2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Interpretable Machine Learning Model to Classify Coronary Bifurcation Lesions

Xiaoqian Liu[1], Madhurima Vardhan[2], Qinrou Wen[3], Arpita Das[2], Amanda Randles[2], and Eric C. Chi[1]

Abstract— **Coronary bifurcation lesions are a leading cause of Coronary Artery Disease (CAD). Despite its prevalence, coronary bifurcation lesions remain difficult to treat due to our incomplete understanding of how various features of lesion anatomy synergistically disrupt normal hemodynamic flow. In this work, we employ an interpretable machine learning algorithm, the Classification and Regression Tree (CART), to model the impact of these geometric features on local hemodynamic quantities. We generate a synthetic arterial database via computational fluid dynamic simulations and apply the CART approach to predict the time averaged wall shear stress (TAWSS) at two different locations within the cardiac vasculature. Our experimental results show that CART can estimate a simple, interpretable, yet accurately predictive nonlinear model of TAWSS as a function of such features.**

Clinical relevance— **The fitted tree models have the potential to refine predictions of disturbed hemodynamic flow based on an individual's cardiac and lesion anatomy and consequently makes progress towards personalized treatment planning for CAD patients.**

I. INTRODUCTION

Bifurcation lesions are one of the most difficult types of coronary lesions to treat and are encountered in 15-20% of percutaneous coronary interventions (PCI) [8]. The anatomic complexity of bifurcation lesions non-trivially disrupts normal hemodynamics, strongly hindering the success of PCI [4]. Prior studies have established that poor PCI treatment outcomes, such as the risk of vessel occlusion, are associated with lesion features such as bifurcation angle, degree of stenosis and lesion length [2], [15], [16], [18]. Despite our knowledge of these associations, it remains unclear how these anatomic features jointly disrupt local hemodynamic quantities, which in turn can lead to development of atherosclerotic lesions [7]. Consequently, we aim to build an interpretable model to understand the impact of different lesion-specific parameters, such as bifurcation angle, lesion length and severity, on local hemodynamic flow.

We model the time averaged wall shear stress (TAWSS) as a representative metric of disturbed hemodynamic flow [7], but the modeling framework that we describe can be used to discover how lesion specific parameters influence other metrics of ischemic burden such as the fractional flow reserve (FFR), instantaneous wave free ratio (iFR) and resting gradient. TAWSS is the frictional force that acts in the tangential direction to blood flow [9], [7]. In this paper, we report on the effectiveness of the classification and regression tree (CART) [3] for fitting a nonlinear model of TAWSS as a function of patient and lesion specific parameters or features. CART is a class of non-parametric models, which split the feature space into regions where the metric-of-interest (TAWSS in this paper) is roughly constant. CART models have been shown to strike an attractive balance between prediction accuracy and straightforward interpretability of the resulting tree based rules in many clinical decision making problems [5], [13], [1], [12].

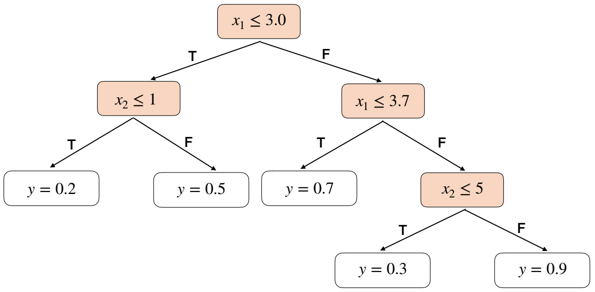

Fig. 1. An example regression tree with the maximal depth of three.

Figure 1 shows an example of a tree model estimated via CART for predicting a variable y using two feature variables x_1 and x_2. The estimated tree provides an easy to interpret prediction model. If an observation has features $x_1 \leq 3.0$ and $x_2 \leq 1$, then its predicted value of y is 0.2. If an observation has features $x_1 \leq 3.0$ but $x_2 > 1$, then its predicted value of y is 0.5. The tree model has partitioned the whole two-dimensional feature space into five regions and assigned a common predicted value of y for each of the five regions. We will briefly overview how trees like the one shown in Figure 1 are estimated from the data and how tuning parameters, such as the height or maximal depth of the tree, are chosen in later sections.

The rest of this paper is organized as follows. In section II, we describe how the synthetic arterial database used in this paper is generated through computational fluid dynamic (CFD) simulations. We then briefly review how CART estimates a regression tree from data. We also discuss how to tune the tree complexity using cross-validation. In section III, we apply the CART approach to the synthetic arterial data. We end this paper with a discussion in section IV.

[1] Xiaoqian Liu and Eric C. Chi are with the Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA xliu62@ncsu.edu, eric_chi@ncsu.edu

[2] Madhurima Vardhan, Arpita Das and Amanda Randles are with the Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA madhurima.vardhan@duke.edu, arpita.das@duke.edu, amanda.randles@duke.edu

[3] Qinrou Wen is with the School of Mathematical Science, Zhejiang University, Hangzhou, Zhejiang 310027, China wqr_olivia@163.com

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

II. Methodology

A. *Performing computational fluid dynamic simulations in a synthetic arterial database*

This study does not involve any experiments on humans or the use of human tissue samples and used image-derived vascular geometries. We created a synthetic arterial database in which we created different combinations of geometries. Vascular anatomies were artificially modified with curvature, length, and occlusion severity by changing the lesion falloff, length and percent stenosis across the bifurcation lesion using Blender, an open source mesh software. Variations for each classification were created: for curvature, smooth and sharp; for length, 10mm (focal), 15mm (tubular), and 20mm (diffuse); and for occlusion severity, 50%, 75% and 95% reduction in vessel diameter. We also varied the bifurcation angle from 30 to 83 degrees and the number of side branches (2, 3 or 4) in the left anterior descending artery. We completed the study for 10 different initial anatomies to minimize bias that could be introduced by using only one underlying template. Therefore, our synthetic database consists of 176 arterial geometries with different types of bifurcation lesion anatomies for the treated and untreated groups. A synthetic database enables a systematic investigation of how local hemodynamics varies with lesion geometry and the isolation of specific anatomic features that alter local blood flow variables.

The arterial geometries were used as inputs to perform 3D physiological simulations using HARVEY, a parallel application based on the lattice Boltzmann method, an alternative to traditional Navier-Stokes solvers [11]. Arterial simulations were performed by modeling blood as an incompressible Newtonian fluid with a density of 1.06 kg/m^3 and dynamic viscosity of 4 cP [17]. The lateral blood vessels walls were modeled using a no-slip boundary condition. At the outlets, a lumped parameter model was prescribed using microcirculation resistance and at the inlets a Poiseuille profile was imposed with transient flow using velocity waveform from literature [14]. From these CFD simulations, TAWSS was computed at the bifurcation site in two locations 1) the main branch in the left anterior descending vessel and 2) the side branch in the diagonal vessel.

B. *Estimating a Tree Model with CART*

A CART model is estimated or fit to the data through recursive binary partitioning followed by a pruning step and the enforcement of multiple stopping rules. For simplicity, we focus constructing binary partitions as shown in Figure 1.

Throughout, we assume our dataset consists a set of p feature variables $S = \{x_1, \dots, x_p\}$ and a response variable y to be predicted. To construct a binary regression tree, we recursively create binary partitions as follows. Starting at the root node, we consider a splitting variable x and a cut-off value c to divide the space into two half-planes, $\{x \le c\}$ and $\{x > c\}$, and then model the response y by its sample mean over each region. We seek the splitting variable and cut-off value that achieves the best fit in a least squares sense.

Algorithm 1 Pseudocode for tree construction

```
1: Start at the root node
2: For each feature variable in $S$, find the cut-off value
   that minimizes the sum of the fitted squared-error loss
   in the two child nodes, and choose the variable $x'$ and
   the corresponding cut-off value $c'$ that minimizes the
   squared prediction error over all $x \in S$.
3: If a stopping criterion is reached, exit. Otherwise, apply
   step 2 to each child node in turn.
```

Splitting stops once a stopping rule is satisfied. For instance, we can stop partitioning if the relative decrease in the least squares prediction error falls below a prespecified threshold. Algorithm 1 summarizes this tree construction procedure.

The resulting tree partitions the feature space into disjoint regions. We model the variable y as the sample mean in each region, which leads to a piece-wise constant function over the feature space. We will see in section III that the final CART model is quite intuitive and mirrors how clinical decisions are often made based on thresholds in biomarkers.

C. *Tuning Tree Complexity with Cross-Validation*

Despite its advantages in simplicity and flexibility, the CART model may run into issues when presented with many irrelevant features. As the number of features increases, the size of the tree grows rapidly, potentially leading to overly complex models and nullifying the model's attractive interpretability. An additional serious issue is that including more variables in a model will always lead to better fits to the idiosyncrasies of the data used to train the model but poorer generalization, or predictive, performance on data not used to train the model. This is the so-called overfitting problem.

To address these issues, we seek to balance the trade-off between the tree complexity, quantified in its depth and number of splits, and the model's goodness-of-fit to the training data. We control the fitted tree's complexity by setting parameters, such as the maximum depth of the tree and the minimum number of samples required to be at a leaf node. This raises the problem of tuning parameter selection.

In this paper, we use cross-validation to tune the complexity of the CART model. Cross-validation is commonly used for selecting tuning parameters in statistics and machine learning. The basic idea of cross-validation is to evaluate the model performance using a resampling procedure.

The details of the K-fold cross-validation procedure are as follows. Suppose we wish to select a maximal tree depth γ from a set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_m\}$ of m candidate maximal tree depths. Given a sample of data, we randomly split the full dataset into K roughly equal-sized groups. Choices of K are typically $5, 10$, or possibly even the sample size n, which corresponds to leave-one-out cross-validation. We set aside one group as the validation set and use the remaining $K - 1$ groups as the training set. We next apply the model to the training set for each γ_j for $j = 1, \dots, m$, and then we calculate the mean-squared prediction error of each fitted model on the held out validation set. The process is repeated K times, so that each group is used exactly once as the validation set. As a result, we obtain K estimates of the prediction error for each $\gamma_j \in \Gamma$. We select the maximal depth $\gamma_j \in \Gamma$ that minimizes the average prediction error.

III. Experiments

In this section, we apply the CART approach on the synthetic TAWSS database as described in Section II. The response variables that we are interested in predicting are the TAWSS in 1) the main branch and 2) the side branch. The feature variables we use include curvature, length, occlusion severity, the bifurcation angle, and the number of side branches in the left anterior descending artery.

A. *TAWSS in the Main Branch*

We first focus on predicting the TAWSS in the main branch. We randomly split the data set into a training set (75%) to fit the regression tree and a testing set (25%) to evaluate the prediction performance of the fitted tree. When fitting the regression tree by CART, we give a range of different values for the maximal depth parameter and use five-fold cross-validation to select the best maximal depth for the tree. We evaluate the prediction performance by the mean squared error defined as $\|\mathbf{y}-\hat{\mathbf{y}}\|_2^2/n_{\text{test}}$, where $\mathbf{y}$ is the vector of TAWSS values in the testing set, $\hat{\mathbf{y}}$ is the CART prediction of $\mathbf{y}$, and n_{test} is the number of observations in the testing set. We estimated a final fitted regression tree with a maximal depth of four and a highly accurate prediction error of 0.003 on the testing set.

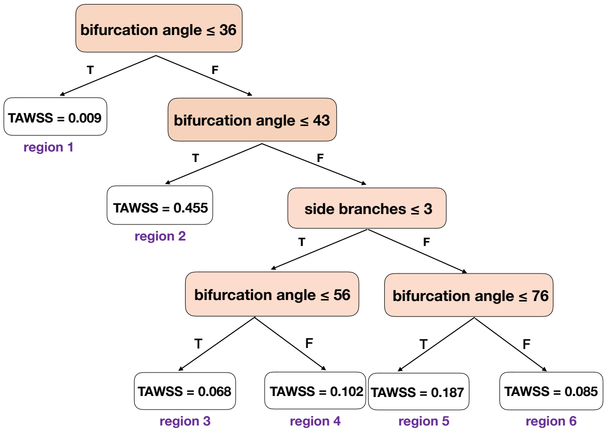

Fig. 2. Fitted regression tree for predicting the TAWSS in the main branch

Figure 2 displays the fitted regression tree for predicting the TAWSS in the main branch. We see that the tree model only includes two features, bifurcation angle and side branches, which indicates that the other three features were less critical for predicting the TAWSS in the main branch. The fitted tree has six leaf nodes, which splits the feature space into six regions. At each leaf node, the TAWSS in the main branch is predicted as a constant value, namely the sample mean of the TAWSS values of the observations falling in that region. Figure 3 displays violin plots of the TAWSS in the main branch in each of the six regions. We see that the distribution of the TAWSS in each region is quite different from each other, which suggests that the CART approach can potentially identify clinically meaningful new classifications based on anatomic geometries. We note in particular that the TAWSS of patients whose features land them in region 2 are noticeably higher than that of the other regions while the TAWSS of patients whose features land them in region 1 are noticeably lower. The number of observations in regions 1 and 2 are on the lower end, 15 and 12 samples respectively, compared to other leaves which have 20 to 40 samples in them. Consequently, the extreme values of TAWSS in these regions may be due to sampling uncertainty. In future work, we plan to generate more geometries from these two regions to see if the pattern persists.

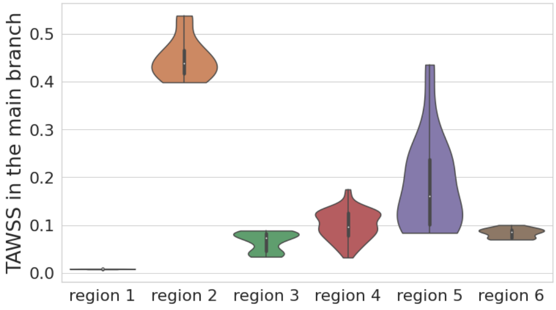

Fig. 3. Violin plots of TWASS in the main branch

B. *TAWSS in the Side Branch*

We next turn to predict the TAWSS in the side branch. Employing the same procedures applied in the main branch, we estimated a fitted regression tree with a maximal depth of three and a prediction error of 0.09 on the testing set.

Figure 4 displays the fitted regression tree for predicting the TAWSS in the side branch. In this case, the fitted tree has four leaf nodes, which splits the feature space into four regions. The side branches tree model, however, uses a different set of three features in making this partition: bifurcation angle, length, and severity. Figure 5 displays violin plots of the TAWSS in the side branch in each of the four regions. We see that the distributions of the TAWSS in the side branch in these four groups are again distinct. Note here that the regions in Figure 5 are different from those in Figure 3 since they are from different feature spaces.

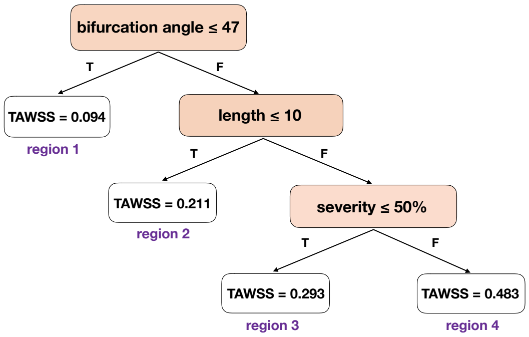

Fig. 4. Fitted regression tree for predicting the TAWSS in side branch

We note that the bifurcation angle is selected as the root or first splitting feature of both fitted trees, indicating that it is the most important factor affecting the TAWSS in main

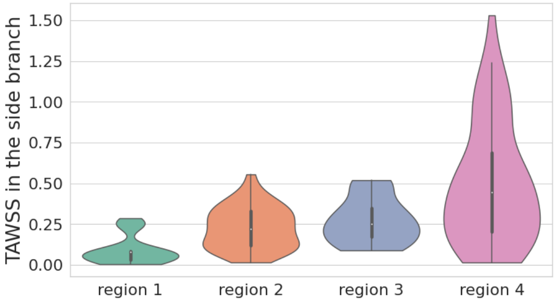

Fig. 5. Violin plots of TAWSS in the side branch

and side branches. This finding is supported by previous clinical evaluations that reported the bifurcation angle as a key criteria in determining the severity of bifurcation lesion and patient outcomes [6], [10]. At the same time the regression trees also differ in important ways, as patient anatomy, namely the number of side branches, appears to more strongly influence the TAWSS in the main branch, while lesion geometry, namely length and severity, appears to more strongly influence the TAWSS in the side branch.

IV. Discussion

In this paper, we sought to better understand how patient-specific and lesion-specific parameters affect an important clinical metric of ischemic burden, namely the TAWSS. We modeled TAWSS as a function of lesion and patient anatomy using a regression tree via CART. The fitted regression tree has the advantage of capturing nonlinear relationships between TAWSS and the geometric features, while producing a simple and highly interpretable model. Such assessment has the potential to help guide physicians to personalize interventional strategies to a patient's cardiac vasculature and lesion configuration. For example, the tree model makes progress towards answering the question whether stenting the main branch with or without the side branch offers hemodynamic advantages over double stenting. Such personalized treatment strategies could provide improved hemodynamics and result in better patient outcomes.

Acknowledgment

Computing support for this work came from Duke OIT and the Duke Compute Cluster. This material is based upon work supported by the NSF under Award Number 1943036 and by the NIH under Award Number U01CA253511, and the AHA Predoctoral Fellowship Award No. 20PRE35211158. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or NIH.

References

[1] Rovlias, A., and Kotsou, S. Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *Journal of Neurotrauma*, 21(7):886–893, 2004.

[2] Darius Aliabadi, Frank V Tilli, Terry R Bowers, Keith H Benzuly, Robert D Safian, James A Goldstein, Cindy L Grines, and William W O'Neill. Incidence and angiographic predictors of side branch occlusion following high-pressure intracoronary stenting. *The American Journal of Cardiology*, 80(8):994–997, 1997.

[3] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and Regression Trees*. CRC press, 1984.

[4] Stephen G Ellis, Michel G Vandormael, Michael J Cowley, Germano DiSciascio, Ubeydullah Deligonul, Eric J Topol, and Thomas M Bulle. Coronary morphologic and clinical determinants of procedural outcome with angioplasty for multivessel coronary disease. implications for patient selection. multivessel angioplasty prognosis study group. *Circulation*, 82(4):1193–1202, 1990.

[5] Judith A Falconer, Bruce J Naughton, Dorothy D Dunlop, Elliot J Roth, Dale C Strasser, and James M Sinacore. Predicting stroke inpatient rehabilitation outcome using a classification tree approach. *Archives of Physical Medicine and Rehabilitation*, 75(6):619–625, 1994.

[6] Demosthenes G Katritsis, Andreas Theodorakakos, Ioannis Pantos, Manolis Gavaises, Nicos Karcanias, and Efstathios P Efstathopoulos. Flow patterns at stented coronary bifurcations: computational fluid dynamics analysis. *Circulation: Cardiovascular Interventions*, 5(4):530–539, 2012.

[7] Konstantinos C Koskinas, Yiannis S Chatzizisis, Aaron B Baker, Elazer R Edelman, Peter H Stone, and Charles L Feldman. The role of low endothelial shear stress in the conversion of atherosclerotic lesions from stable to unstable plaque. *Current Opinion in Cardiology*, 24(6):580–590, 2009.

[8] Azeem Latib and Antonio Colombo. Bifurcation disease: what do we know, what should we do? *JACC: Cardiovascular Interventions*, 1(3):218–226, 2008.

[9] Adel M Malek, Seth L Alper, and Seigo Izumo. Hemodynamic shear stress and its role in atherosclerosis. *Jama*, 282(21):2035–2042, 1999.

[10] Md Foysal Rabbi, Fahmida S Laboni, and M Tarik Arafat. Computational analysis of the coronary artery hemodynamics with different anatomical variations. *Informatics in Medicine Unlocked*, 19:100314, 2020.

[11] Amanda Peters Randles, Vivek Kale, Jeff Hammond, William Gropp, and Efthimios Kaxiras. Performance analysis of the lattice Boltzmann model beyond Navier-Stokes. In *2013 IEEE 27th International Symposium on Parallel and Distributed Processing*, pages 1063–1074. IEEE, 2013.

[12] Aristedis Rovlias, Spyridon Theodoropoulos, and Dimitrios Papoutsakis. Chronic subdural hematoma: Surgical management and outcome in 986 cases: A classification and regression tree approach. *Surgical Neurology International*, 6, 2015.

[13] Nancy R Temkin, Richard Holubkov, Joan E Machamer, H Richard Winn, and Sureyya S Dikmen. Classification and regression trees (CART) for prediction of function at 1 year following head trauma. *Journal of Neurosurgery*, 82(5):764–771, 1995.

[14] Ryo Torii, Nigel B Wood, Nearchos Hadjiloizou, Andrew W Dowsey, Andrew R Wright, Alun D Hughes, Justin Davies, Darrel P Francis, Jamil Mayet, Guang-Zhong Yang, et al. Fluid–structure interaction analysis of a patient-specific right coronary artery with physiological velocity and pressure waveforms. *Communications in Numerical Methods in Engineering*, 25(5):565–580, 2009.

[15] Shengxian Tu, Mauro Echavarria-Pinto, Clemens von Birgelen, Niels R Holm, Stylianos A Pyxaras, Indulis Kumsars, Ming Kai Lam, Ilona Valkenburg, Gabor G Toth, Yingguang Li, et al. Fractional flow reserve and coronary bifurcation anatomy: A novel quantitative model to assess and report the stenosis severity of bifurcation lesions. *JACC: Cardiovascular Interventions*, 8(4):564–574, 2015.

[16] Madhurima Vardhan, Arpita Das, Jonn Gouruev, and Amanda Randles. Computational fluid modeling to understand the role of anatomy in bifurcation lesion disease. In *2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW)*, pages 928–933. IEEE, 2018.

[17] Madhurima Vardhan, John Gounley, S James Chen, Andrew M Kahn, Jane A Leopold, and Amanda Randles. The importance of side branches in modeling 3d hemodynamics from angiograms for patients with coronary artery disease. *Scientific Reports*, 9(1):1–10, 2019.

[18] Dong Zhang, Bo Xu, Dong Yin, Yiping Li, Yuan He, Shijie You, Shubin Qiao, Yongjian Wu, Hongbing Yan, Yuejin Yang, et al. How bifurcation angle impacts the fate of side branch after main vessel stenting: a retrospective analysis of 1,200 consecutive bifurcation lesions in a single center. *Catheterization and Cardiovascular Interventions*, 85(S1):706–715, 2015.