

A Practical Application of Black Box Interpretability



Byungwoo Jang, Mingu Jeong, Hevin Na, Conner Russell (US Military Academy),
Dave Braines, Richard Tomsett, Graham White (IBM UK), Dan Harborne (Cardiff University)

Purpose

Further development into the connection between machine learning and visual recognition platforms.

Additionally, the development of the communication between machine and humans through interpretability of black box events:

- Summarisation
- Explainability (post-hoc)

Tackling Cancer Application

Explaining Black Box Decisions

In regards to the two components of interpretability of a black box event:

- Summarisation: Information directly displayed to the user.
- Explainability: Justification of the results

Practical Application

Taking a series of biopsy images, we used visual recognition and machine learning to attempt to identify and classify cancerous cells.

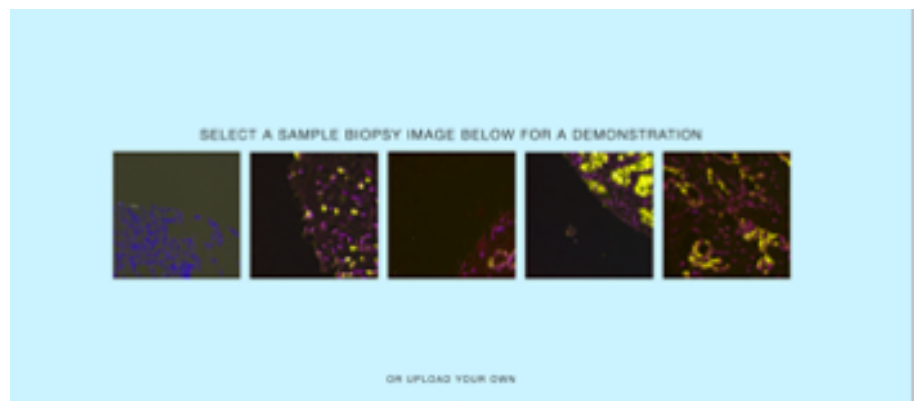


Figure 1: Demonstration of visual recognition on biopsy images.

The end state: biopsy images are put into the program, visual recognition with machine learning is applied, and an output is given. The

output result would include a confidence percent cancerous, and why it thinks a cell is cancerous.



Figure 2: Baseline for what results will appear when the program is run.

Military Relevance

Counter IED Operations

Robots and technology used today in counter IED operations are still operated by a human based off of images from equipment such as radars or basic cameras. This application would mirror the process of identifying cancerous cells, except in this case it would identify and classify IEDs.



Figure 3: TALON Tracked Military Robot locates potential ordnance.