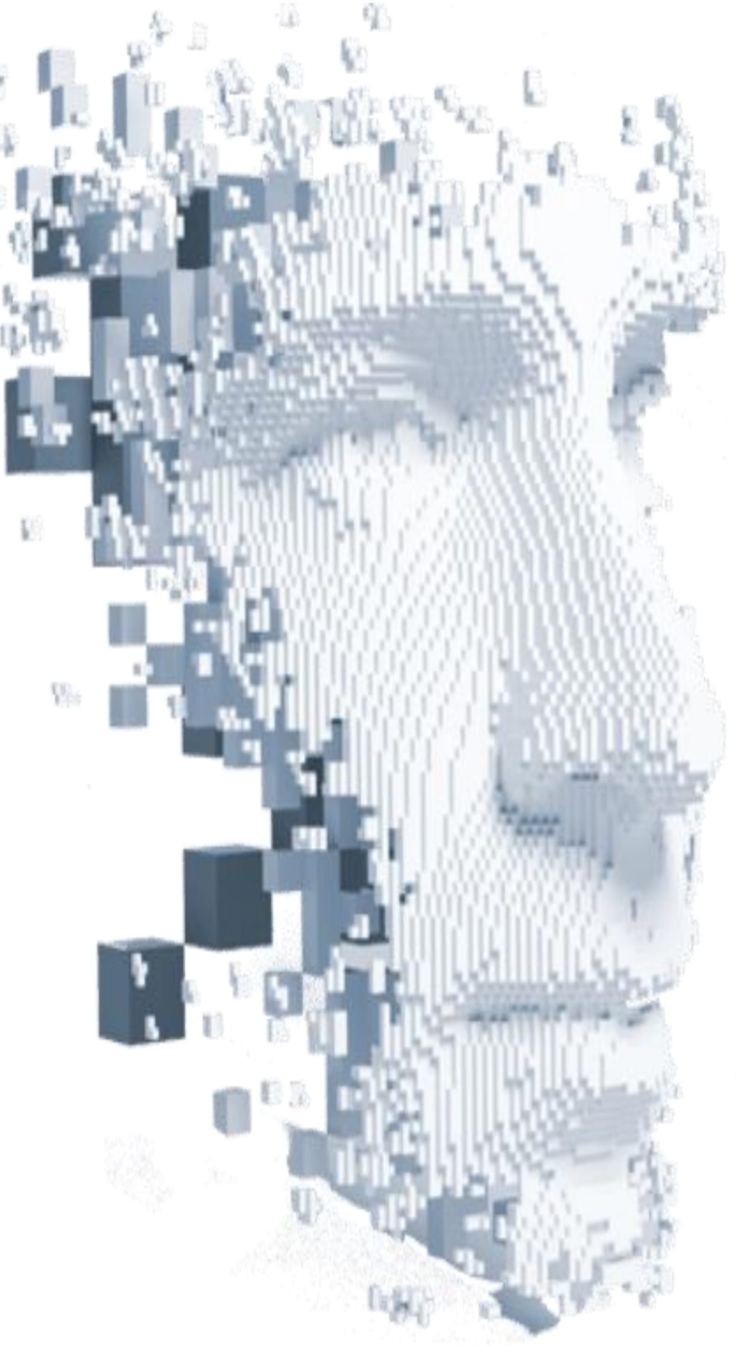
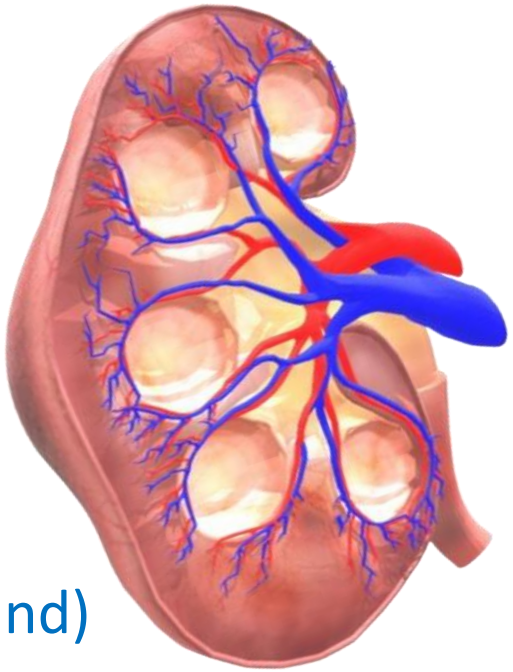


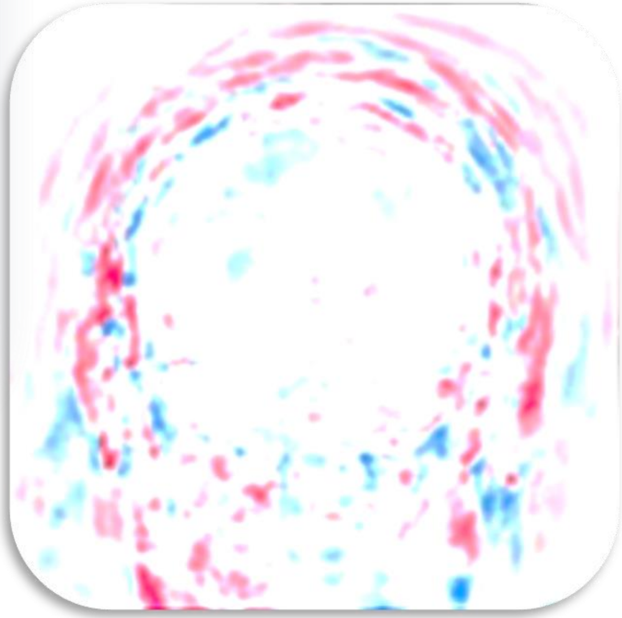
Explainable AI and It's Use in Medical Imaging



MOHAN BHANDARI
(PhD Student, SIIT, Thammasat University, Thailand)

Outlines

- Introduction
- Essence
- XAI Frameworks
- Implementations
- Limitations
- Discussion




Normal Image
Heatmap

Artificial Intelligence

- Simulation of human intelligence processes by machines, especially computer systems.
- Includes learning (acquiring knowledge and skills), reasoning (using logic to make decisions), and self-correction.
- Algorithms

Machine Learning, Deep Learning and Reinforcement Learning

Why many AI algorithms are black box?

- ❑ Decision-making process is often not directly accessible or observable
- ❑ Adapt their internal representations during training, potentially resulting in decisions algorithms
- ❑ Understanding the role and interactions of each layer in the decision-making process can be difficult.
- ❑ The focus on maximizing accuracy can lead to models that are effective but hard to interpret.

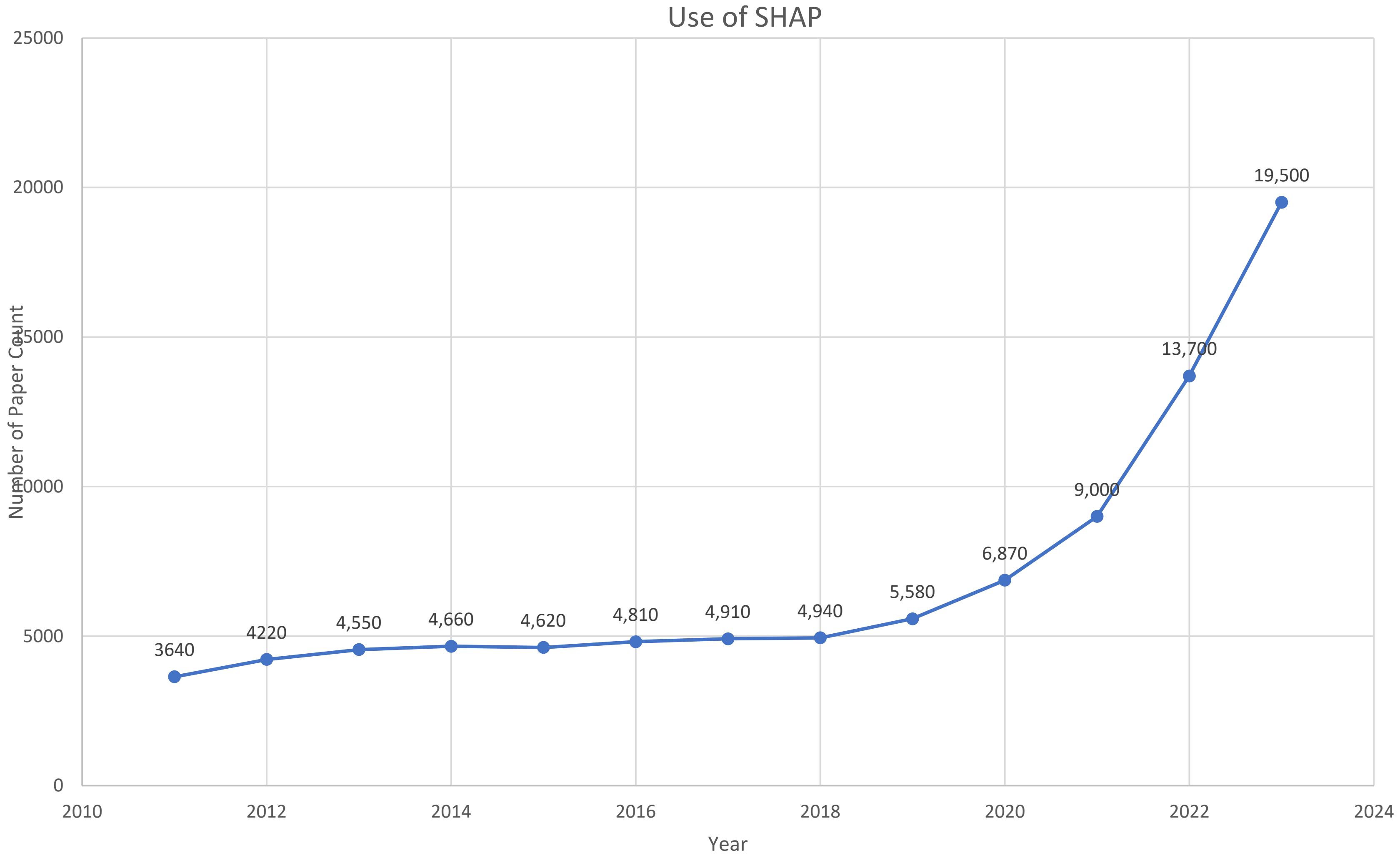
Essence of AI to be explained

Different AI Algorithms

DOG (90%) → Why?

XAI
(LIME)

Gaining Popularity



eXplainable AI (XAI) ??

Framework

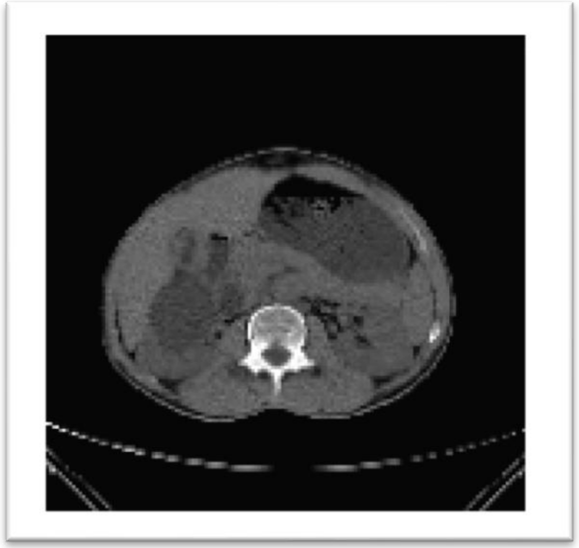
Allow humans to understand and interpret the outcomes and processes of AI systems

Make AI models transparent, understandable, and interpretable for users

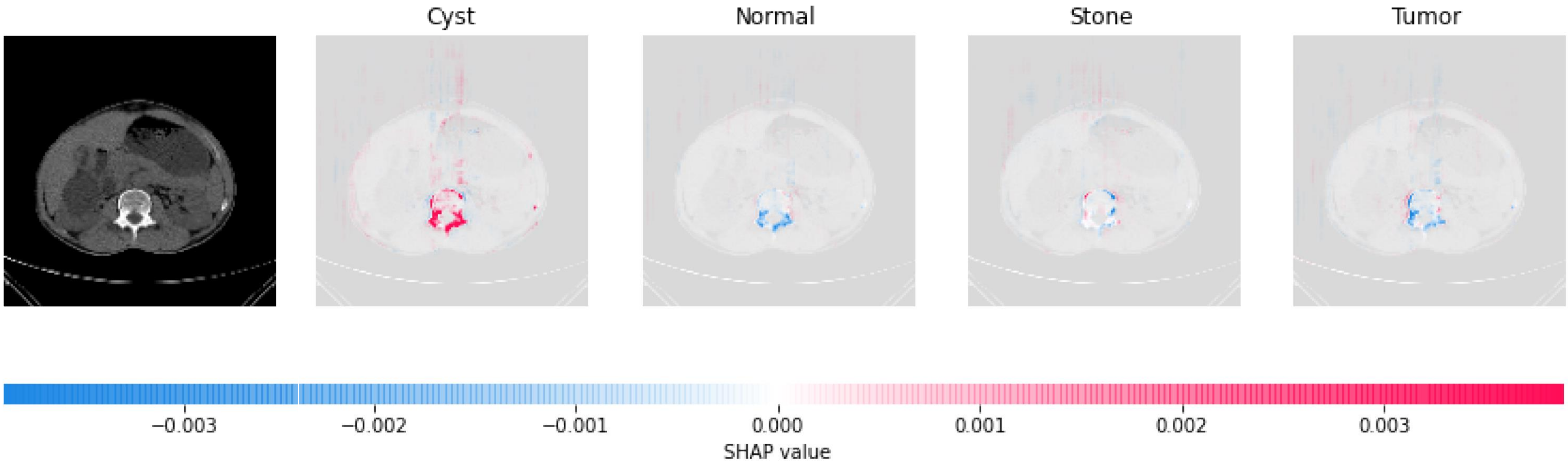
Does not improve the accuracy of AI algorithms

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_5 (Conv2D)           (None, 148, 148, 64)      1792

 max_pooling2d_5 (MaxPooling  (None, 74, 74, 64)       0
 2D)

 conv2d_6 (Conv2D)           (None, 72, 72, 64)        36928

 max_pooling2d_6 (MaxPooling  (None, 36, 36, 64)       0
 2D)

 conv2d_7 (Conv2D)           (None, 34, 34, 64)        36928

 max_pooling2d_7 (MaxPooling  (None, 17, 17, 64)       0
 2D)

 conv2d_8 (Conv2D)           (None, 15, 15, 64)        36928

 max_pooling2d_8 (MaxPooling  (None, 7, 7, 64)         0
 2D)

 conv2d_9 (Conv2D)           (None, 5, 5, 64)          36928

 max_pooling2d_9 (MaxPooling  (None, 2, 2, 64)         0
 2D)

 flatten_1 (Flatten)         (None, 256)               0

 dropout_1 (Dropout)         (None, 256)               0

 dense_2 (Dense)             (None, 128)               32896

 dense_3 (Dense)             (None, 4)                 516

=================================================================
Total params: 182,916
Trainable params: 182,916
Non-trainable params: 0
_________________________________________________________________
```

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Avg ($\mu \pm \sigma$)
TrA	99.33	99.64	99.04	99.31	99.31	99.22	99.64	99.5	99.43	99.24	99.30 ± 0.18
TrL	0.0321	0.0192	0.0445	0.0348	0.0352	0.0350	0.0223	0.2680	0.0335	0.0323	0.0557 ± 0.07
VaA	99.76	99.68	96.47	100	99.76	99.76	99.84	99.68	99.76	99.2	99.39 ± 0.99
VaL	0.0208	0.0192	0.1518	0.0118	0.169	0.0195	0.0142	0.0201	0.0195	0.0452	0.0491 ± 0.06
TsA	99.84	99.76	97.19	100	100	100	100	99.68	99.84	98.88	99.52 ± 0.84
TsL	0.0221	0.0164	0.1106	0.0018	0.0145	0.0131	0.0117	0.0208	0.0178	0.0621	0.0291 ± 0.03



```
pred_single=model.predict(image[0])
pred_single=np.argmax(pred_single, axis=-1)

0.9939, 0.03, 0.001, 0.0051
{Cyst, Normal, Stone, Tumor}
```



Red Pixels increase the classification probability

Blue Pixels Decrease the classification probability

XAI Frameworks

Local Interpretable Model-Agnostic Explanations (LIME)

SHapley Additive exPlanations (SHAP)

Count ++

Local Explanations

focus on interpreting individual predictions

Heatmaps (GradCAM, GradCAM++)

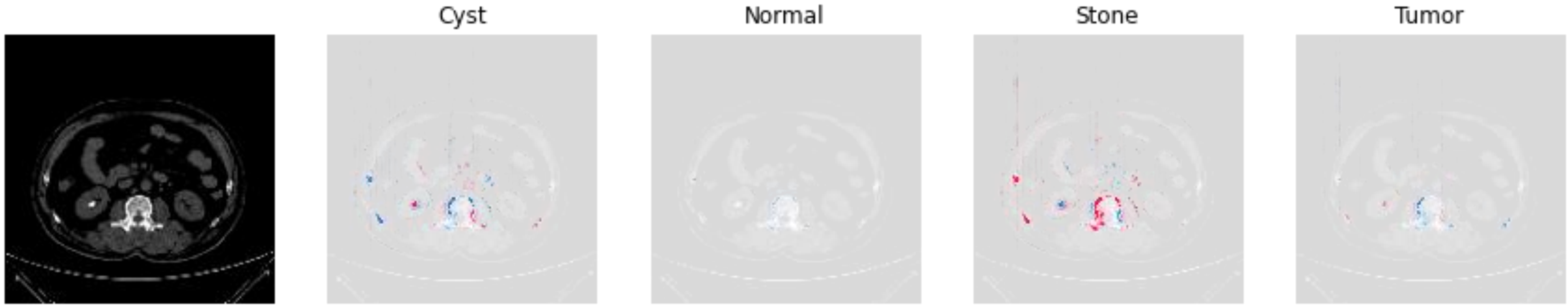
Accumulated Local Effects (ALE)

Count ++

Global Explanations

Aim to provide insights into the overall behavior of the model across different inputs.

Why XAI is important in Medical Images?



Improving Diagnosis and Treatment

Trust and Transparency

Bias Detection and Mitigation

Patient Communication

https://www.mdpi.com/2076-3417/13/5/3125

https://colab.research.google.com/drive/1uCIlNC9aqiAg1LR1ibUykEbcONYzFtoI?authuser=1#scrollTo=qMUyoBkpsbIX https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.822666/full

https://colab.research.google.com/drive/1YKy8WrfLvSR7dn4MSkEGrg4TVq7ODbAh?authuser=1

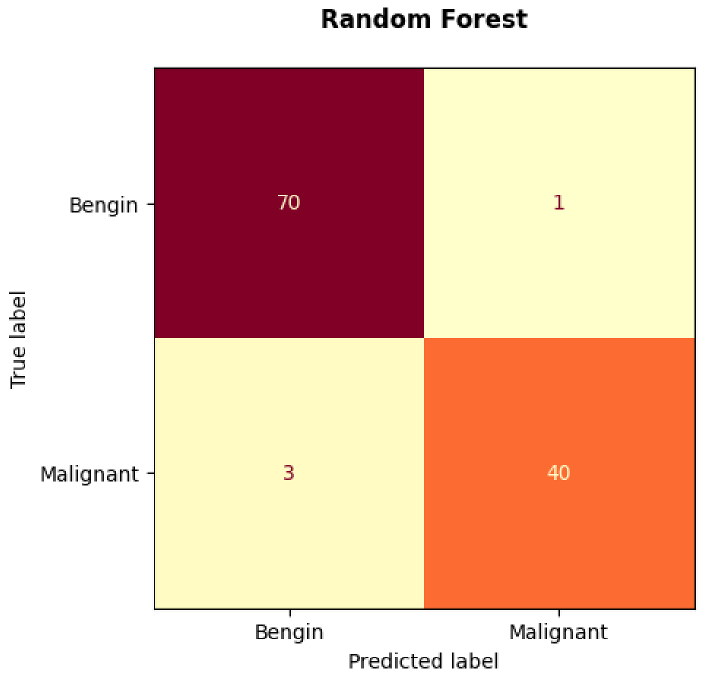
Are XAI only for Image Data?

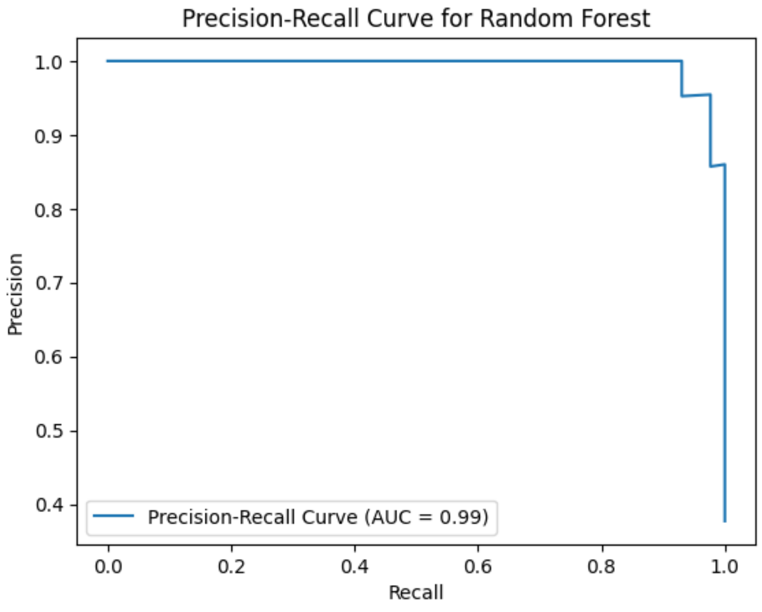
NO

XAI on CSV Data

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...

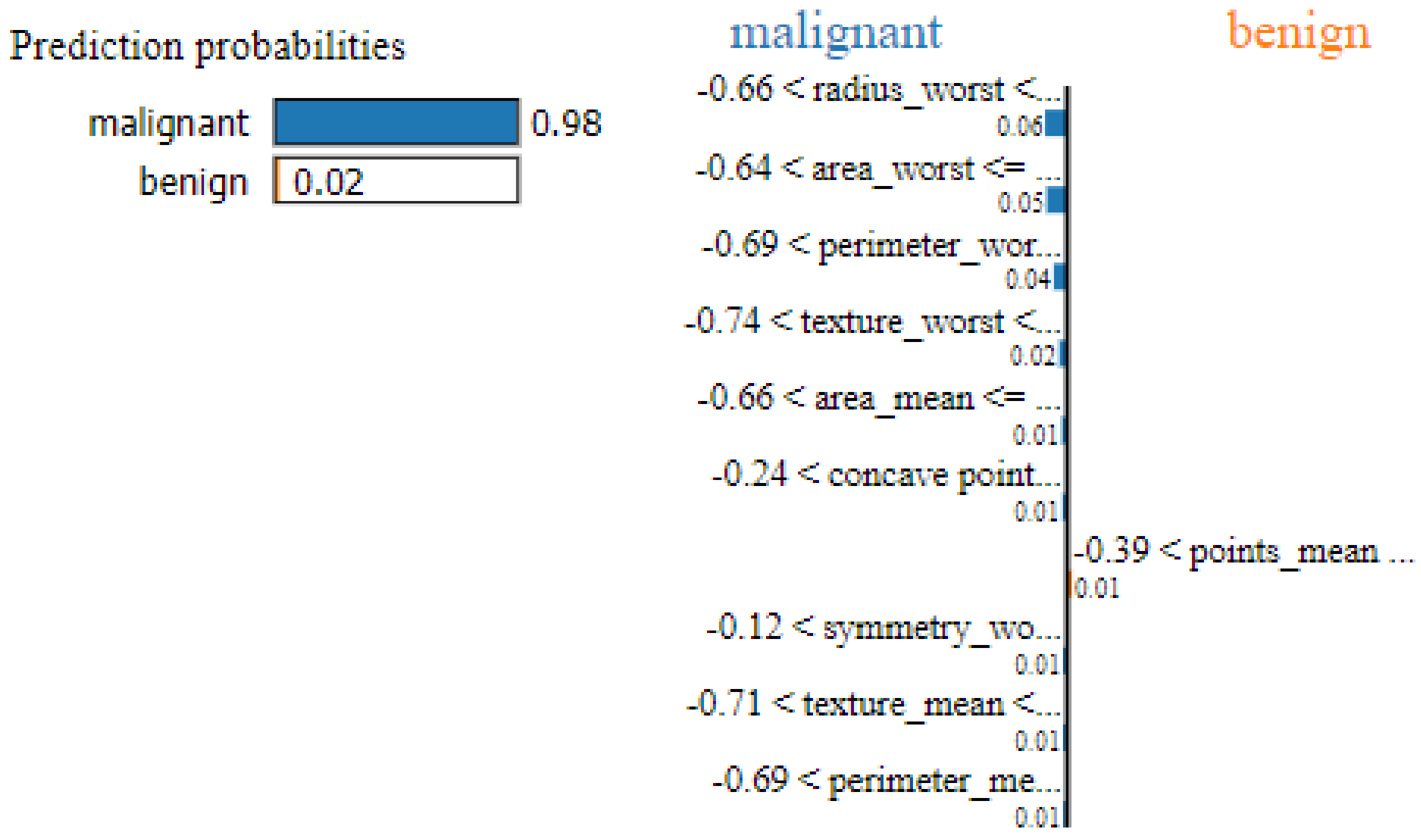
5 rows × 32 columns



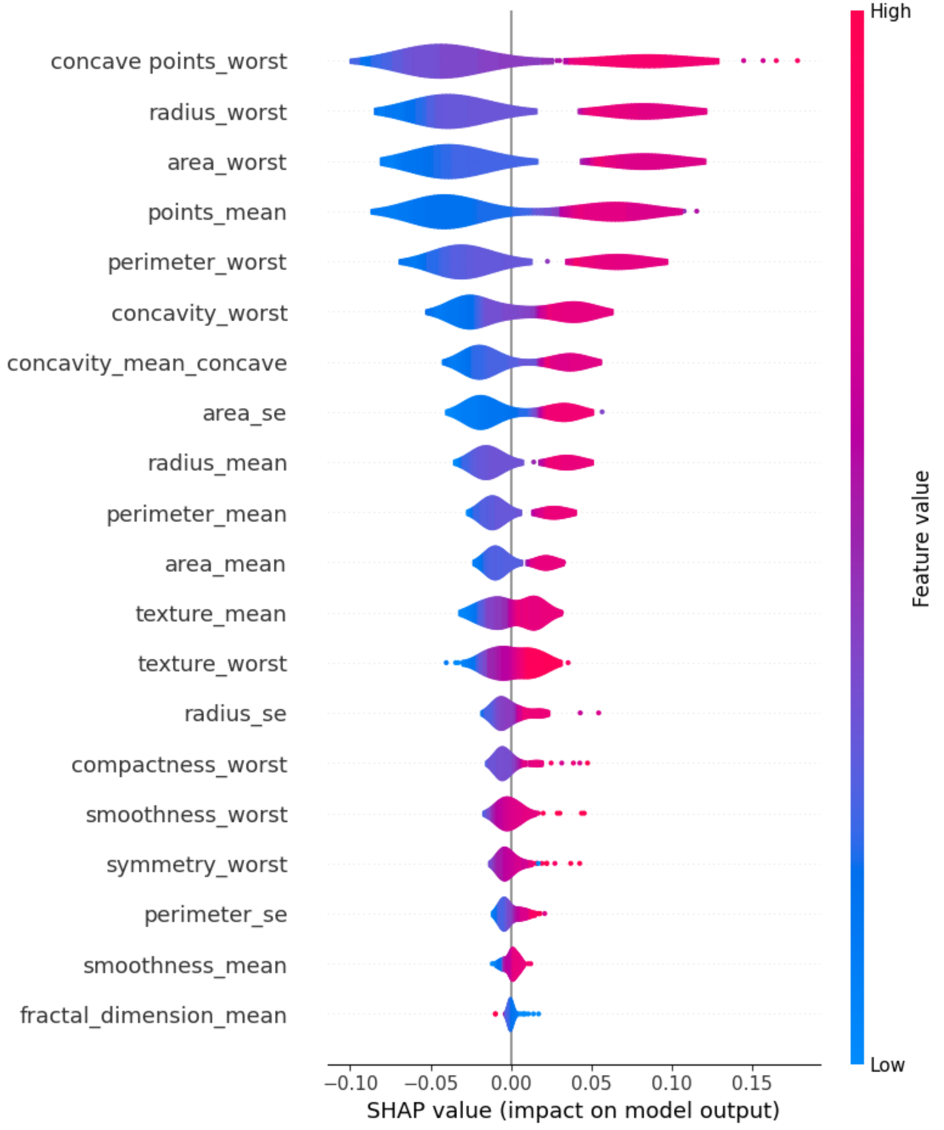


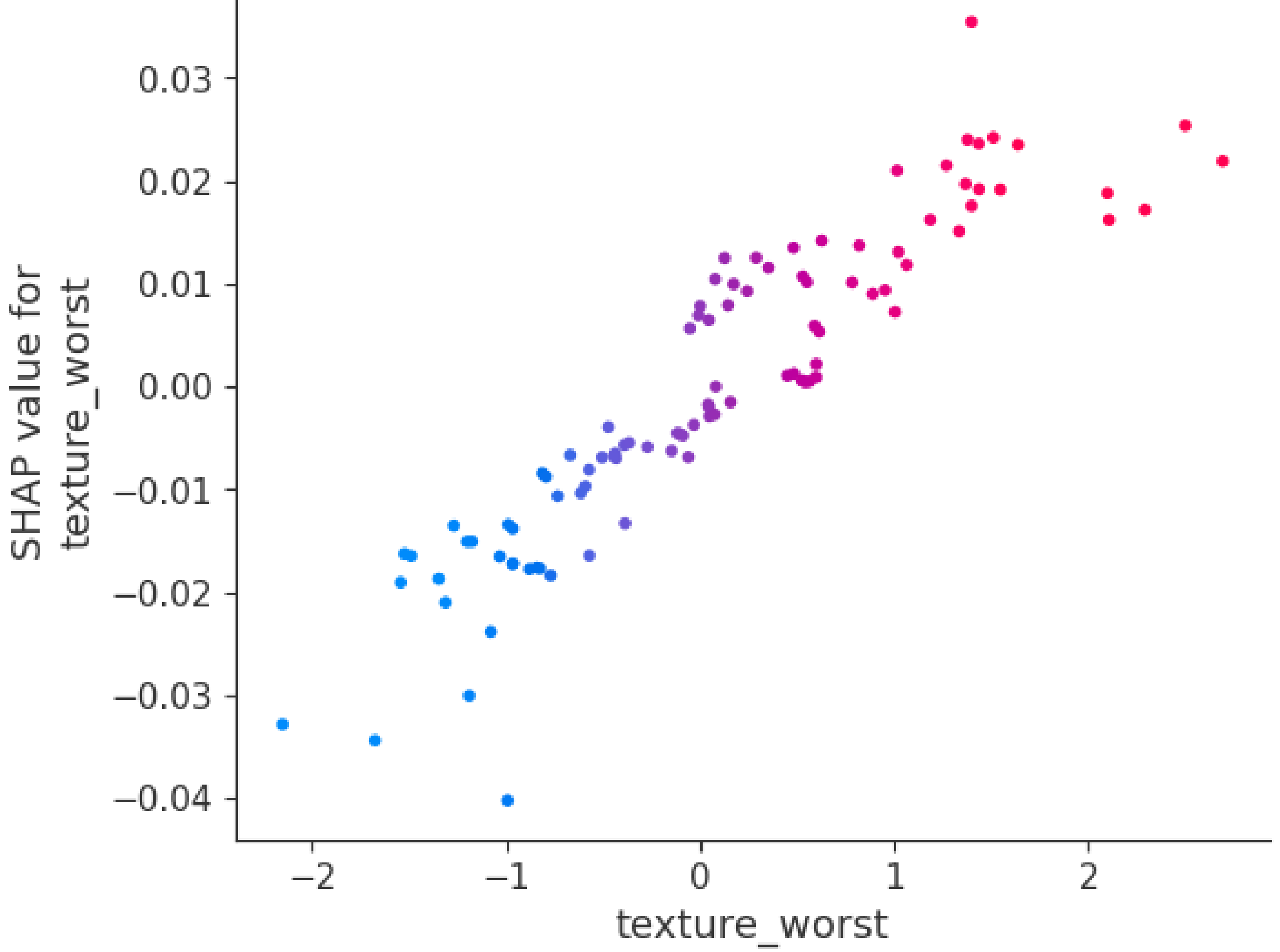
XAI on CSV Data (LIME)

```
exp = explainer.explain_instance(
    data_row=X_test_scaled[0],
    predict_fn=modelRF.predict_proba)
```



XAI on CSV Data (SHAP)





Limitations of XAI algorithms

No Any Standard

Lack of Interpretability for Complex Data Types (For Eg: Multidimension)

Trade-offs with Model Accuracy

Discussion