A Brief Overview of Explainable and Interpretable AI

William Franz Lamberti [1]

University of Virginia

SDSS 2022

[1]MS Statistical Science
PhD Computational Sciences and Informatics

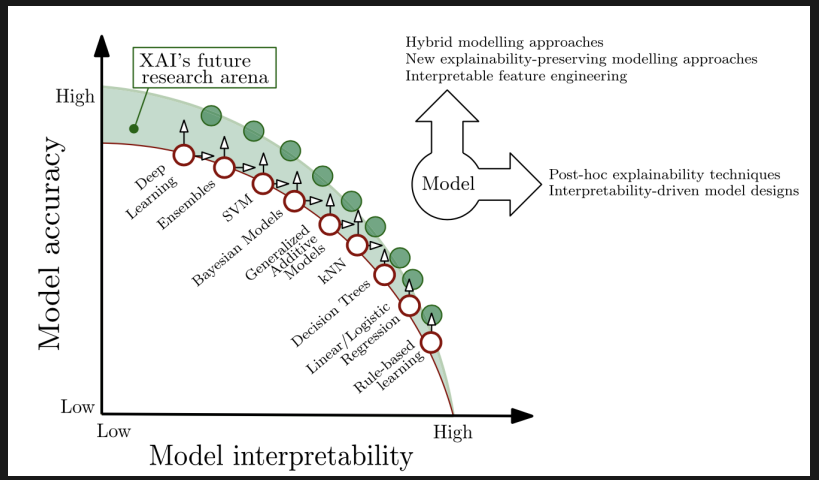
© William Franz Lamberti

Outline

© William Franz Lamberti

Introduction

- Explainable AI (XAI) includes methodologies, statistics, and/or variables that provide insight into how models make predictions
- XAI is a newer idea that depends on interpretability and explainability
- Different definitions, but similar concepts
- Mostly confined to models
- EU and the "Right to Explainability"



Barredo Arrieta et al., 2020

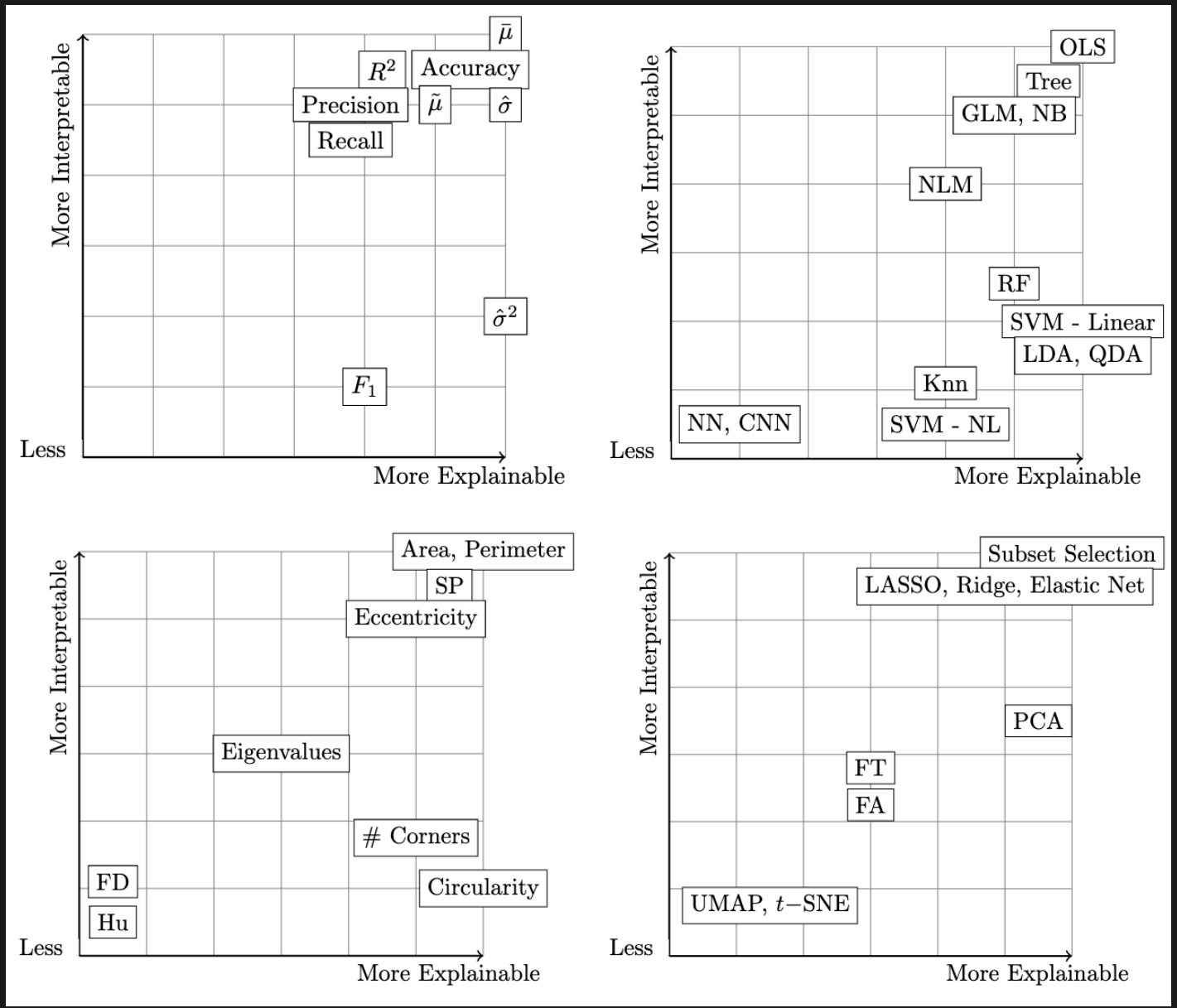
Explainability and Interpretability

- Interpretability = the characteristic of an element to have concrete physical meaning
- Explainability = the property of an element that allows its mechanisms to be explicitly described, understood, and studied

	Model	Variable	Example
Inter.	For every unit increase in X, we expect Y to increase by Δ	X measures the area of a shape	Standard deviation (unit)
Expl.	The model tends to correctly classify class γ with low values as shown by this plot	X measures how circular the shape is	Variance (unit^2)

© William Franz Lamberti

Explainability and Interpretability: Overview

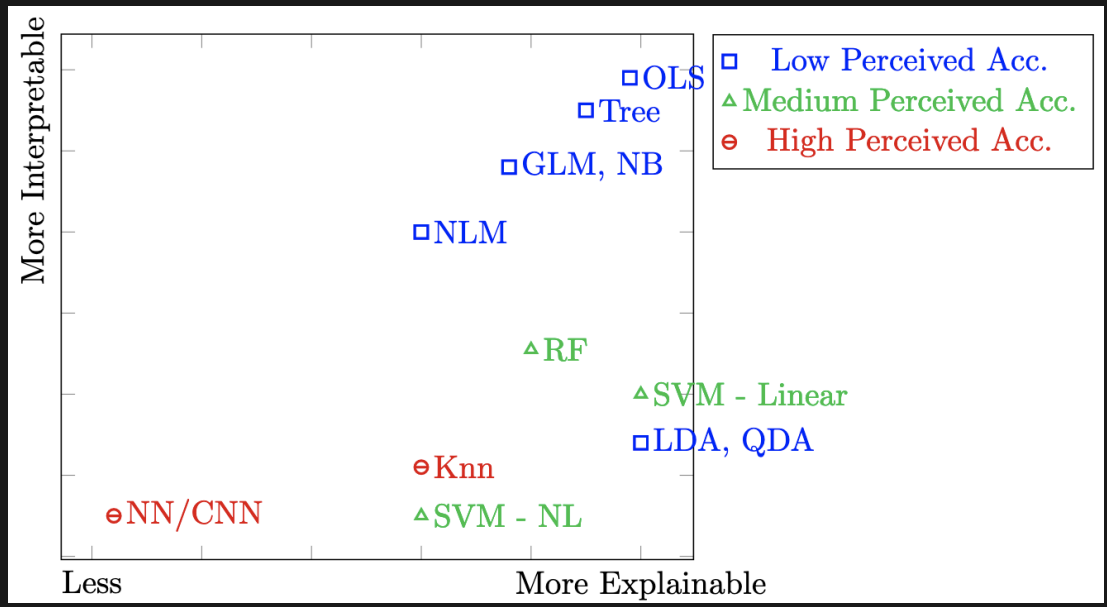


Lamberti, In Press, 2022

© William Franz Lamberti

Explainability and Interpretability: Perceived Accuracy

- More Inter. = less perceived Acc.
- Less Inter. = More perceived Acc.
- Flawed perspective
- Better perspective
  - Understand the problem
  - Occam's Razor

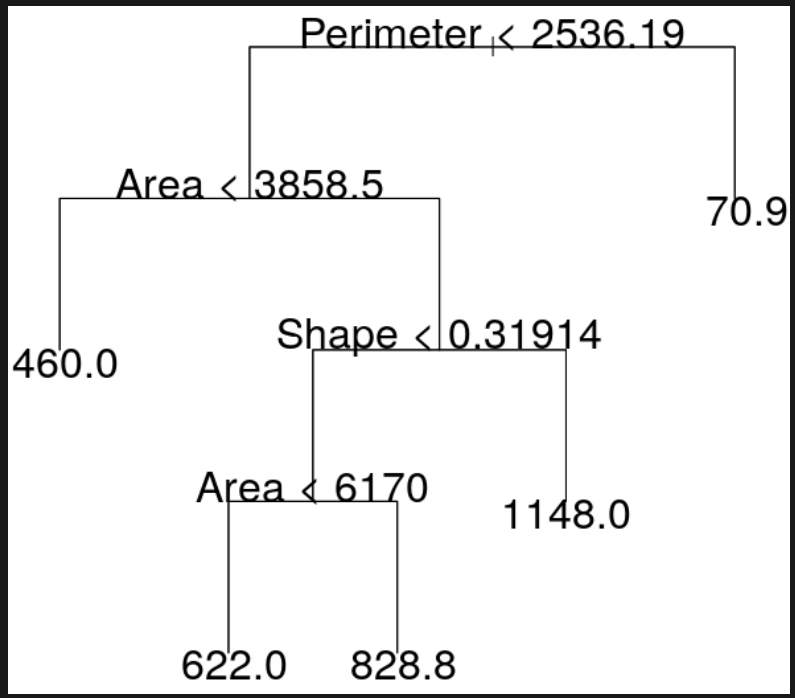


Lamberti, In Press, 2022

"Wherever possible, it makes sense to try the simpler models as well, and then make a choice based on the performance/complexity tradeoff." - ISLR, 2nd Ed.

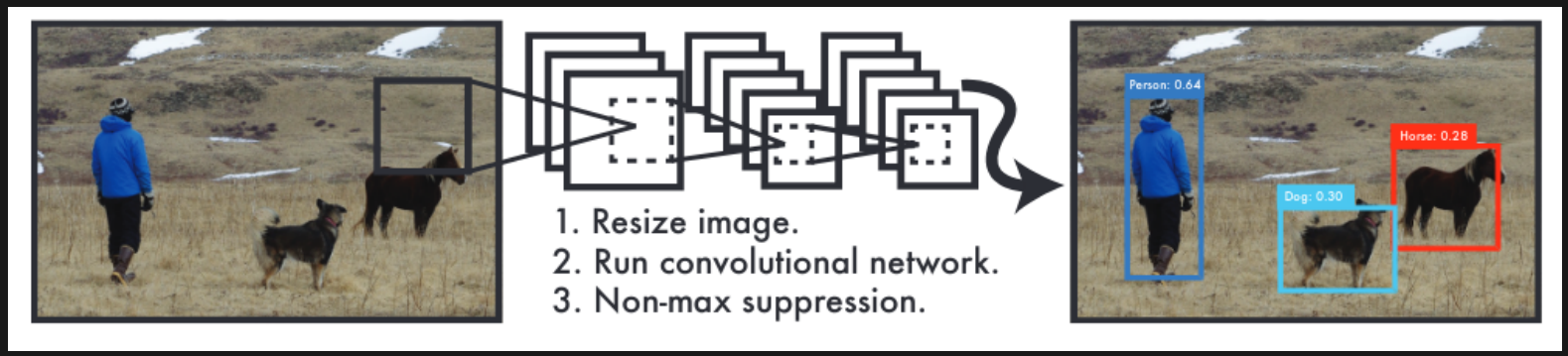
Explainability and Interpretability: Trees

- Pros
  - Easy to visualize
  - Nontechnical audiences easily follow
- Cons
  - Can easily overfit
- High Inter. and Expl.



Lamberti, In Press, 2022

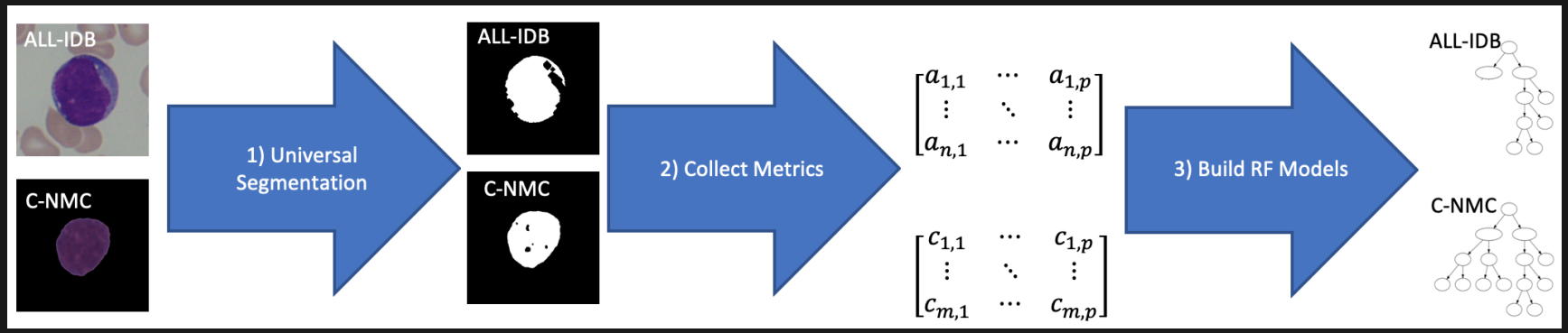
Explainability and Interpretability: CNNs



Redmon, Joseph et al., 2016

- Pros
  - Very powerful
  - Learn features needed for problem
  - Perform segmentation and classification
- Cons
  - Shift specifying features to specifying architecture
  - Costly
  - Difficult to interpret and explain
- Low Inter. and Expl.

WBC: Introduction



- Build model that outperforms state of the art in classifying WBCs as malignant (ALL) or benign (H)
- Use interpretable and explainable metrics
- Use universal segmentation algorithm
- Use as few variables as possible

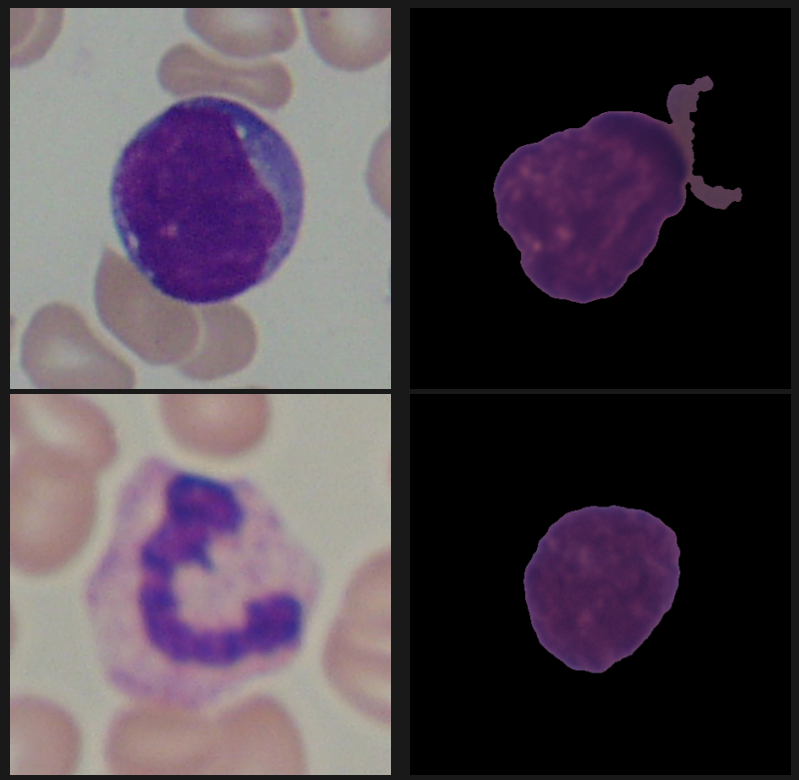
Case Study: Introduction

- Classification of white blood cells (WBCs) as malignant or benign is an important task
- Acute Lumpocytic Leukemia (ALL) is a type of malignant cancer
  - Prone populations
    - Children
    - Elderly
  - ALL Characteristics
    - Less regular in shape
    - Holes in the cytoplasm
    - Circular particles in nuclei
- State of the art approaches use some form of a Convolutional Neural Network (CNN)

© William Franz Lamberti

Case Study: Data

- ALL-IDB2 Classes
  - ALL: 130
  - H: 130
- C-NMC Classes
  - ALL: 7,272
  - H: 3,389
- Universal segmentation would be very useful



Case Study: Segmentation Results

Case Study: Modeling Results

Data	Source	# of Features	Model	Exp.	Inter.	Type	Acc./F_1
ALL-IDB2	Singhal and Singh (2014)	256	SVM	Low	Low	AI	89.72%
	Singhal and Singh (2016)	4096	Knn	Low	Low	AI	93.84%
	Bhattacharjee and Saini (2015)	**8**	Knn	Low	Low	AI	95.24%
	Sahlol, Abdeldaim, et al. (2019)	45	Knn	Low	Low	AI	95.67%
	Sahlol, Kollmannsberger, et al. (2020)	1087	CNN& SVM	Low	Low	AI	96.11%
	William Franz Lamberti (2021)	24	RF	High	Medium	XAI	**100.00%**
C-NMC	Kulhalli et al. Kulhalli et al. (2019)	25×10^6	CNN	Low	Low	AI	85.7
	Ding et al. Ding et al. (2019)	87×10^6	CNN	Low	Low	AI	86.7
	Marzahl et al. Marzahl et al. (2019)	11×10^6	CNN	Low	Low	AI	86.9
	Sahlol et al. Sahlol, Kollmannsberger, et al. (2020)	1,115	CNN&SVM	Low	Low	AI	87.9
	William Franz Lamberti (2021)	**24**	RF	High	Medium	XAI	**90.1**

Conclusion

- Inter. and Expl.
  - Definitions used to describe various aspects of modeling
  - Future Work: Formalized index or metric
- Case Study
  - RF outperforms all other approaches
  - Rethink our conceptions about model performance
- An Overview of Explainable and Interpretable Artificial Intelligence, Lamberti, W. F. In "AI Assurance: Towards Valid, Explainable, Fair, and Ethical AI" (To be Published Fall 2022)

Acknowledgements

- GMU, Office of the Provost
- GMU, Office of Research Computing
- UVA, Center of Public Heath Genomics
- SDSS, Early Career Award

Any Questions?