

# Mining Data from CFD Simulation for Aneurysm and Carotid Bifurcation Models

Radović Miloš, Petrović Dejan, and Filipović Nenad

**Abstract**—Arterial geometry variability is present both within and across individuals. To analyze the influence of geometric parameters, blood density, dynamic viscosity and blood velocity on wall shear stress (WSS) distribution in the human carotid artery bifurcation and aneurysm, the computer simulations were run to generate the data pertaining to this phenomenon. In our work we evaluate two prediction models for modeling these relationships: neural network model and k-nearest neighbor model. The results revealed that both models have high prediction ability for this prediction task. The achieved results represent progress in assessment of stroke risk for a given patient data in real time.

## I. INTRODUCTION

AFTER heart disease and cancer, the third most common cause of death is stroke. The carotid bifurcation stenosis is a significant cause of stroke, producing the infarction in the carotid region by embolization or thrombosis at the site of narrowing. The thrombosis development and embolization is conditioned by the local hemodynamics which can be investigated experimentally and/or by computer modeling.

There are many factors which increase the stroke risk like age, systolic and diastolic hypertension, diabetes, cigarette smoking, etc. It has been shown that changes of the geometrical vessel dimensions in the region of the carotid artery bifurcation certainly affect the blood flow and may lead to stenosis process [1], [2].

Kolachalama used Bayesian Gaussian process emulator to access the relationship between geometric parameters and Maximal Wall Shear Stress (MWSS) and to obtain geometries having maximum and minimum values of the output MWSS [3].

Large changes in the magnitude of maximal wall shear stress can play a role in the embolic mechanism by which carotid lesions can induce stroke [4].

It has been shown that many data mining algorithms like neural networks (NN), linear regression, k-nearest neighbors (k-NN), random forest, and support vector machine have

high potential for modeling relationship between geometric parameters of the carotid bifurcation and the MWSS [5].

The rupture of aneurysm, can cause severe hemorrhage, other complications or death. It has been shown that aneurysm growth occurs at regions of low WSS [6].

An example of data mining application in computational fluid dynamics (CFD) has been shown in Filipovic's paper [7]. In this paper, the focus was to combine the CFD and data mining methods for the estimation of the wall shear stresses in an abdominal aorta aneurysm under prescribed geometrical changes.

In the present work, we evaluate two data mining prediction models (NN model and k-NN model) and test their performance in modeling the relationship between geometric factors, blood density, dynamic viscosity and blood velocity and WSS distribution. The basic idea is to construct probabilistic models for the input variables which will replace classical CFD calculations and to give the output of interest very quickly.

The present approach can be viewed as a computer-based data mining strategy which extracts useful information and synthesizes interesting relationships from data sets generated by running computer simulations on selected cases. The human carotid artery bifurcation and aneurysm were chosen for analysis.

## II. METHODOLOGY

### A. Data Sets for Modeling WSS Distribution

To demonstrate applicability of data mining techniques for assessing relationships between geometric parameters, density, viscosity and velocity one hand and WSS distribution on the other, data sets containing 4779 different examples for aneurysm and carotid bifurcation models are created. Tables I and II show mean values of input variables for two different finite element (FE) models. The perturbation of each parameter was taken as 30% of the corresponding mean value.

TABLE I  
THE AVERAGE VALUES OF INPUT PARAMETERS FOR ANEURYSM MODEL

Description	Mean value	Unit
Length	100	mm
Aneurysm length	40	mm
A	30	mm
Aorta diameter	20	mm
C	20	mm
B	20	mm
Density	0.00105	gr/mm <sup>3</sup>
Dynamic viscosity	0.003675	Pa·s
Velocity	28.13	mm/s

Manuscript received March 25th, 2011. This work was supported in part by the European Commission (Project ARTREAT, ICT 224297) and Serbian Ministry of Science Projects III-41007 and ON-174028.

M. Radović and D. Petrović are with the Research and Development Center for Bioengineering "BioIRC", Kragujevac, Serbia (phone: 381-34-301920; fax: 381-34-500089; e-mail: [mradovic@kg.ac.rs](mailto:mradovic@kg.ac.rs) and [racanac@kg.ac.rs](mailto:racanac@kg.ac.rs)).

N. Filipović, is with the Faculty of Mechanical Engineering, Kragujevac, Serbia and Research and Development Center for Bioengineering "BioIRC", Kragujevac, Serbia (e-mail: [fica@kg.ac.rs](mailto:fica@kg.ac.rs)).

TABLE II  
THE AVERAGE VALUES OF INPUT PARAMETERS FOR CAROTID BIFURCATION MODEL

Description	Mean value	Unit
Angle ICA-CCA	25	deg
Angle ECA-CCA	25	deg
Diameter of CCA	6.2	mm
Diameter of CBRE	3.658	mm
Diameter of CBRI	4.9	mm
Length of CCA	7.44	mm
Length of CBR	7.316	mm
Length of ECA	18.6	mm
Length of ICA	26.04	mm
Diameter at end of ICA	4.34	mm
Diameter of ICB	6.49	mm
Distance to ICB	5.39	mm
Density	0.00105	gr/mm
Dynamic viscosity	0.00367	Pa·s
Velocity	233	mm/s

A 3D FE models with 3D fluid finite elements (8-node isoparametric elements with velocity calculation at all nodes and pressure calculated at the element level) are generated for the carotid artery and aneurysm. The CFD post-processing results give an insight into the local hemodynamics, as well as the blood mechanical action on the vessel walls, such as distributions of shear stress on the wall surfaces. The geometric parameters (Figures 1 and 2) are used for the generation of the blood vessel internal surfaces, which are the boundaries for the blood flow domain. It is assumed that both, carotid bifurcation and aneurysm, have the symmetry plane, hence the FE models are generated for the half of the entire domain. The calculation is performed for this half, but the results can be seen for the entire domain. Steady state simulations with parabolic inflow velocity profile were undertaken and WSS distribution for each geometry was computed. All calculations were performed by using in house software.

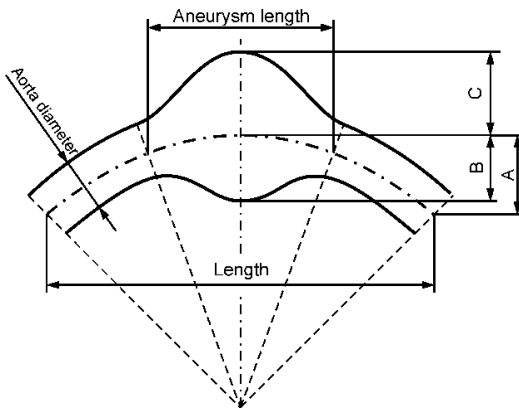


Fig. 1. Geometrical parameters of aneurysm model: 'Length' is the parameter which defines the total horizontal projection of the generated aneurysm model; 'A' is the height of the arc of central line; 'Aorta diameter' is the abdominal aorta diameter; 'B' is the radius from the central line to the inner wall of the aneurysm; 'C' is the radius from the central line to the outer wall of the aneurysm; 'Aneurysm length' is an average length of the aneurysm.

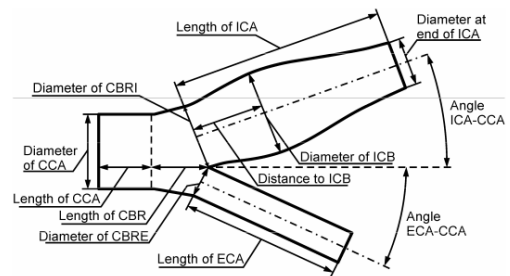


Fig. 2. Geometrical data for the carotid artery model. The abbreviations here are: CCA –common carotid artery, CBR – carotid bifurcation region, CBRE – carotid bifurcation region external, ECA- external carotid artery, CBRI- carotid bifurcation region internal, ICA- internal carotid artery, ICB- internal carotid bulb.

FE model of aneurysm contains 375 nodes from which 195 lie on surface. On the other hand, FE model of carotid bifurcation contains 1854 nodes from which 642 lie on surface. By using CFD simulations WSS values are calculated in surface nodes for each of 4779 different geometries for both models. FE models of aneurysm and carotid artery bifurcation are shown on figure 3.

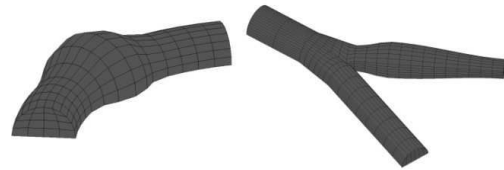


Fig. 3. Finite element models of aneurysm and carotid bifurcation.

### B. Multilayer Perceptron Neural Network

Multilayer perceptron (MLP) neural network is composed of simple elements called *neurons*. The basic structure of the MLP, consists of one or more hidden layers and an output layer.

The objective of the training is to find a set of weights and biases that minimize the error between the neural network predictions and the desired outputs. There are different learning algorithms. The back-propagation algorithm [8] has been the most commonly used training algorithm. The basic algorithm is a gradient descent method in which the network weights and biases are moved along the negative performance function. An iteration of this algorithm can be written as:

$$\Delta X = lr \cdot \frac{dperf}{dX} \quad (1)$$

where  $X$  represents weight and bias variables of the network,  $lr$  is learning rate and  $perf$  is performance function which defines how much real outputs disagree with predicted ones (mean squared error for example).

It has problems with local minima and slow convergence. In the literature, a number of variations of the standard algorithm have been developed [9]. In this study we used backpropagation algorithm with momentum and adaptive learning rate. Each variable is adjusted according to gradient descent with momentum:

$$\Delta X = m_c \cdot \Delta X_{prev} + lr \cdot m_c \cdot \frac{dperf}{dX} \quad (2)$$

where  $m_c$  is momentum constant and  $\Delta X_{prev}$  is the previous change of the weight or bias. For each epoch, if performance decreases toward the goal, then the learning rate is increased by the  $lr_{inc}$  factor. If performance increases by more than the  $max_{inc}$  factor, the learning rate is adjusted by the factor  $lr_{dec}$  and the change that increased the performance is not made. The values of  $m_c, lr_{inc}, lr_{dec}$  and  $max_{inc}$  are given in table III.

TABLE III

$m_c, lr_{inc}, lr_{dec}$  AND  $max_{inc}$  VALUES USED FOR MLP TRAINING

$m_c$	$lr_{inc}$	$lr_{dec}$	$max_{inc}$
0.9	1.05	0.7	1.04

MLP with as few as one single hidden layer is indeed capable of universal approximation in a very precise and satisfactory sense [10].

### C. K-Nearest Neighbors Algorithm

K nearest neighbors algorithm belongs to a class of lazy learning methods. When a new example is presented to a nearest neighbor predictor, a subset of learning examples most similar to the new example is used to make a prediction. This method is described in [11].

For regression problems the mean target variable value from the set of nearest neighbors is predicted:

$$c_x = \frac{1}{k} \sum_{i=1}^k c_i \quad (3)$$

where  $k$  is the number of nearest learning examples which influence the prediction of k-NN algorithm.

Type of distance measure has big impact on determining which set of learning examples are closest to the new example. In the most cases, Euclidean distance is used:

$$D(t_l, t_j) = \sqrt{\sum_{i=1}^a d(v_{i,l}, v_{i,j})^2} \quad (4)$$

In (4),  $D(t_l, t_j)$  is Euclidean distance between 2 examples  $t_l$  and  $t_j$ , and  $a$  is the total number of attributes.

Before calculating Euclidean distance all attributes are scaled to the [0,1] interval. For continuous attributes the distance between two attributes  $v_{i,l}$  and  $v_{i,j}$  is defined as:

$$d(v_{i,l}, v_{i,j}) = |v_{i,l} - v_{i,j}| \quad (5)$$

## III. RESULTS

In this paper we used MLP neural networks trained with backpropagation algorithm and k-NN algorithm for predicting wall shear stress distribution for the two different FE models. The problem that we are solving is multi-target prediction problem, and because of that for each surface node of the models we created one MLP. This means that our model consists of 195 different neural networks in case

of aneurysm model and 642 different neural networks in case of carotid bifurcation model, one for each surface node. For training this model and k-NN model we randomly chose 70% of the total data (3346 learning examples). Remaining 30% of data is used for testing (1433 testing examples).

MLPs with 5 neurons in hidden layer, bipolar sigmoid activation functions in hidden neurons and linear activation function in the output neuron are used. The stopping criterion was defined as the maximum number of learning epochs (1000). Input layer has nine input neurons (in case of aneurysm model) and fifteen input neurons (in case of carotid bifurcation model) corresponding to input parameters (see tables I and II). The output layer consists of one neuron corresponding to WSS value of the node for which MLP is created.

k-NN model predicts the target values that are averaged from the 5 most similar learning examples (nearest neighbors) in the problem space.

We evaluated the performance of the models by computing their relative mean squared error (RMSE). RMSE is computed as a sum of the squared differences between the true and the predicted values of the outputs for all of 1433 testing examples and is afterwards normalized with the sum of the squared errors of the default predictor (i.e. a model which always predicts average values of the outputs).

For j-th testing example squared error is given as:

$$ERR_j = \sum_{i=1}^n (f_{j,i} - \hat{f}_{j,i})^2 \quad (6)$$

where  $n$  is the number of surface nodes (195 or 642),  $\hat{f}_{j,i}$  is the predicted WSS value for i-th node for j-th example and  $f_{j,i}$  is the true value of WSS for i-th node of j-th example.

In the same way, squared error for default predictor for j-th learning example is calculated as:

$$\overline{ERR}_j = \sum_{i=1}^n (f_{j,i} - \bar{f}_i)^2 \quad (7)$$

where  $\bar{f}_i$  is the average value of WSS for i-th node among training examples:

$$\bar{f}_i = \frac{1}{N_{train}} \sum_{j=1}^{N_{train}} f_{j,i} \quad (8)$$

where  $N_{train}$  is the number of training examples (3346).

Finally, RMSE is calculated as:

$$RMSE = \frac{\sum_{j=1}^{N_{test}} ERR_j}{\sum_{j=1}^{N_{test}} \overline{ERR}_j} \quad (9)$$

where  $N_{test}$  is the number of testing examples (1433).

The lower RMSE is, the more accurate the model is. The RMSE values for the tested models are shown in tables IV and V for aneurysm and carotid bifurcation model respectively.

TABLE IV  
RELATIVE MEAN SQUARED ERROR OF THE TESTED MODELS FOR ANEURYSM MODEL

Model	RMSE
MLP	0.0351
k-NN	0.1008

TABLE V  
RELATIVE MEAN SQUARED ERROR OF THE TESTED MODELS FOR CAROTID BIFURCATION MODEL

Model	RMSE
MLP	0.0305
k-NN	0.2416

Figures 4 and 5 show calculated and predicted WSS distribution for three randomly chosen test examples for aneurysm and carotid bifurcation models (Unfortunately due to lack of space is not possible to show all results).

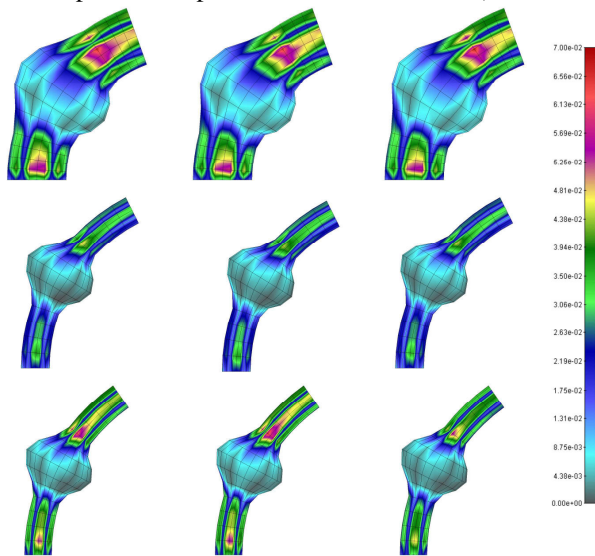


Fig. 4. WSS distribution for aneurysm model (3 randomly chosen geometries out of 1433 testing ones are shown): left-calculated, middle-MLP predicted, right-k-NN predicted [units Pa].

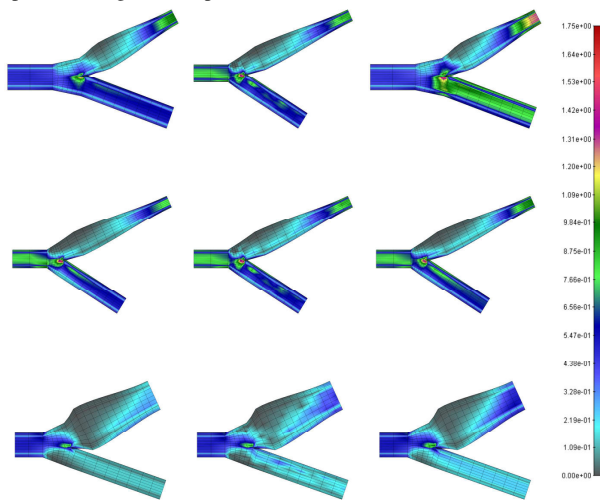


Fig. 5. WSS distribution for carotid bifurcation model (3 randomly chosen geometries out of 1433 testing ones are shown): left-calculated, middle-MLP predicted, right-k-NN predicted [units Pa].

From tables IV and V we can see that both, neural network and k-NN models, showed high potential in WSS

distribution prediction task. Same thing we can conclude from figures 4 and 5, where we can see that calculated and predicted wall shear stress distribution is quite similar. For both, carotid bifurcation and aneurysm models neural network model showed higher precision in modeling WSS distribution than k-NN model.

#### IV. CONCLUSION

This work presented an application of data mining methodology to a hemodynamic problem in which the relationship between geometric parameters, blood density, dynamic viscosity and blood velocity of the human carotid bifurcation and aneurysm, and the wall shear stress distribution was modeled. The results obtained from computer simulations were used as training data to evaluate two different regression models, which both exhibited capabilities of being used for this task. The neural network model showed better results than k-NN model. The achieved results can be used to aid the assessment of stroke risk for a given patient's data in real time.

Our ideas for further work include training and testing other regression models like support vector machines (SVM) and linear regression. Also, further research will focus on real life situations where applicability of created data mining applications will be tested on real patient data.

#### REFERENCES

- [1] U.G.R. Schulz, P.M. Rothwell, "Sex differences in carotid bifurcation anatomy and the distribution of atherosclerotic plaque," *Stroke*, vol. 32, no. 7, pp. 1525-1531, 2001.
- [2] U.G.R. Schulz, P.M. Rothwell, "Major variation in carotid bifurcation anatomy. A possible risk factor for plaque development?," *Stroke*, vol. 32, no. 7, pp. 2522-2529, 2001.
- [3] V.B. Kolachalama, N.W. Bressloff, P.B. Nair, "Mining data from hemodynamic simulations via Bayesian emulation," *BioMedical Engineering OnLine*, vol. 6, no. 47, 2007.
- [4] S. Lorthois, P.Y. Lagree, J.P. Marc-Vergnes, F. Cassot, "Maximal wall shear stress in arterial stenoses: Application to the internal carotid arteries," *ASME Journal of Biomechanical Engineering* Vol. 122, pp. 661-666, 2000.
- [5] M. Radovic, N. Filipovic, Z. Bosnic, P. Vracar and I. Kononenko, "Mining Data from Hemodynamic Simulations for Generating Prediction and Explanation Models," Conference paper, ITAB 2010 Corfu-Greece
- [6] L. Bousset, V. Rayz, C. McCulloch, A. Martin, G. Acevedo-Bolton, M. Lawton, R. Higashida, W.S. Smith, W. Young, D. Saloner, "Aneurysm growth occurs at region of low wall shear stress: Patient-specific correlation of hemodynamics and growth in a longitudinal study," *Stroke*, vol. 39, no. 11, pp. 2997-3002, 2008.
- [7] N. Filipovic, M. Ivanovic, D. Krstajic, M. Kojic, "Hemodynamic Flow Modeling Through an Abdominal Aorta Aneurysm Using Data Mining Tools," *Information Technology in Biomedicine, IEEE Transactions on*, Vol. 15, No. 2, pp. 189 - 194, 2011.
- [8] D.E. Rumelhart, G.E. Hinton, R.J. Williams, "Learning internal representations by error propagation," MIT Press, Cambridge, MA, USA, 1986, pp. 318-362.
- [9] S. Haykin, "Neural Networks: A Comprehensive Foundation," Prentice Hall, New Jersey, USA, 1999.
- [10] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural network*, vol. 4, no.2, pp. 251-257, 1991.
- [11] I. Kononenko, M. Kukar, "Machine learning and data mining," Horwood Publishing Chichester, UK, 2007.