

Explainable AI in Healthcare

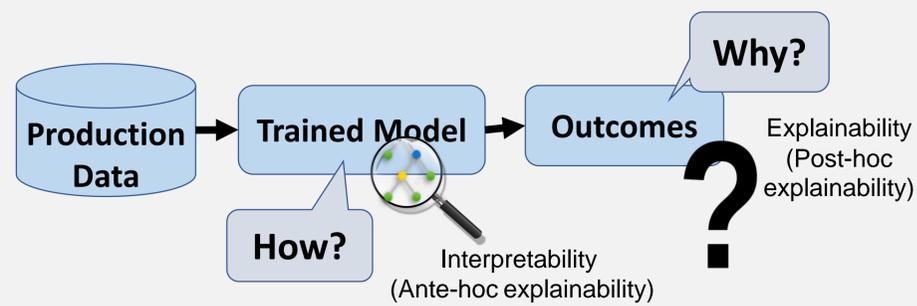
Urja Pawar, Donna O'Shea, Christian Beder, and Ruairi O'Reilly
Munster Technological University (MTU)

urja.pawar@mycit.ie, donna.oshea@cit.ie, christian.beder@cit.ie, ruairi.oreilly@cit.ie

What is eXplainable AI (XAI) ? - Techniques to interpret and explain AI-based systems

Why XAI?

- Understanding why an outcome was generated
- Explaining operation of black-box models
- Understanding the underlying training data

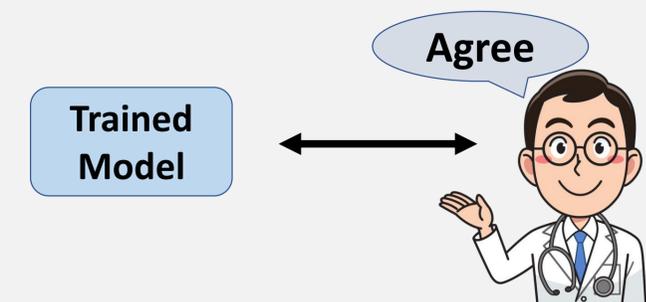


XAI Techniques

Ante-hoc XAI	Post-hoc XAI
Explainability is achieved without any additional operation after training phase	Explainability is achieved with additional operations after training phase
Transparent models are intrinsically interpretable and come under ante-hoc XAI models	Post-hoc XAI can be applied to deep learning models that have complicated internal operation
Examples: Linear/Logistic regression, decision trees	Examples: Local Agnostic Model Explanations (LIME), Shapley values

Why XAI is important in healthcare?

Clinical Validation



Understanding of AI-based systems can enable a validation from medical experts regarding a model's operation

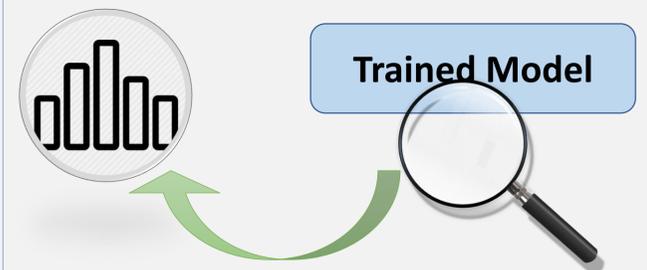
Enhanced Improvements

Debugging of AI-based systems can be enabled via explanations regarding inaccurate outcomes



The explanations can also enable identification of limitations in the medical dataset used for training

Knowledge Discovery

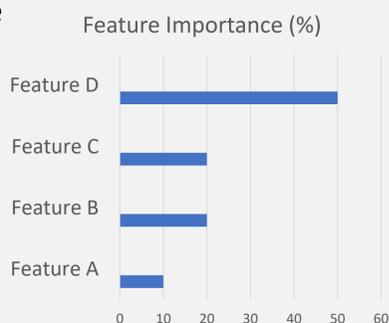


Discovering interesting insights in the healthcare domain with the knowledge of what a model learned

About My Research

Feature Importance

Feature importance techniques assign scores to participating features based on their impact on model's outcome



In 28th Irish Conference on Artificial Intelligence and Cognitive Science, 2020, a work was published to demonstrate how feature importance can enable explainability in the diagnosis of cervical cancer

Medical workflows in hierarchical settings

- The healthcare system is hierarchical – small clinics for general check ups and hospitals for more advanced examinations
- AI-based systems should consider the hierarchical structure and the dynamic nature of medical workflows
- Current work focuses on enabling explainability in diagnostic workflows across the hierarchical medical settings in terms of performance gains and feature importance scores assigned to each diagnosis

