International Journal of Computer Assisted Radiology and Surgery (2020) 15:141–150
https://doi.org/10.1007/s11548-019-02065-2

ORIGINAL ARTICLE

Comparison of statistical learning approaches for cerebral aneurysm rupture assessment

Felicitas J. Detmer[1] · Daniel Lückehe[2] · Fernando Mut[1] · Martin Slawski[3] · Sven Hirsch[4] · Philippe Bijlenga[5] · Gabriele von Voigt[2] · Juan R. Cebral[1]

Received: 16 January 2019 / Accepted: 29 August 2019 / Published online: 4 September 2019
© CARS 2019

Abstract

Purpose Incidental aneurysms pose a challenge to physicians who need to decide whether or not to treat them. A statistical model could potentially support such treatment decisions. The aim of this study was to compare a previously developed aneurysm rupture logistic regression probability model (LRM) to other machine learning (ML) classifiers for discrimination of aneurysm rupture status.

Methods Hemodynamic, morphological, and patient-related information of 1631 cerebral aneurysms characterized by computational fluid dynamics simulations were used to train support vector machines (SVMs) with linear and RBF kernel (RBF-SVM), *k*-nearest neighbors (*k*NN), decision tree, random forest, and multilayer perceptron (MLP) neural network classifiers for predicting the aneurysm rupture status. The classifiers' accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were evaluated and compared to the LRM using 249 test cases obtained from two external cohorts. Additionally, important variables were determined based on the random forest and weights of the linear SVM.

Results The AUCs of the MLP, LRM, linear SVM, RBF-SVM, kNN, decision tree, and random forest were 0.83, 0.82, 0.80, 0.81, 0.76, 0.70, and 0.79, respectively. The accuracy ranged between 0.76 (decision tree,) and 0.79 (linear SVM, RBF-SVM, and MLP). Important variables for predicting the aneurysm rupture status included aneurysm location, the mean surface curvature, and maximum flow velocity.

Conclusion The performance of the LRM was overall comparable to that of the other ML classifiers, confirming its potential for aneurysm rupture assessment. To further improve the predictions, additional information, e.g., related to the aneurysm wall, might be needed.

Keywords Cerebral aneurysm · Risk factors · Hemodynamics · Shape · Prediction · Machine learning

Introduction

Cerebral aneurysms are a common vascular disease occurring in about 2–3% of the overall population [1, 2]. While most aneurysms remain asymptomatic and never rupture, the event of an aneurysm rupture leads to hemorrhagic stroke, which is associated with high mortality and morbidity and, consequently, a large economic burden [3, 4]. Increased use of diagnostic medical imaging has led to more frequent incidental findings of unruptured aneurysms [5]. The risk that is associated with different treatment options to prevent a future rupture outweighs the natural aneurysm rupture risk of about 1% per year [6]. Therefore, the assessment of a patient's individual aneurysm rupture risk is essential when deciding whether to treat an aneurysm or conservatively observe it by follow-up imaging.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11548-019-02065-2) contains supplementary material, which is available to authorized users.

✉ Felicitas J. Detmer
fdetmer@gmu.edu

1 Bioengineering Department, Volgenau School of Engineering, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

2 Computational Health Informatics, Leibniz University, Hannover, Germany

3 Statistics Department, George Mason University, Fairfax, VA, USA

4 Institute of Applied Simulation, ZHAW University of Applied Sciences, Wädenswil, Switzerland

5 Neurosurgery, Clinical Neurosciences Department, Faculty of Medicine, University of Geneva, Geneva, Switzerland

To address this challenge, we recently developed and evaluated an aneurysm rupture probability model that discriminates between ruptured and unruptured aneurysms based on hemodynamic, morphological, and patient-related information [7]. The model showed a good predictive performance both in the 1631 aneurysms used for model training as well as in the 249 aneurysms of two external datasets [8]. The developed probability model is a logistic regression model that had been trained using logistic group lasso regression [9]. The aim of the current study was to assess whether the prediction could be improved by using other machine learning (ML) classifiers that are trained and evaluated on the same datasets as the group lasso model. Additionally, to identify characteristics of aneurysm with a high or low probability of being ruptured, respectively, important variables for discrimination between these two classes were identified.

Methods

Dataset

For model training and external validation, the same datasets as previously used for the probability model were applied [7, 8]. Briefly, 1631 aneurysms (492 ruptured) from 1061 patients in 5 hospitals in the USA, Japan, and Colombia were used for model training. The external validation was performed with 249 aneurysms (66 ruptured) harbored by 203 patients in two European hospitals (the *AneuRisk* dataset from the Niguarda Ca' Granda Hospital in Milan [10] and the *AneuX* dataset obtained from the Geneva University Hospitals). Computational fluid dynamics (CFD) simulations based on the segmented three-dimensional rotational angiography images had been conducted for each of the aneurysms using an in-house finite element solver[1] [11]. Subsequently, the aneurysm hemodynamic environment was characterized by 22 parameters describing the flow complexity, temporal stability, surface forces, and inflow strengths. Additionally, 25 geometric features were computed capturing the aneurysm size, elongation, and irregularity in shape. Both datasets further contained information about the patient age and gender as well as the aneurysm location in the cerebral vasculature (see Table 3 in Online Suppl. Material for the parameters' definitions). The detailed description of all computed features can be found in [7] and the references therein.

[1] Segmentations of the raw 3D-DRA images for the AneuX dataset were performed using the Aneufuse platform. Data are stored at the Swiss Institute of Bioinformatics and available to the scientific community by written request at adb@itis.ethz.ch.

Training of classifiers

To compare the previously developed rupture probability model to different ML approaches, several classifiers were trained on the data of the 1631 aneurysms characterized by the 50 features mentioned above. The classifiers included *k*-nearest neighbors (kNN), decision trees, random forests (RFs), and support vector machines (SVMs) with linear and RBF-kernels [12] as well as a simple multilayer perceptron (MLP) with one hidden layer containing 34 neurons with nonlinear sigmoid activation functions and two output neurons with a softmax function following [13]. Table 1 shows an overview of the different classifiers. For all the classifiers but the MLP, certain tuning parameters (see Table 1) were determined using fivefold (single) cross-validation on the training data and choosing the parameter value that maximizes the mean area under the ROC curve (AUC) over all folds. The grid of tuning parameters is shown in Table 1 in Online Suppl. Material. For decision trees, RFs, and the MLP, the training and evaluation of the classifiers was repeated 100 times to account for the stochastic nature of these classifiers.

Evaluation

For all six classifiers, the AUCs were evaluated in the training and test datasets (see Table 1 for the definition of the continuous probability measures). Furthermore, the sensitivity, specificity, and accuracy were calculated. For the logistic regression model, these metrics had been previously computed using a threshold for classification of cases based on the receiver operating characteristic (ROC) curve [7]. Confidence intervals for the AUCs as well as statistical comparisons of the ROC curves between different classifiers were determined based on the asymptotically exact method of DeLong et al. [15] implemented in the pROC R-package [16].

Variable importance

A further aim was to identify relevant features for the determination of the aneurysm rupture status. RFs and linear SVMs enable the assessment of the prediction strength of each feature in an established and interpretable manner based on the weights determining the separating hyperplane for the SVM and "feature importance" in case of the RF. Here, "feature importance" was defined based on the mean reduction of impurity of the tree nodes where the feature to be assessed is used as the splitting criterion (Gini importance) [17]. The larger the reduction of impurity, the more the feature reduces uncertainty when classifying an aneurysm as ruptured or unruptured and hence the more

Table 1 Overview of different classifiers trained for prediction of aneurysm rupture status based on [12]

Classifier	Brief description	Tuning parameter	Probability measure
K-nearest neighbors (kNN)	Classifies unknown patterns based on the labels of the *k* neighboring patterns. In case of different labels, a majority vote is taken	*k*-number of neighboring patterns used for classification	Relative number of *k* neighboring patterns with positive labels
Decision tree	Builds a binary tree in which each node splits the dataset based on a condition for one feature. An unknown pattern is classified based on the leaf in which it ends up	Maximum depth of tree	Fraction of patterns of the same class in a leaf
Random forest	Uses multiple decision trees to classify an unknown pattern. Thereby, each tree uses only a subset of the original dataset. Based on the results of the individual trees a majority vote is taken	Maximum depth of trees and number of trees	Mean predicted class probabilities of the trees in the forest (class probability of a single tree defined as for decision tree)
Linear SVM	Separates the feature space with a hyperplane. Unknown patterns are classified based on which side of the plane they are located	Cost parameter (C) indicating how much the separating hyperplane can be violated by the training data	Probability from Platt scaling based on predicted class [14]
SVM with RBF kernel	Like a SVM with linear kernel, but due to the RBF kernel, nonlinear separations are possible	Cost parameter (C) indicating how much the separating hyperplane can be violated by the training data and γ-parameter of kernel function	Probability from Platt scaling based on predicted class [14]
Multilayer perceptron (MLP)	Feed-forward artificial neural network with few hidden layers that can be represented using a network diagram (see, e.g., [13]) and is able to learn nonlinear functions	–	Outcome of softmax function [12] at output neuron

Table 2 Comparison of accuracy measures for different classifiers in the training data

Classifier	AUC	Accuracy	Sensitivity	Specificity
Group Lasso [7]	0.855	0.786	0.774	0.791
SVM (linear), $C=0.1$	0.852	0.817	0.600	0.911
SVM (RBF), $\gamma=0.0001$, $C=1000$	0.860	0.823	0.626	0.908
kNN, $k=26$	0.839	0.786	0.437	0.937
Decision Tree, max depth = 4	0.821 ± 0.000	0.798 ± 0.000	0.467 ± 0.000	0.941 ± 0.000
Random Forest, max depth = 6, nr. of trees = 130	0.939 ± 0.001	0.870 ± 0.003	0.679 ± 0.008	0.952 ± 0.003
MLP	0.873 ± 0.001	0.827 ± 0.002	0.633 ± 0.009	0.910 ± 0.004

The tuning parameters were selected based on cross-validation. For the decision tree, RF, and MLP, the shown values are the mean ± standard deviation obtained from 100 repetitions of training the classifier

Table 3 Comparison of accuracy measures for different classifiers in the test data

Classifier	AUC	Accuracy	Sensitivity	Specificity
Group Lasso	0.824	0.767	0.758	0.770
SVM (linear)	0.805	0.791	0.485	0.902
SVM (RBF)	0.812	0.787	0.470	0.902
kNN	0.756	0.767	0.364	0.913
Decision tree	0.702 ± 0.000	0.755 ± 0.000	0.348 ± 0.000	0.902 ± 0.000
Random forest	0.793 ± 0.005	0.780 ± 0.008	0.378 ± 0.021	0.925 ± 0.008
MLP	0.826 ± 0.003	0.786 ± 0.005	0.484 ± 0.016	0.895 ± 0.007

"important" it is. The training of the RF classifier was repeated 100 times and the mean Gini indices of reduction of node impurity were noted for each of the features. For a quantitative comparison of the rankings based on the feature importance for each classifier, Spearman's footrule distance (SFD) was computed [18].

All ML-processing was implemented in Python 3.5.3 using scikit-learn and TensorFlow [19, 20].

Results

Performance of classifiers

Tables 2 and 3 show the accuracy metrics for the different classifiers when evaluated in the training and test data, respectively. To evaluate the classifiers, the performance in the test data is essential. For the purpose of completeness, we report the accuracy metrics in the training data as well in Table 2. In the test data, the AUC was marginally larger for the MLP (0.826, CI [0.768, 0.883]) compared to the group lasso (0.824, CI [0.766, 0.882]), followed by the RBF-SVM (0.812, CI [0.749, 0.875]), linear SVM (0.805, CI [0.740, 0.869]), and RF (0.793, CI [0.731, 0.854]). It was notably lower for the kNN and decision tree with AUCs of 0.756 (CI [0.688, 0.823]) and 0.702 (CI [0.626, 0.777]), respectively. ROC curves were significantly different between the group lasso and linear SVM ($p=0.02$), kNN ($p=0.007$), and decision tree ($p<0.001$) as well as between the RBF-SVM and kNN ($p=0.02$) and decision tree ($p=0.002$), between the linear SVM and decision tree ($p=0.004$), decision tree and RF ($p=0.002$) and MLP and decision tree ($p=0.01$; see Fig. 1 and Table 2 in Supplementary Material).

Based on the default thresholds for classification, the accuracy was largest for the linear SVM with 0.791 and lowest for the decision tree with 0.755. The sensitivity ranged between 0.348 (decision tree, specificity of 0.902) and 0.758 (group lasso, specificity of 0.770), and the specificity between 0.770 (group lasso) and 0.925 (RF, sensitivity of 0.378).

Important variables

For identifying important features based on the weights of the linear SVM, the obtained weights were normalized by the maximum weight of all features. The features with the ten largest relative weight in absolute value are shown in Table 4. Figure 2 and Table 3 in Suppl. Material show a comparison of the importance of variables based on repeated cross-validation for the group lasso model [7], normalized weights of the linear SVM, and the normalized Gini importance obtained from the RF classifier, where a higher rank refers to a larger importance (Fig. 2) and the two categorical variables were excluded for this comparison. There was no clear correlation between the different importance metrics for most of the features. However, the L2-norm of mean

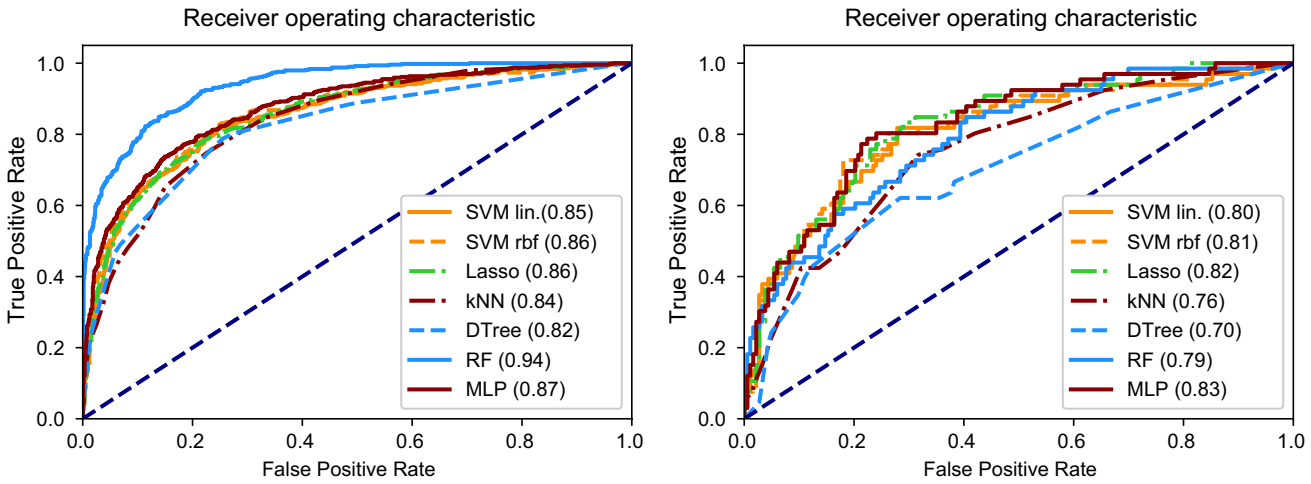

Fig. 1 ROC curves of different classifiers for training (left) and test data (right). The numbers in parentheses refer to the corresponding AUCs. DTree = decision tree

Table 4 Features with ten largest normalized weights of the linear SVM

Feature	Normalized weight
position_ACOM (anterior communicating artery)	1.00
position_PCOM (posterior communicating artery)	0.60
position_ICA-CAV (cavernous segment of internal carotid artery)	−0.46
MLN (mean surface curvature)	0.41
VOR (volume-to-ostium ratio)	−0.38
position_VA (vertebral artery)	0.37
Asize (aneurysm size)	0.36
position_ICA-BIF (bifurcation of the internal carotid artery)	−0.34
position_ICA-OPH (Internal carotid artery-ophthalmic segment)	−0.32

surface curvatures (MLN) and the peak velocity (Vmax) were among the most important variables for each of the three measures.

Aneurysm location was important both based on the weights of the SVM (the three most important features were anterior communicating artery (ACOM), posterior communicating artery (PCOM), and cavernous internal carotid artery (ICA-CAV) location) and the repeated cross-validation for the lasso model (inclusion in 100% of the cross-validation models [7]). Continuous variables with large importance for the SVM and group lasso (> 0.7 of the respective metric) included MLN, volume-to-ostium ratio (VOR), aneurysm width (Awidth), and Vmax.

Based on the Gini importance obtained from the RF classifier, the ten most important variables were isoperimetric ratio (IPR), non-sphericity index (NSI), Gaussian surface curvature (GLN), size ratio (SizeR), position_ACOM, MLN, ellipticity index (EI), height-to-width ratio (HWR), and the two aspect ratios (AR and Aspect). The mean and standard deviation of the Gini importance for all the features are shown in Table 4 in Suppl. Material.

When comparing the variable importance rankings based on the three classifiers, SFDs were 686, 668, and 706 for group lasso vs. RF, RF vs. linear SVM, and linear SVM vs. group lasso, respectively. Based on the asymptotically normal distribution of SFD, these distances were smaller than the expected value of 767.67 given a simple chance model, but these differences were not significant (sd = 70.89, p values of 0.25, 0.16, and 0.38, respectively).

Discussion

The assessment of a patient's aneurysm rupture risk is essential when deciding on treatment of incidental aneurysms. While a plethora of risk factors has been suggested in the literature [21], only a few statistical models have been proposed for assessing the aneurysm rupture risk or

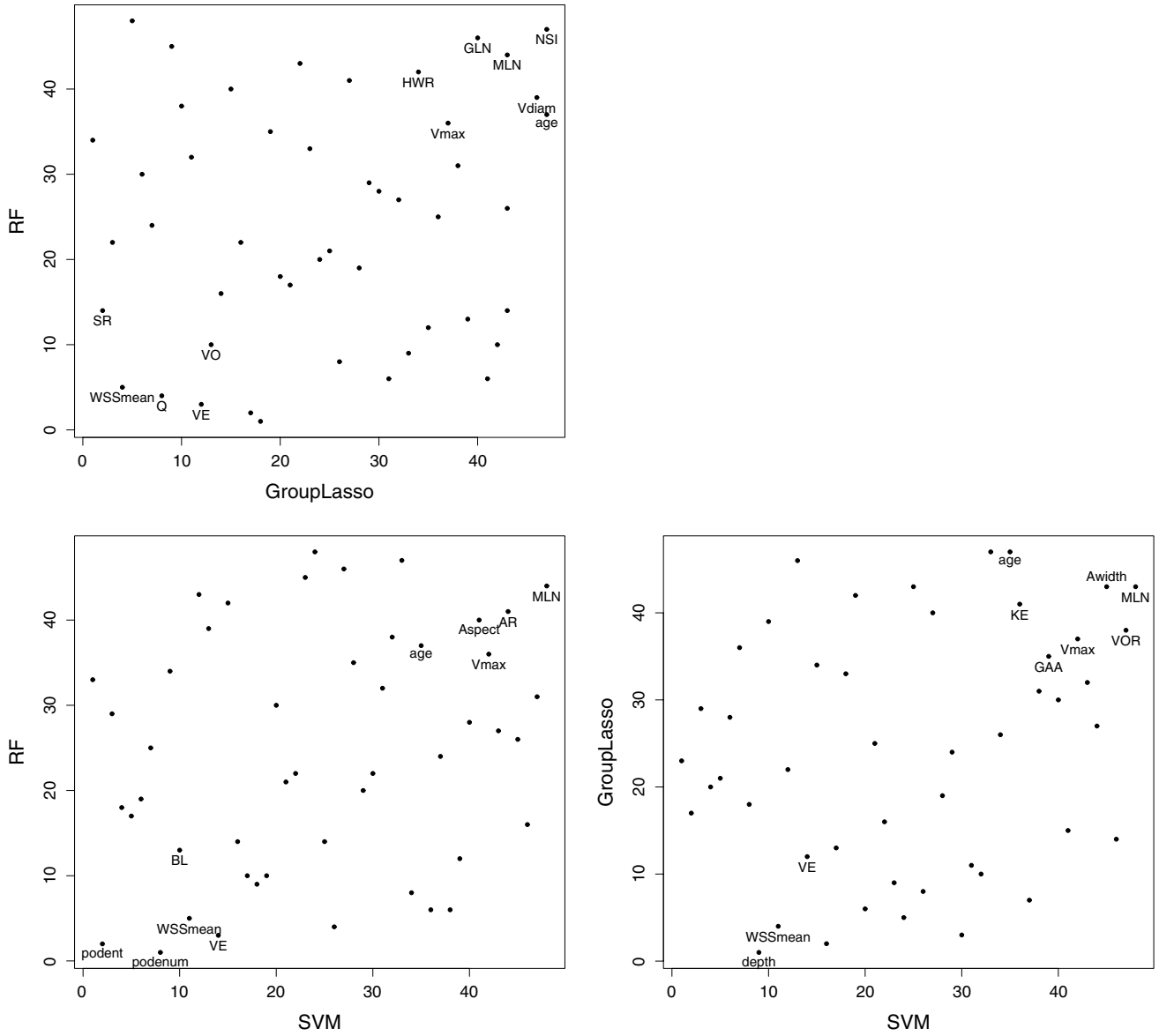

Fig. 2 Comparison of importance metrics of the different classifiers. The axes of the scatterplots show the rank of each feature based on the respective importance measure, where higher numbers refer to a larger importance. The most and least important features for both of the compared methods are indicated by their labels

discriminating between ruptured and unruptured aneurysms [22–25]. The former do not include hemodynamic or morphological information [22, 23], whereas the latter are based on small sample sizes of less than 250 cases [24, 25]. To address this issue, we recently developed a logistic regression model for rupture status prediction based on a large patient cohort including various proposed risk factors from different domains (hemodynamics, morphology, patient characteristics) [7]. This model achieved good results both in its internal and external validation [8]. The current study shows that the predictive performance of this logistic regression model was also comparable to most and better than some of the other ML classifiers.

Comparison of different classifiers

Overall, the discriminatory ability of the classifiers was similar for the MLP, group lasso model, linear SVM, RBF-SVM, and the RF with AUC values between 0.79 and 0.83, and the largest AUCs for the MLP and group lasso model. The accuracy metrics were likewise similar, with the linear SVM achieving the largest value based on the given

threshold for classification. Generally, all the trained classifiers except for the group lasso had a low sensitivity in the test data (between 0.35 for the decision tree and 0.49 for the linear SVM). This finding could be explained by the unbalanced classes in the training data, where 492 of the 1631 aneurysms were ruptured (30.17%). For evaluating an aneurysm's rupture probability, it would be particularly important to have a statistical model with a high sensitivity since high-risk aneurysms that are falsely classified as having a low risk of a future rupture present a hazard for the patient if leaving untreated. For improving the sensitivity of the classifiers, a different threshold of the continuous probability measure (see Table 1) could be applied, e.g., instead of taking the majority class of the leave in the decision tree, one could classify an aneurysm as ruptured already if 30% of the samples in the leave are ruptured.

The ROC curve of the group lasso model was significantly different from the linear SVM, kNN, and decision tree, indicating a superior discriminatory ability compared to these three classifiers. It was not significantly different from the one of the SVM-RBF, RF, and MLP, implying a similar performance. Likewise, the confidence intervals of the computed AUCs were relatively large, suggesting that a larger test sample size would be needed for improving the estimation of the AUC.

Characteristics associated with ruptured and unruptured aneurysms

When looking at important variables for classifying an aneurysm as ruptured or unruptured, both based on the weights of the linear SVM and the relative frequency of variable inclusion in the cross-validation models for the lasso model, the aneurysm location in the cerebral vasculature was particularly influential. This finding is consistent with the previous studies, assigning comparatively high rupture risks to aneurysms in the posterior circulation as well as ACOM aneurysms [23]. In accordance with that, the weights of the SVM were positive for the ACOM, PCOM, anterior cerebral artery (ACA), vertebral artery (VA), and BA-TIP, and negative for all other positions.

Interestingly, the Gini importance as an importance measure for the features of the RF was relatively low for all dummy features of aneurysm location except for the ACOM. One possible explanation for this observation could be that the nodes involving aneurysm location might have always split only a small subset of aneurysms located at that particular artery (see Fig. 1 in [7] for the distribution of aneurysms by location in the training set). This might have led in a low reduction of impurity of the nodes resulting overall in a small variable importance.

Among the continuous features, the mean surface curvature (MLN), aneurysm width, volume-to-ostium ratio, and maximum velocity were frequently selected (lasso model) and had large weights (linear SVM). Of those, MLN and maximum velocity were also important based on the RF. Figure 3 illustrates four aneurysms whose values of their parameters had the largest distance from the separating hyperplane of the linear SVM. The two aneurysms at the left show the two aneurysms from the training data that were obviously classified as ruptured and unruptured, respectively. The two cases on the right display the same information for the test data. As can be seen from the figure as well as Table 5, the cases classified as unruptured were exposed to lower flows (lower Vmax), had a less complex shape (lower MLN), and were located at the internal carotid artery (ICA). They further had a high VOR, indicating that the large ratio of the aneurysm volume to the aneurysm neck decreased the probability of being ruptured. These findings are consistent with the previous ones based on the lasso model [7], which was also supported by the similarly high and low predicted probabilities of the lasso model for the four cases.

Interestingly, the only influential hemodynamic variable according to the lasso model and linear SVM was the maximum velocity, which was also the most important hemodynamic variable based on the RF importance. Different hemodynamic variables have previously been associated with rupture [26]. The finding that mainly morphological features had large values of the defined variable importance metrics could potentially be explained by the association between hemodynamic and geometric variables, where hemodynamics both results from and influences geometry through biomechanical signaling mechanisms in the vessel wall [27].

At the same time, the computed SFDs showed that the feature importance rankings of the three classifiers were not "significantly similar", indicating that a potential randomness of the agreement between rankings (e.g., for MLN and Vmax) cannot be excluded. This finding could possibly be explained by redundancy of the features, i.e., certain features containing similar information and not being ranked consistently among the different classifiers. For example, for the RF, IPR was the most important variable, but one of the least important ones for the group lasso. IPR is largely linearly correlated to NSI (Pearson correlation coefficient of 0.99 in the training data), which could explain why IPR was rarely included in the group lasso cross-validation models. The observed importance of NSI for aneurysm rupture status discrimination is consistent with the literature, where NSI has previously been associated with aneurysm rupture [28, 29].

Additionally, it is important to note that the definitions of "variable importance" using the group lasso, linear SVM, and RF were based on highly different metrics (relative frequency of inclusion in cross-validation models versus weights of trained classifier versus reduction of impurity when splitting sample at respective nodes). Since each of

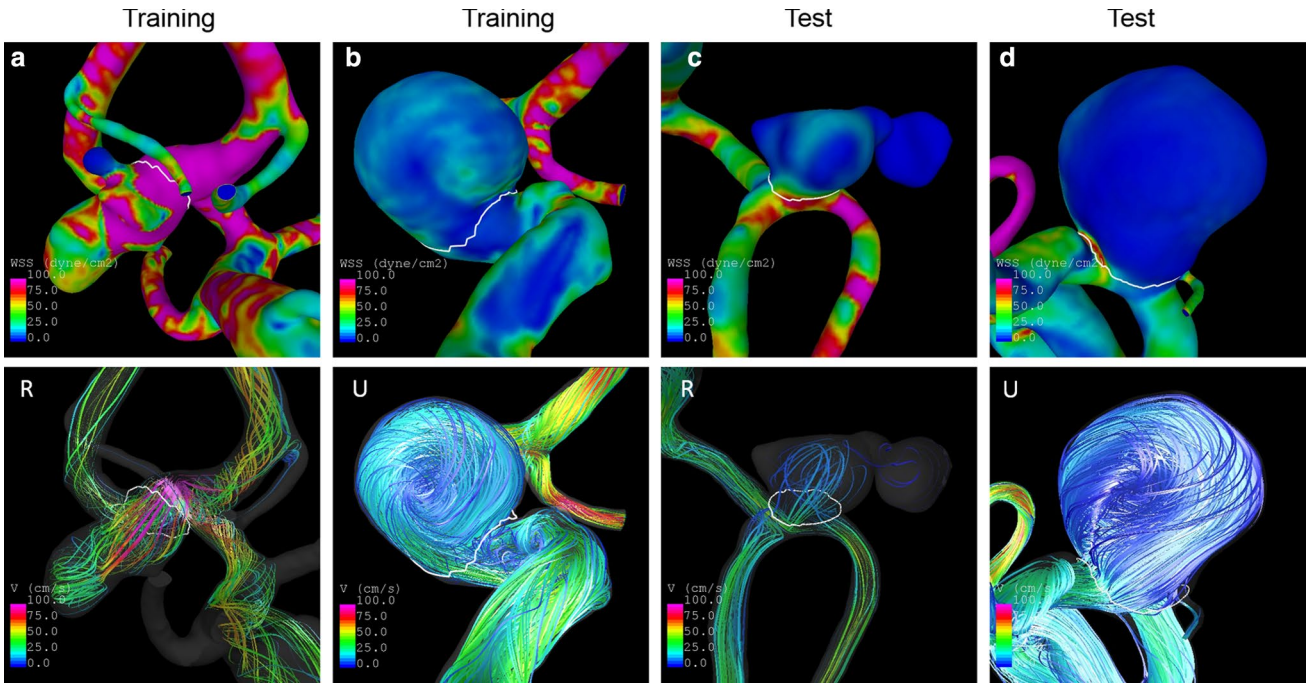

Fig. 3 Illustration of aneurysms with largest distance from hyperplane of linear SVM. The top and bottom row show the wall shear stress distribution and streamlines, respectively, at half of the cardiac cycle. The two aneurysms at the left are cases from the training data, while the two aneurysms at the right are test cases. Selected parameter values of the four cases are shown in Table 5

Table 5 Characteristics of aneurysms illustrated in Fig. 3

Case	Rupture Status	Classification SVM	Position	MLN	VOR	Asize	Vmax	Predicted probability group lasso model
a	R	R	ACOM	0.442824	0.549904	0.766406	282.374	0.9809
b	U	U	ICA-CAV	0.36357	0.928894	1.55401	51.5257	0.0025
c	R	R	ACOM	0.449617	0.867911	1.03828	56.1019	0.9121
d	U	U	ICA-OPH	0.389818	3.99754	1.68195	46.7027	0.0154

the classifiers pursues a different "strategy" for classification, a common value of "importance" for all the features is hence not to be expected. In contrast to the RF, both the linear SVM and the group lasso model assume linear relations between the independent and the dependent variable(s). This could also explain the differences in identified variables between the three classifiers, which becomes clear in Fig. 2.

The aspect of feature redundancy needs to be assessed more in future for yielding a better understanding of important features for aneurysm rupture status discrimination.

Clinical aspects

The results show that the predictive performance of the "best" evaluated classifiers was comparable to the group lasso model. All but the linear SVM differ from the group lasso model by the fact that they do not assume linear associations between the features and the outcome. Furthermore, decision trees and RFs allow for taking interactions into account. The fact that different classifiers performed overall similarly indicates that for improving rupture assessment based on statistical models further, additional information might be needed. This hypothesis is further supported by the fact that none of the classifiers achieved an AUC close to 100%. Rupture of an aneurysm occurs when the forces exerted on the vessel wall exceed the vessel wall strength. Information about the vessel wall is not included in our current data and only indirectly taken into account by the morphological parameters assuming that the shape of the aneurysm is related to different properties of the vessel wall. Hence, it might be possible to improve the predictive performance of the model in future by including information about the vessel wall, e.g., about its thickness, orientation of fibers, or inflammation. Additionally, some genetic

risk factors have been identified whose incorporation might also improve the rupture assessment [30]. Finally, clinical parameters like hypertension have been associated with an increased risk of hemorrhagic stroke and were not included in our models [23].

The presented results justify the use of the group lasso model for aneurysm rupture assessment. Compared to the other classifiers, it has the advantage that its coefficients can be interpreted in an easier way compared to, e.g., the MLP or trained RF. At the same time, different settings of MLPs could be explored more in future to potentially improve its current predictive performance.

Limitations

The data that were used for training and evaluation of the ML classifiers as well as the group lasso model are subject to a selection bias [7, 8]. Only aneurysms of patients that underwent angiographic imaging were included in the study. Hence patients who died of an aneurysm rupture before reaching the hospital as well as those undergoing only computed tomography angiography (CTA) or magnetic resonance angiography (MRA) were not part of the study population.

The differences in the performance of the classifiers could be in part explained by overfitting (especially for the RF). At the same time, the training and test data were obtained from different populations (USA vs. European), and the vessel lumen segmentation of the 3DRA images as a first step for the CFD simulations was performed by independent researchers using a different approach compared to the segmentation of the training data (see [8] for details). These differences could also have contributed to the slightly poorer performance of the lasso, MLP, and SVM models in the test data.

For assessing the performance of a neural network for the aneurysm rupture status classification problem, a MLP with a pre-defined configuration [24] was trained. Alternatively, a deep neural network (DNN) could have been considered. When exploring this option and training a DNN with four hidden layers having 100 neurons each, the DNN was largely overfit with an AUC of 1 (CI [1,1]) in the training and 0.708 (CI [0.6631,0.8016]) in the test data. These results indicate that more data might be needed and other configurations should be considered to achieve an accurate prediction with a DNN, which will be part of future work.

As previously discussed [7], the lasso model as well as the trained ML classifiers were trained and evaluated in cross-sectional data. The lasso model yields the predicted probability of an aneurysm being ruptured. Using the model for risk assessment of a future aneurysm rupture is thus based on the assumption that aneurysms with a high rupture risk resemble those that have already ruptured, which needs to be confirmed. To that aim, the evaluation of the model with longitudinal data is planned in future. Furthermore, an eventual translation of the model into clinical practice would require providing physicians with a tool to apply the model, which will also be part of future work.

Conclusions

The predictive performance of the trained ML classifiers overall was comparable to the group lasso model, with only the MLP (not significantly) outperforming the lasso model. This finding justifies the consideration of the sparse logistic regression model for aneurysm risk assessment. It also indicates that for further improvement of the model, additional information about the aneurysm and patient may be needed. According to the trained classifiers, ruptured aneurysms are characterized by a more complex shape and stronger flows. Additionally, aneurysms at certain locations like the ACOM or PCOM have a higher chance of being ruptured.

Future work aims at assessing the incorporation of other variables such as aneurysm vessel wall enhancement. Furthermore, an evaluation of the model with longitudinal data is planned.

Acknowledgements The authors would like to thank Olivier Brina, Norman Juchler, Rafik Ouared, Diana Sapina, and Mari Cruz Villa-Uriol for helping with the collection and image segmentation of the AneuX data.

Funding SH and PB were supported by SystemsX.ch project AneuX evaluated by the Swiss National Science Foundation. Data for the AneuX dataset was collected and processed in the context of the @neurIST project funded by the EU commission (IST-2004-027703) and AneuX project evaluated by the Swiss National Science Foundation and funded by the SystemsX.ch initiative (MRD 2014/261).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Raw 3D-DRA of the AneuX test data set were provided by the University Hospital of Geneva and collected with formal patient consent according to the @neurIST protocol and ethics authorization PB_2018-00073 (previously CER 07-05) released June 1st 2007 and renewed April 13th 2010, August 19th 2014 and February 28th 2018 initially by the Geneva Cantonal Ethics Commission for Research involving Humans and renewed by Swissethics in 2018.

References

1. Rinkel GJ, Djibuti M, van Gijn J (1998) Prevalence and risk of rupture of intracranial aneurysms: a systematic review. Stroke 29:251–259

2. Vlak MH, Algra A, Brandenburg R, Rinkel GJ (2011) Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. Lancet Neurol 10:626–636. https://doi.org/10.1016/s1474-4422(11)70109-0
3. Rivero-Arias O, Gray A, Wolstenholme J (2010) Burden of disease and costs of aneurysmal subarachnoid haemorrhage (aSAH) in the United Kingdom. Cost Eff Resour Alloc 8:6. https://doi.org/10.1186/1478-7547-8-6
4. Wang G, Zhang Z, Ayala C, Dunet DO, Fang J, George MG (2014) Costs of hospitalization for stroke patients aged 18–64 years in the United States. J Stroke Cerebrovasc Dis 23:861–868. https://doi.org/10.1016/j.jstrokecerebrovasdis.2013.07.017
5. Gabriel RA, Kim H, Sidney S, McCulloch CE, Singh V, Johnston SC, Ko NU, Achrol AS, Zaroff JG, Young WL (2010) Ten-year detection rate of brain arteriovenous malformations in a large, multiethnic, defined population. Stroke 41:21–26. https://doi.org/10.1161/STROKEAHA.109.566018
6. Juvela S, Poussa K, Lehto H, Porras M (2013) Natural history of unruptured intracranial aneurysms: a long-term follow-up study. Stroke 44:2414–2421. https://doi.org/10.1161/strokeaha.113.001838
7. Detmer FJ, Chung BJ, Mut F, Slawski M, Hamzei-Sichani F, Putman C, Jiménez C, Cebral JR (2018) Development and internal validation of an aneurysm rupture probability model based on patient characteristics and aneurysm location, morphology, and hemodynamics. Int J Comput Assist Radiol Surg. https://doi.org/10.1007/s11548-018-1837-0
8. Detmer FJ, Fajardo-Jiménez D, Mut F, Juchler N, Hirsch S, Pereira VM, Bijlenga P, Cebral JR (2018) External validation of cerebral aneurysm rupture probability model with data from two patient cohorts. Acta Neurochir 160:2425–2434. https://doi.org/10.1007/s00701-018-3712-8
9. Meier L, Van De Geer S, Bühlmann P (2008) The group lasso for logistic regression: Group Lasso for Logistic Regression. J R Stat Soc Series B Stat Methodol 70:53–71. https://doi.org/10.1111/j.1467-9868.2007.00627.x
10. Sangalli LM, Secchi P, Vantini S (2014) AneuRisk65: a dataset of three-dimensional cerebral vascular geometries. Electron J Stat 8:1879–1890. https://doi.org/10.1214/14-EJS938
11. Cebral JR, Castro MA, Appanaboyina S, Putman CM, Millan D, Frangi AF (2005) Efficient pipeline for image-based patient-specific analysis of cerebral aneurysm hemodynamics: technique and sensitivity. IEEE Trans Med Imaging 24:457–467
12. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer, New York
13. Liu J, Chen Y, Lan L, Lin B, Chen W, Wang M, Li R, Yang Y, Zhao B, Hu Z, Duan Y (2018) Prediction of rupture risk in anterior communicating artery aneurysms with a feed-forward artificial neural network. Eur Radiol 28:3268–3275. https://doi.org/10.1007/s00330-017-5300-3
14. Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 10:61–74
15. DeLong E, DeLong D, Clarke-Pearson D (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845. https://doi.org/10.2307/2531595
16. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011) pROC: an open-source package for R and S+to analyze and compare ROC curves. BMC Bioinform 12:77
17. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Boston
18. Diaconis P (1988) Group representations in probability and statistics. Institute of Mathematical Statistics, Hayward
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
20. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2016) TensorFlow: large-scale machine learning on heterogeneous systems. arXiv:1603.04467
21. Kleinloog R, de Mul N, Verweij BH, Post JA, Rinkel GJE, Ruigrok YM (2017) Risk factors for intracranial aneurysm rupture: a systematic review. Neurosurgery. https://doi.org/10.1093/neuros/nyx238
22. Tominari S, Morita A, Ishibashi T, Yamazaki T, Takao H, Murayama Y, Sonobe M, Yonekura M, Saito N, Shiokawa Y, Date I, Tominaga T, Nozaki K, Houkin K, Miyamoto S, Kirino T, Hashi K, Nakayama T, Unruptured Cerebral Aneurysm Study Japan Investigators (2015) Prediction model for 3-year rupture risk of unruptured cerebral aneurysms in Japanese patients. Ann Neurol 77:1050–1059. https://doi.org/10.1002/ana.24400
23. Greving JP, Wermer MJ, Brown RD, Morita A, Juvela S, Yonekura M, Ishibashi T, Torner JC, Nakayama T, Rinkel GJ, Algra A (2014) Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: a pooled analysis of six prospective cohort studies. Lancet Neurol 13:59–66. https://doi.org/10.1016/s1474-4422(13)70263-1
24. Xiang J, Yu J, Snyder KV, Levy EI, Siddiqui AH, Meng H (2016) Hemodynamic-morphological discriminant models for intracranial aneurysm rupture remain stable with increasing sample size. J Neurointerv Surg 8:104–110. https://doi.org/10.1136/neurintsurg-2014-011477
25. Bisbal J, Engelbrecht G, Villa-Uriol M-C, Frangi AF (2011) Prediction of cerebral aneurysm rupture using hemodynamic, morphologic and clinical features: a data mining approach. In: Hameurlain A, Liddle SW, Schewe K-D, Zhou X (eds) Database and expert systems applications. Springer, Berlin, pp 59–73
26. Cebral JR, Raschi M (2013) Suggested connections between risk factors of intracranial aneurysms: a review. Ann Biomed Eng 41:1366–1383. https://doi.org/10.1007/s10439-012-0723-0
27. Meng H, Tutino VM, Xiang J, Siddiqui A (2013) High WSS or low WSS? Complex interactions of hemodynamics with intracranial aneurysm initiation, growth, and rupture: toward a unifying hypothesis. AJNR Am J Neuroradiol. https://doi.org/10.3174/ajnr.A3558
28. Raghavan ML, Ma B, Harbaugh RE (2005) Quantified aneurysm shape and rupture risk. J Neurosurg 102:355–362
29. Dhar S, Tremmel M, Mocco J, Kim M, Yamamoto J, Siddiqui AH, Hopkins LN, Meng H (2008) Morphology parameters for intracranial aneurysm rupture risk assessment. Neurosurgery 63:185–197
30. Zhou S, Dion PA, Rouleau GA (2018) Genetics of intracranial aneurysms. Stroke 49:780–787. https://doi.org/10.1161/STROKEAHA.117.018152

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.