Thornton Tomasetti

August 2022

AI Engine for Thornton Tomasetti in Silico Trials: Aortic Graft Stent Demonstrator

Engineer Intern Sofia Lima, Project Engineer Elham Ramyar, Principal John Mould, Vice President Ashley Peterson, and Vice President Kristian Debus

1 Introduction

The medical device industry recognizes the general motivation for in silico trials. The CORE.AI project at Thornton Tomasetti (TT) aims to create an automated workflow for medical device in silico trials. In this work, the goal of the in silico trial is to assess the efficacy and safety of a generic, representative thoracic stent graft for the treatment of a thoracic aortic aneurysm (TAA). We created a virtual cohort of TAA patients from CT image data. This data was processed using various tools, including Rhino scripting in Python, allowing the virtual anatomy of the patient and stent to interact. The data was then analyzed using FEA and the TT Nitinol model in FLEX in order to simulate the stent implantation trial and generate output data. Finally, machine learning techniques using SciKit-learn packages in Python were applied to determine input importance.

2 Project Outline

2.1 Dataset

2.1.1 Parametrize Patient Anatomy

Virtual anatomies were parameterized from CT image data. Briefly, a generic, representative thoracic aorta was defined from available image, the centerline through the lumen of the aorta was calculated using Simpleware image segmentation software. Then, ellipses were fit to points along the centerline. Due to the limited availability of clinical data, virtual patients were generated by varying the size, shape, and location of an aneurysm in the virtual anatomy of the TAA (Figure 1). For these variations, we used ranges that would be expected in the real patient population. In practice, this data would be generated from a real patient dataset if provided by the client.

The stent graft device was designed as a combination of a number of stent rings and tube of graft material. The stent graft length can be adjusted based on the number of rings appropriate for the aneurysm.

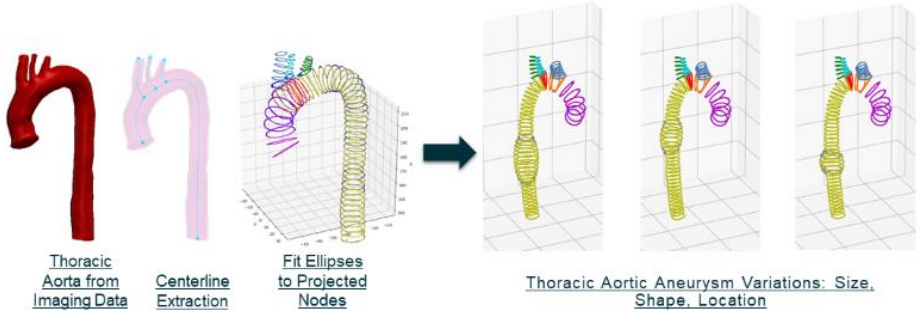

Figure 1. TT's Parametric Anatomy Modeler. Parameterization of TAA based on limited imaging data using TT's parametric anatomy modeler.

2.1.2 Rhino Python Scripting

In order to allow the interaction between the patient anatomy and device to be characterized, Rhino scripting in Python was used to automate the data collection process. For each virtual anatomy defined by 3D data, we process the following in Rhino: auto-loft the anatomy, auto-mesh the anatomy for FEA, auto-create the stent graft, auto-select and -align the stent graft based on clinical guidelines (Figure 2). The selection and placement of the stent graft was based on the best fit stent diameter and length of available prototypes. In our case, we have a single diameter of 36 mm at two lengths of 202 and 253 mm; we select the best stent length to be at least the length of the aneurysm + 30cm on the proximal end + 30 cm on the distal end; and we properly position the stent relative to the aorta.

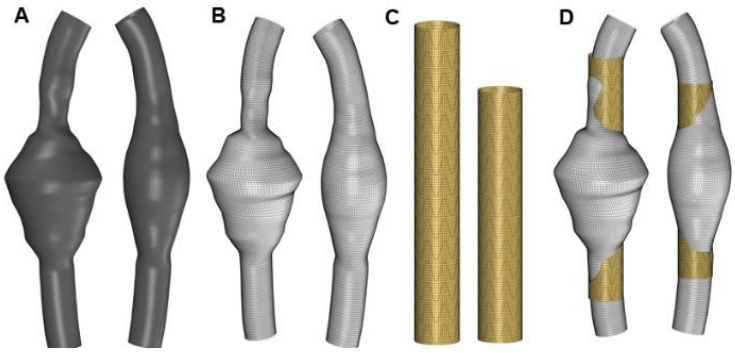

Figure 2. Processing anatomies with Rhino. A) Two anatomies after creating a loft. **B)** Mesh for FEA. **C)** Stent graft. **D)** Aligned stent graft.

2.1.3 FLEX FEA Analysis

The TT Nitinol model was used in this simulation experiment with an assumed friction coefficient of 0.1 between the graft and vessel walls and with dynamic relaxation to obtain a static solution with an explicit FEA code. The TT Nitinol model is able to fit the material properties of nitinol (metal alloy of nickel and titanium) on par with industry standards (Figure 3) and includes additional flexibility to capture feature not represented by the industry standard model. Using this tool, we are able to accurately simulate the loading and unloading response of the device during deployment. Here, we simulated surgical crimping and deployment by radially compressing the stent, properly positioning, and releasing (Figure 4). Once released, the stent graft is subjected to pressure changes in the cardiac cycle. Specifically, we applied a pressure difference between the inside and outside of the graft and a static bias pressure towards the largest deviation of the aneurysm to simulate the loading due to blood momentum.

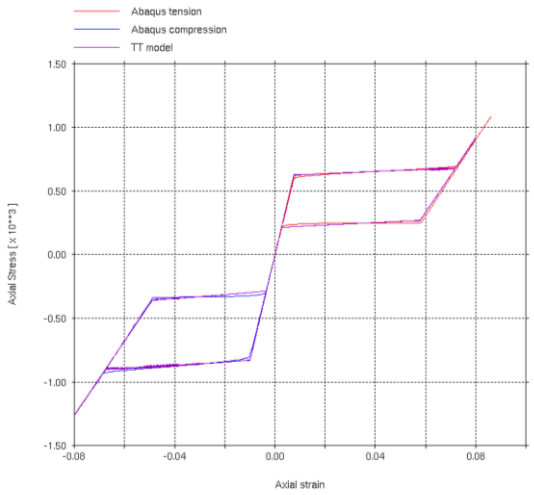

Figure 3. Comparing stress-strain curve for mechanical loading/unloading on nitinol from TT Nitinol Model and Abaqus. The stress-strain curve represents mechanical properties for a material. As positive stress (tension) is applied to nitinol, the material is deformed such that the strain increases. This relationship is somewhat different when compression stress is applied. As shown here, the TT Nitinol model is fit to a state-of-the-art Abaqus model in tension and compression.

Note that though hysteretic, these curves are reversible (elastic) up to strains exceeding 8%. This is why Nitinol is widely used in medical devices.

The FEA models are able to simulate the structural response of the stent graft and anatomy, which leads to the computation of representative trial outcome quantities (Figure 4). To extract the output data, the FLEX FE analyses from the entire cohort of 77 TAA virtual patients were interrogated. Descriptions of input data collected from parameterization of the thoracic aorta and output data extracted via FEA analysis are summarized in Figure 5. We choose these inputs because they are readily available in the operating room and hence clinically and physically tractable.

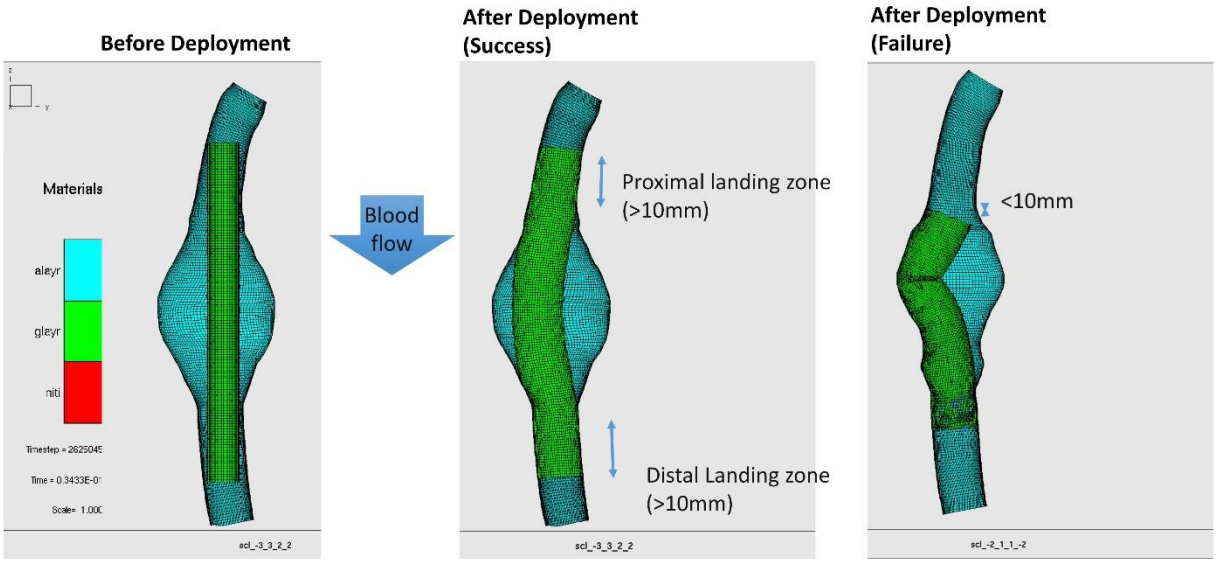

Figure 4. FLEX FE Analysis used to extract output data from deployment successes and failures. Output parameters shown on a cutaway view. Anatomies of stent and aorta shown after auto-aligning (left) and after deployment (right). The stent deforms due to momentum of internal blood flow, resisted by frictional contact with the aorta. FLEX FE analysis is performed to obtain output parameters such as proximal landing length in mm. Final stent graft positions in context of the cardiac cycle in two different trials: one successful with a good seal along the proximal and distal axis (left), and one failure with a proximal seal failure (right).

6 input variables:

X_1 = function of F, and J
X_2 = function of B
X_3 = function of A
X_4 = function of C
X_5 = function of D
X_6 = function of E

1 output variable:

Y = Proximal landing zone length of stent graft

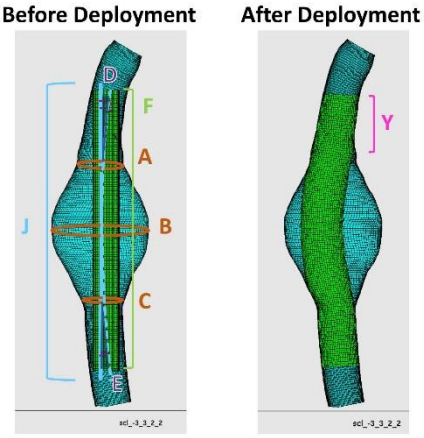

Figure 5. Variable for mathematical analysis. We focus our analyses on the set of inputs above, representing the geometry of the patient. Because our cohort is hypothetical, there is no age, sex, medical history, etc. to consider. We focus our classification analysis on discretized outcomes, and our regression analysis on the proximal landing zone in mm.

2.2 Data Analysis with Supervised Learning methods

The main objective in our analysis is to determine which inputs have the most effect on the output so that improved patient outcomes can be obtained. For this purpose, we use supervised learning methods to fit mathematical models with our data and split the 77 data samples into 55 train subjects and 22 test subjects. We discretize the outcomes by labeling failures as those trials where either the proximal or distal landing distance is less than 10 mm. In total, we find 11 instances of failure, and 66 successful cases.

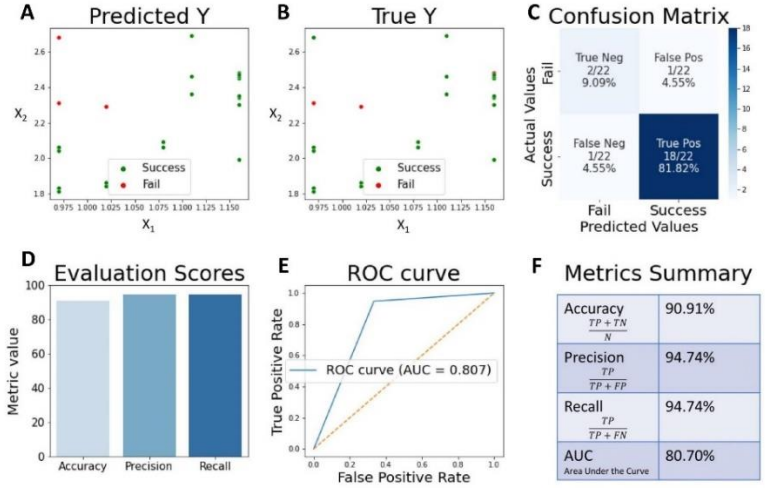

Figure 6. Binary classification analysis with Decision Tree. To evaluate the performance of the model built with our training data, we predicted outcomes using the test data and compared to the known labels. **A** and **B)** Test datapoints plotted along the X_1 and X_2 axes, where the color of the datapoint indicates the outcome. **C-F)** We calculated various evaluation metrics to analyze the performance of our classifier. The confusion matrix counts the number of correct and falsely labeled predictions in our test dataset (n=22). Accuracy, precision, recall, and ROC/AUC are calculated from the confusion matrix.

2.2.1 Classification

Using our discretized outcome data, we used a decision tree classifier, which is a popular method because of its intuitive decision-making algorithm. This model, which we implemented using SciKit-learn, learns a set of 'if-else' rules from the training dataset and builds a tree by splitting on attributes that result in the 'purest' leaf nodes. We used a variety of popular evaluation metrics for binary classification (Figure 6). We evaluated how well our model can be generalized to unseen data and found that our decision tree model predicts outcomes with 90.91% accuracy on this test dataset. Accuracy is generally interpreted as an overall performance. The receiver operating characteristic (ROC) curve compares the false positive rate (FPR) and the true positive rate (TPR); a good curve has a low FPR and a high TPR such that the area under the curve (AUC) is close to (or equal to) 1.

2.2.2 Regression

We also explored regression models to predict outcomes of continuous proximal landing zone values (mm), and we evaluate these models by comparing the true labels to the predicted values using root mean squared (RMS) error. We performed a ridge regression analysis with Scikit-learn's RidgeCV method using different values of the hyperparameter alpha. We chose ridge regression (also known as linear regression with L2 regularization where alpha is the regularization term) because it is a popular method for analyzing which inputs have the most influence on the continuous output (proximal landing zone in mm). We find that this method has RMS of about 6 mm. Ideally, we want this value to be as low as possible in order to be confident in the model's prediction abilities.

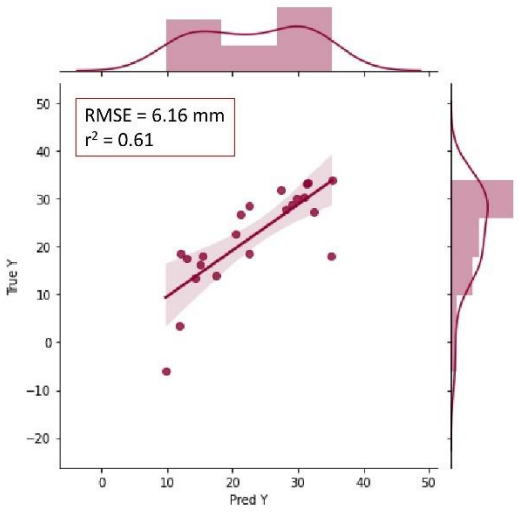

Figure 7. Ridge Regression model performance. Using our selected inputs and continuous value y data generated by FEA analysis, we found that ridge regression has the best performance in terms of RMS error compared to other models we tested with this dataset (data not shown). We show here a scatterplot of the correlation, and the distribution of the predicted and true values on the marginal axes.

We evaluated our performance using root mean squared error (RMSE) and the correlation coefficient (r^2). We found that the RMSE here is 6.16 mm. This can be interpreted as the following: for any prediction, the model will be off by about 6 mm on average.

3 Discussion and Future Directions

With our dataset, we were able to make a binary classification with 80-90% accuracy using various classifiers, with the decision tree model resulting in the highest accuracy. Ideally, a larger dataset with more virtual patients would allow us to make even better predictions for all models, especially for predicting continuous y values. We limit our dataset to n=77 due to budget constraints for this proof of concept effort. Nonetheless, with only 77 samples we are able to capture important relationships between the important factors in this surgical experiment.

We designed our in silico experiment similarly to a clinical trial, with relatively few fail cases. We could bootstrap, a common up-sampling method in clinical trial experimentation, from the set of fail cases or generate more fail cases in order to address the imbalance in our dataset. One strong advantage for in silico trials over in vivo clinical trials is the ability to generate fail cases. This could help enormously with the product design process by being able to capture not only what makes a success a success, but also what causes failures. Given the limitations of experimentation in the real world, the virtual world could play a key role for efficient development of therapies in many areas of biomedicine.

Our dataset here is generated as a hypothetical diseased population. Here, the role of simulation will allow us to identify key parameters in the anatomies in order to engineer a better design. Proper validation of the medical device for pushing the device forward in human trials would require real clinical data provided by the client, and/or analyzing 3D printed models.

We show quantitatively that the anatomy of the patient and device influences therapeutic outcomes in this simulation experiment. We find that the length of the stent compared to the length of the aneurysm in the patient should be an important factor when deciding therapeutic effectiveness. Intuitively, this is likely to be a human doctor's most deciding factor as well, providing us with this very mathematically satisfying demonstration.

In order to interpret the feature learned importance in the ridge regression model, we plot the absolute values of the coefficients associated with each input variable in the regression model. The higher the absolute value of the coefficient, the more influence the associated input variable has on the output (the y value in a standard linear equation). We plotted the coefficients of our model built on the training data for three values of alpha, where alpha=0.005 was chosen as the best estimator based on RMSE (Figure 8, right panel). We found that X_1 and X_2 have the highest absolute value coefficients, consistent with our decision tree analysis for discrete outcomes.

From our decision tree analysis, we found that X_1 and X_2 are indeed the most important features in this tree-building algorithm. We plotted using the '.feature_importances_' attribute on the decision tree classifier object in SciKit-learn (Figure 8, left panel).

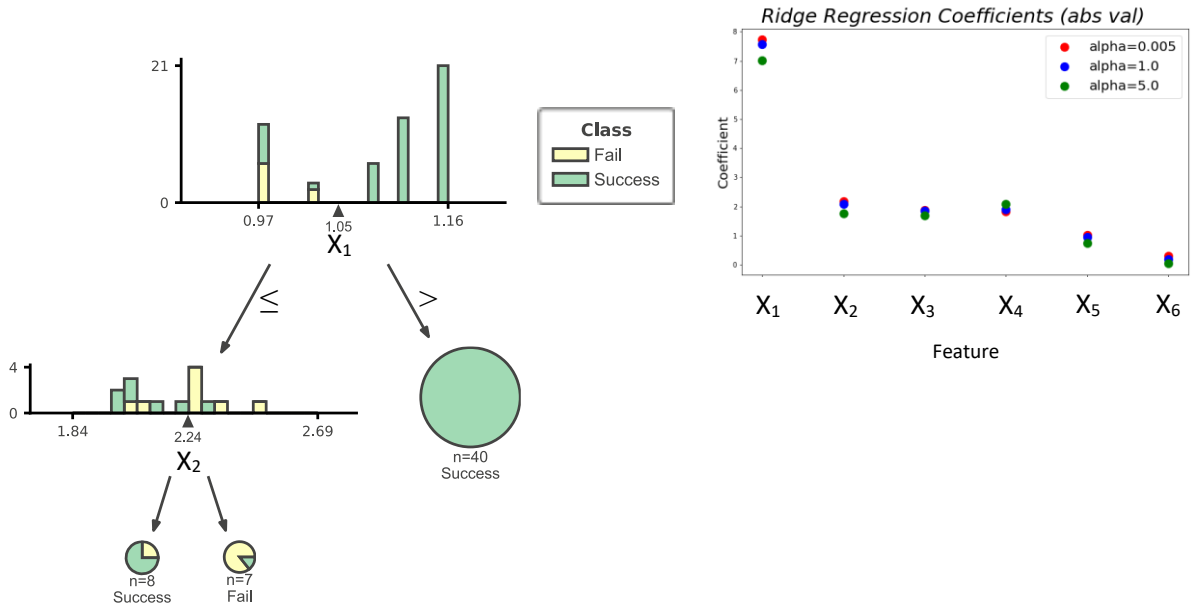

Figure 8. Classification and regression agree. Comparing results from decision tree for predicting discrete outputs and ridge regression for continuous outputs, we see agreement not only among the mathematical models but also with intuition. **Left)** Importantly, the splits in a decision tree model are decided based on the optimal information gain from all possible splits. Thus, we can make very interesting inferences regarding things such as relative feature importance. The visualized tree shows the value at which attribute the split was decided, the total number of samples in the node, and the counts of success and failures (values). From our input variables, the most important features inferred here are X_1 and X_2 described above in Figure 5. **Right)** The coefficients associated with each input in the ridge regression model are plotted. X_1 and X_2 have the coefficients farthest from zero in this instance.

In this project, we developed an exemplar methodology to leverage available data to overcome practical limitations in clinical trials in the medical device industry. The demonstrator here is critical step for making in silico trials a reality, and the Life Sciences team is eager to apply the sophisticated simulation capabilities available at Thornton Tomasetti to solve real-world health problems.