Quantile Loss Function Empowered Machine Learning Models for Predicting Carotid Arterial Blood Flow Characteristics

T. RAJA RANI[1], WOSHAN SRIMAL[1], ABDULLAH AL SHIBLI[2],
NOOH ZAYID SUWAID AL BAKRI[3], MOHAMED SIRAJ[4], T. S. L. RADHIKA[5]
[1]Foundation Programme Department,
Military Technological College,
Muscat,
OMAN

[2]Applied & Research Department,
Military Technological College,
Muscat,
OMAN

[3]MTC Clinic,
Military Technological College,
Muscat,
OMAN

[4]Systems Engineering Department,
Military Technological College,
Muscat,
OMAN

[5]Department of Mathematics,
BITS Pilani Hyderabad,
Hyderabad,
INDIA

Abstract: - This research presents a novel approach using machine learning models with the quantile loss function to predict blood flow characteristics, specifically the wall shear stress, in the common carotid artery and its bifurcated segments, the internal and external carotid arteries. The dataset for training these models was generated through a numerical model developed for the idealized artery. This model represented blood as an incompressible Newtonian fluid and the artery as an elastic pipe with varying material properties, simulating different flow conditions. The findings of this study revealed that the quantile linear regression model is the most reliable in predicting the target variable, i.e., wall shear stress in the common carotid artery. On the other hand, the quantile gradient boosting algorithm demonstrated exceptional performance in predicting wall shear stress in the bifurcated segments. Through this study, the blood velocity and the wall shear stress in the common carotid artery are identified as the most important features affecting the wall shear stress in the internal carotid artery, while the blood velocity and the blood pressure affected the same in the external carotid artery the most. Furthermore, for a given record of the feature dataset, the study revealed the efficacy of the quantile linear-regression model in capturing a possible prevalence of atherosclerotic conditions in the internal carotid artery. But then, it was not very successful in identifying the same in the external carotid artery. However, due to the use of idealized conditions in the study, these findings need comprehensive clinical verification.

Key-Words: - Machine-learning algorithms, Quantile loss function, Boosting algorithms, Numerical model, Blood flow, Carotid artery, Wall Shear Stress.

Received: June 12, 2022. Revised: September 5, 2023. Accepted: September 27, 2023. Published: October 10, 2023.

Abbreviations/Acronyms:
AI: Artificial Intelligence
CCA: Common Carotid Artery
D: Blood Density
ECA: External Carotid Artery
ICA: Internal Carotid Artery
InPr: Inlet Pressure (Blood pressure at the CCA entrance)
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
ML: Machine Learning
MSE: Mean Squared Error
Out1Vel: Blood velocity at outlet1, i.e., ICA
Out2Vel: Blood velocity at outlet2, i.e., ECA
PARDISO: Parallel Direct Sparse Solver
RFR: Random Forest Regressor
RMSE: Root Mean Square Error
SVR: Support Vector Regressor
Vel_CCA: Blood velocity in CCA
Vel_ECA: Blood velocity in ECA
Vel_ICA: Blood velocity in ICA
VIF: Variance Inflation Factor
Vis: Blood viscosity
WSS: Wall Shear Stress
WSS_CCA: Wall Shear Stress in CCA
WSS_ECA: Wall Shear Stress in ECA
WSS_ICA: Wall Shear Stress in ICA

1 Introduction

The prediction of carotid artery blood flow has garnered significant attention in the medical community owing to its relevance in cerebrovascular diseases and neurological disorders. Conventional regression models for estimating blood flow have shown limitations, as they neglect outliers and heteroscedasticity, leading to suboptimal predictions. However, recent progress in machine learning techniques utilizing the quantile loss function presents promising avenues to overcome these challenges and improve prediction accuracy.

A comprehensive review of the existing literature reflecting the current state of the art was undertaken and detailed as follows: [1], explored the potential of artificial intelligence (AI) in predicting cardiovascular disease (CVD). Their study summarizes machine learning (ML) applications in CVD, including direct prediction based on risk factors and medical imaging and ML-based hemodynamics for indirect CVD assessment. Their review discusses research challenges and envisions future AI technology development in cardiovascular diseases. The study, [2], proposed a simulation-based framework to achieve Deep Learning (DL) based hemodynamic prediction of healthy and diseased carotid arteries. The methodology demonstrated accurate DL predictions by utilizing high-quality point cloud datasets and an advanced DL network, aligning well with CFD simulations while significantly reducing computational costs.

The review of the work by, [3], highlights the application of AI in cardiovascular imaging, focusing on coronary atherosclerotic plaque analysis. It encompasses various areas such as plaque component analysis, identification of vulnerable plaque, myocardial function detection, and risk prediction. The review discusses current evidence, strengths, limitations, future directions for AI in cardiac imaging of atherosclerotic plaques, and insights from other fields. The study, [4], presented a valuable guide for readers approaching AI algorithms in carotid atherosclerosis. Their study revealed that the application of AI using US (UltraSound), CTA (Computed Tomography Angiography), and MRI (Magnetic Resonance Imaging) would offer a new strategy for rapid and objective diagnosis. However, limitations such as small cohorts, noise, and difficulties in model comprehension might hinder widespread clinical use. They suggested the necessity of multi-center studies to validate AI's role in symptomatic carotid plaque detection, especially when considering MRI techniques.

The study, [5], developed a machine-learning model to predict the blood flow waveform in the internal carotid artery (ICA). The model was trained on patient data, relying on state-of-the-art Doppler manometry measurements for obtaining accurate results. The study, [6], used machine learning models to detect arterial disease, including stenoses and aneurysms, from peripheral measurements of pressure and flow rates across the network. The study, [7], proposed a model to elucidate individual patients' cerebral circulation using blood flow simulation incorporating clinical data. Their approach enabled obtaining the probability of different outputs, considering patient condition uncertainty. By combining machine learning with blood flow simulation, predictions were performed 43,000 times faster on a desktop computer, allowing real-time surgical risk assessment. Their prediction results revealed the relationship between collateral circulation and life-threatening surgical outcomes.

The study, [8], investigated and compared the performances of different ML techniques for detecting the presence of a carotid disease by analyzing the Heart Rate Variability parameters of opportune electrocardiographic signals selected

from the available databases. The study, [9], introduced a 50-layer residual network as a feature generator for identifying carotid stenosis. The Deep convolutional neural networks classified sonographic images into four categories based on the features related to the ICA blood flow rate. The study, [10], modeled high-fidelity blood flow in understanding CVD. The potential for data-driven patient-specific blood flow modeling in computational and experimental cardiovascular research has been highlighted with an emphasis on the challenges and opportunities in the field. The study, [11], combined both statistical and machine learning methods to reduce information redundancy. By this, they could enhance accuracy in disease diagnosis. Their study developed Graph theory-inspired ML models to identify significant features for prediction models.

Motivated by the capability of ML algorithms to gain valuable insights from data, this research endeavors to employ these algorithms on medical datasets generated from the simulations of the numerical model for the human carotid artery developed in this study. The model used an incompressible Newtonian fluid representation for blood and considered the artery as an elastic pipe, allowing for altering its material properties during simulations facilitated by the COMSOL software. Machine-learning regression models were trained following the development of the numerical model for an idealized human carotid artery and data generation.

The structure of this article is as follows: Section 2 introduces the problem identification and research objectives. In Section 3, a comprehensive explanation of the adopted methodology is provided. Sections 4 and 5 present the research results, along with their interpretation. Finally, Section 6 presents the conclusive findings of the research, highlighting its limitations, practical applications, and potential future research directions in the field.

2 Problem Formulation

While most of the studies in the field of medical data analysis concentrated on medical image classification or hemodynamic prediction by utilizing available clinical datasets on the carotid artery, this work takes a different approach. It integrates Computational Fluid Dynamics (CFD) with data analysis, aiming to analyze datasets generated from the developed models simulating the carotid artery under various anatomical and physiological conditions. The primary focus of this study is on translational research, aiming to predict blood flow characteristics accurately while avoiding the need for invasive or costly methods such as MRI and Ultrasound. By addressing this, the study aims to fulfill the requirement for non-invasive techniques in evaluating vascular changes in the carotid artery, as highlighted in the work by, [12].

A quantitative and experimental approach is being proposed, with computer-based simulations utilized to achieve this goal. The research question is defined as descriptive and seeks an answer to the following:

Is it feasible for Machine Learning models trained on simulated data to accurately detect the presence of atherosclerosis in the carotid artery?

2.1 Research Objectives

Carotid artery blood flow is paramount in maintaining cerebral perfusion and brain health. Accurate carotid artery blood flow prediction is valuable in understanding cerebrovascular diseases and guiding clinical interventions. Therefore, the primary focus of this research is to explore the efficacy of machine learning models in predicting carotid artery blood flow and to assess the effectiveness of the quantile loss function in improving prediction accuracy and robustness.

1. To develop and validate a numerical model for the human carotid artery.
   - Simulate the model for both healthy and unhealthy conditions.
   - Compute the wall shear stress (WSS) and compare their values with the clinical results.

2. To estimate WSS in the carotid artery using ML algorithms.
   - Develop ML models that best fit the data.
   - Compute the metrics and quantile loss function on the train and test datasets to identify the most reliable ML prediction model.

3. Since WSS acts as an initial and independent indicator of atherosclerotic changes, [12], the goal is to forecast these changes based on the calculated estimates of WSS.

Exposures and Outcomes:

The independent variables in this study encompass fluid-related parameters such as blood density and viscosity, flow-related parameters like the inlet pressure and outlets' blood velocities, and material parameters such as the artery's density, Bulk

Modulus, and Poisson ratio. The outcome to be measured is the WSS in all three arterial segments.

3 Methodology

This section presents a detailed, step-by-step procedure from data collection to data analysis designed to accomplish the proposed objectives.

3.1 Prepare Data Collection Instrument

The carotid artery is a vital supplier of blood to the face, neck, and brain. Its main vessel, the Common Carotid Artery (CCA), undergoes bifurcation into the External Carotid Artery (ECA) and Internal Carotid Artery (ICA), as depicted in Figure 1. The ECA primarily supplies blood to the face and neck, whereas the ICA delivers blood to the brain.

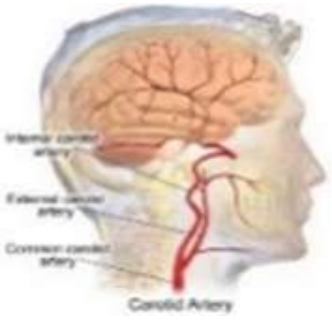

Fig. 1: Anatomy of Carotid Artery

This study has formulated a numerical model to accurately depict the blood flow dynamics in the carotid artery. The model incorporates essential elements of the blood circulatory system, where the blood is represented as a Newtonian fluid, and the artery is treated as an elastic circular pipe with bifurcation. This model was implemented using the COMSOL Multiphysics software, enabling the simulation of the blood flow problem within an idealized artery, as illustrated in Figure 2. Data collection has relied on clinical data about the artery's anatomy and the physiological behavior of blood flow to ensure realism.

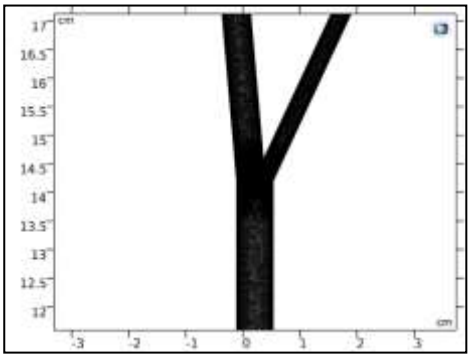

Fig. 2: Geometry of the Artery in COMSOL

The current work focused on the stationary Navier-Stokes equation and implemented it using the laminar flow interface. Viscosity and density values were determined based on available human blood data to ensure accuracy. The elastic behavior of the artery was characterized using anatomical data specific to the carotid artery.

The P1-P1 linear finite element Galerkin method was chosen for the discretization process for effectively handling velocity and pressure variables. To create a physics-based mesh, 40637 triangular elements and 4192 quadrilateral elements were employed, resulting in 76758 degrees of freedom. Among these, 51172 degrees of freedom were utilized to determine the velocity, while the remaining 25586 degrees of freedom were used for the pressure variables. For more detailed mesh parameters, please refer to Table 1.

Table 1. Parameters of the Extremely Fine Mesh

Number of elements	44829
Number of vertex elements	17
Number of edge elements	2466
Average element quality	0.8004
Minimum element quality	0.08677
Mesh area	15.76cm^2

Newton's method has been employed to tackle the resulting non-linear flow problem, which proved effective in finding a solution. Zero initial and no-slip boundary conditions were enforced on the boundary walls, along with no backflow phenomena. The PARDISO solver has been used to efficiently solve the system, greatly facilitating the computation process and leading to reliable results. Additionally, we activated the stability settings within the software to guarantee the numerical stability of the generated Galerkin finite element model. Furthermore, convergence of solutions for a given set of model parameters was ensured throughout the data collection process.

Table 2 presents the input parameters used in the COMSOL model builder to simulate the human carotid arterial blood flow, [13], [14], [15], [16], [17]. Values corresponding to human blood characteristics are assigned to the model parameters, density, and viscosity to represent blood as a Newtonian fluid. Regarding the elastic properties of the carotid artery, the density, bulk modulus, and Poisson ratio are assigned values relevant to the human carotid artery, as indicated in Table 2. The dimensions of the artery, including length, diameter, and other relevant measurements, are derived from information provided in the literature.

Furthermore, Table 2 and Table 3 provide essential ranges for the inlet pressure and output velocities necessary for generating realistic blood flow patterns in the simulation. These ranges are crucial in achieving biologically meaningful results in the model.

Tables 3 (a) and (b) present the data necessary for conducting simulations. The inlet pressure is set to 100, 120, and 130 mm Hg. As for the blood velocities at outlet 1 (ICA) and outlet 2 (ECA), they are assigned values within the specified range (normal or healthy), as indicated in the eighth row (third and fourth columns) of Table 2.

Table 2. Data on Characteristics of Blood and Carotid Artery, [13], [14], [15], [16], [17]

Fluid (blood) Properties		**Artery Properties**	
Density (kg/m^3)	1060	Density	960
Viscosity (Pa. s)	0.004	Bulk Modulus ($N/m^{2)}$	1.2×10^8
		Poisson Ratio	0.45
ARTERY	**CCA**	**ICA**	**ECA**
Diameter (mm)	6.10 ± 0.8	4.8 ± 0.3	3.0 ± 0.6
Length (cm)	13.6 ± 1.2	8.6 ± 1.4	8.6 ± 1.4
Velocity (m/sec)	-	0.187-0.295	0.121-0.185
Bifurcation Angle (36 ±11)°			

Table 3(a). Dimensions of the Simulated Artery

ARTERY	**CCA**	**ICA**	**ECA**
Diameter (mm)	6.2	4.8	3.0
Length (cm)	14.2	9.0	9.0

Table 3(b). Data for running simulations

InPr (mm Hg)	**Out1vel (ICA) (m/s)**	**Out2vel (ECA) (m/s)**	**D (kg/m^3)**	**Vis (Pa. s)**	**Bifurcation Angle (degrees)**
100	0.241	0.153	1060	0.0035	30°
120	0.193	0.122	1075	0.004	
130	0.1205	0.077		0.0045	

3.2 Review the Collected Data for Quality and Completeness

The data types and counts of the numerical model parameters in segment CCA are summarized in Table 4. It is evident from the table that there is no missing data in this segment. Similar findings were observed in the other two segments, ECA and ICA.

Exposures:
Vis: Blood viscosity (Pa.s)
D: Blood Density (kg/m^3)
InPr: Blood pressure at the CCA entrance (mm Hg)
Out1Vel: Blood velocity at exit 1, i.e., ICA (mm Hg)
Out2Vel: Blood velocity at exit 2, i.e., ECA(mm Hg)

Outcomes:
Vel_CCA: Blood velocity in CCA(m/s)
Vel_ICA: Blood velocity in ICA(m/s)
Vel_ECA: Blood velocity in ECA(m/s)
WSS_CCA: Wall Shear Stress in CCA (Pa)
WSS_ICA: Wall Shear Stress in ICA (Pa)
WSS_ECA: Wall Shear Stress in ECA (Pa)

Table 4. Information on numerical model Parameters from Python Code
dtypes: float64(9), int64(3) , memory usage: 8.6 KB

	Column	**Non-null Count**	**Dtype**
0	Vis	90non-null	float64
1	D	90non-null	int64
2	InPr	90non-null	int64
3	Out1vel	90non-null	float64
4	Out2vel	90non-null	float64
5	Vel_CCA	90non-null	float64
6	Vel_ICA	90non-null	float64
7	Vel_ECA	90non-null	float64
8	WSS_CCA	90non-null	float64
9	WSS_ICA	90non-null	float64
10	WSS_ECA_	90non-null	float64

3.3 Data Management

Following the format presented in Table 5, the data has been systematically collected from the numerical model. Each set of exposures was evaluated using the COMSOL software, and the corresponding results were recorded in their respective columns within the table.

Table 5. Data collection format

	Exposures:					Outcomes:					
Sl. No	Vis	D	InPr	Out1vel	Out2vel	Vel_CCA	Vel_ICA	Vel_ECA	WSS_CCA	WSS_ICA	WSS_ECA
1											
2											

3.4 Data Analysis

The first step is to compute the statistics of exposures to get some basic information on the exposures. Table 6 provides a comprehensive breakdown of this data.

Table 6. Descriptive Statistics of Exposures

Statistics	Vis	D	InPr	Out1vel	Out2vel
count	90	90	90	90	90
mean	0.0042	1062	112.2	0.1736	0.1102
std	0.0004	5.0	11.4	0.0675	0.0427
min	0.0035	1060	100	0.0723	0.046
25%	0.004	1060	100	0.1205	0.077
50%	0.004	1060	120	0.193	0.122
75%	0.0045	1060	120	0.241	0.153
max	0.0005	1075	130	0.241	0.153

3.5 Data Validation

Following the completion of the previous stage, the subsequent step involves data validation. During this phase, the reliability of the developed numerical model as a data source is thoroughly assessed and evaluated. This critical process ensures the accuracy and trustworthiness of the generated data for further analysis and interpretation. The current research also examined scenarios where the blood flow in ICA or ECA experienced a reduction due to conditions like atherosclerosis. Consequently, states with 20%, 50%, or 70% reductions in blood velocity at the outlets were analyzed; these findings would shed light on the impact of various health conditions on blood flow distribution within the carotid artery network.

For validating the model, WSS values were computed for the CCA under healthy conditions and found in the interval (0.850 Pa to 3.464 Pa), which is towards the right of the mean, 0.850 +/- 0.195 Pa, as reported by [18].

The subsequent section introduces an in-depth analysis of the collected (simulated) data utilizing advanced ML algorithms. The flow chart in Figure 4 illustrates the step-by-step process employed during this analysis, offering a clear and systematic overview of the entire procedure.

4 Results and Discussions for WSS in CCA

This section presents plots depicting the shear stress in the carotid artery for specific parameters. These visualizations offer a comprehensive understanding of the flow dynamics within the artery. Subsequently, we delve into data analysis using machine learning (ML) algorithms. Figure 3 depicts shear stress contours when InPr is 100, out1vel =0.241, out2vel is 0.077, D =1060, and Vis=0.0045. The initial step involved is conducting a univariate analysis of the features within the dataset. This analysis encompassed an exploration of their range and central tendency. Furthermore, descriptive statistics were thoroughly examined to identify any potential outliers in the data. This meticulous scrutiny determined that noteworthy disparities between the exposures' maximum and 75th percentile values were absent. This absence of outliers underscores the robust nature of the dataset, rendering it well-suited for subsequent analysis. Following the comprehensive univariate analysis of the features, a parallel assessment was executed on the target variables using bar charts. The results of this evaluation indicated an absence of skewness in the data distribution, thereby signifying a well-balanced distribution of the target variables.

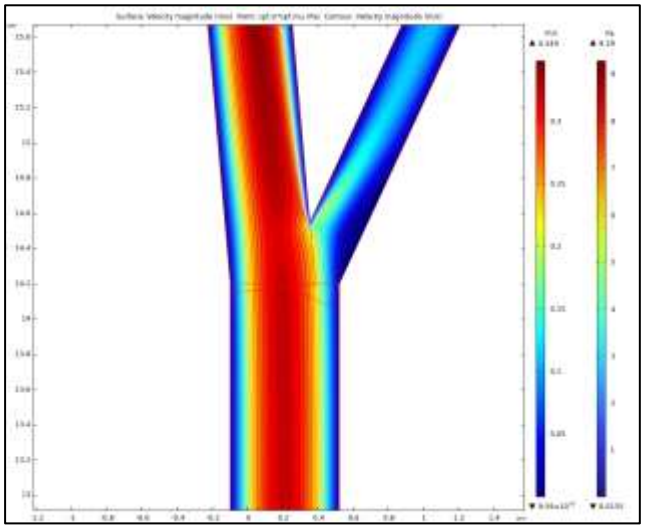

Fig. 3: Shear Stress Contour

Moving forward, the next phase involved conducting a multivariate analysis. This phase encompasses the exploration of interdependencies among multiple dependent variables or features concerning an outcome or target variable. This approach facilitates a more holistic comprehension of the intricate relationships embedded within the dataset. Each feature underwent computation of its Variance Inflation Factor (VIF) score to gauge the extent of correlation. This process allowed for assessing the degree of correlation among the features. Notably, the analysis revealed that the features exhibited a modest level of correlation, as evidenced by VIF scores below 5 for each feature. Subsequently, the first model, namely the linear regression model, was developed using Vis, D, InPr, Out1vel, and Out2vel as features and WSS_CCA as the target variable.

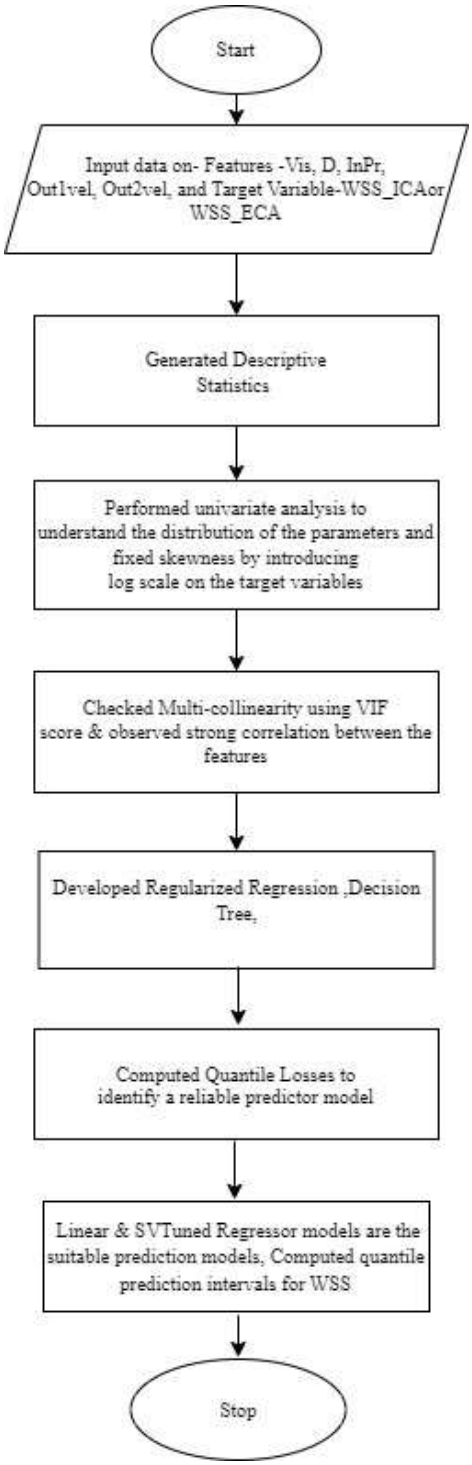

Fig. 4: Flow Chart

Various metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Square Error), $R^{2,}$ and adj-R^2 scores were calculated for both train and test data sets and presented in Table 7. The assumptions on the linear regression model have been carefully examined and deemed satisfactory. As a result, the linear regression model appears to be well-suited for the dataset. Nonetheless, in addition to this, non-linear regression models, ensembled, and boosting models were also applied and evaluated.

A Python code was developed to evaluate the metrics of polynomial regressors with various degrees. To determine the optimal-fit polynomial, graphs depicting the Mean Squared Error (MSE) versus the polynomial degree were plotted for both the training and testing data, as illustrated in Figure 5.

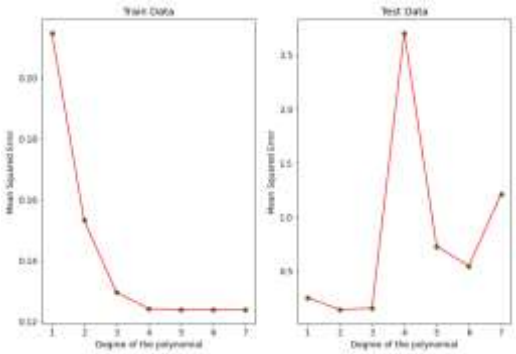

Fig. 5: Graphs depicting MSE vs. degree of the polynomial

The analysis of these graphs showed that a third-degree polynomial exhibited the best performance as a non-linear prediction model for the given data. In the subsequent steps, the SVR (Support Vector Regressor), RFR (Random Forest Regressor), Adaboost, Gradient boosting, and XG boost algorithms were utilized, and their corresponding metrics were calculated and displayed in Table 7.

Observations:

1. The RMSE on the test dataset is the lowest for the polynomial model, showing that this model can predict the target value most accurately.
2. MAE is the average error between the predicted and actual values and is the minimum for the polynomial model.
3. A significant reduction in MAE was observed after hyper-tuning the SVR model using the RBF (Radial Basis Function) kernel with the parameters C=100 and gamma=0.01. This improved performance
4. R^2 for the training dataset is significantly higher for the SVR, decision tree, RFR, RFR-tuned, and all the boosting models, showing that all these models have over-fitted the data.
5. MAPE is 1.1 for the SVR-tuned model, which means, on average, the prediction is off by 1%.

Table 7. Metrics of Regressor Models

Regressor Models	Data Set	RMSE	MAE	R^2	Adj-R^2	MAPE
Linear	**Train**	0.2148	0.1829	0.9525	0.9488	-
	Test	0.2559	0.2373	0.9405	0.9157	-
Poly degree 3	**Train**	0.1295	0.0768	0.9827	0.9814	-
	Test	0.1574	0.0983	0.9775	0.9681	-
Support Vector	**Train**	0.1961	0.1412	0.9603	0.9573	8.2001
	Test	0.2913	0.2194	0.9229	0.8907	7.4285
Support Vector Tuned	**Train**	0.1640	0.1191	0.9723	0.9702	9.0159
	Test	0.1764	0.1241	0.9717	0.9599	1.1255
Decision Tree	**Train**	0.3523	0.2884	0.8770	0.8677	38.0824
	Test	0.4239	0.3322	0.7948	0.7093	-27.4147
Random Forest	**Train**	0.1336	0.0938	0.9823	0.9810	17.5722
	Test	0.2481	0.1907	0.9297	0.9004	-29.4441
Random Forest (Tuned)	**Train**	0.1383	0.1002	0.9810	0.9796	17.6054
	Test	0.2491	0.1945	0.9292	0.8997	-32.1883
Ada Boost	**Train**	0.2281	0.1887	0.9484	0.9445	29.3324
	Test	0.2935	0.2300	0.9017	0.8607	-33.2171
Gradient Boosting	**Train**	0.1177	0.0660	0.9863	0.9852	15.7419
	Test	0.2186	0.1615	0.9454	0.9227	-24.6019
XG Boost	**Train**	0.1104	0.0441	0.9879	0.9870	14.7957
	Test	0.2697	0.1939	0.9169	0.8823	-31.9448

While MAE or MAPE plays a crucial role in identifying acceptable regressor models, it is equally essential to have a more comprehensive evaluation performed to ascertain the most dependable regressor model, [19]. By integrating the quantile loss function, prediction uncertainty can be adequately considered, and deeper insights can be gained into the performance of the models. Moreover, with this comprehensive approach, reliable predictions can be made on unfamiliar datasets.

During the crucial final stage, the main objective was to quantify the prediction uncertainty associated with the model. The Quantile loss function was utilized to accomplish this and provide valuable insights into the degree of uncertainty surrounding the point estimation. The outcomes for the 0.1, 0.5, and 0.9 quantiles are illustrated in Figure 6, offering a comprehensive view of the models' performance concerning uncertainty. The plot analysis revealed that the linear Regression model exhibited the lowest quantile loss, signifying its superior capability to capture prediction uncertainty effectively.

Developing a Python code to implement the quantile linear regressor model also enabled the computation of interval estimates for the target variable. The results of these interval estimates are presented in Table 8. These results provide insights into the efficacy of the linear regression model in capturing the possible prevalence of atherosclerotic conditions in ICA. Notably, the WSS values highlighted in Table 8 were not in the range associated with healthy cases, as detailed in section 3.5. However, the identification of a similar condition in the ECA proved to be less successful.

5 Results and Discussions for ICA and ECA

Figure 7 (Appendix) presents the flow chart for data analysis generated for ICA and ECA. In contrast to the case of CCA, where only the fluid and blood vessel properties act as exposures, here we assumed that the features and target variables in CCA and ECA are the features of ICA. Similarly, the features and target variables of CCA and ICA constitute the feature set for ECA. This assumption is considered appropriate for the present problem, as any stenotic conditions in one of the segments impact the blood flow characteristics in the other carotid artery segments.

As a consequence of this assumption, multicollinearity was observed among the feature sets of both ICA and ECA. Consequently, regularized regression models, including Ridge, Lasso, and Elasticnet, were applied to the dataset. Additionally, the dataset was subjected to a Decision tree, Random forest, and boosting algorithms to determine the most suitable predictor model for assessing the WSS in both bifurcated segments. This process also aided in identifying the significant features of the target variables WSS_ICA and WSS_ECA. It was found that WSS_CCA and Vel_CCA had the most significant impact on WSS_ICA, whereas Vel_CCA and blood pressure in CCA influenced WSS_ECA, see Table 9(a) & Table 9(b).

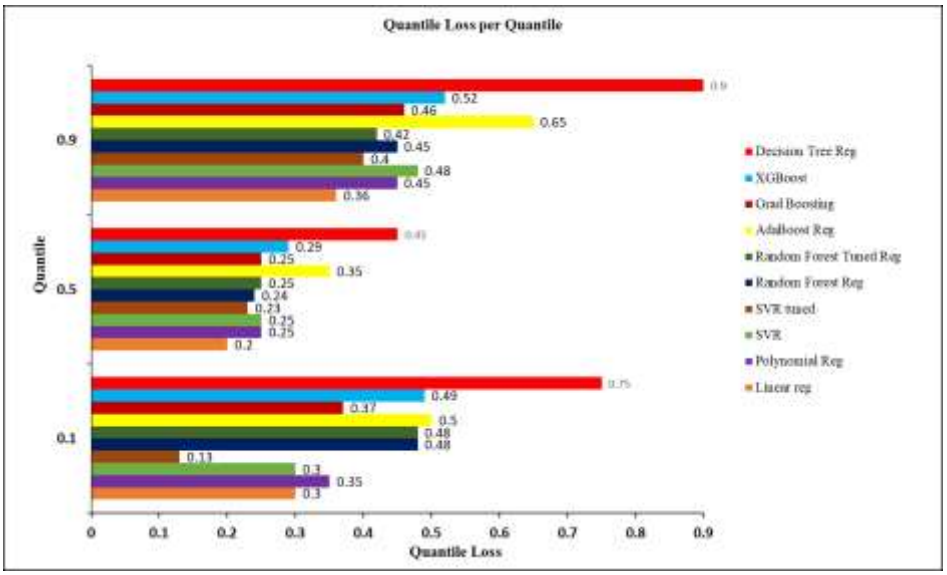

Fig. 6: Plot of Quantile Loss function for the ML developed to predict

Table 8. Quantile predictions of WSS_CCA

Exposures					Target Variable (WSS_CCA)			
Vis	**D**	**InPr**	**Out1vel**	**Out2vel**	**Actual**	**0.1 Quantile**	**0.5 Quantile**	**0.9 Quantile**
0.004	1060	100	**0.1202**	**0.153**	**0.679323**	**0.660288**	**0.684708**	**0.845187**
0.004	1060	100	**0.0717**	**0.153**	**0.638733**	**0.490863**	**0.476632**	**0.638733**
0.0045	1060	100	**0.0717**	**0.153**	**0.576934**	**0.576934**	**0.571987**	**0.765744**
0.0035	1060	100	0.1931	0.153	0.853178	0.829056	0.902331	1.028715
0.0045	1060	120	0.2414	0.046	1.128829	1.000332	1.128829	1.18801
0.005	1060	120	0.2414	0.153	1.61143	1.28635	1.458091	1.61143
0.0035	1060	100	0.2414	0.077	0.85576	0.85576	0.943403	1.023779
0.0035	1060	100	**0.0717**	**0.153**	**0.454918**	**0.404791**	**0.381278**	**0.511722**
0.0045	1075	120	**0.1202**	**0.077**	**0.640726**	**0.640726**	**0.640726**	**0.640726**
0.004	1060	130	**0.0717**	**0.046**	**0.336452**	**0.336452**	**0.336452**	**0.336452**
0.004	1060	100	0.1931	0.122	0.874334	0.857199	0.929918	1.06985
0.004	1060	120	0.2414	0.046	1.033475	0.91426	1.033475	1.061
0.005	1060	120	**0.0717**	**0.153**	**0.693363**	**0.693363**	**0.729826**	**0.88884**
0.0045	1060	120	**0.0717**	**0.153**	**0.634472**	**0.607292**	**0.634472**	**0.761829**
0.004	1060	100	**0.1202**	**0.077**	**0.518568**	**0.51827**	**0.518568**	**0.634654**
0.004	1060	100	0.2414	0.046	1.064914	0.883903	0.97099	1.064914
0.004	1060	120	0.2414	0.122	1.271533	1.056278	1.199615	1.271533
0.0045	1060	100	0.2414	0.046	0.969974	0.969975	1.066344	1.191925

Table 9(a). Important Features for WSS_ICA Identified by Gradient Boosting Model

Feature Importance (ICA)		
0	Out1vel	0.834239
1	WSS_CCA	0.064805
2	Vel_CCA	0.058523
3	Vel_ECA	0.020121
4	log WSS_ECA	0.011729
5	Out2vel	0.008123
6	Vis	0.001929
7	InPr	0.000529
8	D	0

The reliable predictor model for all the target variables was identified in the subsequent steps. The quantile loss function was plotted for all the ML models developed to predict the target variables. The plot of the quantile loss function for the models developed to predict WSS in ICA is shown in Figure 8 (Appendix), while Figure 9 (Appendix) displays the same for WSS in ECA.

Table 9(b). Important Features for WSS_ECA Identified by Gradient Boosting Model

Feature Importance (ECA)		
0	InPr	0.500923
1	Out2vel	0.292016
2	Vel_CCA	0.142237
3	log WSS_ICA	0.02052
4	Vel_ICA	0.019159
5	WSS_CCA	0.014632
6	Vis	0.008388
7	Out1vel	0.002126
8	D	0

These graphs indicate that the most reliable predictor model is the Gradient Boosting algorithm. Consequently, the quantile intervals were computed using the Gradient Boosting Quantile algorithm for WSS_ICA and WSS_ECA, and they are presented in Table 10 (Appendix) and Table 11 (Appendix).

The importance of recognizing these observations relies on a numerical model of an idealized carotid artery, formulated with specific assumptions detailed in this paper's methodology section, which cannot be overstated. Consequently, additional investigation and validation are imperative to corroborate the findings.

5.1 Limitations of this Study

The model proposed in this paper is specifically designed for an idealized carotid artery. Several key assumptions have been considered to ensure numerical tractability and practical feasibility. These assumptions include treating the blood as a Newtonian fluid, assuming the arterial segments to be circular elastic pipes with a constant radius, and considering the blood flow steady. These simplifications enabled efficient computational handling and facilitated reasonably reliable insights into the blood flow dynamics within the carotid artery.

6 Conclusions

This study focused on developing a numerical model to simulate blood flow in the carotid artery under different medical conditions associated with stenosis or atherosclerosis. ML models were tailored to the simulated data and used to predict blood flow characteristics in the artery. Utilizing the quantile loss function designed to assess prediction uncertainty, the Linear regression model better estimates wall shear stress within the CCA. On the other hand, when considering the ICA and ECA, the gradient boosting algorithm emerged as the most effective model for predicting wall shear stress.

Through this study, we identified the blood velocity and the WSS in CCA as the most important features affecting the wall shear stress in ICA, while the blood velocity and the blood pressure affected the WSS in ECA the most. Results also showed that the quantile linear regression model could effectively detect atherosclerotic conditions in ICA for a given set of exposures. But then, such a condition in ECA was not successfully identified. However, it is important to recognize certain limitations in this study, as specific assumptions were made to ensure numerical tractability for the problem. Therefore, the future scope of this work involves extending it to a patient-specific model and utilizing advanced data-analytic tools like ANN to enhance its reliability as a reference source for the medical community. The model's accuracy and applicability would significantly improve for personalized medical assessments and interventions by incorporating patient-specific data.

Acknowledgment:
The authors thank their respective institutions for providing the necessary facilities for this work.

References:

[1] Li X., Liu X., Deng X., & Fan Y., Interplay between artificial intelligence and biomechanics modeling in the cardiovascular disease prediction, *Biomedicines,* 10(9), 2022, 2157. doi:10.3390/biomedicines10092157.

[2] Wang S, Wu D, Li G, Zhang Z, Xiao W, Li R, Qiao A, Jin L and Liu H (2023) Deep learning-based hemodynamic prediction of carotid artery stenosis before and after surgical treatments. *Front. Physiol.* 13:1094743. doi: 10.3389/fphys.2022.1094743.

[3] Zhang J., Han R., Shao G., Lv B., Sun K.. Artificial Intelligence in Cardiovascular Atherosclerosis Imaging. J Pers Med. 2022 Mar 8;12(3):420. doi: 10.3390/jpm12030420.

[4] Miceli G., Rizzo G., Basso M.G., Cocciola E., Pennacchio AR, Pintus C, Tuttolomondo A. Artificial Intelligence in Symptomatic Carotid Plaque Detection: A Narrative Review. Applied Sciences. 2023; 13(7):4321. https://doi.org/10.3390/app13074321.

[5] Fillingham P., Levitt M., Kurt M., Lim D., Federico E., Keen J., Aliseda A., E-177 machine learning model for the prediction of patient-specific waveforms of blood flowthrough the internal carotid artery, *SNIS 19th annual meeting electronic poster abstracts* [Preprint], 2022. doi:10.1136/neurintsurg-2022-snis.288.

[6] Yeh C. Y., Lee H. H., Islam M. M., Chien C. H., Atique S., Chan L., & Lin M. C., Development and validation of machine learning models to classify artery stenosis for automated generating ultrasound report, *Diagnostics*, 12(12), 2022, 3047. doi:10.3390/diagnostics12123047.

[7] Yuhn, C., Oshima, M., Chen, Y., Hayakawa, M., & Yamada, S., Uncertainty quantification in cerebral circulation simulations focusing on the collateral flow: Surrogate Model Approach with machine learning, *PLOS Computational Biology*, 18(7), 2022. doi:10.1371/journal.pcbi.1009996.

[8] Verde, L. and De Pietro, G., A machine learning approach for carotid diseases using heart rate variability features, *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies* [Preprint], 2018. doi:10.5220/0006730806580664.

[9] Lindsey, T. and Garami, Z., Automated stenosis classification of carotid artery sonography using Deep Neural Networks, 2019 18th IEEE International *Conference On Machine Learning and Applications (ICMLA)* [Preprint], 2019a. doi:10.1109/icmla.2019.00302.

[10] Arzani, A. and Dawson, S.T., Data-driven cardiovascular flow modelling: Examples and opportunities, *Journal of The Royal Society Interface*, 18(175), 2021. doi:10.1098/rsif.2020.0802.

[11] Chen Z., Yang M., Wen Y., Jiang S., Liu W., & Huang, H., Prediction of atherosclerosis using machine learning based on Operations Research, *Mathematical Biosciences and Engineering*, 19(5), 2022, pp.4892-4910. doi:10.3934/mbe.2022229.

[12] Diego Gallo, Payam B. Bijari, Umberto Morbiducci, Ye Qiao, Yuanyuan (Joyce) Xie, Maryam Etesami, Damiaan Habets, Edward G. Lakatta, Bruce A. Wassermanand David A. Steinman. Segment-specific associations between local haemodynamic and imaging markers of early atherosclerosis at the carotid artery: An in vivo human study, *Journal of The Royal Society Interface*, 15(147), 20180352, 2018. doi:10.1098/rsif.2018.0352.

[13] S.W.I. Onwuzu, A.C. Ugwu, G.C.E. Mbah, I.S. Elo,, Measuring wall shear stress distribution in the carotid artery in an African population: Computational fluid dynamics versus ultrasound Doppler velocimetry', *Radiography*, 27(2), 2021, 581–588. doi:10.1016/j.radi.2020.11.018.

[14] Marshall, I., Papathanasopoulou, P. and Wartolowska, K., Carotid flow rates and flow division at the bifurcation in Healthy Volunteers, *Physiological Measurement*, 25(3), 2004, 691–697. doi:10.1088/0967-3334/25/3/009.

[15] Fojas, J.J. and De Leon, R.L., Carotid artery modeling using the navier-stokes equations for an incompressible, newtonian and Axisymmetric Flow, *APCBEE Procedia*, 7, 2013, 86–92. doi:10.1016/j.apcbee.2013.08.017.

[16] Ogoh S., Washio T., Paton J. F. R., Fisher J. P., & Petersen L. G., Gravitational effects on intracranial pressure and blood flow regulation in young men: A potential shunting role for the external carotid artery, *Journal of Applied Physiology,* 129(4), 2020, 901–908. doi:10.1152/japplphysiol.00369.2020.

[17] Kamenskiy A. V., MacTaggart J. N., Pipinos I. I., Bikhchandani J., & Dzenis Y. A., Three-dimensional geometry of the human carotid artery, *Journal of Biomechanical Engineering*, 134(6), 2012. doi:10.1115/1.4006810.

[18] Sui B., Gao P., Lin Y., Gao B., Liu L., & An J., Assessment of wall shear stress in the common carotid artery of healthy subjects using 3.0-tesla magnetic resonance, *Acta Radiologica*, 49(4), 2008, 442–449. doi:10.1080/02841850701877349.

[19] Xi Chen, Weidong Liu, Xiaojun Mao, Zhuoyi Yang, Distributed High-dimensional Regression Under a Quantile Loss Function, *Journal of Machine Learning Research* 21, 2020, 1-43.

APPENDIX

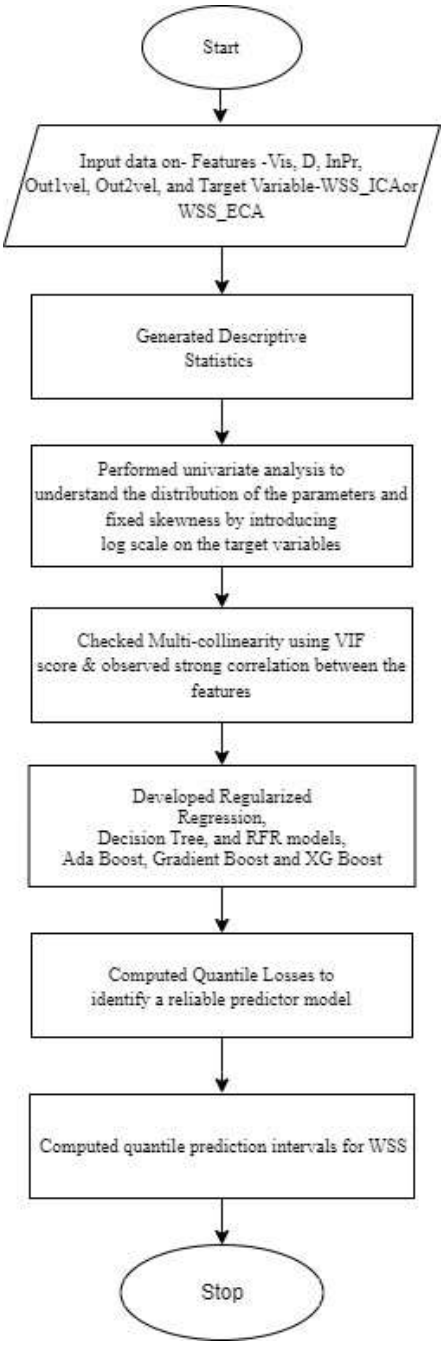

Fig. 7: Flow Chart

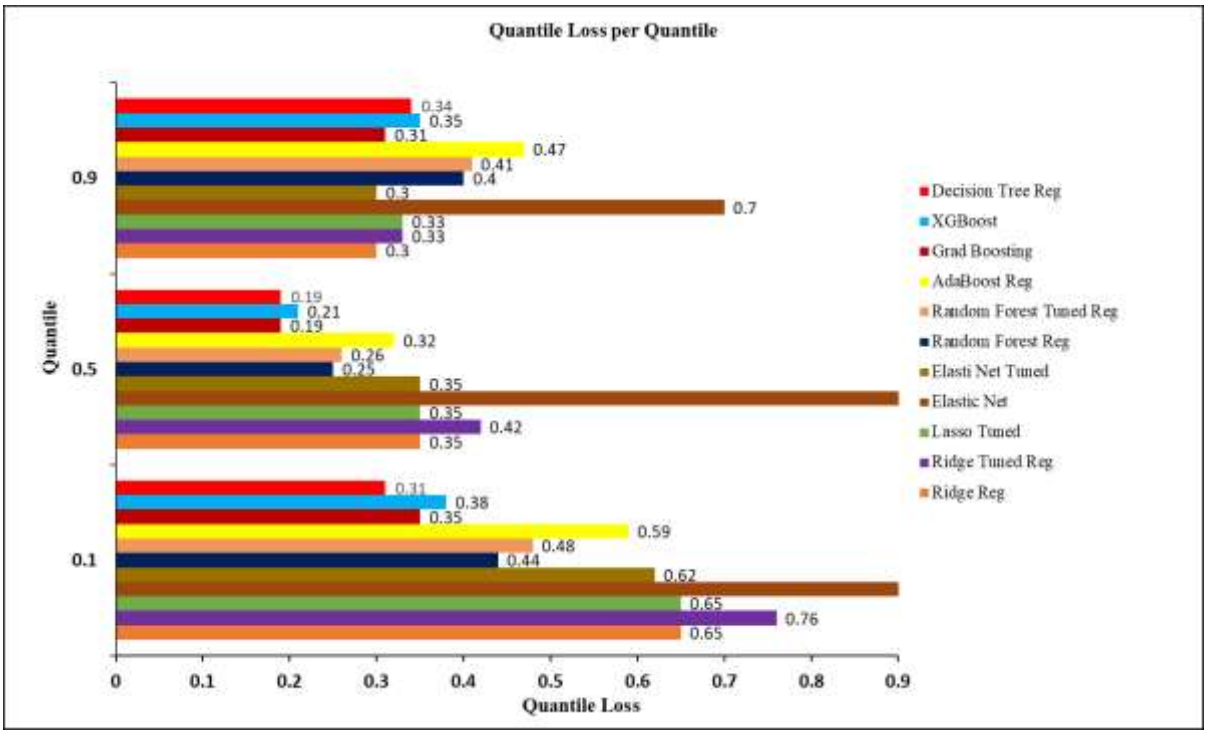

Fig. 8: Plot of Quantile Loss function for the ML developed to predict WSS_ICA

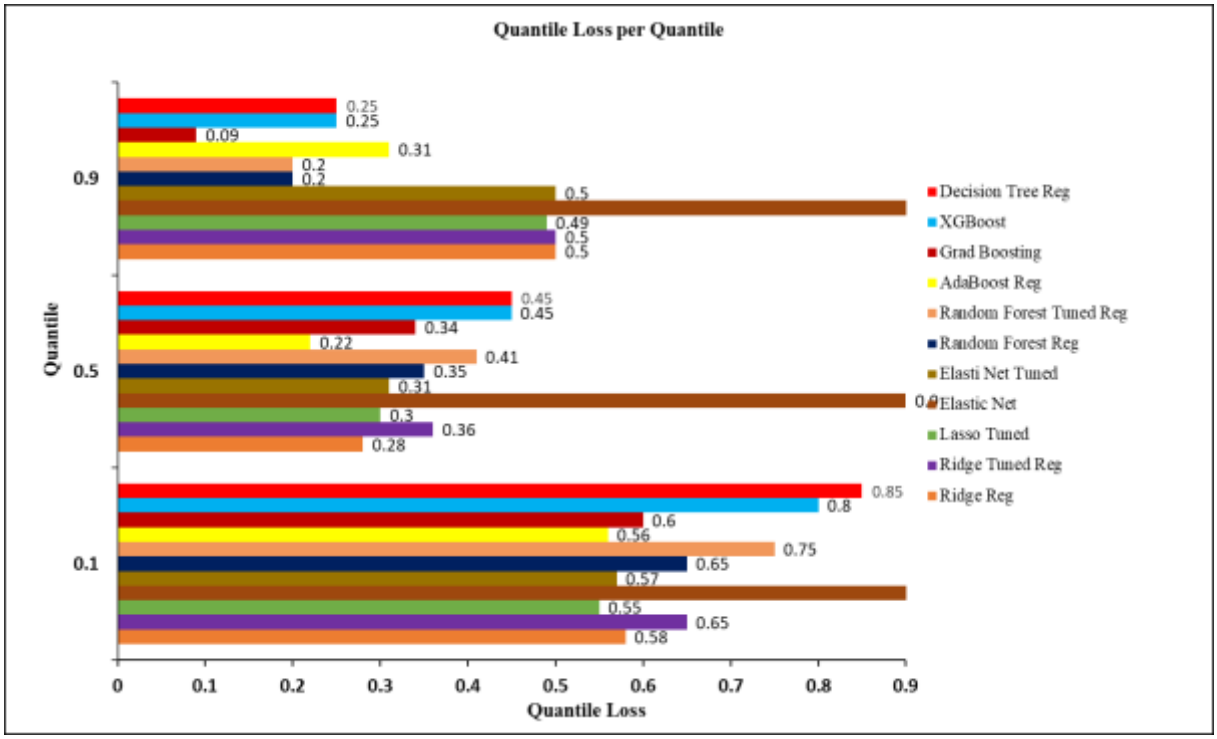

Fig. 9: Plot of Quantile Loss function for the ML developed to predict WSS_ECA

Table 10. Quantile predictions of WSS_ICA

Exposures									**Target Variable (WSS_ICA)**			
Vis	D	InPr	Out1vel	Out2vel	Vel_CCA	Vel_ECA	WSS_CCA	WSS_ECA	Actual	0.1 Quantile	0.5 Quantile	0.9 Quantile
0.0045	1060	100	0.1202	0.077	0.074543	0.088532	0.580231	0.662184	0.226674	0.160194	0.233145	0.280177
0.0045	1075	120	0.1202	0.077	0.075314	0.144331	0.640726	2.1063	0.254805	0.174545	0.228197	0.285173
0.0035	1060	100	0.0717	0.153	0.108956	0.179861	0.454918	2.235889	0.051677	0.06892	0.067079	0.229342
0.0045	1060	100	0.2414	0.077	0.108724	0.219893	1.050035	1.060807	0.600673	0.443052	0.664104	0.644709
0.004	1060	130	0.1931	0.122	0.122154	0.231282	0.949304	5.115034	0.402931	0.344504	0.378675	0.372254
0.0045	1075	120	0.2414	0.122	0.134627	0.267775	1.388943	4.024413	0.536366	0.471021	0.533134	0.553694
0.0045	1060	120	0.2414	0.153	0.15496	0.29696	1.492572	4.715618	0.464327	0.410518	0.465193	0.491097
0.0045	1060	100	0.1202	0.153	0.12139	0.211839	0.756643	2.518693	0.151853	0.130723	0.16931	0.229326
0.0035	1060	100	0.2414	0.046	0.091876	0.193223	0.786211	0.537146	0.557896	0.289502	0.501484	0.508154
0.005	1060	120	0.0717	0.046	0.043588	0.084415	0.391744	1.232727	0.178367	0.144455	0.161412	0.272236
0.0035	1060	100	0.2414	0.153	0.162901	0.294481	1.035599	1.815765	0.348964	0.344433	0.389457	0.479634
0.005	1060	120	0.1931	0.153	0.141493	0.262393	1.328196	4.738597	0.37826	0.323486	0.390723	0.332452
0.0045	1060	120	0.0717	0.046	0.043801	0.084641	0.355445	1.133036	0.159392	0.14206	0.158376	0.291015
0.004	1060	120	0.1931	0.122	0.12296	0.235956	1.029243	3.286256	0.337066	0.355958	0.361726	0.393369
0.005	1060	120	0.1202	0.077	0.074938	0.143758	0.698544	2.289476	0.286675	0.174522	0.230012	0.271204
0.0045	1060	120	0.1202	0.077	0.07527	0.144261	0.639244	2.1008	0.255183	0.16861	0.223877	0.284382
0.004	1060	100	0.2414	0.153	0.155719	0.290189	1.369311	2.47078	0.491992	0.410518	0.433442	0.485729
0.004	1060	100	0.1202	0.153	0.12307	0.2124	0.679323	2.457015	0.131503	0.131523	0.140757	0.229243

Table 11. Quantile predictions of WSS_ECA

Exposures									**Target Variable (WSS_ECA)**			
Vis	D	InPr	Out1vel	Out2vel	Vel_CCA	Vel_ICA	WSS_CCA	WSS_ICA	Actual	0.1 Quantile	0.5 Quantile	0.9 Quantile
0.0045	1060	100	0.1202	0.077	0.074543	0.149824	0.580231	1.742598	0.081353	0.119975	0.128186	0.29356
0.0045	1075	120	0.1202	0.077	0.075314	0.146314	0.640726	1.946125	0.277067	0.260029	0.294618	0.37747
0.0035	1060	100	0.0717	0.153	0.108956	0.14498	0.454918	0.431412	0.295152	0.247443	0.278725	0.369031
0.0045	1060	100	0.2414	0.077	0.108724	0.251491	1.050035	4.373594	0.134003	0.126406	0.134317	0.318121
0.004	1060	130	0.1931	0.122	0.122154	0.235182	0.949304	2.999843	0.709028	0.379176	0.712675	0.698942
0.0045	1075	120	0.2414	0.122	0.134627	0.267429	1.388943	3.930097	0.550008	0.410095	0.54931	0.548763
0.0045	1060	120	0.2414	0.153	0.15496	0.282304	1.492572	3.429711	0.650533	0.435107	0.645038	0.655274
0.0045	1060	100	0.1202	0.153	0.12139	0.187694	0.756643	1.193764	0.33483	0.252674	0.342368	0.358923
0.0035	1060	100	0.2414	0.045	0.091876	0.232939	0.786211	4.078895	0.065181	0.083116	0.078782	0.294
0.005	1060	120	0.0717	0.045	0.043588	0.090326	0.391744	1.389675	0.157107	0.211608	0.16612	0.410847
0.0035	1060	100	0.2414	0.153	0.162901	0.281886	1.035599	2.618967	0.236766	0.209832	0.235724	0.34013
0.005	1060	120	0.1931	0.153	0.141493	0.24371	1.328196	2.826104	0.653889	0.438416	0.613141	0.566112
0.0045	1060	120	0.0717	0.045	0.043801	0.090234	0.355445	1.249647	0.143685	0.134091	0.142067	0.37729
0.004	1060	120	0.1931	0.122	0.12296	0.227379	1.029243	2.534565	0.443784	0.379176	0.48513	0.496302
0.005	1060	120	0.1202	0.077	0.074938	0.146681	0.698544	2.175203	0.302649	0.260029	0.294055	0.411042
0.0045	1060	120	0.1202	0.077	0.07527	0.14636	0.639244	1.948855	0.276301	0.260029	0.294618	0.37747
0.004	1060	100	0.2414	0.153	0.155719	0.289628	1.369311	3.622354	0.328089	0.267149	0.325061	0.496317
0.004	1060	100	0.1202	0.153	0.12307	0.187177	0.679323	1.042109	0.326153	0.260591	0.311447	0.358339

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research at all stages, from formulating the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors declare no conflicts of interest.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US