AMPLIVISION: AUTOMATED ANALYSIS OF MULTIPLEXED DIAGNOSTICS TO DETECT DISEASES

Matheus Nascimento Berbet[2], Josselyn Mata Calidonio [1], Daniel Haehn [2], Kimberly Hamad-Schifferli [1]

[1] Department of Engineering, University of Massachusetts Boston, Boston
[2] Department of Computer Science, University of Massachusetts Boston, Boston

Background

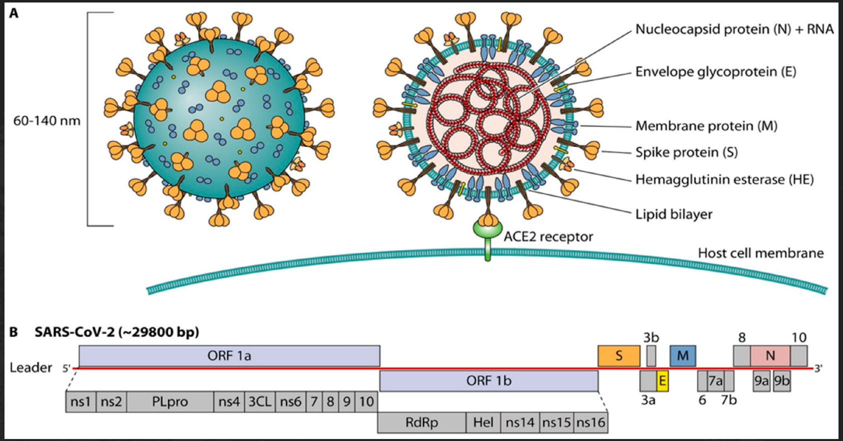
01

A Global Health Need For Automated Reconfigurable Rapid Diagnostics

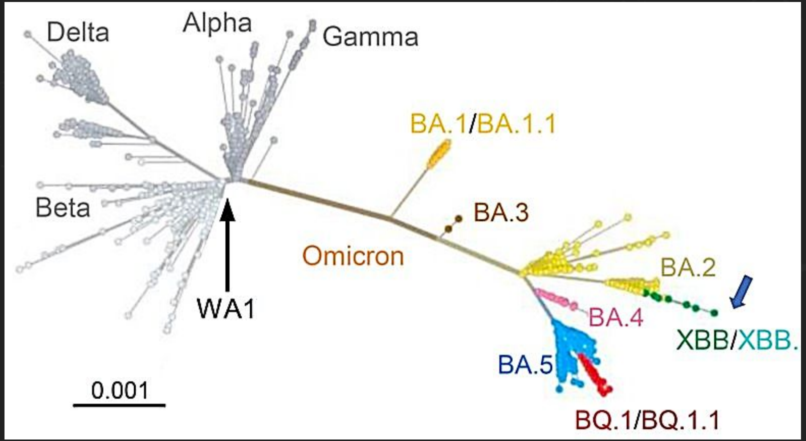
Background and Motivation: Global Health Need

Lateral flow tests **(LFAs)** can go from sample to answer within minutes leveraging **protein binding**.

However, the COVID-19 pandemic proved rapid test development **cannot keep up** with rate of mutation.



Protein Binding



COVID-19 variants

Background and Motivation: Global Health Need

Diagnostics, such as LFAs, help reduce disease transmission, especially in areas where resources are limited.

Dengue

41 Million new cases in India & Brazil in 2021. Source: IHME

Covid-19

103.4 Million total cases in the US alone. Source: WHO

Malaria

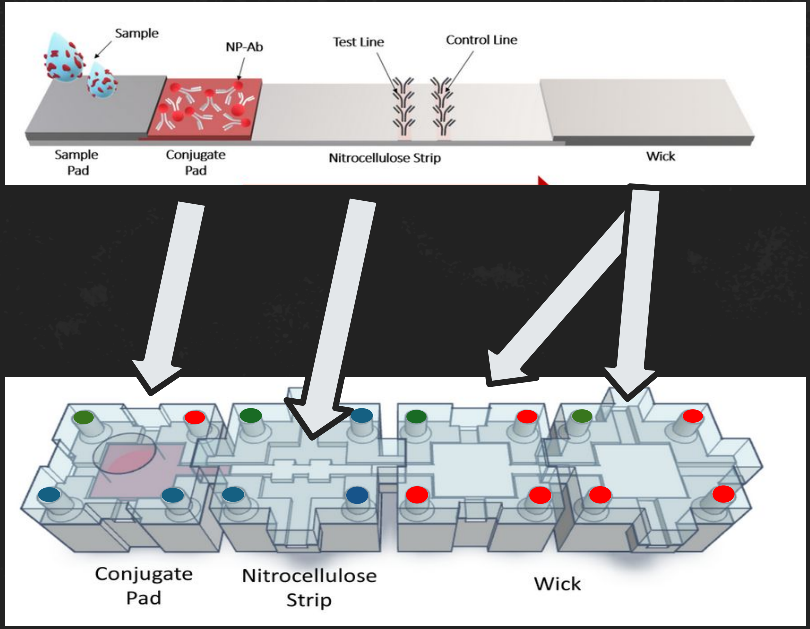
602 K new cases in Africa as of 2020. Source: IHME

Background and Motivation: Global Health Need

[1] AMPLI: A novel way to extension of LFAs.

- LFA components broken down into modular, LEGO-like, pre-assembled blocks
- Blocks can be snapped together on a grid in different, multiplex, configurations.
- This allows for easy modification of the test configuration as the target changes.

Components of a traditional LFA



Components of a Reconfigurable LFA (AMPLI blocks) [1]

[1] Elizabeth A Phillips, Anna K Young, Nikolas Albarran, Jonah Butler, Kaira Lujan, Kimberly Hamad-Schifferli, and Jose Gomez-Marquez, “Ampli: a construction set for paperfluidic systems,” Advanced healthcare materials, vol. 7, no. 14, pp. 1800104, 2018.

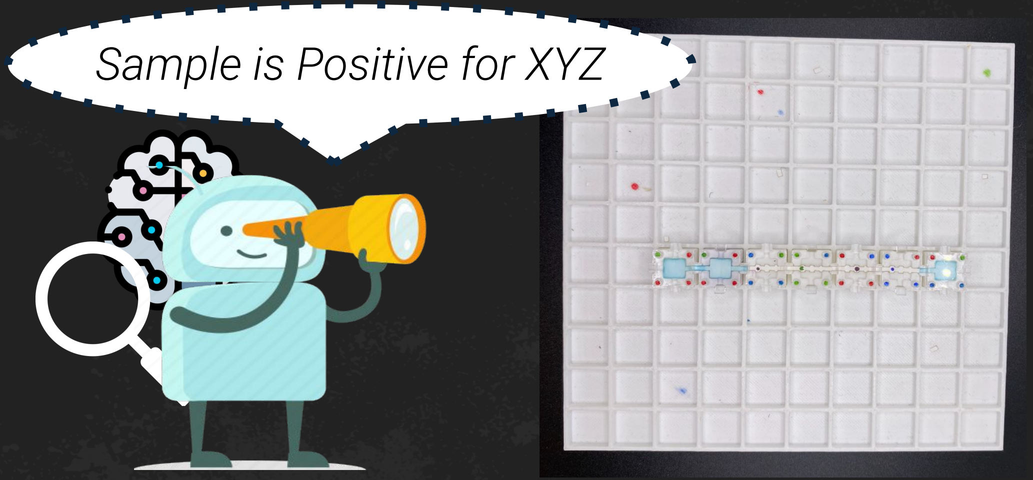
Background and Motivation: Global Health Need

Soooooooo what exactly do I have?

Multiplex tests bring added **complexity**. Underscoring the need for an **automated readout** as users may not be familiar with the test setup. Especially **visually impaired** users who may not be able to distinguish between results.

Objective

Build machine learning based software capable of **analyzing** and **classifying** test outcomes with high accuracy, including **outlier detection** for untrained patterns.



Methods

02

Methods

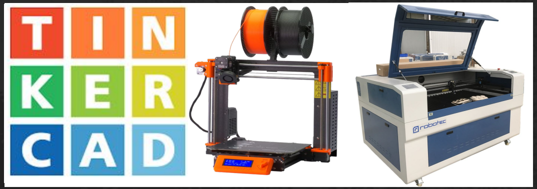
AmpliVision

- **Test Preparation** — **Designing**, **3D printing**, and creating **example tests**.
- **Data Generation** — **Identifying** test components, **extracting** results, and **generating** synthetic data.
- **Machine Learning** — Training a **LENET CNN** model to classify potential results.
- **Outlier Detection** — Recognizing outlier results, symbolizing **new variants**.

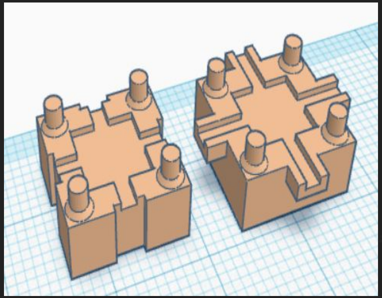
Methods: Test Preparation

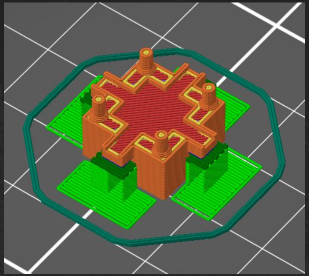
Designing and **3D printing**

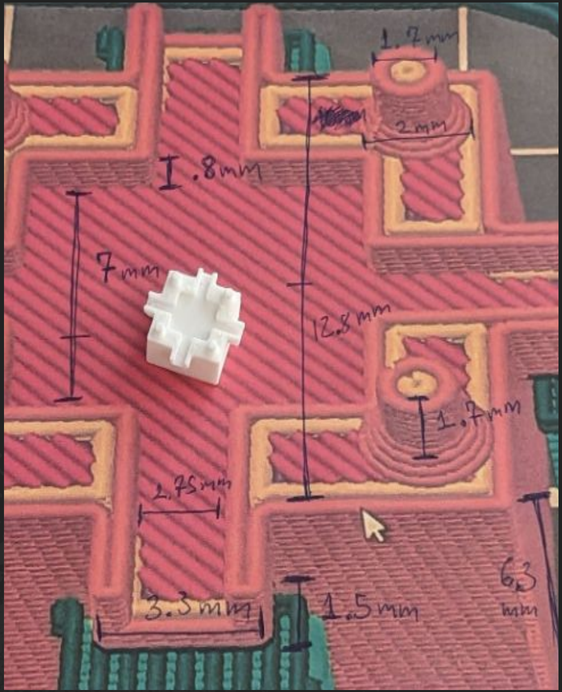
Toolset



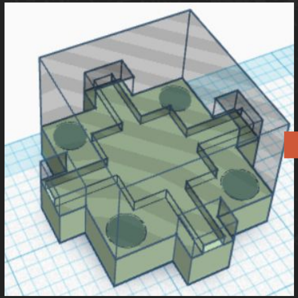
AMPLI Blocks

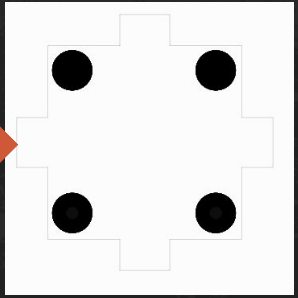




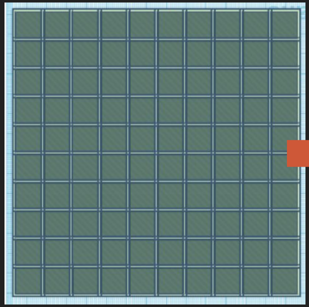


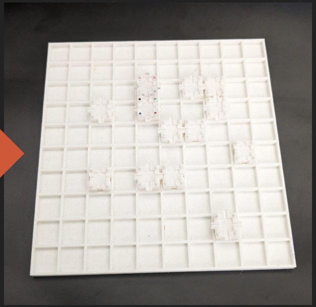
Block Lids





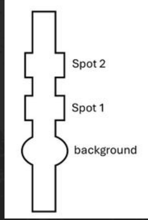
Grids





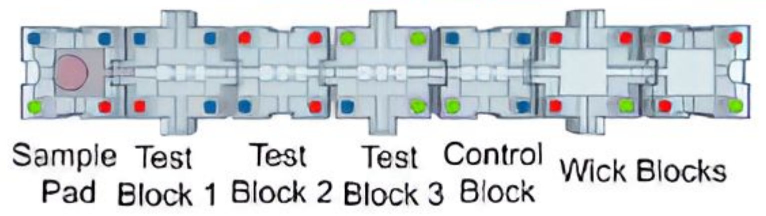
Methods: Test Preparation

Creating **example tests**

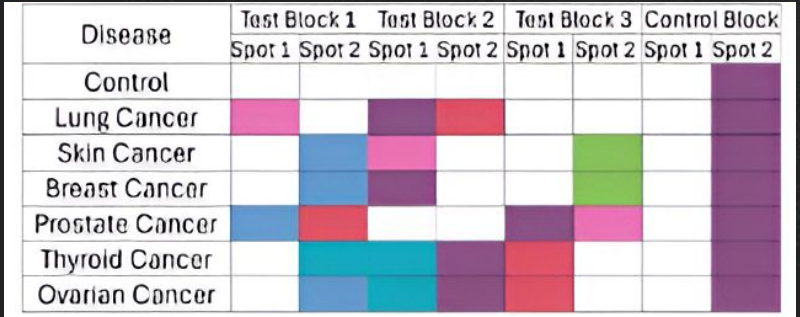


We colored the 4 pins on the block that secure the lid, where each color pattern was **unique** to each **block type**.

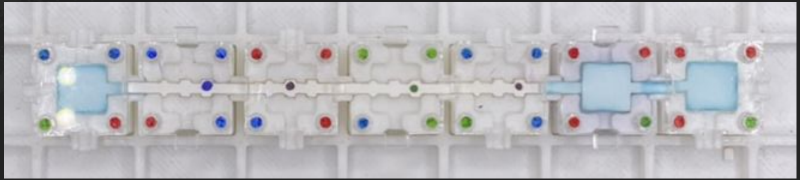
Patterns are **not rotationally degenerate** so block orientation can be read.



Block ID pattern library



We designed **simulated colorimetric assays** with a set of results representing 7 cancer classes.



Example simulated assay of class "Breast Cancer"

Methods

AmpliVision

~~Test Preparation~~

~~Designing, 3D printing, and creating example tests.~~

Data Generation

Identifying test components, **extracting** results, and **generating** synthetic data.

Machine Learning

Training a **LENET CNN** model to classify potential results.

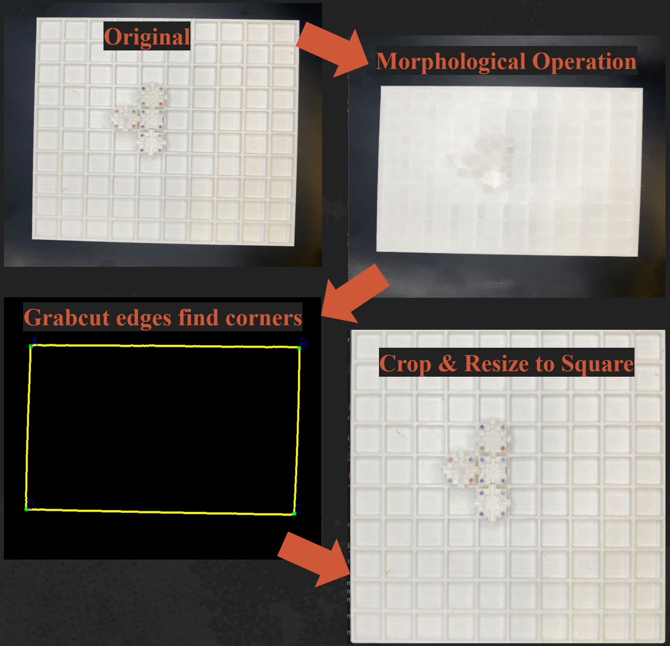
Outlier Detection

Recognizing outlier results, symbolizing **new variants**.

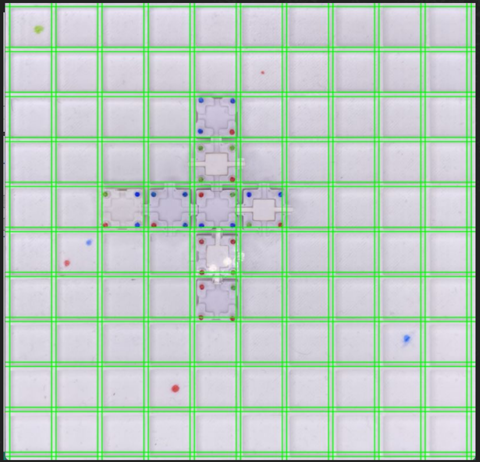
Methods: Data Generation

Preparing to Identify test components

Normalizing Images



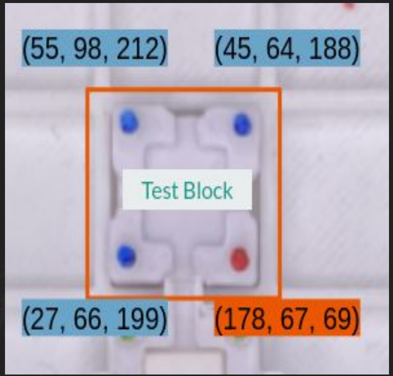
Virtualizing the Grid



Knowing the dimensions of the physical grid, and knowing that the normalized image only includes the grid, we can calculate where **each grid cell** is.

Methods: Data Generation

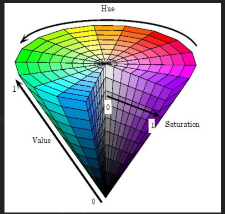
Identifying test components

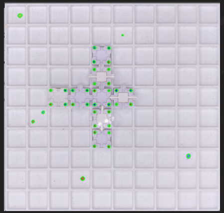


Using the **virtual grid**, we can:

- Find which colored contours (block pins) belong where.
- We can **identify** the block through the pin sequence.

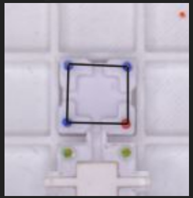
3-Step Identification

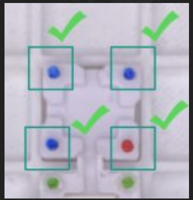


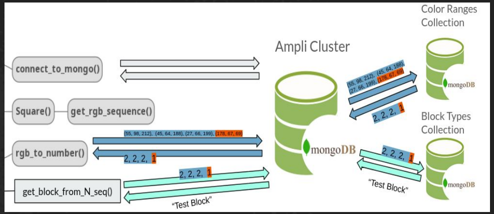


a. Find Pins through **HSV mask** + **Canny edge detection.**

b. Locate block by **arrangement** of pins.



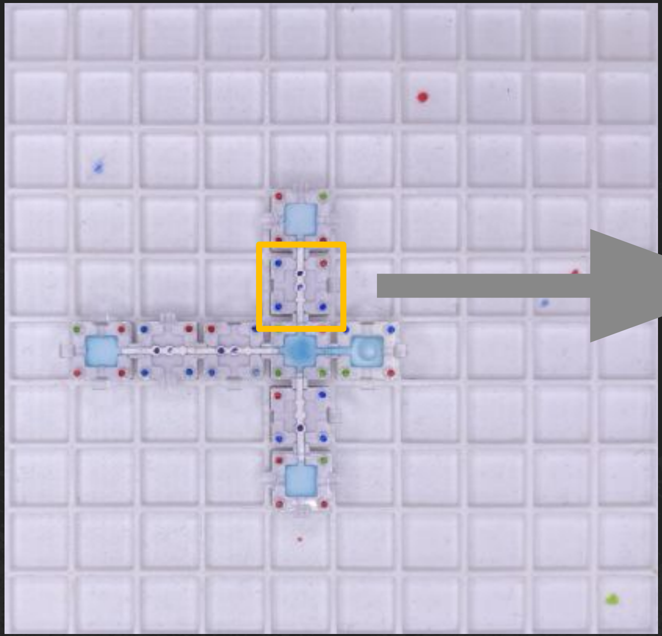




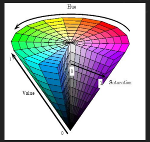
c. Identify block by RGB sequence.

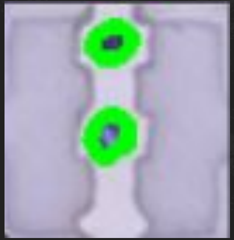
Methods: Data Generation

Extracting results



Positive Spot





Negative Spot

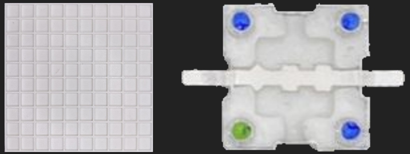
Knowing where the blocks are, we can zoom inside the test area and extract:

- **Positive spots** with an **HSV mask**.
- Negative spots by following the center line of test area divided by three.
- These results are **exported to CSV** file.

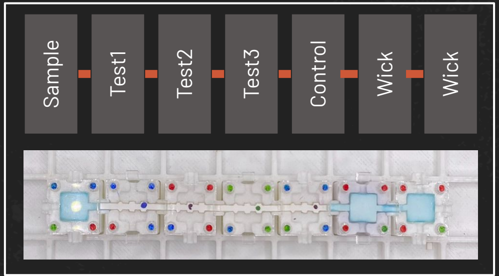
Methods: Data Generation

Generating synthetic data with a "collage" method

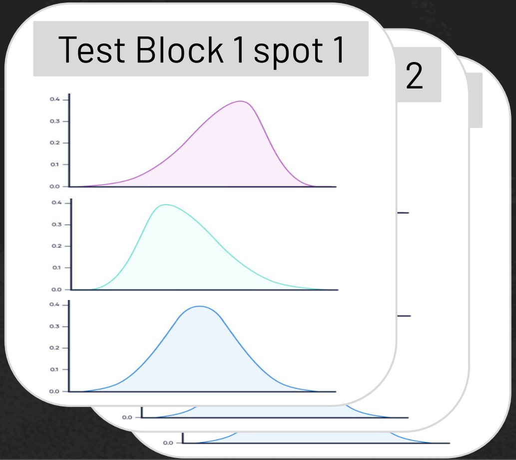
No BKG Components



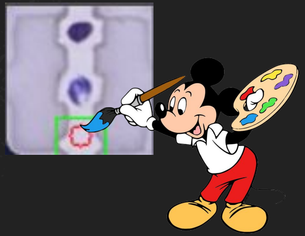
Virtual Grid Graph



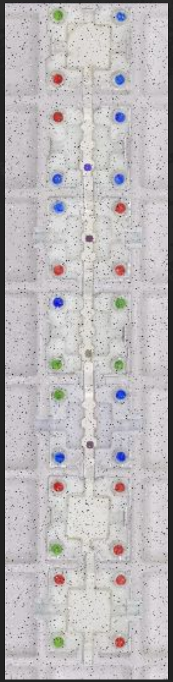
Extracted Fingerprint from CSV



Painting Spots



Generated Images



Methods

AmpliVision

- ~~**Test Preparation**~~ — ~~Designing, 3D printing, and creating example tests.~~
- ~~**Data Generation**~~ — ~~Identifying test components, extracting results, and generating synthetic data.~~
- **Machine Learning** — Training a **LENET CNN** model to classify potential results.
- **Outlier Detection** — Recognizing outlier results, symbolizing **new variants**.

Methods: Machine Learning

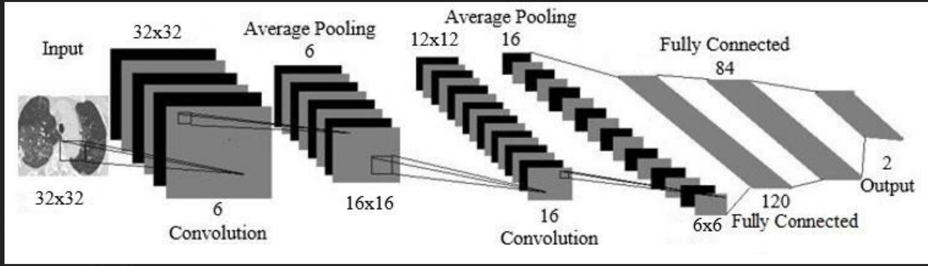
Training a **LENET CNN** model to classify potential results

Toolset

UMB's Chimera HPC

HIGH PERFORMANCE

LENET CNN Architecture [2]



Our workflow is adaptable to work with other models, LENET CNN was chosen as the initial proof of concept implementation.

Tensorflow Training Pipeline

Image Generator

LENET

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998

Methods

AmpliVision

- ~~Test Preparation~~ — ~~**Designing**, **3D printing**, and creating **example tests**.~~
- ~~Data Generation~~ — ~~**Identifying** test components, **extracting** results, and **generating** synthetic data.~~
- ~~Machine Learning~~ — ~~Training a **LENET CNN** model to classify potential results.~~
- Outlier Detection — Recognizing outlier results, symbolizing **new variants**.

Methods: Outlier Detection

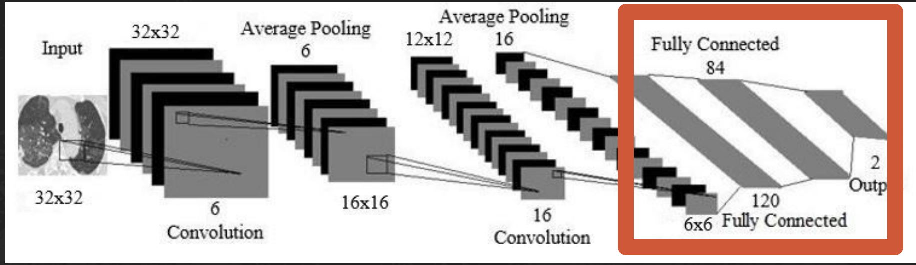
Recognizing outlier results, symbolizing **new variants**.

Toolset

UMB's Chimera HPC

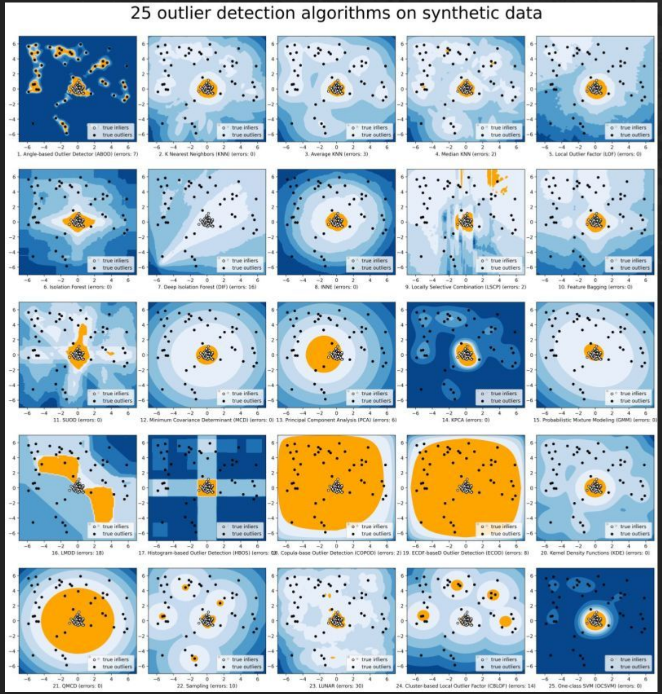
HIGH PERFORMANCE

Image Generator



Using dense of trained model as feature extractor.

PyOD (Python Outlier Detection) [3]



[3] Yue Zhao, Zain Nasrullah, and Zheng Li, "Pyod: A python toolbox for scalable outlier detection," CoRR, vol. abs/1901.01588, 2019.

Methods

AmpliVision

- ~~**Test Preparation**~~ — ~~Designing, 3D printing, and creating example tests.~~
- ~~**Data Generation**~~ — ~~Identifying test components, extracting results, and generating synthetic data.~~
- ~~**Machine Learning**~~ — ~~Training a LENET CNN model to classify potential results.~~
- ~~**Outlier Detection**~~ — ~~Recognizing outlier results, symbolizing new variants.~~

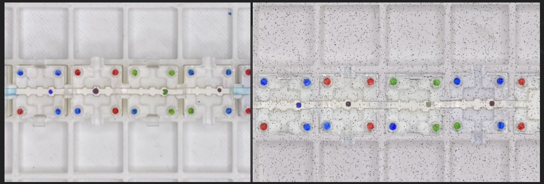
Results

03

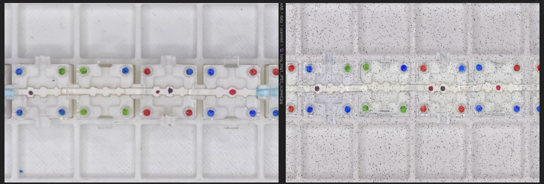
Results: Data Generation

Analyzing test pictures, and generating synthetic data

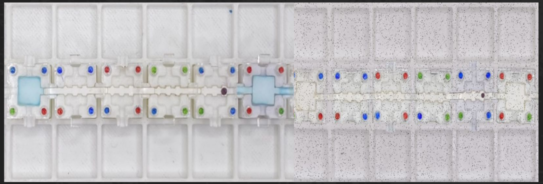
Breast Cancer



Lung Cancer

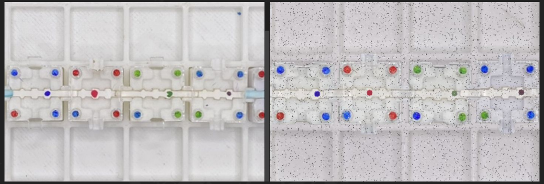


Control

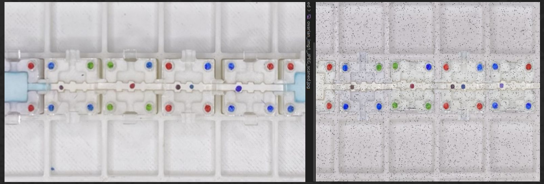


Prostate Cancer

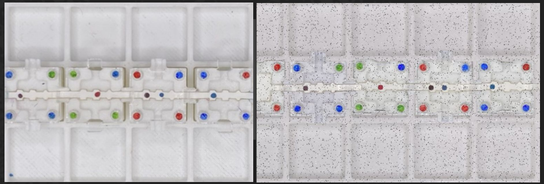
Skin Cancer



Ovarian Cancer



Skin Cancer

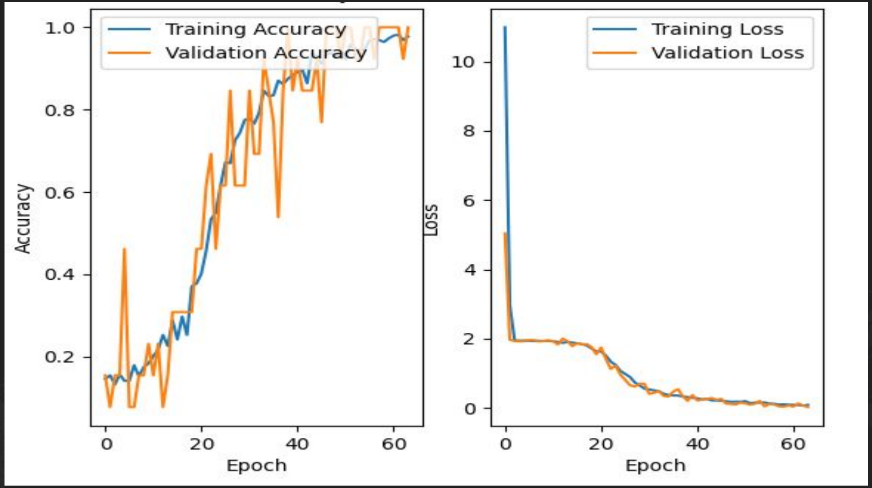


Input images (Left) compared with **generated images (Right)**, shows AmpliVision analyzes the tests **perfectly**.

Results: Machine Learning

Classifying results

Learning Curves



Generated image dataset sizes:

- **~30,000** for **training.**
- **~4,000** for **validation**
- **~1,300** for **testing**

Set	Accuracy(%)	F1 Score	Loss
Training	100	0.9933	0.0785
Validation	100	1.00	0.0475
Test	99.85	0.9970	0.0552

Results: Outlier Detection

Recognizing outliers

7 Best PyOD Results

Model	Accuracy (%)	F1 Score
Avg KNN	**98.18**	**.9091**
HBOS	94.54	.7273
IForest	92.72	.6364
CBLOF	96.36	.8182
KNN	96.36	.8182
MedKNN	**98.18**	**.9091**
Sampling	**98.18**	**.9091**

- Dataset of **100 inliers** and **10 outliers**.
- Highest-performing models had an **accuracy of 98%** and an **F1 score of .9091**.
- More testing is needed.

Conclusions

04

Conclusions

We have demonstrated:

- Software for **automated readout** of a Ampli tests which can be trained to recognize the colorimetric pattern of different targets.

These reconfigurable tests have the potential to make a **significant impact** on infectious disease control as **widely distributed user-centric diagnostics** by **improving diagnostic** rapid response, and **expand disease surveillance** tools for a public health impact.

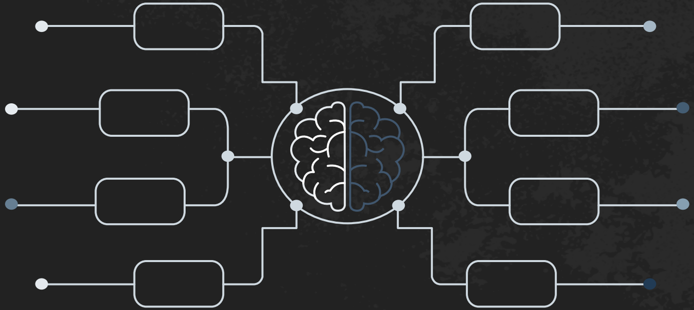
Future Work

05

Future Work

Utilize Yolo v3 for more robust block and test area detection.

Try the workflow on a real dataset such as DENV



Build a secure platform that allows collaborators to optimize our **pre-trained** CNN model for the specific AMPLI tests **they develop**. Ensuring that researchers can conveniently expand diagnostic targets AmpliVision supports.

Bibliography

[1] Cristina Rodriguez-Quijada, Jose Gomez-Marquez, and Kimberly Hamad-Schifferli, “Repurposing old antibodies for new diseases by exploiting cross-reactivity and multicolored nanoparticles,” ACS Nano, vol. 14, no. 6, pp. 6626–6635, 2020.

[2] Josselyn Mata Calidonio, Arianna I Maddox, and Kimberly Hamad-Schifferli, “A novel immunoassay technique using principal component analysis for enhanced detection of emerging viral variants,” Lab on a Chip, vol. 24, no. 16, pp. 3985–3995, 2024.

[3] Josselyn Mata Calidonio, Jose Gomez-Marquez, and Kimberly Hamad-Schifferli, “Paper-powered synthesis of gold nanoparticles and metronidazole for bioassays,” ACS Applied Nano Materials, 2024.

[4] Elizabeth A Phillips, Anna K Young, Nikolas Albarran, Jonah Butler, Kaira Lujan, Kimberly Hamad-Schifferli, and Jose Gomez-Marquez, “Ampli: a construction set for paperfluidic systems,” Advanced healthcare materials, vol. 7, no. 14, pp. 1800104, 2018.

[5] Siddarth Arumugam, Jiawei Ma, Uzay Macar, Guangxing Han, Kathrine McAulay, Darrell Ingram, Alex Ying, Harshit Harpaldas Chellani, Terry Chern, Kenta Reilly, et al., “Rapidly adaptable automated interpretation of point-of-care covid-19 diagnostics,” Communications Medicine, vol. 3, no. 1, pp. 91, 2023.

[6] Jiayi Chen, Pengjie Luo, Zhiwei Liu, Zixian He, Yi meng Pang, Hongtao Lei, Zhenlin Xu, Hong Wang, and Xiangmei Li, “Rainbow latex microspheres lateral flow immunoassay with smartphone-based device for simultaneous detection of three mycotoxins in cereals,” Analytica Chimica Acta, vol. 1221, pp. 340138, 2022.

Bibliography

[7] Jongwon Park, "Smartphone based lateral flow immunoassay quantifications," Journal of Immunological Methods, p. 113745, 2024.

[8] Josselyn Mata Calidonio and Kimberly Hamad-Schifferli, "Biophysical and biochemical insights in the design of immunoassays," Biochimica et Biophysica Acta (BBA) - General Subjects, vol. 1867, no. 1, pp. 130266, 2023.

[9] Josselyn Mata Calidonio, Jose Gomez-Marquez, and Kimberly Hamad-Schifferli, "Nanomaterial and interface advances in immunoassay biosensors," The Journal of Physical Chemistry C, vol. 126, no. 42, pp. 1780417815, 2022.

[10] Yue Zhao, Zain Nasrullah, and Zheng Li, "Pyod: A python toolbox for scalable outlier detection," CoRR, vol. abs/1901.01588, 2019.

[11] Ramin Dehghanpoor, Fatemeh Afrasiabi, Charles Fogel, Tung Dao, Suman Gautam, Aanab Nehela, Ahmad Nehela, Daniel Haehn, and Nurit Haspel, "Classifying protein families with learned compressed representations," in Proceedings of International Conference on Bioinfor, 2023, vol. 92, pp. 47–57.

[12] Daniel Haehn, James Tompkin, and Hanspeter Pfister, "Evaluating 'graphical perception' with cnns," IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 641–650, 2019.

[13] Zhenxing Cui, Lu Chen, Yunhai Wang, Daniel Haehn, Yong Wang, and Hanspeter Pfister, "Generalization of cnns on relational reasoning with bar charts," IEEE Transactions on Visualization and Computer Graphics, pp. 1–15, 2024.

Acknowledgements

- U54 DF/HCC-UMass Boston (NIH Grant number U54CA156734)
- Sloan Foundation as part of the UMass Amherst/UMass Boston Institute of Diversity (IDS) fellowship program
- Beacon Student Success Fund.
- The UMB Makerspace for use of their rapid prototyping facilities.

Thank You!

Do you have any questions?

Connect with me

linkedin.com/in/matheusnberbet

matheus.berbet001@umb.edu

matheusberbet.site