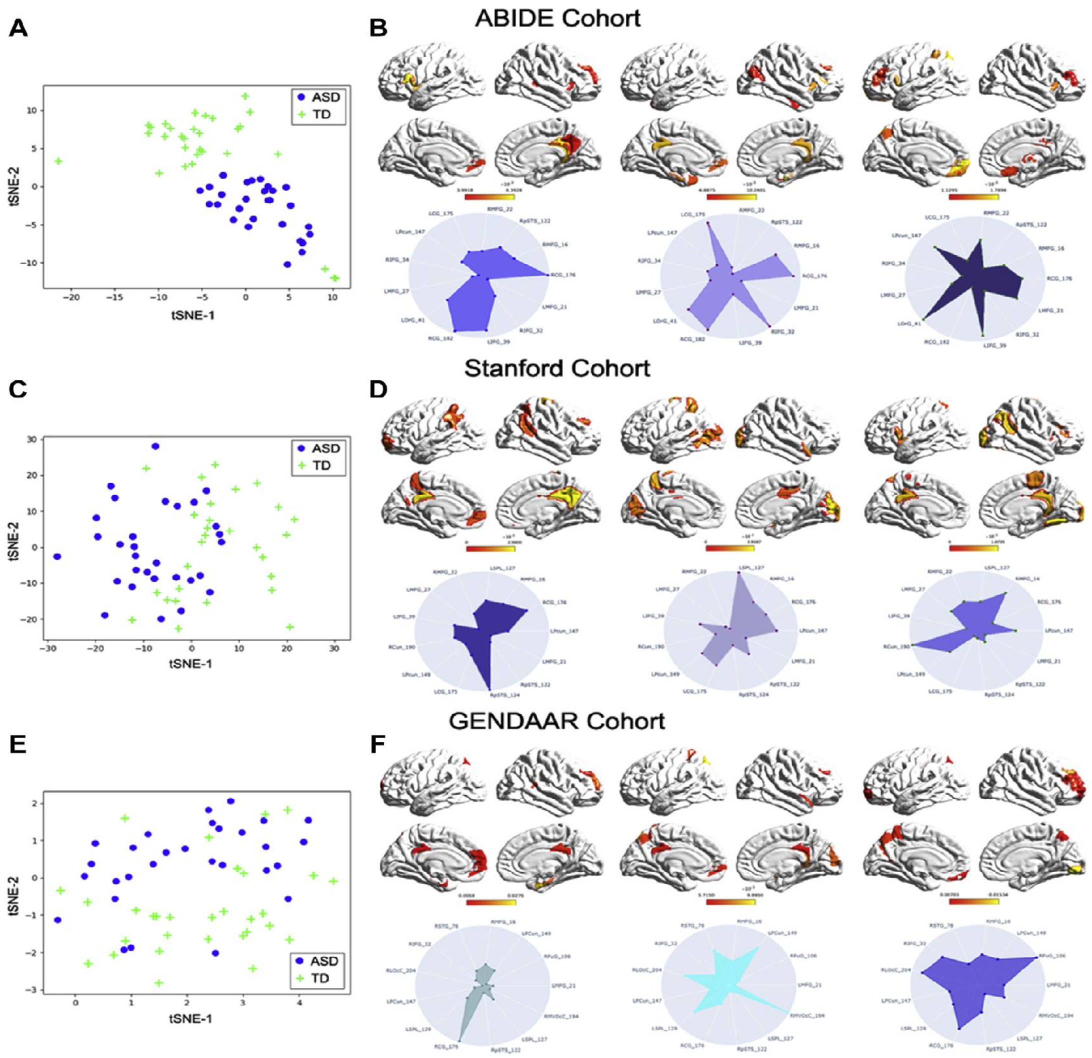
(Right bottom)

+ Contribution scores for each visit and patient details are shown after the updated results of the what-if analysis

+ In the middle, an area chart shows aggregated contribution scores of nine medical codes over time. It shows mean and standard deviation as an area. Users can also see the medical codes and their mean contribution scores in bar chart " $$script]

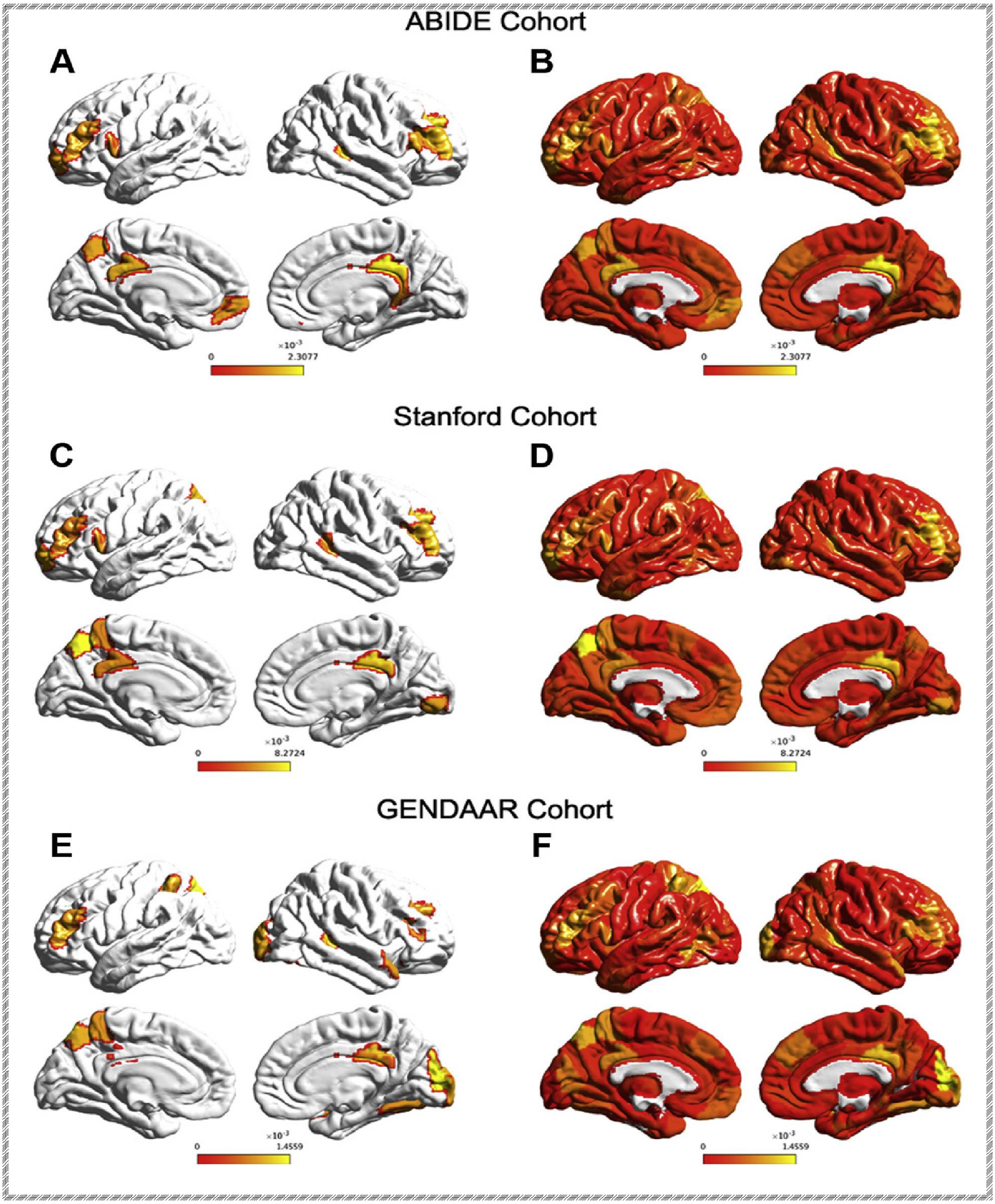
xAI.Med.	tSNEplot; spatiotemporal deep neural network (stDNN)	2022-53

Autism spectrum disorder patients

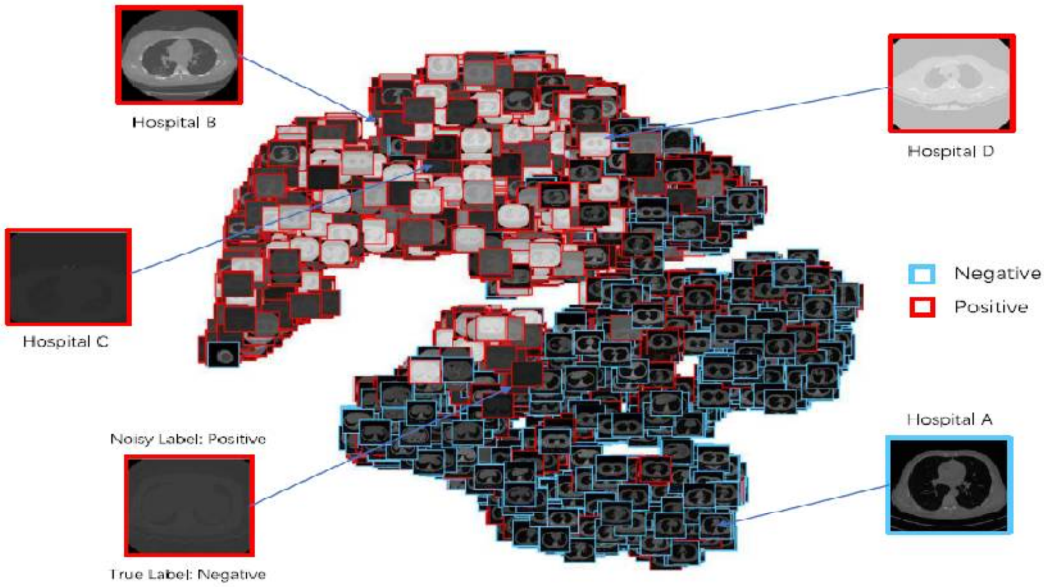


xAI.Med.		2022-53

Autism Brain Imaging Data Exchange

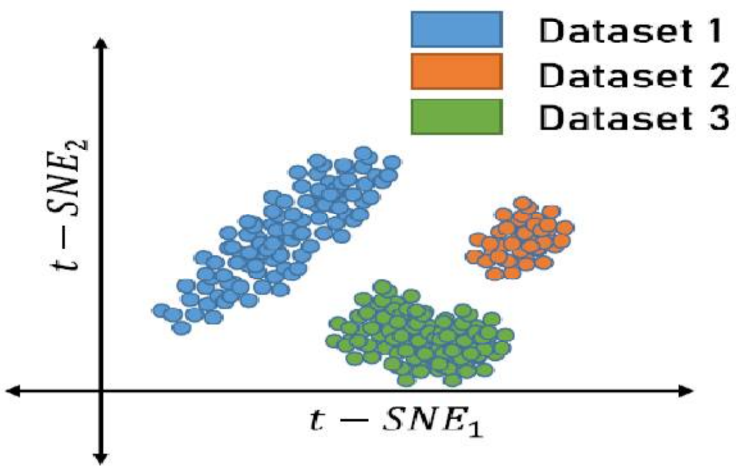


xAI.Med.		2022-62
	T-SNE visualisationof learnt features from CT images	



xAI.Med.		2022-68

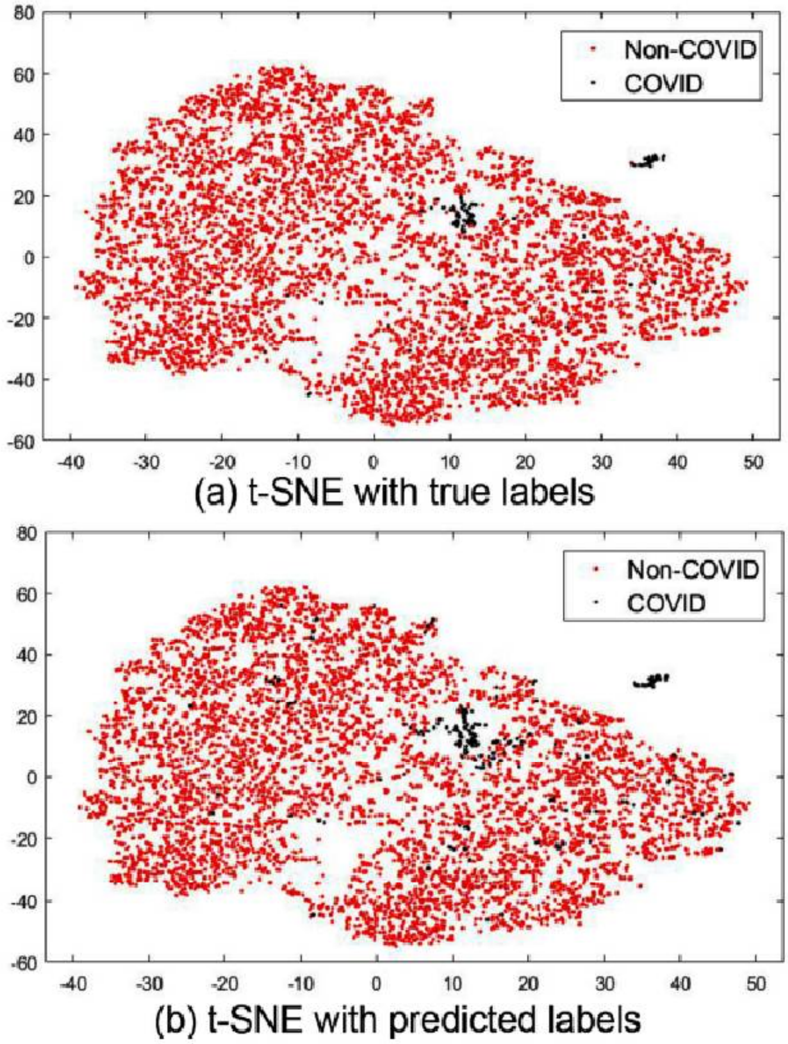
Low dimensional visualization of a fully connected layer of a CNN
Usingt-SNE for interpretability



! Plot shows dataset bias in the DL model due to differences in acquisition

xAI.Med.		2022-75

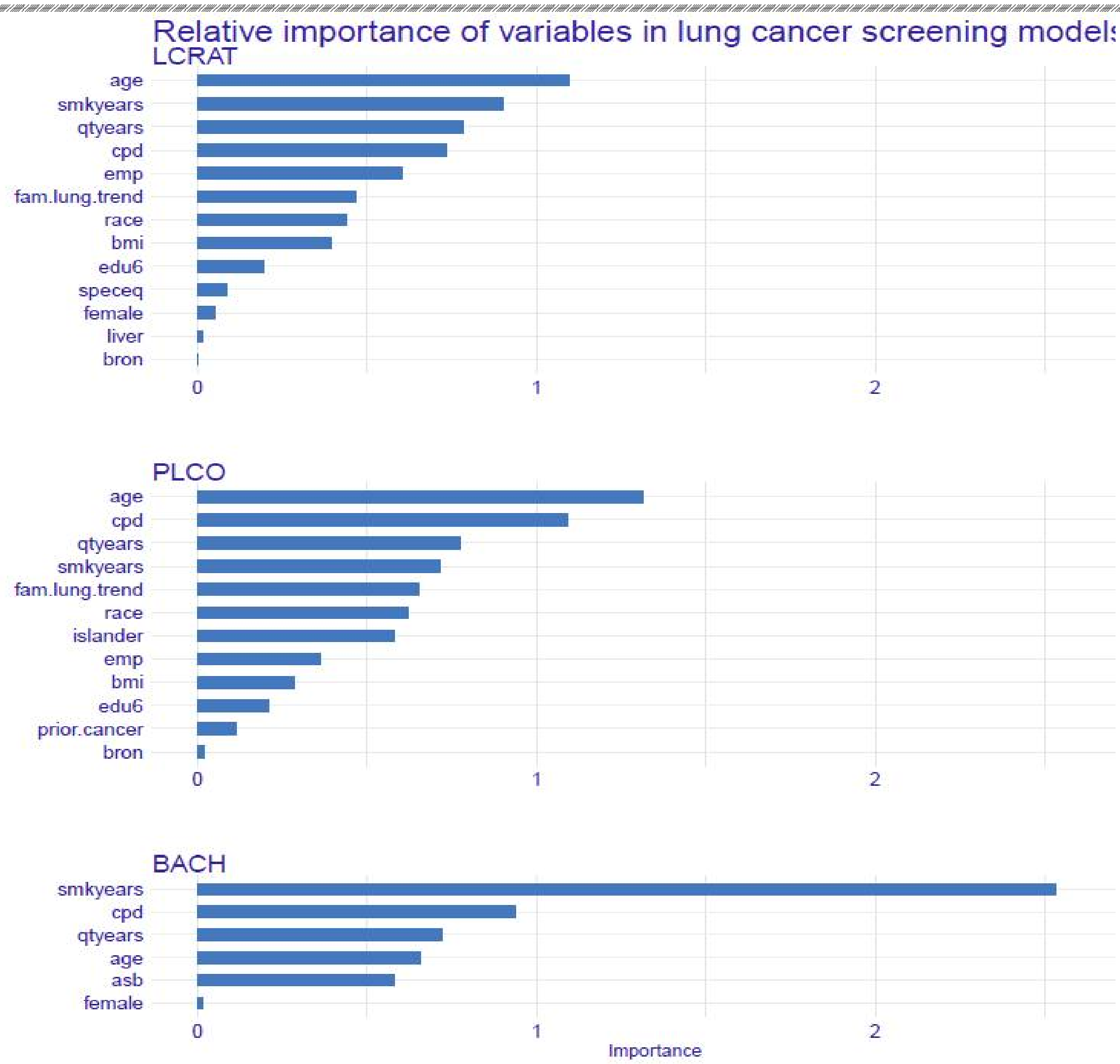
COVID/Non-COVID classification



! Interpretation of feature representation based on (a) ground-truth and (b) predicted labels using t-SNE plot

Variable ImportancePlot (VIP)

xAI.Med.		2022-209

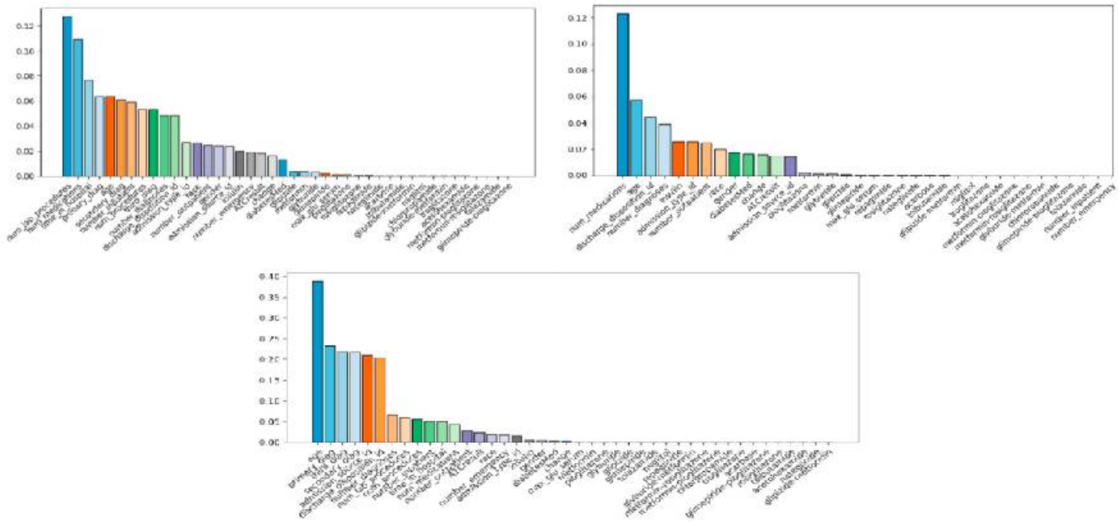


Longer the bar, more important is corresponding variable

✓ asb : asbestos exposure, cpd : cigarettes per day, islander : Pacific Islander ethnicity

xAI.Med.		2022-40

Feature Importance plots (Fip)

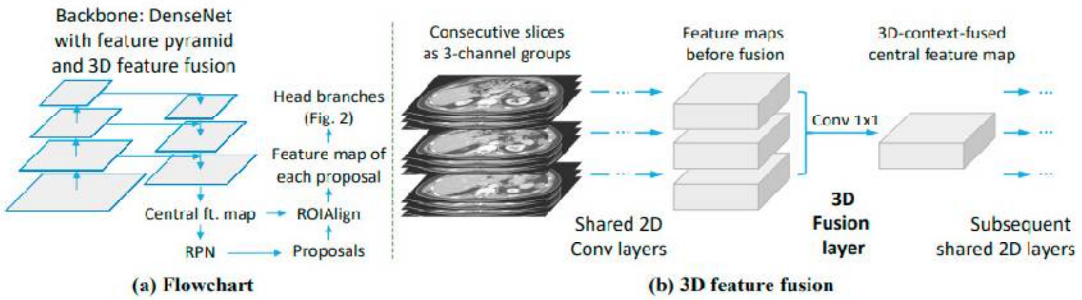


a) Random Forest, b) AdaBoost, and c) K-Nearest Neighbors

Feature Maps

xAI.Med.		2022-59

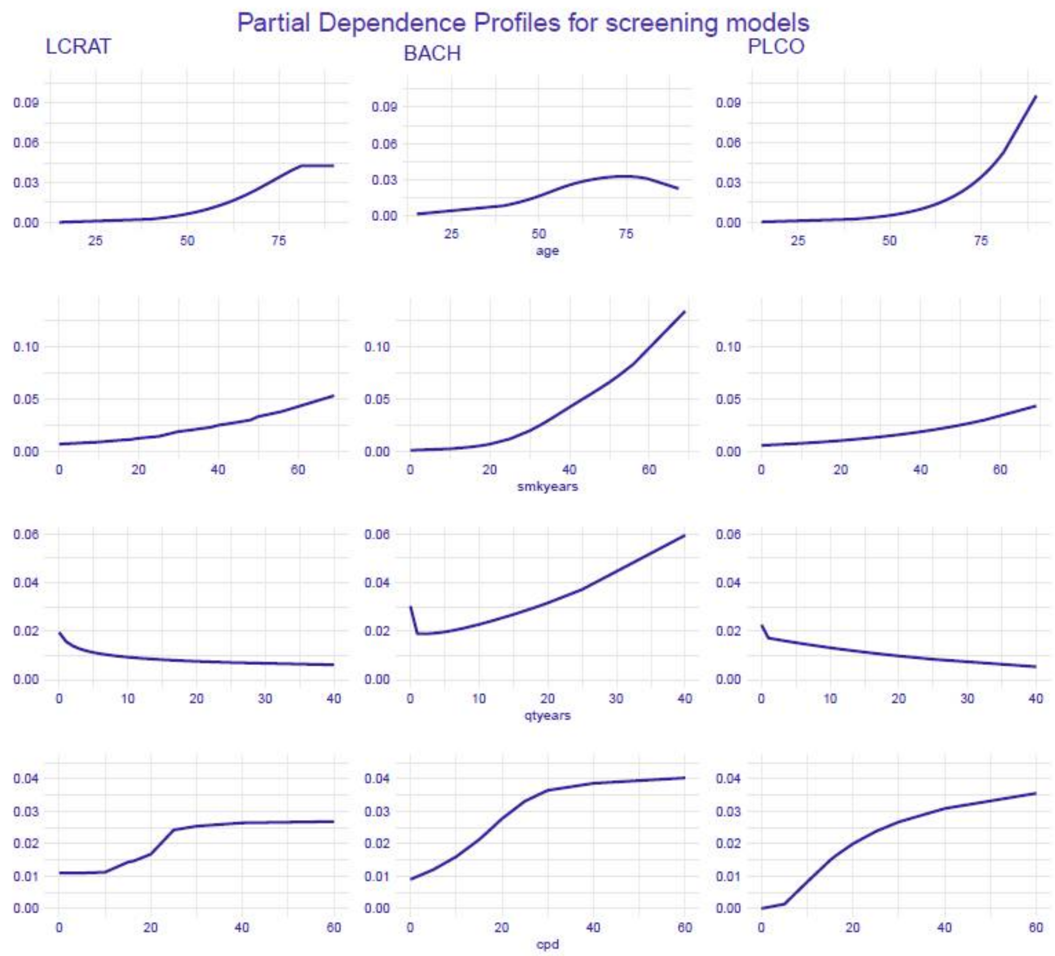
Flow and fusion strategy of feature maps



Partial depencence Model

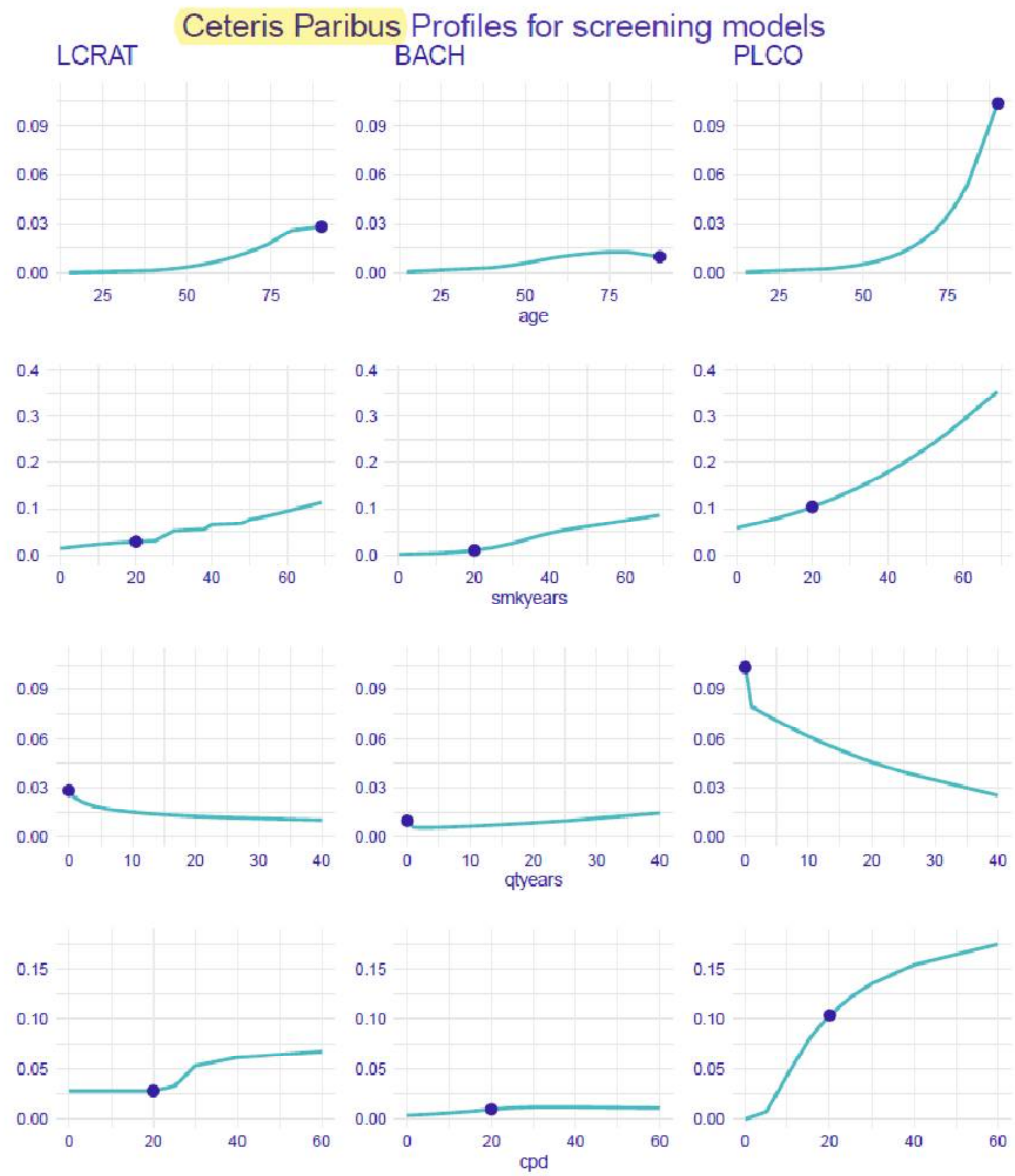
xAI.Med.		2022-209

Relationship between a variable and an average model prediction



x-axis : range of available values of the analyzed variable
y-axis : values of the mean prediction

xAI.Med.	Ceteris Paribus Profiles	2022-209
	Dependencies between continuous variables vs. Prediction for the selected patient	



- ✓ Blue dot : real value of each variable
- ✓ Green line: change of prediction with the change of the variable's value

Tables xAI

XAI APPROACHES USED IN MEDICINE AND HEALTHCARE AS WELL AS THEIR TYPES

Type	Method	Post-hoc	Intrinsic	Local	Global	Application	Ref.
Dimension reduction using XAI	Optimum feature selection	✗	√	✗	√	Prediction of enhancers that are cell type specific	[109]
	Laplacian Eigenmaps	✗	√	✗	√	Classification of Brain tumor from MRI	[107]
	Sparse-balanced SVM	✗	√	✗	√	Diagnosis of type 2 diabetes at an early stage	[158]
	Cluster analysis and LASSO	✗	√	✗	√	Classification of lung cancer patients	[108]
	Sparse DL	✗	√	✗	√	Glioblastoma multiforme long-term survival prediction	[110]
	Optimal feature selection	✗	√	✗	√	Estimation of side effect for drug	[106]
Rule extraction and knowledge	Decision rules	✗	√	✗	√	Stroke Prediction	[126]
	Rule-based system	✗	√	✗	√	Forecasts for 30-day readmissions and pneumonia risk.	[14]
	Mimic learning	√	✗	√	√	Predicting outcomes in the ICU for severe lung injury	[120]

Type	Method	Post-hoc	Intrinsic	Local	Global	Application	Ref.
distillation using XAI	Textual or visual justification	√	✗	√	√	Classification of breast mass	[125]
	Visualization rules	√	✗	✗	√	Clinical diagnosis of diabetes and breast cancer	[121]
	Lists of Bayesian rule	✗	√	✗	√	Stroke prediction	[120]
	Fuzzy rules	✗	√	✗	√	Prediction of in-hospital mortality for all cases	[123]
Feature importance or selection using XAI	DeepLIFT	√	✗	✗	√	Detection of splice site	[133]
	Feature weighting	√	✗	✗	√	Prediction of ICU mortality for all-causes	[128]
	DeepLIFT	√	✗	√	√	Ophthalmic diagnosis	[137]
	Feature marginalization	√	✗	√	√	Diagnosis of inflammatory or microbiota bowel diseases from skin and gut	[131]
Attention mechanism in XAI	Attention	✗	√	√	✗	Prediction of future hospitalization from EHR	[143]
		✗	√	√	√	ICU clinical events predictions	[140]
		✗	√	√	✗	HIV genome integration site prediction	[142]
		✗	√	√	√	Prediction of clinical risk for cardiac cataract or failure	[145]
		✗	√	√	✗	Prediction of heart failure	[144]
	MLCAM	✗	√	√	✗	Localization of brain tumor	[145]
	Grad-CAM	✗	√	√	✗	Appendicitis diagnosis	[147]
	Grad-CAM	✗	√	√	✗	Hypoglycaemia detection using ECG	[149]
Surrogate representation using XAI	LIME	√	✗	√	✗	Prediction of early puberty in the central region	[152]
		√	✗	√	✗	Construction of survival models	[153]
		√	✗	√	✗	Diagnosis of autism spectrum disorder	[159]
	Rule-based XAI	√	✗	√	✗	Patient medications, diagnosis and readmission prediction	[155]

DISCUSSION OF SHAP XAI ALGORITHMS IN THE LITERATURE

Ref	Disease Diagnosis	ML Models Used	Work Summary
[185]	Chronic Obstructive Pulmonary Disease	Gradient boosting machine (GBM)	The work ranks the attributes crucial to the GBM model according to their average absolute SHAP values that reflect the impact of each feature on a prediction.
[186]	Parkinson's disease diagnosis	deep forest (gcForest), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM) and random forest (RF)	The classifiers determine how much the SHAP value contributes to the features. When 150 features are considered, SHAP-gcForest achieves 91.78 percent classification accuracy. With 50 features, SHAP-LightGBM combined with LightGBM yields 91.62% accuracy.
[187]	predicting acute kidney injury progression	XGboost , Logistic Regression (LR)	The SHAP value increases as creatinine levels climb until it reaches about 5 mg/dL. As the Furosemide Stress Test (FST) is increased until it reaches 100 ml/min, the SHAP result decreases.
[188]	COVID-19 Prediction	XGboost model, LGBM, gradient boosting classifier (GBC), categorical boosting (CatBoost), RF	Harris hawks optimization (HHO) approach is employed to fine-tune the hyperparameters of a few cutting-edge classifiers, including the HHO-RF, HHO-XGB, HHO-LGB, and ensemble methods, in order to increase classification accuracy.
[189]	Prediction of COVID-19 diagnosis based on symptoms	GBM	A gradient-boosting machine model created with decision-tree base-learners was used to produce predictions. SHAP values calculate the contribution of each feature to overall model predictions by average across samples
[190]	Efficient analysis of COVID 19	Support Vector Machine (SVM), Naive Bayes (NB), Multiple Linear Regression (MLP), K-Nearest Neighbors (KNN), RF, LR, and Decision Tree (DT), Keras Classifier	Simple classification techniques like random forest can outperform keras with Boruta for feature selection when the dataset size is not too large. Computing the information gain values for each attribute in Clinical Data1 shows the importance of each attribute.
[191]	COVID 19 Diagnosis	SqueezeNet, LIME	LIME and SHAP are compared for COVID diagnosis using SqueezeNet to recognize pneumonia, COVID-19, and normal lung imaging. Results show LIME and SHAP can boost the model's transparency and interpretability.
[192]	COVID-19 vaccine prioritization	Random Forest, XGBoost classifiers, XAI	CovidXAI predicts a user's risk group using Random Forest and XGBoost classifiers. CovidXAI uses 24 criteria to define an individual's risk category and vaccine urgency
[193]	COVID-19 Pneumonia Classification	XGBoost, Random Forest	The goal is to provide grounds for understanding the distinctive COVID-19 radiographic texture features using supervised ensemble ML methods based on trees through the interpretable SHAP approach. SHAP recursive feature elimination with cross-validation is used to select features. The best classification model was XGBoost, with an accuracy of 0.82 and a sensitivity of 0.82.
[194]	Lung cancer hospital length of stay prediction	Random Forest, logistic regression	This paper introduces a predictive LOS framework for lung cancer patients utilizing ML models. Using SHAP, the output of the predictive model (RF) with SMOTE class balancing approach is understood demonstrating the most relevant clinical factors that contributed to predicting lung cancer LOS using the RF model.

xAI.Med.		2023-101

xAI approaches explainable in terms of the characteristics of explainability

Method	Explanation type	Scope	Fidelity		Interpretability	
			Soundness	Completeness	Parsimony	Clarity
Post-hoc explanation	Attribution	Local	General quantitative metric	Unavailable matrices	Satisfied if an instance or feature is human-intelligible.	General quantitative metric
	Attribution	Global	General quantitative metric	Unavailable matrices	Satisfied if an instance or feature is human-intelligible.	Satisfied
	Model	Local	General quantitative metric	Satisfied	General quantitative metric	Unavailable matrices
	Model	Global	General quantitative metric	Satisfied	General quantitative metric	If the model is incapable of providing various rationales, the model is satisfied.
	Example	Local	Unavailable matrices	Unavailable matrices	Satisfied if an instance or feature is human-intelligible.	Unavailable matrices
	Example	Global	Unavailable matrices	Unavailable matrices	Satisfied if an instance or feature is human-intelligible.	Satisfied

xAI.Med.		2022-10

Table 1. Categories of ML, concepts, typical methods, and their representative applications

Learning category	Concepts	Representative methods	Applications
Supervised	learning from labeled data to predict class/clinical measures	SVM, random forest, sparse learning, ensemble learning	Disease diagnosis, prognosis, treatment outcome prediction
Unsupervised	learning from unlabeled data to uncover structure and identify subgroups	Hierarchical clustering, K-means, PCA, CCA	Disease subtyping, normative modeling, identify behavioral and neurobiological dimension
Semi-supervised	learning from both labeled and unlabeled data to perform supervised or unsupervised tasks	multi-view learning, Laplacian regularization, semi-supervised clustering	multi-modal analysis, joint disease subtyping and diagnosis, prediction with incomplete data
Deep	learning hierarchies and non-linear mappings of features for higher-level representations, can be either supervised or unsupervised	CNN, deep autoencoder, GCN, RNN, LSTM, GAN	a large class of generic learning problems
Reinforcement	solving temporal credit assignment problems, optimal control, trial-and-error learning	temporal difference learning, Q-learning, actor-critic model, dynamic programming	online control, modeling of decision-making and choiced behaviors

xAI.Med.		2022-

Methods for explainability

- Stage (Ah: ante hoc; Ph: post hoc)
- Scope (L: local; G: global)
- Forms of explanations (N: numeric; R: rules; T: textual; V: visual)
- Type of models

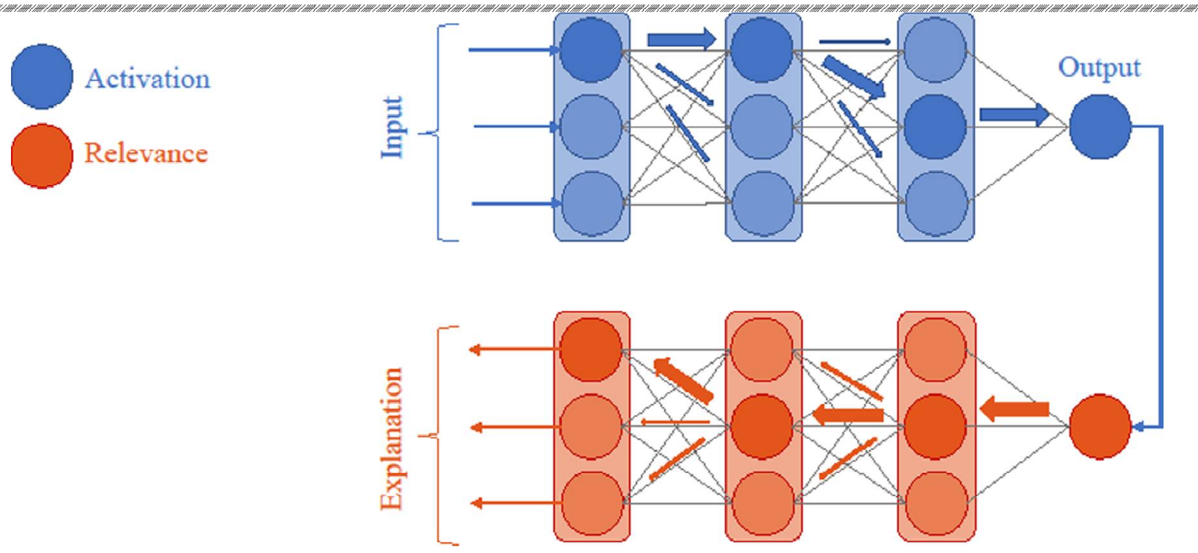
Methods for Explainability	References	Stage		Scope			Form			Models for Primary Task
		Ah	*Ph*	*L*	*G*	*N*	*R*	*T*	*V*	
Ada-WHIPS	[112]		✓	✓			✓			EM
ALMMo-0*	[116]	✓		✓	✓		✓			NFM
Anchors	[96]		✓	✓	✓	✓	✓	✓	✓	MA
ANFIS	[66,118,137,169]		✓	✓	✓	✓	✓	✓	✓	FM; GA; NN;
ApparentFlow-net	[124]	✓		✓					✓	NN
Attention Maps	[79,149,151,152]		✓	✓					✓	NN
BB-BC IT2FLS	[178]	✓		✓		✓	✓		✓	FM
BEN	[69]		✓	✓					✓	NN
BN	[71,165]	✓		✓				✓	✓	BM
BRL	[126]	✓			✓		✓			BM
CAM	[140,141]		✓		✓				✓	NN
Candlestick Plots	[161]		✓		✓				✓	NN
CART	[127]	✓		✓			✓			TM
CASTLE	[47]		✓	✓	✓	✓			✓	MA
Causal Importance	[68]		✓		✓	✓				NN
CFCMC	[61]	✓			✓			✓		FM
CGP	[122]	✓			✓	✓				UM
CIE	[49]		✓	✓	✓	✓		✓		EM; NN; SVM
CIT2FS	[134]	✓		✓				✓	✓	FM
Concept Attribution	[132]		✓	✓	✓				✓	NN
Counterfactual Sets	[7,55]		✓	✓				✓		EM; NN
CTree	[108,127]	✓		✓			✓			TM
DeconvNet	[83]		✓	✓	✓				✓	NN
Decision Tree	[54,75]		✓	✓	✓		✓			NN; TM
Deep-SHAP	[143]		✓	✓	✓	✓			✓	MA
DTD	[85]		✓	✓					✓	NN
DIFFI	[145]		✓		✓	✓				EM
ELI5	[139,142]		✓	✓	✓	✓			✓	MA
Encoder–Decoder	[133]	✓		✓					✓	NN
eUD3.5	[102]	✓		✓	✓		✓			EM
ExNN	[60]	✓		✓	✓				✓	NN
FACE	[78]		✓	✓					✓	NN
FDE	[170]		✓	✓	✓	✓				EM; NN; NNM; SVM
Feature Importance	[67,128,144,172]		✓	✓	✓	✓		✓	✓	MA
Feature Pattern	[163]		✓		✓		✓			EM
FFT	[127]	✓		✓			✓			TM
FINGRAM	[88]	✓		✓					✓	TM
FormuCaseViz	[97]		✓	✓					✓	CBR
FURIA	[52]	✓		✓			✓			FM
Fuzzy LeNet	[82]	✓		✓					✓	FM
Fuzzy Relations	[64,104]	✓		✓				✓		FM
gb4-HIPS	[50]	✓		✓	✓		✓			EM
Generation	[159]		✓	✓				✓		NN
GLAS	[77]		✓	✓					✓	MA
GRACE	[58]		✓	✓				✓		NN

Methods for Explainability	References	Stage		Scope			Form			Models for Primary Task
		Ah	*Ph*	*L*	*G*	*N*	*R*	*T*	*V*	
Template-Based Natural Language Generation	[164]	✓		✓				✓		UM
Time-Varying Neighbourhood	[101]		✓	✓		✓			✓	NN
TreeExplainer	[84]		✓	✓	✓	✓			✓	MA
TREPAN	[89]		✓	✓			✓			NN
Tripartite Graph	[162]	✓		✓		✓			✓	UM
WM Algorithm	[157]	✓			✓		✓			FM
xDNN	[116]	✓			✓			✓		NN
XRAI	[147]		✓	✓					✓	NN

Output of Each of feature layers (CNN)

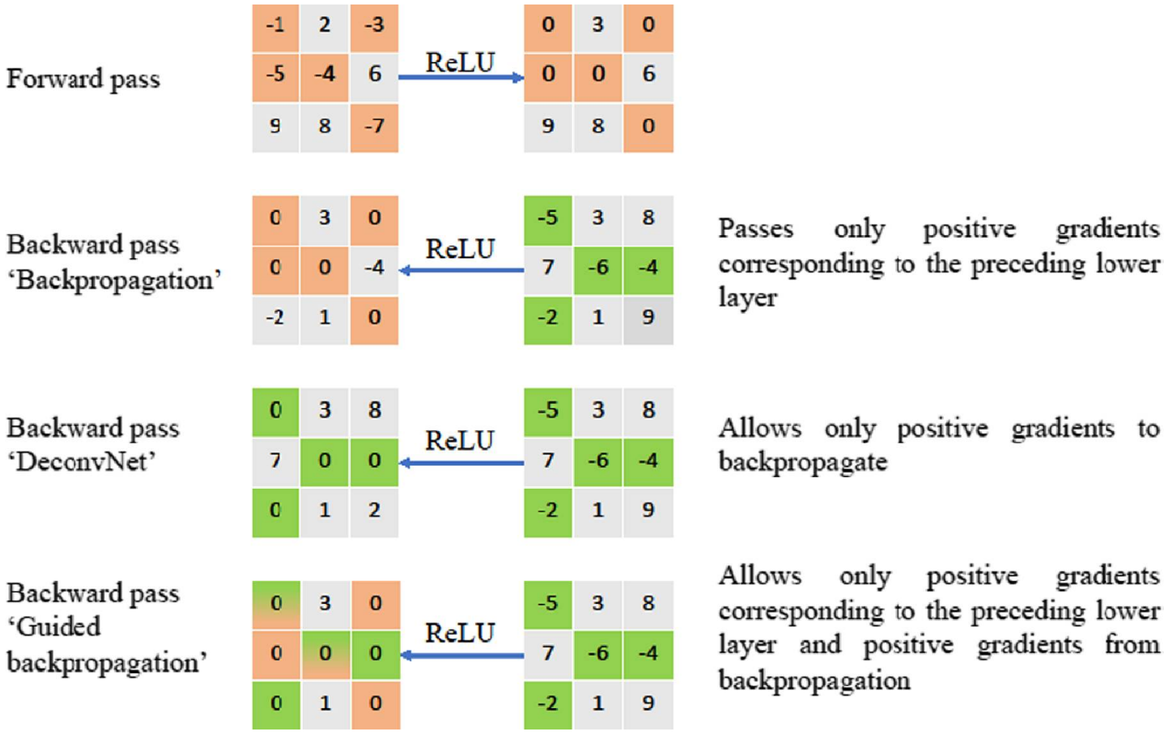
xAI.Med.	2022-58

Layer-wise Relevance Propagation (LRP)



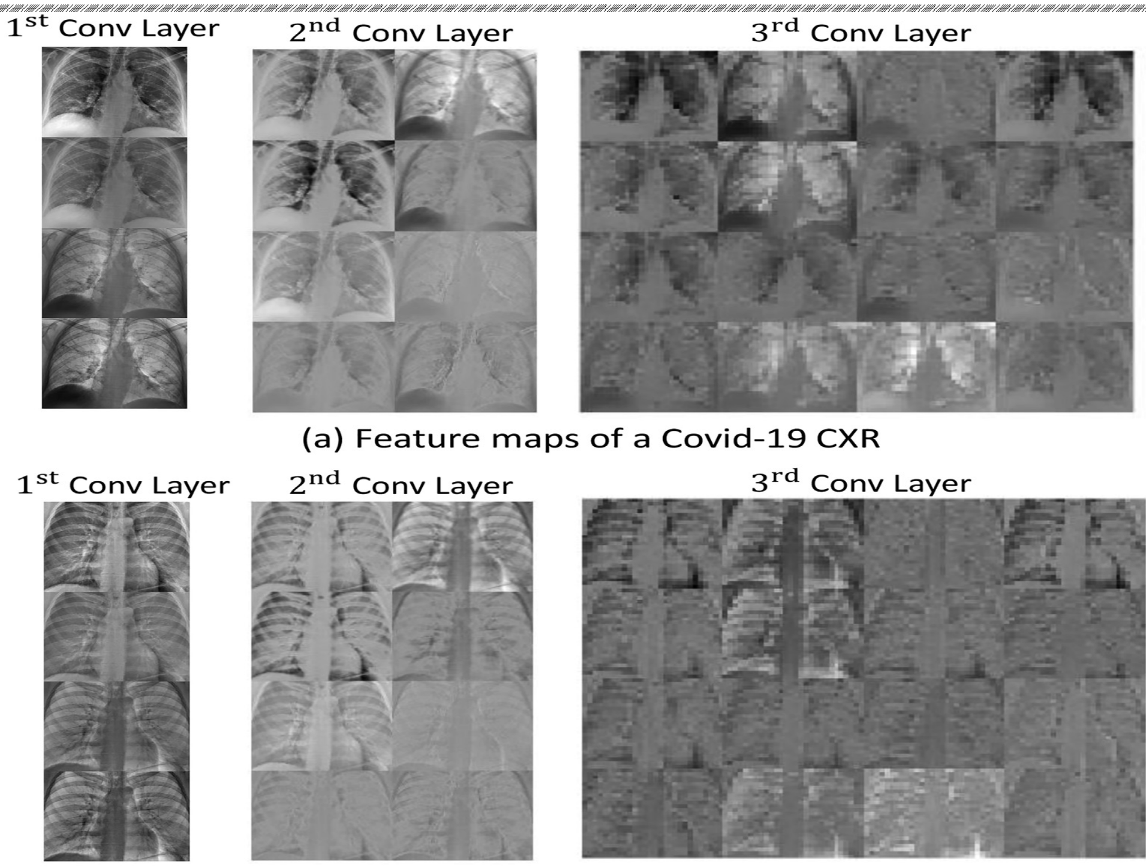
xAI.Med. | 2022-58

Main differences between backpropagation, DeconvNet and Guided backpropagation



xAI.Med. | 2022-46

Feature maps learned by the three convolutional layers of CovNNet on
(a) Covid-19 and (b)No-Covid-19 CXR image

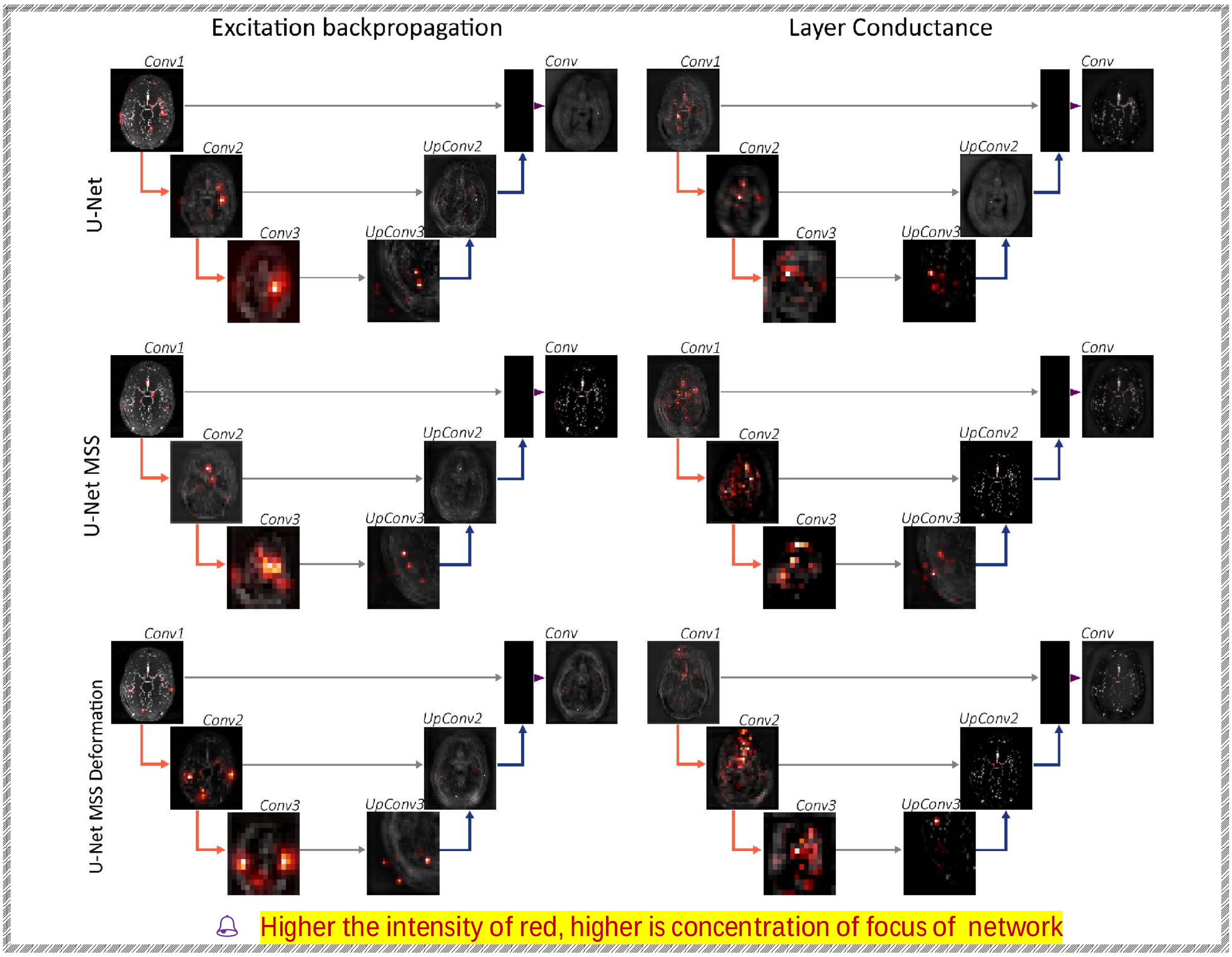


(a) Feature maps of a Covid-19 CXR

(b) Feature maps of a No-Covid-19 CXR

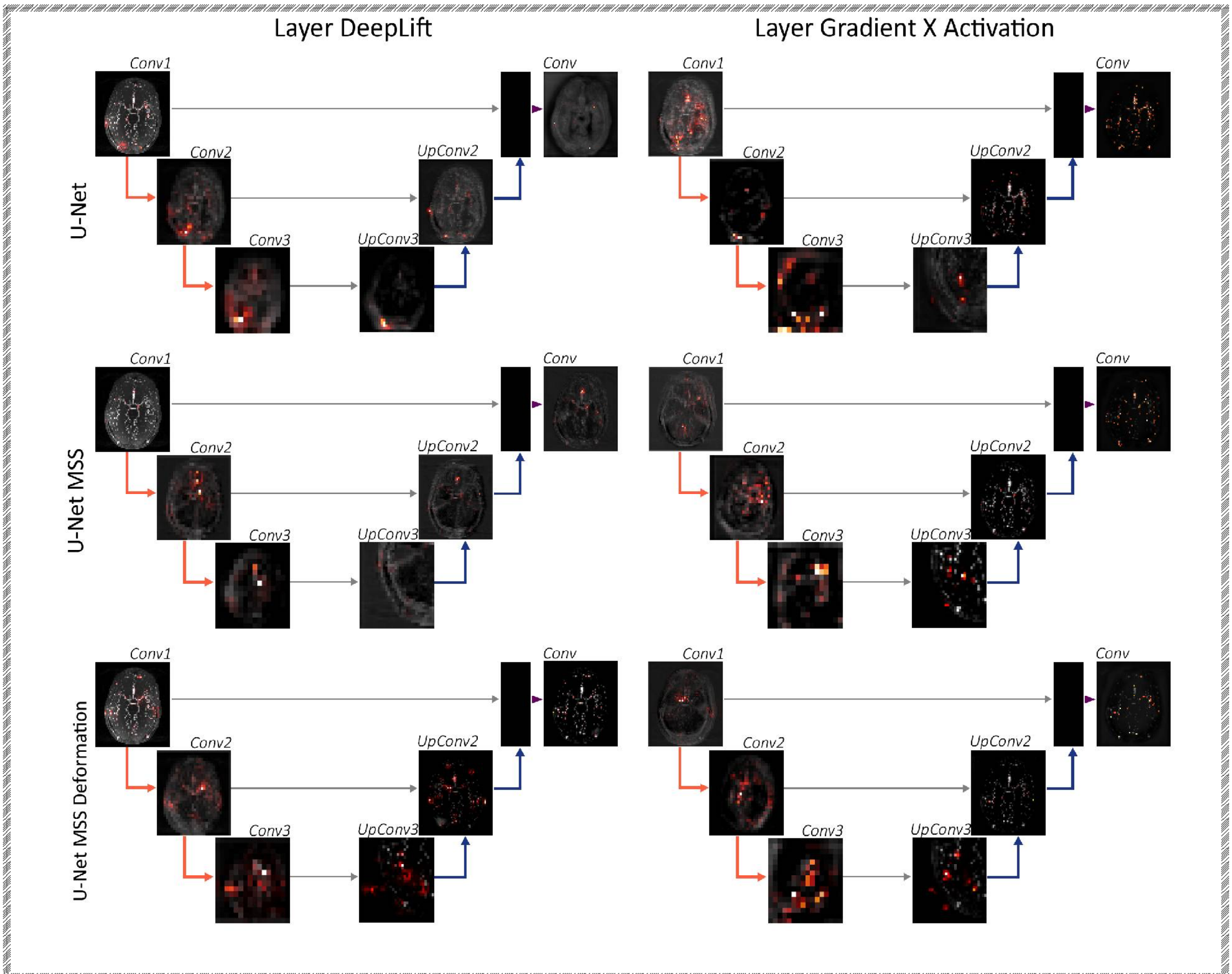
Layer Conductance

xAI.Med.		2022-211
Layer-based interpretability methods Focus of network changes in each layer for three different models		



Layer Gradient activation

xAI.Med.		2022-211



Linguistic Summaries

xAI.Med.		2022-66

Relative linguistic summaries based on short protoforms for mania and hypomania episodes (LS with $T = 1.0$) and extended protoforms for mania and hypomania episodes (LS with $T > 0.5$).

Relative LS based on short protoform	T
Most calls in the state of mania have low spectrum compared to the state of euthymia.	1.0
Most calls in the state of mania have low quality compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low loudness compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low qualty compared to the state of euthymia.	1.0
Relative LS based on extended protoform - HYPOMANIA	*T*
Most calls with low loudness in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with low loudness in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with high loudness in hypomania have high spectrum compared to the state of euthymia.	1.0
Most calls with high loudness in hypomania have high quality compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with low spectrum in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low spectrum in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with high spectrum in hypomania have high loudness compared to the state of euthymia.	1.0
Most calls with high spectrum in hypomania have high quality compared to the state of euthymia.	1.0
Most calls with low quality in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low quality in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with high quality in hypomania have high loudness compared to the state of euthymia.	1.0
Most calls with high quality in hypomania have high spectrum compared to the state of euthymia.	1.0
Relative LS based on extended protoform - MANIA	*T*
Most calls with low loudness in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low loudness in mania have low pitch compared to the state of euthymia.	0.6
Most calls with low loudness in mania have low quality compared to the state of euthymia.	1.0
Most calls with high loudness in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low quality compared to the state of euthymia.	1.0
Most calls with low spectrum in mania have low pitch compared to the state of euthymia.	0.7
Most calls with low spectrum in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low spectrum in mania have low quality compared to the state of euthymia.	1.0
Most calls with medium spectrum in mania have high loudness compared to the state of euthymia.	0.8
Most calls with medium spectrum in mania have low quality compared to the state of euthymia.	0.6
Most calls with high spectrum in mania have high loudness compared to the state of euthymia.	1.0
Most calls with low quality in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low quality in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with high quality in mania have high loudness compared to the state of euthymia.	1.0
Most calls with high quality in mania have high spectrum compared to the state of euthymia.	1.0

xAI.Med.		2022-213

Natural Language explanations generated using GPT-3 text-davinci-003 model for scents observed in the Leffingwell Odor Dataset

Dataset.

Scent	Why the scent?
alcoholic	The molecular property "alcoholic scent" can be explained by the presence of an ethyl/ether O group and the absence of acetal like/methyl groups, two CH2 groups separated by any three bonds, an alkyne group, and an S. These are all very important for the property.
aldehydic	The molecular property "aldehydic scent" can be explained by the presence of an oxygen atom, a lack of an oxygen atom bonded to a secondary carbon atom, the absence of an aromatic/ether oxygen group, the absence of more than two oxygen atoms, and the lack of a sulfur atom.
alliaceous	The molecular property "alliaceous scent" can be explained by the presence of more than one CH2 group that is bonded to two neighbors, one of which is a heteroatom, an atom bonded to three other atoms, one of which is an S, and the presence of an S. The lack of an O and an oxymethylene group (-CH2O-) is also very important for the property.
almond	The molecular property "almond scent" can be explained by the presence of an oxygen atom, an aldehyde/aromatic group, and the absence of an atom bonded to two methyl groups, a CH2 group bonded to two neighbors by non-ring bonds, and an S atom. These are particularly important structure-property relationships for the almond scent.

Continued on next page