Explainable AI for Digital Histopathology

University of Southampton | The Alan Turing Institute

Contact: J.A.Early@soton.ac.uk @JosephAEarly

Joseph Early, Christine Evers, Gopal Ramchurn

Motivation

Processing high resolution digital histopathology images with AI is difficult:

- **Large images**: The images are so large that conventional machine learning cannot be used.
- **Lack of labels**: Labelling these images is very expensive – annotating individual cells is time consuming and requires a trained clinician.
- **Detail required**: If the images are down-sampled, their important details are lost, so accurate predictions cannot be made.

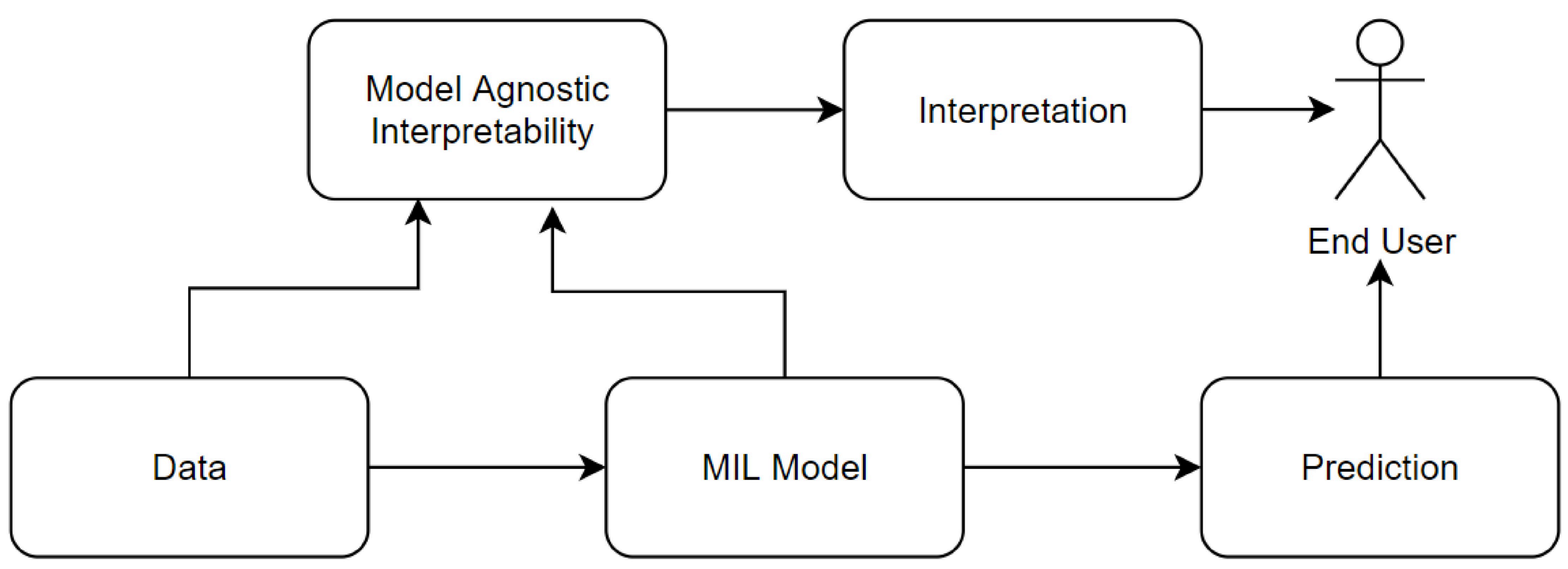
As such, we have to break the whole slide images into patches and use a technique known as **Multiple Instance Learning (MIL)**.

Our work investigates:

- **How MIL models make their decisions**. Which patches are the important ones and what outcomes do they support?
- **The role of explainable AI in digital histopathology**. Can we understand how a classifier makes decisions?

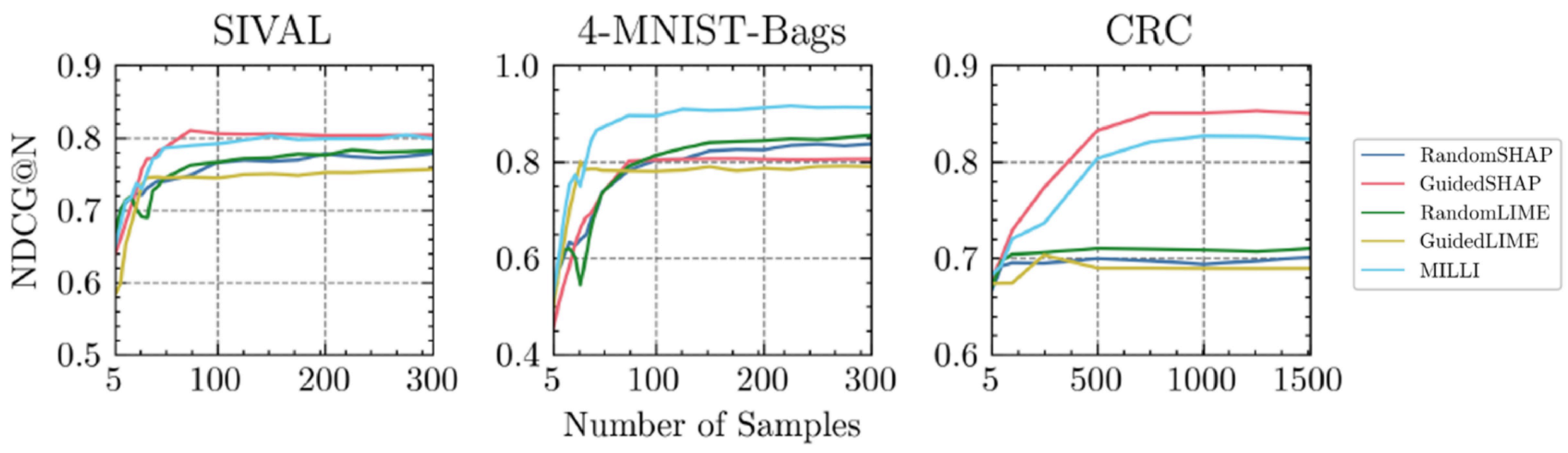
Explainable AI for MIL

Explainable AI allows both **technical and non-technical users** to better understand how an AI system makes decisions, which helps **facilitate trust**.

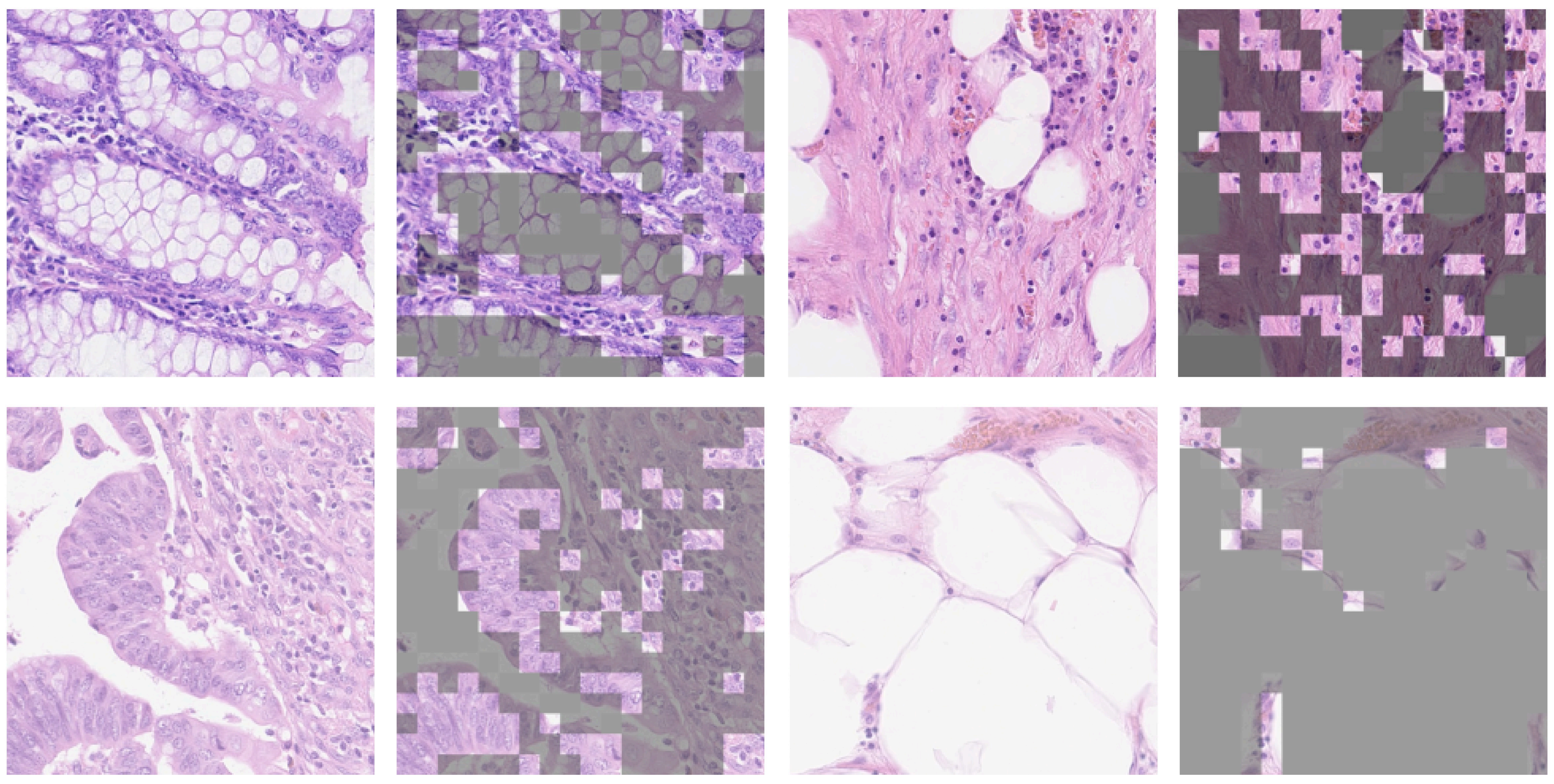


Our approach, **Multiple Instance Learning Local Interpretations (MILLI)**, determines the importance of instances by analysing the effect of removing patches from the original image:

- **Patch contribution**: Different patches contain different data and will contribute to different outcomes. By **removing patches and observing any changes in prediction**, we can determine which outcomes they contribute to.
- **Context aware**: The patches cannot be considered in isolation, therefore we **sample collections of patches to be removed together**.
- **Optimisation**: It is expensive to determine the change in prediction for every possible combination of patches, therefore we **optimise our sampling** by first identifying the patches that are discriminatory and then biasing our sampling to choose more informative collections of patches.
- **Improved efficiency**: On average, compared to existing interpretability methods, MILLI is both **more accurate and more efficient**. It requires fewer samples to determine the contribution of patches, meaning it can generate accurate interpretations in a shorter amount of time.



Explainable Colon Cancer Detection



We applied our methods to classifying tissue in colorectal cancer:

- **Cell identification**: Colorectal cancer originates from epithelial cells. Our approach means these cells can be found **without requiring manual annotation, saving money and time**. Furthermore, our approach also annotates other cell types as part of its explanations.
- **Better performance**: MILLI out-performed existing methods. Our interpretations were **30% more accurate** on average than existing approaches such as interpretable models (we found that these had poor general performance), or post-hoc methods such as LIME and SHAP.

University of Southampton | The Alan Turing Institute

Joseph Early, Christine Evers, and Gopal Ramchurn. "Model Agnostic Interpretability for Multiple Instance Learning." *International Conference on Learning Representations* (2022).