Penn Medicine

The 20th IEEE International Conference on BioInformatics And BioEngineering

Estimating Hard-tissue Conditions from Dental Images via Machine Learning

Authors: Jingxuan Bao, Mansu Kim, Qing Sun, Anderson T. Hara, Gerardo Maupome and Li Shen

Presenter: Jingxuan Bao
Dr. Li Shen's Lab Student Researcher
Applied Mathematics and Computational Science
University of Pennsylvania

Oct. 2020

Motivation

- Machine Learning applications to dental hard tissue conditions are under-explored, in particular for analyzing photographic images.

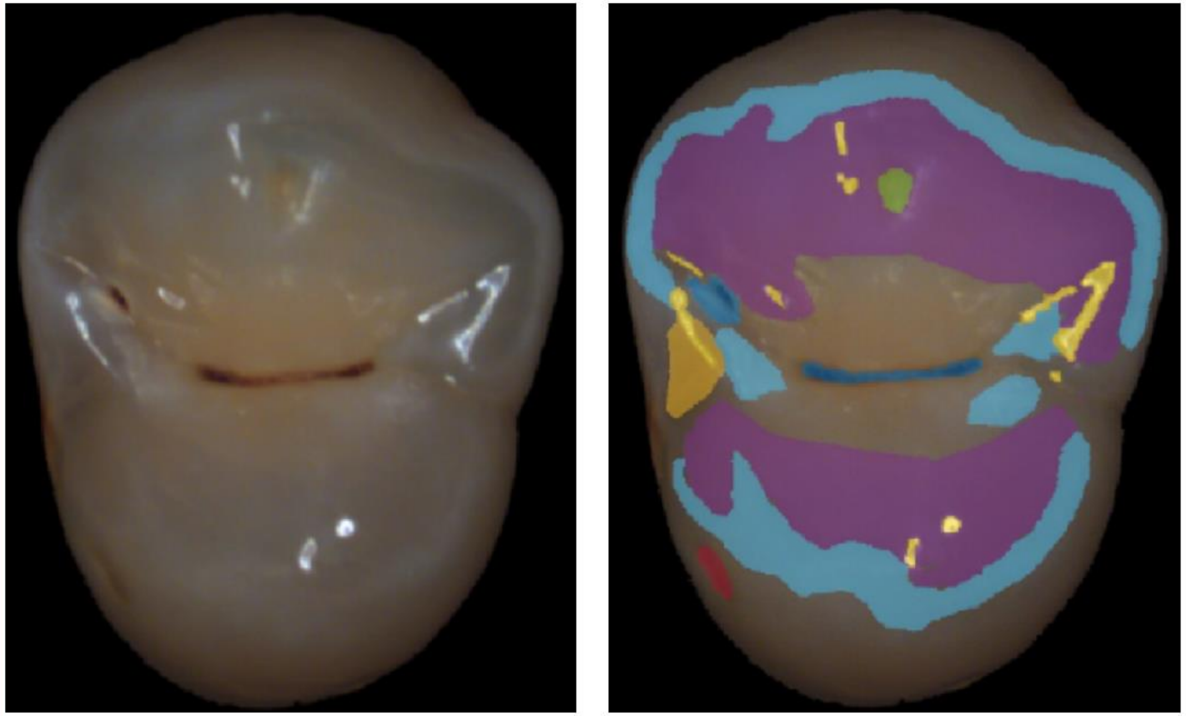
- The clinical diagnostics of dental hard-tissue conditions is routinely performed by visual examination but is often limited by its subjectivity.

- The very limited amount of dental photographic images leads to overfitting when using deep neural networks.

Aim and Data

Given a tooth image, we apply regression-based machine learning methods to estimate the **probability** and **severity** of a hard-tissue condition on the tooth image.

(a) Tooth image (b) Labeled image

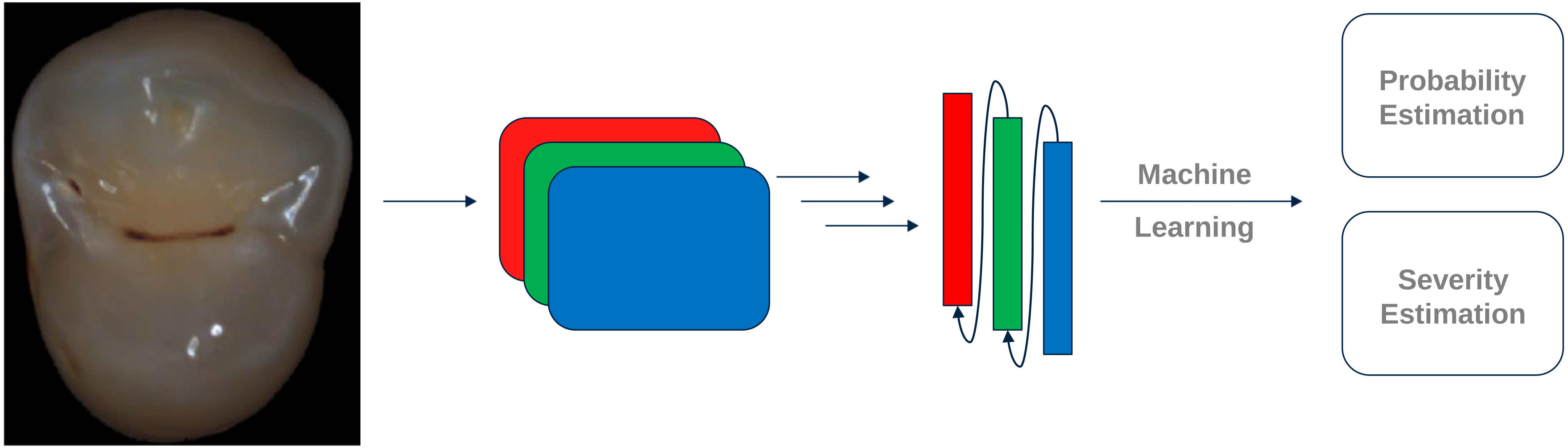


(c) Label information

Condition	Label	Color
Caries	Caries_Enamel	
	Caries_Dentin	
	Caries_Under	
ETW	ETW_Enamel	
	ETW_Dentin	
Fluorosis	Fluorosis	
	Fluorosis_pitted	
Other	Hypoplasia	
	Cracking	
	Restoration	
	Reflection	

Aim and Data

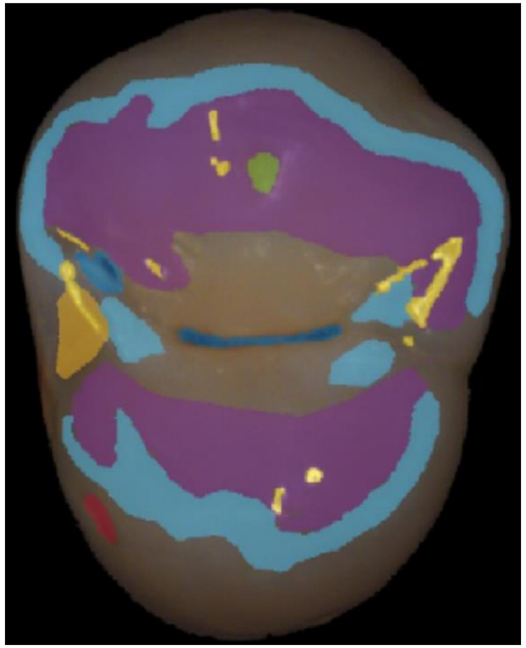
Given a tooth image, we apply regression-based machine learning methods to estimate the **probability** and **severity** of a hard-tissue condition on the tooth image.



Data Preparation

1. Extract the proportions of Caries, ETW, and Fluorosis out of the dental image using the labeled dataset.

2. For each different category, we separate the dataset into case and control datasets.



	A	B	C	D
1	Name	Caries	ETW	Fluorosis
2	100_ETW_904-A.png	0.002005683	0.077111138	0
3	101_ETW_907-A.png	0.00398403	0.123345976	0
4	102_ETW_911-A.png	0.002348158	0.161229451	0
5	103_ETW_919-A.png	0	0.113853905	0
6	104_ETW_928-A.png	0.003160265	0.108988444	0
7	105_ETW_976-A.png	0.002756755	0.085084703	0
8	106_ETW_994-A.png	0.002951728	0.10716078	0
9	107_fluorosis_1030-A.png	0	0	0.243492554
10	108_fluorosis_1077-A.png	0.002697168	0	0.299936925
11	109_fluorosis_108-A.png	0	0	0.235179527
12	10_caries_1315-A.png	0.010821386	0	0
13	110_fluorosis_1383-A.png	0.00348349	0	0.251692283
14	111_fluorosis_1416-A.png	0.004132839	0	0.238091453
15	112_fluorosis_1438-A.png	0.001535441	0	0.241875944
16	113_fluorosis_1466-A.png	0	0	0.216644287
17	114_fluorosis_249-A.png	0.001878717	0	0.237193525
18	115_fluorosis_275-A.png	0.001305474	0	0.257310655
19	116_fluorosis_307-A.png	0	0	0.203521729
20	117_fluorosis_309-A.png	0.000473484	0	0.179147763
21	118_fluorosis_328-A.png	0.014872233	0	0.303166707
22	119_fluorosis_370-A.png	0	0	0.304509428
23	11_caries_1341-A.png	0.007173283	0	0.027399507

Data Preparation

1. Extract the proportions of Caries, ETW, and Fluorosis out of the dental image using the labeled dataset.

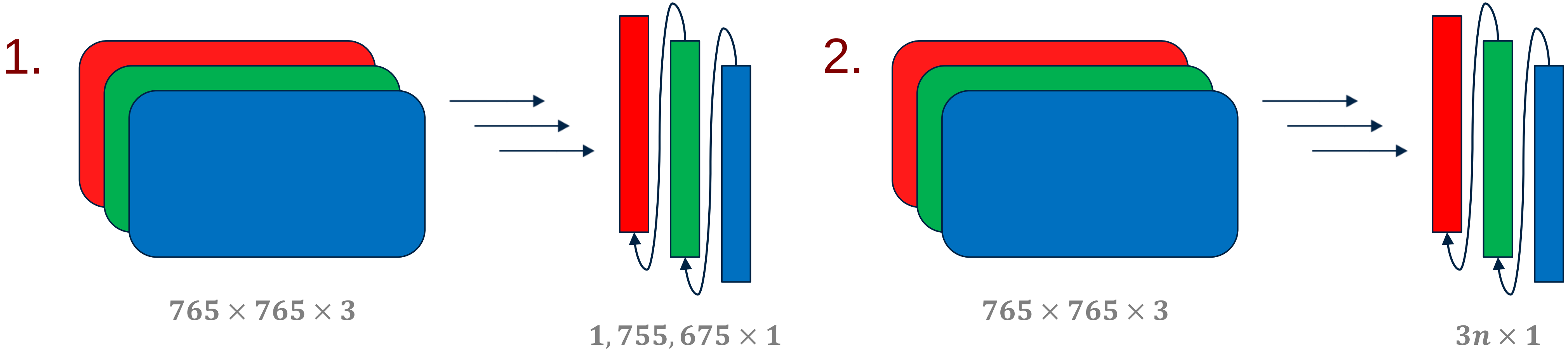
2. For each different category, we separate the dataset into case and control datasets.

Category	Cases	Control	Total
Caries	136	26	162
ETW	76	86	162
Fluorosis	66	96	162

Feature Representation

1. **Color Vector of the Whole Image**: We convert every color image into a single vector to keep all the original color channel values across the entire image.

2. **Distribution Vector with A Given Number of Bins**: The distribution vector is computed by concatenating the histogram of each channel with n bins into a single vector.

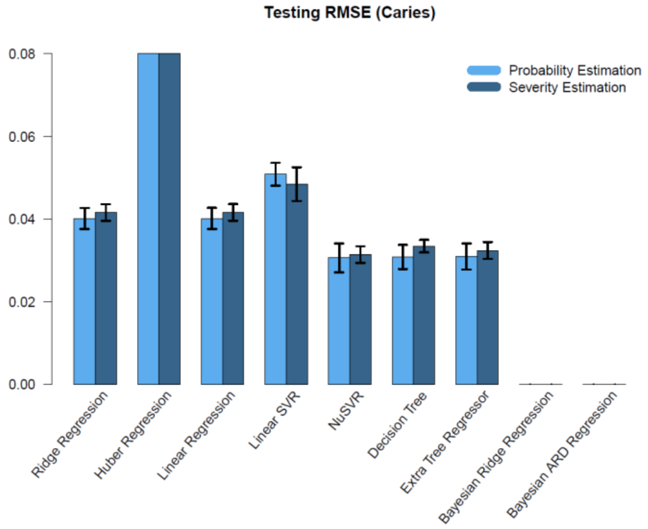


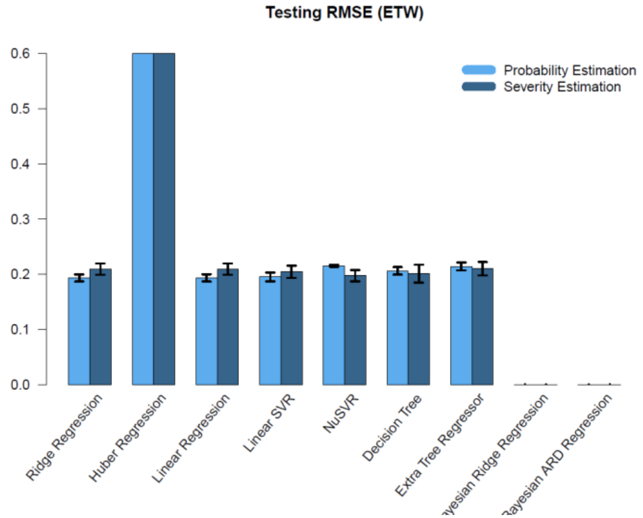
Machine Learning Methods

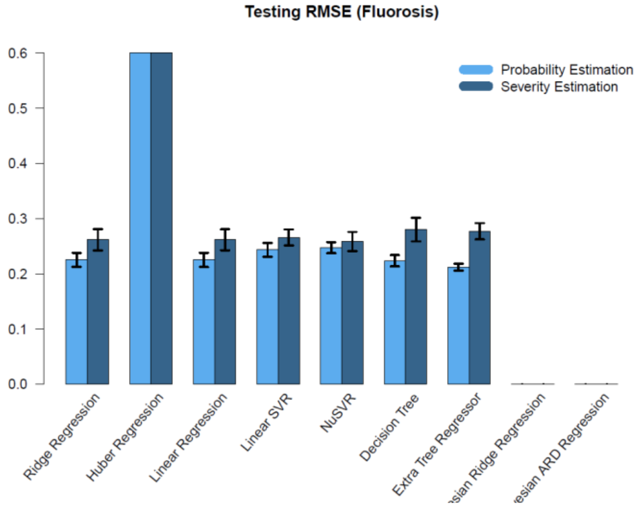
1. **Regression Based Methods**
   - Linear Regression
   - Ridge Regression
   - Huber Regression
2. **Bayesian Based Regression Methods**
   - Bayesian Ridge Regression
   - Bayesian Automatic Relevance Determination (ARD) Regression
3. **Support Vector Based Regression Methods**
   - Linear Support Vector Regression (Linear SVR)
   - Nu Support Vector Regression (NuSVR)
4. **Tree Based Regression Methods**
   - Decision Tree
   - Extra Trees Regression

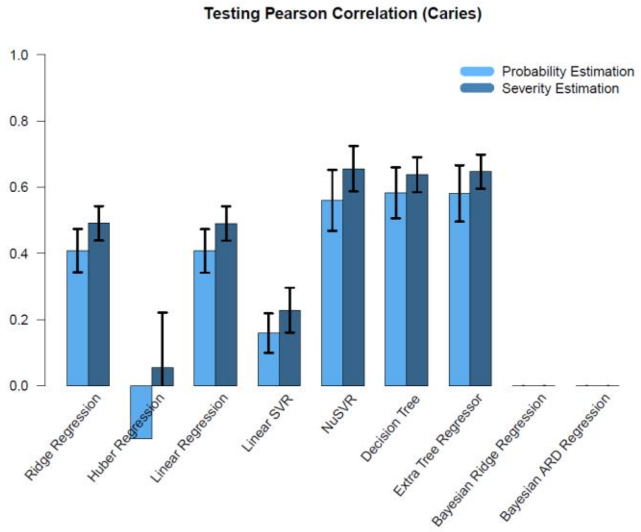
Results

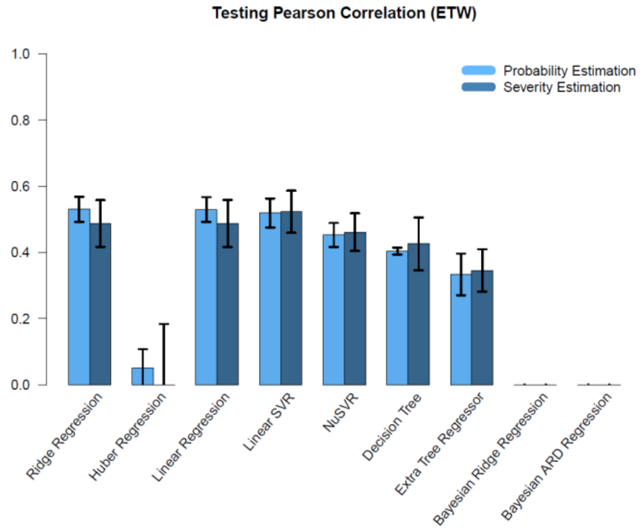
1. Results of Whole Image Color Vector Feature Representation

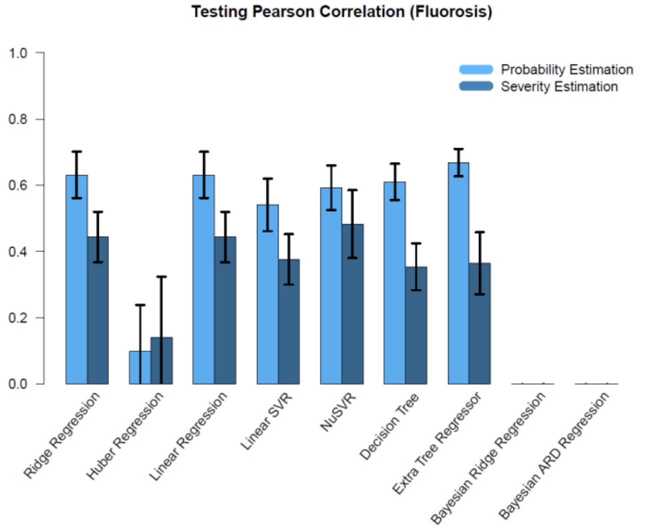






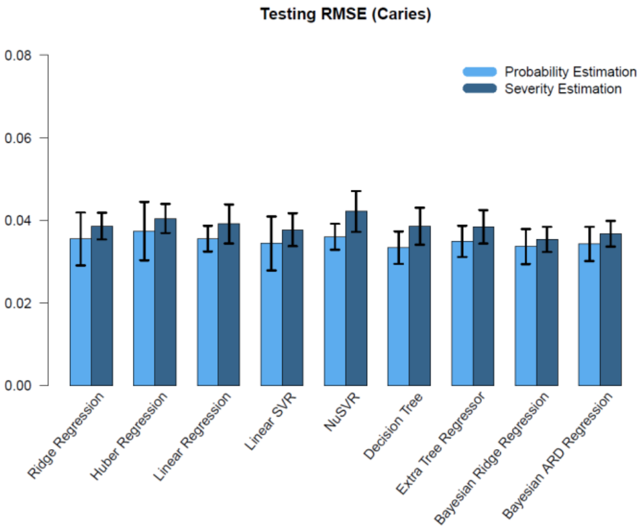


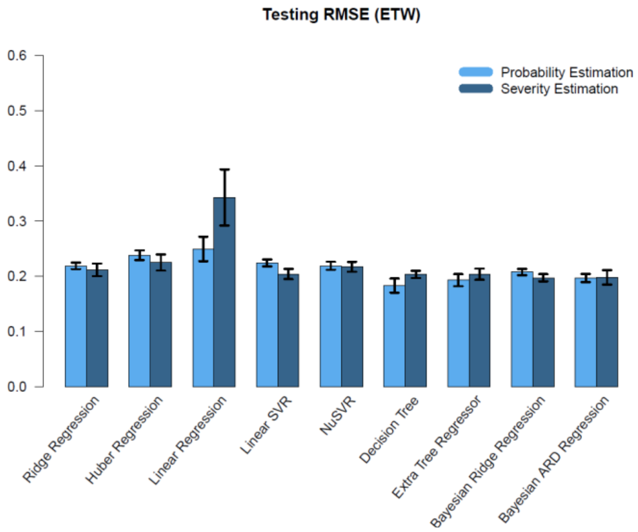


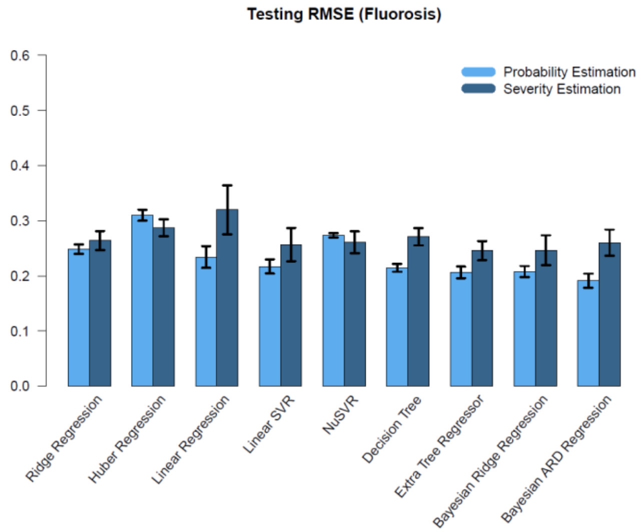


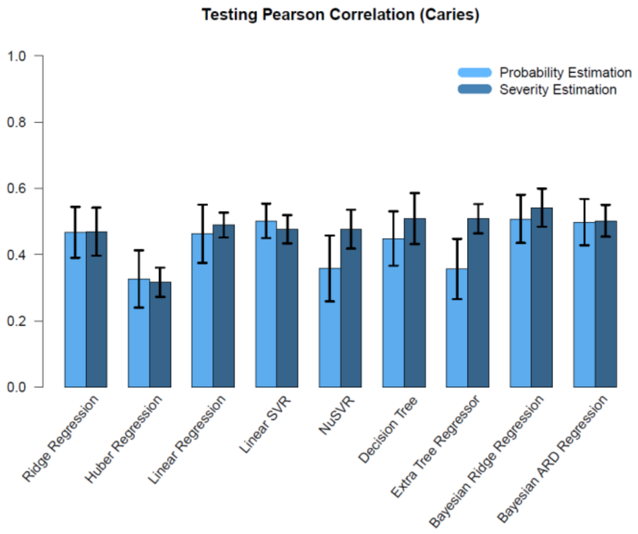
Results

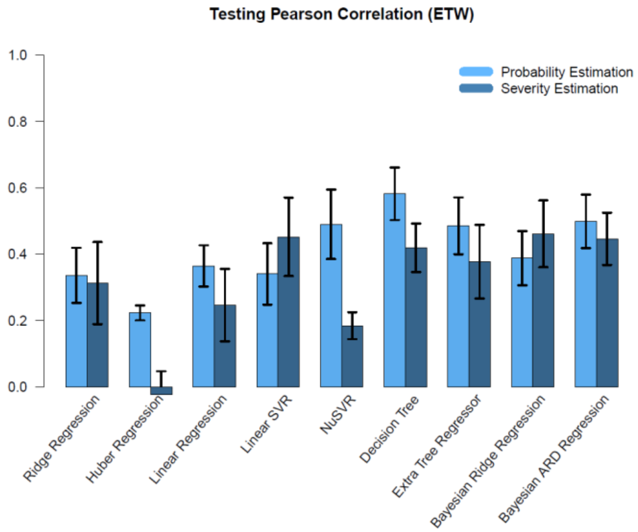
2. Results of Ten-Bin Distribution Vector Feature Representation

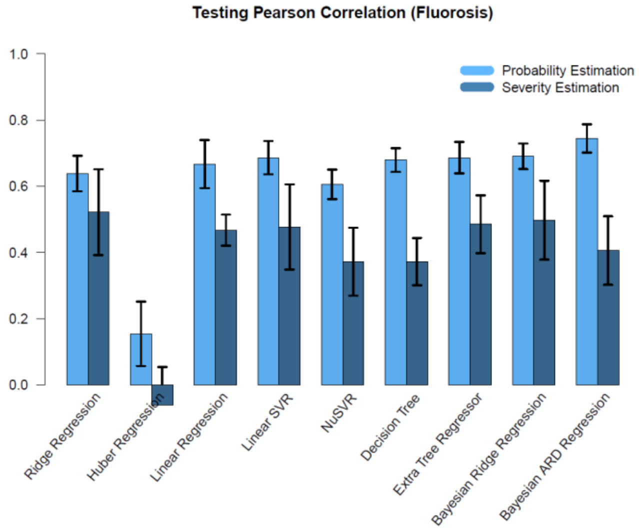






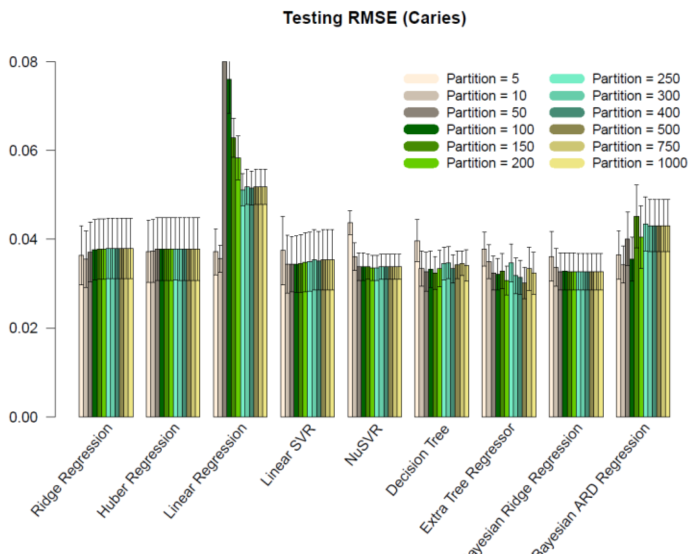


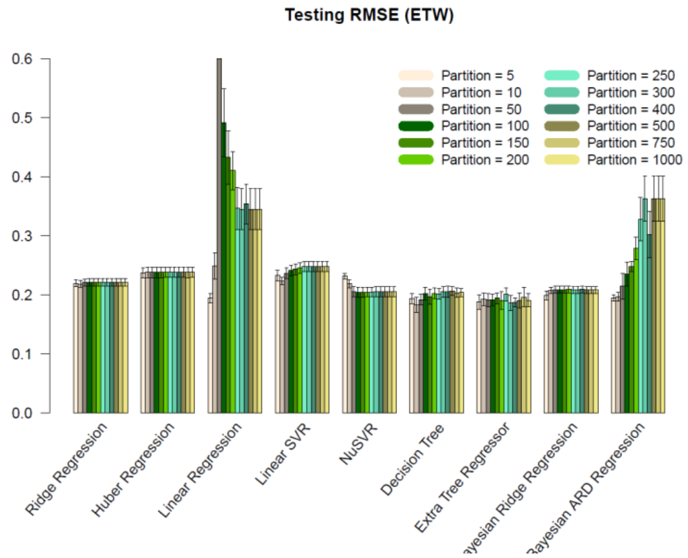


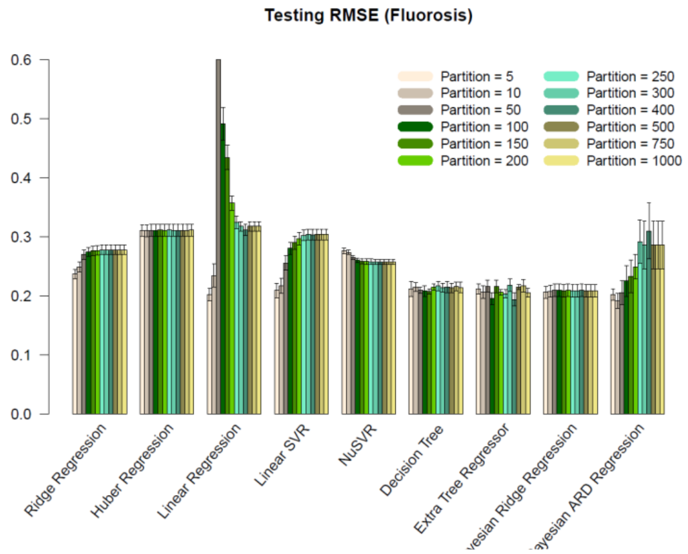


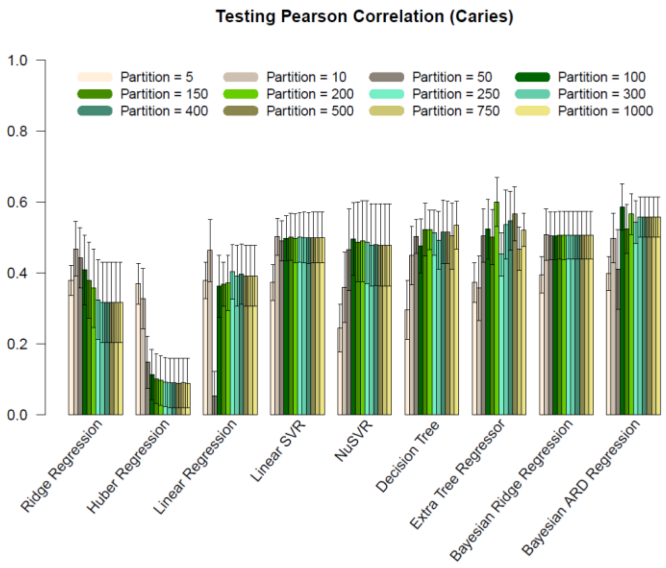
Results

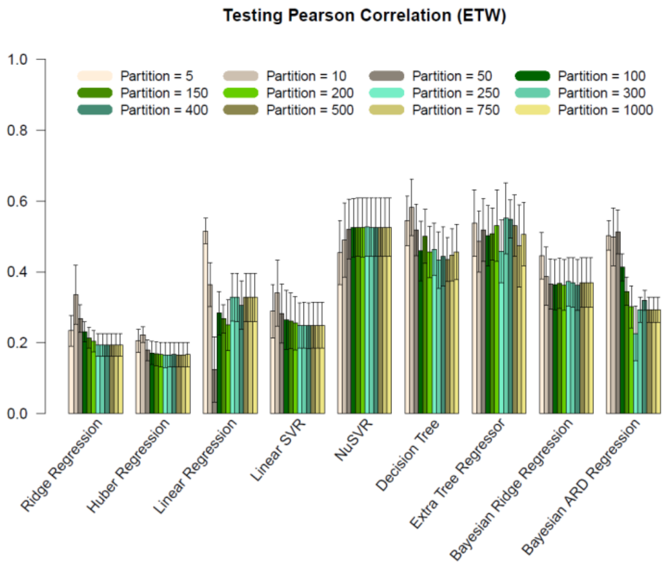
3. Results of Distribution Vector Feature Representation with Various Bin Numbers

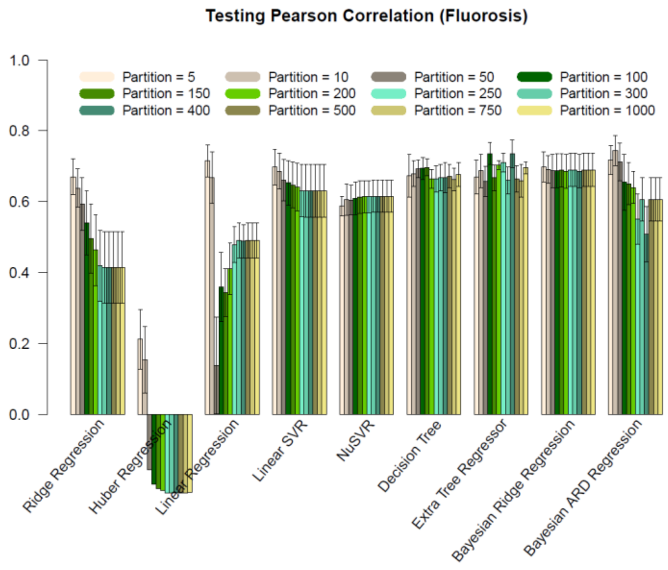






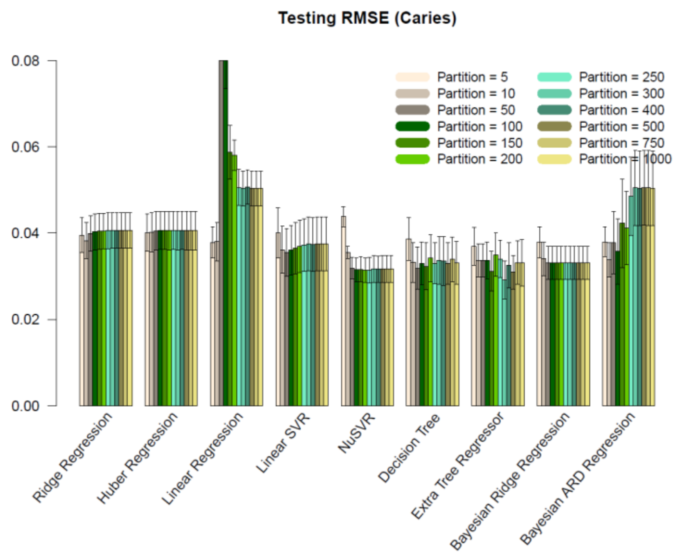


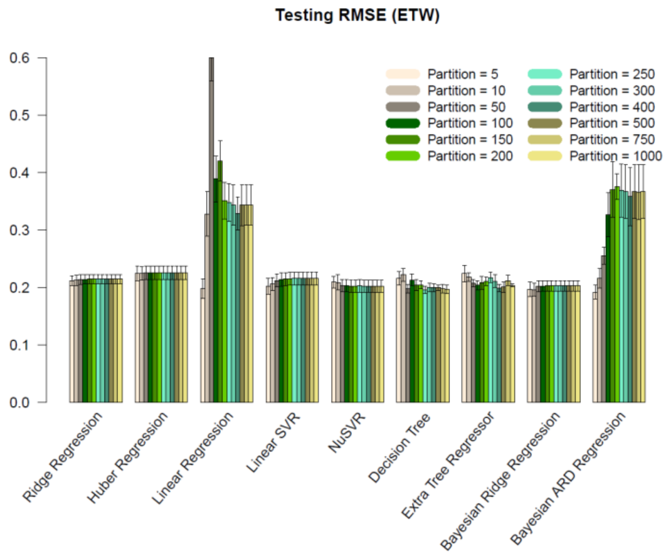


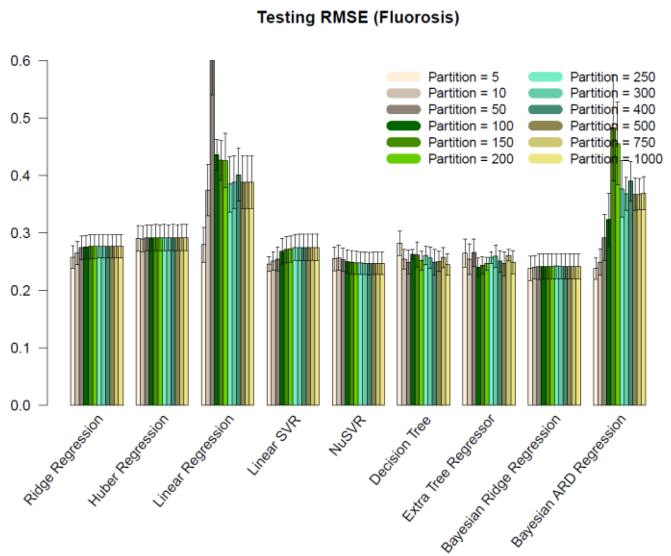


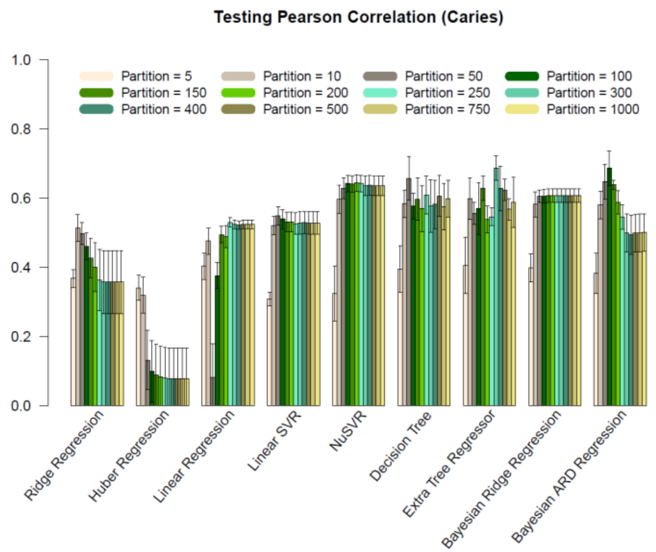
Results

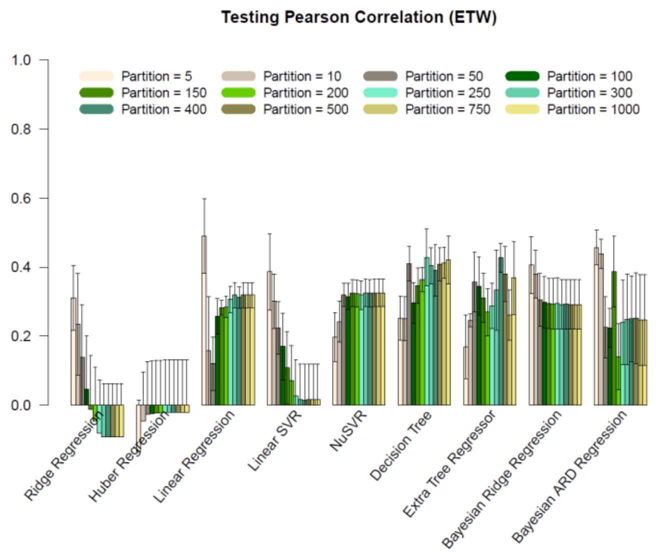
3. Results of Distribution Vector Feature Representation with Various Bin Numbers

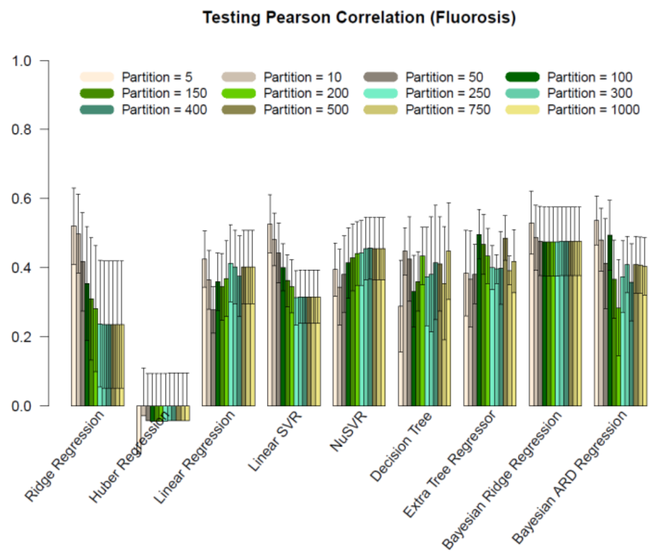












Results

3. Results of Distribution Vector Feature Representation with Various Bin Numbers

Outcome to Estimate	Category	Range of Outcome	Root Mean Square Error (RMSE)			Pearson Correlation		
			Method	Bins	Value	Method	Bins	Value
Probability	Caries	0.030±0.040	Extra Trees Regression	500	0.030	Extra Trees Regression	200	0.600
	ETW	0.158±0.223	Decision Tree	10	0.183	Decision Tree	10	0.581
	Fluorosis	0.184±0.283	Bayesian ARD Regression	10	0.191	Bayesian ARD Regression	10	0.745
Severity	Caries	0.035±0.041	Extra Trees Regression	300	0.029	Extra Trees Regression	300	0.687
	ETW	0.337±0.215	Bayesian ARD Regression	5	0.192	Linear Regression	5	0.490
	Fluorosis	0.451±0.276	Bayesian ARD Regression	5	0.238	Bayesian ARD Regression	5	0.537

Table: Best Prediction Results for Probability and Severity Estimation.

Conclusion & Outlook

1. **Conclusion**
    - Most of the nine machine learning methods are not sensitive to the number of bins.
    - NuSVR, Decision Tree, Extra Trees Regression and Bayesian Ridge Regression show an overall stable and good performance for all three categories and two different estimation tasks.
    - The best probability and severity estimation results are concluded, indicating that machine learning models provide promising opportunities to help clinical evaluation and save resources in the management of these dental conditions.

2. **Outlook**
    - Implement Nested Cross Validation for tuning hyperparameters.
    - Obtain more data to improve the estimation accuracy.
    - Explore new feature representations.
    - Study semantic segmentation problem by segmenting our pixels with different hard tissue conditions.

Acknowledgment

I would like to acknowledge:

- Dr. Li Shen, Ph.D.
- Dr. Mansu Kim, Ph.D.
- Dr. Qing Sun, Ph.D.

Penn Medicine