Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

Mechanistic Learning Based Surrogate-Aided Optimisation: Application to Hydrocephalus Shunt Systems.

Elizabeth Hayman
Department of Engineering Science, Institute of Mathematics, University of Oxford

José-Maria Peña
Lurtis Ltd, Oxford

Jayaratnam Jayamohan
John Radcliffe Hospital, Oxford

Sarah Waters*
Institute of Mathematics, University of Oxford
Email: sarah.waters@maths.ox.ac.uk

Antoine Jerusalem*
Department of Engineering Science, University of Oxford
Email: antoine.jerusalem@eng.ox.ac.uk

For many modern design problems, e.g., in clinical device design, evaluating designs prior to experimental testing typically requires high-fidelity multi-physics models that may involve fluid–structure interaction (FSI), multiple spatial and temporal scales, and dependencies on open systems. Such simulations can take hours or days making them impractical for incorporation into computational optimisation protocols. Surrogate models that provide rapid assessments are therefore essential.

Traditional data-driven surrogates are fast but offer limited interpretability, while reduced-order models capture physical mechanisms but depend on simplifying assumptions and prior knowledge. Mechanistic learning (MxL) addresses these limitations by combining machine learning with reduced-order modelling to create a surrogate that retains mechanistic insight while achieving high computational efficiency.

Here, we propose to integrate MxL based surrogates into a computational optimisation protocol. We showcase this methodology in the design of hydrocephalus shunt systems. Starting from a 3D finite element FSI model, we motivate the use of a fast surrogate model in which fluid stresses on the FSI boundary are combined with ML techniques to predict solid deformation. The resulting MxL surrogate is integrated into a hybrid optimisation algorithm, reducing evaluation time from 3 days to 10 minutes and enabling tractable optimisation of a high-dimensional design space. The resulting designs are ultimately tested in the full FSI model to confirm their improved efficacy. This methodology is applicable to engineering design problems characterised by computationally expensive multi-physics models, particularly those involving dominant FSI effects.

Key words. *Computational optimisation, fluid-structure interaction, mechanistic learning, medical device design, multi-physics design, surrogate-aided optimisation.*

1 INTRODUCTION

Design optimisation is central to many areas of engineering and increasingly supported by high-fidelity computational modelling for maximised time and cost efficiency [1, 2]. These models can naturally be coupled with automated search procedures, allowing candidate designs to be generated algorithmically under user-defined constraints [3, 4, 5, 6]. Such methods reduce reliance on iterative experimentation and enable systematic exploration of high-dimensional design spaces [7].

The potentially quixotic complexity of design opti-

*Corresponding authors

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

misation is best epitomised in the context medical device development, where computational models are used by designers to evaluate performance prior to particularly expensive and lengthy experimental or clinical testing, often further conditioned by ethics and trial considerations [8, 9]. In these applications, device performance often depends on complex multi-physics interactions between fluids and compliant evolving biological tissues. Fluid–structure interaction (FSI) models are therefore required to capture the governing mechanisms [10, 11]. These models typically use the finite element and finite volume methods to solve complex 3D geometries. While FSI models provide mechanistically rich evaluation of a device's performance, their computational cost makes them impractical for direct use in an optimisation requiring hundreds or thousands of candidate evaluations [12, 13, 14]. This thus motivates surrogate-aided optimisation strategies, in which a computationally inexpensive approximation replaces the high-fidelity model within the search procedure [15, 16].

In this paper, we propose to replace high-fidelity multi-physics simulations—particularly those involving FSI and multiscale interactions [10, 11, 17, 18, 19]—with a reduced-order model (ROM) preserving the key mechanistic drivers of deformation, in turn wrapped in a machine learning (ML) mapping trained on data generated from the full FSI model. This augmentation of mechanistic models with data-driven techniques is often referred to as mechanistic learning (MxL) [20, 21, 22, 23]. We note that this approach is distinct from physics-informed learning methods [24], as mechanistic knowledge is embedded structurally through the ROM rather than imposed within the loss function of the ML layer.

The clinical application chosen to highlight this approach is ventricular catheter design in hydrocephalus shunt systems. Hydrocephalus is characterised by excess cerebrospinal fluid (CSF) within the ventricles of the brain [25]. Treatment typically involves implantation of a shunt system that diverts CSF from the ventricles to the abdomen [26]. A major failure mechanism arises when the choroid plexus (ChP) tissue, which produces CSF and is attached to the ventricular wall, is drawn into the inlet holes of the catheter and occludes it. This complication occurs in over 50% of shunt systems [27]. Figure 1 illustrates the biological setting and an example ventricular catheter geometry. Previous computational studies modelled CSF flow through catheters using fluid-only approaches [28, 29, 30, 31, 32]. These studies employed flow-based metrics, such as inflow uniformity, as proxies for ChP deformation. In contrast, we previously introduced a 3D FSI model of the ventricle–catheter–ChP system [33], which demonstrated that several designs proposed in fluid-only studies could potentially be outperformed by a novel design identified through a 2D FSI parameter sweep. This work suggested that fluid-only metrics are not reliably correlated with ChP deformation, and that FSI effects are central to accurately assessing occlusion risk.

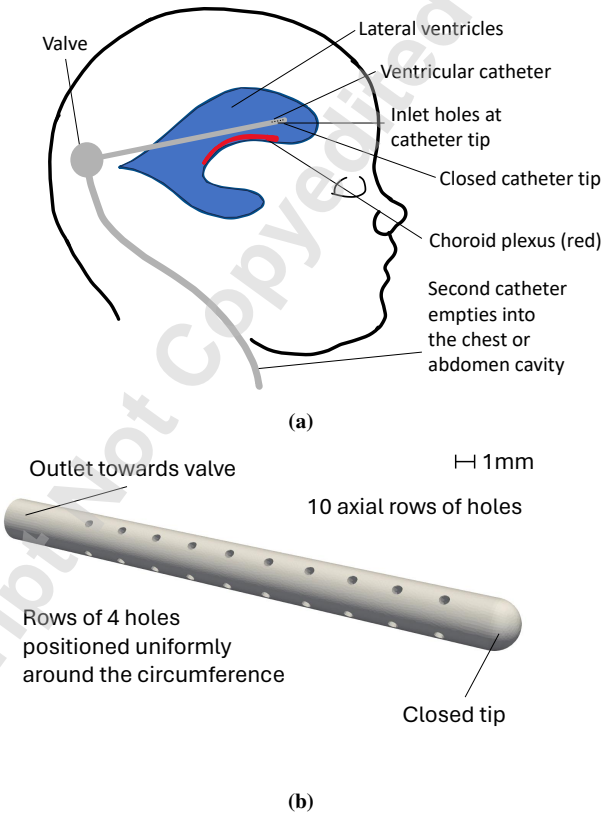


Figure 1: Figures reproduced with permission from Ref. [33] showing: a) biological domain of the shunt system (blue, lateral ventricle; red, ChP tissue; grey, shunt system); b) visualisation of an example ventricular catheter, showing the 30 mm closest to the closed tip.

The ROM employed here consists of a fluid-only simulation of the ventricle–catheter–ChP system with rigid boundaries, from which the fluid stress field on the ChP surface is extracted. This stress field on the FSI interface is the primary mechanical driver of ChP deformation. An ML layer then learns the relationship between this stress field and the maximum equilibrium ChP deformation predicted by the full FSI model. The resulting surrogate is sufficiently fast for use within an optimisation loop, though less accurate than the full model. The ROM is then coupled with the computational opti-

misation protocol proposed in Ref. [7] to systematically identify promising catheter designs beyond those previously considered. Here, an improved design is defined as one that reduces the likelihood of ChP occlusion by minimising the maximum equilibrium deformation of the ChP induced by CSF flow. This target is adopted from Ref. [33] and captures the potential for blockage of the catheter by the ChP tissue (the primary cause of post-operative catheter blockage). We consider designs with precisely four inlet holes per circumferential ring, consistent with previous catheter designs, as well as designs with a variable number of holes. With further knowledge of manufacturing constraints, additional terms could be incorporated into the fitness function to reflect practical considerations, such as limiting the total number of holes. Promising designs identified by the surrogate-aided optimisation are subsequently evaluated in the full FSI model to determine their true fitness and are shown to consistently outperform existing designs from Refs. [33, 32].

In addition to advancing knowledge in hydrocephalus catheter design, this work establishes a flexible and generalisable optimisation framework applicable to multi-physics design problems where high-fidelity simulations are too computationally expensive for direct use in iterative search. By combining a reduced fluid-only model that captures stresses on the FSI interface with an ML layer trained to predict the resulting structural response, the methodology exploits the governing physics of interface-driven deformation while maintaining computational tractability. The resulting surrogate model enables optimisation in settings where repeated full-model evaluations are prohibitive. Additionally, we show that embedding mechanistic structure via ROMs improves predictive robustness relative to purely data-driven surrogates. We note, however, that, while this optimisation methodology successfully decreases the number of designs passed to expensive evaluation in the full model, it does not remove the requirement for experimental validation since both the surrogate and the full FSI model remain idealised representations of the physiological system.

Figure 2 summarises the surrogate-aided optimisation methodology presented in this work.

2 MATERIALS AND METHODS

2.1 Problem statement

The optimisation problem addressed in this paper is to identify the catheter geometry that minimises the maximum equilibrium deformation of the ChP. Let $\boldsymbol{d} \in \Omega \subset \mathbb{R}^p$ denote the vector of geometric parameters defining a catheter design (often referred to as the genotype [34]). The number of parameters needed to describe a design is p, which is consequently the dimension of the search space. The admissible design space Ω imposes geometric and manufacturing constraints to ensure that all candidate designs are physically realisable. The precise physical interpretation of $\boldsymbol{d}$ used here and its mapping to a physical catheter geometry are described in Section 2.4.2.

As demonstrated in Ref. [33], the orientation of the catheter relative to the ChP significantly influences the resulting deformation. Since catheter orientation is not prescribed during surgery, performance must be assessed over all possible orientations. We therefore introduce a rotation angle $\alpha \in [0°, 360°)$, measured from the horizontal as defined in Figure 3, to parameterise the relative orientation between the catheter and the ChP.

The maximum equilibrium deformation of the ChP $D(\boldsymbol{d}, \alpha)$ thus depends on both design and orientation. The clinically relevant performance metric is the worst-case deformation over all orientations,i.e., $\max_{\alpha \in [0°, 360°)} D(\boldsymbol{d}, \alpha)$, and the optimisation problem is therefore formulated as

$$\boldsymbol{d}^* = \underset{\boldsymbol{d} \in \Omega}{\operatorname{argmin}} \left(\max_{\alpha \in [0°, 360°)} D(\boldsymbol{d}, \alpha) \right). \tag{1}$$

Because direct evaluation of $D(\boldsymbol{d}, \alpha)$ requires repeated full FSI simulations, we introduce a surrogate approximation $\widehat{D}(\boldsymbol{d}, \alpha)$ constructed using the MxL framework described below. The optimisation is performed using the surrogate objective

$$\hat{\boldsymbol{d}}^* = \underset{\boldsymbol{d} \in \Omega}{\operatorname{argmin}} \left(\max_{\alpha \in [0°, 360°)} \widehat{D}(\boldsymbol{d}, \alpha) \right). \tag{2}$$

Since $\widehat{D}$ is an approximation of D, there is no guarantee that $\boldsymbol{d}^* = \hat{\boldsymbol{d}}^*$. Instead, the surrogate-aided optimisation identifies candidate designs that are subsequently evaluated using the full FSI model to determine their true performance. The predictive accuracy of $\widehat{D}$ depends on the quality and representativeness of the training data, and its uncertainty is quantified during surrogate validation.

The optimisation problem above is solved using a surrogate-aided optimisation pipeline, illustrated in Figure 2. The framework comprises three components: i) the full 3D FSI model, which defines the true fitness function and supplies training data; ii) a MxL surrogate that approximates this fitness by combining a ROM with an ML layer; and iii) an optimisation algorithm that searches the design space using the surrogate objective. The full model is not used directly within the optimisation loop,

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

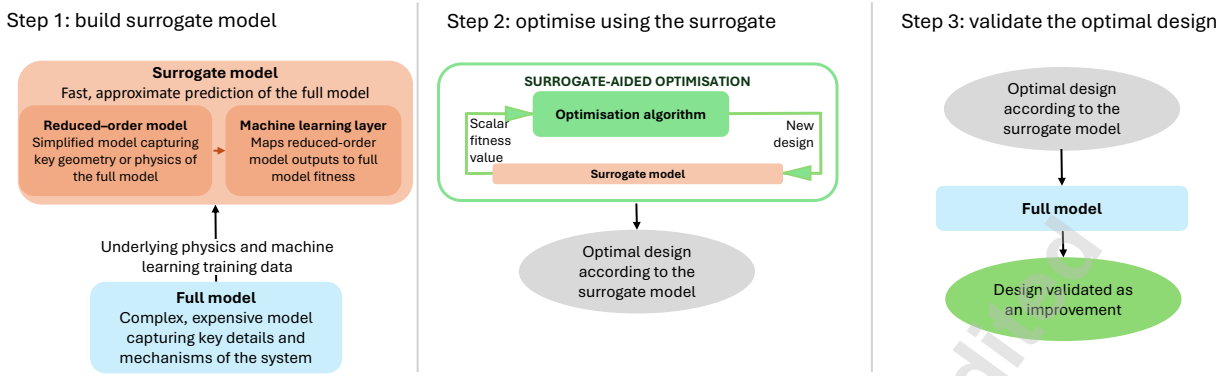


Figure 2: Schematic of the surrogate-aided optimisation methodology presented in this work.

but is employed to construct the surrogate and to evaluate the final candidate designs.

2.2 Optimisation algorithm

The optimisation framework used in this work is the softFEM methodology previously introduced by the authors in Refs. [35, 7, 36, 37]. In this framework, a multiple offspring sampling (MOS) method is employed [38]. MOS is a hybrid, self-adaptive, derivative-free strategy that combines exploitative and exploratory algorithms to efficiently search high-dimensional design spaces.

Exploitatory algorithms typically include gradient-based search methods, which refine design parameters by systematically decreasing the fitness value [39]. Here, the multi-trajectory search (MTS) method is used, progressing step by step by varying one parameter at a time to descend the fitness landscape [40]. Exploratory algorithms, by contrast, investigate solutions across the search space while incorporating information about local structure [41]. In this work, differential evolutionary (DE) strategies are employed, in which new candidate solutions are generated through mutation, recombination, and selection [35, 42, 43]. We use an adaptation of the DE method known as success-history based adaptive DE (SHADE), which allows the parameters controlling mutation and crossover to vary dynamically and stores historical parameter values to guide future updates [44, 45]. By dynamically balancing MTS and SHADE, MOS efficiently searches the space of design parameters, identifying promising regions while maintaining diversity. Full details of the MOS implementation are provided in Ref. [36, 37].

The optimisation is implemented using the python-bound `PyMOS` package [46]. The MOS protocol is initialised with a fixed budget of 1,000 surrogate fitness evaluations and a population size of 15, consistent with previous work [46]. The stopping criterion is the exhaustion of the evaluation budget. The total budget is split equally between the two component algorithms, MTS and SHADE. SHADE is initialised using default hyperparameters (history of 100 population members, initial mutation and crossover rates of 0.5 for all population members, as described in Ref. [45]). MTS does not require additional hyperparameter initialisation, as discussed in Ref. [36, 37].

All optimisations in this study are performed on the surrogate objective. The full FSI model is not evaluated within the optimisation loop, but is used subsequently to assess the true performance of the final candidate designs.

2.3 Full FSI model

The detailed FSI model of the ventricle–catheter–ChP system is described in Ref. [33], which provides full details of the geometric assumptions, material properties, and numerical validation. Briefly, this model considers an idealised 3D geometry of the ventricle-catheter-ChP domain and uses finite element and finite volume methods to solve for the equilibrium displacement of the ChP and the flow of CSF, respectively. In this 3D FSI implementation, the fluid stresses exerted by the CSF on the ChP surface drive its deformation. The shapes of the ventricle and ChP are biologically inspired by measurements of infant ventricles reported in Ref. [48, 49, 50], with dimensions selected to lie within ranges documented in the clinical imaging literature. The idealised domain geometry and key modelling assumptions are presented in Figure 3.

The CSF within the ventricle is modelled as a viscous, Newtonian, incompressible fluid with rheological properties analogous to water, with density of 1000

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

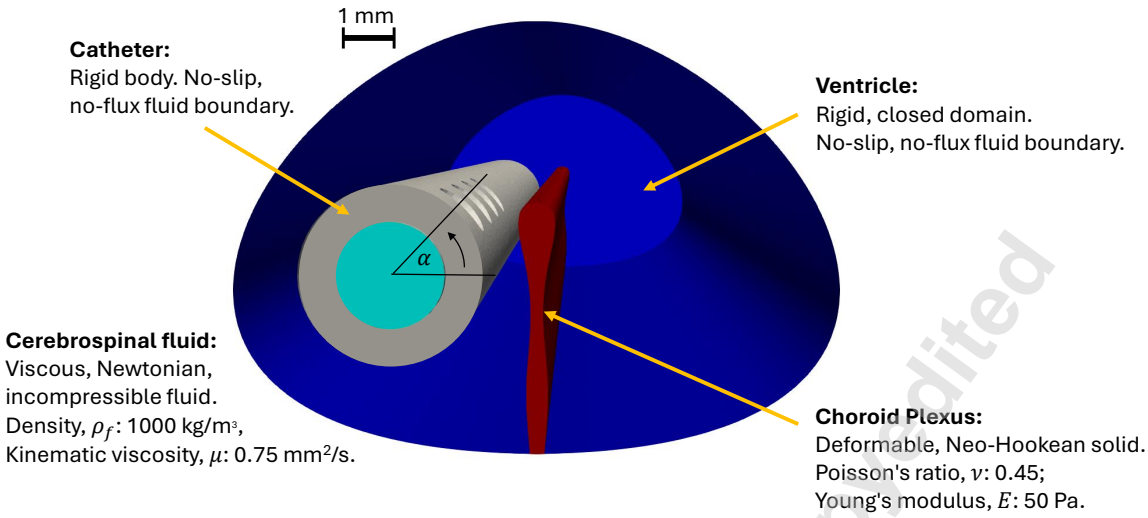


Figure 3: 3D FSI model geometry and key modelling assumptions. Dark blue: ventricle domain; red: ChP; grey: catheter; light blue: catheter outlet. The parameter α (the angle between the positive x axis and the first inlet hole) controls the rotation of the catheter around its centreline relative to the ChP. Image adapted from Ref. [33] with the addition of modelling assumptions in text, under a Creative Commons Attribution 4.0 International License (CC BY 4.0 [47]).

kg/m^{-3} and viscosity of 0.75 mm^2/s [51]. The ChP is modelled as a Neo-Hookean hyperelastic solid, with a Poisson's ratio of 0.45 [52] and Young's Modulus of 50 Pa [33]. This value of Young's Modulus is chosen from the large biologically realistic range discussed in the literature, and was selected to allow for successful discrimination between catheter designs, as discussed in Ref. [33]. The ChP is rooted to the base of the ventricle and is allowed to slip along the ventricle wall at its end faces.

The catheter and ventricle boundary are both modelled as rigid and impermeable, as their respective Young's Moduli are orders of magnitude greater than the ChP. The assumption of impermeability neglects the porous nature of brain tissue, assuming the most serious scenario of hydrocephalus where all the CSF must drain through the catheter and not through natural outlets [33]. The catheter is positioned to the left of the ChP, with the parameter α controlling its relative rotation (α is the angle between the positive horizontal and the first inlet hole anticlockwise, as shown in Figure 3). The full governing equations and boundary conditions of the 3D FSI model are given in Ref. [33] and presented in Appendix 5 for convenience.

This model accounts for the FSI effects between the CSF and ChP, and uses a weakly-coupled methodology with static solvers to find the equilibrium state of the ventricle-catheter-ChP system. As the ChP is additionally the source of CSF into the ventricle system, the ChP boundary is both the FSI interface and the fluid source, see Ref. [33] for more details. The model uses `gmsh` [53] to mesh the fluid and solid domains, `openFOAM` [54] to solve the fluid part using the finite volume method, and an in-house solver `MuPhiSim` [55] to solve the solid part with the finite element method. The coupling between fluid and solid domains is written in `python` and included as an example in `MuPhiSim` [55].

In this work the fitness of a catheter design is the maximum deformation of the ChP in its equilibrium position. This is a single scalar value which we use to quantify the efficacy of a given catheter design. This value is extracted from the FSI model using `ParaView` [56] for postprocessing.

2.4 Surrogate model development

The idealised 3D FSI model is considered here as the ground-truth for predicting the ChP displacement. However, it has a runtime of over 3 days when parallelised optimally over six cores on a 12th generation i7-1265U core laptop or shared university PowerEdge R7525 cluster node equipped with AMD EPYC CPU cores, making it inadequate for multiple evaluations of the fitness function. This motivates the development of a surrogate model, replacing the 3D FSI in the calculation of ChP displacement.

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

2.4.1 *Reduced-order model*

We postulate here that the fluid stress field taken from the ChP surface in its initial, relaxed position can be used to infer the final ChP displacement prediction.

The stress field on the surface of the ChP is obtained from a purely fluid dynamics simulation, modelling the ChP as rigid, in its initial, upright position (as shown in Figure 3). The same system of fluid equations, given in Appendix 5, are solved in the ventricle-catheter-ChP geometry, now with every boundary treated as rigid (no ChP deformation). As with the full FSI model, `gmsh` [53] is used to mesh the fluid domain, `openFOAM` [54] to solve the fluid problem, and `ParaView` [56] to postprocess the simulation and extract the fluid stress field on the surface of the ChP.

This simplified fluid-only model takes under 10 minutes to run when parallelised across four cores on a 12th-generation Intel i7-1265U laptop. This is a runtime of 0.23% of that of the full model when tested on the same machine—a significant speed-up. Fitness evaluations of 10 minutes are satisfactory for incorporation into an optimisation algorithm. By avoiding simulating the full FSI effects, this ROM is thus significantly faster than the full FSI model.

2.4.2 *Ground-truth dataset*

A ML layer is used to make a prediction for ChP deformation given the fluid stress data from the ROM as input. Creating the ML layer requires data to establish relationships between the fluid stresses and ChP deformation, and library of the 3D FSI simulations thus needs to be established.

The 3D FSI model is used to create a synthetic dataset of 180 simulations to use in the training and fitting of the ML layer. All catheter designs have inlet holes placed in the 30 mm of catheter nearest the tip [57]. This is required to ensure that inlet holes remain contained in the ventricle and do not interact with the surrounding brain tissue. An additional constraint of a maximum of 40 inlet holes is imposed to match the maximum number of inlet holes seen in physical examples.

The catheter designs contained in the ground-truth dataset are created as follows. First, the number of inlet holes is chosen as a random integer between 1 and 40. For each inlet hole, the radii and positions on the catheter surface are chosen uniformly and independently at random. Any proposed designs with overlapping holes are discarded before simulation. As observed in Ref. [33], the value of α, which prescribes relative rotation of the catheter to the ChP and is independent of the physical design, significantly affects the maximum ChP deformation. Clinicians do not account for α during shunt insertion, and as such, each design needs to *a priori* succeed for all values of α. Therefore several catheter designs are tested at different values of α and included as separate independent data points (see Appendix, Figure 10).

From each simulation, the maximum equilibrium displacement of the ChP is extracted, along with the initial fluid stress field on the surface of the ChP. To extract the initial stress field, the surface of the ChP is divided into a grid of 891 (= 9 × 99) rectangles, from which the average field value over each one of the rectangles is extracted. The fields of interest are the fluid pressure, and the three components of the fluid stress vector (where the stress vector is given by $\mu\left(\nabla\boldsymbol{v}+(\nabla\boldsymbol{v})^T\right)\cdot\boldsymbol{n}$ with μ the dynamic viscosity of the fluid, $\boldsymbol{v}$ the fluid velocity and $\boldsymbol{n}$ the surface normal vector). Each simulation therefore gives rise to 9×99×4 = 3,564 features for the input data. These are stored as a vector of length 3,564 for each data point, and can be concatenated together to form the input data matrix, a matrix of size 3,564 × 180 (features by data points).

It is worth emphasising that the proposed ML model solely learns the relationship between the ChP deformation and the fluid stress field, and that the design parameters and α are not included as inputs into the ML layer. Therefore the same design at a different value of α gives a different fluid stress field, and consequently is a novel data point to be learnt by the ML layer. As will be discussed below, the worst-case behaviour over α is thus enforced during optimisation by explicitly evaluating multiple α values for each proposed design.

Before being used for the training of the ML model, each feature of the input data is scaled so that values of this feature across the dataset have mean and variance of zero and one, respectively. This removes scale effects and allows features to be directly compared for relative importance. Due to the high ratio of features to data points, a dimensionality investigation is performed. A low-dimensioned feature space is beneficial as the training dataset will cover a higher proportion of the space, and lessen the chance of multi-collinear input data. Here we use principal component analysis (PCA) [58] to investigate the "explanatory" dimension of the data, although other algorithms such as UMAP and t-SNE are well established [59].

The "explanatory" dimension of the input data is defined here as the dimension of the PCA component vector set that contains the majority of the sample variance. The fluid stress field data are highly correlated; although these have a dimension of 3,564, only 54 and 101 dimensions are required to explain 95% and 99% (respectively) of the sample variance. The explanatory dimension is signifi-

cantly smaller than the size of the ground-truth dataset (180). This increases confidence that the ML models will be able to predict from the training set without overfitting.

It is sometimes beneficial to transform input data by projecting onto the subspace spanned by the PCA component vectors that explain the majority of data variance. This is investigated in the following section, projecting onto the 54 PCA vectors that explain 95% of the data variance.

2.4.3 *Machine learning layer*

An ML layer is used to correlate the fluid stress field to the maximum equilibrium ChP deformation. The fluid stress data is the input $\boldsymbol{x}$ (a vector of length 3,564) and the maximum equilibrium ChP deformation is the output y (a scalar). The ML algorithm and hyperparameters are chosen to maximise predictive power. A range of ML models are tested, all available in the `python3` module `scikit-learn`[60] with the exception of the extreme gradient boosted regressor XGBoost which is taken from the package `XGBoost` [61]. The list of tested models is (given in order of increasingly complexity): linear regression, ridge regression, lasso regression, support vector regression, decision tree, random forest, gradient boosted regressor, XGBoost and multi-layer perceptron (a small neural network).

Given the small sizes of ground-truth datasets available (where size is limited to ensure a reasonable time period) from expensive computational models or experimental work, overfitting is a concern. Overfitting is the detrimental hyper-specialisation of the model, which often occurs with an overly complex model with a high number of internal parameters capturing too much of the details of the training data, including noise and internal features, and consequently being less able to generalise to new data points. To prevent overfitting, we use a robust validation process with nested layers of cross validation, described in detail below. The simplest ML models are tested first, as simpler models with fewer internal parameters are less likely to overfit the data. For this reason, neural networks more complex than the MLP are not investigated due to their high number of internal parameters which make them more susceptible to overfitting.

Model selection is performed using the R^2 score to compare between models, where

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{f}(\boldsymbol{x}_i))^2}{\sum_i (y_i - \overline{y})^2}, \tag{3}$$

i indexes the training data set (so each i refers to a different catheter design), y_i is the maximum ChP deformation of FSI simulation i, and $\boldsymbol{x}_i$ is the fluid stress data of FSI simulation i. The prediction of the ML layer applied to input stress data $\boldsymbol{x}_i$ is $\hat{f}(\boldsymbol{x}_i)$, and $\overline{y}$ is the mean over y_i. $R^2 = 1$ means the data is perfectly explained by the model, while $R^2 \leq 0$ gives a model with no explanatory power. We additionally consider other accuracy metrics: root mean squared error (RMSE), mean absolute error (MAE) and Spearman's correlation coefficient ρ to obtain a holistic view of the ML performance.

Optimal hyperparameters for each model are determined through standard five-fold internal cross-validation, with 10% of the dataset held back for validation. Hyperparameter tuning is carried out using the `GridSearchCV` function from `scikit-learn`, which performs an exhaustive search over the specified parameter grid and selects the combination that maximises the R^2 score. Each model is trained and evaluated with five-fold cross-validation, while the held-back validation set is used to assess performance on unseen data with the optimal hyperparameters. A summary of the model selection algorithm is given in the Appendix (Algorithm 2. The standard choice of five folds is used in the cross validation.

Given the limited size of the ground-truth dataset, a nested cross-validation procedure is used: the data are partitioned into ten outer folds. In each round, one fold serves as the validation set while the remaining nine provide training data for inner cross-validation and hyperparameter tuning (Algorithm 2). After hyperparameter optimisation, the models with the greatest mean R^2 across the outer fold validation sets are taken forward (results for all models shown in the Appendix, Table 6, alongside tested hyperparameter ranges).

The highest validation R^2 is achieved by the gradient boosting algorithm, which also produced the lowest RMSE and MAE. The strongest correlation metric (ρ) is obtained by the random forest, while XGBoost performed comparably across all metrics. Across these three methods, the between-fold standard deviation of R^2 (on the order of 0.1) is similar to or larger than the differences in mean R^2, rendering the models statistically indistinguishable. All three are tree-based learners; however, they differ in ensemble strategy: random forests use bagging, whereas gradient boosting and XGBoost are boosting methods. These three algorithms are therefore carried forward as the most promising.

We also integrate PCA preprocessing of the input data into the training pipeline, and give the accuracy scores across each model combined with PCA in the Appendix (Table 7). The PCA projection helps to remove noise from the input data by prioritising features that have the highest variance, and gives a substantial increase in

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

performance across the simpler regression models. However the tree-based models see a small reduction in performance with PCA preprocessing as these models are inherently more robust to outliers, and lose performance with the dimensionality reduction. None of the PCA-aided models outperform the tree-based models discussed above. Subsequently, we thus avoid further PCA preprocessing.

A single hyperparameter optimisation search (Algorithm 2) is then rerun for each of the three best models, now without reserving a separate validation set. The optimal hyperparameters for each model are reported in Table 10. Finally each model, with its optimal hyperparameters (as selected above), is retrained on the full dataset. This maximises training data for the final surrogate, improving predictive performance. The accuracy metrics are calculated again for the final model, analysing the accuracy of the models performance on the whole dataset. In this case only, the model is tested on the same data that it is trained on ("in-sample"). Table 1 reports both the fold-averaged scores from the second hyperparameter search, and the in-sample accuracy when the model is retrained on the full data set. Table 1 shows that further hyperparameter tuning improved the accuracy metrics in every case. Gradient boosting and XGBoost behave similarly, as expected for two gradient-boosted tree implementations. The random forest performs slightly worse than the boosted methods. Figure 4 shows a visualisation of these models' performance, with the predicted values $\hat{f}(\boldsymbol{x}_i)$ plotted against the ground-truth value y_i for all data points i in the training dataset. Points lying close to the red line $\hat{f}(\boldsymbol{x}_i) = y_i$ indicate good predictions. Results are shown for the three models (columns) using cross-validated (out-of-fold) predictions (top row) and after refitting in-sample on the full dataset (bottom row). The top row reflects performance on data that are distinct from training folds, whereas the bottom row reflects in-sample performance. In-sample performance are, by definition, expected to exceed out-of-fold performance, explaining why the bottom row metrics are higher. However, the extreme improvement of the boosted models to $R^2 \geq 0.999$ suggests some level of overfitting to the training data, which is a particular concern given the small dataset size. Note that all conclusions regarding the predictive capability and model selection are based on out-of-sample performance, and in-sample results are reported here solely to illustrate model behaviour and diagnose potential overfitting.

For these reasons, we select the random forest model for integration into the surrogate, sacrificing a small amount of accuracy (slightly lower R^2) to reduce the risk of deploying an overfitted model. The optimal hyperparameters of this random forest are given in Table 10.

Feature importance Feature importance is a useful metric to examine the internal workings of an ML layer, and will be used here to further justify our choice of the Random Forest model. Each entry of the feature importance vector gives a measure of how important the corresponding feature is for the model predictions. In ensemble tree-based models, a feature importance is based on how often that feature is a decision node in any one tree. Here, each feature corresponds to the value of a fluid stress field at a specific point on the ChP surface. Analysing feature importance can therefore give some mechanistic insight into a physical problem: here features correspond to the stress field at different points of the ChP surface. Features being identified as important therefore identifies specific locations of the stress field on the ChP surface that are decisive in determining ChP deformation.

Figure 5 shows feature importance for the random forest (subfigure a) and XGBoost (subfigure b); the gradient boosting algorithm's importances are visually similar to XGBoost and are hence not shown. Figure 5 a) shows that the random forest attributes greatest importance to the pressure field, with a particularly influential region between 25 and 30 mm. The y-component of the fluid stress field also has elevated importance in this range. This aligns with Ref. [33], where the largest ChP deformations occurred near the catheter outlet. The most informative features form spatial clusters instead of single hot spots. This is expected as the underlying pressure and stress fields change smoothly across the surface, so neighbouring locations tend to carry similar information. This means the random forest is detecting physically consistent patterns, not just individual pixels.

In contrast, Figure 5 b) shows the gradient boosted model's feature importances, which exhibit sharp, isolated peaks: consistent with overfitting, where a few features can dominate the algorithm and correspond to noise specific to the training data.

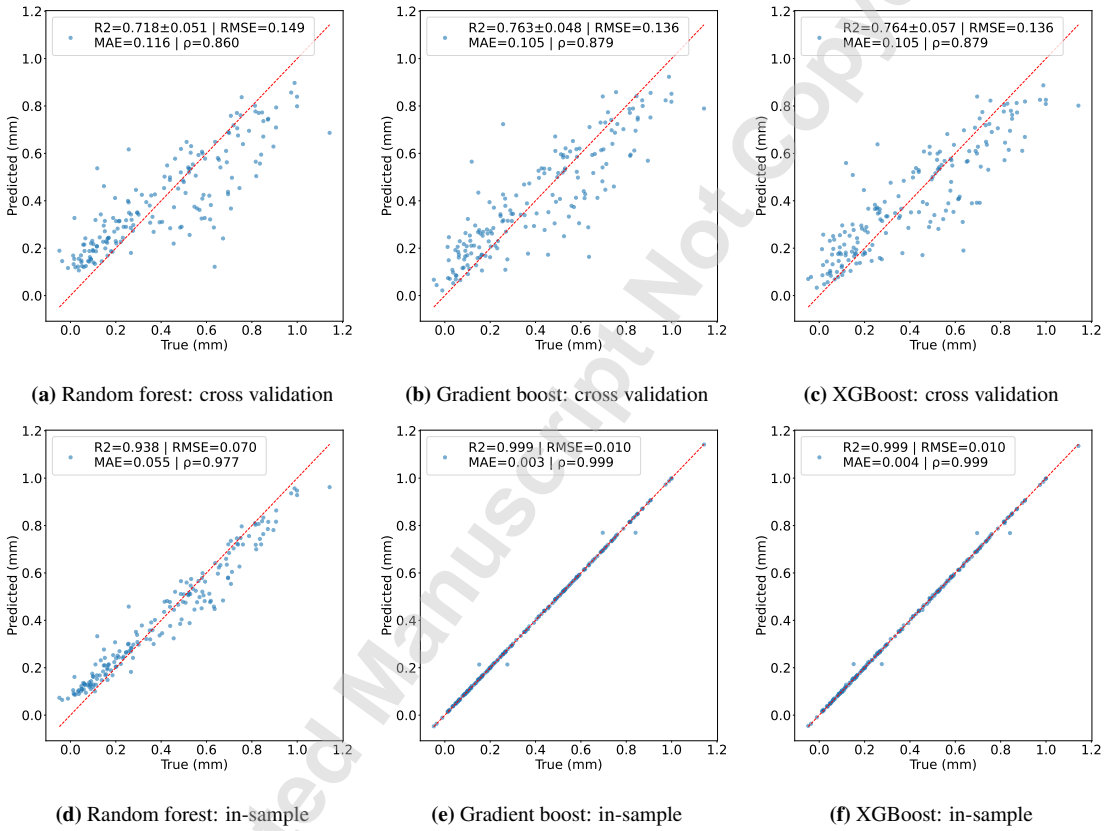
The random forest surrogate (comprised of feature weights and tree structures) is saved in a `pickle` file (∼5 MB). The file is small and can be moved and reloaded easily. Given a new instance of fluid stress data, the model returns a ChP deformation prediction in under 1 ms on one core of an i7 laptop.

Efficacy of ROM features The surrogate developed above uses fluid stress field features generated by a fluid-only ROM as inputs to the ML layer. To assess the importance of this mechanistic component, we additionally trained surrogate models in which the ML layer only takes as input the 40 catheter design parameters described in Section 2.4.2. This study evaluates the predictive performance of a surrogate model without the ROM component.

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

	Accuracy averaged over validation folds				In-sample accuracy			
Model	R^2	RMSE (mm)	MAE (mm)	ρ	R^2	RMSE (mm)	MAE (mm)	ρ
Random forest	0.718 ± 0.051	0.149	0.116	0.860	0.938	0.070	0.055	0.977
Gradient boost	0.763 ± 0.048	0.136	0.105	0.879	0.999	0.010	0.003	0.999
XGBoost	0.764 ± 0.057	0.136	0.105	0.879	0.999	0.010	0.004	0.999

Table 1: Performance metrics with optimal hyperparameters, showing both the accuracy averaged across validation folds (left) and the in-sample accuracy after retraining on the full dataset. Where applicable, R^2 shows mean$\pm$ standard deviation across folds; RMSE/MAE/ρ show fold means.



(a) Random forest: cross validation **(b)** Gradient boost: cross validation **(c)** XGBoost: cross validation

(d) Random forest: in-sample **(e)** Gradient boost: in-sample **(f)** XGBoost: in-sample

Figure 4: Predicted vs. ground-truth of the final ML models with optimal hyperparameters for the cross validation (a,b,c) and when retrained on the full dataset (d,e,f). Red line, prediction=ground-truth.

The same nested cross-validation procedure (Algorithm 2) is used as in the ROM-based surrogate. Table 2 reports the mean R^2 performance across the ten outer validation folds for each ML model, both with and without PCA preprocessing. Negative R^2 values are truncated at 0. These results should be compared directly with Table 6, which reports performance when ROM-derived stress features are used.

Table 2 shows that the predictive performance is uniformly poor when the ML models are trained directly on catheter design parameters. Although the boosted tree methods (Gradient Boost and XGBoost) perform marginally better than other models, R^2 values remain below 0.11. In contrast, substantially higher predictive

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

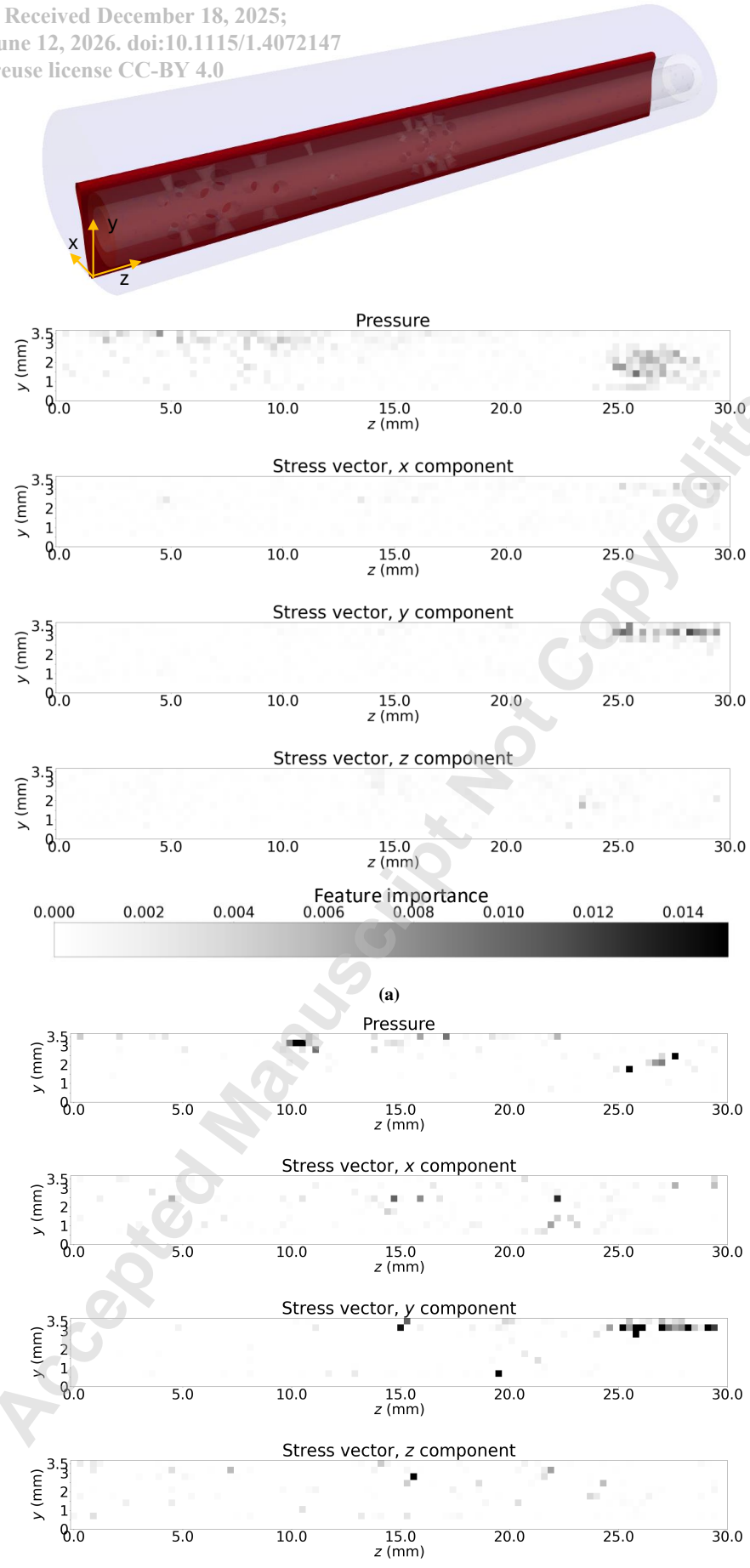


Figure 5: Feature importance plotted on the projected ChP surface for (a) random forest and (b) XGBoost. The feature importance colourbar is common to both algorithms, as is the visualisation of the ventricle geometry shown at the top of the figure. Dark pixels show areas of the stress field on the ChP which are significant in determining ChP displacement.

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

	Linear Regression	Ridge Regression	Lasso Regression	Decision Tree	Support Vector Regressor
Without PCA	0.000	0.000	0.000	0.000	0.030
With PCA	0.005	0.008	0.039	0.000	0.034
	Random Forest	Gradient Boost	XGBoost	Multi-layer Perceptron	
Without PCA	0.007	0.102	0.107	0.000	
With PCA	0.000	0.000	0.000	0.000	

Table 2: Mean R^2 performance across ten outer validation folds for ML models trained directly on catheter design parameters. Negative scores are truncated at 0.

performance is achieved when ROM-derived stress features are used as inputs, see Table 6. These results demonstrate that the ROM component provides essential mechanistic information that is not recoverable from geometric parameters alone. The stress field representation captures physically meaningful features governing ChP deformation, and its inclusion significantly enhances surrogate predictive capability.

2.5 Design encoding

The optimisation results are applied to the design of a hydrocephalus shunt system. For this specific design problem, a parameter encoding and problem-specific fitness function must first be defined.

2.5.1 Parameter encoding

An encoding describes the map between the vector used by the optimisation algorithm (often known as the genotype) and the physical catheter design implemented in the ROM (often called the phenotype) [34].

We first build on the results of Ref. [33], which show that equally sized inlet holes spread uniformly around the catheter circumference are beneficial in reducing both ChP displacement and rotational dependence. A catheter design is therefore constructed by considering circumferential rings of inlet holes. A restriction of up to six circumferential rings of inlet holes is implemented, consistent with commercially available specimens [62] and the computational investigations of Ref. [32, 28]. Each ring k of inlet holes has: a hole radius r_k (the same for all holes on the ring), a number of inlet holes n_k evenly spaced around the circumference, an offset angle ϕ_k that defines the angle between the first hole on the ring k and the horizontal axis, and a pitch length $zPos_k$ specifying the axial distance along the catheter between neighbouring rings, see Figure 6. The design dimension is therefore six rings $\times$ four parameters per ring = 24. The design parameter vector is a vector of length 24, with entries $(r_1, .., r_6, n_1, ..., n_6, \phi_1, ..., \phi_6, zPos_1, ..., zPos_6)$. The design parameter vector is both a member of 24 dimensioned Euclidean space (the genotype), as required by the optimisation algorithm, and also prescribes the physical catheter design which can be incorporated into the ventricle-catheter-ChP geometry of the reduced and full models (the phenotype).

Natural bounds on each design parameter are defined in Table 3. By allowing both the hole radius and the number of holes to take a zero value one of the rings of inlet holes can be omitted if this results in a favourable design. The upper bound of the hole radius is chosen to be similar to the maximum hole radii observed in physical examples, to account for potential, *a priori* unknown, manufacturing constraints. The pitch length and the number of holes in a ring are likewise bounded to be less than 10 mm and 10, respectively. These values are deliberately large, exceeding the ranges seen in physical specimens, which defines a larger search space than that found in existing designs, thus allowing new areas of the design space to be explored.

Table 3: Parameter value ranges for the optimisation encoding.

Variable	Lower bound	Upper bound
Hole radius: r_k	0 mm	0.6 mm
Number of holes: n_k	0	10
Offset angle: ϕ_k	0°	360°
Pitch length: $zPos_k$	0.05 mm	10 mm

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

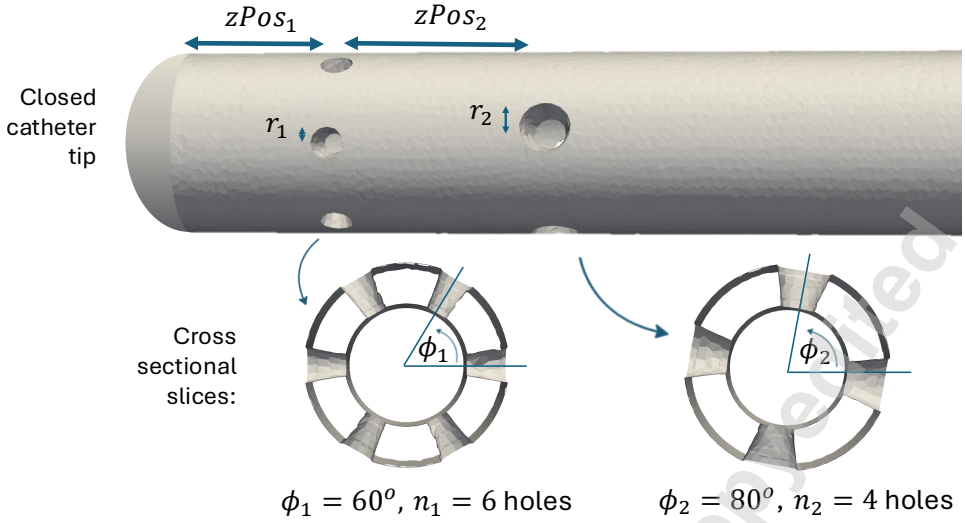


Figure 6: Visualisation of the design encoding used in optimisation, illustrating the design parameters for the first two rings of inlet holes (those closest to the closed tip). Cross sectional slices perpendicular to the centreline of the catheter are shown to illustrate the effect of parameters ϕ_k and n_k. The inlet holes are constrained to lie within the 15 mm segment of the catheter nearest the closed tip.

2.5.2 *Parameter preprocessing*

A preprocessing check is required to ensure that inlet holes do not overlap. This involves checking the values of r_k and n_k to ensure that holes do not overlap around the catheter's circumference. Specifically, the requirement is $2r_k n_k < 2\pi ac \ \forall k$ where $a = 1.25$ mm is the external radius of the catheter, and $0 \leq c \leq 1$ is a constant related to manufacturing constraints. For structural integrity, it is desirable to maintain a reasonable quantity of catheter material between neighbouring holes. Here, $c = 2/3$ to match the physical specimens seen in Ref. [33, 28, 32]. The value of c can be easily updated as improved information on the manufacturing process becomes available. Axial overlap is not possible as the encoding enforces positive $zPos_k$.

A second preprocessing check ensures that holes are placed within the closest 15 mm of the catheter tip, by enforcing that $\sum_k zPos_k \leq 15$ mm. This requirement is also based on examination of available specimens and previous work, ensuring that inlet holes remain contained within the ventricle and do not interact with the surrounding brain tissue.

Each design that fails these checks is immediately assigned an infinite fitness, and the optimisation passes to the next proposed design. We note that the modular nature of the `PyMOS` codebase makes it easy to alter the preprocessing function if new information on the biomedical or manufacturing requirements needs to be incorporated.

2.5.3 *Fitness function*

As surgical insertion is done without prescribing a particular angle α, the maximum fitness evaluation over the range of α is considered the "worst-case" scenario, and consequently is considered the final fitness to be minimised in the optimisation, see Algorithm 1.

This fitness evaluation requires one fluid simulation to be run for each catheter design at each value of α. The fitness evaluation is easily parallelised, running each value of α concurrently, and additionally parallelising within each `openFOAM` simulation using inbuilt methods. For this work each fluid simulation is parallelised over four cores, meaning each optimisation run required four $\times$ 15 = 60 cores. The optimisations are run on a university cluster with 60 CPU cores with each fitness evaluation taking 15 minutes, with the full optimisation of 1,000 evaluations taking approximately two weeks. Note that the 180 3D FSI required to train the ROM took c. two months, which, taken together with the MxL optimisation, still remains significantly less than the years-long endeavour of optimising with the full 3D FSI model.

3 RESULTS

3.1 Optimisation convergence

The optimisation algorithm contains stochastic effects, and hence multiple realisations are run, each with a different random seed. Figure 7 a) shows the conver-

Algorithm 1 Fitness function for optimisation

New parameter vector is proposed by `PyMOS`,
Preprocess parameter to test for a physically reasonable design,
if not physically reasonable **then**:
  Return fitness = infinity
else:
  for $\alpha \in [0°, 360°)$ **do**,
    Create the catheter geometry for given α and inlet hole design,
    Run a fluid simulation in `openFOAM`,
    Extract fluid pressure and stress at the ChP surface,
    Use the pre-trained random forest to predict ChP displacement,
  end for
  Return fitness = maximum predicted ChP displacement across all α.
end if

gence history of each of the optimisation realisations. For each run, the fitness decreases rapidly during the first 200 iterations, with a few, smaller, fitness gains later in the algorithm. This confirms that the budget of 1,000 is sufficient for this problem, with a converged fitness value of around 0.2 mm, despite different starting points. While all realisations reported here used 1,000 iterations, the rapid plateau in performance suggests that, in future applications, a budget of approximately 500 evaluations would likely achieve comparable improvement. Figure 7 b) shows the performance of a single optimisation realisation, with the blue line showing the fitness at each iteration. For the first 500 iterations, the explorative SHADE algorithm is used, which searches a large area of the fitness landscape. This can be seen in the large spread of fitness values in this section of the optimisation. In the second half of the optimisation, MTS is used, which leverages exploitative techniques to improve the current fitness. This can be seen with the blue trajectory, with values much closer to the red line, as the MTS algorithm looks for nearby designs with small incremental improvements to the current fitness. The cut out shows the current best fitness in the second half of the optimisation, amplified to show that fitness improvements are made throughout the optimisation, though the most significant gains in fitness occur in the first 200 iterations.

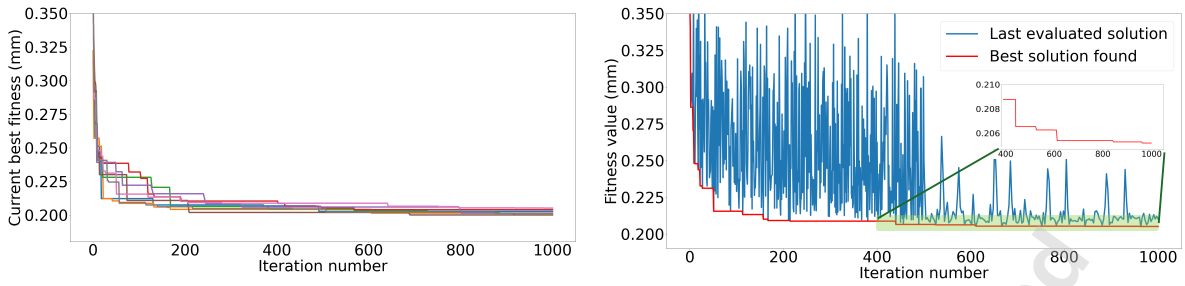
3.2 Optimised designs

Each completed optimisation realisation suggests a new candidate design that the surrogate-aided optimisation has identified as having the lowest fitness. Figure 7 shows that all the optimisation realisations suggest candidate designs where the surrogate model has predicted fitnesses of around 0.2 mm. However, from the accuracy metrics evaluated during the training procedure, the ML layer has an average error of 0.1 mm (Table 1). In particular, Figure 4 a) shows the performance of the ML layer on new data. For low values $\leq$0.3 mm, the scatter plot is below the red line (that indicates perfect predictions), meaning that the ML layer here tends to overestimate predictions in this region. As a result, while the multiple concurrent optimisation realisations produce a set of candidate designs which are predicted by the surrogate model to have a similar, low fitness value of 0.2 mm, these fitness predictions come with an uncertainty of 50%. To confirm and quantify the improved performance of these candidate designs, we test their performance in the full FSI model at all values of α. Specifically, we consider the two most promising designs (as evaluated by the surrogate model), though a greater number could be examined as computational time and resources allow. Figure 8 shows visualisations of these two designs.

The two identified designs both show a cluster of holes close to the closed tip, and a second cluster close to the midpoint of the catheter, where $z = 15$ mm. There is a visible difference in the radii of the inlet holes, with design (a) having consistently larger inlet holes than those in design (b). Interestingly, both designs recover the findings of Ref. [33], in that larger n_k values (catheter with a high number of holes placed circumferentially for each ring of holes) are desirable.

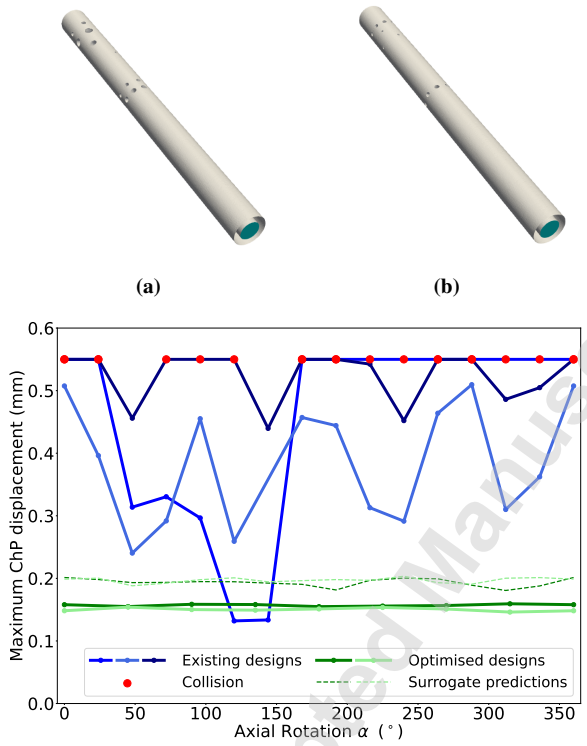
These improved designs are then tested in the full 3D FSI model to evaluate their true performance. Figure 8 shows the maximum ChP displacement of the improved designs evaluated by the full model plotted at various values of α in shades of green, along with the surrogate predictions (green, dashed), compared to the results of previous designs in current clinical use, presented in Ref. [33] in shades of blue. These results confirm that promising results proposed by the surrogate model are indeed true design improvements after testing in the 3D FSI.

The optimised designs outperform all three of the previous designs at almost every value of α. There are no ChP collision events, and the optimised designs additionally show close to no sensitivity to α, a highly desirable characteristic, as it is not possible to select α during implant surgery. The two designs, despite differences in their inlet holes lead to very similar ChP displacements. This is a positive for the design process as it suggests that



(a) (b)

Figure 7: a) Optimisation convergence history: current minimum fitness for each iteration number. Colours: different optimisation realisations, run concurrently from different seeds. b) Algorithm performance for a particular realisation. Blue: current iteration fitness; red: current minimum fitness. The minimum fitness for the final 600 iterations is amplified in the cut out.



(a) (b) (c)

Figure 8: a) and b) Optimal catheter designs identified by the surrogate-aided optimisation framework. c) Maximum ChP displacement against α, comparing designs suggested by the optimisation (green shades) to the physical design examples investigated in Ref. [33, 32] (blue shades). The surrogate model predictions for the new designs are shown in dashed lines. Collision with the ChP is marked by red dots.

the fitness of the catheter is not overly sensitive to small changes in the design parameters (for example, the small errors introduced during the manufacturing process).

In both cases, the true performance of the designs, as evaluated by the full model, is better than the surrogate prediction (ChP displacement of 0.15 mm compared to surrogate predictions of 0.2 mm). As discussed earlier, we quantified the uncertainty in the surrogate model predictions as an absolute error of 0.1 mm, and hence this discrepancy is unsurprising. This difference emphasises the need to test the optimal designs identified by the surrogate-aided optimisation in the full model, so as to identify the true fitness of the design and remove the uncertainty inherent to the surrogate fitness evaluation. Importantly, Figure 8 c) confirms that the optimised designs achieve reductions in deformation that exceed the average surrogate prediction error, demonstrating the meaningful improvement provided by the proposed MxL approach.

3.3 Four-hole designs

In the encoding described above inlet holes are positioned in rings around the circumference of the catheter, with any number of holes allowed in a ring (up to 10), and rings allowed to be offset from each other (parameter ϕ_k). However, the physical designs reviewed in Ref. [33, 28, 63] all restrict their investigation to designs with four inlet holes in each ring and no relative circumferential offset between the rings. We now consider an optimisation restricted to this subspace of designs. This is of interest: i) for comparison against previous work, ii) because of the likelihood that the current manufacturing process has a preference for four-hole designs given that all current examples are of this form, and iii) to evaluate whether designs generated with this restriction can perform as well as the more general designs previously

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

studied. Under these conditions $n_k = 4$ and $\phi_k = 0 \ \ \forall k$ for each ring of inlet holes, reducing the design space to a dimension of 12.

Figure 9 shows two designs with the four-hole configuration generated from two separate runs of the optimisation algorithm. As with the designs presented in Figure 8, both four-hole designs have one group of large holes clustered close to the closed tip and another group close to the axial midpoint of the catheter, near $z = 15$ mm.

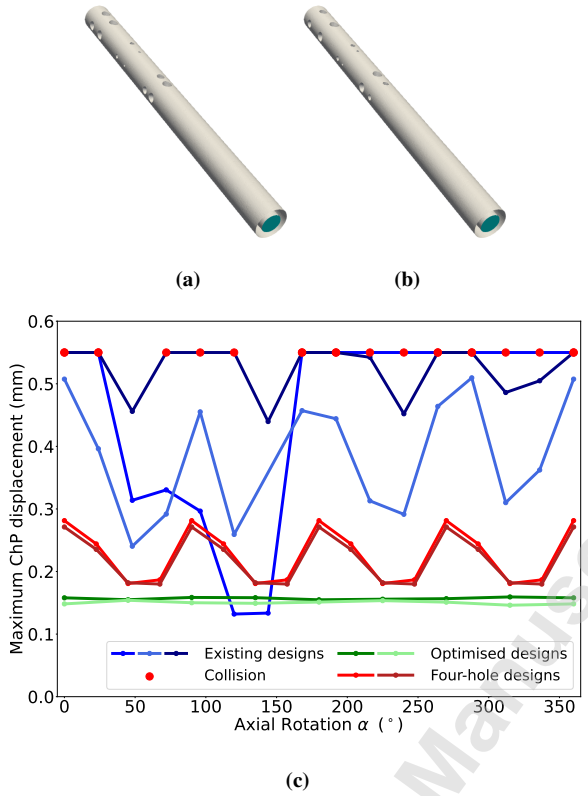


Figure 9: a) and b) Optimal catheter designs enforcing a regular four-hole arrangement on each ring of inlet holes. c) Maximum ChP displacement against α, comparing the four-hole designs (red) to designs suggested by the general encoding (green) and to the physical design examples investigated in Ref. [33, 32] (blue). Collision with the ChP is marked by red dots.

The optimised four-hole design is then tested in the full model for 16 values of α across the range $(0°, 360°)$. Figure 9 shows the maximum ChP displacement plotted against values of α.

The four-hole designs identified by the optimisation are again improvements over the existing designs. There is still dependence on α, but the overall ChP displacement is reduced. The displacement of the ChP for the four-hole designs varies between a minimum displacement of ~ 0.2 mm and a maximum displacement of just under 0.3 mm. While the variation due to α is less for the four-hole design than for the pre-existing examples shown in blue, this optimisation suggests it is not possible to completely remove the sensitivity to α with a four-hole design. Again, despite differences in the design of the two catheters shown in Figure 9, the ChP displacement curve is very similar (red lines), suggesting a variety of designs lead to similar improvement in catheter performance. These are however still outperformed by the previously identified designs of Figure 8 a) and b).

4 CONCLUSION

This paper presented a surrogate-aided MxL methodology for computational optimisation, and illustrated its use in the context of hydrocephalus catheter design. Design problems can often be defined as a computational optimisation, but using the solutions of high-fidelity, complex computational models as fitness functions in optimisation algorithms is expensive. Accordingly, earlier work on hydrocephalus catheter design neglected the use of computationally expensive FSI models, instead making use of fluid-only models and evaluating performance using fluid-based metrics [64, 28, 29, 30, 31]. In our previous study [33], we introduced the first 3D FSI model of the hydrocephalus scenario and demonstrated that fluid-only metrics, when used without accounting for tissue mechanics, are not reliably correlated with ChP deformation. As the full 3D FSI model is too expensive to be used for more than a few tens of runs and thus cannot be directly incorporated into an optimisation sweep, we proposed here the development of a MxL-based surrogate model to directly replace the full model at every fitness evaluation.

A key advantage of this methodology is that it preserves the mechanistic structure and low computational cost of fluid-only models, while augmenting them with an ML layer trained on outputs from the full FSI simulations. Doing so, physically interpretable fluid features are correlated with the quantity of interest (here, the ChP deformation) as trained with high-fidelity data. The resulting surrogate model therefore combines the capability of the full model at a fraction of the computational cost, enabling its integration within optimisation workflows that would otherwise be computationally prohibitive. While the ML component does not impose explicit physics constraints within its loss formulation (as is done in physics-informed neural networks, among others), the surrogate structure remains mechanistically grounded through the

ROM, which encodes the governing fluid physics and supplies physically interpretable features to the ML layer. An important final step consists in testing the design proposed by the optimisation in the high-fidelity full model due to the error in the surrogate fitness evaluations (which here can reach c. 50%). The final ranking of candidate designs is therefore based on a full model evaluation rather than surrogate prediction.

For the specific case of hydrocephalus shunt design, our methodology suggests new designs that are clear improvements over examples currently used in the UK National Health Service, as evaluated by our fine-grained model. Our results, which recover those of Ref. [33], suggest that a large number of inlet holes spread around the catheter circumference are particularly beneficial in reducing the impact of the insertion angle α. We additionally found that the proposed optimal designs cluster holes near the closed tip and the axial midpoint of the catheter, suggesting that such arrangement participates in reducing ChP deformation.

The surrogate models constructed using this methodology can only be as accurate as the full model used to generate the training data. In this work, the ground truth is a 3D FSI model, which remains an idealisation of the in vivo ventricle–catheter scenario and incorporates simplified geometry, material representations, and boundary conditions. Consequently, the optimisation results are conditioned on these modelling assumptions and should be interpreted within this context. Experimenta and/or clinical validation is therefore a necessary next step to assess the real-world performance of the proposed designs. Furthermore, if other clinically relevant objectives were to become of interest, the framework could readily accommodate alternative fitness definitions. For example, maintaining low intracranial pressure is clinically desirable, and a modified fitness function could incorporate an additional term to account for this. With further knowledge of manufacturing constraints, additional terms could be incorporated into the fitness function to reflect practical considerations, such as limiting the total number of holes.

While the present work focuses on hydrocephalus shunt occlusion, the underlying methodology is not application-specific. The surrogate constructed here maps features of the initial fluid stress field on the fluid–solid interface to the resulting equilibrium solid deformation. This structure applies to problems in which (i) design performance is evaluated by a computationally expensive multi-physics model, (ii) a ROM is available to capture key physical mechanisms, and (iii) the quantity of interest cannot be sufficiently well recovered from the ROM alone. Many engineering and biomedical design problems satisfy these conditions, and particularly those involving FSI where device performance is determined by solid deformation. Examples include the optimisation of heart valves [65], implant devices (focussing either on porous [66] or patient-specific properties [67, 68]), skeletal joint replacement [69], and ureteric flow washout devices [70]. In such settings, full multi-physics simulations are often too expensive for direct use within optimisation loops, while ROMs lack sufficient predictive accuracy. The present framework provides a systematic approach to embedding high-fidelity information within a computationally efficient surrogate suitable for computational optimisation.

More broadly, the framework is independent of the specific ML architecture, optimisation algorithm, or governing physics. Its contribution is therefore methodological as well as application-specific, providing a general strategy for enabling tractable optimisation in multi-physics design problems where direct use of the full model is infeasible.

5 ACKNOWLEDGMENTS

The authors would like to acknowledge funding from the EPSRC CDT in Sustainable Approaches to Biomedical Science: Responsible and Reproducible Research - SABS:R3 (EP/S024093/1). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

All codebases developed for this work are publicly available. The 3D FSI model can be found as the example `py-FSI-3D` in https://github.com/muphisim, [55], and the surrogate development code is available in https://TBD/mxl-gen.

REFERENCES

[1] Kapoor, A., Ray, T., Jepson, N., and Beier, S., 2025, "A surrogate-assisted multiconcept optimization framework for real-world engineering design," *J. Mech. Des.*

[2] Jiang, L., Gu, X., and Chen, S., 2020, "Generative design of bionic structures via concurrent multiscale topology optimization and conformal geometry method," *J. Mech. Des.*

[3] Krish, S., 2011, "A practical generative design method," *Comput. Aided. Des.*

[4] Gradišar, L., Dolenc, M., and Klinc, R., 2024, "Towards machine learned generative design," *Autom. Constr.*

[5] Smorzhenkov, N., and Ignatova, E., 2021, "The use of generative design for the architectural solutions synthesis in the typical construction of residential buildings," *E3S Web Conf.*

[6] Zhou, S., Li, W., Chen, Y., Sun, G., and Li, Q., 2011, "Topology optimization for negative permeability metamaterials using level-set algorithm," *Acta Mater.*

[7] Peña, J.-M., LaTorre, A., and Jérusalem, A., 2019, "Softfem: The soft finite element method," *Int. J. Numer. Methods Eng.*

[8] Fogel, D., 2018, "Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review.," *Contemp. Clin. Trials Commun.*

[9] Chauhan, V., Shukla, V., and Sharma, N., 2024, "Review of medical device design," *Int. J. Sci. Res.,* **13**.

[10] Cardillo, G., and Camporeale, C., 2021, "Modeling fluid–structure interactions between cerebro-spinal fluid and the spinal cord," *J. Fluids Struct.*

[11] Gholampour, S., 2018, "FSI simulation of CSF hydrodynamic changes in a large population of non-communicating hydrocephalus patients during treatment process with regard to their clinical symptoms," *PLoS One*.

[12] Chen, Z.-Q., Wang, R.-L., Sanchez, R.-V., De Oliveira, J., and Li, C., 2018, "An adaptive genomic difference based genetic algorithm and its application to memetic continuous optimization," *Intel. Data Anal.*

[13] Holloman, C., Lee, H., and Higdon, D., 2006, "Multiresolution genetic algorithms and markov chain monte carlo," *J. Comput. Graph. Stat.*

[14] Sáez, Y., 2009, "Optimization using genetic algorithms with micropopulations," *Optimization Techniques for Solving Complex Problems*.

[15] Jin, Y., Olhofer, M., and Sendhoff, B., 2001, "Managing approximate models in evolutionary aerodynamic design optimization," *IEEE Trans. Evol. Comput.*

[16] Liu, S., Wang, H., Peng, W., and et al., 2024, "Surrogate-assisted evolutionary algorithms for expensive combinatorial optimization: a survey.," *Complex Intell. Syst.*

[17] Feng, Y., Chen, Y., Yao, Y., and et al., 2022, "The brain as a structure: A model of how fluid–structure interactions stiffen brain tissue after injury," *Eng. Struct.*

[18] Lin, C.-Y., 2020, "Alternative form of standard linear solid model for characterizing stress relaxation and creep: Including a novel parameter for quantifying the ratio of fluids to solids of a viscoelastic solid," *Front. Mater. Sci.*

[19] Su, L., Wang, M., Yin, J., and et al., 2023, "Distinguishing poroelasticity and viscoelasticity of brain tissue with time scale," *Acta Biomater.*

[20] Metzcar, J., Jutzeler, C., Macklin, P., Köhn-Luque, A., and Brüningk, S., 2024, "A review of mechanistic learning in mathematical oncology," *Front. Immunol.*

[21] van Aalst, M., Nies, T., Pfennig, T., and Matuszyńska, A., 2025, "Mxlpy - python package for mechanistic learning in life science," *bioRxiv*.

[22] Benzekry, S., 2020, "Artificial intelligence and mechanistic modeling for clinical decision making in oncology.," *Clin. Pharmacol. Ther.*

[23] Procopio, A., Cesarelli, G., Donisi, L., Merola, A., Amato, F., and Cosentino, C., 2023, "Combined mechanistic modeling and machine-learning approaches in systems biology – a systematic literature review," *Comput. Methods Programs Biomed.*

[24] Raissi, M., Perdikaris, P., and Karniadakis, G., 2019, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*

[25] Linninger, A., Tangen, K., Hsu, C.-Y., and et al., 2016, "Cerebrospinal fluid mechanics and its coupling to cerebrovascular dynamics," *Annu. Rev. Fluid Mech.*

[26] Vinchon, M., Rekate, H., and Kulkarni, A., 2012, "Pediatric hydrocephalus outcomes: a review.," *Fluids Barriers CNS*.

[27] Aschoff, A., Kremer, P., Hashemi, B., and et al., 1999, "The scientific history of hydrocephalus and its treatment.," *Neurosurg. Rev.*

[28] Galarza, M., Giménez, A., Pellicer, O., and et al., 2015, "New designs of ventricular catheters for hydrocephalus by 3-d computational fluid dynamics.," *Childs Nerv. Syst.*

[29] Galarza, M., Giménez, A., Valero, J., and et al., 2015, "Basic cerebrospinal fluid flow patterns in ventricular catheters prototypes.," *Childs Nerv. Syst.*

[30] Galarza, M., Giménez, A., Valero, J., and et al., 2014, "Computational fluid dynamics of ventricular catheters used for the treatment of hydrocephalus: a 3d analysis.," *Childs Nerv. Syst.*

[31] Galarza, M., Pellicer, O., and et al., 2016, "Influence of the hole geometry on the flow distribution in ventricular catheters for hydrocephalus.," *BioMed. Eng.*

[32] Weisenberg, S., 2016, "Improving ventricular catheter design through computational fluid dynam-

ics," PhD thesis, University of Tennessee Thesis submission.

[33] Hayman, E., Nguyen, V.-D., McFarlane, I., and et al., 2025, "Reduced model aided fluid-structure interaction design framework for shunt systems," *Med. Eng. Phy.*

[34] Fischer, S., De Oliveira, J., Mumford, J., and Kell, L., 2021, "Using a genetic algorithm to optimize a data-limited catch rule," *ICES J. Mar. Sci.*

[35] LaTorre, A., Muelas, S., and Peña, J.-M., 2011, "A mos-based dynamic memetic differential evolution algorithm for continuous optimization: a scalability test," *Soft Comput.*

[36] Nguyen, V.-D., Sanchez Naharro, P., Pena, J.-M., and Jerusalem, A., 2025, "Soft-finite-element-method-driven multiphysics airfoil optimization," *A. I. A. A.*

[37] Nguyen, V.-D., Sanchez Naharro, P., Pena, J.-M., and Jerusalem, A., 2025, "Softfem driven metamaterial design," *J. Mech. Phys. Solids*.

[38] Latorre, A., 2009, "A framework for hybrid dynamic evolutionary algorithms: Multiple offspring sampling (mos)," PhD thesis, University of Tennessee.

[39] Liu, D., and Nocedal, J., 1989, "On the limited memory bfgs method for large scale optimization.," *Math. Program.*

[40] Tseng, L.-Y., and Chen, C., 2008, "Multiple trajectory search for large scale global optimization," *IEEE Trans. Evol. Comput.*

[41] Venter, G., 2010, "Review of optimization techniques," *J. Aerosp. Eng.*

[42] Behera, S., Sahoo, S., and Pati, B., 2015, "A review on optimization algorithms and application to wind energy integration to grid.," *Renewable Sustainable Energy Rev.*

[43] Kumar, A., Dutta, S., and Gandomi, A., 2022, "Surrogate model-driven bio-inspired optimization algorithms for large-scale and high-dimensional problems," In *Biomimicry for Aerospace*. Elsevier.

[44] Tanabe, R., and Fukunaga, A., 2013, "Evaluating the performance of shade on cec 2013 benchmark problems," In IEEE Trans. Evol. Comput.

[45] Tanabe, R., and Fukunaga, A., 2013, "Success-history based parameter adaptation for differential evolution," In IEEE Trans. Evol. Comput.

[46] Naharro, P., Toharia, P., LaTorre, A., and et al, 2022, "Comparative study of regression vs pairwise models for surrogate-based heuristic optimisation," *Swarm Evol. Comput.*

[47] Creative Commons Attribution 4.0 International License http://creativecommons.org/licenses/by/4.0/ Accessed: 2025-10-03.

[48] Boucher, M.-A., Lippé, S., Dupont, C., and et al., 2018, "Computer-aided lateral ventricular and brain volume measurements in 3d ultrasound for assessing growth trajectories in newborns and neonates," *Phys. Med. Biol.*

[49] Lu, Z., He, J., Yu, Y., and et al., 2019, "Measurement of lateral ventricle volume of normal infant based on magnetic resonance imaging.," *Chin. Neurosurg. J.*

[50] Schuh, A., Makropoulos, A., Robinson, E., Cordero-Grande, L., Hughes, E., Hutter, J., Price, A., Murgasova, M., Teixeira, R. P. A. G., Tusor, N., Steinweg, J. K., Victor, S., Rutherford, M. A., Hajnal, J. V., Edwards, A. D., and Rueckert, D., 2018, "Unbiased construction of a temporally consistent morphological atlas of neonatal brain development," *bioRxiv*.

[51] Gupta, S., Soellinger, M., Boesiger, P., and et. al., 2008, "Three-dimensional computational modeling of subject-specific cerebrospinal fluid flow in the subarachnoid space.," *J. Biomech. Eng.*

[52] Morin, F., Chabanas, M., Courtecuisse, H., and et al, 2017, "Biomechanical modeling of brain soft tissues for medical applications," In *Biomechanics of Living Organs*. Academic Press.

[53] Geuzaine, C., and Remacle, J.-F., 2009, "Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities," *Int. J. Numer. Meth. Engng.*

[54] Weller, H., Tabor, G., Jasak, H., and et. al., 1998, "A tensorial approach to computational continuum mechanics using object-oriented techniques," *Comput. Phys.*

[55] Nguyen, V.-D., Kirchhelle, C., Abdollahi, A., and et al., 2023, "A numerical framework coupling finite element and meshless methods in sequential and parallel simulations," *Finite Elem. Anal. Des.*

[56] Ahrens, J., Geveci, B., and Law, C., 2005, *ParaView: An End-User Tool for Large Data Visualization, Visualization Handbook* Elsevier.

[57] Tomei, K., 2017, "The evolution of cerebrospinal fluid shunts: Advances in technology and technique.," *Pediatr. Neurosurg.*

[58] James, G., Witten, D., Hastie, T., and et al., 2013, *Introduction to Statistical Learning : with Applications in R.* Springer.

[59] Draganov, A., and Dohn, S., 2024, Relating tsne and umap to classical dimensionality reduction.

[60] Pedregosa, F., Varoquaux, G., Gramfort, A., and et al., 2011, "Scikit-learn: Machine learning in Python," *J. M. L. R.*

[61] Chen, T., and Guestrin, C., 2016, "Xgboost: A scalable tree boosting system," *J.A.C.M.*
[62] Life, I. Neurocritical care, ventricular catheters.
[63] Scarff, J., 1966, "Evaluation of treatment of hydrocephalus: Results of third ventriculostomy and endoscopic cauterization of choroid plexuses compared with mechanical shunts," *Arch. Neurol.*
[64] Weisenberg, S., TerMaath, S., Barbier, C., and et al., 2018, "A computational fluid dynamics simulation framework for ventricular catheter design optimization.," *J. Neurosurg.*
[65] Danilov, V., Klyshnikov, K., Onishenko, P., Proutski, A., Gankin, Y., Melgani, F., and Ovcharenko, E. A., 2023, "Perfect prosthetic heart valve: generative design with machine learning, modeling, and optimization," *Front. Bioeng. Biotechnol.*
[66] Egan, P., Bauer, I., Shea, K., and Ferguson, S., 2019, "Mechanics of three-dimensional printed lattices for biomedical devices," *J. Mech. Des.*
[67] Kanagalingam, S., Dalton, C., Champneys, P., and et al., 2022, "Detailed design for additive manufacturing and post processing of generatively designed high tibial osteotomy fixation plates.," *Prog. Addit. Manuf.*
[68] Ricotta, V., Campbell, R., Ingrassia, T., and et al., 2020, "A new design approach for customised medical devices realized by additive manufacturing.," *Int. J. Interact. Des. Manuf.*
[69] Hernandez-Aristizabal, D., Arroyave-Tobón, S., Marquez-Florez, K., and Linares, J.-M., 2025, "Bio-inspired generative design of contact interfaces subjected to time-dependent loading conditions," *Results Eng.*
[70] Williams, J., Wechsung, F., Turney, B., and et al., 2021, "Shape optimisation for faster washout in recirculating flows," *J. Fluid Mech.*
[71] Madhukar, M., Choudhary, A., Boal, D., and et al., 2012, "Choroid plexus: normal size criteria on neuroimaging.," *Surg. Radiol. Anat.*
[72] Segura, M., and Enríquez, R., 1990, "Measurement of the choroid plexus of the lateral ventricles in full-term newborns using ultrasonics.," *Bol. Med. Hosp. Infant. Mex.*
[73] Sweetman, B., Xenos, M., Zitella, L., and et al., 2011, "Three-dimensional computational prediction of cerebrospinal fluid flow in the human brain," *Comput. Biol. Med.*
[74] Brydon, H., Hayward, R., Harkness, W., and Bayston, R., 1995, "Physical properties of cerebrospinal fluid of relevance to shunt function. 2: The effect of protein upon csf surface tension and contact angle.," *Br. J. Neurosurg.*

APPENDIX A: 3D FSI GOVERNING EQUATIONS

Here, we present the full set of governing equations for the 3D FSI model. These equations are solved using the finite element and finite volume approaches, as detailed in Ref. [33], with the geometry shown in Figure 3. The variables used in these equations are summarised in Table 4, with constant parameter values given in Table 5.

Fluid equations:

$$\nabla \cdot \boldsymbol{v} = 0, \tag{4}$$

$$\rho_f(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\nabla p + \mu \nabla^2 \boldsymbol{v}. \tag{5}$$

Solid equations:

$$\nabla \cdot \boldsymbol{\sigma} = \mathbf{0} \tag{6}$$

$$\text{where} \quad \boldsymbol{\sigma} = \frac{E}{3(1-2\nu)}(J-1)\mathbb{I} \tag{7}$$

$$+ \frac{E}{2(1+\nu)} J^{-\frac{5}{3}}(\mathbb{B} - \frac{1}{3}(\text{tr}(\mathbb{B}))\mathbb{I}), \tag{8}$$

$$\text{and} \quad \mathbb{B} = (\mathbb{I} + \nabla \boldsymbol{u})(\mathbb{I} + \nabla \boldsymbol{u})^T. \tag{9}$$

Fluid boundary conditions:

$$\boldsymbol{v} = \mathbf{0} \ \text{ on ventricle walls and catheter material,} \tag{10}$$

$$\boldsymbol{v} = \frac{Q}{A_{ChP}}\boldsymbol{n} \ \text{ on ChP surface,} \tag{11}$$

$$\boldsymbol{v} - (\boldsymbol{v} \cdot \boldsymbol{n})\boldsymbol{n} = \mathbf{0}, \ \text{ and} \tag{12}$$

$$-p + \boldsymbol{n} \cdot \mu \left(\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T\right) \cdot \boldsymbol{n} = 0 \ \text{ at fluid outlet,} \tag{13}$$

Solid boundary conditions:

$$\boldsymbol{u} = \mathbf{0} \ \text{ on ChP base,} \tag{14}$$

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} - (\boldsymbol{n} \cdot \boldsymbol{\sigma} \cdot \boldsymbol{n})\boldsymbol{n} = \mathbf{0} \ \text{and} \tag{15}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on ChP end-faces,} \tag{16}$$

$$\left(-p\mathbb{I} + \mu \left(\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T\right)\right) \cdot \boldsymbol{n} = \boldsymbol{\sigma} \cdot \boldsymbol{n} \ \text{ on ChP surface.} \tag{17}$$

Table 4: Variables within the full governing equations

Fluid velocity:	$\boldsymbol{v}$	Solid point displacement: $\boldsymbol{u}$
Fluid pressure:	p	Solid stress tensor: $\boldsymbol{\sigma}$
Normal vector:	$\boldsymbol{n}$	Solid Jacobian: $J = \det(\mathbb{I} + \nabla \boldsymbol{u})$

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

Table 5: Model geometry parameters.

Parameter	Value	Reference
A_{CP} ChP surface area	220.33 mm^2	[71, 72]
ρ_s Solid density	$1.05 \cdot 10^3$ kg/m^3	[19]
E Young's modulus	50 kg /(m s^2)	[52]
ν Poisson's ratio	0.45	[52]
ρ_f Fluid density	10^3 kg/m^3	[32]
Q Fluid production	6 mm^3/s	[73]
μ Fluid viscosity	$7.5 \cdot 10^{-4}$ kg/(m s)	[74]

APPENDIX B: GROUND-TRUTH ALPHA DEPENDENCE

In the ground-truth dataset used in the training of the ML model (Section 2.4), several data points relate to the same physical catheter design, tested at different values of α. Figure 10 shows a histogram of the number of values of α tested for each catheter design. Where $\alpha > 1$ the values of α are chosen uniformly in the range [0°, 360°]. Most designs have between 1 and 4 values of α tested, which allows the investigation of a wide variety of catheter designs, while explicitly including some α dependence into the dataset.

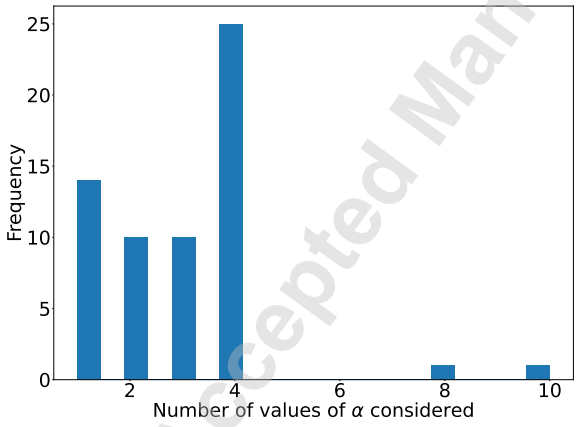


Figure 10: Frequency histogram showing the number of α values considered for different catheter designs in the ground-truth dataset.

APPENDIX C: MACHINE LEARNING MODEL SELECTION AND HYPERPARAMETER OPTIMISATION

When training the ML layer of the surrogate model, initial model selection is done using the standard cross-validation procedure, with the algorithm given below.

Algorithm 2 Hyperparameter optimisation

Remove 10% of the ground-truth data and save as the validation data,
Initialise five folds, splitting the data into five equally sized, distinct groups,
for Hyperparameters values **do**
  for *fold* $\in \{1, \ldots, 5\}$ **do**
    fold is the test set,
    The remaining four folds become the training data.
    Fit the ML model to the training data,
    Make predictions on the inputs of the test data,
    Compare predictions with outputs of the test data,
    Calculate the R^2 score for the fold.
  end for
  Calculate the mean of the fold R^2 scores.
end for
Identify the best hyperparameters by the highest mean fold R^2 score.
Retrain the model with optimal hyperparameters on all fold data,
Evaluate model performance by calculating the R^2 score on the validation data.

As discussed in the main text, the small training dataset size requires the use of a nested cross-validation procedure. Table 6 gives the mean over the ten validation datasets of each of the accuracy metrics, for each of the ML models after hyperparameter optimisation for each model with no initial PCA preprocessing. A good model has high R^2 and ρ scores and low RMSE and MAE. The best value for each metric is highlighted in **bold**. Full details of the tested hyperparameter ranges are provided below, alongside further visualisations of the performance of the models on the validation data (Figure 11).

Table 7 shows the accuracy scores from the nested cross-validation hyperparameter sweep where a PCA preprocessing step is included in the pipeline are given. Again, the best score across all the ML models considered is shown in bold.

Model: without PCA	R^2	RMSE (mm)	MAE (mm)	ρ
Linear regression	-2.411 ± 4.308	0.462	0.296	0.526
Ridge	-0.257 ± 1.564	0.284	0.195	0.653
Lasso	0.525 ± 0.169	0.188	0.155	0.761
Decision tree	0.538 ± 0.263	0.182	0.142	0.699
Support vector regressor	0.652 ± 0.139	0.162	0.132	0.788
Random forest	0.699 ± 0.093	0.152	0.118	**0.838**
Gradient boost	$\mathbf{0.721 \pm 0.099}$	**0.144**	**0.111**	0.827
XGBoost	0.705 ± 0.122	0.147	0.114	0.816
MLP	0.298 ± 0.423	0.222	0.171	0.706

Table 6: Mean of performance metric across ten outer validation folds. R^2 shows mean$\pm$ standard deviation across folds; RMSE/MAE/ρ show fold means.

Model: with PCA	R^2	RMSE	MAE	ρ
Linear regression	0.478 ± 0.136	0.143	0.202	0.781
Ridge	0.473 ± 0.139	0.143	0.205	0.778
Lasso	0.533 ± 0.118	0.152	0.193	0.751
Decision tree	0.387 ± 0.205	0.172	0.221	0.640
Support vector regressor	$\mathbf{0.687 \pm 0.097}$	**0.120**	**0.158**	0.823
Random forest	0.567 ± 0.126	0.155	0.186	0.811
Gradient boost	0.679 ± 0.102	0.125	0.160	0.828
XGBoost	0.670 ± 0.108	0.126	0.162	0.815
MLP	0.486 ± 0.174	0.136	0.202	**0.861**

Table 7: Mean of performance metric across ten outer validation folds with a PCA preprocessing step. R^2 shows mean$\pm$ standard deviation across folds; RMSE/MAE/ρ show fold means.

Tables 8 and 9 give a brief summary of each hyperparameter for each ML model, and the values tested in the grid-search. The three graph based models: random forest, gradient boost and XGBoost share several hyperparameters due to their similar structure. The values chosen for investigation are taken from a small range around default values.

Figure 11 shows the performance of each optimised ML model on the validation data. As discussed in the main text, the random forest model is chosen for the surrogate as it has the highest predictive power. The optimised hyperparameters for this model are given in Table 10.

Table 8: Hyperparameter optimisation (part one).

ML model	Hyperparameter	Values tested
Linear regression	`fit_intercept`: Whether to include intercept	True, False
	`positive`: Restrict coefficients to be positive	True, False
Ridge	`alpha`: Regularisation strength	0.01, 0.1, 1.0, 10.0, 100.0
	`fit_intercept`: Whether to include intercept	True, False
	`solver`: Optimisation algorithm	auto, svd, cholesky, lsqr, sag, saga
Lasso	`alpha`: Regularisation strength	1e-4, 1e-3, 1e-2, 1e-1, 1.0
	`fit_intercept`: Whether to include intercept	True, False
	`selection`: Coordinate descent update order	cyclic, random
Decision tree	`max_depth`: Maximum depth of tree	None, 2, 3, 5, 7, 10
	`min_samples_split`: Minimum samples to split node	2, 5, 10
	`min_samples_leaf`: Minimum samples at each leaf	1, 2, 4, 8
	`max_features`: Features considered at each split	None, sqrt, log2
	`splitter`: Split selection strategy	best, random
Support vector regressor	C: Regularisation strength	0.1, 0.5, 1
	γ: Influence of individual samples	1e-5, 1e-4, 0.001
	ϵ: Width of the no-penalty region	0.1, 0.3, 0.5
Random forest	`n_estimators`: Number of trees	100, 200, 300
	`max_depth`: Maximum depth of trees	2, 3, 5
	`max_features`: Features considered at each split	sqrt, log2
	`min_samples_split`: Minimum samples to split node	2, 3, 5
	`min_samples_leaf`: Minimum samples at each leaf	1, 2, 4
Gradient boost	`n_estimators`: Number of trees	100, 200, 300
	`learning_rate`: Contribution of each tree	0.05, 0.1, 0.15
	`max_depth`: Maximum depth of trees	2, 3, 5
	`subsample`: Fraction of samples per tree	0.8, 0.9, 1.0
	`min_samples_leaf`: Minimum samples at each leaf	1, 2, 4
XGBoost	`n_estimators`: Number of trees	100, 200, 300
	`learning_rate`: Contribution of each tree	0.05, 0.1, 0.15
	`max_depth`: Maximum depth of trees	2, 3, 5
	`subsample`: Fraction of samples per tree	0.8, 0.9, 1.0
	`colsample_bytree`: Fraction of features per tree	0.8, 0.9, 1.0

Journal of Mechanical Design. Received December 18, 2025;
Accepted manuscript posted June 12, 2026. doi:10.1115/1.4072147

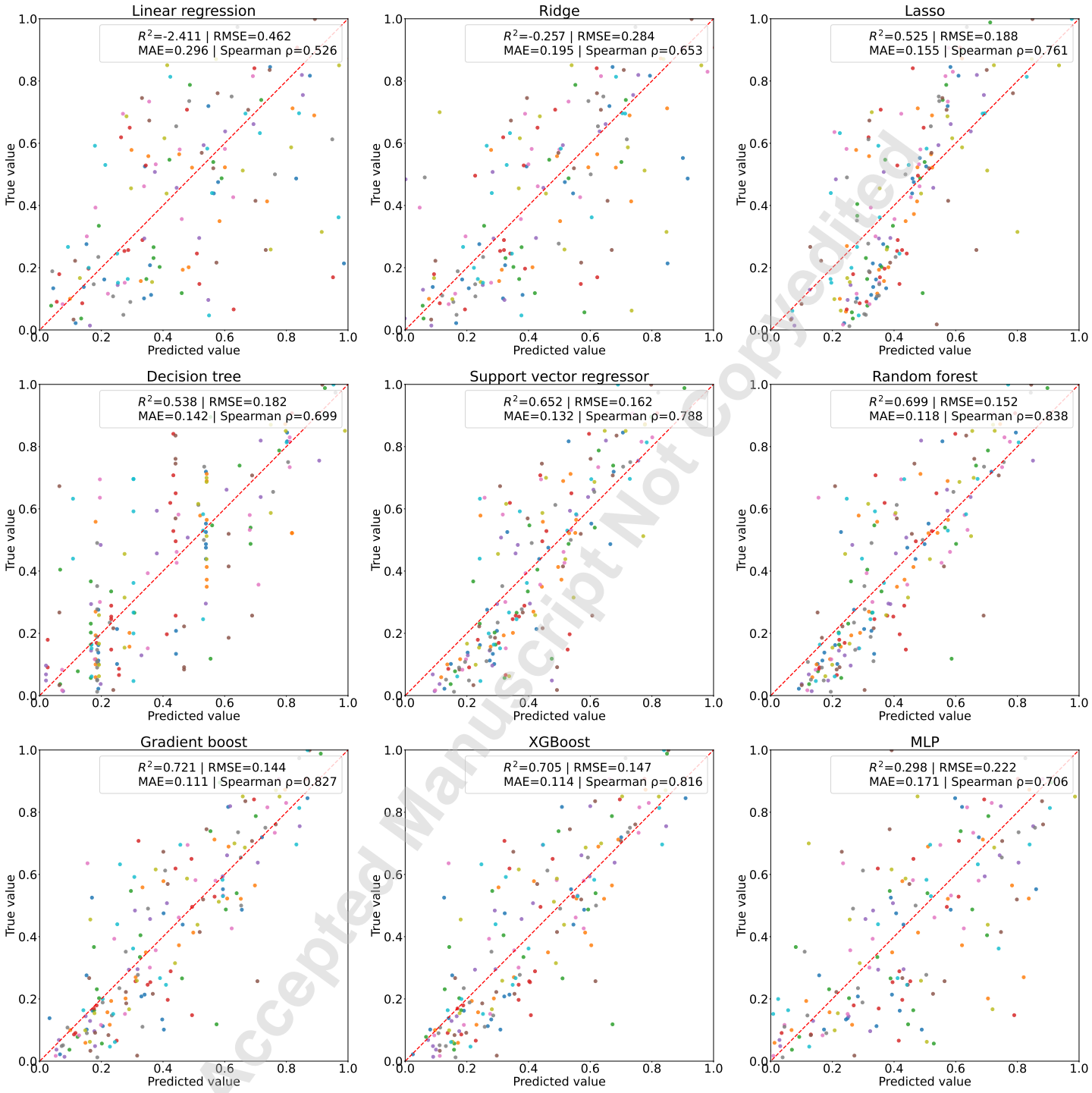


Figure 11: Predicted and ground-truth data for the validation set, with optimal hyperparameters, for each model (colours: different validation folds).

Table 9: Hyperparameter optimisation (part two).

ML model	Hyperparameter	Values tested
MLP	`hidden_layer_sizes`: Number/size of hidden layers	(50,), (100,),(100,50), (50,50,50)
	`activation`: Hidden layer activation function	relu, tanh
	`solver`: Optimisation algorithm	adam, lbfgs
	`alpha`: Regularisation strength	1e-5, 1e-4, 1e-3, 1e-2
	`learning_rate`: Learning rate schedule	constant, adaptive
	`learning_rate_init`: Initial learning rate	1e-3, 1e-2
	`early_stopping`: Use validation to stop training	True, False

Hyperparameter	Optimal value
`n_estimators`	200
`min_samples_split`	3
`min_samples_leaf`	4
`max_features`	`sqrt`
`max_depth`	5

Table 10: Optimised hyperparameters for the random forest.