Structural and Multidisciplinary Optimization (2025) 68:175
https://doi.org/10.1007/s00158-025-04104-w

RESEARCH PAPER

Interpretable SHAP-bounded Bayesian optimization for underwater acoustic metamaterial coating design

Hansani Weeratunge[1] · Dominic M. Robe[2] · Elnaz Hajizadeh[2]

Received: 2 March 2025 / Revised: 9 July 2025 / Accepted: 28 July 2025 / Published online: 8 September 2025

Abstract

We present an interpretability-informed Bayesian optimization framework for the inverse design of underwater acoustic coatings composed of polyurethane elastomers with embedded metamaterial features. A data-driven model was used to capture the relationship between acoustic performance, specifically, sound absorption and the corresponding geometrical design variables. To interpret these relationships, we applied SHapley Additive exPlanations (SHAP), enabling the identification of key parameters influencing the objective function and providing both global and local insights into their effects. The insights from the SHAP analysis were used to automatically refine the bounds of the design space, guiding the optimization process toward more promising regions. This approach was tested on two polyurethane materials with different hardness levels and yielded improved optimal designs compared to standard Bayesian optimization without increasing the number of simulations. This work underscores the effectiveness of combining interpretability techniques with optimization for the efficient and cost-effective design of underwater acoustic metamaterials under strict computational constraints and can be generalized towards other materials and engineering optimization problems.

Keywords Interpretable machine learning · SHAP · Metamaterials · Underwater acoustic coatings

1 Introduction

Advances in acoustic materials technologies play a crucial role in reducing the acoustic signature and likelihood of detecting naval platforms by adversaries, ensuring their survival and effective operation. The emergence of advanced sonar technologies demands rapid development of novel materials solutions with effective sound attenuation over an extended range of frequencies, particularly the challenging low frequencies corresponding to long wavelengths. This requirement makes state-of-the-art coating technologies impractically thick and compromises their resistance to hydrostatic pressure resulting from operations at depth.

Acoustic metamaterials - an array of voids embedded in an inherently damping matrix material such as rubber or polyurethane elastomer (PU) (Robe et al. 2025; Shireen et al. 2023) have emerged as a promising class of materials to obtain sub-wavelength attenuation (Sharma et al. 2017). To enhance their acoustic performance over low frequency range (1kHz-10kHz), the design of these coatings often involves embedding resonant inclusions, such as disk cavities (Calvo et al. 2015), cylindrical voids (Ranjbar and Bayer 2024), hard spheres (Meng et al. 2012), and combination of both voids and hard inclusions (Sharma et al. 2019). However, the effectiveness of these materials depends on the availability of an elastomeric matrix with fine-tuned frequency-dependent viscoelastic behavior as well as optimized size and layout of voids across the matrix.

Previously, we developed experimentally validated finite element models (FEMs) to simulate the acoustic behavior of polyurethane coatings with embedded voids (Weeratunge et al. 2022), where for the first time, we accounted for the frequency-dependent viscoelastic properties of the polyurethane matrix materials. By employing FEM, we were able to

Responsible Editor: Lei Wang.

✉ Elnaz Hajizadeh
ellie.hajizadeh@unimelb.edu.au

[1] Department of Mechanical Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

[2] Soft Matter Informatics Research Group, Department of Mechanical Engineering, Faculty of Engineering And Information Technology, Soft Matter Informatics Research Group, The University Of Melbourne, Melbourne, Australia

model the complex interactions within the coatings and gain valuable insights into the effects of various design parameters, including void size, matrix modulus, and backing materials (Weeratunge et al. 2022). These models demonstrated the significant influence of these parameters on the absorption coefficient, which must be considered in their design.

The design of novel acoustic materials is traditionally relied on "forward" design strategies (Hajizadeh and Garmabi 2008), where designers iteratively search through the parameter space to identify configurations that yield desired material characteristics. This often involves brute-force search guided by domain knowledge, intuition, or trial-and-error. However, as the number of design parameters increases, the combinatorial explosion of possible configurations makes this process inefficient and impractical (Cai et al. 2006).

To address these limitations, optimisation-based "inverse" design strategies have emerged (Abdolahi et al. 2025; Weeratunge et al. 2023). In this approach, designers specify a target performance (ground truth), and the goal is to mathematically search for the optimal set of design parameters that achieve the target. This is typically formulated as a numerical optimization problem, where a forward model is used to compare candidate designs against the desired outcome and guide the search (Weeratunge et al. 2022).

However, using physics-based forward models such as FEM in optimization can be prohibitively expensive due to the large number of time and resource-intensive simulations required in optimization loops. With recent development of Machine Learning (ML), this has motivated the development of ML-based surrogate models to approximate the behavior of complex simulations at a fraction of the computational cost (Panisilvam et al. 2023). By training models on large datasets, surrogate ML models can be used for efficient predictions and further uncover patterns and relationships that may not be evident through traditional analytical methods (Panisilvam et al. 2023; Abdolahi et al. 2025). Among them, the application of Deep Neural Networks (DNN) has become widespread in various fields due to their ability to capture nonlinear dynamics in complex models (Shireen et al. 2022). Their potential has also been demonstrated in the accelerated and accurate prediction of the acoustic performance of materials (Kumar et al. 2023; Fu et al. 2021; Jeon et al. 2020; Weeratunge et al. 2022).

These surrogate models enable scalable and efficient inverse design, especially when integrated with Bayesian Optimisation (BO), an efficient strategy well-suited for resource-intensive problems. BO excels at optimizing "black box" functions with limited data by building a probabilistic surrogate model, typically using a Gaussian process to estimate the objective function. This active learning framework balances the exploration of new regions in the parameter space with the exploitation of previously gathered information using an acquisition function to guide the search. Importantly, BO incorporates uncertainty quantification through its probabilistic model, which helps prioritize areas in the design space that are either uncertain or promising. As more evaluations are conducted, the accuracy of the surrogate model improves, which results in refining the optimization process and reducing the number of simulations needed to locate optimal solutions. This way, we can efficiently navigate the design space, accelerating the convergence to an optimal solution (Liang et al. 2021).

One of the key reasons BO is practically well-suited for such problems is its ability to efficiently handle expensive simulations within a limited budget. Unlike metaheuristic algorithms, which require a large number of evaluations to explore the design space, BO uses probabilistic models to strike an optimal balance between exploration and exploitation. This approach significantly enhances efficiency, making BO a powerful tool for optimization in computationally demanding applications.

A significant challenge with ML models lies in their lack of interpretability, making it difficult to understand how individual input parameters influence model predictions. This limitation can hinder the practical application of these models in fields where transparency is essential for informed decision-making. Although intrinsically interpretable models like linear regression are easy to interpret, they typically lack accuracy, in contrast to more complex, less interpretable models. To address this challenge, recent advances in machine learning have focused on developing techniques that interpret the relationship between input variables and predicted results, thus improving the transparency of "black-box" models. Several tools have been developed to interpret black-box models, including partial dependence plots (PDP), individual conditional expectation (ICE), and local interpretable model-agnostic explanations (LIME). These techniques provide post hoc explanations at the global or local level, breaking down complex models to make their inner workings more comprehensible (Salih et al. 2025). For example, PDPs illustrate the average impact of an input variable on the model's prediction, revealing the relationship between the input and the output, while ICE plots offer similar insights for individual predictions. LIME approximates complex models locally with interpretable models, identifying influential features that drive specific predictions (Palar et al. 2023).

SHapley Additive exPlanations (SHAP) has gained significant attention as an interpretability method due to its strong theoretical foundation in cooperative game theory (Lundberg and Lee 2017). Based on Shapley values, SHAP fairly distributes a model's prediction across its input features, providing a unified measure of feature importance (Khalid et al. 2023). This approach ensures that both local explanations for individual predictions and global insights

into overall model behavior are captured, making SHAP a powerful tool for understanding complex machine learning models (Stender et al. 2021). Moreover, SHAP adheres to essential interpretability properties such as local accuracy and consistency, enabling transparent and reliable explanations of model predictions (Lundberg and Lee 2017).

Despite its capabilities, SHAP remains underutilized in engineering design and optimization, where it could identify the most influential features and clarify how changes in parameter values affect the objective function, allowing a more focused search strategy. This targeted refinement can potentially reduce the number of required simulations, conserving computational resources and accelerating convergence (Vimbi et al. 2024; Salih et al. 2025).

Recent studies have applied SHAP to optimize engineering designs, revealing crucial factors driving material properties and performance (Islam et al. 2024). For example, machine learning techniques combined with SHAP values have been used to assess the impact of architectural characteristics on acoustic performance in educational buildings, highlighting the influence of wall material and room dimensions on sound diffusion and clarity (Tabatabaei Manesh et al. 2024). SHAP was applied to deep neural network models to enhance interpretability by identifying and removing low impact features, thus improving the accuracy of the model and emphasizing significant parameters in acoustic design assessments (Abarghooie et al. 2021).

Dalal et al. (2024) used a method that integrates synthesis, characterization, and machine learning to optimize polymer design for nucleic acid delivery (Hajizadeh and Larson 2017; Xiang et al. 2021; Zhu et al. 2016). By combining parallel experimentation with machine learning, including SHAP and Bayesian optimization, this research demonstrates the utility of these methods individually. Wang et al. (2024) employs ML models to predict the compressive strength of self-compaction concrete, where SHAP analysis identified the impact of different additives on the compressive strength of concrete, potentially reducing the need for extensive experimental tests. SHAP has also found use in metamodeling, to extract insight about a global surrogate function (Abdolahi et al. 2025). By highlighting the parameters with the greatest impact, SHAP can enable designers to prioritize the most critical factors, potentially saving time and resources during material development. These works demonstrate the usefulness of SHAP for material design problems, but none of these integrate SHAP analysis into the optimization loop to assess whether it can accelerate the design process.

We, for the first time, integrate SHAP analysis algorithmically into a Bayesian Optimization routine by modifying the bounds of the search space. We compare the performance of this SHAP-bounded optimization to an off-the-shelf domain reduction algorithm. We examine SHAP values to identify the most influential topological parameters that affect the acoustic performance of polyurethane coatings with metamaterial features. By constraining the design factors with effectively monotonic effect on performance, SHAP is found to accelerate the design process, reducing the need for exhaustive trial-and-error.

The manuscript is organized as follows. Section 2 provides an overview of several key facets of the methodology we have developed. Subsection 2.1 presents a simplified illustrative toy problem that demonstrates the value of SHAP. Subsection 2.2 explains the FEM model developed to simulate the polyurethane (PU) acoustic material system. Subsection 2.3 details our objective function, which quantifies acoustic absorption performance while incorporating practical constraints. Section 2.4 introduces the SHAP-Bounded Bayesian Optimization framework. Subsection 2.4.1 describes the development of a DNN to model the relationship between design variables and acoustic performance, while Subsection 2.4.2 outlines the application of SHAP for model interpretability. Subsection 2.4.3 presents the SHAP-guided parameter space refinement strategy, which iteratively narrows the search space to enhance optimization efficiency. Section 3 discusses the results of our approach, including performance improvements and key insights obtained from the optimization process. Finally, some concluding remarks and suggestions for further research are contained in Section 4.

2 Methodology

This section describes the problem formulation and the inverse design methodology aimed at enhancing the acoustic performance of underwater coatings by optimizing their design variables. The objective is to maximize the sound absorption capabilities of a polyurethane (PU) matrix embedded with voids while ensuring manufacturability.

To build intuition for our proposed interpretability-informed optimization framework, we first present a simplified illustrative toy problem that demonstrates the value of SHAP in guiding the optimization process. This toy example serves to conceptually validate our approach by highlighting how SHAP can identify non-intuitive parameter sensitivities and support data-driven refinement of the design space. The insights gained from this example help motivate the subsequent application of our method to the full-scale acoustic metamaterial design problem.

Following this illustrative example, we detail the main methodology. We first formulate an optimization problem applying BO to explore the design space. We then train a DNN using the same data points collected during the optimization process. This DNN serves as an interpretability tool, allowing us to apply SHAP to analyze the contribution

of each design variable to acoustic absorption. By leveraging SHAP insights, we iteratively refine the bounds of the optimization problem, directing the search toward the most influential regions of the design space. This approach improves optimization efficiency without increasing the number of simulations. The following subsections detail the problem setup, the formulation of the objective function, and the SHAP-guided refinement process that enhances the design of high-performance acoustic coatings.

2.1 Illustrative toy example of SHAP-enhanced optimization

An illustration of the proposed advantage of this approach is presented in Fig. 1. In Fig. 1A a toy 2D objective function is rendered with many local maxima. For completeness, the precise expression for the toy objective is

$$y = x + 2y - xy/8 - x^2/10 - y^2/2 + \sin 5x \sin 5y. \quad (1)$$

The black points indicate 20 evaluations of that function with random parameters and 10 subsequent evaluations guided by classical BO. Fig. 1B shows the result of a SHAP analysis of a model trained on the 30 samples. The conceptual takeaway from this figure is that SHAP picks out the largest scale trends and separates the contributions of each parameter. In Fig. 1C by contrast we see the predictions of a Gaussian Process Regression (GPR) model trained on the same 30 samples. In a rugged landscape, a GPR model can miss local maxima and fail to represent the large-scale trend. The problem is clarified in Fig. 1D where we see that the acquisition function has prioritized several lower local maxima ahead of the best one. The crux of the present work is to exclude from the search space those parameter ranges with strictly negative SHAP values. In this toy example, we would impose $x > 2$ and $y < 3$ on subsequent sample suggestions. There is another algorithmic detail discussed below to avoid over-tightening bounds. The ideal result would be for the majority of the misguided peaks in Expected Improvement (EI) in Fig. 1D to be ignored, allowing BO to more rapidly discover the global maximum. It should be noted that the objective function in Fig. 1 was crafted deliberately to illustrate the plausibility of combining SHAP with BO. Qualitative differences exist between a hilly 2D landscape and a 10D space with high order critical points. To genuinely evaluate the effectiveness of this optimization strategy, we will apply it to a material design problem.

2.2 Model problem

The objective of this study is to optimize the design variables of a polyurethane (PU) matrix slab embedded with voids to maximize its underwater acoustic absorption performance. The model includes circular cross section of two layers of voids within a PU matrix, attached to a steel backing and submerged in water. For a detailed description of the model, including schematic representation, material properties, and parameters readers are referred to our previous publication (Weeratunge et al. 2022).

The design involves ten independent geometrical parameters, including the thickness of the slab, the radius of the voids, and parameters that define their placement, illustrated in Fig. 2A. To accommodate practical manufacturing constraints, the maximum thickness of the slab was limited to 100 mm, which limited the use of two layers of voids. Additionally, a minimum distance of 1 cm was maintained between the edges of the coating and the voids to ensure feasibility in fabrication. The detailed upper and lower bounds for these design variables are provided in Table 1.

The performance of a particular layout varies nonlinearly with the frequency-dependent rheology of the PU matrix, so different PU materials will require separate optimization

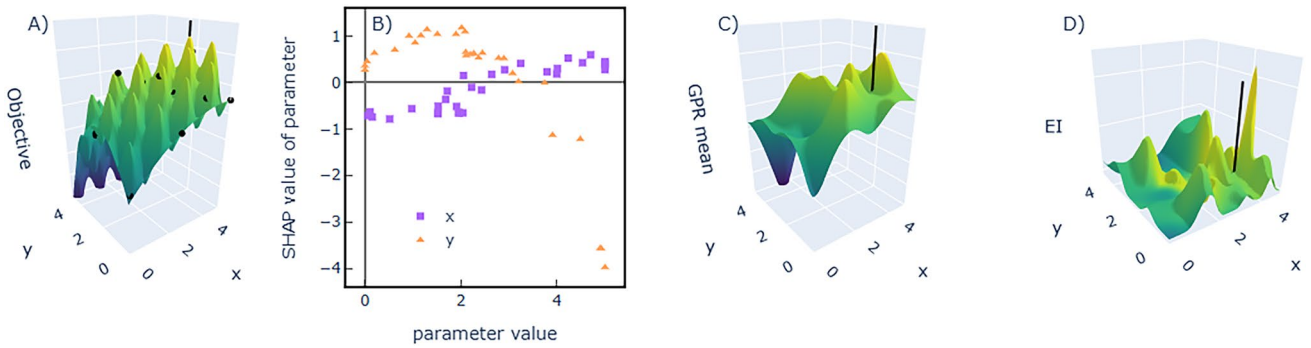

Fig. 1 Illustration of the complementary information provided by SHAP and BO. **A** An objective function with many local optima and a large-scale trend. Black points represent function evaluations during an optimization. The vertical black line indicates the global maximum. **B** The SHAP values for the two parameters X and Y at the evaluated points. **C** A Gaussian Process Regression (GPR) model fit to the evaluated points. **D** The Expected Improvement (EI) calculated from the GPR model

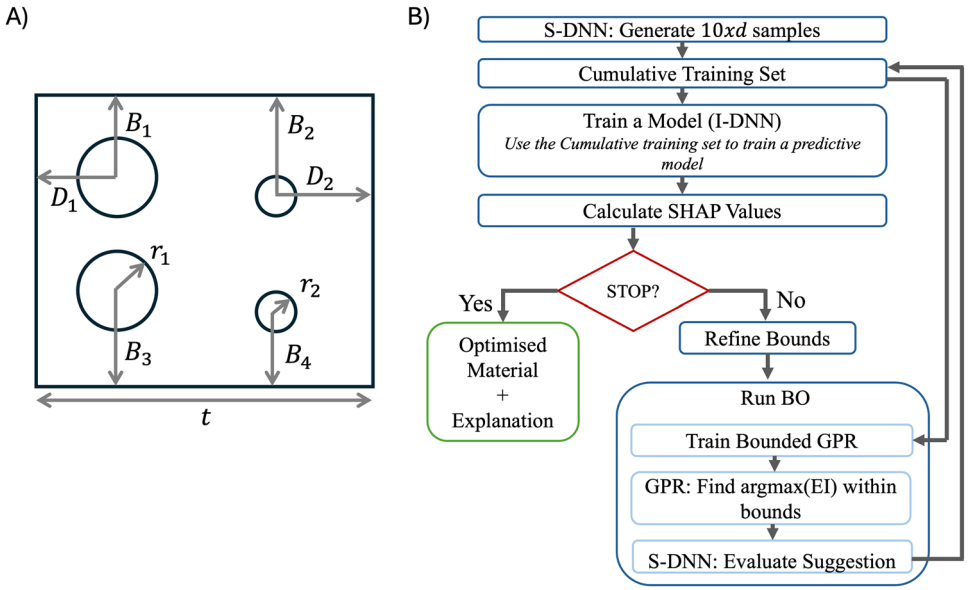

Fig. 2 The design of the current study. **A** the structural parameters for the metamaterial design being optimized. r_1 and r_2 are the radii of the first and second layers of voids. D_1 and D_2 are the horizontal distances of the layers from their respective edges. B_1, B_2, B_3, and B_4 are the vertical distances of each void from their respective edges. **B** Overview of the proposed optimization methodology. The S-DNN is used as a surrogate to the FEM model for efficient evaluation of the objective function. An I-DNN is trained on the cumulative dataset and used exclusively for SHAP-based interpretability to identify influential design variables. If the sampling budget is not exhausted, the design space boundaries are refined based on SHAP insights. A BO loop is then initiated to iteratively sample promising regions of the refined design space. A GPR model is trained within this refined region, and the argmax of the Expected Improvement (EI) acquisition function is selected and added to the dataset. This iterative BO process continues until the sampling budget is reached

Table 1 Lower and upper bounds of the values of design variables

Design Variable	Lower Bound/(*mm*)	Upper Bound/(*mm*)
r_1, r_2	2	15
D_1, D_2	10	80
B_1, B_2, B_3, B_4	10	80
h	30	100
t	30	100

runs. Two commercial PU materials with different rheological properties, summarized as a "hardness" rating have been previously characterized so that their rheological data could be used as FEM constitutive equations, as discussed in Weeratunge et al. (2022). Here we optimize two materials with hardness 80 and 90, dubbed PU80 and PU90.

In this study, we build upon the optimization framework developed in our previous work (Weeratunge et al. 2022), where a computationally expensive FEM was used to model the acoustic performance. To address the high computational cost, we previously developed a surrogate DNN model, referred to as the S-DNN, which achieved a Pearson correlation coefficient of 0.999 with respect to FEM predictions. In that study, we also validated the S-DNN by comparing the absorption coefficients predicted by the S-DNN and FEM, demonstrating excellent agreement. Given this high accuracy, the objective function values calculated using the surrogate DNN (S-DNN) are nearly identical to those obtained from FEM. Therefore, in the present work, we use the S-DNN to calculate the objective function which drastically reduces computational time. This substitution enables us to run multiple independent optimization routines and to average the results, allowing for a robust assessment of the proposed algorithm's consistency and reliability.

2.3 Objective function

Acoustic attenuation depends non-linearly on the frequency of incident sound, so the optimization of acoustic materials is generally a multi-objective problem with trade-offs in the attenuation at different frequencies. We construct an aggregate objective function using a weighted average of the absorption at various frequencies, weighted toward lower frequencies.

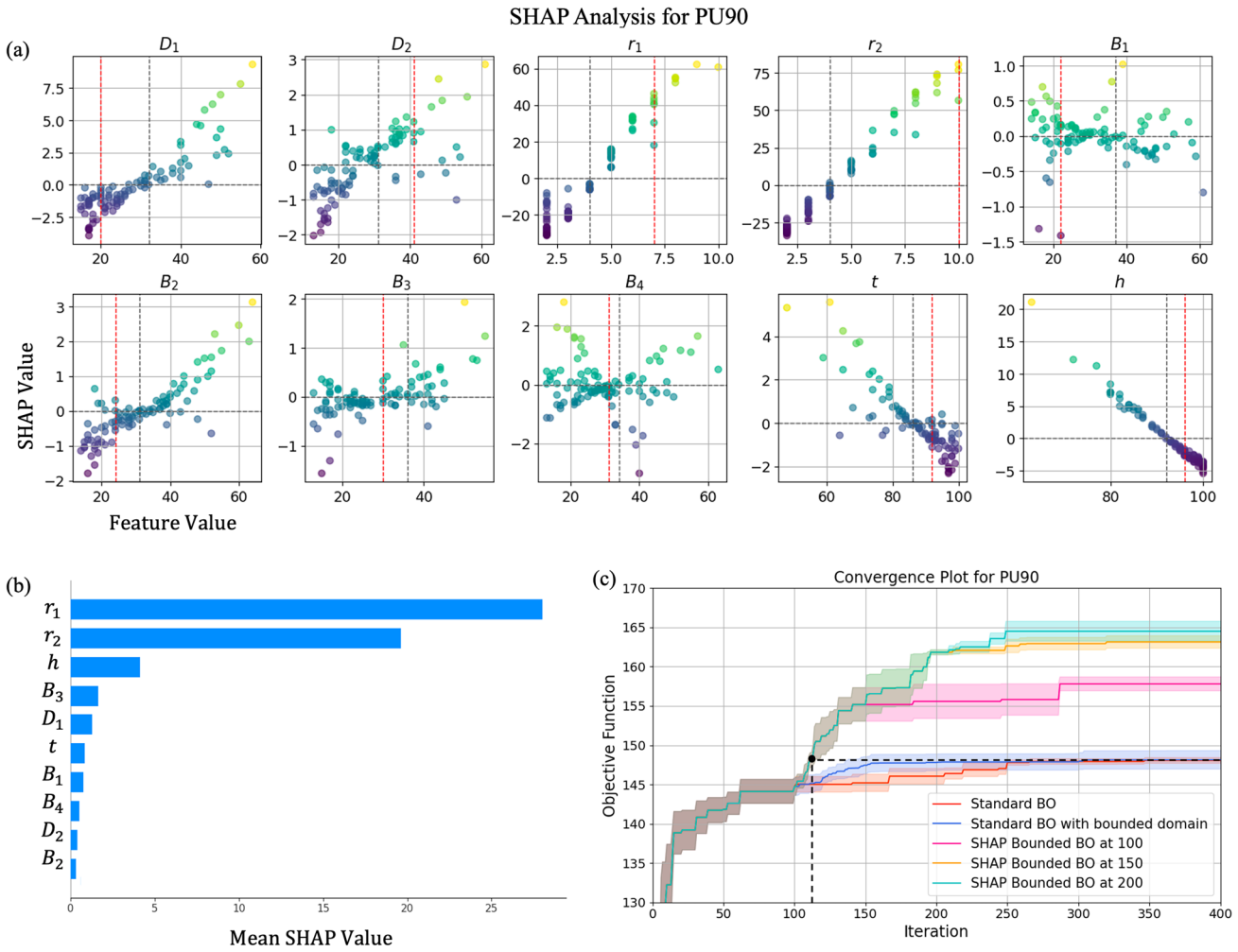

Fig. 3 Results of PU90 **a** Variation of the SHAP value for each feature. Lighter colors emphasize higher SHAP values. Red vertical dashed lines indicate the parameter values for the best-so-far layout. Black vertical dashed lines indicate the crossover from regions with negative to positive SHAP values. **b** Mean SHAP values, ranking features by their overall impact. **c** Convergence plot illustrating the mean and standard deviation of the best objective function value at each iteration. The black point and broken line indicate the number of iterations at which SHAP-informed BO surpassed standard BO's optimal solution after 400 iterations

$$\text{Objective function} = \max\left(\sum_{i=1}^{N} w_i a_i - p\right), \qquad w_i = \frac{N+1-i}{N}, \tag{2}$$

where w_i and a_i represent the weight and absorption coefficient of the ith frequency, respectively, and N is the number of frequencies. A total of 1000 frequency points were considered, ranging from 10 Hz to 10 kHz with equal intervals of 10 Hz. The penalty term p ensures manufacturability by penalizing infeasible solutions. The initial bounds of the design variables, based on manufacturing limitations, are provided in Table 1.

2.4 SHAP-bounded Bayesian optimization

This section presents the methodology employed to optimize the acoustic performance of polyurethane coatings with embedded voids by integrating Bayesian Optimization (BO), DNN, and SHAP. The objective is to efficiently explore the high-dimensional design space and uncover the most influential design parameters. The approach begins by training a DNN (I-DNN) model using the data points sampled during the BO process (Calculated using the S-DNN). The I-DNN serves as an interpretable model that approximates the mapping between design variables and objective values observed during optimization. SHAP is applied to the I-DNN to extract both global and local insights into how individual design parameters influence the objective function. This interpretability enables focused refinement of the

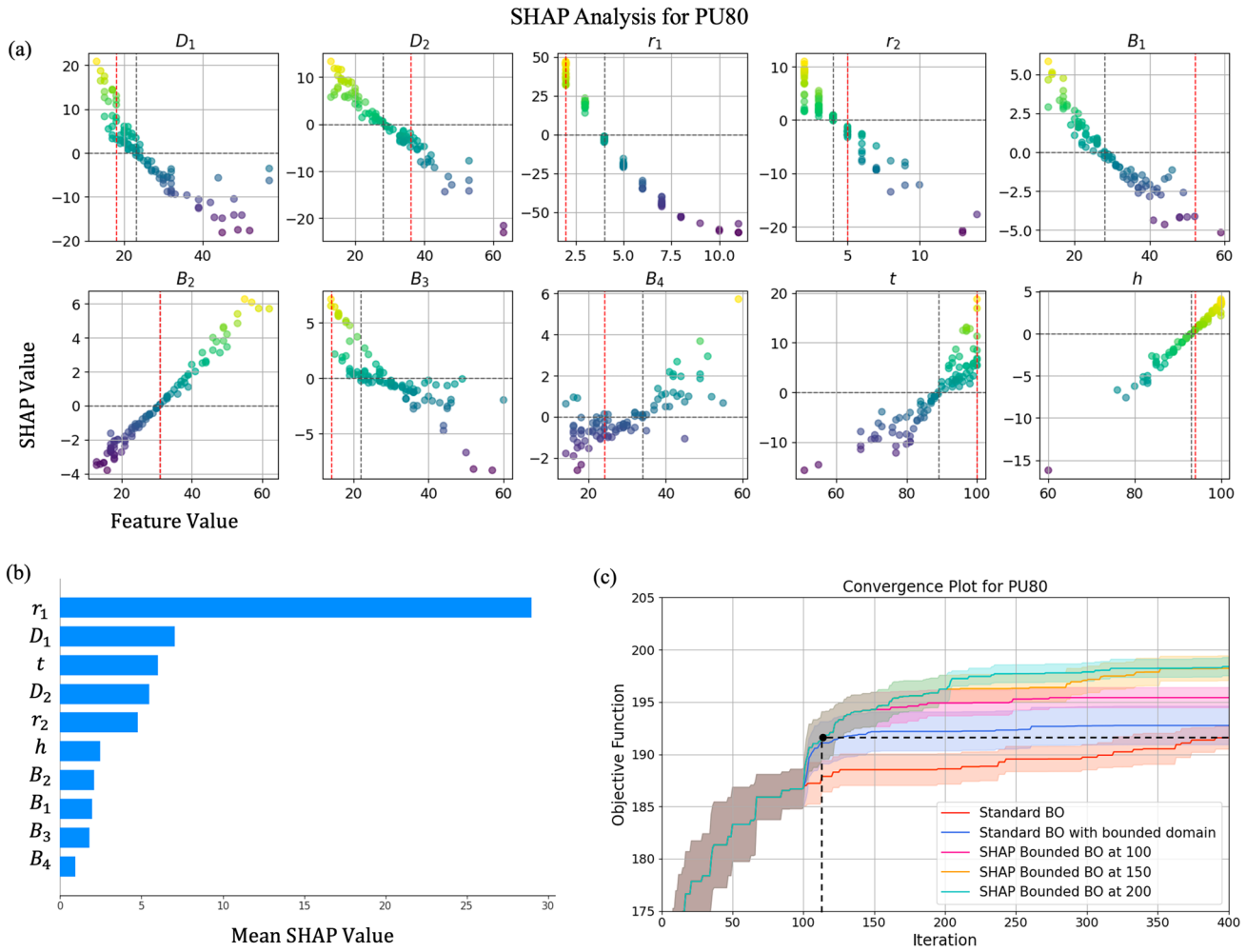

Fig. 4 Results of PU80 **a** Variation of the SHAP value for each feature. Lighter colors emphasize higher SHAP values. Red vertical dashed lines indicate the parameter values for the best-so-far layout. Black vertical dashed lines indicate the crossover from regions with negative to positive SHAP values. **b** Mean SHAP values, ranking features by their overall impact. **c** Convergence plot illustrating the mean and standard deviation of the best objective function value at each iteration. The black point and broken line indicate the number of iterations at which SHAP-informed BO surpassed standard BO's optimal solution after 400 iterations

parameter space, enhancing the efficiency of the optimization process. The methodology integrates these techniques into a coherent framework that accelerates optimization and ensures transparency and clarity in decision-making.

2.4.1 Interpretable deep neural network (I-DNN) for mapping design variables to acoustic performance

This study employed DNNs to model the relationship between the objective function and the geometrical design variables of underwater acoustic coatings with embedded voids using the data sampled during BO. The primary goal was to establish an interpretable mapping between design variables and their influence on the objective function. This relationship forms the foundation for the SHAP analysis in the next section.

Although GPR is commonly used within BO due to its built-in uncertainty modeling, which is valuable for guiding exploration, it has notable limitations in the context of interpretability. Specifically, GPR lacks the ability to provide accurate local explanations of model predictions, which is essential for interpretability. Additionally, GPR often struggles to represent complex, and highly nonlinear relationships. To overcome these challenges, we employed a DNN trained on the data generated during the optimization process, referred to here as the Interpretable DNN (I-DNN). The I-DNN provides improved predictive accuracy and is well-suited for integration with SHAP to produce both global and local explanations of variable influence.

The architecture of the I-DNN consists of fully connected layers with ReLU activation functions and is trained using the mean squared error loss. The network takes ten geometrical design variables as inputs and predicts the objective function, i.e., the weighted sum of the absorption coefficient as the output. Using I-DNN in combination with SHAP, this study enables a deeper understanding of the design space and supports informed parameter space refinement beyond what GPR alone can offer.

2.4.2 Interpretable machine learning with SHAP

For DNNs, SHAP employs a Deep SHAP, which combines Shapley values with DeepLIFT (Deep Learning Important FeaTures) (Shrikumar et al. 2017). DeepLIFT is a gradient-based approach that assigns scores to features based on their contributions. Deep SHAP uses these scores to approximate Shapley values efficiently, making it feasible to apply SHAP to complex and non-linear neural networks. This allows researchers to determine the importance of each input feature, providing insight into the behavior of the neural network, and identifying critical factors that influence predictions (Linardatos et al. 2021).

Visual tools like summary plots and dependence plots further enhance interpretability by showing the distribution of SHAP values for each feature and the relationships between features and model outputs.

This study employs SHAP to interpret the I-DNN model that predicts the objective function. The interpretability provided by SHAP is crucial for high-dimensional design spaces such as ours, enabling a deeper understanding of how geometric parameters, such as void radii, influence sound absorption. Using SHAP, we improve the transparency of the model and identify critical parameters, facilitating informed decisions for narrowing the search space for optimization.

2.4.3 SHAP-guided parameter space refinement

In this study, a simulation budget of 400 evaluations was used. The optimization process, illustrated in Fig. 2B and detailed in Algorithm 1, begins by initializing the full range of possible values for all design parameters, representing the complete parameter space in which the optimization will take place.

To explore this parameter space, samples of 10 times the number of design parameters (10 * D, where D is the number of design variables) is randomly generated. These samples serve as initial design configurations and are used to train a surrogate model, such as a DNN or another appropriate model, capable of predicting the objective function based on the design parameters.

Following the model training, SHAP is applied to analyze the predictions of the trained model. SHAP helps to identify the contribution of each design parameter to the model's output, revealing the top most influential features: those design variables that significantly impact the optimization objective. This analysis provides valuable insights into how different design parameters interact and affect the overall performance.

Using the insights from SHAP, the bounds of the design parameters are adjusted to narrow down the design search space. For each of the most influential features, if the best solution found so far has a positive Shapley value for that feature, the bounds are adjusted accordingly. Specifically:

- If all SHAP values to the right of the best-so-far sample are positive, the lower bound is updated to be 10% lower than the best value
- If all SHAP values to the left of the best-so-far sample are positive, the upper bound is updated to be 10% higher than the best value

This ensures that the optimization process focuses on the most promising regions of the design space while respecting the original parameter constraints. Parameters with less clear trends or minimal impact on the objective function retain their initial bounds.

After these adjustments, BO is performed within the newly defined, more focused search space. This process is repeated iteratively. In our problem, the SHAP analysis and the adjustment of bounds were done every 50th iteration starting from the 100th, to ensure that the search space remained focused on the most promising regions. However, for this specific problem, no significant changes in the bounds were observed after the 250th iteration. The iterative process continues until the simulation budget is exhausted or the optimization meets a predefined tolerance level. By leveraging SHAP for ongoing refinements and adjusting the bounds, the optimization process is able to focus on the most relevant regions of the parameter space, making more efficient use of the remaining 300 evaluations and improving the overall performance of the optimization.

Algorithm 1 SHAP-bounded Bayesian Optimization Algorithm

Initialize Parameter Space: Start with the full allowed parameter space for all design variables.
Generate Initial Samples: Randomly generate a set of $10 \times D$ samples, where D is the number of design parameters.
Train Surrogate Model: Use the generated samples to train a predictive model (e.g., a Deep Neural Network or another suitable surrogate model).
Apply SHAP Analysis: Apply SHAP to interpret the model and evaluate the contribution of each feature to the objective function.
Identify the top 6 features that most significantly influence the model's predictions.
Adjust Parameter Bounds Based on SHAP Values
for each of the top 6 features **do**
    if the best-so-far sample has a positive Shapley value for this feature **then**
        if all SHAP values to the right are positive **then**
            Set the new lower bound to `best value - 10% of best value`.
        else if all SHAP values to the left are positive **then**
            Set the new upper bound to `best value + 10% of best value)`.
        else
            No adjustment needed for this feature.
        end if
    end if
end for
Bayesian Optimization (BO) within Refined Bounds: Perform Bayesian Optimization (BO) within the newly defined parameter bounds, filtering samples inside new bounds.
Repeat Iteration: Return to step 3 and repeat the process until the simulation budget is exhausted or the optimization reaches a predefined tolerance level.

3 Results and discussion

This section presents the results obtained by applying the proposed optimization algorithm to two different polyurethane materials: PU80 and PU90.

3.1 SHAP analysis for PU90

Figures 3a and b present the results of the SHAP analysis for PU90 conducted at the 100th iteration. This analysis revealed that the most influential features are r_1 and r_2. The observed trend shows that higher values of r_1 and r_2 contribute more to the objective function. Among the 100 samples, the design with the best objective had $r_1 = 7$ and $r_2 = 10$. Based on these findings, the lower bounds of r_1 and r_2 were adjusted, allowing for a 10% margin around the best solution, thereby narrowing down the search space. Consequently, the lower bounds were set to 6 and 9 for r_1 and r_2, respectively.

The next most influential feature, h, demonstrated a clear trend, but since the optimal value, indicated by the red dashed line in the figure, had a negative SHAP value, the bounds for h were not adjusted. For the remaining features, the SHAP plots displayed scattered patterns, indicating no clear influence, and therefore, their bounds were left unchanged.

In addition to the SHAP-based refinements, some bounds were further adjusted to meet imposed constraints. For instance, the minimum allowable value for h was calculated as $h_{min} = min(30 + 4r_1^{min}, 30 + 4r_2^{min})$, to ensure sufficient space between interfaces for manufacturing. This resulted in a revised lower bound of 54. Similarly, the lower bounds of other parameters were also adjusted to comply with geometric constraints, further narrowing the design space.

3.2 SHAP analysis for PU80

Figures 4a and b present the results of the SHAP analysis for PU90 conducted at the 100th iteration. For this material, the parameters D_1, r_1, B_2, B_3, t, and h all had refined bounds since there were consistently positive SHAP values to the right or left of the parameter value for the best sample so far.

There are differences between the SHAP analyses for PU80 and PU90 for almost every parameter. The SHAP values for PU90 indicate clearly that higher values of D_1, D_2, r_1, and r_2 increase the objective function, and increased t or h

decrease the objective. PU80 shows the opposite trends for these six parameters. B_1, B_3, and B_4 show no clear impact for PU90, but have distinct effects on PU80. Only B_2's SHAP values increase with the parameter for both materials. Considering the magnitudes as well as direction of the SHAP value dependencies, even this parameter varies significantly between the materials. Only the range of SHAP values for r_1 and h are similar for the two materials. The ranges of D_1, D_2, B_1, B_2, B_3, B_4, and t, all increase noticeably from PU90 to PU80, indicating that the performance of PU80 is in general much more sensitive to the feature geometry. In particular, PU90 is generally not sensitive to the vertical positions B_1, B_2, B_3, B_4, while PU80 is.

3.3 Discussion

We now discuss the efficiency of various optimization strategies. An ever-present concern when comparing optimization routines is that randomized initial data leads to a stochastic optimization result. A particular routine might perform well occasionally, but usually perform poorly. In practice, with an expensive objective function, one does not have the privilege of running multiple algorithms multiple times. Here we seek to robustly compare methods on a non-trivial but manageable objective function, to clearly inform decisions about which methods should be applied to problems with strictly limited optimization budgets. To this end, we execute each of the following optimization protocols several times with varied initial samples, to obtain confidence intervals for the effectiveness of each method as a function of how many iterations have been carried out. These results are presented in Figs. 3c and 4c. The lowest performing method (orange online) is a standard Bayesian Optimization routine. The next curve up (blue online) shows the result for BO with an off-the-shelf domain reduction technique applied after the first 100 random samples. For both materials the domain reduced optimization reaches the same performance after <150 iterations that standard BO reached after 400 iterations (indicated by the dashed line). For PU80, the domain reduced runs match BO's 400 iteration performance after only a few of iterations, and this method has a possibility (at the border of statistical significance) of discovering a better solution than BO within the 400 iteration budget. The Next curve up (pink online) represents the result of applying our SHAP-informed algorithm to reduce the search domain after the first 100 samples. This step is followed by a rapid improvement of performance for both materials, relative to standard BO, and both materials have surpassed the standard BO 400 iteration result after approximately 10 samples (indicated by the vertical dashed line). Even after this single application of the SHAP-informed domain reduction, the optimized objective function is roughly 6% higher for PU90 and 2% higher for for PU80 after 400 iterations. The remaining two curves in each figure represent the further improvement by applying the SHAP-informed domain reduction again after 150 iterations (yellow online) and 200 iterations (green online). The final results after 400 iterations yield on average an 11% improvement over BO for PU90, and 3% for PU80.

Notably, every time any of the domain reduction steps is applied, it is followed by an accelerated improvement, then a more sudden plateau. This character lends itself to early stopping criteria. For instance, consider the top curve in Fig. 3, corresponding to SHAP-bounding applied repeatedly up to 200 iterations. In this case, no further improvements are observed for the next 150 iterations, so the optimization could have been halted after 300 iterations, saved 25% of the budget, and still produced an 11% better solution than regular BO.

The application of these domain reduction steps could be considered tantamount to weighting exploitation over exploration, and there are other methods to tune Bayesian Optimization in this way. We surveyed various acquisition functions and biasing factors for off-the-shelf BO, and none of them performed significantly better than the implementation presented here.

One could of course consider variations to the schedule on which the domain reductions are applied. The primary concern is to avoid inadvertently excluding the global optimum from the search domain. We have attempted to mitigate this by only applying the domain reduction to a particular parameter if the SHAP value of that parameter is positive for the best geometry so far. The intuition is that a truly optimal solution would be expected to have positive contributions from all parameters. This is not strictly true, as parameters have nonlinear interactions which are not captured by the uni-variate SHAP dependence plots. However, we interpret the negative SHAP value in the optimal solution as an ambiguous signal, which should not be acted upon. A second precaution that our algorithm follows is to ensure that the best observed solution is included, with a margin, in the new domain. That is, if a particular parameter has high SHAP values in a region, while the best known solution's SHAP value for that parameter is positive, but outside that region, then we expand the new bounds to include both the higher SHAP values and the best known solution. With considerations such as these encoded in the algorithm, it could plausibly be applied as frequently as every iteration.

Another important consideration is the reliability of SHAP analysis on limited data sets. Figs. 3a and 4a present clean trends for some variables, and noisy trends for others. It is not obvious how to judge, much less how to encode, a criterion for judging a trend as meaningful or noisy. We implemented the criterion that all SHAP values to the right

or left of the best-so-far result must be positive for a parameter to be constrained. This is a very coarse criterion that errs on the side of distrusting a trend if there is any ambiguity about the positive impact of a region of parameter space. It also would tend to avoid restricting a parameter with a non-monotonic SHAP dependence. This choice of criterion was informed by the coarsest observation that for this optimization problem, the SHAP dependencies seem to usually be monotonic. Regardless of the specific criteria, though, there is a concern that a sparse sampling of the parameter space could yield a misleading SHAP analysis (in addition to a misleading initial optimum), which could result in bounds that exclude the global optimum. To avoid this we followed the standard practice in Bayesian Optimization of initially sampling ten times the problem dimensionality. Under this protocol, none of the optimization runs became locked out from a highly optimal solution, as evidenced by the narrow confidence intervals for the top curves in Figs. 3c and 4c. It would of course be advantageous to apply the accelerating effect of domain reduction earlier, but this is clearly risky without some heuristic to expect that the sampling has not missed an important domain. The authors can find no studies of the robustness of SHAP analysis for small data sets, so it is an open question if interpretability tools could provide some of that reassurance, and allow earlier activation of intelligent search algorithms if there is a clear "interpretation" of the existing data.

4 Conclusion

Bayesian Optimization is a popular algorithm with many applications and variations. We will mention here a few promising directions for further development of SHAP-informed optimization.

The most common kernels for GPR models in BO routines are radial basis functions or the more general Matern functions, because they perform reasonably well on many objective functions. However, this kernel assumes that the objective function is "flat" at the largest scales. That is, the model prediction far away from a training point always approaches the global mean. It is plausible that a kernel suited to the large-scale structure of a particular objective function could have a similar effect as the routine we have implemented here, but the authors are not aware of any algorithm for automatically searching the space of possible kernel functions to find one that is most appropriate for a particular problem. The algorithm presented here has the strength that it is straightforward to implement. Also, it can be wrapped around any choice of kernel or acquisition function to combine acceleration effects.

Another common strategy when dealing with multidimensional optimization task is to reduce the dimensionality using methods like principle component analysis, discriminant analysis, and autoencoders. In a qualitative way, these methods leverage the same information as the analysis presented here. They attempt to isolate the most effective directions to move through parameter space, and focus on those. Notably, these methods to modify the nature of the parameters are complementary to domain reduction, which seeks to modify the bounds on the parameters. Further investigation is needed to determine the effectiveness of combining these two strategies. They could be redundant due to leveraging the same broad trends. Alternatively, dimensional reduction could amplify the SHAP-bounding algorithm by constructing a parameter space with clearly defined SHAP trends.

Another potential expansion of the work presented here would be to include two-parameter SHAP interaction values. SHAP values for a particular parameter can vary systematically with other parameters, which sometimes makes it impossible to identify a domain with purely positive or negative SHAP values. See, for instance, the SHAP values for the parameter B_3 in Fig. 4. There are clearly two clusters of points, one with a distinct negative slope, and one that is more noisy with smaller magnitudes. In this case it was still possible to identify a range of B_3 for which the SHAP values were strictly positive, but other optimization tasks might not be so lucky. In fact, the apparent multi-modal character of SHAP values for a particular parameter value indicate precisely the kind of coupling between parameters that makes optimization problems generally challenging. In some cases, the bifurcation of SHAP values for a particular parameter can be primarily due to a second particular parameter. SHAP interaction values can be computed to examine the influence of varying any two parameters simultaneously. It would require some planning to implement a boundary reduction algorithm that utilizes this information, but it could be beneficial for problems with inseparable parameters.

Various methods have been reported for constructing acquisition functions to modify the behavior of BO (Liu et al. 2018). Future work may consider using SHAP values to bias an acquisition function without strictly enforcing a boundary. Such a method would have the advantage of eliminating the possibility of accidentally locking out the global optimum, but would require implementation effort for each combined acquisition function.

In this work, we have developed an optimization algorithm which leverages SHAP analysis to automatically modify the bounds of the BO search space. This investigation was motivated by the observation that Gaussian process models sometimes fail to capture the large-scale trends in a function, while SHAP identifies these trends well. While this is not a rigorous analysis of the interaction between the methods (particularly in high dimensions), it was sufficient to motivate the investigation. We applied this algorithm to the problem of optimizing the

layout of structural features in an acoustic metamaterial, and found that if classical BO was given a budget of 400 iterations, then SHAP-informed BO could match its performance with less than 1/3 of the iterations. After the SHAP-informed BO exhausted its full budget, it consistently achieved a 3% improvement for one material and an 11% improvement for the other.

Author contributions HW: conceptualization, methodology, software, validation, analysis, investigation, writing - original draft, visualization. DR: conceptualization, methodology, validation, writing - original draft, visualization. EH: ideation, conceptualization, supervision, writing - reviewing and editing, funding acquisition.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability Source code may be obtained from the corresponding author upon reasonable request.

Declarations

Replication of results Data and algorithms can be requested from the corresponding author.

References

Abdolahi J, Robe D, Larson R, Kirley M, Hajizadeh E (2025) Interpretable active learning meta-modeling for the association dynamics of telechelic polymers on colloidal particles. J Rheol 69:183–199. https://doi.org/10.1122/8.0000930

Abarghooie R, Zomorodian ZS, Tahsildoost M, Shaghaghian Z (2021) A machine-learning framework for acoustic design assessment in early design stages. In: Proceedings of Building Simulation 2021: 17th Conference of IBPSA, 1548–1555. https://doi.org/10.26868/25222708.2021.30401

Cai C, Hung KC, Khan MS (2006) Simulation-based analysis of acoustic absorbent lining subject to normal plane wave incidence. J Sound Vib 291(3):656–680

Calvo DC, Thangawng AL, Layman CN, Casalini R, Othman SF (2015) Underwater sound transmission through arrays of disk cavities in a soft elastic medium. J Acoust Soc Am 138(4):2537–2547. https://doi.org/10.1121/1.4931446

Dalal RJ, Oviedo F, Leyden MC, Reineke TM (2024) Polymer design via SHAP and bayesian machine learning optimizes pDNA and CRISPR ribonucleoprotein delivery. Chem Sci 15(19):7219–7228

Fu Y, Kabir II, Yeoh GH, Peng Z (2021) A review on polymer-based materials for underwater sound absorption. Polym Testing 96:107115. https://doi.org/10.1016/j.polymertesting.2021.107115

Hajizadeh E, Garmabi H (2008) Response surface based optimization of toughness of hybrid polyamide 6 nanocomposites. Int J Chem Biomol Eng 1:40–44

Hajizadeh E, Larson RG (2017) Stress-gradient-induced polymer migration in taylor–couette flow. Soft Matter 13(35):5942–5949. https://doi.org/10.1039/C7SM00821J

Islam MM, Das P, Rahman MMEA (2024) Prediction of compressive strength of high-performance concrete using optimization machine learning approaches with shap analysis. J Build Rehabil 9:94. https://doi.org/10.1007/s41024-024-00445-z

Jeon JH, Yang SS, Kang YJ (2020) Estimation of sound absorption coefficient of layered fibrous material using artificial neural networks. Appl Acoust 169:107476. https://doi.org/10.1016/j.apacoust.2020.107476

Khalid S, Song J, Azad MM, Elahi MU, Lee J, Jo SH (2023) Kim HS prediction of compressive strength of high-performance concrete using optimization machine learning approaches with shap analysis. J Build Pathol Rehabil 8:1–12. https://doi.org/10.1007/s42107-023-00689-z

Kumar S, Jin H, Lim KM, Lee HP (2023) Comparative analysis of machine learning algorithms on prediction of the sound absorption coefficient for reconfigurable acoustic meta-absorbers. Appl Acoust 212:109603. https://doi.org/10.1016/j.apacoust.2023.109603

Liang Q, Gongora AE, Ren Z, Tiihonen A, Liu Z, Sun S, Deneault JR, Bash D, Mekki-Berrada F, Khan SA, Hippalgaonkar K, Maruyama B, Brown KA, Fisher J III, Buonassisi T (2021) Benchmarking the performance of bayesian optimization across multiple experimental materials science domains. NPJ Comput Mater 7(1):188. https://doi.org/10.1038/s41524-021-00656-9

Linardatos P, Papastefanopoulos V, Kotsiantis S (2020) Explainable ai: a review of machine learning interpretability methods. Entropy 23(1):18. https://doi.org/10.3390/e23010018

Liu H, Ong Y-S, Cai J (2018) A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Struct Multidisc Optim 57(1):393–416

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Guyon I, Luxburg U, Bengio S, Wallach HM, Fergus R, Vishwanathan SVN, Garnett R (eds) Advances in Neural Information Processing Systems 30. Curran Associates Inc, Red Hook, pp 4765–4774

Meng H, Wen J, Zhao H, Wen X (2012) Optimization of locally resonant acoustic metamaterials on underwater sound absorption characteristics. J Sound Vib 331(20):4406–4416. https://doi.org/10.1016/j.jsv.2012.05.027

Panisilvam J, Hajizadeh E, Weeratunge H, Bailey J, Kim S (2023) Asymmetric cyclegans for inverse design of photonic metastructures. APL Machine Learn 1(4):046105

Palar PS, Zuhal LR, Dwianto YB, Shimoyama K, Morlier J (2023) Shapley Additive Explanations for Knowledge Discovery via Surrogate Models. In: AIAA SciTech Forumv, National Harbor, United States, p. 0. https://doi.org/10.2514/6.2023-0332. https://hal.science/hal-04034297

Ranjbar M, Bayer MU (2024) Design of acoustic coating for underwater stealth in low-frequency ranges. J Braz Soc Mech Sci Eng 46:140. https://doi.org/10.1007/s40430-024-04720-5

Robe D, Menzel A, Phillips A, Hajizadeh E (2025) Smiles to scattering: automated high-throughput atomistic polyurethane simulations compared with waxs data. Comput Mater Sci 256:113931. https://doi.org/10.1016/j.commatsci.2025.113931

Salih AM, Raisi-Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, Menegaz G (2025) A perspective on explainable

artificial intelligence methods: SHAP and LIME. Adv Intell Syst 7(1):2400304. https://doi.org/10.1002/aisy.202400304

Sharma GS, Skvortsov A, MacGillivray I, Kessissoglou N (2017) Sound transmission through a periodically voided soft elastic medium submerged in water. Wave Motion 70:101–112. https://doi.org/10.1016/j.wavemoti.2016.10.006

Sharma GS, Skvortsov A, MacGillivray I, Kessissoglou N (2019) Sound absorption by rubber coatings with periodic voids and hard inclusions. Appl Acoust 143:200–210. https://doi.org/10.1016/j.apacoust.2018.09.003

Shireen Z, Hajizadeh E, Daivis P, Brandl C (2023) Linear viscoelastic shear and bulk relaxation moduli in poly(tetramethylene oxide) (PTMO) using united-atom molecular dynamics. Comput Mater Sci 216:111824

Shireen Z, Weeratunge H, Menzel A, Hajizadeh E (2022) A machine learning enabled hybrid optimization framework for efficient coarse-graining of a model polymer. NPJ Comput Mater 8:224. https://doi.org/10.1038/s41524-022-00914-4

Shrikumar A, Greenside P, Kundaje A (2017) Learning important features through propagating activation differences. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, 70, 3145–3153. PMLR, Sydney, Australia. https://proceedings.mlr.press/v70/shrikumar17a.html

Stender M, Adams C, Wedler M, Grebel A, Hoffmann N (2021) Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube. J Acoust Soc Am 149(3):1932–1945. https://doi.org/10.1121/10.0003755

Tabatabaei Manesh M, Nikkhah Dehnavi A, Tahsildoost M, Alambeigi P (2024) Acoustic design evaluation in educational buildings using artificial intelligence. Build Environ 261:107476. https://doi.org/10.1016/j.buildenv.2024.111695

Vimbi V, Shaffi N, Mahmud M (2024) Interpreting artificial intelligence models: a systematic review on the application of lime and shap in Alzheimer's disease detection. Brain Inform 11(1):10. https://doi.org/10.1186/s40708-024-00222-1

Wang Z, Liu H, Amin MN, Khan K, Qadir MT, Khan SA (2024) Optimizing machine learning techniques and shapley additive explanations (shap) analysis for the compressive property of self-compacting concrete. Mater Today Commun 39:108804. https://doi.org/10.1016/j.mtcomm.2024.108804

Weeratunge HZ, Shireen I, Sagar A, Menzel AW, Phillips S, Halgamuge R, Hajizadeh E (2022) A machine learning accelerated inverse design of underwater acoustic polyurethane coatings. Struct Multidisc Optim 65:213. https://doi.org/10.1007/s00158-022-03322-w

Weeratunge H, Robe D, Menzel A, Phillips AW, Kirley M, Smith-Miles K, Hajizadeh E (2023) Bayesian coarsening: rapid tuning of polymer model parameters. Rheol Acta 62(10):477–490. https://doi.org/10.1007/s00397-023-01397-w

Xiang J, Hajizadeh E, Larson RG, Nelson D (2021) Predictions of polymer migration in a dilute solution between rotating eccentric cylinders. J Rheol 65(6):1311–1325. https://doi.org/10.1122/8.0000330

Zhu G, Rezvantalab H, Hajizadeh H, Wang X, Larson RG (2016) Stress-gradient-induced polymer migration: perturbation theory and comparisons to stochastic simulations. J Rheol 60(2):327–343. https://doi.org/10.1122/1.4942252