

Received October 12, 2020, accepted November 1, 2020, date of publication November 9, 2020, date of current version November 20, 2020. *Digital Object Identifier* 10.1109/ACCESS.2020.3036692

Machine Learning Application With Quantitative Digital Subtraction Angiography for Detection of Hemorrhagic Brain Arteriovenous Malformations

JIA-SHENG HONG¹, CHUNG-JUNG LIN^{2,3}, YUE-HSIN LIN¹, CHENG-CHIA LEE^{2,4}, HUAI-CHE YANG^{2,4}, LING-HSUAN MENG⁵, TE-MING LIN^{2,3}, YONG-SIN HU^{2,6}, WAN-YUO GUO^{2,3}, WEI-FA CHU³, AND YU-TE WU^D1,^{7,8}

¹Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei 11221, Taiwan

²School of Medicine, National Yang-Ming University, Taipei 11221, Taiwan

³Department of Radiology, Taipei Veterans General Hospital, Taipei 11217, Taiwan

⁴Department of Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, Taipei 11217, Taiwan

⁵Department of Advanced Therapy, Siemens Healthineers, Taipei 11503, Taiwan
⁶Department of Radiology, Taipei Veterans General Hospital, Taoyuan Branch, Taoyuan 330023, Taiwan

⁸Brain Research Center, National Yang-Ming University, Taipei 11221, Taiwan

Corresponding author: Yu-Te Wu (ytwu@ym.edu.tw)

This work was supported in part by the Siemens Healthcare Ltd., (Taiwan Division); in part by the National Yang-Ming University; in part by the Taipei Veterans General Hospital; and in part by the Taiwan's Ministry of Science and Technology under Grant 108J029, Grant V106-C-058, and Grant MOST-106-2314-B-010-015-MY2.

ABSTRACT Clinical features are the primary measures used for risk assessment of cerebrovascular diseases. However, clinical features, especially angioarchitecture, in digital subtraction angiography require further interpretation by specialized radiologists. This approach for risk assessment requires multivariable analysis and is, therefore, challenging when completed manually. In this study, we employed three machine learning models, namely the random forest, naïve Bayes classifier, and support vector machine, for the detection of hemorrhagic brain arteriovenous malformations using digital subtraction angiography. Quantitative measurements from digital subtraction angiography were used as features, and the chi-squared test, minimum redundancy maximum relevance, ReliefF, and two-sample t tests were used for feature selection. Bayesian optimization was conducted to optimize the hyperparameters of the three models. The random forest model outperformed the other two models. As a human control, three radiologists diagnosed an independent testing data set. The random forest model had a computation time of less than a second for the whole data set for classification. Accuracy and the area under the receiver operating characteristic curve were 92.7% and 0.98 for the training data set and 85.7% and 0.97 for the independent testing data set, respectively. Compared with the mean diagnosis time of approximately half a minute per patient and the highest accuracy of 76.2% for the three radiologists, the random forest model was faster and more accurate for our data set. These results suggest that the machine learning model based on hemodynamic features from quantitative digital subtraction angiography is a promising tool for detecting hemorrhagic brain arteriovenous malformations.

INDEX TERMS Brain arteriovenous malformation, digital subtraction angiography, machine learning, quantitative analysis, rupture risk, time-density curve.

I. INTRODUCTION

In digital subtraction angiography (DSA), continuous X-ray acquisition with contrast agent injection in the target vessels and subtraction of images without contrast are employed to eliminate all background structures, except for the enhanced

The associate editor coordinating the review of this manuscript and approving it for publication was Hongjun Su.

vasculature. Because it has the highest spatial and temporal resolution among all clinical imaging modalities, it is considered the gold standard for the diagnosis of cerebrovascular diseases, such as brain arteriovenous malformations (BAVMs). A BAVM is an abnormality of the vasculature occurring when an artery directly connects to a vein without a buffer capillary. This abnormality leads to a high-flow shunt and eventually results in the rupture of the vessel,

⁷Institute of Biophotonics, National Yang-Ming University, Taipei 11221, Taiwan

causing intracranial hemorrhage with a high risk of death. This disease is rare but fatal, occurring in 1.42 people (95% confidence interval 1.3–1.6) and being fatal in 0.70 people (95% confidence interval 0.6–0.8) per 100,000 personyears [1]. The precise diagnosis of hemorrhagic BAVMs is crucial for treatment planning. Many treatment strategies are available for BAVMs, including medication, embolization, microsurgery, and stereotactic radiosurgery. A randomized trial revealed that medication is superior to any other intervention therapy for an unruptured BAVM [2]. Therefore, the level of hemorrhagic risk is a critical factor in treatment decisions.

Relevant clinical features, including clinical presentation and angioarchitecture [3], are associated with rupture incidence. Although such clinical features are straightforward, their accurate assessment requires complex computed tomography (CT) or magnetic resonance imaging (MRI) protocols, long scanning time, and experienced radiologists. Besides, intranidal aneurysms and complexed venous outlets are occasionally beyond the scope of CT and MRI. Furthermore, the overall clinical features of individual BAVMs are too complex, with approximately 30 variables, for clinical radiologists to fully comprehend. Additionally, inconsistent results have been obtained regarding the relationship between clinical features and rupture risk. For example, one study revealed the size of the AVM nidus to be associated with intracerebral hemorrhage [4], whereas another indicated that it is not a risk factor [5]. The lack of consistency in previous study results might be caused by differences in the population of the patient cohort, subjective clinical confounders, or radiologists with different levels of expertise. Although angioarchitectural analysis is conventionally used in hemorrhagic risk analysis, it is occasionally criticized for being too subjective, and many years of training are required to master the necessary interpretive skills.

Quantitative DSA (QDSA) is a different approach for assessing risk levels. QDSA defines the salient features from the time-density curve (TDC) of a selected region of interest (ROI) as an objective method for quantifying the hemodynamics of the cerebrovasculature. QDSA provides a surrogate hemodynamic marker and has shown promising results for evaluating the severity of several cerebrovascular diseases, such as dural arteriovenous fistula, carotid stenosis, and BAVMs [6]–[8]. A ruptured BAVM leads to hemorrhage and results in the change of intracranial pressure as well as the hemodynamic conditions. QDSA features can reflect the functional discrepancy between unruptured and ruptured BAVMs. For example, in our previous study [8], we found that the higher the stasis index (SI; a QDSA feature determined by the wash-in slope [WI] divided by the absolute value of the wash-out slope [WO] of the TDC), the higher the level of stagnant outflow. Compared with clinical features, the quantification of temporal information is less intuitive; thus, it has not been commonly used in clinical practice. Nonetheless, QDSA is clinically easier to assess than clinical features; thus, it has more potential for clinical application.

In the clinical setting, it is challenging for radiologists to assess multiple clinical features and hemodynamic QDSA features simultaneously. Machine learning is a powerful approach for managing massive data with high-dimensional features; it has been used to predict the outcomes of several cerebrovascular diseases [9], [10]. To our knowledge, no researcher has used the temporal features of DSA to build machine learning models for the assisted diagnosis of hemorrhagic BAVMs and compared its performance with that of clinical radiologists. In this study, we aimed to detect the ruptured BAVMs by using QDSA features and machine learning algorithms. Furthermore, we compared model performance to the diagnosis by clinical radiologists.

II. STUDY COHORTS AND IMAGING PROTOCOL

This study was approved by the local institutional review board. Cases of 171 patients with BAVMs between 2011 and 2019 were collected retrospectively and randomly separated into the training and testing data sets for the analysis. A total of 150 cases were used in the data set used to train the models, and a separate data set comprising 21 cases was used for testing. No significant differences were found in age, gender ratio, and hemorrhage BAVM percentage between the training and testing data sets (see Table 1). All BAVM diagnoses were confirmed using DSA acquisition. The intracranial hemorrhages caused by the BAVMs were further validated using imaging modalities such as CT and MRI as the gold standard.

 TABLE 1. Demographics of participants in the training and testing data sets.

	Total	Training	Testing
Number of Patients	171	150	21
Mean Age at Acquisition	41.1	37.3	41.1
	(4-80)	(4-80)	(8-61)
Gender	99/72	87/63	12/9
(Male/Female)	(58/42)	(58/42)	(57/43)
BAVM Hemorrhage	94/77	82/68	12/9
(No/Yes)	(55/45)	(55/45)	(57/43)

Values in brackets are the minimum and maximum values for age and percentages for other variables.

Images for all cases were collected using the same scanner (Artis zee, Siemens Healthcare) following a standardized routine protocol. The anterior-posterior and lateral projections were simultaneously acquired because of the biplane design of the scanner. The power injector (Angiomat ILLUMENA, Liebel-Flarsheim) with a 4-French size catheter was preset to 1000 psi and located at the C4-6 vertebral body level for common carotid and vertebral artery injection. In order for standardization, we choose those placed at the C4 level with good image quality and without complications after the examination. At 0.2 s after image acquisition commencement, diluted contrast agent (Omnipaque-350 with the total volume of 12–14 mL) was injected for 1.5 s to form a bolus. Two frame rates were used for acquiring DSA images.

IEEEAccess



FIGURE 1. Quantitative DSA. ROIs in the anterior-posterior and lateral projection (A, B). Schematic of TDC analysis and hemodynamic parameters (C, D).

One was constant at a rate of 6 frames per second (FPS) to the end. Another was dynamic at 7.5 FPS for the initial 5 s, followed by 4 FPS for 3 s, then 3 FPS for 2 s, and finally 2 FPS to the end. The radiographer determined the duration of acquisition based on adequate visualization of the venous structure. Overall acquisition times ranged from 5 to 15 s.

III. QUANTITATIVE DSA

Fig. 1 displays the key features of QDSA analysis. Clinical radiologists delineated the ROI at the nidus of the BAVM for the time-density analysis (Fig. 1A, B). The TDC denotes the average grayscale density within the ROI at each time point (Fig. 1C). It was fitted by the gamma-variate function [11] with the trust-region-reflective least square algorithm [12] to reduce random noise and recirculation caused by overlapping vessels. The form of the fitted function is given as follows:

$$\mathbf{D}(t) = \mathbf{K} \times (t - t_0)^{\alpha} \times e^{-(t - t_0)/\beta}$$

where D is the function of grayscale density with respect to time *t*, *t*₀ is the appearance time, K is the scaling constant, and α and β are the arbitrary parameters. **D** = {*d*₁, *d*₂, ..., *d*_N} corresponds to time **T** = {*t*₁, *t*₂, ..., *t*_N}, where N is the total number of temporal points.

Eight hemodynamic parameters (Fig. 1D) were quantified from the fitted TDC: peak density (PD), bolus arrival time (BAT), time to peak (TTP), WI, WO, SI, full width at half maximum (FWHM), and area under the curve (AUC_{TDC}). They were defined as follows:

PD = max (**D** (t))
BAT = t_a, (**D**(t_a) = 20)
TTP = argmax (**D** (t))
WI = max
$$\left(\frac{d\mathbf{D}}{dt}\right)$$

WO = min $\left(\frac{d\mathbf{D}}{dt}\right)$
SI = WI/ |WO|
FWHM = t_{h2} - t_{h1}, (t_{h2} > t_{h1})
AUC_{TDC} = $\int_{t_1}^{t_N} \mathbf{D}$ (t) dt,

PD is the maximum value of the density. BAT is the time point at which the density reaches $D(t_a)$, which is the arrival

density and is set at 20 arbitrary units. The argmax operator refers to the argument of the maximum in a function, and TTP is the argument, which is the time point at which the maximum value of the fitted function occurs. WI and WO are the increasing and decreasing slopes of the fitted TDC, respectively. FWHM is the duration between the time points at the densities $\mathbf{D}(t_{h1})$ and $\mathbf{D}(t_{h2})$, which are the densities with values that are half of the PD value. AUC_{TDC} is the integral of the fitted function, which is the area under the TDC. SI is calculated by the division of WI by the absolute value of WO [8]. There were two projections with 8 parameters each; thus, the total number of QDSA features was 16 for each participant in the study.

As described in the imaging protocol section, two frame rate types were used: one was constant and the other was dynamic. The original TDC conducted using a constant rate was resampled using a dynamic one; then, curve fitting was applied to both curves. The quantitative parameters derived from the fitted curves using constant and dynamic rates were the same.

IV. MACHINE LEARNING APPROACHES

Feature selection is a prerequisite to building machine learning models, and redundant features used in training may degrade model performance for prediction. Here, the filter type method [13] was used, and features were all normalized to zero mean and unit variance. Three commonly used machine learning algorithms, namely random forest, naïve Bayes, and support vector machine (SVM), were used for training and testing. The machine learning models have different hyperparameters that are external to the model and must be determined before model training. Table 2 lists the hyperparameters that were optimized in our study, and the following section describes the algorithms and models used.

A. FEATURE SELECTION

Four types of algorithm were used to rank the features: the chi-squared test, minimum redundancy maximum relevance (MRMR) algorithm [32], ReliefF [33], and two-sample *t* test.

Each feature ranking algorithm is described in brief as follows:

1) Chi-squared test: This test was used to examine the individual dependency between the QDSA feature and

TABLE 2. Hyperparameters of machine learning models in this study.

Model Hyperparameters	
Random Forest	Number of trees; minimum leaf size
Naïve Bayes	Kernel estimator type; width
Support Vector Machine	Kernel mapping type; box constraint

the target variable (i.e., BAVM hemorrhage). If dependency was found between the feature and target variable, the *p* value would be small. The ranking was based on the value of $-\log(p)$.

- 2) MRMR: This minimizes the redundancy of a set of features and maximizes the relevance of the feature set corresponding to the target variable based on mutual information so that feature importance is ranked.
- 3) ReliefF: This algorithm is used to perform scoring based on the relationship between a random observation and its nearest neighbors. It iteratively updates the score by a positive term if a feature is different from its neighbors with different target variables and by a negative term if a feature is different from neighbors with the same target variable.
- 4) Two-sample *t* test: This is a statistical test of the difference in feature means between the target variables. If the mean of a QDSA feature in one target variable was different from the other, the *p* value would be small. log (*p*) was applied for feature ranking.

B. RANDOM FOREST

In a random forest [17] the concept of a majority vote is applied. A random forest combines many decision trees, and the final classification depends on the class that has the most votes where a tree gives each vote. Each decision tree is built using a bootstrap data set, which comprises randomly sampled data with replacement from the training data set, and is evaluated using out-of-bag data. Two hyperparameters are to be optimized in the random forest, namely the number of trees and minimum leaf size. Generally, the more trees there are in a random forest, the higher is the performance, but performance decreases beyond a certain point. The minimum leaf size is the smallest number of observations in a decision tree, and the model tends to overfit when the leaf size is too large. The random forest is a classic ensemble learning method with proven applicability in many settings [19], [30], [31].

C. NAÏVE BAYES

The naïve Bayes classifier [18] is based on Bayes' theorem, and it assumes that features are independent and Gaussian distributed. The classification results for a new data set are determined by its likelihood of being posteriorly distributed, which is proportional to the multiplication of the prior and likelihood in the training data set. The prior of the data is the distribution of each class, and the likelihood is the distribution of each predictor. Because the assumption of a Gaussian distribution may not hold for features, the use of a kernel estimator in naïve Bayes has been suggested to provide higher performance than the Gaussian naïve Bayes [19].

D. SVM

The SVM [20] is an algorithm that finds a hyperplane to separate each class with the maximal margin. It is easy to find the boundary in linearly separable data distribution; however, the data is not always ideally distributed. Generally, a kernel method is applied to map the data to a higher dimension and make it separable, and the type of the kernel is a hyperparameter. The distance between the data and the hyperplane is the margin. A hard margin can make a complete separation but is sensitive to the outlier and the overlapping of classes. On the other hand, a soft margin allows for misclassifications within the margin and is more robust in real-world applications. The box constraint is a penalty term that decides the cost of the misclassification, and the higher it is, the harder the margin becomes.

E. BAYESIAN OPTIMIZATION

A fundamental method for adjusting the hyperparameters involves testing every possible combination to find the one with the highest performance, but this is time consuming. The aim of Bayesian optimization [14] is to optimize the loss function composited using different hyperparameters. It is based on the Gaussian process and assumes a multivariate normal distribution of the loss function at every point of the hyperparameters. Standard deviation, which represents the uncertainty of the loss value at a certain point, decreases once that point is known. An acquisition function is applied to decide the next point of the hyperparameters by accounting for the mean and standard deviation of the estimated loss function, and the estimated loss eventually approaches the true loss. The estimated minimum loss was obtained through an acceptable number of iterations, that is, 100 loops in our experiment. Compared with the trial of every possible combination of the hyperparameters, Bayesian optimization was an efficient approach for the optimization of the hyperparameters [14]. Additionally, we applied 5-fold cross-validation for the training data set to prevent the model from overfitting. The whole training data set was trained again using the optimal hyperparameters to achieve the final model for the classification of the training and testing data sets.

V. EXPERIMENTAL SETUP AND WORKFLOW

The experiment was performed on a personal computer with intel CPU Core-i7 9700 and 64 gigabytes of RAM. The computer programming language MATLAB v. R2019b (MathWorks, Inc.) was used to perform time-density analysis and the model learning process. The Statistics and Machine Learning Toolbox and Optimization Toolbox in MATLAB were utilized to fit the TDC, rank the features, build the machine learning models, and optimize the model hyperparameters.



FIGURE 2. Experimental workflow with the steps of feature extraction, feature selection, training machine learning models, and performance evaluation.

Fig. 2 depicts the overall workflow of the experiment. First, an experienced clinical radiologist delineated the nidus of the BAVM and then conducted a semiautomatic TDC analysis using in-house software developed under MATLAB with a graphical user interface [8]. Second, the four feature selection algorithms were performed on ODSA features to rank the features for machine learning. Machine learning models were trained using the ranked features, and the hyperparameters of each model were optimized using Bayesian optimization through 5-fold cross-validation. The classification of both the training and testing data sets were generated for evaluation. Accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC_{ROC}) , and computation time were obtained to evaluate model performance. Because the optimal model was assumed to have the ability to obtain the highest value and lowest deviation between the training and testing data sets, we chose the models based on the following criteria: the highest average performance of the training and testing data sets subtracted by the difference between them. To provide a comparison, three experienced radiologists diagnosed the same DSA testing images based on their visual inspection and empirical knowledge, as they do in clinical practice.

VI. RESULTS

A. STATISTICS AND FEATURE RANKING

The mean and standard deviation of QDSA values corresponding to the BAVM hemorrhage and the feature ranking are presented in Table 3. Of 16 hemodynamic parameters, 9 and 13 parameters exhibited significant differences in the chi-squared test and two-sample t test, respectively. The MRMR and ReliefF are not statistical tests; thus, no significant p value is available.

The mean PD was higher in the nonhemorrhage group, and the difference was statistically significant for both TABLE 3. QDSA results in the training data set.

	NonHx.	Hx.	Fe	eature	Ranki	ng
Anterior-posterior Projection			(1)	(2)	(3)	(4)
PD	372.1±103.2	293.7±111.0	5*	10	7	4*
BAT	1.4 ± 0.3	1.5 ± 0.4	12	2	15	11*
ТТР	3.0±0.5	3.1±0.7	15	11	11	15
WI	330.9±96.2	291.2±118.5	14	16	12	13*
WO	-152.1±57.9	-107.1±54.3	7*	8	14	3*
SI	2.3 ± 0.8	3.1±1.8	3*	3	4	7*
FWHM	2.8 ± 0.7	3.0 ± 0.8	13	13	3	14
AUCTDC	1077.5±343.4	926.5±360.6	9*	4	13	10*
Lateral P	Lateral Projection					
PD	463.3±142.6	340.8±139.2	2*	6	5	2*
BAT	1.4±0.3	1.5±0.4	10	15	16	9*
ТТР	3.0 ± 0.5	3.0±0.6	16	12	10	16
WI	430.4±158.5	337.2±138.2	6*	9	9	6*
wo	-209.1±81.2	-133.5±63.2	1*	1	6	1*
SI	2.2 ± 0.8	2.8 ± 0.9	4*	5	1	5*
FWHM	2.5±0.6	2.8 ± 0.7	11	7	2	12*
AUCTDC	1226.0±412.8	1001.7±453.4	8*	14	8	8*

"*" indicates that the p value is less than 0.05. "Hx." is an abbreviation of hemorrhage. Index (1), (2), (3), and (4) under the feature ranking were ranked by the chi-squared test, MRMR, ReliefF, and two-sample t test, respectively.

anterior-posterior and lateral projections in both statistical tests. The mean BAT was earlier in the nonhemorrhage group, and this was statistically significant for both projections in the *t* test. An earlier mean TTP was observed for the anterior-posterior projection, but the difference was not statistically significant. The mean inflow and outflow gradients, WI and WO, respectively, were significantly steeper in the nonhemorrhage group. A higher mean SI (the ratio of WI and the absolute value of WO) was observed in the hemorrhage group, and the difference was statistically significant for both projections in both statistical tests. The mean FWHM was narrower in the nonhemorrhage group, but the difference was only significant for the lateral projection in the *t* test. A smaller mean AUC_{TDC} was observed in the hemorrhage group, and the difference was statistical tests.

B. MODEL PERFORMANCE

The models' output was the probability of rupture and set at 0.5 to calculate accuracy, sensitivity, and specificity. The highest value and lowest deviation between the training and testing data sets were the criteria for selecting the optimal model. Moreover, the models with the least features were preferred if their performance was the same. The hyperparameter values were different for every model with different numbers of features. Table 4 lists the hyperparameters of the chosen machine learning models, and Table 5 presents their performance before and after feature selection. All models had higher performance after feature selection than before it.

 TABLE 4. Hyperparameters and feature ranking algorithm of the chosen models after Bayesian optimization.

Model	Hyperparameter / Feature Ranking Algorithm	(1)	(2)
	Number of trees	484	500
RF	Minimum leaf size	22	4
	ReliefF	-	Top 6 features
	Kernel estimator type	normal	normal
NB	Width	0.21006	0.11426
	ReliefF	-	Top 8 features
	Kernel mapping type	linear	polynomial
SVM	Box constraint	0.25766	0.057676
	Two-sample t-test	-	Top 7 features

RF, NB, and SVM are abbreviations of random forest, naïve Bayes, and support vector machine, respectively. Index (1) and (2) for the hyperparameters represent the model before and after feature selection, respectively, and (2) for the feature ranking algorithm stands for the ranked feature included.

Before feature selection (all 16 features involved in training), accuracy of the training data set was the highest 81.3% for the random forest and the lowest 74.7% for the SVM. The accuracy of the testing data set was the highest 85.7% for the random forest, and the lowest 76.2% for the naïve Bayes classifier and SVM. The sensitivity of the training

FABLE 5. Performance of the mod	els for training	and testing	data sets.
--	------------------	-------------	------------

Mod	lel	Accuracy	Sensitivity	Specificity	AUC _{ROC}
Before Feature Selection					
DE	(1)	81.3%	76.5%	85.4%	0.93
KF	(2)	85.7%	88.9%	83.3%	0.89
ND	(1)	78.7%	73.5%	82.9%	0.88
NB	(2)	76.2%	55.6%	91.7%	0.84
CLUM	(1)	74.7%	64.7%	82.9%	0.84
SVM (2)	(2)	76.2%	66.7%	83.3%	0.88
After Feature Selection					
DE	(1)	92.7%	91.2%	93.9%	0.98
RF	(2)	85.7%	100%	75.0%	0.97
ND	(1)	81.3%	76.5%	85.4%	0.91
NB	(2)	81.0%	66.7%	91.7%	0.86
CX/M	(1)	83.3%	83.8%	82.9%	0.91
5 V M	(2)	81.0%	88.9%	75.0%	0.96

RF, NB, and SVM are abbreviations of random forest, naïve Bayes, and support vector machine, respectively. Index (1) and (2) show the performance for the training and testing data sets, respectively.

data set was the highest 76.5% for the random forest and the lowest 64.7% for the SVM. For the testing data set, sensitivity was the highest 88.9% for the random forest and the lowest 55.6% for the naïve Bayes classifier. The specificity of the training data set was the highest 85.4% for the random forest and the lowest 82.9% for the naïve Bayes and SVM models. For the testing data set, the specificity was the highest 91.7% for the naïve Bayes classifier and the lowest 83.3% for the random forest and SVM. For the training data set, the highest AUC_{ROC} was 0.93 for the random forest and the lowest one was 0.84 for the SVM. For the testing data set, the highest AUC_{ROC} was 0.89 for the random forest, and the lowest one was 0.84 for the naïve Bayes classifier.

After feature selection, the highest performance of the random forest and naïve Bayes models was obtained using the top six features and the top eight features ranked by ReliefF, respectively. In the SVM, the highest performance was obtained using the top seven features ranked by the t test. The optimal model among the three machine learning algorithms was the random forest, and its accuracy, sensitivity, specificity, and AUC_{ROC} were 92.7%, 91.2%, 93.9%, and 0.98 for the training data set, respectively. For the testing data set, they were 85.7%, 100%, 75.0%, and 0.97, respectively. For the training data set, the accuracy, sensitivity, specificity, and AUC_{ROC} of the naïve Bayes model were 81.3%, 76.5%, 85.4%, and 0.91, respectively. For the testing data set, they were 81.0%, 66.7%, 91.7%, and 0.86, respectively. The SVM model had accuracy, sensitivity, specificity, and AUC_{ROC} of 83.3%, 83.8%, 82.9%, and 0.91, respectively. For the testing data set, they were 81.0%, 88.9%, 75.0%, and 0.96,

respectively. After feature selection, all the models had fewer features, and their performance was improved. The computation time of the machine learning models for the whole testing data set was less than 1 s.

C. DIAGNOSIS BY RADIOLOGISTS

Three radiologists (one neuroradiologist with 15 years of experience, one neuroradiology fellow, and one resident with 4 years of training) diagnosed all cases in the testing data set based on DSA images. Table 6 describes their performance in terms of accuracy, sensitivity, specificity, and AUC_{ROC}. The average accuracy was 66.7%, with a minimum value of 52.4% and a maximum value of 76.2%. The fellow had the highest performance, with sensitivity of 66.7% and specificity of 91.7%. Their mean diagnosis times ranged from 17 to 33 s, with an overall mean of 28 s. An AUC_{ROC} of 0.78 was obtained from the average of the three diagnoses.

TABLE 6. Radiologists' performance for the testing data set.

	Accuracy	Sensitivity	Specificity	Time (s)
Neuro- radiologist	71.4%	55.6%	83.3%	17
Fellow	76.2%	66.7%	91.7%	33
Resident	52.4%	44.4%	58.3%	33
Average	66.7 %	55.6%	77.8%	28

VII. DISCUSSION

A. SUPERIOR PERFORMANCE BY THE PROPOSED MODEL

In this study, we used QDSA, which provides temporal hemodynamic information, to train machine learning models to detect hemorrhagic BAVMs. Of our models, the random forest model had the highest performance, with accurate detections of 92.7% and 85.7% and AUC_{ROC} values of 0.98 and 0.97 for the training and testing data sets, respectively. The sensitivity and specificity of the random forest model were 91.2% and 93.9% for the training data set and 100% and 75.0% for the testing data set. Fig. 3 displays the receiver operating characteristic curves for the random forest model for the training and testing data sets as well as the mean diagnostic results by radiologists for the testing data set, with AUC_{ROC} values of 0.98, 0.97, and 0.78, respectively. The results showed that the random forest model trained with hemodynamic features from QDSA analysis was superior to the diagnoses of radiologists and could help clinical radiologists with complex risk assessment.

B. RISK ASSESSMENT WITH CLINICAL FEATURES

Clinical features are currently the primary measures used for hemorrhagic risk assessment, and such features have been well investigated. Guidelines concerning the clinical features of BAVMs are available and provide a standard



FIGURE 3. Receiver operating characteristic curve of the random forest model for the training and testing data sets, and the average of diagnoses by radiologists.

for researchers and trials [3]. The Spetzler-Martin grade, based on the eloquent area, nidus size, and deep location, has been used for the risk assessment of BAVM-related brain surgery [21]. Nevertheless, this grading system is not intended for hemorrhagic risk assessment. One study revealed that input flow pressure and venous outflow restriction are meaningful predictors of hemorrhagic BAVMs [22]. A meta-analysis showed that the deep location of the BAVM, a small number of drainage veins, flow-related aneurysms, and venous ectasia are all independent factors related to hemorrhage risk [5]. Many researchers have obtained the same results in terms of the relationship between venous angioarchitecture and the risk of BAVM hemorrhage [4], [5], [23]. Nonetheless, despite the proven relationship between clinical features and hemorrhagic risk, a precise evaluation of the angioarchitecture requires a clinical radiologist with years of experience, and such an evaluation takes considerable interpretation time. Other modalities with 3-dimensional spatial information are also required under certain circumstances. In a clinical emergency, this is not an ideal approach for hemorrhagic risk assessment.

C. STUDIES REGARDING QUANTITATIVE DSA

In contrast to clinical features, QDSA can be conducted immediately by defining vascular ROIs directly on DSA images. The quantitative parameters of TDC objectively reflect the hemodynamics of the cerebrovasculature. The difference in hemodynamics is important, and a study demonstrated the difference within various brain vessels through an independent component analysis to separate the artery, capillary, and vein using TDC [24]. Similarly, normal and abnormal brain vessels have different hemodynamic properties. A relative TTP was prolonged in a patient with the stenoocclusive arterial disease before stent placement compared with a normal control [25]. BAT, TTP, WO, FWHM, and AUC_{TDC} were reduced in patients with carotid stenosis after stenting [7], indicating increased flow through the carotid artery after treatment. Additionally, TTP has been used to estimate the cerebral circulation time for the detection of vasospasms in patients with subarachnoid hemorrhages [26]. A double peak sign of the TDC was observed in dural arteriovenous fistula and is used for the evaluation of venous drainage function [6].

Regarding BAVMs, a previous study found that the TTP of the feeding artery was lower in patients with hemorrhages than in patients with seizures, and TTP was used to assess hemorrhagic risk in BAVMs by QDSA [27]. Another study revealed that the mean transit time and the ratio of the mean transit time of the draining to the feeding vessels showed statistical differences between hemorrhagic and nonhemorrhagic BAVMs [34]. A recent study also found a prolonged mean transit time in restricted venous drainage and a prolonged cerebral circulation time in BAVMs with hemorrhage [35]. Additionally, in a previous study, we introduced the SI and found that it was associated with supratentorial BAVM rupture at the most dominant drainage vein [8]. In that research, we also demonstrated that BAT, TTP, SI, and FWHM at the cavernous internal carotid artery as well as PD, BAT, SI, and FWHM at the drainage vein all differed significantly between patients with and without hemorrhage. In the present study, instead of measuring the feeding artery and drainage vein, we directly assessed the nidus of the BAVM. We found that almost all QDSA features (except for TTP for both anterior-posterior and lateral projections as well as FWHM for the anterior-posterior projection) were significantly different between patients with and without hemorrhage. On average, PD, WI, and AUCTDC were higher in nonhemorrhagic patients, and BAT, WO, SI, and FWHM were lower in nonhemorrhagic patients. This phenomenon was related to more stagnant outflow in the hemorrhagic BAVM [8], [35]. All of the aforementioned findings thus support the efficacy of QDSA for assessing cerebrovascular diseases.

D. PREVIOUS MODELS

Most relevant studies have focused on finding predictors of BAVM hemorrhage from clinical features [4], [5], [8], [23], [36], [37]. Feghali et al. developed a scoring model termed R₂eD [36] based on multivariable logistic regression to detect hemorrhagic BAVMs based on race (nonwhite vs. white), exclusive deep location, BAVM size (small vs. large), venous drainage (exclusive deep vs. other), and monoarterial feeding (1 vs. > 1 feeding artery). They used the hold-out method, in which half of the data was split as the validation data set to observe if the model was overfitting; the AUCROC was 0.685. Although clinical features are more intuitive, variation in hemodynamics can directly alert clinicians to the risk of a hemorrhagic BAVM. Our previous study introduced the SI [8], which reflects hemodynamic differences between hemorrhagic and nonhemorrhagic BAVMs, to build a predictive model using logistic regression, and the AUC_{ROC} was 0.75. Huang et al. combined clinical features with TTP to build a statistical model using multivariable logistic regression, and the AUC_{ROC} was 0.838 [37]. In that study, the combination of clinical features with QDSA improved the predictive model. Table 7 of this article summarizes the predictors and performance of the predictors from the aforementioned studies. In this study, we used QDSA from BAVM nidus through machine learning algorithms with proper feature selection and hyperparameter optimization and found AUC_{ROC} values of 0.98 and 0.97 for the training and testing data sets, respectively. Our study verified the feasibility and application of machine learning models for detecting hemorrhagic BAVMs using QDSA.

TABLE 7.	Comparison	with	other	studies.
----------	------------	------	-------	----------

Study	Feature	AUCROC
J. Feghali <i>et al.</i> [36]	Race, exclusive deep location, AVM size, venous drainage, and monoarterial feeding	0.685
H. C. Yang <i>et al.</i> [8]	Stasis index of drainage vein and exclusive deep venous drainage	0.750
Z. Huang <i>et al.</i> [37]	Age, gender, location, TTP, aneurysm related with BAVM, nidus volume, cross-sectional area of feeding arteries and draining veins, venous drainage (cortical, deep, or mixed), draining types (single or multiple)	0.838
Proposed model	SI, FWHM of anterior-posterior projection; PD, WO, SI, FWHM of lateral projection	0.97

E. REDUNDANCE OF CLINICAL FEATURES

The proposed machine learning method can efficiently detect hemorrhagic BAVMs using QDSA features. Compared with clinical radiologists, it performed faster and more accurately. It can assist in diagnosis in clinical settings. Because machine learning algorithms can effectively manage high-dimensional features, it is reasonable to hypothesize that the use of QDSA combined with clinical features may improve detection accuracy. However, we found that the inclusion of clinical features, namely seizure, headache, focal deficit, supratentorial location, dilated feeder artery, sprouting angiogenesis, nonsprouting angiogenesis, single drainage, multiple drainage, number of drainage veins, venous rerouting, venous ectasia, exclusive deep veins, and venous sacs, did not result in a clear improvement in results. The accuracy, sensitivity, specificity, and AUC_{ROC} of the model with clinical features were 71.4%, $55.6\%,\,83.3\%,\,and\,0.91,\,respectively,\,for$ the testing data set. Given the direct availability of QDSA, it could be more useful than clinical features.

F. SUPERIORITY OF THE RANDOM FOREST MODEL

Our experimental result revealed that the random forest model outperformed the naïve Bayes and support vector

models. The main limitation of the naïve Bayes model is the assumption of independence among the predictor features. Because QDSA features are quantified from the same TDC, they do not meet the independence assumption. In terms of SVM, performance is limited when features between classes have overlapping distributions. Despite the presence of statistically significant differences in QDSA features between the nonhemorrhagic and hemorrhagic groups, an overlap was observed from their mean and standard deviation values.

A random forest model involves the implementation of the universal concept that two heads are better than one, and it does not require any assumptions. By combining multiple decision trees, the random forest model can balance the bias-variance tradeoff associated with most machine learning methods, and this enhances its performance. Additionally, because resampling bootstrap data sets are used, the random forest model can robustly manage data sets with limited sample sizes [30]. Earlier studies have demonstrated its feasibility and efficacy in practical application [19], [31]. Studies have also shown the superiority of the random forest model to other single-learner-based machine learning models. A study compared the performance of different classification algorithms applied to asphalt pavement deterioration data and reported that the random forest model outperformed the naïve Bayes model [19]. Another study that assessed groundwater potential demonstrated that the random forest model was superior to kernel-mapped SVM models [31]. In the present study, the random forest model was trained using QDSA features and effectively detected hemorrhagic BAVMs on DSA, thereby improving the rupture assessment.

G. FEATURE SELECTION AND HYPERPARAMETER OPTIMIZATION

The purpose of feature selection is to eliminate redundant or irrelevant features such that the model exhibits higher performance and becomes simpler to train and easier to interpret. A study showed higher model performance than other models in the detection of Parkinson disease from voice signals, and the related researchers used the chi-squared model to rank features for the removal of noisy features [38]. Another study developed a two-stage model to diagnose heart failure by using a statistical model based on mutual information at the first stage and a neural network at the second stage, and the performance of this model was higher than the other 28 prediction models [39]. A feature ranking algorithm named ReliefF [33] was used to predict compressive strength and concrete composition, and the result revealed higher correlation and fewer errors than the model without feature selection using ReliefF [40]. In addition, a ReliefFbased artificial neural network had 100% cross-validation accuracy for fault diagnosis in ball bearings [41]. In the present study, we used the aforementioned algorithms to rank features and perform selection to find the optimal subset. Higher performance was found after feature selection. Moreover, the optimal model was the random forest trained using the top six features ranked using the ReliefF algorithm. Fig. 4 presents the violin plot of the top two and last two features ranked using ReliefF. The more distinct distribution between nonhemorrhage and hemorrhage was observed for the top two features than for the last two features.



FIGURE 4. The violin plot for the top two (SI_{LAT} and FWHM_{LAT}) and last two (BAT_{AP} and BAT_{LAT}) QDSA features ranked by the ReliefF algorithm. Subscript LAT and AP are the abbreviations of lateral and anterior-posterior projections.

Hyperparameter optimization is another important concern in the development of an efficient machine learning model. The common scheme for finding appropriate hyperparameters involves empirical manual tuning. With this method, the result highly depends on the operator's experience, and finding the optimal solution cannot be guaranteed every time. Grid search is another method for testing every combination of hyperparameters and for finding the globally optimal hyperparameters; however, it is time consuming and impractical for the current application. Alternatively, Bayesian optimization [14] and the genetic algorithm [43] are two efficient methods for hyperparameter optimization. The genetic algorithm is an efficient approach inspired by Darwin's theory of natural evolution. It evaluates and selects parent candidates with favorable performance from a given population. Then, the offspring take the properties from the parent through crossover, and some properties are altered through mutation to prevent the model from becoming stuck in a local optimum. The genetic algorithm has been used in machine learning and achieves satisfactory performance [45], [46]. In this study, we attempted to use both Bayesian optimization and the genetic algorithm with the same computational costs (100 models) to optimize the hyperparameters of the random forest with all QDSA features. The accuracies were 81.3% and 85.7% for the training and testing data sets, respectively, for Bayesian

optimization and 78.7% and 81.0%, respectively, for the genetic algorithm. Because the Bayesian optimization had performance superior to that of the genetic algorithm for our data sets, we adopted it in this study. Nevertheless, a previous study demonstrated that the genetic algorithm could be more effective than Bayesian optimization [47]. Therefore, depending on the data set, either the Bayesian optimization or genetic algorithm can be considered for hyperparameter optimization when machine learning models are being developed.

H. STUDY LIMITATIONS

Despite the favorable outcomes, this study has some limitations. First, we did not examine the uncertainties caused by different protocols and modalities. A previous study reported the complete mixing of the contrast agent with blood flow and that it did not separate once mixed [28]. This study proved that the TDC could represent the real hemodynamics of blood flow, and it did not show variability under typical imaging protocols. Second, the selection of ROIs was not the same every time, and we did not account for intraobserver and interobserver reliability. Previous studies, however, have demonstrated the consistency of QDSA by showing its robust and clinical reproducibility [6]-[8], [24]-[26]. Third, even though two projections of 2D-DSA were acquired simultaneously, it was not possible to fully compensate for the overlapping of vasculatures. A 4D-DSA technique has been developed recently, offering both three-dimensional structure and temporal cerebrovasculature information [29]. This is a solution to the problem of overlapping vessels and has the potential to achieve better performance than 2D-DSA in the application of QDSA. Future studies are warranted on the external validation and reliability of ROI selection. Finally, although the proposed model trained by QDSA features has superior performance, modern angiography machines can perform "flat-panel CT" to confirm intraoperative bleeding during angiography in one or two minutes. Our proposed model may assist radiologists in determining the hemorrhage risk before additional radiological examinations such as flat-panel CT.

VIII. CONCLUSION

In clinical settings, the evaluation of angioarchitecture needs to be performed by specialized radiologists. This is time consuming and tends to be subjective. Conversely, QDSA objectively quantifies the temporal information of the cerebrovasculature by delineating the nidus of the BAVM on DSA images without further assessment by clinical radiologists. In this study, the accuracy of detecting hemorrhages through machine learning models based on QDSA features exceeded that of diagnosis by trained radiologists in the detection of hemorrhagic BAVM. QDSA combined with the random forest model is thus a promising approach for the evaluation of hemorrhagic BAVMs and has the potential to be used for detecting other cerebrovascular diseases.

ACKNOWLEDGMENT

This manuscript was edited by Wallace Academic Editing.

REFERENCES

- R. A. Gabriel *et al.*, "Ten-year detection rate of brain arteriovenous malformations in a large, multiethnic, defined population," *Stroke*, vol. 41, no. 1, pp. 21–26, Jan. 2010, doi: 10.1161/STROKEAHA.109. 566018.
- [2] J. P. Mohr *et al.*, "Medical management with or without interventional therapy for unruptured brain arteriovenous malformations (ARUBA): A multicentre, non-blinded, randomised trial," *Lancet*, vol. 383, no. 9917, pp. 614–621, Feb. 2014, doi: 10.1016/S0140-6736(13)62302-8.
- [3] Joint Writing Group of the Technology Assessment Committee American Society of Interventional and Therapeutic Neuroradiology; Joint Section on Cerebrovascular Neurosurgery a Section of the American Association of Neurological Surgeons and Congress of Neurological Surgeons; Section of Stroke and the Section of Interventional Neurology of the American Academy of Neurology *et al.*, "Reporting terminology for brain arteriovenous malformation clinical and radiographic features for use in clinical trials," *Stroke*, vol. 32, no. 6, pp. 1430–1442, Jun. 2001, doi: 10.1161/01.str.32.6.1430.
- [4] J. A. Hernesniemi, R. Dashti, S. Juvela, K. Väärt, M. Niemelä, and A. Laakso, "Natural history of brain arteriovenous malformations: A long-term follow-up study of risk of hemorrhage in 238 patients," *Neurosurgery*, vol. 63, no. 5, pp. 823–831, Nov. 2008, doi: 10.1227/01.NEU.0000330401.82582.5E.
- [5] H. Kim, R. Al-Shahi Salman, C. E. McCulloch, C. Stapf, W. L. Young, and F. the MARS Coinvestigators, "Untreated brain arteriovenous malformation: patient-level meta-analysis of hemorrhage predictors," *Neurology*, vol. 83, no. 7, pp. 590–597, Aug. 2014, doi: 10.1212/WNL.000000000000688.
- [6] W.-Y. Guo *et al.*, "Quantifying the cerebral hemodynamics of dural arteriovenous fistula in transverse sigmoid sinus complicated by sinus stenosis: A retrospective cohort study," *Amer. J. Neuroradiol.*, vol. 38, no. 1, pp. 132–138, Jan. 2017, doi: 10.3174/ajnr.A4960.
- [7] M. M. H. Teng, F.-C. Chang, C.-J. Lin, L. Chiang, J.-S. Hong, and Y.-H. Kao, "Peritherapeutic hemodynamic changes of carotid stenting evaluated with quantitative DSA in patients with carotid stenosis," *Amer. J. Neuroradiol.*, vol. 37, no. 10, pp. 1883–1888, Oct. 2016, doi: 10.3174/ajnr.A4813.
- [8] H. C. Yang *et al.*, "Stasis index from hemodynamic analysis using quantitative DSA correlates with hemorrhage of supratentorial arteriovenous malformation: A cross-sectional study," *J. Neurosurg.*, vol. 132, no. 5, pp. 1574–1582, 2019. [Online]. Available: https:// thejns.org/view/journals/j-neurosurg/aop/article-10.3171-2019.1. JNS183386.xml.
- [9] H. C. Kim *et al.*, "Machine learning application for rupture risk assessment in small-sized intracranial aneurysm," *J. Clin. Med.*, vol. 8, no. 5, p. 683, May 2019, doi: 10.3390/jcm8050683.
- [10] E. K. Oermann *et al.*, "Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations," *Sci. Rep.*, vol. 6, no. 1, p. 21161, Feb. 2016, doi: 10.1038/srep21161.
- [11] H. K. Thompson, C. F. Starmer, R. E. Whalen, and H. D. Mcintosh, "Indicator transit time considered as a gamma variate," *Circulat. Res.*, vol. 14, no. 6, pp. 502–515, Jun. 1964, doi: 10.1161/01.res.14.6.502.
- [12] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, May 1996, doi: 10.1137/0806023.
- [13] D. Koller and M. Sahami. (Feb. 1996). Toward Optimal Feature Selection. Accessed: Jun. 15, 2020. [Online]. Available: http://ilpubs.stanford.edu:8090/208/
- [14] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2012, pp. 2951–2959.
- [15] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4, doi: 10.1109/ICPR.2008.4761297.

- [16] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting (With discussion and a rejoinder by the authors)," *Ann. Statist.*, vol. 28, no. 2, pp. 337–407, Apr. 2000, doi: 10.1214/aos/1016218223.
- [17] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [18] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [19] S. M. Piryonesi and T. E. El-Diraby, "Role of data analytics in infrastructure asset management: Overcoming data size and quality problems," *J. Transp. Eng., B, Pavements*, vol. 146, no. 2, Jun. 2020, Art. no. 04020022. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/JPEODX.0000175
- [20] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1. New York, NY, USA: Springer, 2001.
- [21] R. F. Spetzler and N. A. Martin, "A proposed grading system for arteriovenous malformations," *J. Neurosurg.*, vol. 65, no. 4, pp. 476–483, Oct. 1986, doi: 10.3171/jns.1986.65.4.0476.
- [22] D. H. Duong *et al.*, "Feeding artery pressure and venous drainage pattern are primary determinants of hemorrhage from cerebral arteriovenous malformations," *Stroke*, vol. 29, no. 6, pp. 1167–1176, Jun. 1998, doi: 10.1161/01.str.29.6.1167.
- [23] M. A. Stefani, P. J. Porter, K. G. terBrugge, W. Montanera, R. A. Willinsky, and M. C. Wallace, "Angioarchitectural factors present in brain arteriovenous malformations associated with hemorrhagic presentation," *Stroke*, vol. 33, no. 4, pp. 920–924, Apr. 2002, doi: 10.1161/01.str.0000014582.03429.f7.
- [24] J.-S. Hong, Y.-H. Kao, F.-C. Chang, and C.-J. Lin, "Validating the automatic independent component analysis of DSA," *Amer. J. Neuroradiol.*, pp. 540–542, Jan. 2019, doi: 10.3174/ajnr.A5963.
- [25] C. J. Lin *et al.*, "Monitoring peri-therapeutic cerebral circulation time: A feasibility study using color-coded quantitative DSA in patients with steno-occlusive arterial disease," *Amer. J. Neuroradiol.*, vol. 33, no. 9, pp. 1685–1690, Oct. 2012, doi: 10.3174/ajnr.A3049.
- [26] C. F. Lin, S. P. C. Hsu, C. J. Lin, W. Y. Guo, C. H. Liao, W. F. Chu, S. C. Hung, Y. S. Shih, and Y. T. Lin, "Prolonged cerebral circulation time is the best parameter for predicting vasospasm during initial CT perfusion in subarachnoid hemorrhagic patients," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0151772, doi: 10.1371/journal.pone.0151772.
- [27] J. S. Norris, T. A. Valiante, M. C. Wallace, R. A. Willinsky, W. J. Montanera, K. G. terBrugge, and M. Tymianski, "A simple relationship between radiological arteriovenous malformation hemodynamics and clinical presentation: A prospective, blinded analysis of 31 cases," *J. Neurosurg.*, vol. 90, no. 4, pp. 673–679, Apr. 1999, doi: 10.3171/jns.1999.90.4.0673.
- [28] B. B. Lieber, C. Sadasivan, Q. Hao, J. Seong, and L. Cesar, "The mixability of angiographic contrast with arterial blood," *Med. Phys.*, vol. 36, no. 11, pp. 5064–5078, Nov. 2009, doi: 10.1118/1.3243079.
- [29] K. L. Ruedinger, E. C. Harvey, S. Schafer, M. A. Speidel, and C. M. Strother, "Optimizing the quality of 4D-DSA temporal information," *Amer. J. Neuroradiol.*, pp. 2124–2129, Oct. 2019, doi: 10.3174/ajnr.A6290.
- [30] N. Gunduz and E. Fokoue, "Robust classification of high dimension low sample size data," 2015, arXiv:1501.00592. [Online]. Available: http://arxiv.org/abs/1501.00592
- [31] S. A. Naghibi, K. Ahmadi, and A. Daneshi, "Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping," *Water Resour. Manage.*, vol. 31, no. 9, pp. 2761–2775, Jul. 2017, doi: 10.1007/s11269-017-1660-3.
- [32] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005, doi: 10.1142/s0219720005001004.
- [33] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1, pp. 23–69, Oct. 2003, doi: 10.1023/A:1025667309714.
- [34] T. Todaka, J.-I. Hamada, Y. Kai, M. Morioka, and Y. Ushio, "Analysis of mean transit time of contrast medium in ruptured and unruptured arteriovenous malformations: A digital subtraction angiographic study," *Stroke*, vol. 34, no. 10, pp. 2410–2414, Oct. 2003, doi: 10.1161/01.STR.0000089924.43363.E3.
- [35] J.-K. Burkhardt, X. Chen, E. A. Winkler, D. L. Cooke, H. Kim, and M. T. Lawton, "Delayed venous drainage in ruptured arteriovenous malformations based on quantitative color-coded digital subtraction angiography," *World Neurosurg.*, vol. 104, pp. 619–627, Aug. 2017, doi: 10.1016/j.wneu.2017.04.120.

- [37] Z. Huang, K. Peng, C. Chen, F. Zeng, J. Wang, and F. Chen, "A reanalysis of predictors for the risk of hemorrhage in brain arteriovenous malformation," *J. Stroke Cerebrovascular Diseases*, vol. 27, no. 8, pp. 2082–2087, Aug. 2018, doi: 10.1016/j.jstrokecerebrovasdis.2018.03.003.
- [38] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019. [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S095741741930452X
- [39] L. Ali and S. A. C. Bukhari, "An approach based on mutually informed neural networks to optimize the generalization capabilities of decision support systems developed for heart failure prediction," *IRBM*, May 2020, doi: 10.1016/j.irbm.2020.04.003.
- [40] V. Vakharia and R. Gujar, "Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques," *Construct. Building Mater.*, vol. 225, pp. 292–301, Nov. 2019, doi: 10.1016/j.conbuildmat.2019.07.224.
- [41] V. Vakharia, V. K. Gupta, and P. K. Kankar, "A comparison of feature ranking techniques for fault diagnosis of ball bearing," *Soft Comput.*, vol. 20, no. 4, pp. 1601–1619, Apr. 2016, doi: 10.1007/s00500-015-1608-6.
- [42] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281–305, Feb. 2012. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/2188385.2188395
- [43] A. Fraser and D. Burnell, *Computer Models in Genetics*. New York, NY, USA: McGraw-Hill, 1970.
- [44] P. I. Frazier, "A tutorial on Bayesian optimization," 2018, arXiv:1807.02811. [Online]. Available: http://arxiv.org/abs/1807.02811
- [45] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshtkar, and S. A. C. Bukhari, "LDA-GA-SVM: Improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Comput. Appl.*, Jul. 2020, doi: 10.1007/s00521-020-05157-2.
- [46] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, 2019, Art. no. 2000410, doi: 10.1109/jtehm.2019.2940900.
- [47] N. Mori, M. Takeda, and K. Matsumoto, "A comparison study between genetic algorithms and Bayesian optimize algorithms by novel indices," in *Proc. Conf. Genetic Evol. Comput. (GECCO)*, Washington DC, USA, 2005, pp. 1485–1492, doi: 10.1145/1068009.1068244.



JIA-SHENG HONG received the B.S. degree from the Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, Taiwan, R.O.C., in 2015, and the M.S. degree from the Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan, R.O.C., in 2017, where he is currently pursuing the Ph.D. degree.

He has been certified as a Medical Radiation Technologist by the Taiwan's Ministry of Health and Welfare. His main research interests include medical image processing, machine learning, and interventional imaging.



CHUNG-JUNG LIN received the M.D. degree from National Taiwan University, in 2001, and the M.S. degree in interventional radiology from Paris Descartes University.

He is currently an Attending Physician with the Taipei Veterans General Hospital and a Professor of radiology with the School of Medicine, National Yang-Ming University, Taipei, Taiwan, R.O.C. He is also the current Director of the Imaging Center, Taipei Veterans General Hospital. His

research interests include neuroradiology, interventional imaging, hemodynamics, endovascular treatments, and arteriovenous malformation.



YUE-HSIN LIN received the B.S. and M.S. degrees from the Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan, R.O.C., in 2016 and 2018, respectively.

He has been certified as a Medical Radiation Technologist by the Taiwan's Ministry of Health and Welfare. His master's thesis is on the use of machine learning in the classification of cerebral arteriovenous malformation hemorrhage.



CHENG-CHIA LEE is currently a Functional Neurosurgeon at the Taipei Veteran General Hospital, who specializes in Gamma Knife radiosurgery. He has been the Secretary-General of the Taiwan Society for Stereotactic Functional Neurosurgery and Radiosurgery (TSSFNR), since 2015. His main research interest includes radiation effects after Gamma Knife radiosurgery.

Mr. Lee has received several awards, including the Outstanding Paper Award from the Taiwan

Neurosurgical Society, the Young Researcher Scholarship from the Taiwan Epilepsy Society and the Taiwan Pain Society, and the Young Neurosurgery Award (Top 15) from the World Federation of Neurosurgical Societies (WFNS), since 2014.



HUAI-CHE YANG is currently the Attending Neurosurgeon of the Department of Functional Neurosurgery, Taipei Veteran General Hospital. He specializes in Gamma Knife radiosurgery and stereotactic functional neurosurgery. He was the Secretary-General of the Taiwan Society for Stereotactic Functional Neurosurgery and Radiosurgery, from 2011 to 2013. He completed his functional neurosurgery and radiosurgery Fellowship at the Department of Neurological Surgery,

University of Pittsburgh, from 2008 to 2010. He has published more than 60 peer-reviewed articles.



LING-HSUAN MENG received the M.D. and Ph.D. degrees from the Department of Neurosurgery, China Medical University, Shenyang, China, in 2011 and 2018, respectively.

She is a Collaboration Scientist at Siemens Healthcare. She coordinates the research project between Siemens Healthcare, the Taipei Veterans General Hospital, and National Yang-Ming University.







TE-MING LIN received the M.D. degree from National Yang-Ming University, in 2012, and the M.S. degree in biomedical electronics and bioinformatics from National Taiwan University, in 2015.

He is currently a Radiology Resident at the Taipei Veterans General Hospital. His research interests include the hemodynamic change in brain arteriovenous malformation and the physics of MR imaging.

YONG-SIN HU received the M.D. degree from National Yang-Ming University, Taipei, Taiwan, R.O.C., in 2013.

He is currently an Attending Physician with the Department of Radiology, Taipei Veterans General Hospital, Taoyuan Branch. His research interests include neuroradiology, interventional imaging, hemodynamics, vascular malformations, and radiosurgery.

WAN-YUO GUO is currently the Head Radiologist at the Taipei Veterans General Hospital and a Professor of radiology with the School of Medicine, National Yang-Ming University. He received training and conducted research in stereotaxy and neuroimaging at the Karolinska Institute, Sweden. He is an enthusiastic imaging Practitioner and a Researcher in neuroradiology and neurosciences. His publications include various topics, including neurosurgery, obstetrics,

WEI-FA CHU received the B.S. degree from the Yuanpei University of Medical Technology.

He is currently serving as a Radiation Technol-

ogist with the Taipei Veterans General Hospital.

He has participated in various research projects

and has assisted in collecting clinical data.

pediatrics, perinatology, pathology, fetal brain volumetry and morphometry, neurology, anesthesiology, psychiatry, neurophysiology, radiation biology, molecular imaging, medical physics, signal analysis, computer science, and biostatistics. He was the President of XXI SNR. He is also the President-Elect of the World Federation of Neuroradiological Societies.





YU-TE WU received the B.S. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, R.O.C., in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pittsburgh, Pittsburgh, PA, in 1992 and 1997, respectively.

He was a Research Associate with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, from 1997 to 1998. He is currently the Dean of the Office of Research and Development and

a Distinguished Professor with the Institute of Biophotonics, National Yang-Ming University, Taipei, Taiwan, R.O.C. He has published more than 100 articles in the areas of neuroimaging, machine learning, and brain–computer interface.