On-Device Prediction for Chronic Kidney Disease

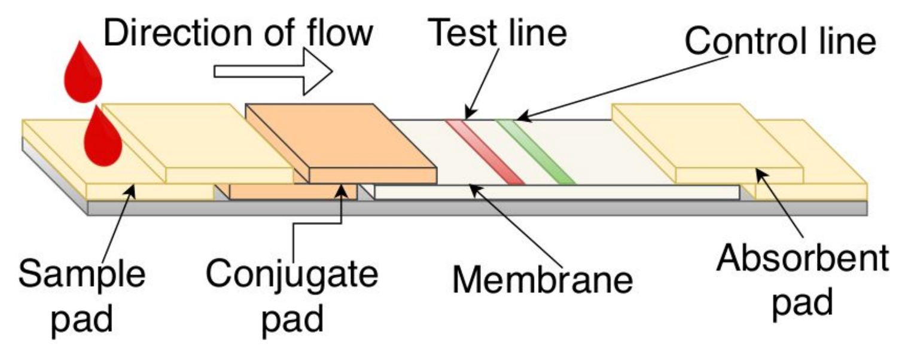
Alex Whelan, Soham Phadke, Alessandro Bellofiore, and David C. Anastasiu

Santa Clara University

Problem Description

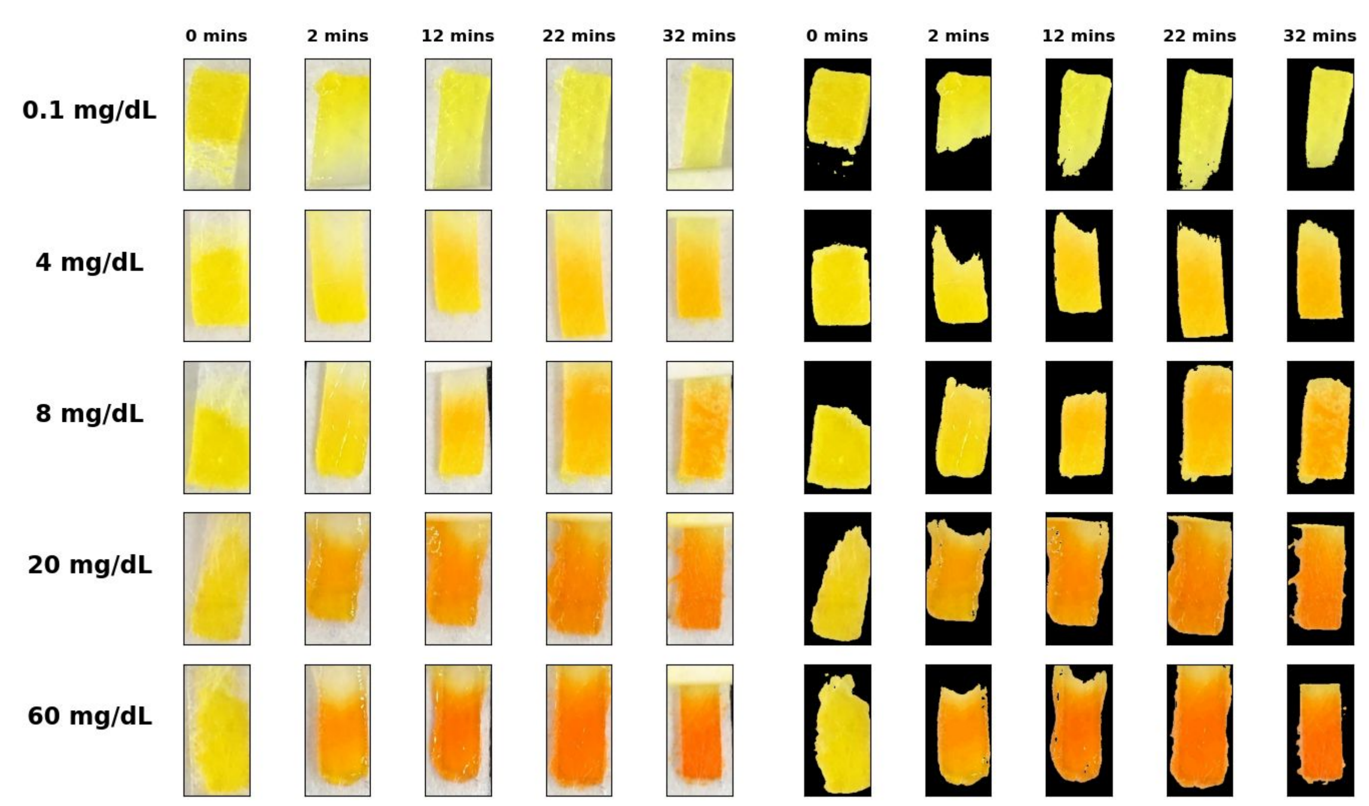
Problem: Predicting solution concentrations (SC) of LFAs in order to help diagnose early stages of Chronic Kidney Disease (CKD)

- Challenges
  - Existing CKD screening methods are expensive and/or non-accessible or unavailable
  - Color regression is an open area of research
  - LAB testing is difficult to scale as a PoCT
- Goal
  - Develop a model for fast and accurate CKD screening
  - Point-of-Care Testing (PoCT) for underdeveloped countries



Lateral Flow Assay (LFA) design

Dataset



Mask by Thresholding

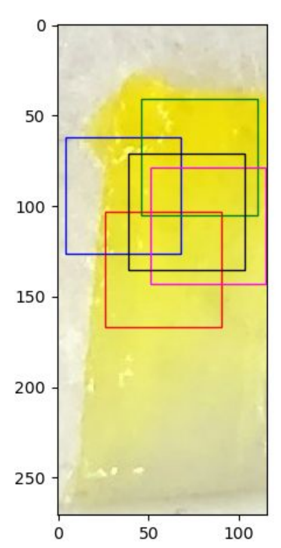
$$\text{logical mask} = \prod_{i,j,k} T_{k,min} \leq I_{i,j,k} \leq T_{k,max}$$

Where **i,j,k** represent the image's height, width and color channel channel

$$hue \in (0, 40)$$
$$saturation \in (128, 255)$$
$$value \in (128, 255)$$

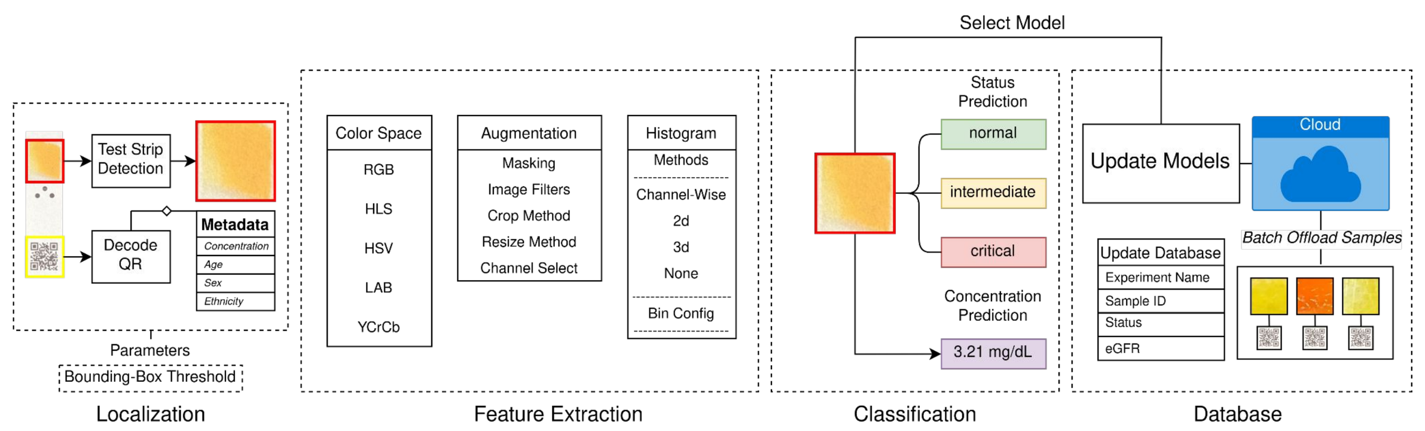
Dataset Augmentation

- Augmented crops are are assigned new metadata from the population distribution and treated as a new sample
  - Centralized Random Crop (right)
    - min,max IOU threshold b/w center crops
  - Tiling
    - Partition sample into uniform sub-crops
  - Random Crops

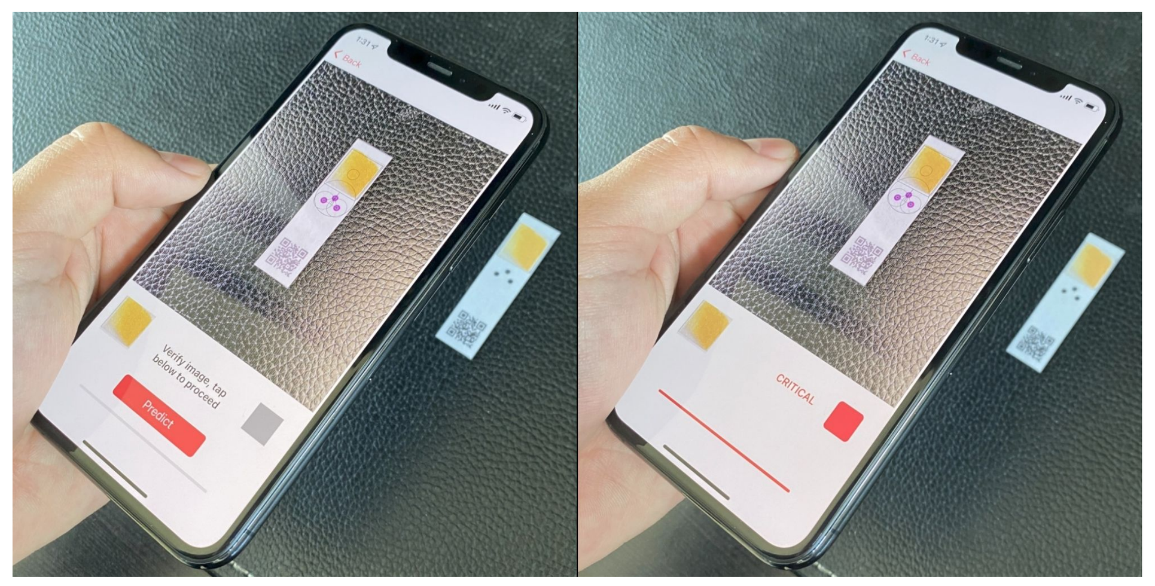


Proposed Framework

- **Model Pipeline & Application Design**
  - Our framework allows for rapid model training and evaluation using various configurations
  - A pre-trained model is then exported into Apple's CoreML library for use in our application
  - Predicted samples are uploaded into Firebase for future model re-training



Model Framework

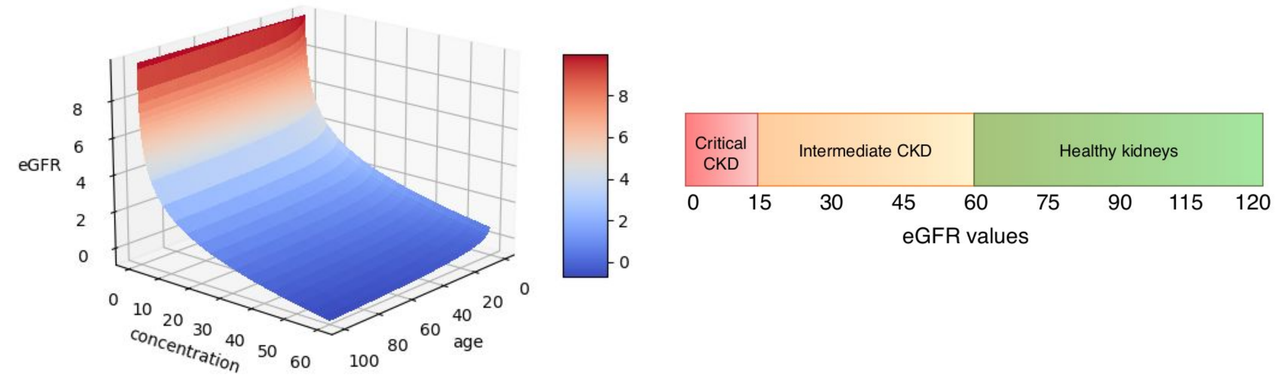


KHP Application

Predicting CKD Status (Classification)

- **Estimated Glomerular Filtration Rate (eGFR)**
  - Using a population distribution we assign samples an age, sex, and ethnicity
  - We predict status & serum creatinine concentration (S_{cr})

$$eGFR = 175 \times S_{Cr}^{-1.154} \times \text{age}^{-0.203} \times 0.742 \text{ if female} \times 1.212 \text{ if African born}$$

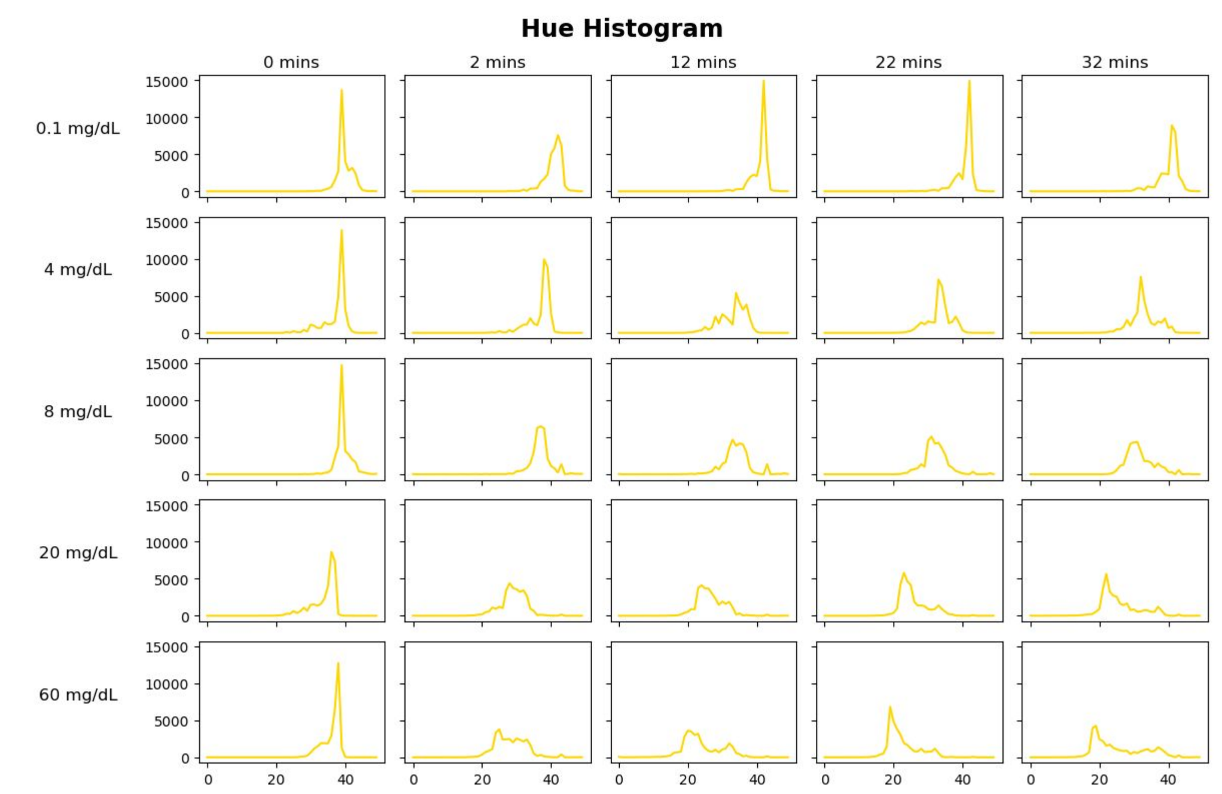


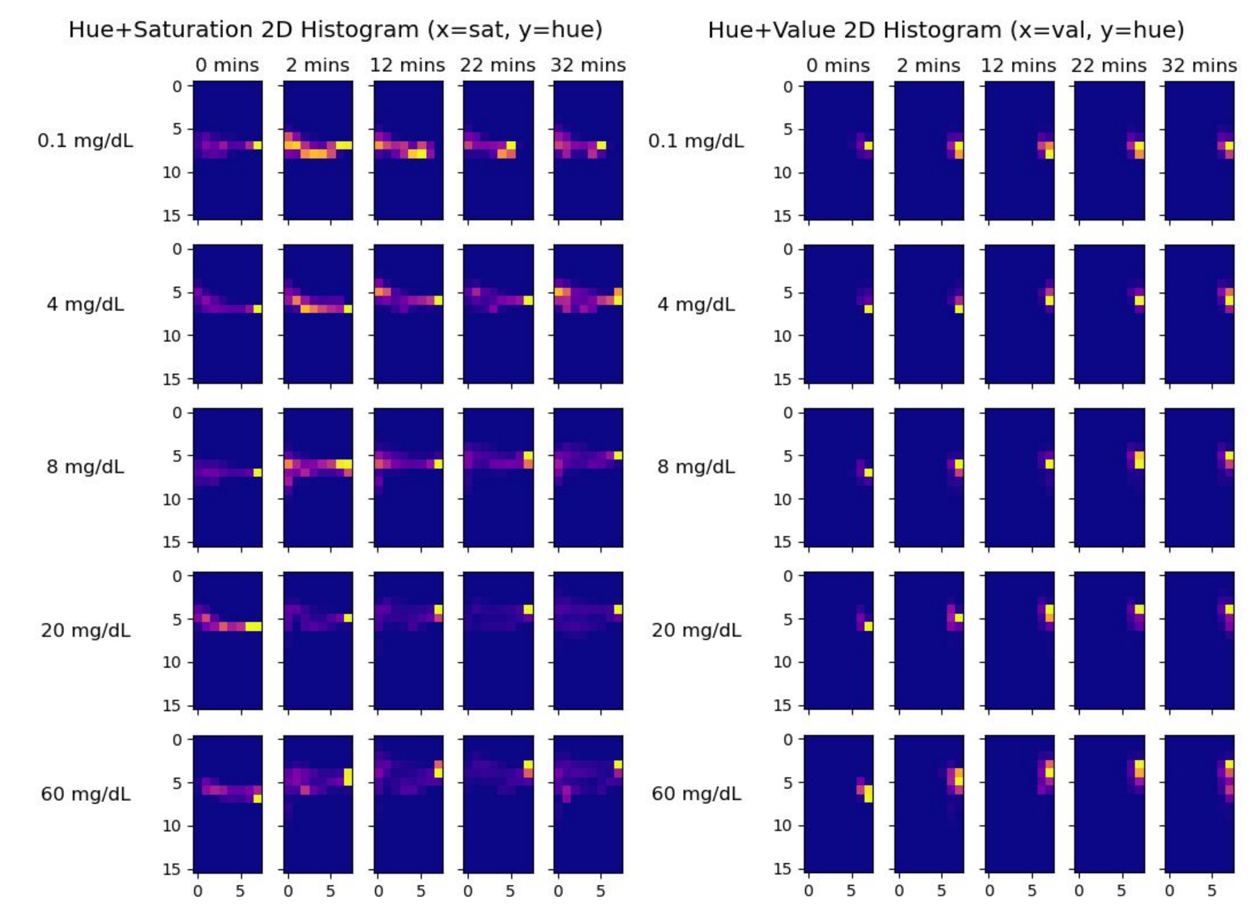
Stage of CKD	STAGE 1	STAGE 2	STAGE 3A	STAGE 3B	STAGE 4	STAGE 5
eGFR	90 or greater	Between 60 and 89	Between 45 and 59	Between 30 and 44	Between 15 and 29	Less than 15
Level of kidney damage	Mild kidney damage	Mild kidney damage	Mild to moderate kidney damage	Mild to moderate kidney damage	Moderate to severe kidney damage	End-stage kidney disease. Kidneys are close to failure or have completely failed. You will need to start dialysis or have a kidney transplant.

https://www.ckdandt2d.com/managing-chronic-kidney-disease

Histogram Construction (Feature Extraction)

- A histogram is the frequency distribution of values in a given data set
  - Binned Histogram is an extension where we map channel values in the range (0,255) to equidistant bins β
  - Histograms are spatially-invariant therefore size of image is irrelevant in the histogram construction

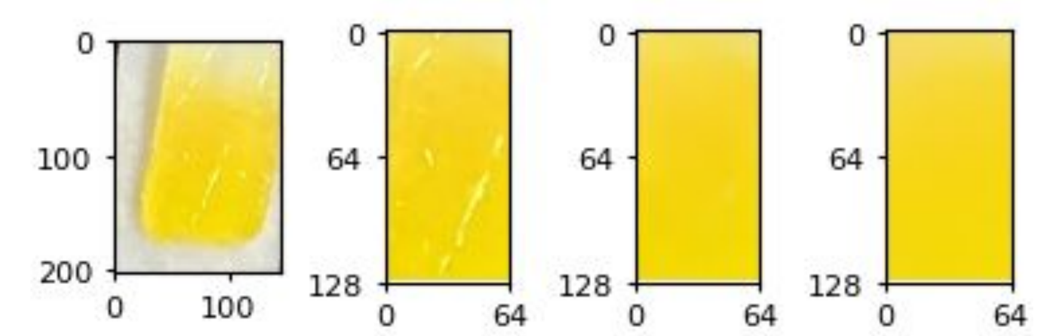




- We construct our histogram in 3 different ways
  - Channel-wise : Single Histogram per channel concatenated
  - 2d/3d : A conjunctive AND construction where a single pixel x is considered if there are bins α,β,γ s.t x is an element of all
  - None : We can also pass the raw pixel values

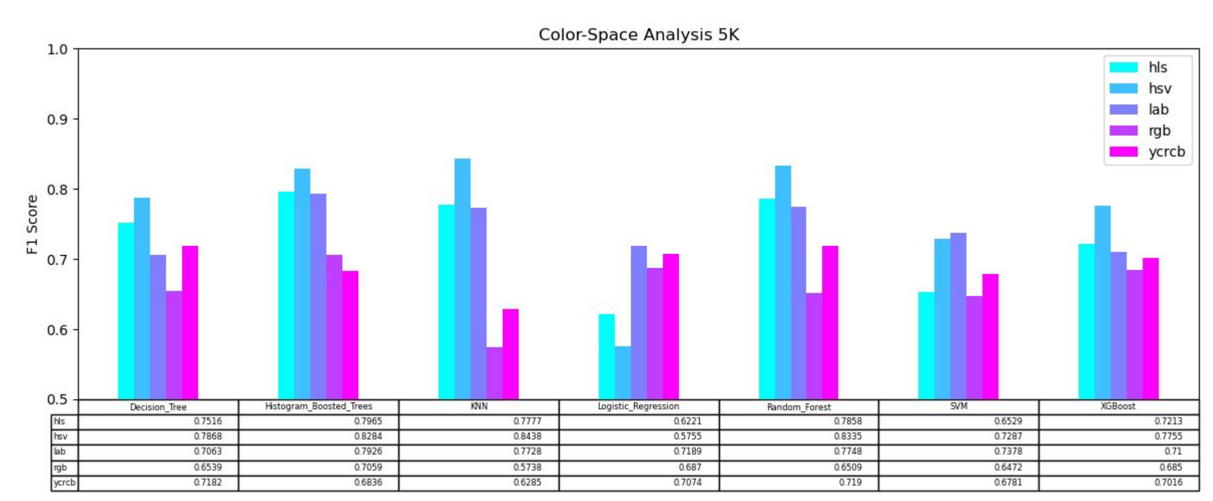
Image Augmentation

- Cropping is preferred to resize because we want to avoid color interpolation
- Artifacts are removed by applying median filtering



Color Spaces

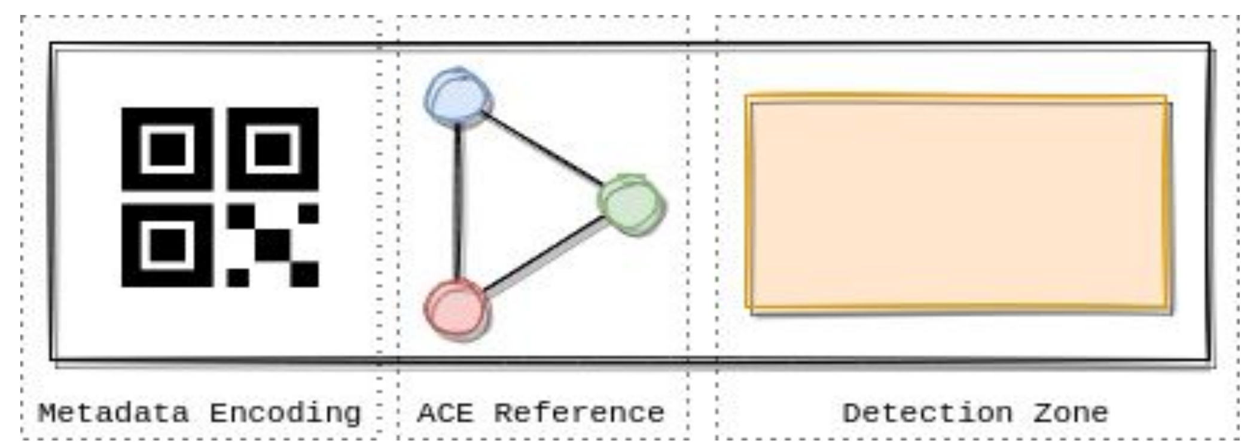
- Experimenting with various color spaces HSV gave best results



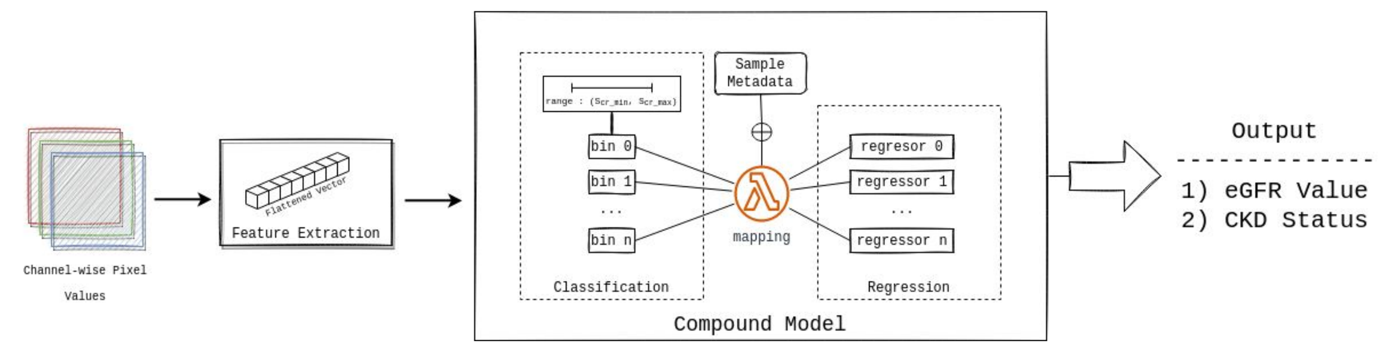
Future Works

- **Automatic Color Equalization**
  - Color constancy due to various light sources
  - Gamma Correction
  - Contrast-limited Histogram Equalization
  - Adaptive Histogram Equalization

Concept Teststrip Design



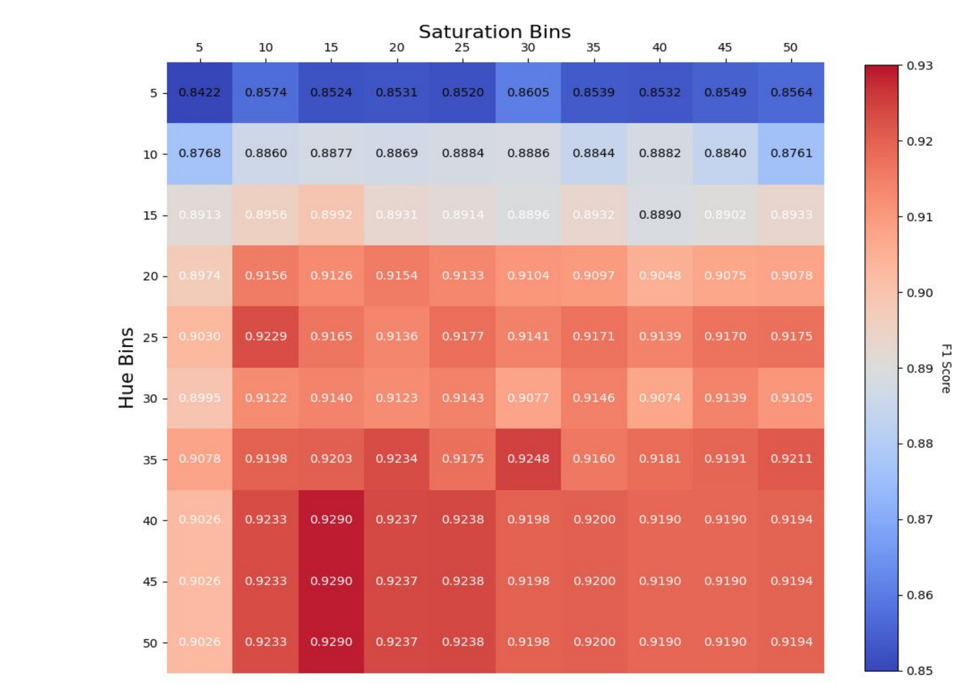
- **Lambda Mapper**
  - Unsupervised method for assigning regressors
  - Mixing sub-estimators



Evaluation

- After a model configuration is chosen we perform an extensive gridsearch
  - Each model population is re-seeded 3 times
  - Hyper-parameters are explored at each bin point (x,y) for 200 times using 10-fold cross-validation
  - Total training approximately 1 week or 168 hours on 2 RTX 3090 GPUs with 24 Gb Memory

Sample KNN Grid



Optimal Bin Sizes

Model Comparison