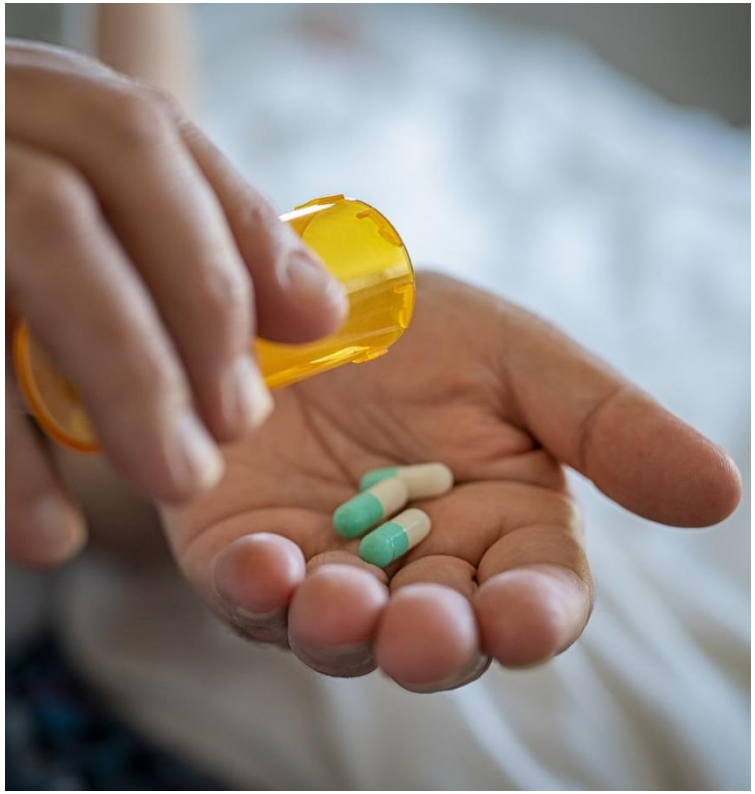
Predicting HIV Viral Load

Hyojin Lim and Tom Lindman

UNIVERSITY *of* WASHINGTON

HIV Viral Load Testing

- Without a cure, antiretroviral therapy (ART) is central to treating HIV[1]
- Viral load testing essential to ART
- Viral load testing essential to combating the HIV epidemic



1. Drain et. Al (2020)

Viral Load Testing: Challenges

- Quantitative polymerase chain reaction (qPCR) the "gold standard" for viral load testing[2]
- qPCR requires:
  - Expensive lab equipment
  - Highly trained staff
  - Time
- Implementation challenges in low- and middle-income countries[3]

2. Zhao et. Al (2019)
3. Greig et. Al (2020)

Viral Load Testing: Potential Solutions

- Isothermal DNA amplification offers potential qPCR alternative
- Reaction produces florescent spots on paper slip
  - Spots correlate with viral load
  - Can process tests using phone application
- Challenge: prediction at high spot concentration
- Project goal: increase prediction range, accuracy

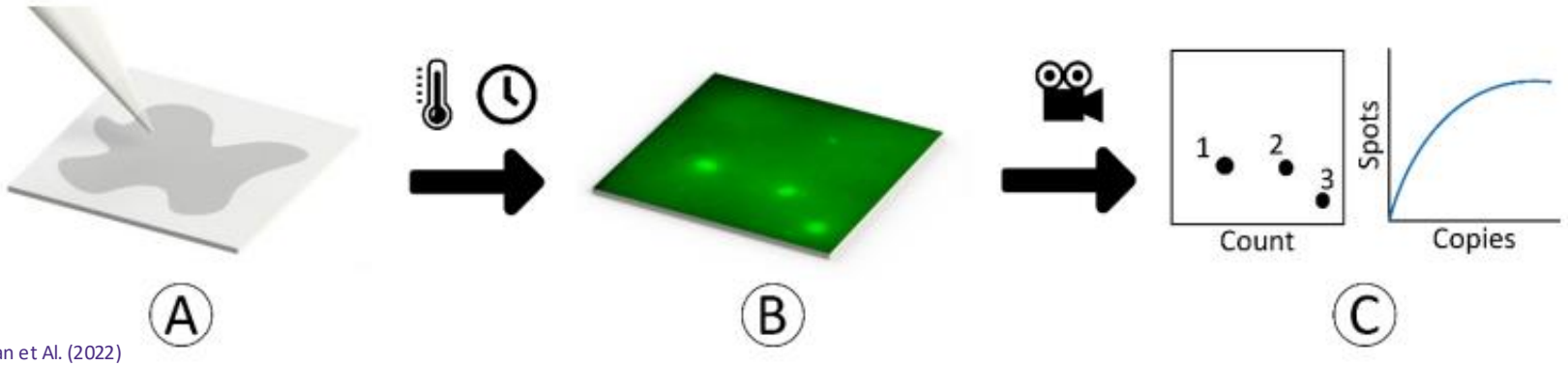


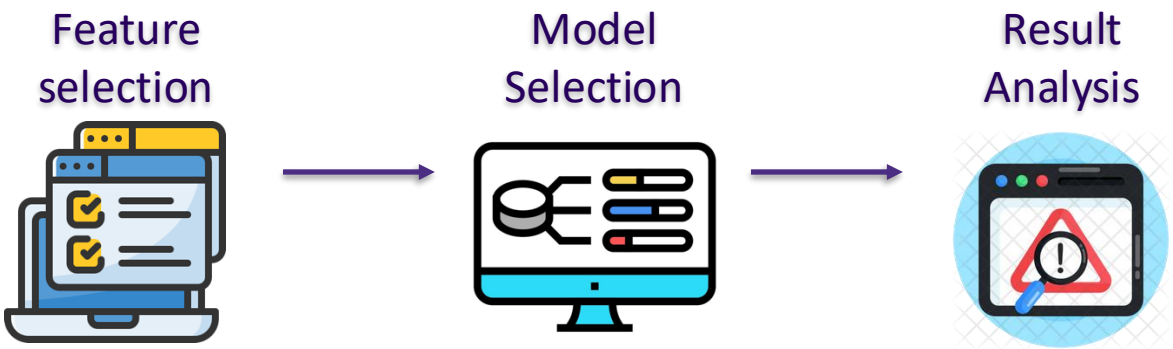
Image credit: Sullivan et Al. (2022)

User and Use Cases

- **Users: Main Project team**
  - **Cole Martin and collaborators (primary users)**
  - **Use cases:**
    - **Update phone app to predict viral load using our model**
    - **Update our model using new data, features, and ML algorithms**
- **Users: Clinicians testing phone app**
  - **Clinicians working with patients on ART**
  - **Use case:**
    - **Monitor patient viral load using our model**
    - **Interact with model exclusively via phone app**

Prediction with ML models

- ML models are effective in analyzing data
- Lightweight ML models can be run on mobile devices
- Key points:
  - Feature Selection: Use minimal, relevant data
  - Model Selection: Opt for lightweight & accurate models.
  - Result Analysis: Evaluate to enhance accuracy.



Feature selection

- **Key features: 11 features in total**
  - **max_fluorescence_change: Maximum change in fluorescence**
  - **max_area_change: Maximum area change**
  - **max_spot_change: Maximum spot change**

Critical features for extending the model's dynamic range

  - **percent_area: Percentage of area covered**
  - **n_spots: Number of spots identified**

Less critical

Model selection

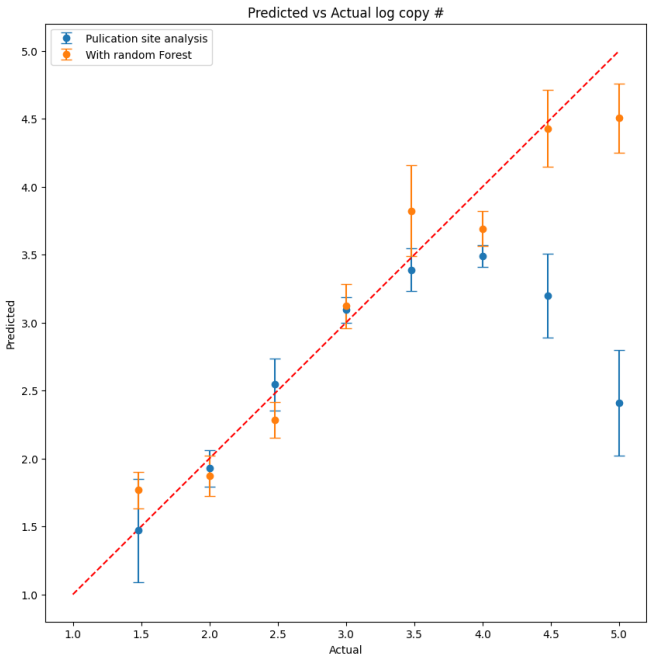
- **Models tested**
  - **Linear Regression: Simple, interpretable**
  - **Random Forest: Handles non-linearity, reduces overfitting**
  - **XGBoost: Fast, high-performance gradient boosting.**

	Linear Regression	Random Forest	XGBoost
Average Train MSE	0.0090	0.0028	0.0000
Average Test MSE	0.0127	0.0152	0.0138

Image credit: Sullivan et Al. (2022)

Result analysis and key takeaways

> Result analysis



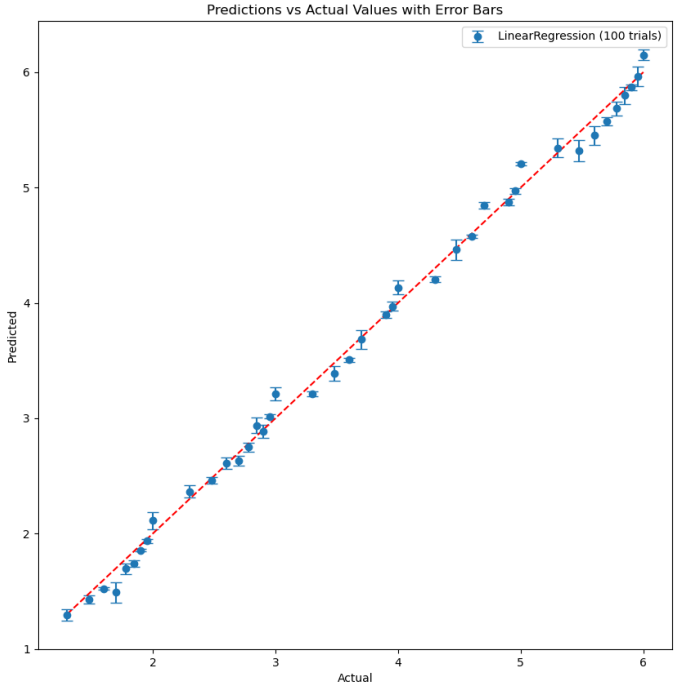


Image credit: Sullivan et Al. (2022)

Future work

- Collect more data
- Conduct sensitivity analysis
- Fully integrate Android App

Image credit: Sullivan et Al. (2022)