Title
• Harnessing Interpretable Machine Learning for Origami Feature Design and Pattern Selection

Authors
Yi Zhu*, Evgueni T. Filipov¹,²

Affiliations
1. Department of Civil and Environmental Engineering, University of Michigan at Ann Arbor
2. Department of Mechanical Engineering, University of Michigan at Ann Arbor
* Corresponding Author (email: yizhu-cee@umich.edu)

Abstract
Engineering design of origami systems is challenging because comparing different origami patterns requires using categorical features and evaluating multi-physics behavior targets introduces multi-objective problems. This work shows that a decision tree machine learning method is particularly suitable for the inverse design of origami. This interpretable machine learning method can reveal complex interactions between categorical features and continuous features for comparing different origami patterns, can tackle multi-objective problems for designing active origami with multi-physics performance targets, and can extend existing origami shape fitting algorithms to further consider non-geometrical performances of origami systems. The proposed framework shows a holistic way of designing active origami systems for various applications such as metamaterials, deployable structures, soft robots, biomedical devices, and many more.

Teaser
Machine learning computes interpretable rules for designing origami systems with different patterns and multi-physical behaviors
Introduction

Origami, the art of folding paper, provides a method to build novel 3D engineering structures from flat 2D surfaces. These origami structures can be used in a variety of applications such as biomedical devices (1, 2), micro/soft robots (3, 4, 5), frequency selective surfaces (6, 7), metamaterials (8, 9), aerospace structures (10, 11), and many more. Over the years, there have been a number of inverse design methods proposed for origami systems (12, 13, 14, 15), but these methods and algorithms only solve kinematical design problems like fitting origami to arbitrary shapes and geometries. Designing active origami structures for general engineering applications is difficult because of categorical variables are needed to represent origami features, there is a lack of clear criteria for comparing different patterns, and highly nonlinear and multi-physical behaviors are common for active origami. Moreover, these origami systems can experience large deformation and change of function during the folding process, which can add additional difficulties to the construction of a rigorous inverse design problem. Mathematically, a generic origami inverse design solution method should be able to consider multiple objectives and the interaction between categorical features and continuous features. This is because comparing competing origami patterns requires using categorical features and solving multi-physics performance targets will naturally introduce multi-objectives problems.

Machine learning has proven itself to be a versatile and powerful method to solve physical science problems (16), financial problems (17), biomedical problems (18), e-sports games (19), etc. Moreover, a large number of different machine learning methods like neural networks (20), rule lists (21), boosting (22), random forest (23), and many others are being explored to solve problems of different size, complexity, and nature. Because of the broadness and diversity of these methods, one key challenge in is to select the appropriate machine learning method for a given engineering problem. For origami type problems, machine learning techniques have been used to predict chaotic dynamic responses (24) and to solve for origami folding motions (25). However, no prior work has effectively tackled the problem of inversed design for origami.

In this work, we show that an interpretable machine learning method called the decision tree and its ensemble version called random forest (26, 23) are particularly suitable for tackling the inverse design problem of active origami. Figure 1 summarizes the fundamental idea of this work. The design of origami can be thought of as a nonlinear function \( f \) to calculate the performance indices of the system (such as stiffness, Poisson’s ratio, material cost, etc.) based on given design features (such as the number of origami cells, the thickness of materials, sector angles of the origami pattern, etc.). Usually, because this nonlinear function is too complex to obtain a closed form solution, numerical simulations are used to obtain the relationship between feature and performance. With this in mind, understanding the origami design means understanding how this nonlinear function behaves and the inverse design problem is resolved by obtaining the inverse function \( f^{-1} \) (this function is not guaranteed to exist mathematically, but we can still use the concept of inverse function to think about the problem). Traditionally, this is done by deliberately picking a set of numbers for the features and calculating the associated performances using simulation methods (27, 28). By plotting the relationships between the features and the performances based on these deliberately generated data points, one can understand the design rules embedded in the nonlinear function \( f \) and potentially understand the inverse relationship to approximate \( f^{-1} \).

Now, with the advancement of machine learning, it is possible to approach the problem differently in a more automated manner. Mathematically, machine learning methods (for
supervised learning problems) can be thought of as function generators that can approximate any nonlinear functions given enough data points to represent the relationship between the features and the performances. To further tackle the inverse design problem, interpretable machine learning methods can be used during the training. Unlike standard “black box” machine learning methods, interpretable machine learning methods can produce human-understandable decision rules so that people understand why certain judgments are made by the machine learning algorithm (29, 30). This allows us to obtain an interpretable approximation of the nonlinear function $f$, so that the inverse design rules can be easily constructed (we can think of the process as finding the inverse function $f^{-1}$, even if this function may not exist). More specifically, we will use one interpretable machine learning method called decision tree and its ensemble version called random forest to analyze the generated origami database (26, 23). What we found is that by selecting the “more informative” branches within different decision trees, useful design rules of origami systems can be computed for the origami inverse design problem.

The decision tree method is highly interpretable and uses a tree-like decision chart to classify a data point (23, 26). The bottom figure in Fig. 1 shows a sample decision tree. When using the decision tree, the data point will flow into the tree from the top branch node. Depending on if the data matches the rule specified in the branch node or not, the data will keep flowing to the left branch or to the right branch. Finally, after a series of judgments, the data point will fall into a leaf node where no more branching occurs, and a label (or prediction of performance) is determined. The structure of trees and the splitting rules in branch nodes are learned by the machine learning algorithm automatically during the training process. Because the decision tree is highly interpretable, we can backtrack the feature selection criteria easily to formulate the inverse design rules. Moreover, the flexible formulation of decision trees allows them to tackle multi-objective and consider the interaction between categorical and continuous features simultaneously. These capabilities make decision trees a perfect method to tackle origami design problems. This work uses a classical decision tree method implemented from the sklearn package (31). More state-of-the-art decision tree methods (like the Generalized and Scalable Optimal Sparse Decision Trees (32) and the Optimal Classification Trees (33)) could also be used in the future to further improve the performance.

In the following sections of this work, we will demonstrate how to use the decision tree method to compute human-understandable design rules for active origami design. We will use three origami design problems to further show that the proposed method can select the better design from different origami patterns, can tackle the multi-objective design of origami with multi-physics behaviors, and can enable origami shape fitting algorithms to further consider non-geometrical properties of origami. More specifically, we will use these three examples to highlight why decision tree based methods are particularly suitable for tackling the origami inverse-design problem because they can handle multi-objective and can consider the interaction between categorical and continuous features simultaneously.

Results

Compute Origami Design Rules with Interpretable Machine Learning

First, we introduce how to compute design rules for active origami systems using interpretable machine learning methods. To demonstrate the methodology, a simple design problem for a Miura origami unit cell is studied and Fig. 2A shows the setup of the problem. A Miura origami pattern is cut out from a square material and folded to 60% extension ratio ($\text{Ext}$). The extension ratio is measured as the ratio between the folded length $L'$ to the flat
length $L$ of the pattern. Four design variables (features) can be adjusted so that the single unit Miura origami pattern can have an axial stiffness $k$ to meet the target range ($k < 6000$ N/m). These four variables are the thickness of the panels ($t_p$), the thickness of the creases ($t_c$), the width of the creases ($W$), and the sector angle of the pattern ($\gamma$). The potential feature ranges are determined by considering practical fabrication limits and material limits and they are $1.0 \text{ mm} < t_p < 6.0 \text{ mm}$, $0.5 \text{ mm} < t_c < 1.0 \text{ mm}$, $1.0 \text{ mm} < W < 4.0 \text{ mm}$, and $50^\circ < \gamma < 80^\circ$.

After setting up the problem, we can randomly sample values of the design features to build a feature database (step number 1 in Fig. 2B). Each column in the table on Fig. 2B is associated with one data point. For this demo example, 1000 data points are generated. After generating the feature data, we compute the axial stiffness (the performance) using origami simulation methods. In this paper, an origami simulation package called SWOMPS (34) is used to compute the origami performance for populating the database. This simulation package is chosen because it is computationally efficient and can compute complex origami behaviors including large folding deformations, nonlinear mechanics, and electro-thermal actuations (35, 36). Finally, we can assign class labels to the data point based on if the data point meets the target performance or not as demonstrated in step number 3 on Fig. 2B. For example, those data points that have a stiffness $k < 6000$ N/m (meet the target) will be assigned label 1 while the remaining samples are assigned label 0.

Next, we can use the design features and the class labels to set up a supervised learning problem and this problem can be solved using the decision tree method. The decision tree method is trained to differentiate those designs that meet the target performances from others using the values of the design features. Figure 2C shows a sample decision tree for the origami design problem. To determine the class of a data point, the data point will be sent into the tree from the top. Each time the data point flows into a branch node, a simple criterion is checked to determine if the point should go to the left branch or to the right branch. For example, a datapoint with $t_p = 1.1 \text{ mm}$ at the first node of Fig. 2C will be further sent to the left branch because we have $t_p = 1.1 \text{ mm} < 2.1 \text{ mm}$. After a series of judgments, the data point will be sent to a leaf node, where no more branching occurs, and a class label is predicted. For example, the datapoint listed in column one of the Table on Fig. 2B will follow the gray arrow (“Rule 1”) downward and be judged as Class 1 data. This means the machine learning algorithm thinks that this feature design ($t_c = 0.65 \text{ mm}$, $t_p = 1.1 \text{ mm}$, $W = 1.2 \text{ mm}$, $\gamma = 52^\circ$) is most likely to produce a single unit Miura origami with axial stiffness $k < 6000$ N/m. The algorithm comes to this conclusion because the feature design values match the rule: $t_c < 0.8 \text{ mm}$, $t_p < 2.1 \text{ mm}$. In the light of this, each branch associated with Class 1 in the decision tree gives a design rule (the inverse function $f^{-1}$ for the inverse design problem) to produce an origami that can meet the target performance. As can be seen, the highly interpretable structure of a tree method is useful for the inverse design because the inverse design relationship (which shows how to pick features based on the target performances) can be constructed easily. Moreover, we will use randomly selected sub-datasets to train multiple different decision trees (this forms an ensemble version of the decision tree method, and it is called random forest). This allows us to create more potential design rules. The splitting criterion, the structure of the tree, and leaf node predictions are learned by the machine learning algorithm during the training process using a machine learning package called sklearn (31). After training multiple decision trees, we can gather the design rules by tracing back the tree branches as shown in Fig. 2D. For example, Rule 1 and Rule 2 gathered in Fig. 2D are correlated to the two different branches in the sample tree marked with the two gray arrows in Fig. 2C.
Although these decision trees are learned by the machine learning method automatically, there are manually specified other variables that control how decision trees are computed. These user-specified variables are referred to as hyperparameters in machine learning and a technique called grid search is usually performed to select these hyperparameters. Basically, different combinations of the hyperparameters are used to train the machine learning algorithms and the best combination is selected. Details on the grid search are provided in the supplementary material section S3.3.

Now that we have collected a number of potential design rules, we need to select those that provide better performances. Figure 2E shows two important aspects for selecting better rules. First, a design rule should have high precision, which means that the data points that follow the design rules should indeed fall within the target performance boundary. The precision is defined as the ratio between the number of accurate predictions of class \( t \) over the number of all predictions of Class \( t \) (37). In this example, Rule 1 (blue dots) predicts 10 data points as Class 1, and 9 of them are correct so it has a precision of 0.9. Moreover, a good design rule should contain a good number of data points so that they can spread the entire target performance region (making it more representative). This can be measured by directly counting the number of data points that satisfy the rule or using the recall index. The recall is defined as the accurate predictions of Class \( t \) over the number of all data of Class \( t \) (37).

Suppose we have a total number of 30 points in the target zone (within the blue box), then Rule 1 will have a recall of 0.3 (9/30). Thus, to ensure the performance of the computed design rules, we select rules that have high precision, high recall, and a larger number of data that fit the rules. For instance, among the three sample rules shown on Fig. 2E, Rule 1 (blue dots) is better than Rule 2 and Rule 3 (orange squares and pink crosses).

In this work, we use the following routine to select design rules with better performances. The rules need to satisfy two thresholds and they are: (1) the precision is greater than 0.9, and (2) the number of data points is greater than 10. Rules that do not satisfy these thresholds are eliminated from further consideration. Next, we rank the rules using the F-score function (37):

\[
F_\beta = (1 + \beta^2) \left( \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \right)
\]

This F-score function creates an average score from both the precision and the recall value, where the recall is seen as \( \beta \) times more significant as the precision value. In this work, we select a value of \( \beta = 0.2 \) because we focus more about the precision of the selected design rules. Finally, we select the rules with the highest F-scores for our origami design and Fig. 2F shows the selected rules for this demonstration example. The darker region of the box chart indicates the computed range of the four design variables and the design rule is \( t_c < 0.78 \text{ mm}, \ t_p < 2.2 \text{ mm}, \ W \geq 1.9 \text{ mm}, \ \gamma \geq 67^\circ \).

Although we managed to find a rule that performs well in the training dataset and get a 1.0 precision, we need to further test it to see if it is really good. The testing is conducted by computing the precision of the rule using another testing dataset that is not used for training the machine learning method. This is usually referred to as the hold-out testing in machine learning because a portion of the data (testing data) is held out from training the machine learning method. Basically, the training data are homework problems for the machine learning algorithm and the testing data is the final exam. Details of the hold-out testing setup can be found in the supplementary material section S3.3. For the demonstration example, the computed design rule has a testing precision of 0.86. This performance is decent for an
unbalanced dataset like the one in the demonstration example (target data only consists of 7% of the total dataset). In addition, a sample design with $t_c = 0.70 \text{ mm}$, $t_p = 1.5 \text{ mm}$, $W = 2.5 \text{ mm}$, $\gamma = 70^\circ$ is studied and presented on Fig. 2F. This design has an axial stiffness $k = 5702 \text{ N/m}$, which indeed meets the target performance.

Unlike optimization-based inverse design methods, the proposed design method provides flexible feature ranges so that engineers and researchers can further fine-tune the design based on their needs. Moreover, this highly interpretable methodology can also provide information on the significance of each design feature. More significant features will have tighter thresholds while the less significant features may not even be used. For the design of this single Miura unit, the thickness of the panel is the most significant feature to control for the design to achieve the target stiffness.

**Compare Different Origami Patterns**

Comparing different origami patterns is difficult because categorical variables are needed to represent different origami patterns. Common optimization-based origami design methods tend to use continuous optimization algorithms (12, 13) which cannot tackle categorical variables. Although Chen et. al. has demonstrated that it is possible to solve the mountain/valley fold assignment problem of origami using mixed integer program that can tackle categorical variables (14), it is unclear how this method can be extended to tackle generic origami inverse design problems. Thus, this subsection will demonstrate how the proposed machine learning based method can be used to compare different origami patterns for inverse design. The capability of comparing different origami patterns is achieved because the decision tree method can simultaneously handle continuous and categorical variables and even consider the potential interaction between them.

In this subsection, we will study a design problem of origami canopy as shown in Fig. 3A. In this study, an origami pattern is cut out from a thin square plate with a footprint of 0.2 m × 0.2 m. Two different patterns are studied, and they are the standard Miura-origami pattern (Pattern 1) and the Tachi-Miura Polyhedron (TMP) (27) (Pattern 2). To represents the two patterns, an integer (binary) variable $p = \{1,2\}$ is introduced. In addition, we further consider the situation where the two patterns can have different numbers of unit cells in them. The numbers of unit cells are represented as $m$ and $n$ in the two directions. In addition to these categorical features, three continuous design features are also added to the problem, and they are the thickness of panels $1.0 \text{ mm} < t_p < 6.0 \text{ mm}$, the thickness of creases $0.5 \text{ mm} < t_c < 1.0 \text{ mm}$, and the width of creases $1.0 \text{ mm} < W < 4.0 \text{ mm}$. An origami database is populated by randomly sampling values of these design features, and the database contains 2000 Miura origami samples and 2000 TMP origami samples. Origami simulations SWOMPS are used to calculate the stiffness performances of these different origami systems, and more details of the simulation setup can be found in the supplementary materials section S2. In this example, we will design for two targets, and they are the axial stiffness $k_a$ at 60% extension and the bending stiffness $k_b$ at 90% extension. For both stiffnesses, we create shifting targets that contain four target zones as indicated in Fig. 3B and 3C.

We first study the design for axial stiffness $k_a$ at 60% extension (see Fig. 3 (b)). Assume that we want to design an origami canopy to have $15000 \text{ N/m} < k_a < 30000 \text{ N/m}$ (zone a1), we can label those data that meet the target as 1 and the rest of the data as 0. After labeling, we can train the tree method to learn the underlying structures of this database and use the procedure in Fig. 2 to compute the design rules for this target. The computed design rule is shown in the left column of Fig. 3B. We can then repeat the process for the other three
targets $30000 \text{ N/m} < k_a < 50000 \text{ N/m}$, $50000 \text{ N/m} < k_a < 80000 \text{ N/m}$, $15000 \text{ N/m} < k_a$. This divides the data into four design target zones, and the computed rules for all targets are shown in Fig. 3B. The training precision, testing precision, and sample realizations are also provided in Fig. 3. This series of results shows how we can design for an origami canopy to have increasingly stiffer behaviors in axial compression. Interestingly, the machine learning method prefers changing the continuous variables to achieve the different stiffness targets in the axial compression. More specifically, the machine learning method thinks that controlling the thickness of creases and the width of the creases are more significant than other parameters because tighter thresholds are used for the two features.

However, the design rules can be very different when we study the 90% bending stiffness. We repeat the same process of labeling the data and computing the rules for 90% bending stiffness. Similarly, four design rules are computed for four different target zones as shown in Fig. 3C. When we investigate the result of this series of design rules, we can see that the machine learning method is paying more attention to the categorical features. As the target moves from one zone to another, the computed design rules change in a non-continuous manner because of the complex interactions between categorical variables and continuous variables. For example, when the target changes from zone b2 to zone b3 (stiffer target in 90% bending), the machine learning method indicates that increasing the thickness of the panel is enough to meet the target (only the shaded range of $t_p$ changes). However, when further increasing the requirement on bending stiffness at 90% extension (from b3 to zone b4), the trained machine learning method thinks it is better to change the categorical features and the continuous features simultaneously. A similar categorical jump is also observed when the target moves from zone b1 to zone b2. The proposed method can capture these complex interactions between the continuous features and categorical features, which are difficult to capture with optimization-based design methods.

Moreover, this example also highlights another advantage of using data science and machine learning method for solving origami design problems. Unlike optimization-based design methods, the database is reusable for different target performances. Thus, we can use the same database to design for different target performances (e.g. the different zones demonstrated in the example). This allows us to save computation time and design for active origami efficiently.

**Design for Active Origami with Multi-Physics Performances**

This subsection shows how we can use the proposed method to design active origami systems with multi-physics behaviors, which requires solving multi-objective problems. Active origami systems are superior to passive origami because they can achieve folding motions on their own without relying on human hands or external mechanisms connected to the origami (3, 4, 38, 39). However, designing active origami systems is difficult because of the coupled multi-physics behaviors that can introduce a multi-objective problem setup. Although these multi-objective problems are difficult to solve using standard optimization-based methods, they can be solved using the proposed decision tree method because the tree method provides high flexibility in labeling data points.

In this subsection, we will study how to design an electro-thermally actuated origami gripper system, and the setup of the problem is demonstrated in Fig. 4A. In this example, three gripper patterns are used to achieve the target gripping motion (closing the gripping tip to less than 1 mm gap as indicated in Fig. 4A Pattern 1). We assume that the gripper is designed using an electro-thermal bi-layer actuator demonstrated in (5). The actuator contains an
active layer on one side and a passive electro heater layer on the other side. When current is passed through the electro heater layer, Joule heating will heat up the region locally. Then, because the active expands more than the passive layer folding motion is generated (see right figure on Fig. 4A). In addition to a categorical variable \( p \) used to describe the pattern, other design features of the grippers include the length \( L_i \) and width \( L_j \) of the gripper arm, the location of the first hinge from the base (measured as a ratio \( Ra \) of the outer arm compared to the total arm length), the thickness of the two layers in the actuator design \( t_1 \) and \( t_2 \), and the width of the actuator creases.

Four indices are used to measure the performance of these grippers and they are (1) fundamental frequency \( \text{freq} \) of the gripper, (2) input heating power \( Q \) needed to close its gripping arm, (3) maximum crease temperature \( T \) during the gripping motion, and (4) stiffness \( k \) of the gripper in resisting loads applied to pry it open. The origami simulation package SWOMPS is used to simulate the behaviors for creating an origami performance database. 2000 samples are computed for each gripper. Details about the setup of the simulation can be found in supplementary materials S2.

These four performance indices contain multi-physical information about the gripper regarding the dynamics, power consumption, thermal properties, and mechanical behaviors. Traditionally, designing such active origami is difficult because we need to compare these separate multi-physical indices simultaneously and it is challenging to weigh their importance appropriately. Because of this difficulty, most existing active origami systems were designed using trial-and-error approaches (2, 3, 4). Here, we show that the proposed interpretable machine learning based method can tackle the multi-objective problem setup using the flexibility in labeling classes.

For example, suppose we want to simultaneously design an origami gripper to have the following performance target: \( 10 \text{ Hz} < \text{freq} < 40 \text{ Hz} \), heating power \( Q < 0.2 \text{ W} \), and maximum temperature \( T < 200^\circ C \) (Target 1). We can just label those data points that satisfy the performance target to be Class 1 and label the rest of the data as Class 0. Then, by computing the more representative decision rules for Class 1, we can obtain an active origami that satisfy all three performance targets simultaneously. Figure 4B shows the computed design rules for this Target 1. In addition to the rule with the highest F-score, the second-highest rule is also computed and plotted in Fig. 4B. Interestingly, we can see that both rules have high precision and are similar to each other (except for a small difference in the selection of \( t_1 \) and \( t_2 \)). Similarly, we can extend the results to simultaneously design for all four performance indices (Target 2: frequency \( 10 \text{ Hz} < \text{freq} \), heating power \( Q < 0.7 \text{ W} \), and maximum temperature \( T < 500^\circ C \), and \( 0.002 \text{ N/m} < \text{stiffness} k \)), and the result is presented in Fig. 4C. If we compare the computed rules for Target 2 with those for Target 1, we can see that the machine learning method has picked another pattern after adding in the stiffness requirement. This is because Pattern 3 selected for achieving Target 1 only has horizontal creases, which cannot provide additional stiffness compared to Pattern 2 selected for achieving Target 2. This high interpretability of the tree methods helps users to better understand and reason about the behaviors of active origami systems. Moreover, the machine learning method also demonstrates that the significant features for designing this origami gripper can be different for different targets. For example, controlling the values of the gripping arm length \( L_i \) and the location of the first creases (defined by \( Ra \)) is only significant for Target 2.

In general, the design rules with the highest few F-scores obtained from the machine learning method tend to be similar to each other. However, Fig. 3D shows an interesting
result where the top two competing rules have relatively large differences between them. This is because the two rules select different patterns, and the machine learning method thinks both are representative. Rule 1 of Target 3 selects Pattern 1 while Rule 2 selects Pattern 2. These results highlight that by computing multiple rules with high F-scores, it is possible to find distinct design alternatives to achieve flexible designs. Because Target 3 is 2-dimensional, we further extract the data points that fit the rules in the training and testing dataset and plot them in Fig. 3E. The result shows that the extracted data points can trace the design boundary nicely and fill the design boundary with reasonable coverage. This highlights the effectiveness of the proposed methodology.

**Integrating Non-Geometrical Properties into Origami Shape Fitting**

Finally, we want to demonstrate how the proposed method can enable origami shape fitting algorithms to further consider non-geometrical properties of active origami. So far, most research on origami inverse design focuses on kinematical shape fitting (e.g. (12, 13, 15)). Usually, the shape fitting problem can be constructed as an optimization problem, and the error between the target geometry and the origami is minimized given certain constraints (12, 13). However, these existing optimization-based studies cannot consider other non-geometrical properties of the origami systems, which are significant for designing active origami. Moreover, these optimization schemes often leave tremendous flexibility for a designer to vary the origami pattern (e.g. number of panels used or maximum size of panels), without showing which combination may be better. Thus, this section will demonstrate how the proposed method can further enable the origami shape fitting algorithms to consider the interaction between the shape fitting design and non-geometrical behaviors of the origami systems.

As a demonstration, we implement our method on top of an existing shape fitting approach introduced in (12), where an analytical solution was derived to build Miura-origami strips to fit arbitrary planar curves. Figure 5A shows how this shape fitting method can generate different origami strips to fit a target planar curve. In this method, the target curves are first separated into different sections and the number of segments is measured by \( m \). Then a planar origami strip geometry is generated by setting the offset length \( l_o \) of the center node and the width of the strip \( W_s \). Finally, the 3D origami is created by extruding the planar geometry to form the Miura geometry with an extrusion \( l_e \). Figure 5A also shows the shape fitting results for three different curves. As can be seen, there is great flexibility in selecting these parameters for shape fitting and the selection should depend on which combination can give the more desirable origami performances (which tend to be non-kinematical). Suppose our target is to build an origami structure that can achieve a given stiffness performance while fitting a half-circle arch shape, how should we select these shape fitting parameters? The proposed machine learning based method will be able to answer questions like this.

Without loss of generality, we will focus on designing a Miura-origami half-circle arch with a 2m radius. The origami arch database is generated by randomly picking shape fitting design features to build different arch geometries. These variables include the number of segments \( m \), the offset length \( l_o \), the strip width \( W_s \), and the extrude dimension \( l_e \). Other design features such as the thicknesses of panels \( (t_p) \) and creases \( (t_c) \) and the width of creases \( (W) \) are also included because they affects the stiffness of the arch. Based on the random set of design features, we compute and record the responses of the origami using the SWOMPS simulation package, and the database consists of 3000 data samples. The performance indices include the stiffness in X-direction \( (k_x) \) and Z-direction \( (k_z) \), the error of shape fitting
(e), and whether the structure will snap under a 5 N load applied vertically ($S_e$). More details on the simulation setup are available in the supplementary material section S2.

With the database established, we apply the interpretable machine learning method to analyze the database. Because the proposed machine learning based design method can handle a mixture of categorical and continuous variables, it can consider the integer variables used in shape fitting algorithm. Moreover, because the method can also tackle multi-objectives, we can design for the shape-fitting errors and the stiffness performance of the origami structure simultaneously. Target 1 on Fig. 5C represents a target with stricter stiffness requirement but a less strict shape fitting target while Target 2 has a more relaxed stiffness requirement but a stricter error target. Both targets contain about 5% of the total data, so they are comparable. Figure 5C shows the computed rules for the two targets and they both have high precision. The result demonstrates that the machine learning method can produce different design rules to accommodate the interactions between shape fitting errors and mechanical performances. More significantly, the proposed machine learning based method is not tied to specific origami patterns or shape fitting methods. Thus, the proposed methodology can be combined with other origami shape fitting approaches (such as those in (12, 13, 15)), enabling them to design for non-geometrical behaviors of origami.

Finally, we show that the proposed method can design origami systems with complex mechanical behaviors such as bistability and multi-stability (40, 41). Target 3 in Fig. 5C shows the computed design rules for an origami arch to exhibit a snap-through behavior under an applied 5N load. Unlike designing for a stiff arch with small fitting errors, the machine learning method shows that introducing more segments (larger $m$) into the origami strip can give rise to the snapping behavior. Moreover, it is necessary to have a low panel thickness and a low crease thickness so that the origami is more likely to snap. The testing precision of this design rule is high, indicating that the design rule is reliable and accurate.

**Discussion**

This work introduces an inverse-design method for origami systems using interpretable machine learning. First, the origami database is populated using randomly generated origami design features and the performances are calculated using origami simulation methods. Next multiple decision trees are trained to learn how to pick design features so that an origami structure can achieve the target performance. Finally, the origami design rules are computed by backtracking the splitting criteria of tree branches and selecting the one with the highest F-score. From a machine learning point of view, the proposed method uses the F-score to select the better tree branches from the rest, which is similar to selecting the better classifiers from a number of trained classifiers using the F-score. Focusing on a single branch in the decision tree can produce sparser design rules and can help us compute the inverse design relationship. Tree methods are suitable for designing origami because they can capture the categorical features needed to represent different origami patterns and can handle multi-objective problems for studying active origami with multi-physics performances.

There are a number of benefits of using data science and interpretable machine learning based approaches for origami design when compared to using optimization-based strategies. First, the generated database can be reused to compute new rules for different targets, and this advantage is demonstrated in all three design examples presented in this work. Second, the method can simultaneously analyze the significance of different design features for a given design target, which is not provided in optimization-based design methods. If a design feature is significant for achieving a given performance target, a relatively tight threshold
of that feature will be computed. Third, the proposed method can demonstrate the complex interaction between continuous variables and categorical variables. This is necessary for designing origami systems because comparing different origami patterns will naturally introduce categorical variables. These categorical data are difficult to capture using the continuous optimization-based design method. In addition, we show that the proposed method can handle multi-objective design targets from active origami systems with multi-physics behaviors. Finally, we demonstrate that the proposed method can extend existing origami shape fitting algorithms to further design for non-geometrical performances of origami structures. This is important because origami engineering needs to focus more on developing structures with superior performances rather than solely folding 3D geometries.

In summary, this work demonstrates how to compute informative design rules of origami systems using interpretable machine learning. A decision tree method based inverse design scheme is developed for active origami systems. Three origami performance databases are built using the origami simulation package SWOMPS for solving three inverse design problems regarding different origami structures. We showed that the proposed method can consider categorical features needed to represent different origami patterns, can handle multi-objective problems from designing origami structures with multi-physics behaviors, and can enhance existing origami shape fitting algorithms to further consider non-geometrical performances of active origami. We envision that the proposed methodology can be used for designing active origami systems with superior performances for various applications in biomedical devices, soft robotics, metamaterials, deployable structures, and many more.

Materials and Methods

Origami simulation:
- This work uses an open-access origami simulation called SWOMPS (34). This origami simulator uses a bar and hinge model to represent the geometry of origami systems, which is a common simulation technique for origami structures. In addition, this simulator can explicitly model compliant origami creases (folds with distributed width) which makes it suitable for simulating the behaviors of active origami structures (35). The simulator package integrates a state-of-the-art multi-physics model to capture the electro-thermal actuation (36) in active origami assemblages. The implementation codes for building origami databases can be found in the supplementary materials or on the GitHub page: https://github.com/zzhuyii/GenerateOrigamiDataSet. A more detailed introduction of the underlying origami simulation method and how it is applied in each design example can be found in the supplementary text sections S1 and S2.

Machine learning method implementation:
- This work uses an open-access package sklearn (31) to implement the decision tree machine learning method. When training decision trees, entropy-based criterion is used to identify the best splitting rules at branch nodes. Because the target class tend to contain only a small number of data (5% to 10%), the balanced class weight is used to tackle the imbalanced dataset. The results computed in the main article is accomplished using the following hyperparameter: the cost-complexity-pruning alpha value is 0.001, the maximum depth of trees is 20, and the number of training trees is 100. The details of the hold-out testing can be found in the supplementary text section S3. All analyzing codes and all three datasets are provided in the supplementary materials or they can be found on GitHub: https://github.com/zzhuyii/TreeForOrigami.
References

1. K. Kuribayashi, K. Tsuchiya, Z. You, D. Tomus, M. Umemoto, T. Ito and M. Sasaki, Self-deployable origami stent grafts as a biomedical application of Ni-rich TiNi shape memory alloy foil, *Mater. Sci. Eng. A*, **419**, 131-137, (2006).

2. T. G. Leong, C. L. Randall, B. R. Benson, N. Bassik, G. M. Stern and D. H. Gracias, Tetherless thermobiochemically actuated microgrippers, *Proc. Natl. Acad. Sci.*, **106**, 703-708, (2009).

3. J. Na, A. A. Evans, J. Bae, M. C. Chiappeelli, C. D. Santangelo, R. J. Lang, T. C. Hull and R. C. Hayward, Programming Reversibly Self-Folding Origami with Micropatterned Photo-Crosslinkable Polymer Trilayers, *Adv. Mater.*, **27**, 79-85, (2015).

4. Q. Liu, W. Wang, M. F. Reynolds, M. C. Cao, M. Z. Miskin, T. A. Arias, D. A. Muller, P. L. McEuen and I. Cohen, Micrometer-sized electrically programmable shape-memory actuators for low-power microrobotics, *Sci. Robot.*, **6**, eabe6663, (2026).

5. Y. Zhu, M. Birla, K. R. Oldham and E. T. Filipov, Elastically and Plasticlly Foldable Electrothermal Micro-Origami for Controllable and Rapid Shape Morphing, *Adv. Funct. Mater.*, **30**, 20003741, (2020).

6. A. Biswas, C. L. Zekios and S. V. Georgakopoulos, Transforming single-band static FSS to dual-band dynamic FSS using origami, *Sci. Rep.*, **10**, 13884, (2020).

7. K. Fuchi, J. Tang, B. Crowgey, A. R. Diaz, R. E. J. and R. O. Ouedraogo, Origami Tunable Frequency Selective Surfaces, *IEEE Antenn. Wirel. Prop. Lett.*, **11**, 473-475, (2012).

8. M. Schenk and S. D. Guest, Geometry of Miura-folded metamaterials, *PNAS*, **110**, 3276-3281, (2013).

9. E. Boatti, N. Vasios and K. Bertoldi, Origami Metamaterials for Tunable Thermal Expansion, *Advanced Materials*, **29**, 1700360, (2017).

10. S. A. Zirbel, B. P. Trease, M. W. Thomson, R. J. Lang, S. P. Magleby and L. H. Howell, "HanaFlex: a large solar array for space," *Proc. SPIE 9467*, *Micro- and Nanotechnology Sensors, Systems, and Applications VII, 94671C*, 22 May 2015.

11. A.-S. Kaddour, C. A. Velez, M. Hamza, N. C. Brown, S. P. Magleby, L. L. Howell and S. V. Georgakopoulos, A Foldable and Reconfigurable Monolithic Reflectarray for Space Applications, *IEEE Access*, **8**, 219355-219366, (2020).

12. L. H. Dudte, E. Vouga, T. Tachi and L. Mahadevan, Programming curvature using origami tessellations, *nature materials*, **15**, 583-588, (2016).

13. X. Dang, F. Feng, P. Plucinsky, R. D. James, H. Duan and J. Wang, Inverse design of deployable origami structures that approximate a general surface, *Int. J. Solids. Struct.*, **234-235**, 111224, (2022).

14. Y. Chen, L. Fan, Y. Bai, J. Feng and P. Sareh, Assigning mountain-valley fold lines of flat-foldable origami patterns based on graph theory and mixed-integer linear programming, *Computers and Structures*, **239**, 106328, (2020).

15. X. Zhou, H. Wang and Z. You, Design of three-dimensional origami structures based on a vertex approach, *Proc. R. Soc. A.*, **471**, 20150407, (2015).

16. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborova, Machine learning and the physical sciences, *Reviews of Modern Physics*, **91**, 045002, (2019).

17. J. D. Spiegeleer, D. B. Madan, S. Reyners and W. Schoutens, Machine learning for quantitative finance: fast derivative pricing, hedging and fitting, *Quantitative Finance*, **18**, 10, 1635-1645, (2018).

18. I. Kononenko, Machine learning for medical diagnosis: history, state of the art, and perspective, *Artificial Intelligence in Medicine*, **23**, 89-109, (2001).

19. O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danhelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, Budden, D. Budden, Y. Sulsy, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wunsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, Silver David and D. Silver, Grandmaster level in StarCraft II using multi-agent reinforcement learning, *Nature*, **575**, 350-354, (2019).

20. Y. LeCun, Y. Bengio and G. Hinton, Deep Learning, *Nature*, **521**, 436-444, (2015).

21. F. Wang and C. Rudin, "Falling Rule Lists," *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, San Diego, CA, USA, 1013-1022, (2015).

22. Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," *Machine Learning: Proceedings of the Thirteenth International Conference*, (1996).
23. L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and regression trees, Monterey CA: Wadsworth and Brooks, 1984.

24. H. Yasuda, K. Yamaguchi, Y. Miyazawa, R. Wiebe, J. R. Raney and J. Yang, Data-driven prediction and analysis of chaotic origami dynamics, *Commun. Phys.*, 3, 168, (2020).

25. A. Namiki and S. Yokosawa, "Robotic Origami Folding with Dynamic Motion Primitives," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sept 28 - Oct 2, 2015. Hamburg, Germany*, (2015).

26. J. R. Quinlan, Induction of decision trees, *Mach. Learn.*, 1, 81-106, (1986).

27. H. Yasuda and J. Yang, Reentrant Origami-Based Metamaterials with Negative Poisson's Ratio and Bistability, *Physical Review Letters*, 114, 185502, (2015).

28. J. L. Silverberg, J.-H. Na, A. A. Evans, B. Liu, T. C. Hull, C. D. Santangelo, R. J. Lang, R. C. Hayward and I. Cohen, Origami structures with a critical transition to bistability arising from hidden degrees of freedom, *Nat. Mater.*, 14, 389-393, (2015).

29. C. Rudin, Stop explaining black box machine learning models for high stakes decision and use interpretable models instead, *Nat. Mach. Intell.*, 1, 206-215, (2019).

30. W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl and Y. Bin, Definitions, Methods, and Applications in Interpretable Machine Learning, *PNAS*, 116, 22071-22080, (2019).

31. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, a. V. Niculae, a. P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt and G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project, *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108-122, (2013).

32. J. Lin, C. Zhong, D. H, C. Rudin and M. Seltzer, "Generalized and Scalable Optimal Sparse Decision Trees," *Proceedings of the 37th International Conference on Machine Learning*, (2020).

33. D. Bertsimas and J. Dunn, Optimal Classification Trees, *Mach. Learn.*, 106, 1039-1082, (2017).

34. Y. Zhu and E. T. Filipov, Sequentially Working Origami Multi-Physics Simulator (SWOMPS): A Versatile Implementation, *IDETC/CIE 2021, August 17-19, Online*, DETC2021-68042, (2021).

35. Y. Zhu and E. T. Filipov, A bar and hinge model for simulating bistability in origami structures with compliant creases, *J. Mech. Robot.*, 12, 021110, (2020).

36. Y. Zhu and E. T. Filipov, Rapid Multi-Physics Simulation for Electro-Thermal Origami Systems, *Int. J. Mech. Sci.*, 202-203, 106537, (2021).

37. A. Tharwat, Classification assessment methods, *NEJE*, 17, 168-192, (2020).

38. S. Felton, M. Tolley, E. Demaine, D. Rus and R. Wood, A method for building self-folding machines, *Science*, 345, 644-646, (2014).

39. B. An, S. Miyashita, O. Aaron, M. T. Tolley, M. L. Demaine, E. D. Demaine, R. J. Wood and D. Rus, An End-to-End Approach to Self-Folding Origami Structures, *IEEE Trans. Robot.*, 34, 1409-1424, (2018).

40. J. A. Faber, A. F. Arrieta and A. R. Studart, Bioinspired spring origami, *Science*, 359, 1386-1391, (2018).

41. J. L. Silverberg, A. A. Evans, L. McLeod, R. C. Hayward, T. Hull, C. D. Santangelo and I. Cohen, Using origami design principles to fold reprogrammable mechanical metamaterials, *Science*, 345, 647-650, (2014).

**Acknowledgments**

**Funding:**

National Science Foundation Grant #2054148
Rackham Research Grant, University of Michigan, Ann Arbor

**Author contributions:**

Conceptualization: YZ
Methodology: YZ
Investigation: YZ
Visualization: YZ
Supervision: ETF
Writing—original draft: YZ
Writing—review & editing: YZ, ETF
Competing interests:
The authors declare they have no competing interests.

Data and materials availability: All data and code used in the analyses are available in the main text or the supplementary material.
Fig. 1. Compute origami design rules with interpretable machine learning. The relationship between origami design features (left) and performances of origami systems (right) can be thought of as a black-box nonlinear function $f$ with no analytical solution. This work shows that it is possible to train an interpretable machine learning method (the bottom decision tree method) to uncover the underlying structures of this black-box nonlinear function $f$, so that we can build human understandable design rules for the inverse design problem (solve for $f^{-1}$ conceptually).
Fig. 2. Compute interpretable design rules for origami assemblages. A. Set up the design problem of a single unit Miura-origami pattern. B. Populate a database of origami features and performances using origami simulation method. Label the data points based on whether they meet the target performance or not. C. Train a number of decision trees to classify the database. D. Gather design rules by collecting the splitting criteria in each branch of the decision tree (follow the gray lines in the sub-figure (c)). E. Select the more representative rules using the precision, the recall, and the number of data points satisfying the rules. F. Choose the final design rule with the highest F-score. In this sub-figure, the design rule is indicated using the darker boxes in the box chart. Training and testing precisions and one sample Miura origami are provided for reference.
Fig. 3. Compare different origami patterns using the interpretable machine learning.

A. Problem setup for the design of origami canopies. Two origami patterns are used to build the canopy. B. Design for the axial stiffness of the canopy at 60% extension. C. Design for the bending stiffness at 90% extension. D. Sample designs for the selected four zones and the stiffness performance of each design.
**Fig. 4.** Design of an origami gripper with multi-physics and multi-objectives. A. Problem setup for building active origami grippers with three different patterns. B-D. The top two computed design rules for the design Targets 1 to 3. The lighter boxes indicate the full ranges of different design features, and the shaded boxes indicate the selected design range computed by machine learning. E. The top two design rules for Target 3 give distinct design alternatives with data point distributions that both meet the given target.
Fig. 5. Integrating Interpretive Machine Learning Design with Shape Fitting. A. Miura-ori design can be modified to fit arbitrary shapes of curves. Geometrical design features for fitting curves include the number of units $m$, the offset length $l_o$, the width of strip $W_s$, and the extrude dimension $l_e$. B. Four performance indices are studied, and they are the error of fitting $e$, the stiffness in X and Z directions $k_x$ and $k_z$, and a binary variable $S_z$ that indicates if the structure snapped under a 5N load in the Z direction. C. Computed rules for the three different targets and the corresponding sample designs.