Protein Crystal Detection

Team

Linda Wymann

Background in Geology with experience in the digitalization of construction. I am eager to create change in this industry using Data Science.

Claudio Cunha

Mechanical engineer recently turned Data Scientist. Interested in the Industry 4.0 revolution.

Our Baseline

- *CrystalsFirst Gmbh* is a biopharmaceutical start up combining machine learning with their proprietary lab technology for drug design


- A crystallographer at *CrystalsFirst Gmbh* screens thousands of images every week to find samples containing crystals

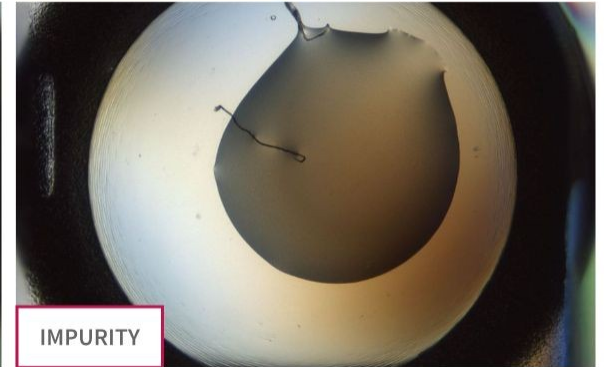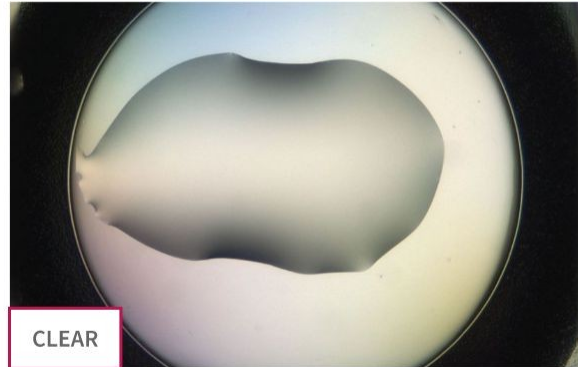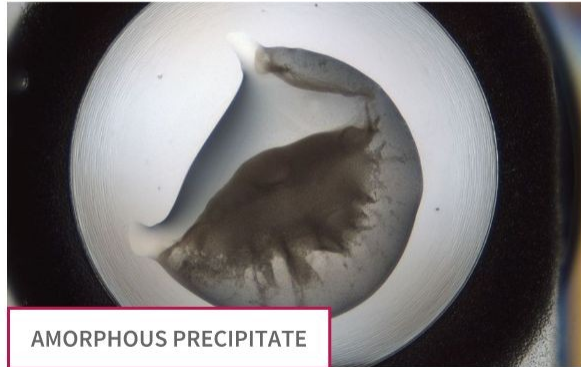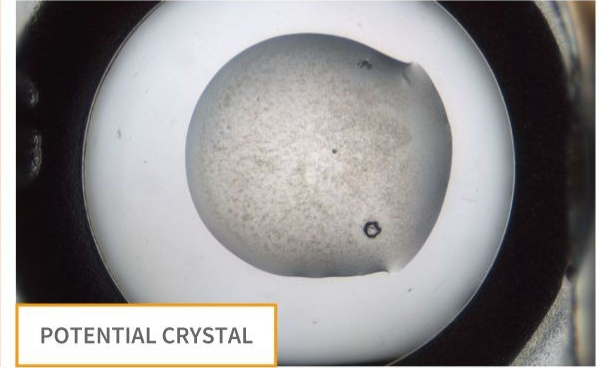- No more than 1% of images contain crystals
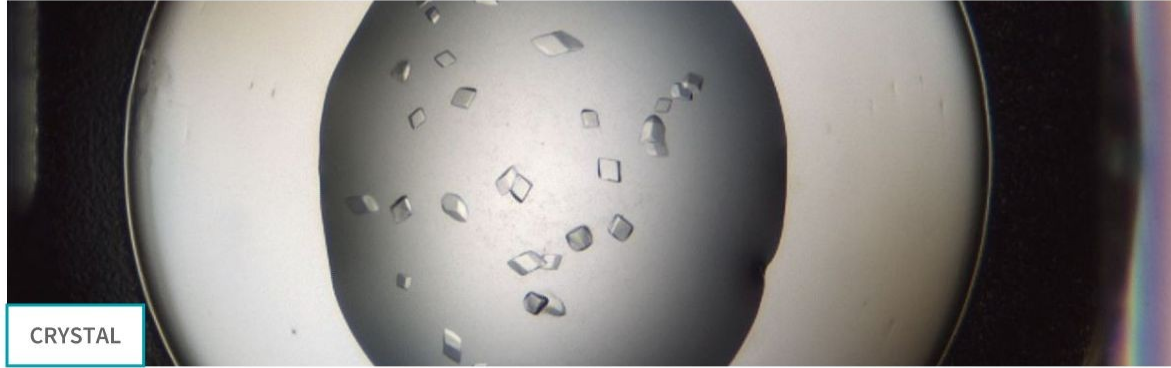
Our Goal

1. Reduce manual labor of a crystallographer

2. Not miss any crystal images

Our Approach

- **Treating it as a binary classification problem**

  crystal or **no crystal** instead of multi classification

- **Manually label images**

- **Compare various models and pick best performing one**

Inception V3
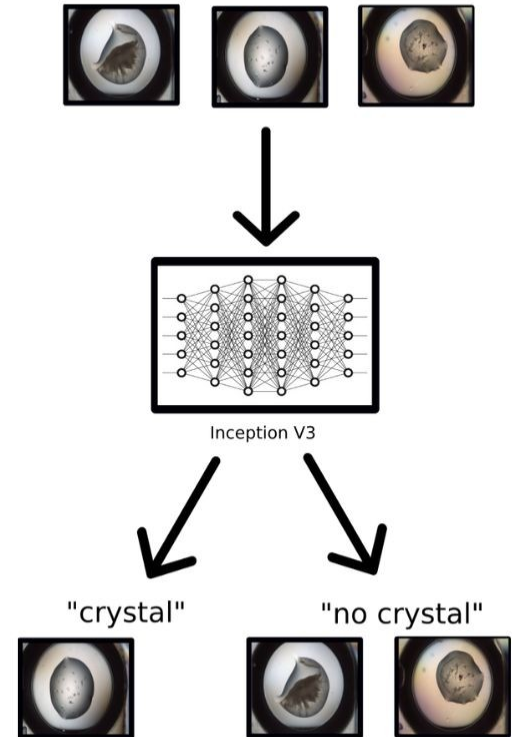
"crystal"   "no crystal"

Our Struggle - What *is* a Crystal?

CRYSTAL

POTENTIAL CRYSTAL

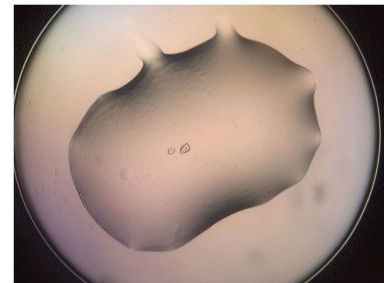AMORPHOUS PRECIPITATE

CLEAR

IMPURITY

Our Achievements - Detecting Crystals

- **9 out of 10 potential crystals were correctly identified on test set**

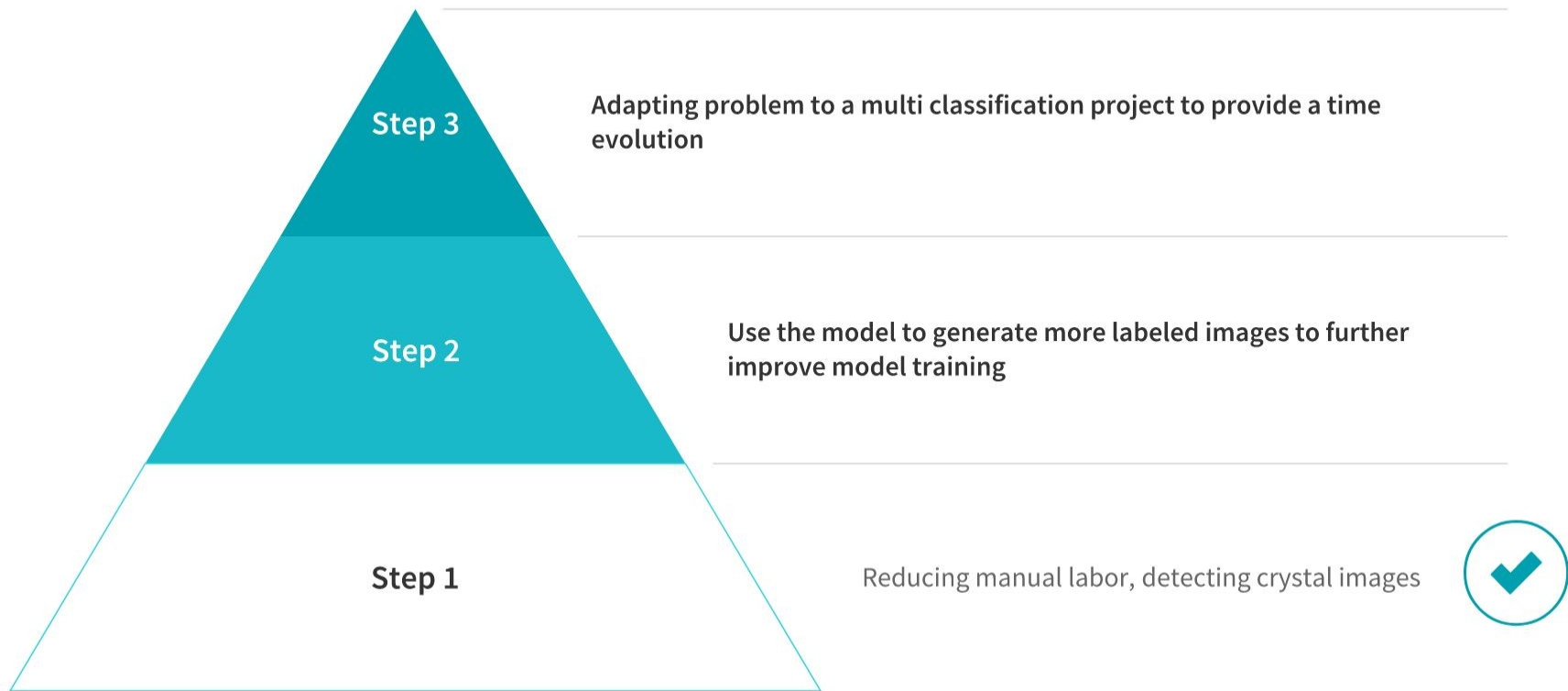 misclassified samples are either:

  - early stage crystal formation ➡ will evolve over time

  - impurity (e.g piece of plastic) ➡ will **not** evolve over time

- **Crystallographer will only inspect images of potential crystals**

  Reduce batch screening time by 99%

{Propulsion}

Further Improvements - Our Suggestions

Step 3

Adapting problem to a multi classification project to provide a time evolution

Step 2

Use the model to generate more labeled images to further improve model training

Step 1

Reducing manual labor, detecting crystal images

Thank you for your support

CRYSTALS FIRST

TURNING **INSIGHTS** INTO **MEDICINES**

MORITZ RUF
&
SERGHEI GLINCA

{Propulsion}

ALBIN PLATHOTTATHIL

BADRU STANICKI

NITIN KUMAR

LAURENT MEYER

Connect with us!

Linda Wymann

linkedin.com/in/linda-p-wymann

lpwymann@gmail.com

Claudio Cunha

linkedin.com/in/claudiojgcunha

claudiojgcunha@gmail.com