Estimating Hard-tissue Conditions from Dental Images via Machine Learning

Jingxuan Bao*, Mansu Kim†, Qing Sun‡, Anderson T. Hara§, Gerardo Maupome¶ and Li Shen†‖
*School of Arts and Sciences, University of Pennsylvania, Philadelphia, USA
†Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
‡Keck School of Medicine, University of Southern California, Los Angeles, USA
§School of Dentistry, Indiana University, Indianapolis, USA
¶Fairbanks School of Public Health, Indiana University, Indianapolis, USA

Email: *bao96@sas.upenn.edu, †mansu.kim@pennmedicine.upenn.edu, ‡qingsun@usc.edu,
§ahara@iu.edu, ¶gmaupome@iu.edu, ‖li.shen@pennmedicine.upenn.edu

***Abstract*—Despite the great success of machine learning in various biomedical domains, applications to dental hard tissue conditions (primarily on dental Caries, Erosive Tooth Wear (ETW), and Fluorosis) are under-explored, in particular for analyzing photographic images. The clinical diagnostics of these dental hard-tissue conditions is routinely performed by visual examination but is often limited by its subjectivity. To bridge this gap, we apply four categories of machine learning strategies including nine different methods with two different feature representations to estimate the probability and severity of dental hard-tissue conditions from photographic tooth images. Our first empirical study is performed on the real dataset containing both controls and cases, and the best probability estimation results are achieved by Extra Trees Regression (RMSE: 0.030, Pearson correlation: 0.600) for Caries, Decision Tree (RMSE: 0.183, Pearson correlation: 0.581) for ETW, and Bayesian ARD Regression (RMSE: 0.191, Pearson correlation: 0.745) for Fluorosis. Our second empirical study is performed on the case only datasets, and the best severity estimation results are achieved by Extra Trees Regression (RMSE: 0.029, Pearson correlation: 0.687) for Caries, Bayesian ARD Regression and Linear Regression (RMSE: 0.192, Pearson correlation: 0.490) for ETW, and Bayesian ARD Regression (RMSE: 0.238, Pearson correlation: 0.537) for Fluorosis.**

These results indicate that machine learning models provide promising opportunities to help clinical evaluation and save resources in the management of these dental conditions.

***Keywords*-photographic dental imaging; computer-aided estimation and diagnosis; machine learning; dental hard-tissue conditions**

I. Introduction

Machine learning techniques have been successfully used for a variety of object detection and classification tasks in natural images [1] as well as computer-aided detection/diagnosis based on medical images [2]–[4]. Significant progress has been made in the deep learning field, where deep learning has quickly become a highly effective method for analyzing medical images in the study of various disorders including neuro, retinal, pulmonary, breast, cardiac, abdominal, and musculoskeletal conditions [3]. However, in some scenarios, it is noted that classical machine learning methods can still outperform deep networks, especially when the dataset is relatively small.

While machine learning and deep learning have been shown effective for various biomedical image recognition problems, their applications for analyzing photographic dental images are still underexplored. There exist prior studies that apply machine learning to the analysis of dental radiographical images [5], [6]. However, few studies using machine learning to analyze photographic dental images [7], [8]. Of note, the clinical diagnostics of the main dental hard-tissue pathologies is routinely performed by visual examination but is limited by its subjectivity.

To bridge this gap, in this paper, we apply nine machine learning methods to automatically analyze photographic images, aid the visual diagnosis of the probability and severity of the dental hard-tissue conditions. Specifically, given a tooth image, we apply regression-based machine learning methods to estimate *the proportion of a hard-tissue condition on the tooth image*. We call this proportion as *probability* if we don't know whether the tooth has the condition or not, and call it as *severity* if we know the tooth has the condition. Empirically, to estimate *probability*, we analyze all the images with or without a certain condition (i.e., including both cases and controls); to estimate *severity*, we analyze only images with the condition (i.e., including cases only). In this work, we focus on the analysis of three hard tissue conditions: (1) Caries, (2) Erosive Toot Wear (ETW) and (3) Fluorosis.

Moreover, we apply two feature representations to estimate the proportion of image pixels with the condition. Using a real dataset with 5-fold cross validation, we demonstrate the feature representations and methods with best performance. We also compare the different machine learning methods for probability and severity estimation tasks. Our promising empirical results suggest that probability and severity estimation with machine learning has the potential to assist dentists to detect the hard-tissue conditions using photographic dental images and improve clinical outcomes.

The rest of the paper is organized as follows. First, we discuss our data and data preprocessing in Section II. Next,

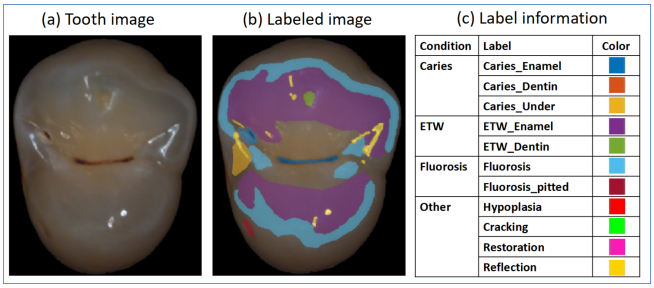
2471-7819/20/$31.00 ©2020 IEEE
DOI 10.1109/BIBE50027.2020.00058

we present feature representation and machine learning methods in Section III. After that, we show our empirical results in Section IV. Finally, we conclude the paper in Section V.

II. Materials

A. Data Description

We collected a dataset containing 162 digital photographs of the occlusal (top-view) surface of extracted human premolars, under identical magnification and lighting conditions. We developed an Image Labeler tool with the MATLAB software. Using this tool, an experienced and trained dental examiner screened, manually segmented and labeled all the images into 11 different descriptors shown in Fig. 1. These labeled images were used as our ground truth. In this work, we study three conditions, including Caries (containing three labels, Caries Enamel, Caries Dentin and Caries Under), Erosive Tooth Wear or ETW (containing two labels, ETW Enamel and ETW Dentin), and Fluorosis (containing two labels, Fluorosis and Fluorosis pitted).



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

Figure 1: **Example of Dental Image and Label.** (a) Example tooth image. (b) Tooth image color-labeled with various conditions. (c) Label information.

B. Data Preparation

To prepare the data for the machine learning analysis. We first extract the proportions of Caries, ETW, and Fluorosis out of the dental image using the labeled dataset. We calculate the proportions by the ratio between the number of pixels for each category and the total number of tooth pixels in each dental image. Under such circumstance, each labeled dental image will return three numbers between 0 and 1, which are the proportions of Caries, ETW and Fluorosis respectively. Note that the sum of the three numbers is not necessary to be 1 and it could be 0 if the teeth is healthy. For each different category, we separate the dataset into case and control datasets, where the control dataset is selected if the proportion of hard-tissue condition is approximately 0 ($< 10^{-5}$) and the rest of the data forms the case dataset.

As a result, our dental dataset can be divided into three comparisons, one for each condition: "Caries" (n=136) vs "Non-caries" (n=26), "ETW" (n=76) vs "Non-ETW" (n=86), and "Fluorosis" (n=66) vs "Non-fluorosis" (n=96). Next, since our image dataset contains dental images with various size, we then resize all images into 765×765 pixels RGB images. For each image, we extract the RGB channels and convert them into three color matrices, where each element among one of the three matrices is a number between 0 and 255, representing the scale value of corresponding color channel at the corresponding position of the image. Then, we extract the representative features of images using two different approaches: 1) stretching the image into a vector with all the image information preserved, and 2) calculating the distribution vector describing the histogram of color channel values. More details about feature representations will be discussed in Section III.

III. Methods

In this work, we apply nine widely used machine learning methods to estimate the severity and probability of dental hard tissue conditions. Our prediction models employ two different feature representations as the predictor and aim to estimate probability or severity score as a response. The performance of each model is evaluated by root-mean-square-error (RMSE) and the correlation between actual and predicted scores.

A. Feature Representation

We implement two different feature representations to train our model for machine learning methods.

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. To be specific, we vectorize the image ($765 \times 765 \times 3$) and concatenate into a single vector ($1{,}755{,}675 \times 1$). We normalize the image by dividing each element of the vector by 255. For each image, the feature is a color vector with length 1,755,675.

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. In other words, we divide the color channel values from 0 to 255 into n equal length intervals, and then count the number of values in each interval. For each color channel, we are able to obtain a length n vector. Next we concatenate all three color channel vectors (3 vectors of length n) into a single vector of length $3n$ to be our feature vector.

B. Machine Learning Methods

We implement nine different machine learning methods in total to estimate the proportion of image pixels with different dental hard-tissue conditions using "color vector" or "distribution vector" as the predictor. The results are evaluated by root mean square error (RMSE) and Pearson correlation using stratified 5-fold cross validation. The nine machine learning methods are described below.

1) Regression Based Method: In regression based method, we implement Linear Regression, Ridge Regression, and Huber Regression. For Linear Regression, we assume the ground truth label is a linear combination of our representative vector plus an error term following normal distribution, and we estimate the parameters by minimizing the sum of square errors [9]. For Ridge Regression, we add an l_2 regularization term when minimizing the least square loss, which is helpful to mitigate the problem of multicolinearity [10] and improves efficiency in exchange for a tolerable amount of bias when estimating parameters [11]. For Huber Regression, we estimate the parameters using Huber loss instead of square loss, which is less sensitive to outliers than the square loss [12].

2) Bayesian Based Regression Methods: In Bayesian based regression method, we implement Bayesian Ridge Regression and Bayesian Automatic Relevance Determination (ARD) Regression. Bayesian Ridge Regression follows the same setup as the Ridge Regression in previous paragraph, but we use Bayesian approach to make inference about the parameters. We estimate the parameters by assuming spherical Gaussian prior for the parameters, which adapts to the data at hand [13]. For Bayesian ARD Regression, it poses a different prior over parameters, which is an axis-parallel, elliptical Gaussian distribution [14]. With such a change in prior distribution, the estimated parameters will become sparser than the Bayesian Ridge Regression [13], [15].

3) Support Vector Based Regression Method: In support vector based regression method, we implement Linear SVR and NuSVR. Linear SVR method trains a supervised-learning model by using a symmetrical loss function, which penalizes high and low misestimates equally, leading to the independence between computational complexity and the dimensionality of the input space by the application of Vapnik's ϵ-insensitive approach [16]. Moreover, it only considers the linear kernel, which provides a faster implementation than SVR [14]. For NuSVR, we introduce a new parameter ν which controls the number of support vectors and margin errors [14].

4) Tree Based Regression Method: In tree based regression method class, we implement Decision Tree and Extra Trees Regression. Decision Tree is a non-parametric supervised learning method aimimg at creating a predictive model to estimate the value of a target variable by learning decision rules from the data features, which performs well even if assumptions are not fully satisfied by the true model where the data were generated [14]. Extra Trees Regression is an algorithm specifically designed for trees, which implements a meta estimator that fits a number of randomized decision trees on a range of subsamples of training dataset. It can improve the estimation accuracy and control overfitting by using average [14].

C. Five-Fold Cross Validation

The performance of the machine learning methods are evaluated using five-fold cross validation. For probability estimation task, we use case and control dataset by first separating the case dataset and control dataset of images and their corresponding labels into 5 folds. Then we iteratively select training set and test set. For each time, we select 4 of the subsets for both case and control dataset as training set, and the other one to be the test set. For severity estimation task, we only use case dataset by first separating the case dataset into three equal spaces according to the values in the label set, where those with the label value at the range of the highest one third proportion are labeled as Group 1, and similarly the middle one third as Group 2, and the lowest one third as Group 3. We separate each of the group into 5 folds and iteratively select 4 of them combining together to be our training set, and the other one combining together to be our test set. We train nine machine learning models using the training set and evaluate the performance using the test set.

IV. Results

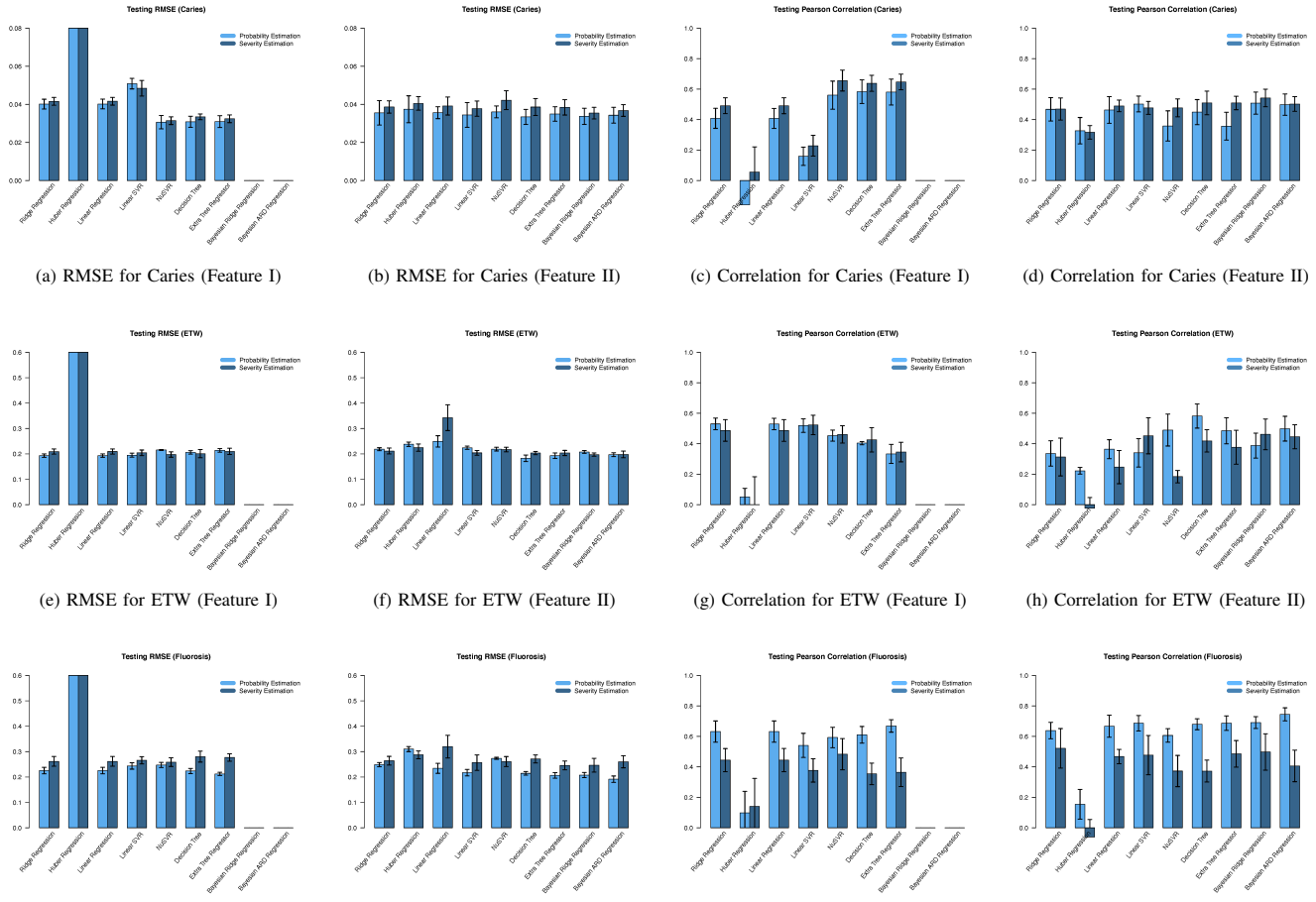
In this section, we present the probability and severity estimation results obtained from nine machine learning methods and two feature representations. The probability results are estimated by using both the case and control data; and severity results are estimated using only the case data.

1) Results of Whole Image Color Vector Feature Representation: We first show the results using the whole image color vector feature representation. We input the whole image color vector together with the outcome value (i.e., proportion of pixels with the condition) for each condition and estimate the predicted probability and severity. We evaluate the performance by calculating the RMSE and Pearson correlation between the estimated and actual outcome values. All the probability and severity estimation results are shown in Fig. 2 (see **Feature I** panels).

The best probability estimation results in terms of Pearson correlation are obtained by Decision Tree (Caries: 0.796), Ridge Regression (ETW: 0.631) and Extra Trees Regression (Fluorosis: 0.771). The best severity estimation results in terms of Pearson correlation are obtained by NuSVR (Caries: 0.842; Fluorosis: 0.688) and Linear SVR (ETW: 0.665). Fig. 3 shows detailed scatter plots to visualize the individualized prediction performance for the test subjects for the above six experiments. In each panel, predicted outcome values are plotted against actual values for testing subjects only.

2) Results of Ten-Bin Distribution Vector Feature Representation: After examining the results using the original color image representation, we further evaluate the results using representative feature vector with ten bin size distribution. All the probability and severity estimation results are shown in Fig. 2 (see **Feature II** panels).

The best probability estimation performance results in terms of Pearson correlation are obtained by Bayesian Ridge



(a) RMSE for Caries (Feature I) (b) RMSE for Caries (Feature II) (c) Correlation for Caries (Feature I) (d) Correlation for Caries (Feature II)

(e) RMSE for ETW (Feature I) (f) RMSE for ETW (Feature II) (g) Correlation for ETW (Feature I) (h) Correlation for ETW (Feature II)

(i) RMSE for Fluorosis (Feature I) (j) RMSE for Fluorosis (Feature II) (k) Correlation for Fluorosis (Feature I) (l) Correlation for Fluorosis (Feature II)

Figure 2: **Five-fold Cross-validation Performance for Estimating Probability/Severity: Comparison between Feature I (Whole Image Color Vector Feature Representation) versus Feature II (Ten-bin Distribution Vector Feature Representation).** In each bar plot, the bar height and error bar indicate the mean and standard error of five testing performance measures respectively. The prediction performance is measured by (1) root mean square error (RMSE) and (2) Pearson correlation between predicted and actual outcome values. (a) - (d) show the results for estimating probability and severity of Caries using nine regression methods and two different feature representations. (e) - (h) show the results for estimating probability and severity of ETW using nine regression methods and two different feature representations. (i) - (l) show the results for estimating probability and severity of Fluorosis using nine regression methods and two different feature representations. Of note, the Bayesian based methods when using the whole image color vector feature representation require huge amount of memory whose results cannot be calculated and thus not shown here.

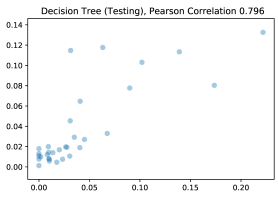
Regression (Caries: 0.701), Decision Tree (ETW: 0.762) and Bayesian ARD Regression (Fluorosis: 0.838). The best severity estimation performance results in terms of Pearson correlation are obtained by Bayesian Ridge Regression (Caries: 0.741; ETW: 0.681) and Ridge Regression (Fluorosis: 0.761). Fig. 4 shows detailed scatter plots to visualize the individualized prediction performance for the test subjects for the above six experiments. In each panel, predicted outcome values are plotted against actual values for testing subjects only.

3) Results of Distribution Vector Feature Representation with Various Bin Numbers: Since different bin numbers yield different distribution vector feature representations, we further vary the bin number when calculating the distribution vector for each image to examine its predictive power. Specifically, we set the bin number to be 5, 10, 50, 100, 150, 200, 250, 300, 400, 500, 750, or 1000. The probability and severity results are summarized in Fig. 5 and Fig. 6 respectively. Moreover, we summarize the best probability and severity estimation results for the three different dental hard tissue conditions in Table I.

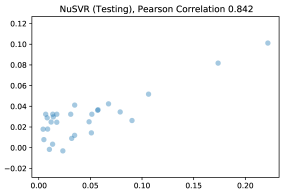
According to Fig. 5, the probability estimation results show that nine machine learning methods are generally not sensitive to the bin number except Linear Regression and

Table I: **Best Prediction Results for Probability and Severity Estimation.** The best results are selected according to the mean RMSE or mean Pearson correlation from five cross validation trials. The outcome range is shown as the mean ± standard deviation. All results are rounded by keeping three decimal places.

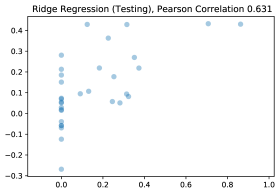
Outcome to Estimate	Category	Range of Outcome	Root Mean Square Error (RMSE)			Pearson Correlation		
			Method	Bins	Value	Method	Bins	Value
Probability	Caries	0.030±0.040	Extra Trees Regression	500	0.030	Extra Trees Regression	200	0.600
	ETW	0.158±0.223	Decision Tree	10	0.183	Decision Tree	10	0.581
	Fluorosis	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



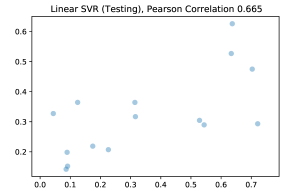
(a) Decision Tree (Caries)



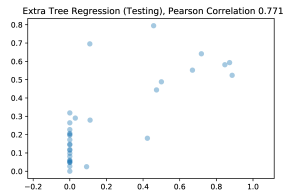
(b) NuSVR (Caries)



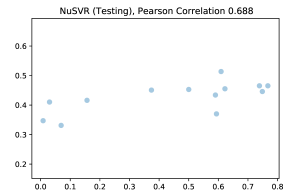
(c) Ridge Regression (ETW)



(d) Linear SVR (ETW)

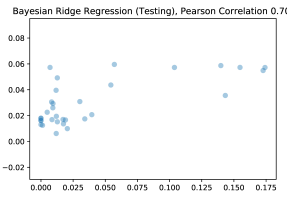


(e) Extra Trees Regression (Fluorosis)

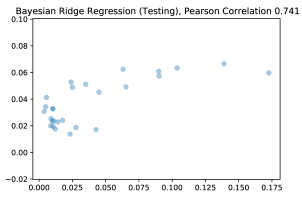


(f) NuSVR (Fluorosis)

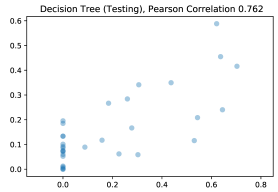
Figure 3: **Best Prediction Results using the Whole Color Vector Feature Representation.** Predicted outcome values are plotted against actual values for testing subjects only.



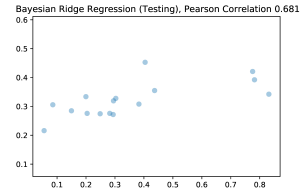
(a) Bayesian Ridge Regression (Caries)



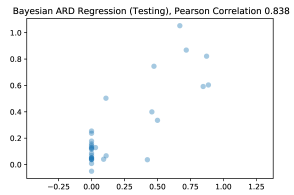
(b) Bayesian Ridge Regression (Caries)



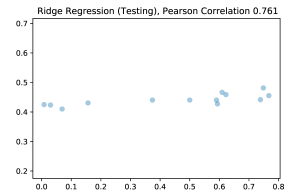
(c) Decision Tree (ETW)



(d) Bayesian Ridge Regression (ETW)



(e) Bayesian ARD Regression (Fluorosis)



(f) Ridge Regression (Fluorosis)

Figure 4: **Best Prediction Results using the Ten-Bin Distribution Vector Feature Representation.** Predicted outcome values are plotted against actual values for testing subjects.

Bayesian ARD Regression for all three hard tissue condition datasets. For Fluorosis dataset, besides Linear Regression and Bayesian ARD Regression, Linear SVR is slightly unstable with respect to the bin number for distribution vector feature representation. Furthermore, Fig. 6 also shows that Linear Regression and Bayesian ARD Regression are more sensitive to the bin number for Fluorosis dataset. Compared to probability estimation results, the estimation performance of Caries is similar between two estimation tasks. However, the severity estimation task shows a worse performance than probability estimation task for ETW and Fluorosis datasets in terms of Pearson correlation measurement.

In summary, most of the nine machine learning methods are not sensitive to the bin number. Among those nine machine learning methods, NuSVR, Decision Tree, Extra Trees Regression and Bayesian Ridge Regression show an overall stable and good performance for all three categories and two different tasks. The comparison between the two estimation tasks show that the machine learning methods perform equally well in terms of RMSE and for Caries in terms of Pearson correlation. However, the probability estimation results outperform the severity estimation results

in terms of Pearson correlation for ETW and Fluorosis.

V. Conclusion and Outlook

A. Conclusion

In this paper, we evaluated the performances of multiple machine learning methods on estimating the probability and severity of dental hard tissue conditions (primarily on dental Caries, Erosive Tooth Wear, and Fluorosis) using photographic tooth images. We applied and compared nine regression methods for estimating dental hard tissue conditions: Linear Regression, Ridge Regression, Huber Regression; Bayesian Ridge Regression, Bayesian ARD Regression; SVR, Linear SVR, NuSVR; Decision Tree, and Extra Tree Regression. Moreover, we employed and compared two different feature representations for each tooth image, including the whole image color vector feature representation and distribution vector feature representation. We also did comparative studies through varying the bin number when calculating the distribution vector to examine whether the prediction result is sensitive to the bin number. We estimated the prediction performance by RMSE and Pearson correlation between predicted and actual outcome values, using five-fold cross-validation.

The probabilities and severity of dental photographic images of the three conditions are examined. Results show that the whole image color vector feature representation tends to cause overfitting. The best probability and severity estimation results are introduced respectively. Overall, most of the nine machine learning methods are not sensitive to the bin number while using the distribution vector feature representation. Among those nine machine learning methods, NuSVR, Decision Tree, Extra Trees Regression and Bayesian Ridge Regression show a generally stable and good performance for all three dental conditions and both probability and severity estimation tasks.

B. Outlook

In this paper, using dental photographic images, we have presented the results to estimate the probability and severity of three dental hard tissue conditions with nine popular machine learning methods. However, for some of these methods such as Ridge Regression and NuSVR, the machine learning model contains hyperparameters. We applied these methods by using the default setting of the hyperparameters. One interesting future direction could be to use nested cross-validation for tuning hyperparameters, which may yield improved prediction results. Moreover, the size of our current dataset is relatively small, which may underestimate the Pearson correlation measurement, in particular for the severity estimation task that analyzes only case images. The estimation results could be strengthen by a larger dataset when available in the future. Other interesting future topics include: (1) to explore new feature representations, (2) to examine advanced deep learning methods, and (3) to study a relevant but more challenging semantic segmentation problem by segmenting out pixels with different hard-tissue conditions on the photographic images.

References

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh *et al.*, "Imagenet large scale visual recognition challenge," *Int. Journal of Computer Vision*, vol. 115, 2015.

[2] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.

[4] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018.

[5] C.-W. Wang, C.-T. Huang *et al.*, "A benchmark for comparison of dental radiography analysis algorithms," *Medical Image Analysis*, vol. 31, pp. 63 – 76, 2016.

[6] A. Katsumata and H. Fujita, "Progress of computer-aided detection/diagnosis (cad) in dentistry," *Japanese Dental Science Review*, vol. 50, no. 3, pp. 63 – 68, 2014.

[7] E. D. Berdouses, G. D. Koutsouri *et al.*, "A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images," *Computers in Biology and Medicine*, vol. 62, pp. 119 – 135, 2015.

[8] L. Ghaedi, R. Gottlieb, D. C. Sarrett, A. Ismail, A. Belle, K. Najarian, and R. H. Hargraves, "An automated dental caries detection and scoring system for optical images of tooth occlusal surface," in *Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 1925–1928.

[9] D. Freedman, *Statistical Models: Theory and Practice*, 2005.

[10] P. Kennedy, *A Guide to Econometrics*. Cambridge, Massachusetts: The MIT Press, 2003.

[11] M. H. and J. Gruber, *Improving Efficiency by Shrinkage: The James–Stein and Ridge Regression Estimators*. New York: Marcel Dekker, 1998.

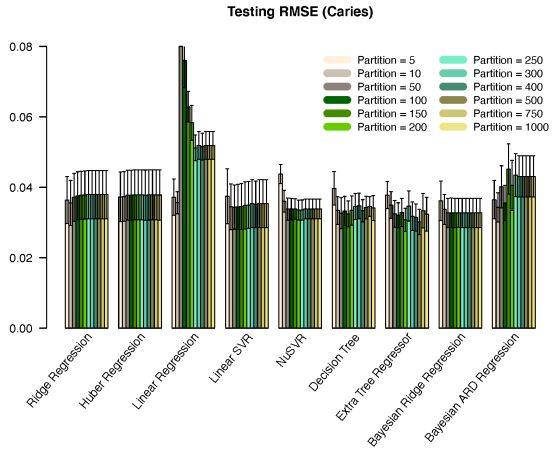
[12] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 03 1964. [Online]. Available: https://doi.org/10.1214/aoms/1177703732

[13] M. Jordan, J. Kleinberg, and B. Scho¨lkopf, *Pattern Recognition and Machine learning*. New York: Springer, 2006.

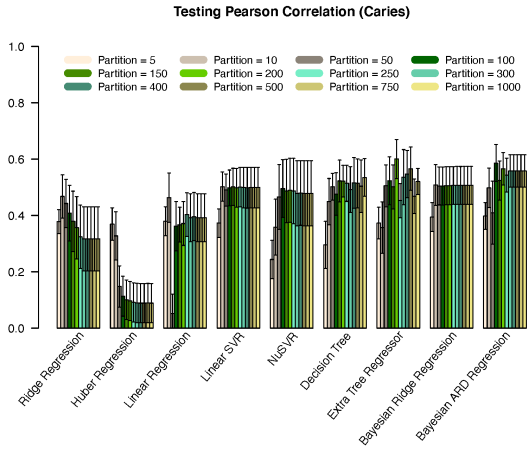
[14] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination," in *Advances in Neural Information Processing Systems 20*, 2008, pp. 1625–1632.

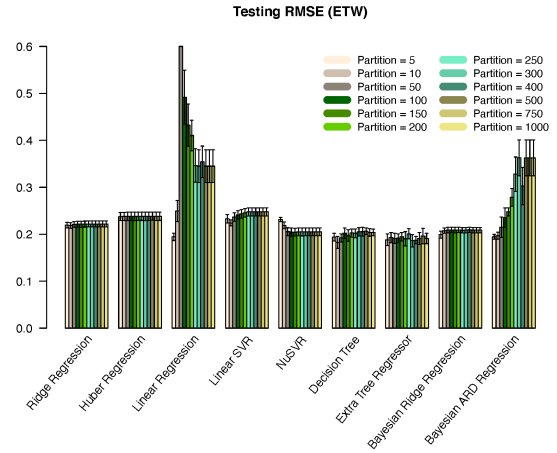
[16] A. M. and K. R., *Efficient Learning Machines*. Berkeley, CA: Apress, 2015.



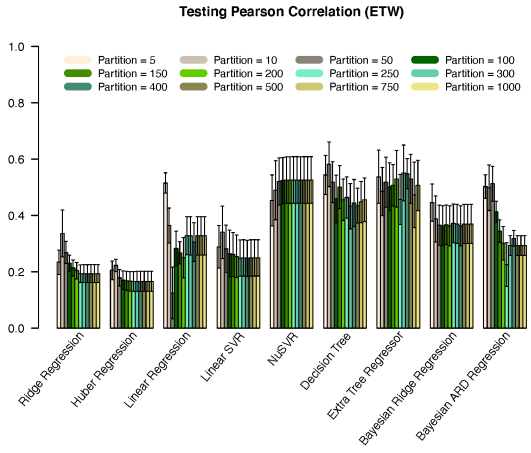
(a) Testing RMSE for Caries



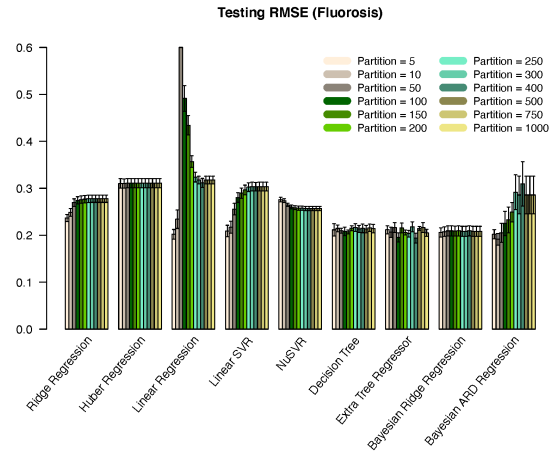
(b) Testing Pearson Correlation for Caries



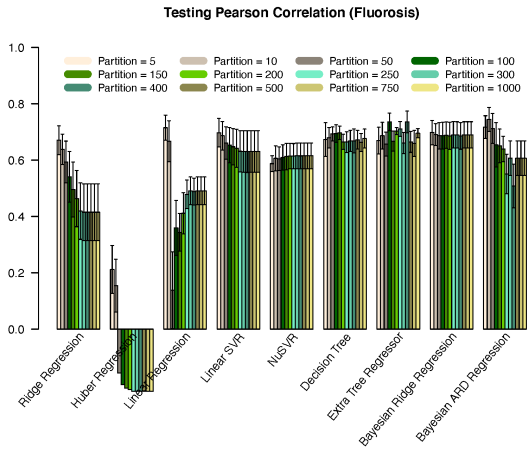
(c) Testing RMSE for ETW



(d) Testing Pearson Correlation for ETW

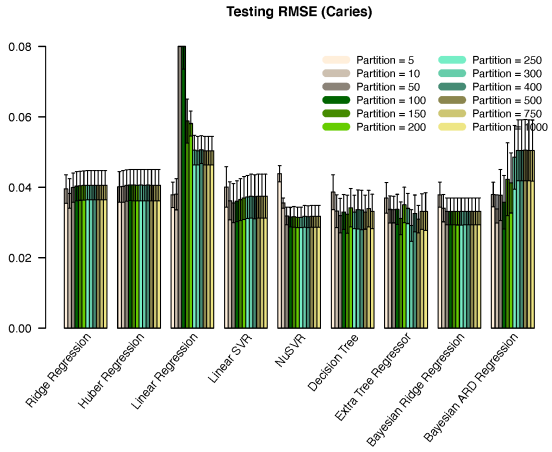


(e) Testing RMSE for Fluorosis

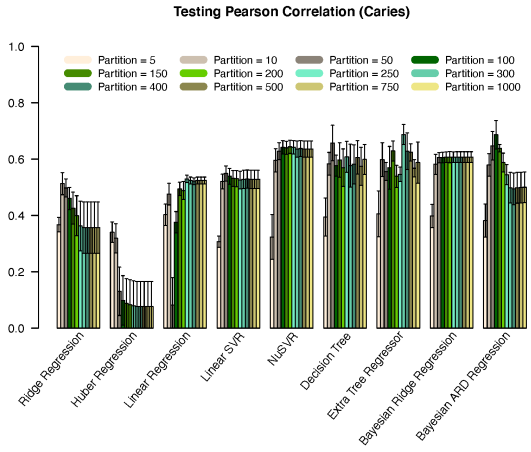


(f) Testing Pearson Correlation for Fluorosis

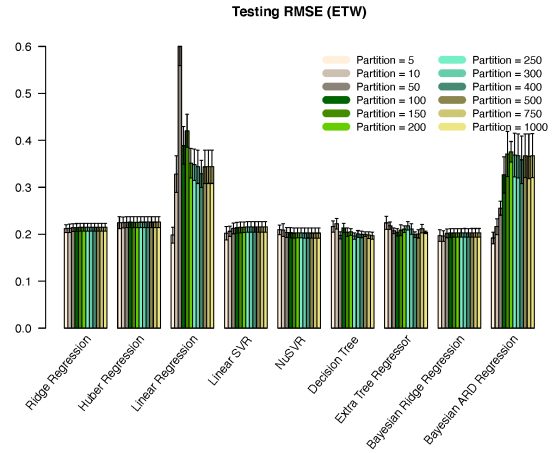
Figure 5: **Five-fold Cross-validation Performance for Estimating Probability: Comparison using Distribution Vector Feature Representation with Varying Number of Bins (i.e., Partitions)**. In each bar plot, the bar height and error bar indicate the mean and standard error of five testing performance measures respectively. The prediction performance is measured by RMSE and Pearson correlation between predicted and actual outcome values.



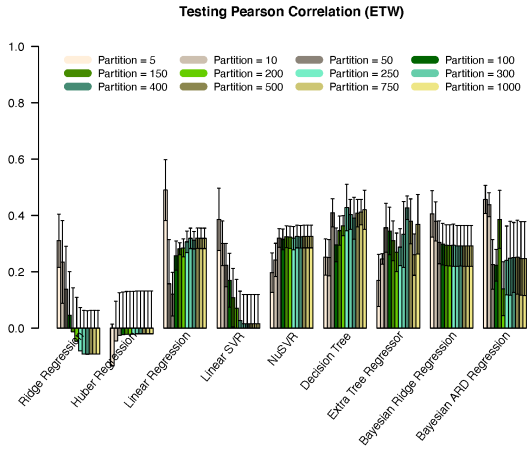
(a) Testing RMSE for Caries



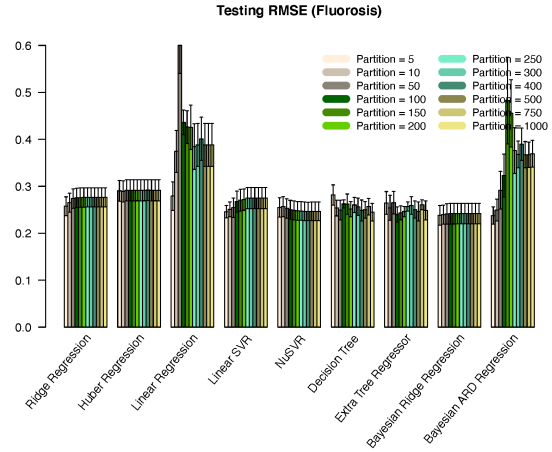
(b) Testing Pearson Correlation for Caries



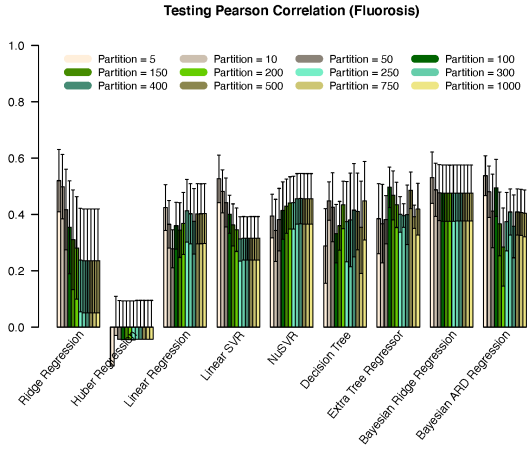
(c) Testing RMSE for ETW



(d) Testing Pearson Correlation for ETW



(e) Testing RMSE for Fluorosis



(f) Testing Pearson Correlation for Fluorosis

Figure 6: **Five-fold Cross-validation Performance for Estimating Severity: Comparison using Distribution Vector Feature Representation with Varying Number of Bins (i.e., Partitions)**. In each bar plot, the bar height and error bar indicate the mean and standard error of five testing performance measures respectively. The prediction performance is measured by RMSE and Pearson correlation between predicted and actual outcome values.