Explainable AI for High Resolution Images

JOSEPH EARLY, CHRISTINE EVERS, and GOPAL RAMCHURN

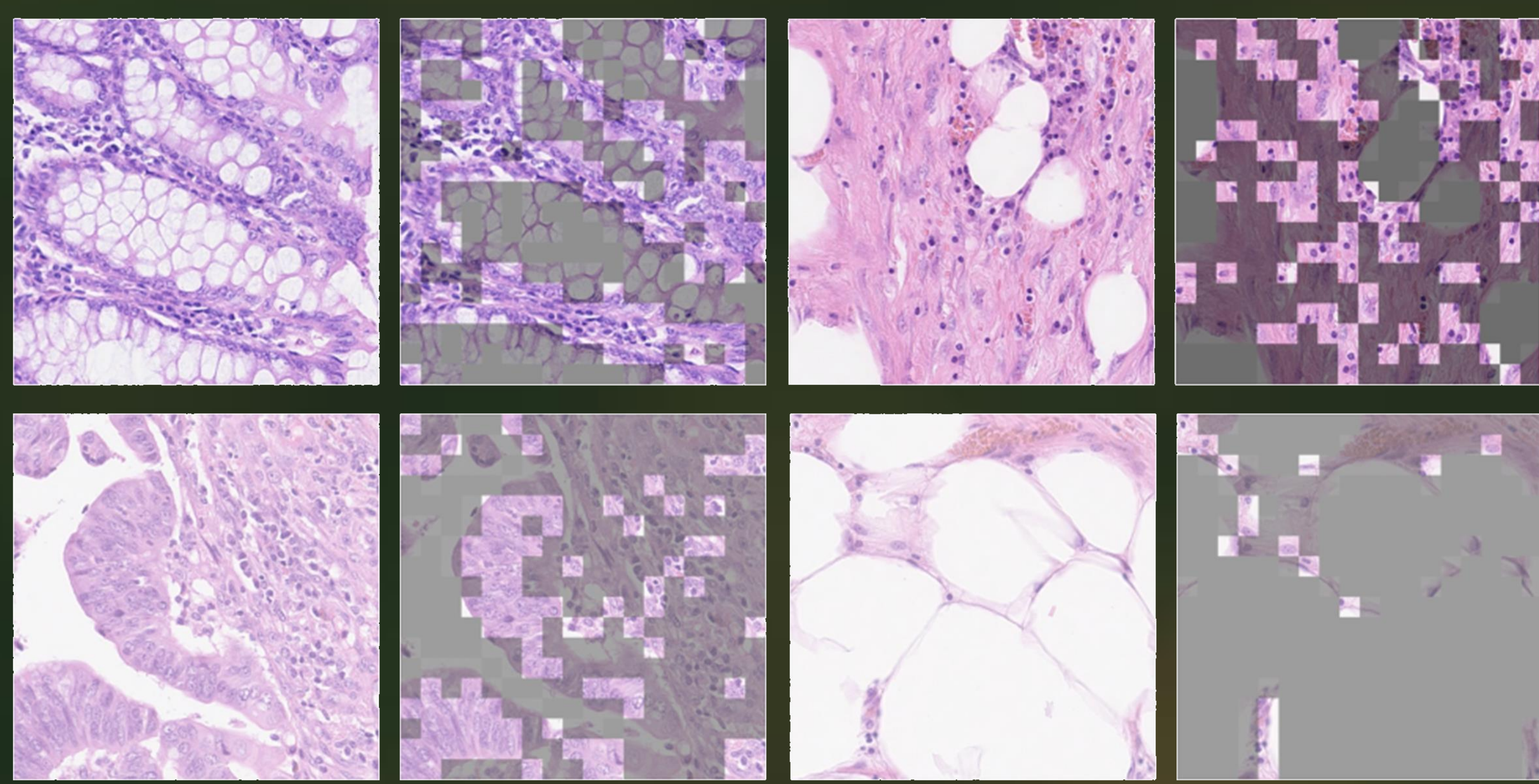
Agents, Interaction, and Complexity Group, University of Southampton, UK
J.A.Early@soton.ac.uk @JosephAEarly

INTRODUCTION

- Conventional machine learning cannot process high resolution images.
- To work with **gigapixel images**, machine learning models would need to be incredibly large, such that they **could not be run, even on supercomputers**.
- Reducing the size of these images leads to important details being lost.
- **AIM:** Establish methods than can not only work with high resolution images, but also **provide explanations** about decision making.

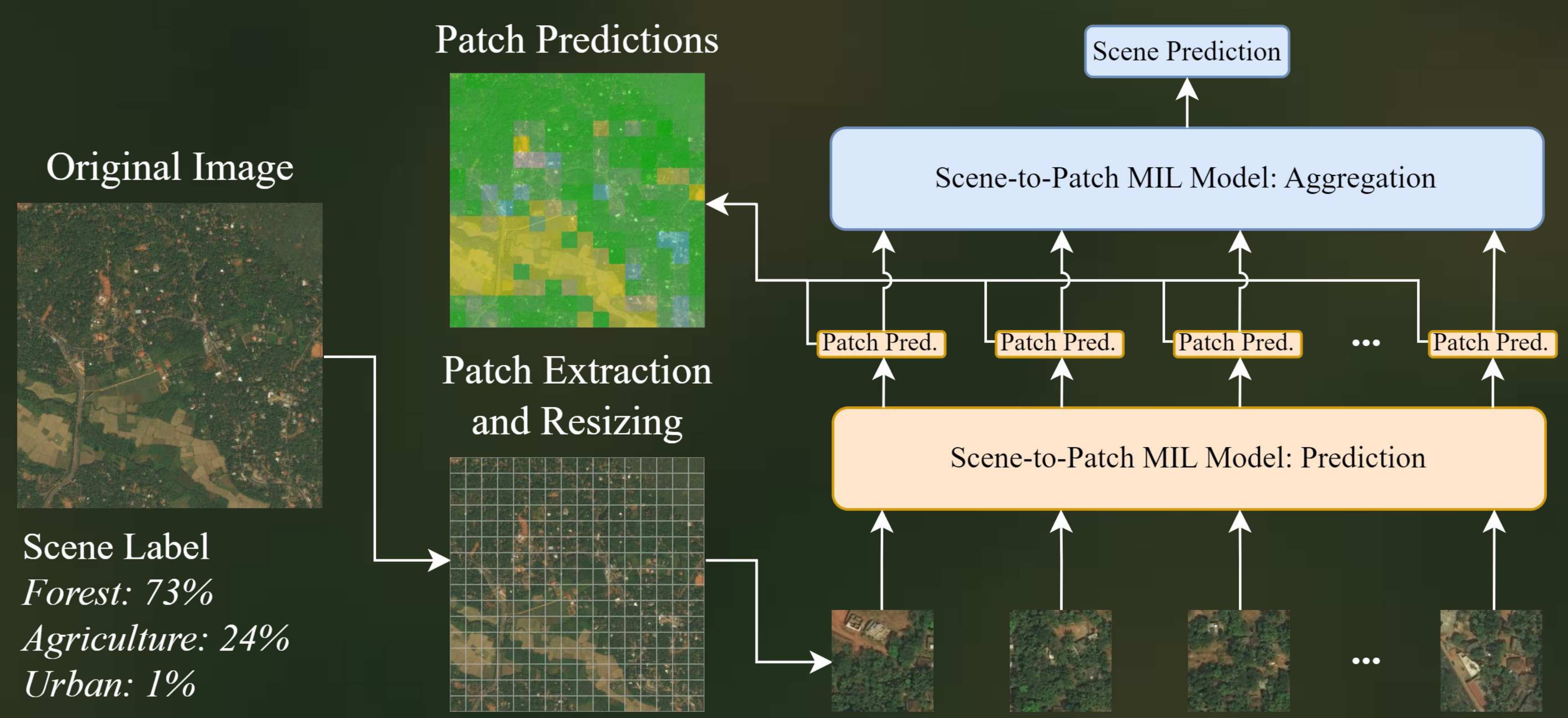
MOTIVATION

- High resolution images come from a variety of sources, for example **satellite data** and **medical imaging**.
- Labelling these images is **very expensive** – it can take a large amount of time and requires trained specialists.
- These images are often used in domains were **trust is crucial**, for example **identifying cancer** in digital histopathology scans.

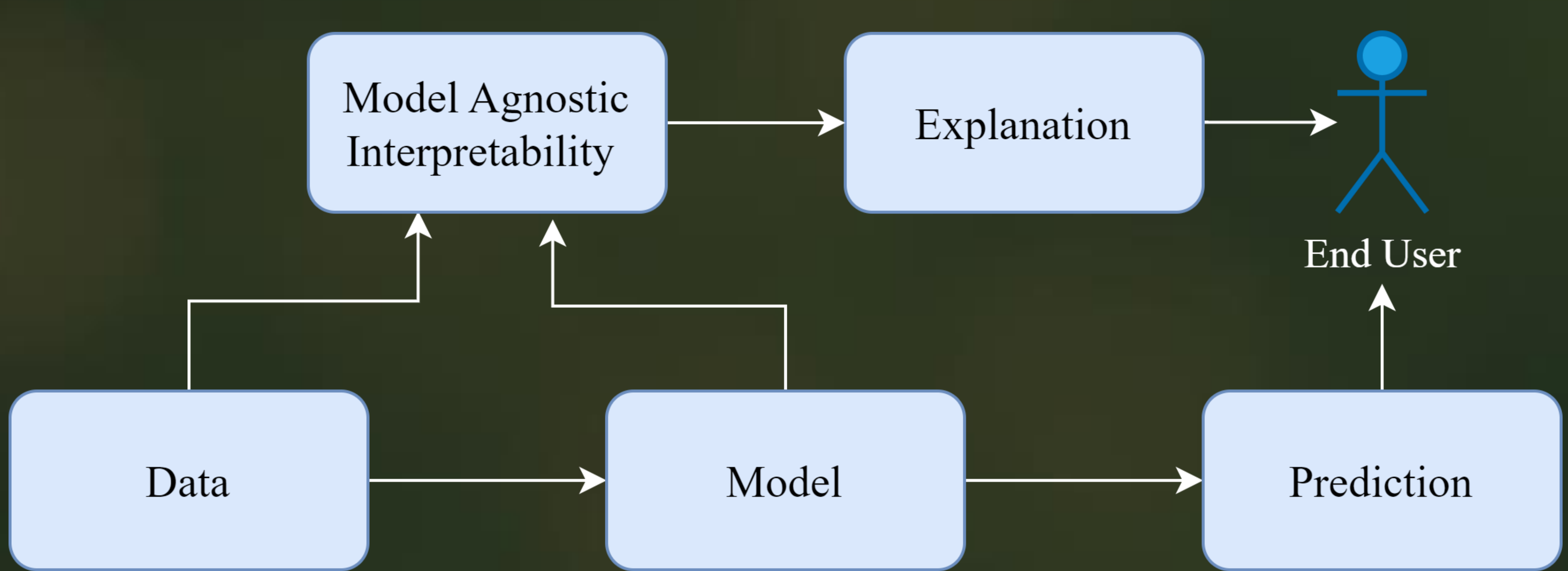


METHODS

- To process high resolution images, we can use a machine learning technique known as **multiple instance learning**.
- This approach breaks the image (scene) into **patches**, which are small enough to processed, but are not labelled.
- The learnt representations for each patch are then **aggregated**, using methods such as **attention**.
- These models are able to learn from just scene-level labels. This **expedites the labelling process**, as patch-level or segmentation labels are not required.



- Certain models are **inherently interpretable** – they create explanations as part of their processing.
- For example, some architectures produce **patch-level predictions** as well as scene-level predictions.
- Other models require **post-hoc explanation**.
- This can be achieved with a **model-agnostic** approach, where the type of underlying model does not matter.



RESULTS

- Our post-hoc model-agnostic explainability method, **Multiple Instance Learning Local Interpretations (MILLI)**, was found to be effective on a wide range of datasets [1].
- It **outperformed existing methods** such as LIME and SHAP, and was also more efficient.
- We used MILLI to produce explanations for **colorectal cancer detection**. It was able to correctly identify epithelial cells (where cancer originates), with up to a **30% improvement in accuracy** over existing methods.
- For satellite images, we developed **interpretable Scene-to-Patch models**, which segment land into different categories – an important use of **AI for climate change mitigation** [2].

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

[2] Joseph Early, Ying-Jung Deweese, Christine Evers, and Sarvapali Ramchurn. "Scene-to-Patch Earth Observation: Multiple Instance Learning for Land Cover Classification." *Tackling Climate Change with Machine Learning: Workshop at NeurIPS (2022).*

University of Southampton
The Alan Turing Institute
UKRI Trustworthy Autonomous Systems Hub