Interactive Machine Learning for Image Data Science

Alexandre Xavier Falcão

Laboratory of Image Data Science - Institute of Computing - University of Campinas

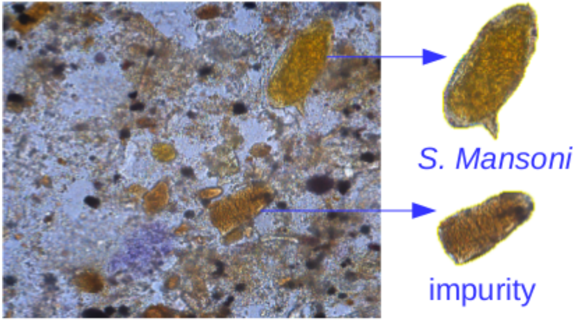
afalcao@ic.unicamp.br
lids.ic.unicamp.br

Agenda

- Why should experts control the machine training process?
- Strategy and research goals.
- Methodology: a case of success that involves Parasitology, Chemistry, Mechanical Engineering, Computer Science, and the Industry.
- List of other under-development applications by the same methodology.

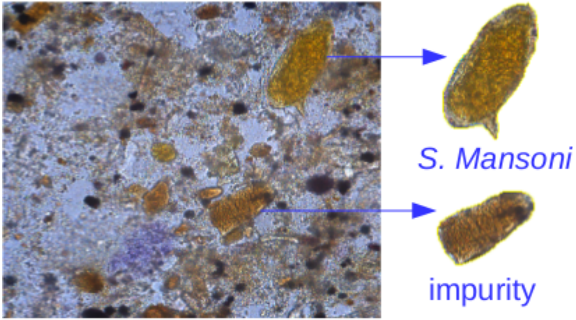
Why should experts control the machine training process?

Machines can be trained to classify images from pre-annotated examples related to a given problem.



Why should experts control the machine training process?

Machines can be trained to classify images from pre-annotated examples related to a given problem.



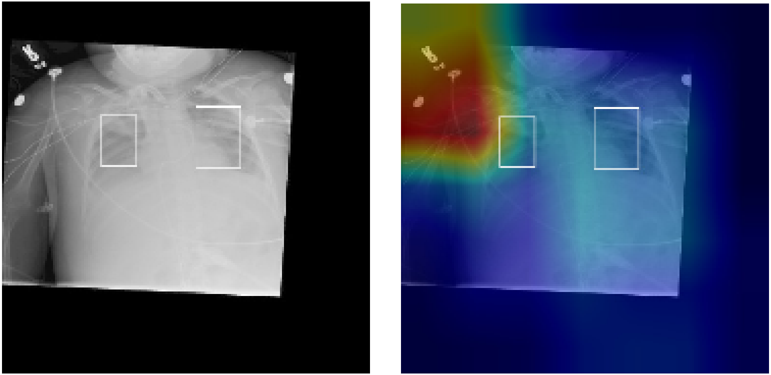
This usually requires considerable human effort in image annotation and leaves several unanswered questions.

Why should experts control the machine training process?

- How many pre-annotated examples are needed and which are the most effective ones?
- What can humans and machines learn from each other?
- Can the machine classify images from other sources with no need for retraining?
- Can the experts explain the decisions of the machine?

Why should experts control the machine training process?

For instance, a result of a neural network trained to classify normal and pneumonia images.



The decision (pneumonia) was correct, but it was clearly based on image characteristics unrelated to the problem.

Strategy and research goals

Strategy: exploit the complementary skills of humans in knowledge abstraction and machines in data processing.

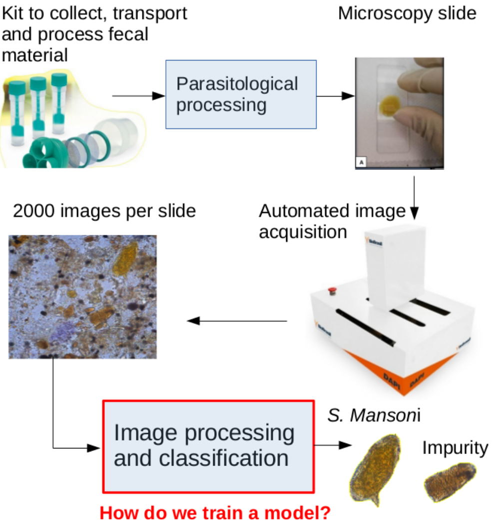
Strategy and research goals

Strategy: exploit the complementary skills of humans in knowledge abstraction and machines in data processing.

Goals: build reliable image-based decision-making systems such that

- the effort of the expert is minimized,
- the expert can understand and control the process, and
- the decisions of the machine are explainable by experts.

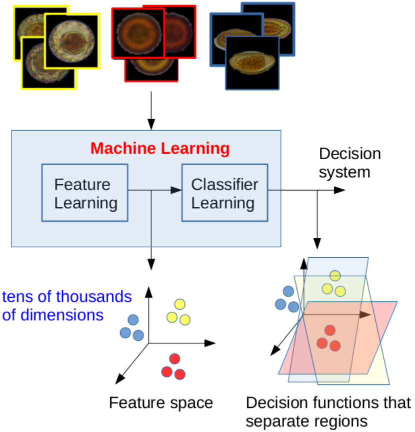
Methodology: a case of success



Automated diagnosis of human intestinal parasites for the 15 most common species in Brazil.

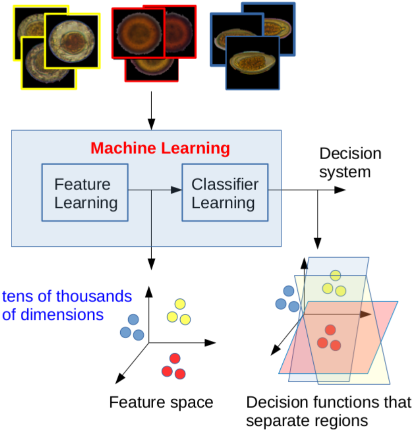
Methodology: what is the key problem?

Examples from distinct classes must be mapped into separated groups of points in some multidimensional feature space.



Methodology: what is the key problem?

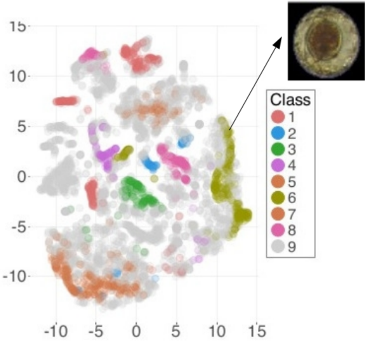
Examples from distinct classes must be mapped into separated groups of points in some multidimensional feature space.



How can the experts understand and intervene in this process?

Methodology: what is the main goal?

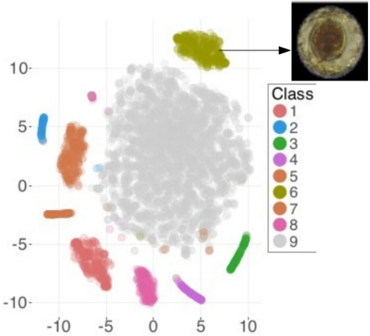
By non-linear projections, an expert can verify the success of feature learning and decide when to end the training process.



Feature projection before training from eight species of helminth eggs and impurities.

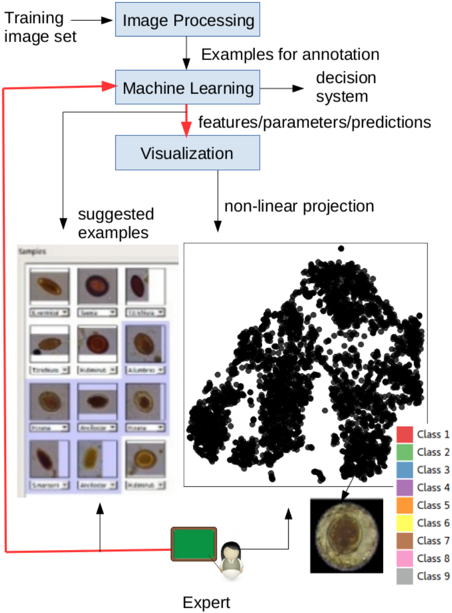
Methodology: what is the main goal?

By non-linear projections, an expert can verify the success of feature learning and decide when to end the training process.

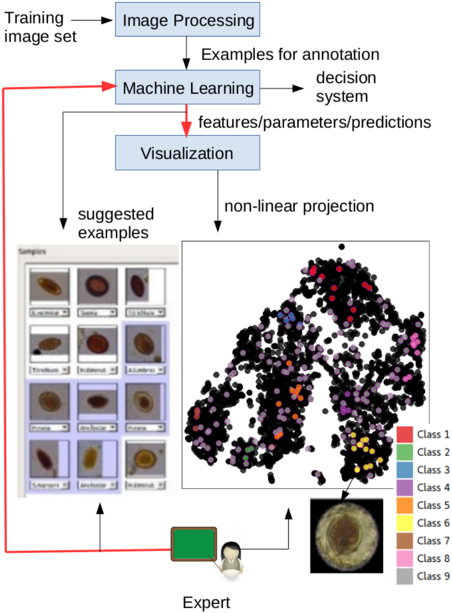


Feature projection after training from eight species of helminth eggs and impurities.

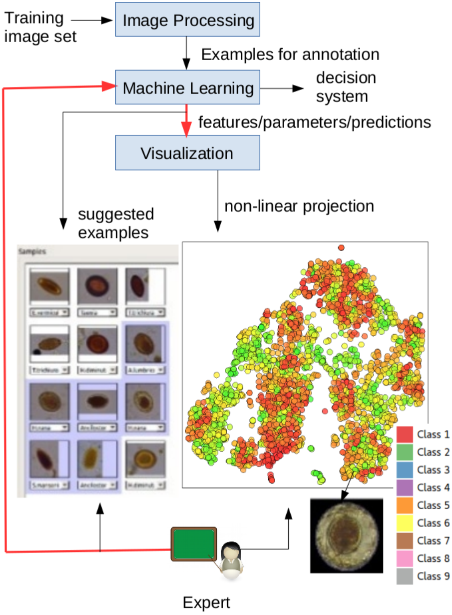
Methodology: how does it work?



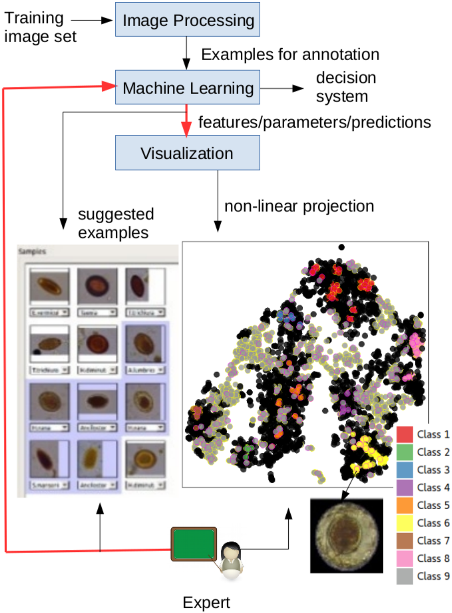
Methodology: how does it work?



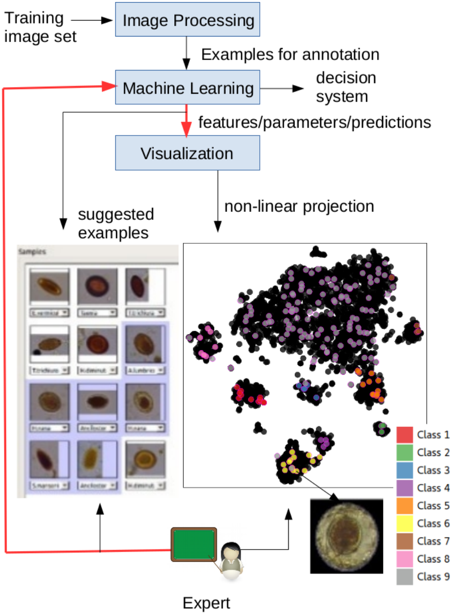
Methodology: how does it work?



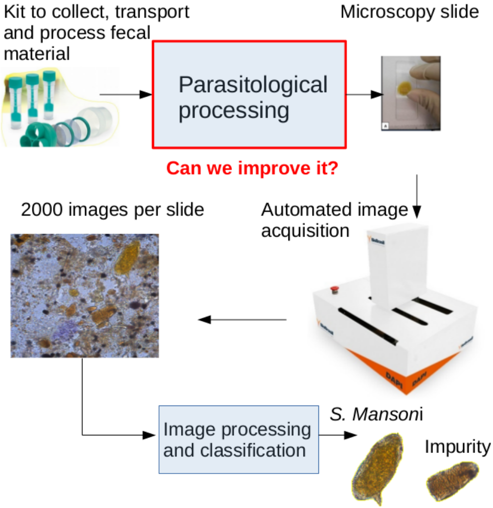
Methodology: how does it work?



Methodology: how does it work?



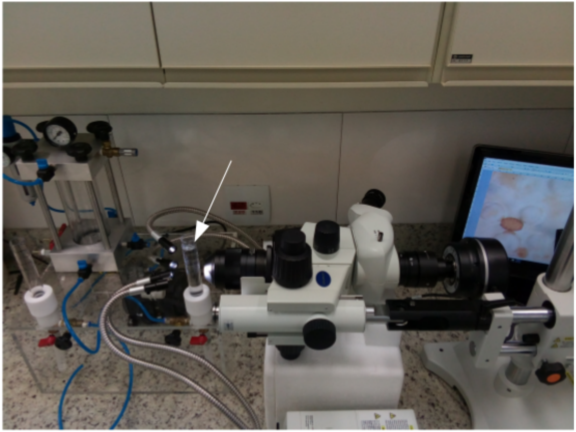
Methodology: a case of success



Automated diagnosis of human intestinal parasites for the 15 most common species in Brazil.

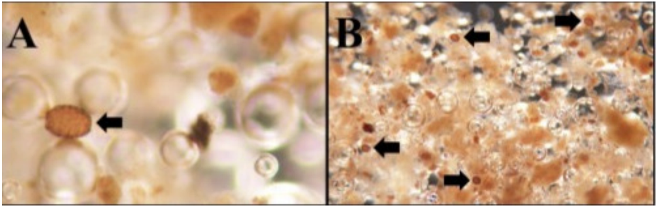
Methodology: interdisciplinary research

We have investigated flotation by dissolved air to create microscopy slides with more parasites and less impurities.



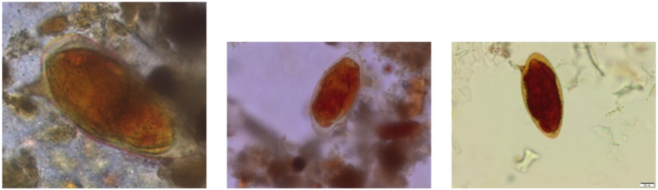
Bubbles of air are injected at the bottom of a tube containing the parasites and impurities.

Methodology: interdisciplinary research



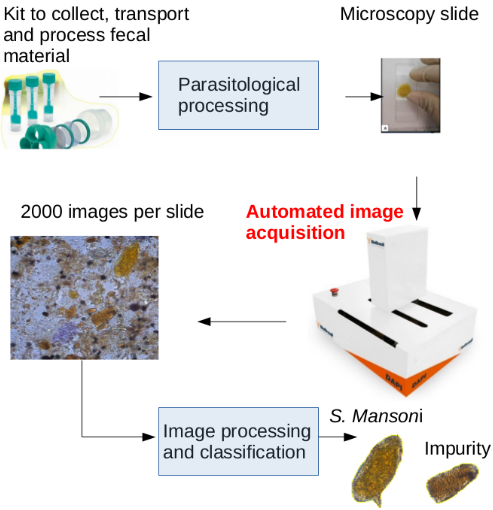
The bubbles carry the parasites upward the top of the tube, leaving most impurities at the bottom.

Methodology: interdisciplinary research



In comparison with the sedimentation principle, the flotation principle (right) can significantly improve image quality.

Methodology: a case of success



Automated diagnosis of human intestinal parasites for the 15 most common species in Brazil.

Methodology: interdisciplinary research

The evolution of the automated image acquisition system.

Other under-development applications

Same methodology, new data and algorithms.

- Remote sensing image classification for applications in Agriculture (e.g., a coconut tree counter).
- Anomaly detection in MR images of the brain (e.g., stroke).
- Anomaly detection in CT images of the thorax (e.g., lung nodules).
- Automated diagnosis of parasites in humans and animals.
- Construction of digital rocky outcrop (geological) models.

Final remarks

Human-machine interactions have potential to reduce human effort and increase human confidence in algorithmic decision.

However, easier said than done.

Our approach sheds light on how to build reliable image-based decision-making systems.

Acknowledgments

Thank you

FAPESP, CNPq, CAPES, ANP, the team of the LIDS and collaborators from different areas and countries.