Title

Artificial neural networks analysis predicts long-term fistula function in hemodialysis patients following percutaneous transluminal angioplasty.

Authors

Chien, Aichi
Lall, Ayush
Patel, Maitraya
et al.

Publication Date

2024-06-01

DOI

10.1016/j.engmed.2024.100010

Peer reviewed

HHS Public Access

Author manuscript

EngMedicine. Author manuscript; available in PMC 2024 July 02.

Published in final edited form as:

EngMedicine. 2024 June ; 1(1): . doi:10.1016/j.engmed.2024.100010.

Artificial neural networks analysis predicts long-term fistula function in hemodialysis patients following percutaneous transluminal angioplasty

Aichi Chien[a,*], **Ayush Lall**[a], **Maitraya Patel**[b], **Lucas Cusumano**[a], **Justin McWilliams**[a]

[a] Department of Radiology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

[b] Division of Ultrasonography, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

Abstract

Kidney failure is particularly common in the United States, where it affects over 700,000 individuals. It is typically treated through repeated sessions of hemodialysis to filter and clean the blood. Hemodialysis requires vascular access, in about 70% of cases through an arteriovenous fistula (AVF) surgically created by connecting an artery and vein. AVF take 6 weeks or more to mature. Mature fistulae often require intervention, most often percutaneous transluminal angioplasty (PTA), also known as fistulaplasty, to maintain the patency of the fistula. PTA is also the first-line intervention to restore blood flow and prolong the use of an AVF, and many patients undergo the procedure multiple times. Although PTA is important for AVF maturation and maintenance, research into predictive models of AVF function following PTA has been limited. Therefore, in this paper we hypothesize that based on patient-specific information collected during PTA, a predictive model can be created to help improve treatment planning. We test a set of rich, multimodal data from 28 patients that includes medical history, AVF blood flow, and interventional angiographic imaging (specifically excluding any post-PTA measurements) and build deep hybrid neural networks. A hybrid model combining a 3D convolutional neural network with a multi-layer perceptron to classify AVF was established. We found using this model that we were able to identify the association between different factors and evaluate whether the PTA

* Corresponding author. aichi@ucla.edu (A. Chien).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Aichi Chien reports financial support was provided by NIH. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval

The institutional review board approved this retrospective study and waived informed consent for participation.

Consent for publication

All authors approved the publication.

CRediT authorship contribution statement

Aichi Chien: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ayush Lall:** Software, Project administration. **Maitraya Patel:** Writing – original draft. **Lucas Cusumano:** Conceptualization. **Justin McWilliams:** Investigation, Conceptualization.

procedure can maintain primary patency for more than 3 months. The testing accuracy achieved was 0.75 with a weighted F1-score of 0.75, and AUROC of 0.75. These results indicate that evaluating multimodal clinical data using artificial neural networks can predict the outcome of PTA. These initial findings suggest that the hybrid model combining clinical data, imaging and hemodynamic analysis can be useful to treatment planning for hemodialysis. Further study based on a large cohort is needed to refine the accuracy and model efficiency.

Keywords

Angioplasty; Clinical predictive models; Deep learning; Fistula; Hemodialysis; Medical informatics

Introduction

HEMODIALYSIS is the most common treatment for kidney failure, which is characterized by a severe loss of kidney function, with diabetes a major underlying cause. Also known as end stage renal disease, in the United States, over 700,000 individuals suffer kidney failure, and hemodialysis is therefore a critical part of healthcare and healthcare spending. To enable purifying blood through hemodialysis, a vascular access within a patient with sufficient rate of flow is necessary. This may take the form of a catheter, a surgical graft, or a surgically created arteriovenous fistula (AVF). Because of their relative durability and low rate of infection, AVF are the preferred vascular access [1,2]. An AVF consists of a direct connection between an artery and a vein, typically created within a patient's arm, and usually requires a few months to mature before it may be used for dialysis. To reach maturity, approximately 50% of AVF require assistance, most commonly a percutaneous transluminal angioplasty (PTA) procedure, also referred to as fistulaplasty, to dilate the vessels which comprise the AVF [3].

It is essential to maintain a sufficient rate of blood flow within the AVF to support dialysis, usually a minimum of 400 mL/min [4]. While all vascular accesses have a limited lifetime, 20–60% of AVF can fail or collapse early, in which case a new fistula may be required [4]. Because some AVF may not mature and during maturation a catheter may be required for dialysis access, PTA is usually tried first to restore blood flow within the existing AVF. Therefore, PTA is a critical, and extremely common procedure in clinical care of hemodialysis patients, both to encourage AVF maturation, and to prolong the period for which an AVF can be used for hemodialysis. Fig. 1 presents a clinical example of AVF stenosis before and after PTA.

Postintervention AVF function (or primary-assisted patency) is the number of days before another intervention is required or the AVF must be abandoned. While there are no strict guidelines, it is generally accepted that the predictors of post-intervention patency are generally the indicators for a need for PTA, specifically blood flow rate and change in fistula diameter [4]. If PTA reverses these, the AVF is more likely to remain useable. The importance of other patient-specific factors, such as fistula configuration and other aspects of patient medical history are accepted, but not clearly defined. Likewise, imaging is typically only utilized to guide intervention, and even Doppler ultrasound (DUS) blood flow

measurements are not an essential part of care. Therefore, synthesizing these different types of data (clinical, flow, and images) through a computer-aided predictive model may greatly improve prediction of postintervention AVF function and hemodialysis planning.

Related work

Research related to PTA and AVF function has generally been more oriented towards description than prediction. To date, few predictive models have focused on the characteristics of PTA. Takahashi et al. recently published a medical study using standard statistical methods to find that both fistula configuration (anatomical location) and stenosis location were related to postintervention patency [5]. Zhu et al. also recently conducted a similar medical study with a larger set of clinical variables, but likewise did not include other types of data, such as blood flow measurements or images [6]. The work of Saucy et al. found that intra-operative blood flow measured following fistula creation via a transit time flow measurement probe provided a positive predictive value for early AVF failure greater than 90%, whereas the negative predictive value was less than 40%. But, a similar study following AVF creation by Genek et al. based on the DUS assessment of flow-mediated dilatation, medical histories, internal diameter, wall thickness, peak systolic flow rate and resistive index of the cephalic vein and radial arteries failed to show any statistical significance for predicting fistula failure [7], and concluded that a multifactorial approach might be successful. Drawing on these studies, we can identify a basic set of features with which to begin constructing our predictive model.

A few researchers have applied more sophisticated analyses and modeling approaches to this problem. Rezapour et al. showed that data mining techniques and machine learning, specifically decision trees, could be used to identify risk factors for early AVF failure [8]. Their predictive model, based on eight clinical parameters, achieved a reported accuracy of 61.66%. This study focused on maturation and early AVF failure and is somewhat limited by the small number of parameters, three of which—hypertension, diabetes, and smoking—are closely correlated. Most recently, the deep learning approach, which has been very successful in the healthcare space, has been applied to study fistula maturation [9]. Kordzadeh et al. reported an accuracy of 77.5% predicting AVF maturation based on 10 clinical parameters: age at the time of surgery, presence of thrill/pulse, gender, diabetes mellitus, ischemic heart disease, congestive heart failure, hypertension, vein size, and artery size [10]. Because Kordzadeh et al. achieved a relatively high accuracy with only these parameters and 266 patients, deep learning is a promising direction of study for predicting post-TPA AVF function.

The related work we report has focused on using clinical parameters, sometimes including DUS or other flow measurements, to predict AVF maturation and early failure. The current approach we will present builds on this by considering the details of PTA, which is used on the majority of AVF, and specifically focusing on post-PTA AVF function. In this study, we introduce the use of hybrid deep learning models to the prediction of AVF function. Hybrid deep learning and hybrid neural networks have been used to describe a variety of different hybridizations of network designs and data. Here, it describes analysis of multi-modal clinical data, through the integration of data-specific neural networks. This

allows our predictive model to add intraprocedural 2D angiography image series to the numerical patient clinical information other researchers have previously studied. The single model can then be trained with a single outcome based on multiple inputs, forcing the model to balance for the strengths and shortcomings of each input source. This type of hybrid model has been shown to work well in other medical research. In a study conducted by M. M. Ahsan et al., hybrid deep learning models analyzing numerical information (age, gender, and temperature) and chest X-Rays were shown to predict with 96.30% accuracy whether a patient was infected with the coronavirus that causes COVID-19 [11]. By utilizing a broader set of available data, hybrid deep learning models have the potential to improve prediction of AVF function. The remainder of the paper is organized as follows: section III describes the proposed hybrid neural network, training strategy, and performance metrics. Section IV describes the data collection, preprocessing, data selection, and hyperparameter optimization, and section V describes the results. Finally, section VI discusses significance of our findings, limitations, and future work, and VII gives our conclusions.

Methodology

To analyze a rich set of numeric and image data, our high-level network design used hybrid deep neural networks to combine output from Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) and parse that output into another ANN, which classifies the survival time. The weights and biases of the hybrid network were optimized using gradient descent and backpropagation. Thus, the parameters of the component ANN's and CNN's were optimized as one network, creating a single model that can offset the limitations of multiple separate models. We explain this in detail in the next section.

Network design—A neuron in an ANN can be represented by the simple equation below, where f is the activation function, b the bias, w_i the weights applied on each input, and x_i the input data.

$$output = f\left(b + \sum_{i=1}^{n} w_i x_i\right)$$

Our multimodal dataset includes 2D digital subtraction angiography (DSA) image series that capture the flow of contrast during PTA, and our network design processed this data using a CNN. CNN are designed to analyze input with spatial or temporal characteristics, such as images or videos because they analyze patches of data together. Thus, they are extensively used in medical imaging and analysis. In a recent study conducted by Zheng et al., CNN's were able to auto-detect and classify low-level colorectal polyps with accuracy of 84.9%, higher than that of visual inspection by endoscopists (74.3%) [12]. In a CNN context, a convolution is a linear operation on patches of data that produces a single output. An activation function is applied to each value in the convolved output. Because the DSA images are in the form of video, we needed a method to analyze spatial data in three dimensions if we consider time to be the third dimension. Thus, our network design utilized 3D Convolutional Layers [13], with time providing the third dimension.

For our CNN, we created convolutional blocks which perform multiple operations (Fig. 2). Four convolutional blocks with progressively larger kernel sizes (5×5×1, 5×5×1, 7×7×1,

9×9×1) were joined together to form a CNN. The four convolutional blocks each contained four layers: 3D convolution, batch normalization, activation (Leaky ReLU), and dropout ($p = 0.08$). Dropout was implemented to reduce network overfitting [14]. 3D max pooling was also applied to the first two blocks to reduce the memory required for the model. The output from the fourth convolutional block was flattened into one single-dimensional array and parsed into a fully-connected layer of 35 neurons. The first ANN consisted of two fully-connected layers, activated by Leaky ReLU. The first layer had ten neurons, and the second five, thus, the network returned an output of five neurons.

To combine the outputs of the ANN and CNN, both were flattened and combined into a single one-dimensional array, which was then used as input for two fully-connected layers to give the final output. The CNN contributed 35 data points and the ANN contributed 5 data points, which were concatenated into a single array of 40. This array formed the input for the final fully connected layers of size 10 and 1, returning a binary output value using a softmax function to normalize the values into a [0, 1] probability. To summarize, the model has an ANN for the numerical data, and a CNN for the imaging data. The outputs of both were combined to become the input for a third ANN which outputs the actual prediction (Fig. 2). Table 1 further details the architecture of the CNN.

Training strategy—The parameters of the combined, hybrid network were trained with respect to the actual output through backpropagation and gradient descent for best possible output prediction. We use He initialization for initial parameter generation [15]. The Gradient Descent method optimized Cross-Entropy (CE) Loss. CE Loss measures the reduction of entropy of an inferential process [16]. Let $y_1, y_2, \cdots, y_n \in \mathbb{R}$ be the possible categories of a predictive model. In a binary prediction scenario, there are only two possibilities, i.e., $y_i \in 0,\ 1$. CE Loss is defined as:

$$Loss(y_1, \cdots, y_n) = -\sum_{y_1, \cdots, y_n} p(y_i) * log_2\ q(y_i)$$

where p is a function that returns the probability of picking a certain classification y_i from the labels (or real output) and q is a function that returns probability of a category y_i being in the prediction from the model. Gradient Descent is an iterative algorithm that updates parameters every iteration based on the gradient of the loss function. Let θ be one of many parameters in the neural network N. Let L be the cross-entropy loss function. As this loss function is based on the prior and posterior probabilities of each category in the output, we can assume that the loss function is related to the parameter θ. Then, we can calculate the partial derivative of the loss function $\frac{\partial L}{\partial \theta}$ with respect to θ. This is back-propagation of the errors. We use the iterative gradient descent algorithm to reduce the size of the cross-entropy loss. To reduce loss, we then shift the value of the parameter in the opposite direction of the gradient. We update θ by subtracting $\frac{\partial L}{\partial \theta}$ multiplied by a constant $\alpha \in (0,\ 1)$ known as the learning rate. Thus, the gradient descent iterative update can be written as follows:

$$\theta_{updated} = \theta - \alpha * \frac{\partial L}{\partial \theta}$$

In this study, we used Stochastic Gradient Descent (SGD) with Momentum. Momentum multiplies the original value of the parameter by a weight $w \in (0,1)$, thus, the iterative equation becomes:

$$\theta_{updated} = w * \theta - \alpha * \frac{\partial L}{\partial \theta}$$

Performance metrics—In this study, we used three metrics – test accuracy, F1-score, and Area Under Receiver Operating Characteristic (AUROC) to analyze the predictive ability of our model. The F1-score is a valuable measurement of the ability of a model to predict true negatives and true positives. The AUROC measures the probability that the value of the model output of a positive case is greater than the value of model output of a negative case. This measurement is calculated as the area under the curve of a Receiver Operating Characteristic (ROC), which is the measurement of the ability of a statistical model to classify at different probability thresholds. The probabilistic output of a softmax layer (final layer of our hybrid network) ranges continuously from 0 to 1. Different thresholds are used to divide the output into binary cases – i.e., if the threshold is 0.5, all cases with softmax output <0.5 can be classified as case '0' and the rest as case '1'. Varying these different thresholds gives us a range of true positive and false positive rates, which, when plotted against each other, form the Receiver Operating Characteristic. Values for AUROC then range from 0.5 to 1. A high AUROC indicates that the predictive model separates the cases very well, as the probability that the softmax value of a case '1' input is higher than a case '0' is high.

Experimental protocol

Two general categories of data were used in this study – patient clinical (numerical) information and images. The hybrid models were built and trained using PyTorch [17]. The experiments were done on an AMD EPYC 7371 16-core processor running at 3.1 GHz, with 256 GB RAM. The actual network training was done using the NVIDIA CUDA framework and an NVIDIA Titan RTX 24 GB VRAM Graphics Processing Unit.

Clinical imaging data collection—The institutional review board approved this retrospective study and waived informed consent for participation. Through retrospective review of our electronic medical records (EMR), we identified 387 patients who underwent PTA for AVF stenosis between January 2013 and December 2018. Archived images for each PTA and pre- and post-procedure DUS were collected and anonymized from our institution's PACS. Baseline demographics and characteristics for patients included in our cohort were collected from EMR.

Cases included in our cohort met the following criteria: PTA interventional DSA with catheter-injected iodinated contrast agent was performed in an identical projection, magnification, and similar catheter position for both pre- and post-PTA acquisitions. Only

single-plane DSA sequences acquired with at least 3 frames per second were included (OEC 9800; GE Healthcare, Waukesha, Wisconsin, or Allura Xper; Philips Healthcare, Inc, Andover, Massachusetts). Cases without pre- and post-procedural ultrasonography were excluded. In total 28 hemodialytic patients had complete data and were included. These included 15 female and 13 male patients. The ethnicities were 32% Black or African American, 32% Asian, 29% White and 7% unknown or refused to provide. The average age was 58 years old with a range of 98 to 19 years old.

Using DSA sequences prior to treatment and following the last treated lesion, interventional AVF blood flow characteristics data were collected using a previously developed angiogram contrast flow analysis tool (PVEC) [17]. Briefly, to quantify blood flow regions of interests (ROIs) are defined in the inflow artery, outflow vein approximately 1 cm from the anastomosis, and the outflow vein at the furthest point from the anastomosis included in the field of view (Fig. 1). The rate of contrast increase in each ROI across the image series is then fit to quantify the blood flow rate. This analysis was performed independently by AL and LC.

Preprocessing for CNN—The imaging data are series of 2D images which capture the flow of injected contrast during the PTA procedure. The raw imaging data are videos of dimensions (512, 512, z) where the time length of the videos is variable. Each patient has two image series: pre- and post-PTA which were, respectively, immediately before PTA was performed and at the completion of PTA. The time length of the videos was normalized by selecting equal chunks centered at the video median from all the videos.

The label used for the supervised learning was ‘Days to Failure’, a discrete number of days following PTA. Currently, fistula failure is classed as early failure when it never developed to the point it can be used within first three months [25]. We converted these into binary classes determined by whether patients showed AVF failure (or required further intervention) within 3 months; because this is roughly the time-scale for maturation of a new AVF, this may be considered a clinical indication of PTA success. We then split the data into 71.4% and 28.6% for training and testing, respectively (20 training cases and 8 testing cases). Image series training data were augmented through standard methods with rotation variations.

Data selection—Decision tree analysis, which separates the dataset iteratively into lower-entropy groups, was implemented to identify the most relevant features for incorporation into the hybrid network. In contrast to existing models, we excluded post-intervention DUS measurements, as these are collected some weeks after PTA [4,18]. While these DUS measurements are highly correlated with AVF time to failure, predicting continued patency so long after PTA limits the clinical value of the model. A selection of features and their correlation to days to failure are summarized in Table 2. The numerical and imaging data were combined and used as input for the hybrid deep learning model.

Hyperparameter optimization—A grid search technique based on a brute-force algorithm was implemented to determine the best hyperparameters in the model training. The grid search program was modeled after Scikit-Learn’s GridSearch program [19] with a

few changes for the hyperparameters of the deep neural network. The optimal permutation of hyperparameters was chosen by picking the models which achieved the lowest average validation loss.

We trained the models in iterative epochs and calculated the loss at the end of each epoch. In each epoch, we split the training data using Leave-One-Out Cross-Validation (LOOCV) and calculated loss for both the training and the validation set [20]. One of the major issues with small datasets is the homogeneity of the feature space, and the limits of the extent augmentation improves that, which can lead to overfitting. LOOCV reduces that potential to overfit by changing the training set at each epoch.

Results

In this section, we show the results of the decision trees, and the most relevant variables selected through exploratory data analysis, the best hyperparameter values picked using grid search, and the training and testing results for different experiments on the hybrid deep learning model.

The decision tree analysis pruned most of the numerical dataset, leaving 8 variables for inclusion in the network (Fig. 3). The variables: Age at Fistulaplasty (years)', 'Reason for Intervention', 'Pre-Proximal Qt 2'(representing the blood flow rate at proximal vein before PTA procedure), 'Pre-Proximal cmax 1' (representing the maximum contrast at proximal vein before PTA procedure), 'Pre-Proximal cmax 2' (representing the maximum contrast at proximal vein before PTA procedure), 'Post-Proximal cmax 2'(representing the maximum contrast at proximal vein post PTA procedure), 'Pre US anastomosis (representing ultrasound flow measurement at anastomosis before PTA procedure)', 'Pre US proximal vein (representing ultrasound flow measurement at proximal vein before PTA procedure)' were the ones found to have the highest feature importance, i.e., provide the most information gain and were thus kept for the analysis. Of these, the first two are clinical variables, and the next four are PVEC blood flow parameters calculated as described above in the Clinical Imaging Data Collection subsection of the Experimental Protocol and in a previous publication [17], where the suffixes '1' and '2' indicate the operator performing the analysis. The last two are DUS with the location of the measurement as a suffix.

The grid search algorithm finds optimal values for hyperparameters by finding the values that minimize average validation loss. Fig. 4 shows the specific ranges for each hyperparameter and plots average validation loss against hyperparameter values for learning rate, number of epochs, batch size, and gradient momentum. The peak values were as follows: Learning rate = 0.1, Number of Epochs = 20, Batch Size = 10, and Momentum = 0.9.

The results for the main experiment containing all streams of data are an accuracy of 0.75, an F1-score of 0.75 and an AUROC of 0.75. Fig. 5(a) shows the Receiver Operating Characteristic and the AUROC of 0.75. This means that in the output predictions, a true positive output (Case '1') has a probability of 0.75 of having a higher output softmax value than a true negative output (Case '0'), indicating that the softmax layer output of the model

is segregated well between the two cases. The precision and recall for the neural network are both 0.75, as well. Fig. 5(b) shows the confusion matrix of the predictions on the testing dataset. The confusion matrix and the F1-score indicate that the predictions have good precision and recall, i.e., the model can pick true positives and true negatives well.

Multiple experiments were conducted to identify the value of flow and imaging data, as shown in Table 3. We removed imaging data, then flow data, and then both from the deep learning models to identify the importance of each modality of data. The worst results (0.5 accuracy, 0.5 F1-score, 0.3 AUROC) were achieved when only clinical data were parsed, and the best results (0.75 accuracy, 0.75 F1-score, 0.75 AUROC) were achieved with all data. Thus, results of the experiments show that the performance of the model is consistently improved by incorporating the interventional blood flow data and imaging data.

Discussion

Significance of findings

In this research, we present a novel method with components of a hybrid model combining a 3D convolutional neural network with a multilayer perceptron to processes multimodal data. The method integrates different computational network structures to process a variety of types of data such as images and clinical records for predictions. The potential impact of this research on patient treatment is that this hybrid model incorporates only pre-procedure multimodal data for prediction, giving it potential as a real-time procedural evaluation tool to provide feedback during the PTA procedure. The value of considering multimodal data is shown by the experiments where different kinds of data are excluded from the network, specifically the value of the intraprocedural blood flow data (PVEC) and the PTA interventional images. In our results, including either flow data or imaging appear to produce the same predictive result from the model, suggesting that including data on blood flow within the AVF is key to identifying whether the AVF will fail within the time period in question. While previous research into AVF has sometimes included DUS or other flow data, the use of interventional image data and incorporating a CNN into a hybrid network is new [7,21,22].

In terms of biomedical findings, our results are generally consistent with previous reports in medical journals. However, with decision trees we found that for our dataset, fistula configuration, specifically brachiocephalic AVF, was not a key parameter for determining long-term AVF function (Fig. 3), unlike the descriptive study of Takahashi et al. using Cox proportional hazards models [5]. Our findings of the importance of blood flow for AVF function prediction are complementary to those of Zhu et al., in that residual stenosis is an important determinant of blood flow through the AVF. As with the work of Ferring et al. and Genek et al., DUS measurements were an important predictive factor, although our multimodal approach to AVF failure appears significantly more accurate than that of Ferring et al. [7,18]. The accuracy (~75%) of the hybrid model we propose on the test dataset is comparable to that of the clinical variable-only neural network of Kordzadeh et al. [10]. A key difference is that their model predicts AVF maturation, whereas our model predicts AVF function following PTA, which to date has received much less study. Considering that their model was also trained on a significantly larger dataset, there is likely still room for

our model performance to improve. Moreover, future work to include treatment-related data such as hospital procedure parameters and different imaging modalities such as 3D DSA, contrast-enhanced computed tomography (CE-CT) images and ultrasound imaging should also be integrated in the model to improve predictions that captures the characteristics of a particular practice group, devices, and patient groups.

Overall, these results for the hybrid neural network approach are encouraging and highlight the need for further research. The high F1-score and AUROC, plus the training and validation graph, indicate that the model has learned some inherent differences in the feature space.

Limitations

The purpose of this study is to present methodology utilizing a hybrid network to combine different types of data including clinical imaging, clinical information, and blood flow. We present a retrospectively collected dataset to test this approach. In general, retrospective studies may suffer from incomplete or inconsistent data. When the dataset is complex and multimodal, this concern is amplified. To minimize this, our study size was 28 patients with complete data identified out of approximately 5 years of PTA cases in our center. Due to the training and testing split, the models were trained on 20 cases (71%). The image data were augmented through standard methods, but ultimately, this is a small sample size to train deep learning models. It is possible that the models may be learning some sample bias. This bias could be one of many possibilities – as we were looking for the cleanest and most complete data; we are possibly including patients for whom the medical care was excellent. While we attempted to mitigate the effects of the small sample size through LOOCV, we cannot assume that these models will perform as well on other fistula data procured from other means, and ultimately further research with a larger cohort is required for future studies.

However, the high F1-Score indicates that the model does well on both precision and recall, meaning a low false positive and false negative error. A high AUROC means that the model's probabilities for each binary case are much separated, thus there is little confusion in the model between separating out the two cases. This is very encouraging as it means that the model has identified a latent division at least in the small sample between the two types of patients based on the streams of data and suggests that Deep Learning techniques could find the latent division in a larger sample as well.

Future work

Our results show that there are unique differences between the classes of post-PTA AVF function, and that they can be identified by a deep learning model. The clearest future direction is to continue this work with a larger sample set. Building on our findings presented here, the sample size limitations of this work will be addressed through homogeneous, prospectively collected data. This will reduce any bias that may be present in the trained model and allow us to quantify the extent to which images and the derived flow information provide unique features for the hybrid neural network.

In this project, we split the continuous variable of post-PTA "Days to Failure" into a discrete, clinically-significant, binary variable (AVF function of 3 months) to do

classification. Another future approach may be to use “Days to Failure” specifically as a continuous variable for regression modeling using the hybrid deep learning models. This could be accomplished with some small modifications to the current network model, such as changing the final softmax layer to a layer that scales the output to correlate with “Days to Failure.” A larger dataset would support this development, which might increase the clinical utility of this predictive model. Other possibilities for improving the utility of our model include modifying our preprocessing methods to perform interpolation between frames, allowing the model to more effectively analyze different interventional DSA imaging framerates.

We used a generic grid search to explore minimum loss solutions in the hyperparameter space, which we believe was sufficient for the current dataset. However, there are other algorithms, such as Randomized Search and Tree-structured Parzen Estimator Approach (TPE), which have been shown to perform better for the convex numerical dataset [23]. With a future dataset, these may allow us to further optimize the hyperparameters for improved performance.

Finally, the work presented here is a single piece for improving clinical care for hemodialysis patients through biomedical informatics. Other related research includes the identification of stenosis, a major cause of AVF failure and indication for PTA. Recently, Wang et al. proposed a wireless blood-flow sound recorder based on stethoscope auscultation for monitoring fistula narrowing and failure, and achieved a precision of 87.84% at detecting stenosis [24]. In combination with new tools such as this, the approach we present here may improve clinical care for kidney failure, which remains a prevalent and increasingly common disease.

Conclusion

Our experiments show that hybrid deep learning models that process multimodal biomedical data are useful for analyzing AVF. We also found that the interventional image-based blood flow measurements improved the performance of the hybrid deep learning models. Based on our preliminary model, we are able to predict with 75% accuracy cases that fail within 3 months of PTA. This information can inform clinical follow-up planning and identify those patients who might be at risk of PTA failure. A future study based on a large cohort could further improve performance, refine the accuracy for clinical application and lead to a viable, real-time product for medical use.

Acknowledgement

This work was supported in part by NIH Grant R01HL152270 and a UCLA Radiology Exploratory Research Grant Award.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to PHI but are available from the corresponding author on reasonable request.

References

[1]. Momeni A, Mardani S, Kabiri M, Amiri M. Comparison of complications of arteriovenous fistula with permanent catheter in hemodialysis patients: a six-month follow-up. Adv Biomed Res 2017;6:106. 10.4103/2277-9175.213666 (in eng). [PubMed: 28904934]

[2]. Schinstock CA, Albright RC, Williams AW, Dillon JJ, Bergstralh EJ, Jenson BM, et al. Outcomes of arteriovenous fistula creation after the fistula first initiative. Clin J Am Soc Nephrol : CJASN 2011;6(8):1996–2002. 10.2215/CJN.11251210 (in eng). [PubMed: 21737851]

[3]. Lee T, Qian JZ, Zhang Y, Thamer M, Allon M. Long-term outcomes of arteriovenous fistulas with unassisted versus assisted maturation: a retrospective national hemodialysis cohort study. J Am Soc Nephrol 2019;30(11):2209–18. 10.1681/ASN.2019030318 (in eng). [PubMed: 31611240]

[4]. Lok CE, Huber TS, Lee T, Shenoy S, Yevzlin AS, Abreo K, et al. KDOQI clinical practice guideline for vascular access: 2019 update. Am J Kidney Dis 2020/04/01/2020;75(4):S1–164. 10.1053/j.ajkd.2019.12.001 (in en). [PubMed: 32778223]

[5]. Takahashi EA, Harmsen WS, Misra S. "Endovascular arteriovenous dialysis fistula intervention: outcomes and factors contributing to fistula failure," (in eng). Kidney Med 2020/06//May-undefined 2020;2(3):326–31. 10.1016/j.xkme.2020.02.004.

[6]. Zhu Z-R, Zou L, Xing Y, Tan Y-C, Xu G-J, He Z-J, et al. Predictors of primary patency after percutaneous balloon angioplasty for stenosis of Brescia-Cimino hemodialysis arteriovenous fistula. Br J Radiol 2020;93(1109):20190505. 10.1259/bjr.20190505 (in eng). [PubMed: 32101462]

[7]. Gibyeli Genek D, Tuncer Altay C, Unek T, Sifil A, Seçil M, Camsari T. Can primary failure of arteriovenous fistulas be anticipated? Hemodial Int 2015;19(2):296–305. [PubMed: 25170532]

[8]. Rezapour M, Khavanin Zadeh M, Sepehri MM. Implementation of predictive data mining techniques for identifying risk factors of early AVF failure in hemodialysis patients. Comput Math Methods Med 2013;2013:830745. 10.1155/2013/830745. [PubMed: 23861725]

[9]. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. Comput Methods Progr Biomed Jul 2018;161:1–13. 10.1016/j.cmpb.2018.04.005 (in eng).

[10]. Kordzadeh A, Esfahlani SS. The role of artificial intelligence in the prediction of functional maturation of arteriovenous fistula. Annals of vascular diseases 2019; 12(1):44–9. 10.3400/avd.oa.18-00129 (in eng). [PubMed: 30931056]

[11]. Ahsan MM, Alam TE, Trafalis T, Huebner P. Deep MLP-CNN model using mixed-datato distinguish between COVID-19 and non-COVID-19 patients. Symmetry 2020; 12.

[12]. Zhang R, Zheng Y, Mak TWC, Yu R, Wong SH, Lau JYW, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. IEEE J Biomed Health Inf 2017;21(1):41–7. 10.1109/JBHI.2016.2635662.

[13]. Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 2013;35(1):10. 10.1109/TPAMI.2012.59.

[14]. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. 2012. arXiv preprint arXiv:1207.0580.

[15]. Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. 30th international conference on machine learning, 28; 2013. 02/13/2013.

[16]. Ramos D, Franco-Pedroso J, Lozano-Diez A, Gonzalez-Rodriguez J. Deconstructing cross-entropy for probabilistic binary classifiers. Entropy 2018;20(3):208. [PubMed: 33265299]

[17]. Chien A, Wang Y-L, McWilliams J, Lee E, Kee S. Venographic analysis of portal flow after TIPS predicts future shunt revision. Am J Roentgenol 2018;211(3):684–8. [PubMed: 30085841]

[18]. Ferring M, Henderson J, Wilmink T. Accuracy of early postoperative clinical and ultrasound examination of arteriovenous fistulae to predict dialysis use. J Vasc Access Jul-Aug 2014;15(4):291–7. 10.5301/jva.5000210 (in eng). [PubMed: 24500848]

[19]. Fabian Pedregosa GV, Gramfort Alexandre, Vincent Michel, Bertrand Thirion, Grisel Olivier, Blondel Mathieu, et al. Scikit-learn: machine learning in Python. JMLR 2011;12:5.

[20]. Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recogn 2015;48(9):7.

[21]. Saucy F, Haesler E, Haller C, Déglise S, Teta D, Corpataux J-M. Is intra-operative blood flow predictive for early failure of radiocephalic arteriovenous fistula? Nephrol Dial Transplant 2010;25(3):862–7. 10.1093/ndt/gfp577. [PubMed: 19892754]

[22]. Ahsan MM, Alam TE, Trafalis T, Huebner P. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. Symmetry 2020/09//2020;12(9):1526. 10.3390/sym12091526 (in en).

[23]. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 2011;24:2546–54.

[24]. Wang H, Wu C, Chen C, Lin B. Novel noninvasive approach for detecting arteriovenous fistula stenosis. IEEE Trans Biomed Eng 2014;61(6):1851–7. 10.1109/TBME.2014.2308906. [PubMed: 24845295]

[25]. Gjorgjievski N, Dzekova-Vidimliski P, Gerasimovska V, Pavleska-Kuzmanovska S, Gjorgievska J, Dejanov P, et al. Primary failure of the arteriovenous fistula in patients with chronic kidney disease stage 4/5. Maced J Med Sci 2019;7(11):1782–7. 10.3889/oamjms.2019.541.

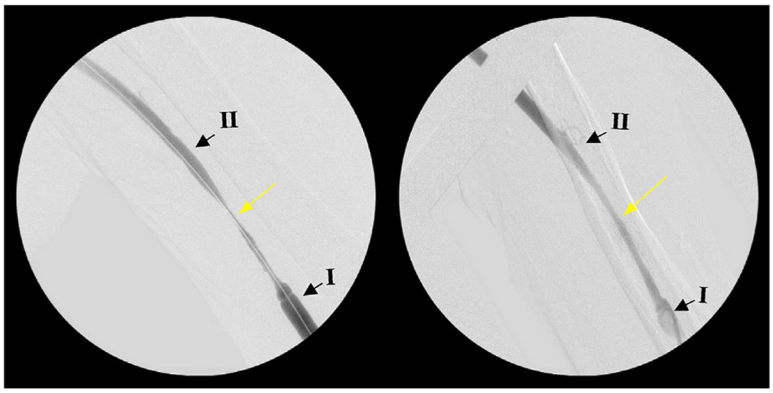

Fig. 1.

Clinical images of AVF in kidney failure patient. PTA interventional digital subtraction angiography (DSA) images pre- and post (left and right images respectively). The yellow arrows indicate a region of stenosis treated with PTA. (I) Indicates ROI for outflow vein approximately 1 cm from the anastomosis, and (II) shows the ROI for the distal outflow vein.

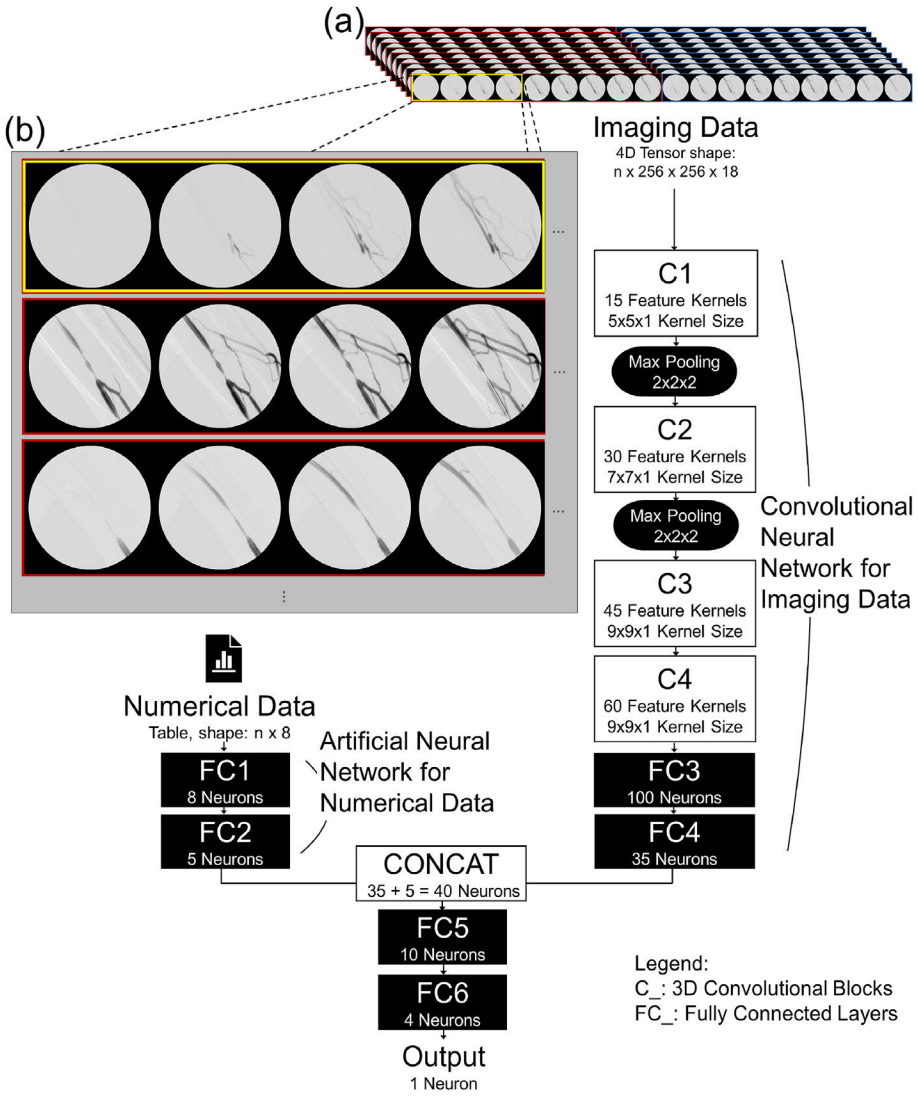

Fig. 2.

(a) Block diagram of the MLP-CNN Hybrid Neural Network. The model analyzes two data streams separately and pools the data by concatenating extracted features. Input imaging data consists of a series of 9 pre-TPA images (outlined in red) and 9 post-TPA images (outlined in blue) (b) Inset shows sample portions of three interventional DSA image series.

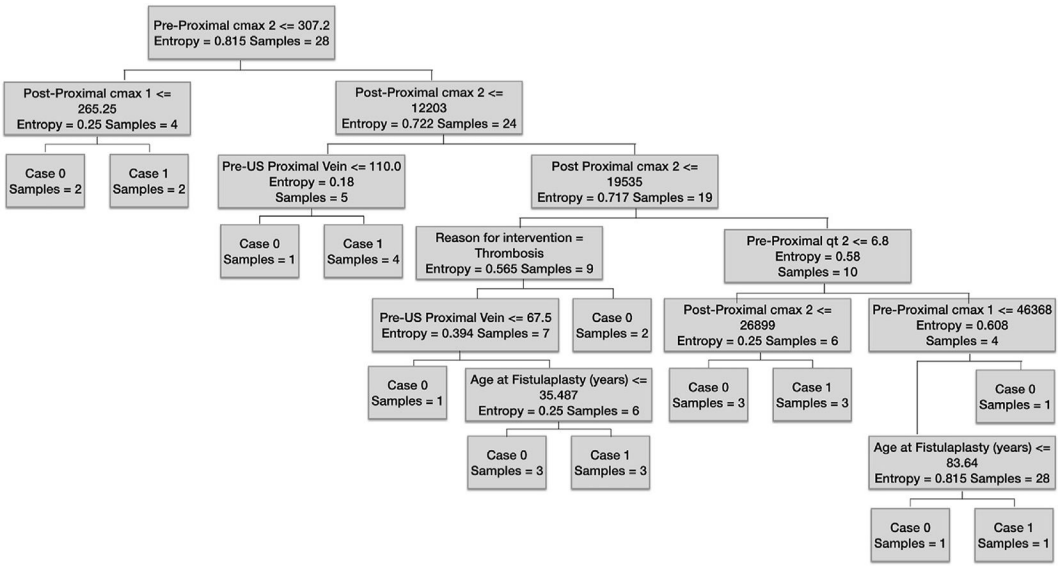

Fig. 3.

Decision tree analysis was used on the numerical features of the dataset to select features for inclusion in the neural network. The selected features were ‘Age at fistulaplasty (years)’, ‘Reason for intervention’, ‘Pre-Proximal Qt 2′, ‘Pre-Proximal cmax 1’, ‘Pre-Proximal cmax 2’, ‘Post-Proximal cmax 2’, ‘Pre US anastomosis’, and ‘Pre US proximal vein’.

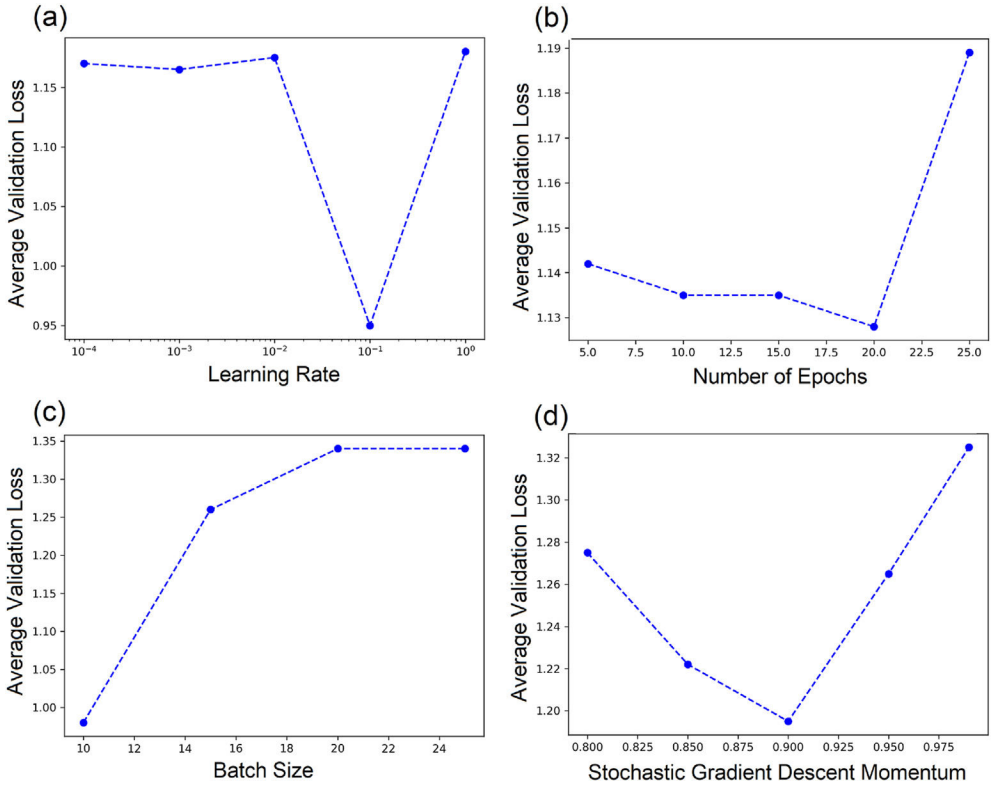

Fig. 4.

GridSearch was used to optimize hyperparameters for the network. Other hyperparameters are kept constant while varying the hyperparameter of interest. The best values for hyperparameters Learning Rate, Number of Epochs, Batch Size, and SGD momentum were determined using GridSearch. (a) Optimization of SGD learning rate = 0.1. (b) Training iterations (epochs) = 20. (c) Batch size = 10. (d) SGD momentum = 0.9.

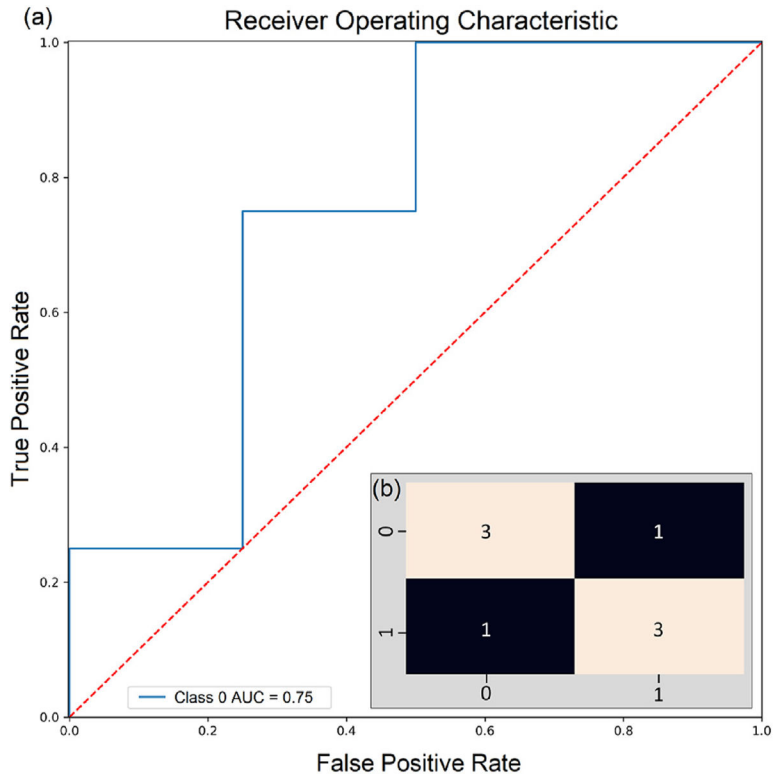

Fig. 5.

(a) ROC curve indicates how well the model differentiates the binary categories with AUROC of 0.75. (b) Inset presents confusion matrix for test cases using fully trained model. The hybrid network model predicts 4 cases as 1 (AVF function > than 3 months) out of which 3 are correct. Network predicts 3 cases as 0 (AVF function < than 3 months) out of which 3 are correct. Accuracy, F1 score, Recall, and Precision = 0.75.

Table 1

Details of convolutional neural network.

Layer	Sub-layers	Hyperparameters
Layer 1	3D Convolution (C1) Batch Normalization ReLU Dropout 3D Max Pooling	Feature Maps: 15, Kernel Size: 5×5×1 p = 0.08 Kernel Size: 2×2×2, Stride:2
Layer 2	3D Convolution (C2) Batch Normalization ReLU Dropout 3D Max Pooling	Feature Maps: 30, Kernel Size: 7×7×1 p = 0.08 Kernel Size: 2×2×2, Stride:2
Layer 3	3D Convolution (C3) Batch Normalization ReLU Dropout 3D Max Pooling	Feature Maps: 45, Kernel Size: 9×9×1 p = 0.08
Layer 4	3D Convolution (C4) Batch Normalization ReLU Dropout 3D Max Pooling	Feature Maps: 60, Kernel Size: 9×9×1 p = 0.08
Layer 5	Perceptron Layer (FC3) Perceptron Layer (FC4)	Neurons: 100 Neurons: 35

(The convolutional section of the hybrid network has four blocks of layers with increasing feature maps and kernel size with a final perceptron block. The output is concatenated with the numerical section that consists of a standard deep neural network of two layers to form the input for the combined deep network).

Table 2

Pearson's correlation coefficient to days to failure.

Feature	Correlation
Age of patient at fistulaplasty	0.03
Age of the fistula	0.21
Time since last fistula intervention	0.34
Number of stenoses treated	0.12
Sex	0.18
Pre-PTA DUS timing (days before fistulaplasty)	0.23
Pre-PTA DUS artery peak systolic flow	0.09
Pre-PTA DUS peak systolic flow through anastomosis	−0.20
Pre-PTA DUS Proximal vein peak systolic flow	−0.13
Pre-PTA (Interventional) PVEC Q_τ proximal vein	−0.15
Post-PTA (Interventional) PVEC Q_τ proximal vein	0.05
Post-PTA DUS timing (days after fistulaplasty)	0.27
Post-PTA DUS artery	0.47
Post-PTA DUS anastomosis	−0.04
Post-PTA DUS proximal vein peak systolic flow	−0.13
Interventional PVEC Q_τ proximal vein change	−0.10

Table 3

Model performance with different subsets of data.

Experiment	Accuracy	F1-Score	AUROC
All	0.75	0.75	0.75
Without Flow Data	0.5	0.5	0.5
Without Imaging Data	0.5	0.5	0.5
Without Flow and Imaging Data	0.5	0.5	0.31

(Flow data refers to blood flow characteristics extracted from images by PVEC, while Imaging Data refers to the DSA image series analyzed by the CNN. Each category listed includes clinical data—both the patient information gathered before or at the time of the procedure, and pre-intervention DUS).