AI

MDPI

Article

Data-Driven Modeling and Classification of Brain Blood-Flow Pathologies

Irem Topal [1,2], Alexander Cherevko [3], Yuriy Bugai [3], Maxim Shishlenin [4], Jean Barbier [5], Deniz Eroglu [2,6], Édgar Roldán [5] and Roman Belousov [7,*]

1 Department of Biomedical Sciences, University of Padova, 35131 Padova, Italy; irem.topalkement@unipd.it
2 Faculty of Engineering and Natural Sciences, Kadir Has University, Istanbul 34083, Türkiye; deniz.eroglu@khas.edu.tr
3 Lavrentyev Institute of Hydrodynamics, 630090 Novosibirsk, Russia; cherevko.ngs@yandex.ru (A.C.); y.v.bugay@gmail.com (Y.B.)
4 Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia; m.a.shishlenin@mail.ru
5 ICTP—The Abdus Salam International Centre for Theoretical Physics, 34151 Trieste, Italy; jbarbier@ictp.it (J.B.)
6 Department of Mathematics, Imperial College London, London SW7 2AZ, UK
7 Cell Biology and Biophysics Unit, EMBL—European Molecular Biology Laboratory, 69117 Heidelberg, Germany
* Correspondence: roman.belousov@embl.de; Tel.: +49-6221-387-8367

Abstract

Cerebral aneurysms and arteriovenous malformations are life-threatening hemodynamic pathologies of the brain. While surgical intervention is often essential to prevent fatal outcomes, it carries significant risks both during the procedure and in the postoperative period, making the management of these conditions highly challenging. Parameters of cerebral blood flow, routinely monitored during medical interventions or with modern noninvasive high-resolution imaging methods, could potentially be utilized in machine-learning-assisted protocols for risk assessment and therapeutic prognosis. To this end, we developed a linear oscillatory model of blood velocity and pressure for clinical data acquired from neurosurgical operations. Using the method of Sparse Identification of Nonlinear Dynamics (SINDy), the parameters of our model can be reconstructed online within milliseconds from a short time series of the hemodynamic variables. The identified parameter values enable automated classification of the blood-flow pathologies by means of logistic regression, achieving a balanced accuracy of 74%. Our results demonstrate the potential of this model for both diagnostic and prognostic applications, providing a robust and interpretable framework for assessing cerebral blood vessel conditions.

Keywords: data-driven modeling; diagnostics; hemodynamics; cerebral aneurysm; cerebral arteriovenous malformation

Academic Editors: Yeliz Karaca, Majaz Moonis and Yudong Zhang

Received: 10 December 2025
Revised: 9 February 2026
Accepted: 3 March 2026
Published: 11 March 2026

1. Introduction

Intracranial arterial aneurysms (AA) and arteriovenous malformations (AVM) are cerebral blood-flow defects that can lead to life-threatening hemorrhage or neurological pathologies [1,2]. Often, these threatening conditions warrant surgical intervention. However, due to the significant uncertainty of the risks associated with both the pathologies and their treatment, determining whether such an intervention is advisable is a challenging medical question even for experienced specialists, who must judge on a wide range of complex factors.

AAs represent swollen areas caused by weakened blood-vessel walls, which can expand and rupture. No unified theory of their genesis exists, though most researchers attribute aneurysms to degenerative changes in the vascular walls caused by various factors. An aneurysm consists of a neck, body, and dome, with the dome formed by a single layer of intima (thinnest)—the most vulnerable part where ruptures typically occur. The risk of hemorrhage from an unruptured aneurysm is about 1% per year, but a recurrent hemorrhage is significantly more frequent: 15–25% within the first two weeks and 50% within the first six months. The risk of hemorrhage also increases with the size of the aneurysm: defects under 5 mm have a lifetime rate of 2.5%. Aneurysms of 6–10 mm in size burst in 41% of cases, and those in the range of 11–15 mm in 87%. Repeated hemorrhages are generally more severe than initial ones, with a fatality rate of 32% in the first week, 43% by the second week, and up to 63% within the first year. Surgical treatment for cerebral AAs involves either open aneurysm clipping or conservative endovascular methods, though the latter is only applicable to pathologies of a small size.

AVMs are congenital pathologies manifested as entangled blood vessels that disrupt the oxygen supply of the surrounding tissues, thereby provoking cell death. In severe cases, they can rupture, resulting in intracranial hemorrhage with the frequency reported in the range of 30–82% [3]. The high morbidity and disability rates highlight the dangers associated with this pathology. A common method of treating AVMs is embolization: embolizing agents, such as N-butyl-2-cyanoacrylate or a non-adhesive copolymer of ethylene and vinyl alcohol (ONYX) [4], are delivered by a micro-catheter to block the pathological vessels. The mortality rate of treated patients with AVMs achieves 9.3% during the period of hospitalization.

The decision on surgical intervention requires a careful risk assessment before and after the treatment, as a re-operation may be deemed necessary [5–7]. Therefore, modeling hemodynamic flows in patients with AAs and AVMs is a mathematical problem of utter practical importance.

Due to the increasing role of computational methods in medicine, with the rapid improvement in dataset quality and methods of their processing, there is a growing interest in applying machine learning to clinical data to assist medical decision-making, treatment planning, and execution [8–12]. In this report, we explore the use of modern machine learning techniques on data acquired during the treatment of AAs and AVMs in real time. In particular, we leverage the Sparse Identification of Nonlinear Dynamical Systems (SINDy) [13] as the model-discovery framework and apply logistic regression to characterize and to automatically classify blood flow with pathologies before and after surgery.

Our proposed methodology offers a novel approach to analyzing cerebral blood flow in patients undergoing surgery for cerebral blood vessel malformations. Classification of the identified model parameters could help medical doctors make more informed decisions in surgical planning. Additionally, our findings highlight the prospects of using larger datasets to design automatic diagnostic tools for blood pathologies or methods for forecasting the outcome of a surgical treatment.

1.1. Background

During surgical interventions, physical parameters of patients' blood flow, such as velocity $v(t)$ and pressure $p(t)$, are continuously monitored [14–16], as illustrated in Figure 1. While detailed physical models for their dynamics are fairly complicated [12,17–21], in this work, we adopt a phenomenological approach from Refs. [15,16], where the time series of

these hemodynamic variables are fitted to a family of parsimonious nonlinear equations of the Liénard type [22]:

$$\ddot{p}(t) + A[p(t)]\,\dot{p}(t) + B[p(t)]p(t) = \epsilon v(t), \tag{1}$$

where $A(p) = a_0 + a_1 p + a_2 p^2 + \cdots$ and $B(p) = b_0 + b_1 p + b_2 p^2 + \cdots$ are algebraic polynomials in $p(t)$ with coefficients $a_{0,1,2\ldots}$ and $b_{0,1,2,\ldots}$.

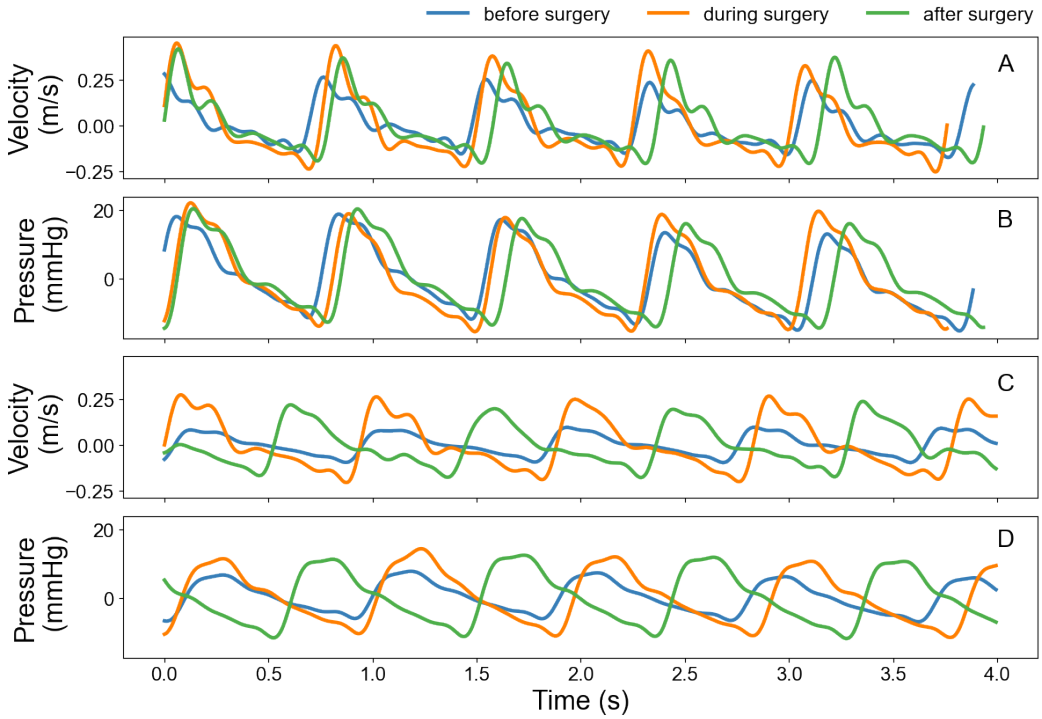

Figure 1. Examples of time series for blood flow velocity (**A**,**C**) and pressure (**B**,**D**) in clinical data before, during, and after surgery: patient with an arterial aneurysm (**A**,**B**), and patient with an arteriovenous malformation (**C**,**D**). The pressure and velocity values (vertical axes) are demeaned and span five cardiac cycles (Section 2.1).

Simulations of Equation (1) represent the direct (well-posed) Cauchy problem with the initial-value conditions $p(0) = p_0$ and $\dot{p}(0) = \dot{p}_0$. In Refs. [15,16], the coefficients of algebraic polynomials $a_{0,1,2\ldots}$, $b_{0,1,2,\ldots}$ and the initial conditions p_0 and $\dot{p}_0$ are determined from experimental data $[p(t_i), v(t_i)]_{i=0,1,2,\ldots,n}$ using techniques from inverse-problem theory [23,24]. However, this approach is computationally expensive and sensitive to initial parameter guesses [25,26], as it often involves iterative, high-dimensional optimization, especially with nonlinear polynomial terms, making it costly.

Common regression techniques, including Least Squares, LASSO, and Ridge regression, often face limitations in online short-latency applications due to high computational demands or sensitivity to initial parameter guesses [27]. Moreover, fitting data to a minimal model risks misinterpretation, as patient-specific variations may not be adequately captured by a universal set of terms. A robust approach would, therefore, benefit from starting with a larger model framework that can adaptively reduce complexity based on individual patient data. To address the computational challenge, L_2-norm-based approaches, like Ridge regression, are often employed for their speed. However, while effective at controlling over-fitting, they do not inherently reduce the number of terms, resulting in models that may remain unnecessarily complex. Conversely, L_1-norm-based methods such as LASSO can achieve sparsity by eliminating terms with minimal contribution, but their iterative nature makes them too slow for online short-latency applications.

1.2. Overview

The data-driven discovery of dynamical systems offers a novel approach to modeling and understanding complex real-world systems [28]. Expanding on the previous works [14–16], we explore this approach and present advances in two directions: data-driven modeling of the blood-flow dynamics with short-latency identification of the system parameters, and the potential clinical applications of the relationships learned from this data.

To address the limitations of the original model (1), we expanded the basis functions by constructing a more comprehensive candidate library, $\mathbf{\Theta}$, which includes not only polynomials in $p(t)$ and its time derivatives, but also other nonlinear terms:

$$\ddot{p} + c_{01}\dot{p} + c_{02}\dot{p}^2 + c_{03}\dot{p}^3 + (c_{10} + c_{11}\dot{p} + c_{12}\dot{p}^2)p + (c_{20} + c_{21}\dot{p})p^2 = \epsilon v, \tag{2}$$

where c_{ij} are the coefficients to be determined. This expanded equation allows us to assess the contribution of higher-order terms to the accuracy of the model and determine whether the original model is optimal in terms of fit and simplicity.

To determine the coefficients c_{ij}, we apply the Sparse Identification of Nonlinear Dynamical Systems (SINDy) method [13] with Sequentially Thresholded Least Squares (STLS) optimization, which is computationally efficient and converges to a sparse solution rapidly as an alternative to the traditional regression techniques [29]. By adjusting the sparsity threshold, we systematically eliminate terms, with the original equation (Equation (1)) being recoverable for appropriate threshold values. This trade-off between simplicity and computational feasibility makes SINDy a suitable method for online short-latency applications. A detailed explanation of the SINDy method is provided in Methods.

Since the inception of the SINDy methodology, its limitations have been explored and extensions have been proposed to address outstanding challenges [30–49]. However, because the clinical data in our possession consist of time series with low levels of noise, and the forcing term $v(t)$ partially incorporates much information about latent variables in our model, these limitations do not present difficulties in our work. Moreover, since the number of candidate basis functions is much smaller than the length of the time series, the regression problem is overdetermined, favoring stable $L2$-regularized solutions [50]. Therefore, we are able to apply the original methodology already implemented in an open-source code [51,52].

SINDy has been shown to work well with libraries composed of truncated Taylor expansions [13,53]. Therefore, we start from the most general Liénard system based on the second-order power series (2) in $(p, \dot{p})$. By varying the sparsity threshold (η), we sought a balance between the model complexity and its accuracy. Observing that the fitting and prediction error remains unchanged or even decreases (for AVM patients) while η increases and, thus, the higher-order terms are progressively eliminated (Section 3.1), we identified the value of the sparsity threshold that robustly renders a linear structure of the governing equation for all patients in our clinical data.

Specifically, in the regression procedure [29], we simplify the inverse problem by setting $p(0)$ to the experimental value $p_0 = p(t_0)$, estimating $\dot{p}(0)$ from the forward-difference formula $p_0 = [p(t_1) - p(t_0)]/(t_1 - t_0)$, thus excluding these values from the fitting parameters. By setting the sparsity threshold at $\eta = 1.0$, the nonlinear terms of order $O(p^3)$ are eliminated, and by $\eta = 5.0$, Equation (2) is reduced to a linear model of a forced damped harmonic oscillator

$$\ddot{p}(t) + a\,\dot{p}(t) + b\,p(t) = \epsilon v(t), \tag{3}$$

where a, b and ϵ are constants. This linear approximation retained only three key parameters—a, b, and ϵ—resulting in a model that strikes the balance between simplicity and accuracy.

Our findings indicate that, while achieving a slightly better forecasting accuracy (Section 3.4), the complex nonlinear model proposed in Refs. [15,16] may be overly detailed. The simplified linear model, with only three parameters, effectively captures the essential dynamics of blood flow and offers a more practical, interpretable solution. We specifically evaluated the reproducibility of parameter values across datasets, observing that, even in the worst case, the relative deviation of the parameters did not exceed 24% (see Results). In addition, by reducing the fitting problem to solving an overdetermined linear system, SINDy does not require an initial parameter guess, which is one of the critical aspects in the original technique of Refs. [15,16]. This efficiency enables rapid model construction at scale, making it well-suited for clinical and short-latency applications.

Further, we explore whether the fitted values of the model parameters, viewed as means of data-dimensionality reduction, could act as predictor variables for pathologies or treatment prognosis in higher-tier machine learning models. To pursue this line of inquiry, we applied logistic regression to classify blood-flow anomalies and surgically treated vessels based on the parameters of Equation (3) inferred from the patient data. In this application, our automatic classifier discriminates between three classes—flows with AA, AVM, and in surgically treated vessels—with a balanced accuracy of 74%, which is a surprisingly promising result given the limited data.

2. Materials and Methods

In this study, we employ a multi-step approach to analyze clinical data and apply machine learning for classification. First, the clinical data is reconstructed using sparse regression methods, which help identify relevant features while reducing model complexity. Next, we apply logistic regression to classify the reconstructed data into distinct categories of blood flow anomalies, based on clinical outcomes. The following subsections describe the details of the data acquisition, the sparse regression procedure, and the classification methodology [29].

2.1. Clinical Data

Blood velocity and pressure were recorded using a Doppler sonography from an intravascular guidewire with a diameter 0.34 mm. The signals were processed through an analogue-to-digital converter at a frequency of 200 Hz. A digital low-pass filter of order $N = 198$ with a Hamming window and a cutoff frequency of 7 Hz was used for noise suppression (Nyquist frequency 99 Hz, sampling frequency 198 Hz). Filtering with identical parameters was applied to both pressure and velocity signals in a single forward pass, followed by compensation of the filter group delay ($N/2$ points), resulting in an effective zero-phase response.

Velocity and pressure data were collected before, during, and after surgery from ten patients: five with AAs, and five with AVMs, all of whom underwent a successful surgical treatment. The boundaries of individual cardiac cycles were determined by the points of local minima in the pressure signal (diastolic minimum). Five consecutive cardiac cycles were extracted from the recording for further analysis. Each time series thus lasted between 3.345–5.375 s, with a time step of $\Delta t = 5$ ms. The time-averaged values of pressure and velocity were subtracted from each trajectory before fitting the data to theoretical models.

2.2. Sparse Identification of Blood-Flow Dynamics

Many dynamical systems of variables $\boldsymbol{x}(t) \in \mathbb{R}^n$ are described by governing equations of the form $\dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t)]$, which contain just a few terms in the expression for $\boldsymbol{f}(\boldsymbol{x})$. Therefore, such equations are sparse in some basis (set) of functions. SINDy was motivated by this observation and aims to solve this sparse identification problem [13,51]. In the same spirit, given an observed dependency of velocity on time, we are looking for a governing equation of pressure dynamics in the form of a second-order differential equation:

$$\ddot{p}(t) = f[p(t), \dot{p}(t), v(t)], \tag{4}$$

given the initial values $p(0)$ and $\dot{p}(0)$.

In vector form, we denote the time series as $\boldsymbol{P} = [p(t_0), p(t_1), \ldots, p(t_n)]$, $\dot{\boldsymbol{P}} = [\dot{p}(t_0), \dot{p}(t_1), \ldots]$, and $\boldsymbol{V} = [v(t_0), v(t_1), \ldots]$ for the consecutive instances of time $t_{i=0,1,2,\ldots,n}$. By performing a sparse regression on the linear equation

$$\ddot{\boldsymbol{P}} = \boldsymbol{\Theta}(\dot{\boldsymbol{P}}, \boldsymbol{P}, \boldsymbol{V})\, \boldsymbol{\Xi}, \tag{5}$$

we seek a vector of coefficients, $\boldsymbol{\Xi}$, that identifies the system's dynamics within the basis of candidate functions such as $p(t)$, $\dot{p}(t)$, $p^2(t)$, and others. These candidate functions are organized into a library, represented by the matrix

$$\boldsymbol{\Theta}(\dot{\boldsymbol{P}}, \boldsymbol{P}, \boldsymbol{V}) = \begin{bmatrix} | & | & | & | & | & | \\ p(t) & \dot{p}(t) & p(t)\dot{p}(t) & p(t)^2 & \dot{p}(t)^2 & \ldots & v(t) \\ | & | & | & | & | & | \end{bmatrix} \in \mathbb{R}^{n+1 \times k}, \tag{6}$$

with each column containing one of the k basis functions evaluated at the consecutive instances of time $t_{i=0,1,2,\ldots,n}$. To compose $\boldsymbol{\Theta}$, we approximate the derivative of the pressure numerically $\dot{p}(t)$ as

$$\dot{p}(t_i) = \begin{cases} \frac{p(t_{i+1}) - p(t_{i-1})}{t_{i+1} - t_{i-1}} + O(\Delta t^3), \text{ if } 0 < i < n \\ \frac{p(t_1) - p(t_0)}{t_1 - t_0} + O(\Delta t^2), \text{ for } i = 0, \\ \frac{p(t_n) - p(t_{n-1})}{t_n - t_{n-1}} + O(\Delta t^2), \text{ for } i = n, \end{cases} \tag{7}$$

and similarly to calculate $\ddot{\boldsymbol{P}}$ given this approximation of $\dot{\boldsymbol{P}}$. The central finite differences in the above equation are accurate up to the second order in $\Delta t = t_{i+1} - t_i$ for all instants $t_{0<i<n}$, except the boundaries t_0 and t_n, where it is not applicable, and we have to resort to the first-order forward and backward finite differences, respectively. The coefficients $\boldsymbol{\Xi}$ are determined by STLS, which starts with an ordinary least squares solution

$$\hat{\boldsymbol{\Xi}} = \arg\min_{\boldsymbol{\Xi}} \left(\|\ddot{\boldsymbol{P}} - \boldsymbol{\Theta}(\boldsymbol{P}, \boldsymbol{V})\boldsymbol{\Xi}\|_2 \right), \tag{8}$$

and eliminates the coefficients $\boldsymbol{\Xi}$ that are smaller than the predefined sparsity threshold η. Then, another least squares solution is obtained for $\boldsymbol{\Xi}$ on the remaining coefficients, which are compared to the threshold again. This procedure is iterated until no more coefficients can be zeroed out.

By increasing the sparsity threshold η, we can select progressively simpler models with fewer significant non-zero terms. In our case, larger thresholds lead to models that consistently highlight the same key terms across different patients, as their data share similar patterns. This ensures that the model captures the most universally relevant dynamics without over-fitting [54]. We have tested values of $\eta \in \{0.1, 1.0, 5.0\}$.

The accuracy of identified models can be characterized by the normalized root mean squared error (NRMSE):

$$\mathrm{NRMSE}(p,\hat{p}) = \frac{1}{\max p(t) - \min p(t)} \sqrt{\frac{1}{n+1}\sum_{i=0}^{n}[p(t_i) - \hat{p}(t_i)]^2}. \tag{9}$$

where $p(t)$ represents the ground-truth time series of pressure, while $\hat{p}(t)$ is the trajectory generated by the identified model, starting from the same initial conditions, $\hat{p}(0) = p(0)$ and $\dot{\hat{p}}(0) = \dot{p}(0)$, and using the best-fit parameter values, $\hat{\boldsymbol{\Xi}}$. NRMSE provides a way to assess how closely the identified model replicates the observed data over time.

To benchmark the computational cost of fitting, we measured the time taken for seven runs of the optimization routine, each consisting of 10^3 iterations, executed on a single core of an Intel Core i7-13700 processor (manufactured by Intel Corporation, Santa Clara, CA, USA).

2.3. Logistic Regression for Classification of Blood-Flow Dynamics

Using several hundred elements of two-dimensional time series of the velocity and pressure $(p(t_i), v(t_i))$, we identify between three and ten coefficients $\hat{\boldsymbol{\Xi}}$ depending on the assumed model, effectively reducing the number of *descriptors* representing the same data.

However, features of blood flow differ between the experimental conditions, i.e., before, during, or after the surgery (Figure 1). Since the pressure–velocity relationship is encoded in the parameters $\hat{\boldsymbol{\Xi}}$, these learned parameters from the patients' data can be used as inputs in a classification problem. Here, we consider three categories of interest for each patient

$$Y(\hat{\boldsymbol{\Xi}}^{(i)}) = \begin{cases} 0 & \text{for AA flow,} \\ 1 & \text{for AVM flow,} \\ 2 & \text{for flow after a successful treatment,} \end{cases} \tag{10}$$

where $Y(\hat{\boldsymbol{\Xi}}^{(i)})$ refers to the actual class of the patient i and the last category is interpreted as a pathology-free condition.

In the context of a multi-class logistic regression [27,54], which we adopt in this report, the probability for the parameter values $\hat{\boldsymbol{\Xi}}^{(i)}$ of the ith patient to belong to the category $m \in \{0,1,2\}$ can be encoded by the softmax function (Section 7.4, Equation (79) [54]) as

$$P_m(\hat{\boldsymbol{\Xi}}^{(i)}) = \frac{\exp\left(\sum_j w_{mj}\hat{\Xi}_j^{(i)}\right)}{\sum_{m=0}^{2}\exp\left(\sum_j w_{mj}\hat{\Xi}_j^{(i)}\right)}, \tag{11}$$

where the summation over j includes all input coefficients $\hat{\Xi}_j$, namely a, b and ϵ. The input arguments $\boldsymbol{Xi}_j$ are extended by a bias term $\hat{\Xi}_0 = 1$, and w_{mj} are the weights that the model optimizes to assign categories. To determine these weights, maximizing the multinomial log-likelihood function (Section 7.4, cf. Equations (80) and (81) [54]),

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} \sum_i \sum_{m=0}^{2} \left\{ y_m^{(i)} \ln P_m(\hat{\boldsymbol{\Xi}}^{(i)}) + (1 - y_m^{(i)}) \ln\left[1 - P_m(\hat{\boldsymbol{\Xi}}^{(i)})\right] \right\}, \tag{12}$$

we use the L-BFGS algorithm [55], which ensures efficient and reliable convergence to the optimal values w_{mj}. In Equation (12), we set $y_m^{(i)} = 1$ if the ith time series has the label m, $Y(\hat{\boldsymbol{\Xi}}^{(i)}) = m$, or $y_m^{(i)} = 0$ otherwise.

2.4. Software

System identification (Section 2.2) and pathology classification (Section 2.3) were implemented in Python 3.12.11 [29] and based on the packages PySINDy [51,52] and scikit-learn [56], respectively. For visualization, we used the Python package MLxtend [57] and Wolfram Mathematica [58].

3. Results

3.1. Efficient Identification of Blood-Flow Dynamics

To identify the most efficient model from the same family as Equation (1), we progressively increased the sparsity threshold η. This process starts with a more general candidate library $\boldsymbol{\Theta}$ (Equation (2)), including polynomials of the pressure and its time derivatives up to the third degree, along with a linear term proportional to the velocity of the blood flow. We varied the sparsity threshold η over a range of values, from a small threshold $\eta = 0.1$ and an intermediate value $\eta = 1.0$ to the maximum $\eta = 5.0$.

The threshold value $\eta = 0.1$ narrows down the candidate library to the form of Equation (1) for all patients' data (Figure S1A). The majority of the nonlinear terms disappear at $\eta = 1.0$ (Figure S1B). Finally, the linear Equation (3) emerges robustly across all patients' data at the threshold $\eta = 5.0$ (Figure S1C).

To validate the identified models, the system was simulated using the inferred parameters. The time series thus generated were then compared against the experiments (Figure 2). While the sparsest solution demonstrates the best performance, reinforcing its utility for generalization, on certain patients' data, the higher-order models may appear indistinguishable from it (Figure 2A), or even strikingly inefficient (Figure 2B). Higher-order models show average NRMSE similar to the simplest linear Equation (3) for AA patients (Figure 2C). However, on AVM patients (Figure 2D), the linear model performs strikingly better than the alternatives. For both AA and AVM data, the linear model prevails by the computational cost.

The inefficiency of higher-order equations may seem counterintuitive, as one might expect more complex models to perform better at intermediate thresholds. However, it can be explained by the fitting process, which aims to reproduce the pressure derivatives evaluated numerically rather than the original time series. The higher-order equations robustly fit the derivative data tighter than the linear model (Figure S2 with the same patients as in Figure 2).

As the time derivatives of pressure in our method are calculated numerically from time series, artifacts due to noise, filtering, or smoothing in the data can occur. While advanced techniques like Kalman filtering or other smoothing methods [52] could potentially reduce such artifacts and improve consistency, finding the optimal method for numerical differentiation is beyond the scope of this study. Moreover, such approaches do not always guarantee a reliable performance, particularly given the variability in clinical data. Instead, selecting a large sparsity threshold η effectively isolates the most significant terms and ensures robustness even in the presence of artifacts related to the numerical differentiation. These features make the combination of a large threshold η with the finite-difference technique for computing time derivatives sufficient to achieve a balance between simplicity, consistency, and accuracy.

When evaluating the statistics and performance across all our clinical data, the *sparsest* model (3), which is computationally most efficient, achieves accuracy close to the higher-order models used in the earlier works. Overall, the linear model performs strikingly better than its more complex alternatives on the AVM data as quantified by the average NRMSE (Figure 2D, orange bars).

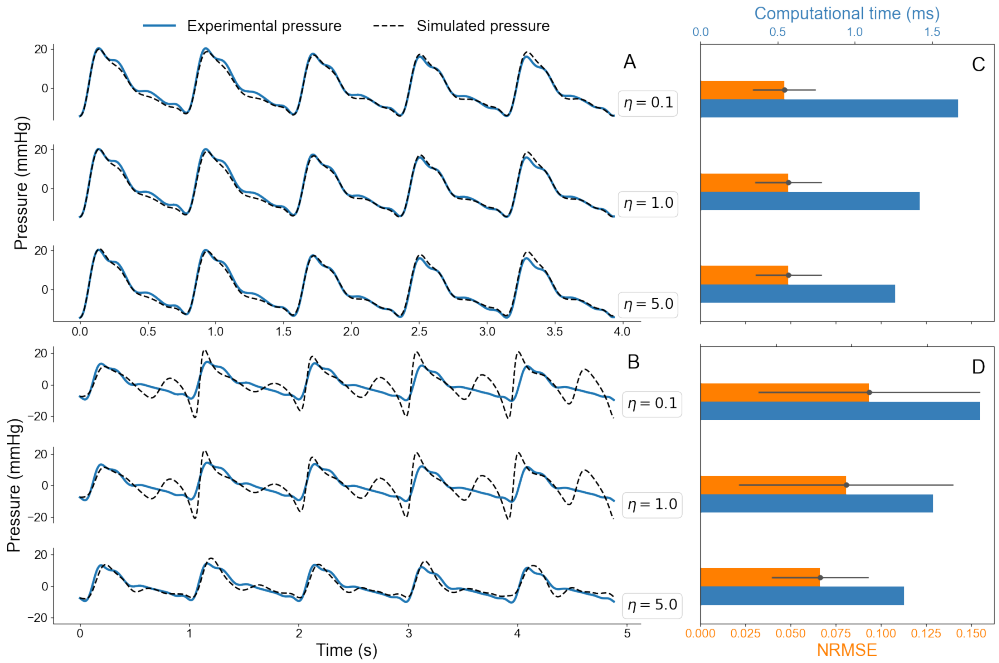

Figure 2. Examples of fitting with various sparsity thresholds η. For illustration, we select one AA (**A**) and one AVM (**B**) patient, whose models display the worst match between the simulations and experiments. NRMSEs averaged over all AA (**C**) and AVM (**D**) patients summarize the overall performance of the learned model (orange bars, standard deviation shown by whiskers). Computational time in (**C**,**D**) is reported for the patients in (**A**,**B**) (blue bars, average over several runs).

3.2. Reproducibility and Forecasting Performance of the Linear Model

The linear model (3) inferred by SINDy accurately captures patients' blood-flow pressure from the velocity time series with a set of just three coefficients (a, b, ϵ). The simulated trajectories are in good agreement with the experimental time series (Figure 3), and display a counterclockwise current—the circulation commonly found in arteries [15,16]. Assessing the reproducibility of the fitted parameters for the same patient is crucial for validating the analysis and demonstrating the robustness of the methodology. To evaluate reproducibility, we divided the time series into two equal parts and independently applied our fitting procedure to each half of the data (Figure 4A). Values of the parameters thus learned differed at the most by 24% relative to the estimate inferred from the whole series across all patients' data considered, underscoring the robustness of the model across all patient data considered (Figure 4B).

To test the effect of the training-set length on the forecasting performance of our method, we inferred the model parameter values from time series of varied length. By *forecasting*, we mean the reconstruction of pressure dynamics conditioned on the measured velocity signal. Because the time series for each patient consists of a total of five cardiac cycles (Section 2.1), we partition each signal into five equal temporal segments, effectively corresponding to the individual cycles. We reserve the fifth segment for validation, and the first three for fitting. The fourth cycle is not used in either training or testing to allow time correlations to decay and prevent temporal leakage between training and validation data. We fitted a total of 1, 2, or 3 consecutive cardiac cycles from the training set. The forecasting performance is then characterized by NRMSE (9) of the simulations reproducing the validation cycle.

Training the model on a dataset covering a single cardiac cycle enables excellent predictions of the experimental time series for the same duration (Figure 5). Extending the training set to encompass up to three cardiac cycles does not significantly affect the NRMSE, though it expectedly increases the computational cost. This vindicates the SINDy method and its robustness in the present medical application due to the low variance of

the experimental time series over different cycles and, thus, to low overfitting, even when training on such short time series.

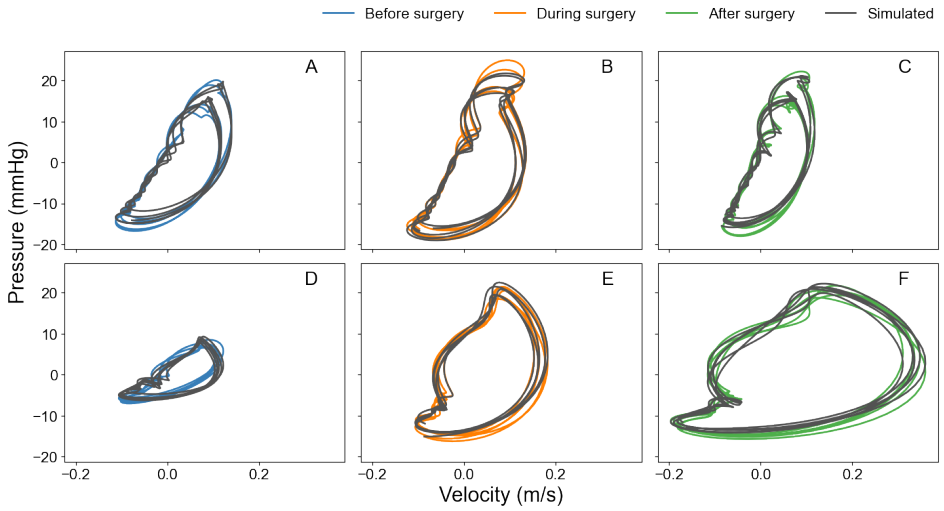

Figure 3. Examples of pressure–velocity diagrams before (blue), during (orange) and after (green) surgery for time series of a patient with an arterial aneurysm (**A**–**C**) and with an arteriovenous malformation (**D**–**F**). We simulate Equation (3) for each patient by using the initial condition of the experimental pressure.

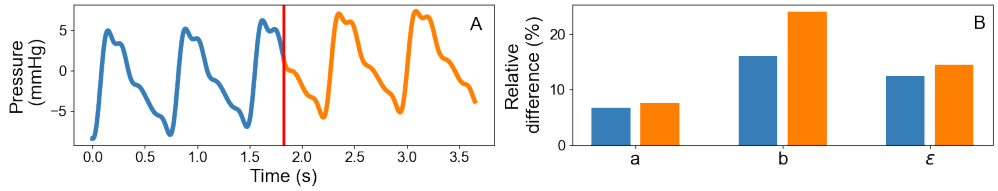

Figure 4. Reproducibility of AA patients' parameter values. (**A**) Example of splitting in half the pressure trajectory taken from an AA patient before surgery. First, we infer parameters from each part (blue and orange trajectories) independently, and then from the whole time series. (**B**) The maximum relative difference of model parameters learned from the two halves shown in (**A**) (blue and orange) with respect to the values inferred from the entire series, emphasizing the reproducibility of our results.

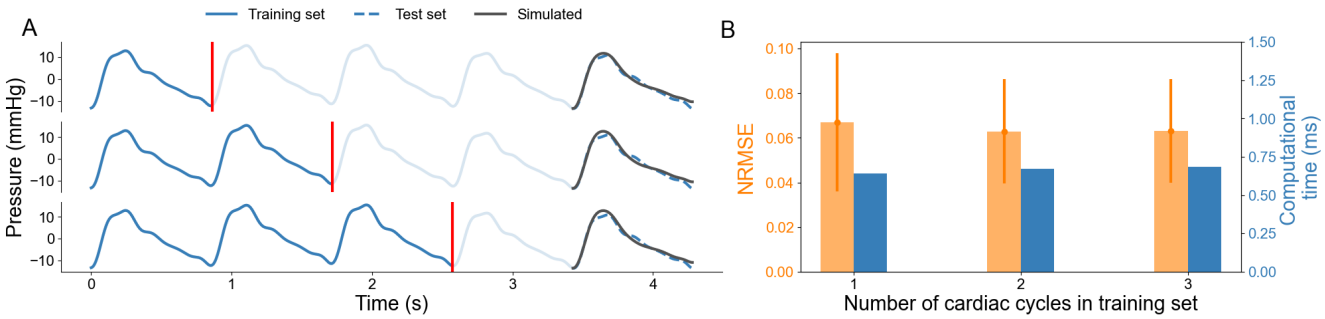

Figure 5. Forecasting performance of Equation (3). We split the pressure time series into training and test sets. Simulations of the model with parameter values learned from the training sets of various lengths are compared with the time series of the test set. (**A**) An example of splitting an AA patient's pressure time series. The training sets (solid blue line) are delimited by the red vertical lines. The last cardiac cycle (dashed blue line), kept as the test set, is compared with the simulated model (solid black line). Averaged NRMSE of the simulated pressure on a test set of a single cardiac cycle across all AA patients indicates negligible changes in the accuracy with the increasing size of the training set, which expectedly offsets the computational cost (**B**). The orange whiskers in (**B**) indicate the standard deviation over all AA patients.

3.3. Machine Learning Classification of Arteriovenous Pathologies

Since the patients' data used in this work contain successfully treated blood-flow defects, the post-surgery time series of pressure and velocity may serve as proxies of healthy vessels. Consequently, we investigate whether the parameters of Equation (3) learned from observations (Tables S1 and S2) can be used to discriminate between the normal and abnormal flow dynamics, and to identify the pathological defect in the latter case.

Using multi-class logistic regression [27,54], we train the softmax classifier Equation (11) with the feature set $\boldsymbol{\Xi} = (\Xi_0, \Xi_1, \Xi_2, \Xi_3) = (1, a, b, \epsilon)$, where Ξ_0 and its associated weight w_{m0} correspond to the constant bias term (see Materials and Methods for details). The classifier is trained to categorize the output into three categories, $m \in \{0, 1, 2\}$: arterial aneurysm, arteriovenous malformations, and treated vessels, respectively, cf. Equation (10).

To exploit a small set of twenty examples—comprising five arterial aneurysms and five arteriovenous malformations, both pre- and post-surgery—we first resample the data into 100 partitions, each with training and test subsets of 16 (80%) and 4 (20%) examples, respectively. Note that in total, there are over 4800 such unique partitions. Then, to obtain a more robust estimate of classifier performance, we apply grouped cross-validation that prevents patient-level data leakage but reduces the number of partitions.

The average accuracy of the logistic regression, evaluated over the 100 random partitions, is $73 \pm 2\,\%$. To investigate further properties of the classification, we constructed decision boundaries by optimizing Equation (11) over the full set of 20 examples in the 3-dimensional feature space (Figure 6), as well as their 2-dimensional projections (Figure 7).

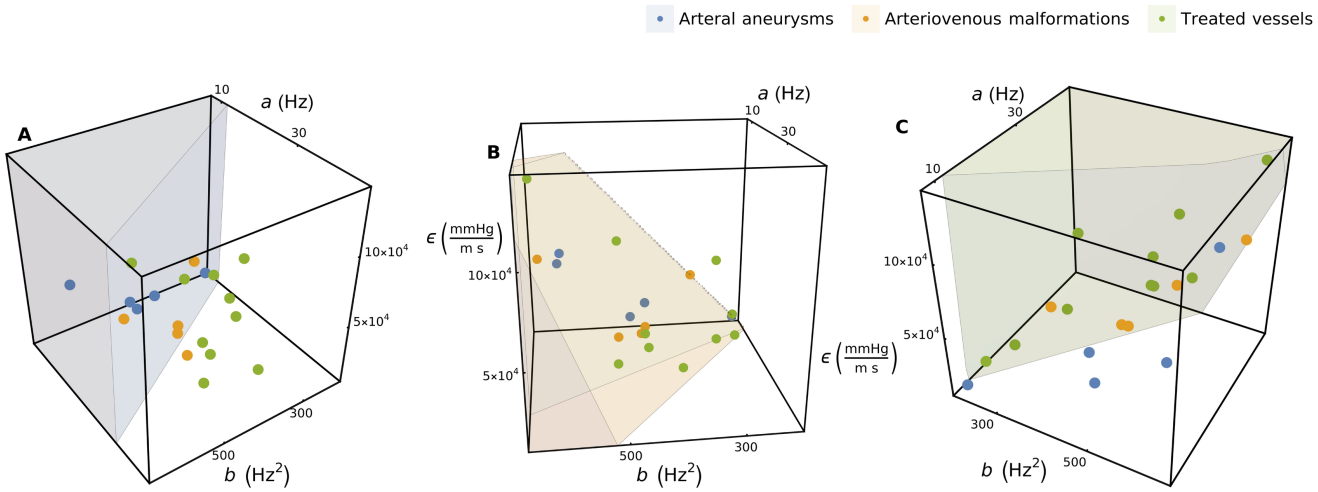

Figure 6. Decision boundaries of the logistic classifier Equation (12) in the 3-dimensional feature space of parameters a, b, and ϵ from Equation (3). These boundaries separate the regions of 5 AA patients (blue circles, blue area in (**A**)) and 5 AVM patients (orange circles, orange area in (**B**)) before the surgery from the same patients' data after the surgery (green circles, green area in (**C**)). The decision boundaries between these three classes are constructed by using Wolfram Mathematica [58].

Most of the patients' data, with a single exception, fall into the regime of an underdamped harmonic oscillator with $a^2 < b$ (Figure S3). We notice that AA occupies the region of small frequencies $\sqrt{b}$ and low dissipation a (Figure 6A). In contrast, examples of the treated vessels reside in the high-dissipation and moderate-frequencies area (Figure 6C). Arteriovenous malformations belong in a layer of relatively smaller volume in between these two categories. In addition, boundaries of the arterial aneurysms' region appear parallel to the direction of the axis ϵ and, therefore, independent of the coupling strength between the pressure and velocity.

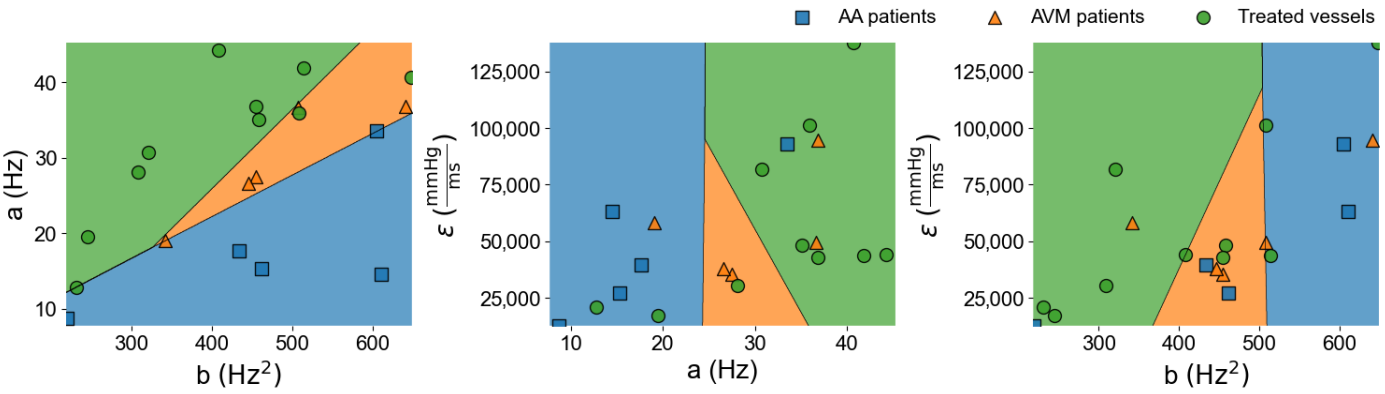

Figure 7. Decision boundaries of pathology classification by logistic regression. Decision boundaries of the logistic classifier Equation (12) are projected onto the ab-, ϵa-, and ϵb-planes cutting the third axis of the full 3-dimensional feature space at the mean values of the parameters ϵ, b, and a over all patients' data, respectively. The full dataset contains model parameters based on 5 AA patients and 5 AVM patients before surgery, and the same patients' data after the surgery. The decision boundaries between these three classes are constructed by using the MLxtend [57] Python module.

To guard against the patient-level leakage, we apply a grouped cross-validation scheme. Specifically, to keep the original 80/20 train–test ratio, we trained the classifier with the data of 8 out of 10 patients, thus leaving 2 patients per fold for testing. Over all possible combinations of patients' data, which amount to 45 folds in total, the classifier achieved an accuracy of 0.75 ± 0.22, a balanced accuracy of 0.74 ± 0.22, and a macro-F1 score of 0.69 ± 0.26. Its performance was the highest for treated vessels, with the lowest and more variable precision and recall for the AVM data (Table 1).

Table 1. Performance of pathology classifier in the grouped leave-two-out cross-validation.

	AA	AVM	Treated Vessel
Recall	0.62 ± 0.46	0.49 ± 0.47	0.79 ± 0.29
Precision	0.62 ± 0.46	0.42 ± 0.43	0.88 ± 0.24

The average confusion matrix (Table 2) exhibits strong diagonal dominance, with misclassifications mainly involving the AVM class. AA pathologies are almost never confused with treated vessels. Perhaps the most critical property for clinical applications, verified in our analysis, is that treated vessels are classified more reliably.

Table 2. Row-normalized confusion matrix of the pathology classifier averaged over 45 folds of the data for the grouped leave-two-out cross-validation.

	AA	AVM	Treated Vessels
AA	0.80	0.20	0.00
AVM	0.18	0.62	0.20
Treated vessels	0.01	0.20	0.79

3.4. Comparative Performance of Linear-Model Fitting with SINDy

To compare the computational complexity and cost of our method with the original approach of Refs. [15,16], we use the fitting results over the full duration of each time series in our AA and AVM datasets. Although the original inverse-problem technique can be parallelized to significantly improve its performance, we refer to its serial implementation, since we do not explore concurrent-execution strategies for SINDy.

On average, the serial implementation of the original inverse-problem technique takes approximately 10 s to reconstruct Equation (1) for a single patient, compared to only the ms-scale performance of our method based on Refs. [51,52] and the linear Equation (3). As expected, the original method, which incorporates higher-order nonlinear terms with more fitting parameters, achieves slightly better accuracy than the linear model proposed in our approach (Figure 8). This difference appears negligible for the intended applications, e.g., in the classification of pathologies. Moreover, when revisiting the original polynomial formulation in Equation (1), we observed that the constant terms a_0 and b_0 dominated in the trend of $A(p)$ and $B(p)p$ averaged over patients, while higher-order terms were negligible (Figure S4), supporting the suitability of the linear model in Equation (3) to capture the core dynamics of blood flow. Therefore, the significant gain in computational efficiency offered by our method outweighs the marginal loss in expressive power, especially for online short-latency applications.

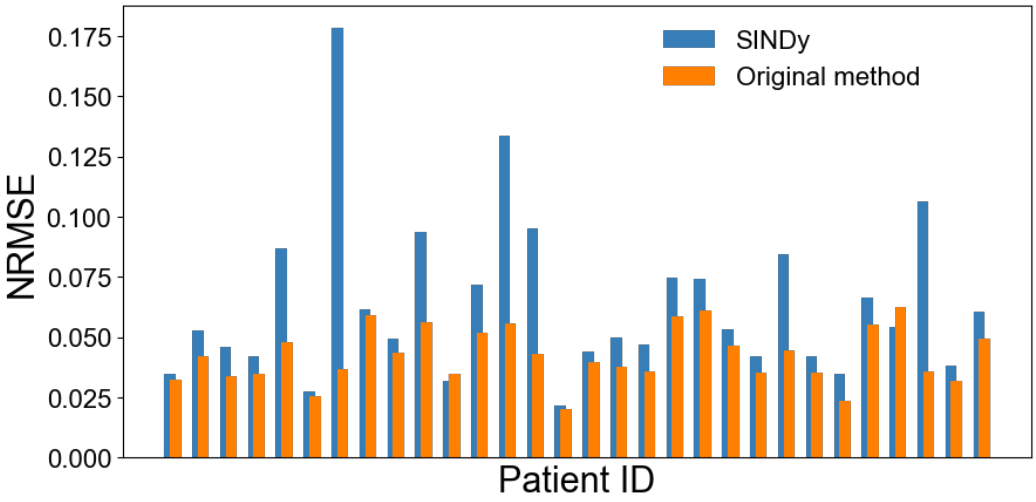

Figure 8. Comparison of NRMSE between the Equations (3) fitted with SINDy and (1) fitted with the original inverse-problem technique from [15,16] (2016) over five cardiac cycles per each AA and AVM patient data.

4. Discussion

Our work illustrates the potential of modern machine learning techniques to enhance the analysis of hemodynamical data routinely monitored during cardiovascular surgeries in real time. By overcoming the computational challenges faced by the original method [15,16], the combination of SINDy renders our procedure faster and more robust (Section 3.4, Figure 8), while describing the system by a simpler linear model (5). These enhancements make our approach well-suited for online short-latency applications. One potential application, the identification of pathologies, we have proposed and tested here.

Through SINDy, the computational cost of handling general models becomes manageable, where the sparsity threshold offers a way to find a balance between accuracy and complexity. Unlike direct methods of linear-model fitting, the model-selection capabilities of SINDy allow additional control over observations that are poorly described by Equation (5). The occurrences of such time series in medical applications can be used to trigger exceptional procedures or fallback to the higher-order models.

Although our study relies on direct measurements, which have the advantage of being more precise, noninvasive techniques of measuring blood-flow variables also exist [59–63]. Therefore our method, which succinctly encodes the observations of hemodynamic data with just a few interpretable parameters and exhibits excellent forecasting capabilities,

could be used in various clinical settings, e.g., for therapeutic assessment before the intervention stage.

Indeed, the hemodynamic descriptors obtained through our procedure prove to be sufficiently informative for the automated detection of pathological blood flows. By using the logistic regression over a small set of training data, we achieved a 74% balanced accuracy of discrimination between arterial aneurysms, arteriovenous malformations, and treated vessels. Moreover, this method, in combination with a linear model, renders it accessible to analysis and physically interpretable decision boundaries. Incorporating the model parameters into surgical planning can enhance medical decision-making and reduce the risk of unnecessary reoperation.

The hemodynamic descriptors produced by our method can also serve as effective means for dimensionality reduction. When combined with complementary techniques—such as physics-informed neural networks (PINNs), which leverage sparse imaging data and physical laws to predict brain hemodynamics [9]—they can form the foundation for a complex analytical framework.

While our findings are based on a limited sample size, the use of cross-validation partially mitigates overfitting risks. Importantly, our grouped cross-validation analysis shows that the successfully treated vessels are identified most reliably. Future work should prioritize the acquisition of larger patient datasets to enable more robust assessment of the classification performance. Moreover, beyond diagnostic applications, illustrated here by the classification of pathologies, one may pursue a more ambitious goal of predicting hemodynamic parameters after surgical intervention. Such predictions would facilitate the development of a therapeutic prognosis, helping in medical decisions and improving the statistics of treatment outcomes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ai7030105/s1, Figure S1: Sparse structure of Equation (2); Figure S2: Accuracy of numerical differentiation; Figure S3: Parameters of the autonomous part (left-hand side) of Equation (3); Figure S4: Coefficients of the Lienard Equation (1); Table S1: Parameter values of Equation (3) inferred from the AA patients' data; Table S2: Parameter values of Equation (3) inferred from the AVM patients' data.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, I.T.; validation, all authors; formal analysis, all authors; investigation, I.T., A.C., Y.B., M.S., and D.E.; data curation, A.C. and Y.B.; writing—original draft preparation, I.T. and R.B.; writing—review and editing, all authors; visualization, I.T.; supervision, J.B., D.E., É.R., and R.B.; and project administration, É.R. and R.B.; All authors have read and agreed to the published version of the manuscript.

Funding: A.C. and Y.B. acknowledge the support of the state assignment of LIH SB RAS (Theme no. FWGG-2021-0009-2.3.1.2.10). M.S. acknowledges the support of the state assignment of IM SB RAS (Theme No. FWNF-2024-0001). É.R. acknowledges financial support from PNRR MUR project PE0000023-NQSTI. D.E. was partially supported by TUBITAK (118C236), UKRI (EP/Z002656/1), and the BAGEP Award of the Science Academy. R.B. acknowledges funding from the EMBL. The APC was funded by EMBL.

Institutional Review Board Statement: The clinical data used in this report were acquired during neurosurgical operations at the Meshalkin Research Institute of Circulation Pathology (Novosibirsk, Russia) between March 2012 and September 2017. The investigation was conducted in accordance with the Helsinki Declaration and approved by the Inspection Commission of the E.N. Meshalkin National Medical Research Center (Session No. 549).

Informed Consent Statement: All patients involved in the study gave their informed consent. The authors do not have access to information that could identify individual participants during or after data collection.

Data Availability Statement: The original data presented in the study are openly available at https://github.com/lffrrnt/HemodynamicsDiscovery/tree/main/data (accessed on 2 March 2026).

Acknowledgments: I.T., J.B., É.R., and R.B. thank the London Mathematical Laboratory school for establishing the initial framework of this collaboration.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AA	Arterial aneurysms
AVM	Arteriovenous malformations
LASSO	Least absolute shrinkage and selection operator
SINDy	Sparse identification of nonlinear dynamical systems
STLS	Sequentially thresholded least squares

References

1. Brisman, J.L.; Song, J.K.; Newell, D.W. Cerebral Aneurysms. *N. Engl. J. Med.* **2006**, *355*, 928–939. [CrossRef] [PubMed]
2. Al-Shahi, R. A systematic review of the frequency and prognosis of arteriovenous malformations of the brain in adults. *Brain* **2001**, *124*, 1900–1926. [CrossRef] [PubMed]
3. Mast, H.; Mohr, J.P.; Osipov, A.; Pile-Spellman, J.; Marshall, R.S.; Lazar, R.M.; Stein, B.M.; Young, W.L. 'Steal' Is an Unestablished Mechanism for the Clinical Presentation of Cerebral Arteriovenous Malformations. *Stroke* **1995**, *26*, 1215–1220. [CrossRef]
4. Solomon, R.A.; Connolly, E.S. Arteriovenous Malformations of the Brain. *N. Engl. J. Med.* **2017**, *376*, 1859–1866. [CrossRef] [PubMed]
5. Martin, N.A.; Bentson, J.; Viñuela, F.; Hieshima, G.; Reicher, M.; Black, K.; Dion, J.; Becker, D. Intraoperative digital subtraction angiography and the surgical treatment of intracranial aneurysms and vascular malformations. *J. Neurosurg.* **1990**, *73*, 526–533. [CrossRef]
6. Thompson, R.C.; Steinberg, G.K.; Levy, R.P.; Marks, M.P. The Management of Patients with Arteriovenous Malformations and Associated Intracranial Aneurysms. *Neurosurgery* **1998**, *43*, 202–211. [CrossRef]
7. Platz, J.; Berkefeld, J.; Singer, O.C.; Wolff, R.; Seifert, V.; Konczalla, J.; Güresir, E. Frequency, risk of hemorrhage and treatment considerations for cerebral arteriovenous malformations with associated aneurysms. *Acta Neurochir.* **2014**, *156*, 2025–2034. [CrossRef]
8. Arsalidou, M.; Skuratov, N.; Khalezov, E.; Bernstein, A.; Burnaev, E.; Sharaev, M. Effects of age, gender, and hemisphere on cerebrovascular hemodynamics in children and young adults: Developmental scores and machine learning classifiers. *PLoS ONE* **2022**, *17*, e0263106. [CrossRef]
9. Sarabian, M.; Babaee, H.; Laksari, K. Physics-Informed Neural Networks for Brain Hemodynamic Predictions Using Medical Imaging. *IEEE Trans. Med. Imaging* **2022**, *41*, 2285–2303. [CrossRef]
10. Pedrett, R.; Mascagni, P.; Beldi, G.; Padoy, N.; Lavanchy, J.L. Technical skill assessment in minimally invasive surgery using artificial intelligence: A systematic review. *Surg. Endosc.* **2023**, *37*, 7412–7424. [CrossRef]
11. Marmarelis, V.Z.; Shin, D.C.; Kang, Y.; Novak, V. Data-based modeling of cerebral hemodynamics quantifies impairment of cerebral blood flow regulation in type-2 diabetes. *J. Cereb. Blood Flow Metab.* **2024**, *44*, 1288–1301. [CrossRef]
12. Vakitbilir, N.; Froese, L.; Gomez, A.; Sainbhi, A.S.; Stein, K.Y.; Islam, A.; Bergmann, T.J.G.; Marquez, I.; Amenta, F.; Ibrahim, Y.; et al. Time-Series Modeling and Forecasting of Cerebral Pressure–Flow Physiology: A Scoping Systematic Review of the Human and Animal Literature. *Sensors* **2024**, *24*, 1453. [CrossRef]
13. Brunton, S.L.; Proctor, J.L.; Kutz, J.N.; Bialek, W. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [CrossRef] [PubMed]
14. Orlov, K.; Panarin, V.; Berestov, V.; Krivoshapkin, A.; Kislitsin, D.; Chupakhin, A.; Baranov, V.; Cherevko, A.; Khe, A.; Telegina, N. Method for Intraoperative Doppler Control of Radicality of Arteriovenous Malformation Embolisation. RU-2511235C2, 2014. Available online: https://patenton.ru/patent/RU2511235C2/en (accessed on 2 March 2026).
15. Cherevko, A.A.; Mikhaylova, A.V.; Chupakhin, A.P.; Ufimtseva, I.V.; Krivoshapkin, A.L.; Orlov, K.Y. Relaxation oscillation model of hemodynamic parameters in the cerebral vessels. *J. Phys. Conf. Ser.* **2016**, *722*, 012045. [CrossRef]
16. Parshin, D.V.; Ufimtseva, I.V.; Cherevko, A.A.; Khe, A.K.; Orlov, K.Y.; Krivoshapkin, A.L.; Chupakhin, A.P. Differential properties of Van der Pol — Duffing mathematical model of cerebrovascular haemodynamics based on clinical measurements. *J. Phys. Conf. Ser.* **2016**, *722*, 012030. [CrossRef]

17. Ye, D.; Krzhizhanovskaya, V.; Hoekstra, A.G. Data-driven reduced-order modelling for blood flow simulations with geometry-informed snapshots. *J. Comput. Phys.* **2024**, *497*, 112639. [CrossRef]
18. Kaczmarek, W.; Magdziarz Ibrahim-El-Nur, J.; Łoś, M.; Roleder, T. Reconstructing cerebral hemodynamics from sparse data using Neural Operator Transformers. *Comput. Biol. Med.* **2025**, *195*, 110492. [CrossRef]
19. Csala, H.; Mohan, A.; Livescu, D.; Arzani, A. Physics-constrained coupled neural differential equations for one dimensional blood flow modeling. *Comput. Biol. Med.* **2025**, *186*, 109644. [CrossRef]
20. Isaev, A.; Dobroserdova, T.; Danilov, A.; Simakov, S. Physically Informed Deep Learning Technique for Estimating Blood Flow Parameters in Four-Vessel Junction after the Fontan Procedure. *Computation* **2024**, *12*, 41. [CrossRef]
21. Blanco, P.J.; Müller, L.O. One-Dimensional Blood Flow Modeling in the Cardiovascular System. From the Conventional Physiological Setting to Real-Life Hemodynamics. *Int. J. Numer. Methods Biomed. Eng.* **2025**, *41*, e70020. [CrossRef]
22. Strogatz, S.H.; Liénard systems. In *Nonlinear Dynamics and Chaos*, 2nd ed.; CRC Press: London, UK, 2019; Chapter 7.4.
23. Kabanikhin, S.I. *Inverse and Ill-Posed Problems*; Inverse and Ill-Posed Problems Series; De Gruyter: Berlin, Germany, 2011.
24. Bugai, Y.V.; Cherevko, A.A.; Shishlenin, M.A. Classification of cerebrovascular pathologies in real-time using nonlinear ODE-based surrogate model. *J. Inverse Ill-Posed Probl.* **2025**, *34*, 71–85. [CrossRef]
25. Hanke, M.; Neubauer, A.; Scherzer, O. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* **1995**, *72*, 21–37. [CrossRef]
26. Kabanikhin, S.; Shishlenin, M. Quasi-solution in inverse coefficient problems. *J. Inverse Ill-Posed Probl.* **2008**, *16*, 705–713. [CrossRef]
27. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer Publishing Company, Incorporated: New York, NY, USA, 2014.
28. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019. [CrossRef]
29. Topal, I. HemodynamicsDiscovery. Source Code Supporting Our Findings. 2025. Available online: https://github.com/lffrrnt/HemodynamicsDiscovery (accessed on 2 March 2026).
30. Kaiser, E.; Kutz, J.N.; Brunton, S.L. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *474*, 20180335. [CrossRef]
31. Thaler, S.; Paehler, L.; Adams, N.A. Sparse identification of truncation errors. *J. Comput. Phys.* **2019**, *397*, 108851. [CrossRef]
32. Zhang, L.; Schaeffer, H. On the Convergence of the SINDy Algorithm. *Multiscale Model. Simul.* **2019**, *17*, 948–972. [CrossRef]
33. Lagergren, J.H.; Nardini, J.T.; Michael Lavigne, G.; Rutter, E.M.; Flores, K.B. Learning partial differential equations for biological transport models from noisy spatio-temporal data. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *476*, 20190800. [CrossRef]
34. Nardini, J.T.; Lagergren, J.H.; Hawkins-Daarud, A.; Curtin, L.; Morris, B.; Rutter, E.M.; Swanson, K.R.; Flores, K.B. Learning Equations from Biological Data with Limited Time Samples. *Bull. Math. Biol.* **2020**, *82*, 119. [CrossRef]
35. Reinbold, P.A.K.; Gurevich, D.R.; Grigoriev, R.O. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Phys. Rev. E* **2020**, *101*, 010203. [CrossRef]
36. Reinbold, P.A.K.; Kageorge, L.M.; Schatz, M.F.; Grigoriev, R.O. Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression. *Nat. Commun.* **2021**, *12*, 3219. [CrossRef]
37. Hirsh, S.M.; Barajas-Solano, D.A.; Kutz, J.N. Sparsifying priors for Bayesian uncertainty quantification in model discovery. *R. Soc. Open Sci.* **2022**, *9*, 211823. [CrossRef] [PubMed]
38. Fasel, U.; Kutz, J.N.; Brunton, B.W.; Brunton, S.L. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2022**, *478*, 20210904. [CrossRef]
39. Goyal, P.; Benner, P. Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2022**, *478*, 20210883. [CrossRef]
40. Lejarza, F.; Baldea, M. Data-driven discovery of the governing equations of dynamical systems via moving horizon optimization. *Sci. Rep.* **2022**, *12*, 11836. [CrossRef]
41. Delahunt, C.B.; Kutz, J.N. A Toolkit for Data-Driven Discovery of Governing Equations in High-Noise Regimes. *IEEE Access* **2022**, *10*, 31210–31234. [CrossRef]
42. Cortiella, A.; Park, K.C.; Doostan, A. A Priori Denoising Strategies for Sparse Identification of Nonlinear Dynamical Systems: A Comparative Study. *J. Comput. Inf. Sci. Eng.* **2022**, *23*, 011004. [CrossRef]
43. Anjos, L.d.; Naozuka, G.T.; Volpatto, D.T.; Godoy, W.A.C.; Costa, M.I.d.S.; Almeida, R.C. A new modelling framework for predator-prey interactions: A case study of an aphid-ladybeetle system. *Ecol. Inform.* **2023**, *77*, 102168. [CrossRef]
44. Brummer, A.B.; Xella, A.; Woodall, R.; Adhikarla, V.; Cho, H.; Gutova, M.; Brown, C.E.; Rockne, R.C. Data driven model discovery and interpretation for CAR T-cell killing using sparse identification and latent variables. *Front. Immunol.* **2023**, *14*, 1115536. [CrossRef]

45. Gutierrez-Vilchis, A.; Perfecto-Avalos, Y.; Garcia-Gonzalez, A. Modeling bacteria pairwise interactions in human microbiota by Sparse Identification of Nonlinear Dynamics (SINDy)*. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; p. 1–4. [CrossRef]
46. Prokop, B.; Gelens, L.; Pelz, P.F.; Friesen, J. Challenges in identifying simple pattern-forming mechanisms in the development of settlements using demographic data. *Phys. Rev. E* **2023**, *107*, 064305. [CrossRef]
47. Prokop, B.; Gelens, L. From biological data to oscillator models using SINDy. *iScience* **2024**, *27*, 109316. [CrossRef]
48. Fung, L.; Fasel, U.; Juniper, M. Rapid Bayesian identification of sparse nonlinear dynamics from scarce and noisy data. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2025**, *481*, 20240200. [CrossRef]
49. Gerardos, A.; Ronceray, P. Principled model selection for stochastic dynamics. *Phys. Rev. Lett.* **2025**, *135*, 167401. [CrossRef] [PubMed]
50. Topal, I.; Eroglu, D. Reconstructing Network Dynamics of Coupled Discrete Chaotic Units from Data. *Phys. Rev. Lett.* **2023**, *130*, 117401. [CrossRef]
51. de Silva, B.M.; Champion, K.; Quade, M.; Loiseau, J.C.; Kutz, J.N.; Brunton, S.L. PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data. *J. Open Source Softw.* **2020**, *5*, 2104. [CrossRef]
52. Kaptanoglu, A.A.; de Silva, B.M.; Fasel, U.; Kaheman, K.; Goldschmidt, A.J.; Callaham, J.; Delahunt, C.B.; Nicolaou, Z.G.; Champion, K.; Loiseau, J.C.; et al. PySINDy: A comprehensive Python package for robust sparse system identification. *J. Open Source Softw.* **2022**, *7*, 3994. [CrossRef]
53. Kaheman, K.; Kutz, J.N.; Brunton, S.L. SINDy-PI: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *476*, 20200279. [CrossRef]
54. Mehta, P.; Bukov, M.; Wang, C.H.; Day, A.G.; Richardson, C.; Fisher, C.K.; Schwab, D.J. A high-bias, low-variance introduction to Machine Learning for physicists. *Phys. Rep.* **2019**, *810*, 1–124. [CrossRef]
55. Zhu, C.; Byrd, R.H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Trans. Math. Softw.* **1997**, *23*, 550–560. [CrossRef]
56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
57. Raschka, S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J. Open Source Softw.* **2018**, *3*, 638. [CrossRef]
58. Wolfram Research, Inc. *Mathematica*, Version 14.1; Wolfram Research, Inc.: Champaign, IL, USA, 2024.
59. van Zijl, P.C.; Eleff, S.M.; Ulatowski, J.A.; Oja, J.M.; Uluğ, A.M.; Traystman, R.J.; Kauppinen, R.A. Quantitative assessment of blood flow, blood volume and blood oxygenation effects in functional magnetic resonance imaging. *Nat. Med.* **1998**, *4*, 159–167. [CrossRef]
60. Lee, D.C.; Simonetti, O.P.; Harris, K.R.; Holly, T.A.; Judd, R.M.; Wu, E.; Klocke, F.J. Magnetic Resonance Versus Radionuclide Pharmacological Stress Perfusion Imaging for Flow-Limiting Stenoses of Varying Severity. *Circulation* **2004**, *110*, 58–65. [CrossRef] [PubMed]
61. Schuchardt, F.; Schroeder, L.; Anastasopoulos, C.; Markl, M.; Bäuerle, J.; Hennemuth, A.; Drexl, J.; Valdueza, J.M.; Mader, I.; Harloff, A. In vivo analysis of physiological 3D blood flow of cerebral veins. *Eur. Radiol.* **2015**, *25*, 2371–2380. [CrossRef]
62. Errico, C.; Pierre, J.; Pezet, S.; Desailly, Y.; Lenkei, Z.; Couture, O.; Tanter, M. Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature* **2015**, *527*, 499–502. [CrossRef] [PubMed]
63. Morgan, A.G.; Thrippleton, M.J.; Wardlaw, J.M.; Marshall, I. 4D flow MRI for non-invasive measurement of blood flow in the brain: A systematic review. *J. Cereb. Blood Flow Metab.* **2020**, *41*, 206–218. [CrossRef] [PubMed]