

Article

Data-Driven Framework for Predicting Airborne Sound Insulation of Recycled Rubber–Polyurethane Composite Panels

Miljan Kovačević ^{1,*} , Anđelko Crnoja ², Borko Bulajić ³  and Predrag Petronijević ⁴ 

¹ Faculty of Technical Sciences, University of Pristina, Knjaza Milosa 7, 38220 Kosovska Mitrovica, Serbia

² Department of Construction, University North, 104 Brigade 3, 42000 Varaždin, Croatia; acrnaja@unin.hr

³ Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia; borkobulajic@uns.ac.rs

⁴ Faculty of Civil Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia; pecap@grf.bg.ac.rs

* Correspondence: miljan.kovacevic@pr.ac.rs; Tel.: +381-606173801

Abstract

The increasing accumulation of end-of-life tires has motivated the development of sustainable construction materials incorporating recycled rubber for acoustic insulation applications. This study proposes a data-driven framework for predicting the weighted airborne sound reduction index (R_w) of recycled rubber–polyurethane composite panels based on a limited experimental dataset. Specimens with varying granulometric composition, material density, and polyurethane adhesive dosage were evaluated in accordance with EN ISO 10140-2:2010 and EN ISO 717-1:2013. To address data scarcity, a regression-oriented SMOTE strategy was applied exclusively to the training set to preserve statistical representativeness and avoid data leakage. Test set representativeness was ensured by systematically evaluating numerous data splits and adopting the one that maximized multivariate statistical consistency. A hierarchical modeling approach was adopted, ranging from classical regression models to tree-based ensemble methods and multigenic symbolic regression. Model performance was evaluated using R^2 , RMSE, MAE, and MAPE on an independent test set. The highest accuracy and robustness were obtained using symbolic regression, with R^2 values close to 0.99 and minimal prediction errors. Shapley value analysis and PDP/ICE plots identified material density as the dominant predictor of R_w , followed by polyurethane adhesive dosage, while granulometric composition exhibited a weaker influence. The proposed framework provides an accurate and interpretable tool for the preliminary design and optimization of recycled rubber acoustic panels.

Keywords: recycled rubber composites; weighted sound reduction index; machine learning; symbolic regression



Academic Editor: Kimihiro Sakagami

Received: 26 January 2026

Revised: 14 February 2026

Accepted: 23 February 2026

Published: 2 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The accumulation of end-of-life (EoL) tires is a growing global environmental concern. Each year, approximately 1.5 billion tires are manufactured and discarded worldwide, contributing to a massive waste stream that poses significant ecological and health risks due to their non-biodegradable nature [1]. In many regions, nearly 1 billion EoL tires are generated annually, constituting more than 2% of total solid waste and overwhelming current waste management systems [2,3].

Improper disposal through landfilling or incineration exacerbates pollution through toxic emissions and soil contamination. As a response, numerous recycling and valorization

strategies have emerged, including the incorporation of crumb rubber into construction materials, which recent life cycle assessments show can reduce carbon emissions and support circular economy objectives [4]. Incorporating recycled rubber into acoustic and thermal insulation panels has demonstrated not only environmental benefits but also improvements in sound absorption, vibration damping, and energy efficiency. These panels typically combine rubber granules with binders such as cement, gypsum, or polyurethane, yielding composites with varying porosity, density, and acoustic response.

Previous studies by Crnoja and Kos have demonstrated that granulometric composition, density, and binder content strongly influence the acoustic performance of recycled rubber-based construction materials, while also highlighting their environmental relevance [5].

Smirnova et al. investigated the acoustic absorption properties of lightweight composite panels composed of waste tire rubber crumbs (1–3 mm) and two types of binders: cement and gypsum [6]. Panels were fabricated with 40%, 60%, and 80% rubber content and thicknesses of 10–40 mm. The samples were tested in an impedance tube in accordance with ISO 10534-2 [7]. The results showed that the highest sound absorption coefficients ($\alpha > 0.7$) were achieved in samples with 80% rubber content and gypsum binder, particularly at frequencies above 1000 Hz [6]. The data confirmed that both binder type and rubber content had a significant impact on acoustic behavior.

Colom et al. investigated the thermal and acoustic properties of composites made from ground tire rubber (GTR) and synthetic binders such as polyurethane and acrylics [8]. Using multivariate linear regression, they modeled thermal conductivity as a function of porosity, density, and thickness, achieving an R^2 value of up to 0.86. Although acoustic absorption was not explicitly modeled, the results indicated that thickness and density also positively influence sound absorption behavior, supporting the suitability of rubber-binder composites for sustainable building insulation [8].

Badida et al. conducted an experimental investigation on the sound insulation performance of compressed recycled rubber crumb materials, evaluating the influence of rubber fraction (10–30%), compaction pressure (200–400 MPa), and sample thickness on sound transmission loss (STL) [9]. The results showed that STL improves with increasing density and thickness, highlighting significant nonlinear dependencies between material parameters and acoustic behavior [9].

Algaifi et al. investigated the prediction of mechanical and acoustic properties of rubberized concrete using artificial intelligence methods, specifically artificial neural networks (ANNs) [10]. The input parameters included rubber aggregate ratio, graphene content, binder composition, density, and porosity. The developed models demonstrated an excellent predictive accuracy, with coefficients of determination exceeding 0.95 for acoustic indicators.

Recent studies have demonstrated the applicability of machine learning techniques for analyzing sound insulation performance in complex composite structures. Wang et al. employed a Random Forest model to evaluate the contribution of multiple material layers and structural parameters to the weighted sound insulation index (R_w) of high-speed train composite floors, achieving coefficients of determination exceeding 0.88 [11]. Their results highlighted the strongly nonlinear relationship between material properties and acoustic performance, emphasizing the limitations of traditional single-parameter analysis approaches in acoustic systems.

Despite extensive experimental investigations on rubber-based acoustic composites, predictive modeling of the weighted airborne sound reduction index (R_w) remains limited, particularly with respect to interpretable data-driven approaches suitable for small-sample conditions. Existing studies largely focus on experimental characterization, whereas the sys-

tematic integration of experimentally validated data with robust and explainable machine learning models remains scarce.

This situation reveals a clear research gap, as limited dataset sizes constrain the direct application of conventional machine learning techniques and necessitate modeling strategies capable of operating reliably with experimentally grounded data. Accordingly, there is a need for a framework that combines statistical rigor, predictive performance, and physical interpretability.

Recent review studies on machine learning in acoustics emphasize the increasing adoption of data-driven methods for sound insulation and noise control, while simultaneously identifying critical challenges related to data scarcity, model robustness, and physical interpretability of predictions [12]. Reviews focusing on machine learning applications in acoustics highlight that many existing approaches rely on large datasets or black-box models, which limit their applicability in laboratory-scale material testing and engineering-oriented interpretation [12,13]. Complementary reviews on acoustic metamaterials and polymer-based sound-insulating materials further indicate that experimental investigations remain dominant, whereas statistically grounded predictive frameworks integrating experimentally validated data with explainable machine learning techniques are still relatively scarce [13,14].

Consequently, a methodological gap exists between experimentally validated acoustic research and the development of reliable, interpretable predictive models capable of operating under limited data conditions. To bridge this gap, the present study proposes a hierarchical experimental–statistical–machine learning framework for predicting the R_w of rubber–polyurethane composite panels. Linear regression is adopted as a transparent baseline, followed by tree-based ensemble models to capture nonlinear interactions and multigene genetic programming (MGGP) to derive explicit analytical expressions. Model robustness is quantitatively assessed, while interpretability is ensured through Shapley Additive Explanations (SHAP), complemented by Partial Dependence (PDP) and Individual Conditional Expectation (ICE) analyses.

In this way, the proposed framework directly addresses the limitations identified in the literature by combining statistical rigor, predictive accuracy, and physical interpretability within a unified and experimentally grounded modeling strategy.

2. Materials and Methods

2.1. Materials

Laboratory measurements of airborne sound insulation were conducted at the accredited acoustics laboratory of the Institute of Civil Engineering of Croatia (IGH) using a two-room acoustic field method. The test facility consisted of a structurally decoupled source room ($\approx 60 \text{ m}^3$) and receiving room ($\approx 56.5 \text{ m}^3$), designed to prevent flanking transmission. Test specimens were mounted in a window-type configuration with an exposed area of approximately 0.8 m^2 , using a modular wooden frame. Installation gaps were fully filled with soft mineral wool, while airtightness and vibration decoupling were ensured using acrylic acoustic sealant and elastic rubber gaskets, eliminating acoustic bridges (Figure 1).

Airborne sound insulation was measured in accordance with EN ISO 10140-2:2010 [15], which defines laboratory test methods for the airborne sound insulation of building elements. Measurements were performed using Brüel & Kjær (Brüel & Kjær, Nærum, Denmark) instrumentation (Figure 2), including a dual-channel analyzer (Type 2270), an omnidirectional sound source (Type 4296), and condenser microphones (Type 4189). A white noise signal with an average sound pressure level of approximately 105 dB was generated in the source room, while sound pressure levels were recorded in the receiving room.

Measurements were conducted in the one-third-octave frequency range from 50 Hz to 5000 Hz using multiple source positions [15–17].

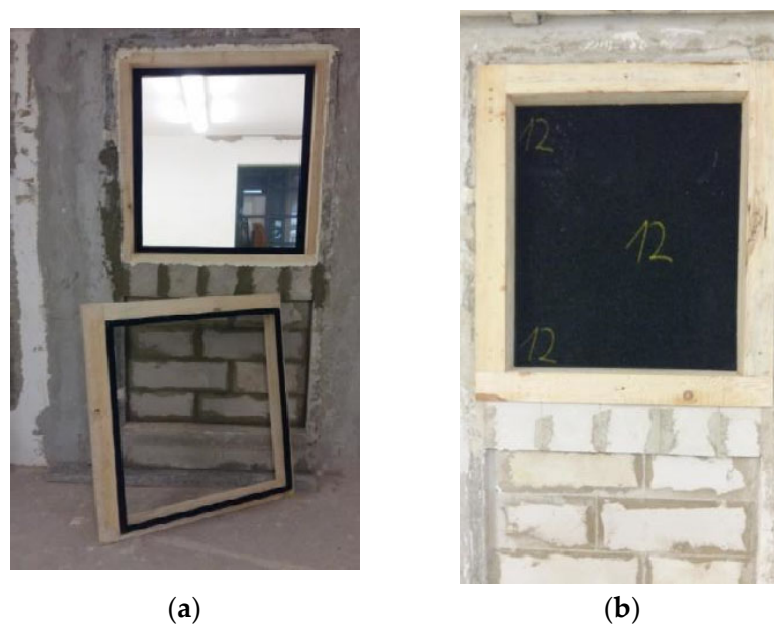


Figure 1. Experimental setup: (a) test frame for specimen installation; (b) installed test specimen; source: authors [16].

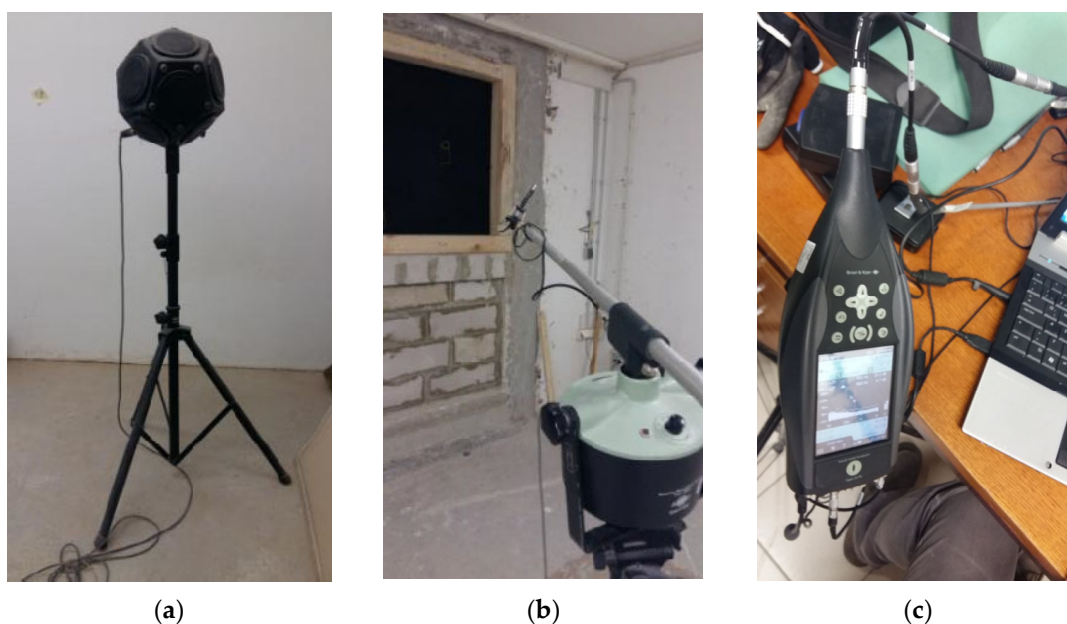


Figure 2. Measurement equipment: (a) omnidirectional sound source (Type 4296, Brüel & Kjær); (b) condenser microphones (Type 4189, Brüel & Kjær); (c) handheld analyzer (Type 2270, dual-channel, Brüel & Kjær) [16].

The evaluation and determination of the single-number weighted sound reduction index (R_w) were carried out in accordance with EN ISO 717-1:2013 [17], based on a comparison of third-octave band sound reduction indices with the standardized reference curve (Figure 3).

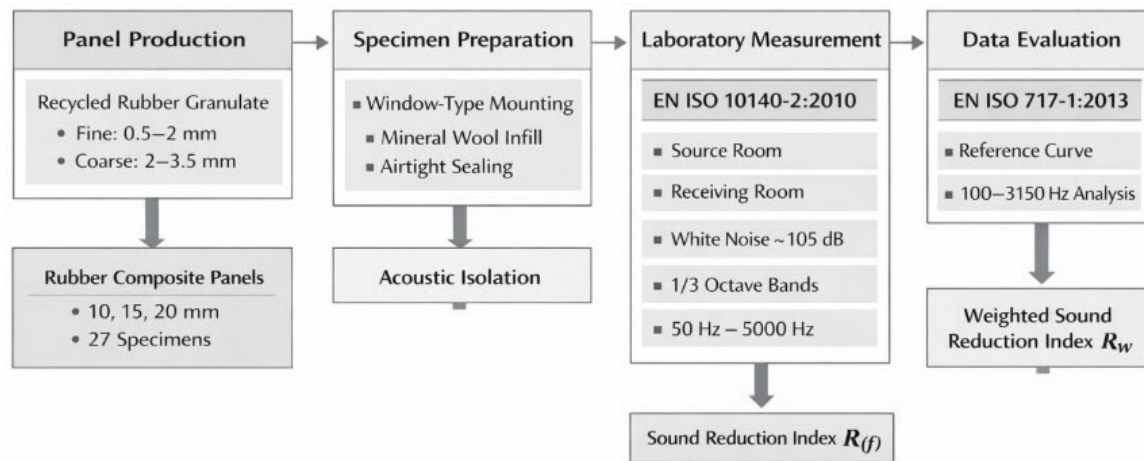


Figure 3. Methodological framework for laboratory measurement and evaluation of airborne sound insulation of recycled rubber composite panels.

The tested specimens were composite panels manufactured from recycled rubber granulate obtained through the mechanical processing of end-of-life tires, followed by sieving, magnetic separation, and additional crushing and classification to remove steel fibers and impurities and to obtain predefined particle size fractions. Two rubber aggregate fractions were used: a fine fraction (0.5–2 mm) and a coarse fraction (2–3.5 mm). Three mixture types were prepared: panels containing only fine aggregate, only coarse aggregate, and a combined mixture consisting of 35% fine and 65% coarse aggregate, as aggregate grading directly influences panel porosity and acoustic energy dissipation [15–17].

Panels with nominal thicknesses of 10, 15, and 20 mm were produced with controlled variations in granulometric composition, bulk density, and polyurethane-based thermoreactive binder content. After homogenization in industrial mixers, the mixture was cast into steel molds and compacted by industrial hot pressing under pressures up to 5 MPa (≈ 30 tonnes surface load) at a controlled production temperature of 80–130 °C (± 5 °C tolerance), ensuring binder activation and curing. Thermal curing was performed either in controlled chambers for 45 min or under ambient storage conditions.

A total of 27 panels were examined (Table 1), covering three thickness levels (10, 15, and 20 mm) and three granulometric compositions (fine, coarse, and combined), with three replicate specimens per configuration.

Table 1. Specimen sampling plan.

Sample Thickness	Granulometric Composition	Number of Specimens
10 mm	0.5–2 mm	3
10 mm	2–3.5 mm	3
10 mm	0.5–2 (35%) + 2–3.5 (65%) mm	3
15 mm	0.5–2 mm	3
15 mm	2–3.5 mm	3
15 mm	0.5–2 (35%) + 2–3.5 (65%) mm	3
20 mm	0.5–2 mm	3
20 mm	2–3.5 mm	3
20 mm	0.5–2 (35%) + 2–3.5 (65%) mm	3
Total specimens		27

This sampling strategy ensures that both geometric and material-related variability are systematically represented, while the use of replicate specimens enables the assessment

of repeatability and limits the influence of local material heterogeneity under controlled laboratory conditions [15–17].

2.2. Methods

The analyzed models are selected to represent a hierarchy of increasing complexity. Linear and polynomial regression models are employed as transparent baselines to assess global trends and the degree of nonlinearity in the data. Tree-based ensemble methods are used to capture localized nonlinear interactions with enhanced predictive capability, while multigene symbolic regression is adopted to achieve a high accuracy together with an explicit analytical formulation suitable for engineering interpretation. Particular emphasis is placed on evaluating the robustness of the optimal model through cross-dataset performance consistency.

In addition, explainable machine learning techniques are applied to interpret the selected model. Shapley value analysis is used to quantify the relative contribution of each input variable to the predicted response, providing physically meaningful insight into the governing factors of airborne sound insulation.

All machine learning analyses were performed in MATLAB R2023b (MathWorks, Natick, MA, USA), utilizing the Statistics and Machine Learning Toolbox for regression modeling, validation, and statistical evaluation. Symbolic regression modeling was conducted using GPTIPS 2, an open source multigene genetic programming framework operating within the MATLAB environment. The selection of these software tools was motivated by their robustness, reproducibility, advanced statistical capabilities, and suitability for small-sample regression analysis with strong interpretability requirements.

2.2.1. Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) was employed as the reference model to establish a baseline representation of the input–output relationship. Beyond its predictive role, MLR provides analytical insight into the global influence of individual input variables. The model assumes a linear dependency between the response variable and the predictors and can be expressed as (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \quad (1)$$

where β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_p$ are regression coefficients, and ϵ denotes the random error term. Model parameters were estimated using the ordinary least squares method by minimizing the sum of squared residuals using the following Equation (2):

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2)$$

where $\hat{\beta}$ represents the estimated parameter vector, X is the predictor matrix, and Y contains the observed responses.

Although linear regression is inherently limited in its ability to represent nonlinear interactions, it serves as an essential benchmark. Differences in performance between the MLR model and more advanced approaches provide a quantitative indication of the degree of nonlinearity present in the investigated problem.

2.2.2. Regression Trees

Regression trees represent the first step toward locally adaptive modeling. The fundamental principle is the recursive partitioning of the input space into disjoint regions (Figure 4), within which the response variable is approximated by a constant value [18,19].

At each split, the algorithm selects the input variable and threshold that minimize the within-region sum of squared errors (3):

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (3)$$

where $R_1(j,s)$ and $R_2(j,s)$ are the regions formed by partitioning the j -th predictor at s , with c_1 and c_2 denoting the corresponding regional mean predictions.

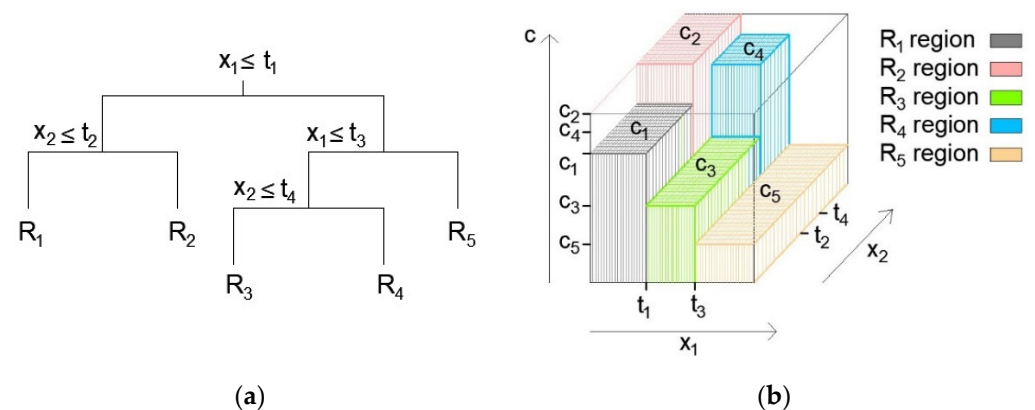


Figure 4. Input-space partitioning (a) and 3D regression surface obtained by a regression tree (b) [19].

2.2.3. TreeBagger and Random Forest Ensembles

While regression trees can effectively capture localized patterns, individual trees are highly sensitive to variations in the training data and therefore tend to exhibit high variance. Consequently, they are primarily used as base learners within ensemble frameworks rather than as standalone predictive models. To mitigate the instability of individual regression trees, an ensemble approach based on bootstrap aggregation (bagging) was adopted [18–20].

Within the TreeBagger framework, a large number of regression trees are trained on different bootstrap-resampled subsets drawn from the original dataset (Figure 5). In the Random Forest variant, additional randomness is introduced by restricting the number of candidate input variables considered at each split, which further reduces correlation among individual trees and improves generalization performance.

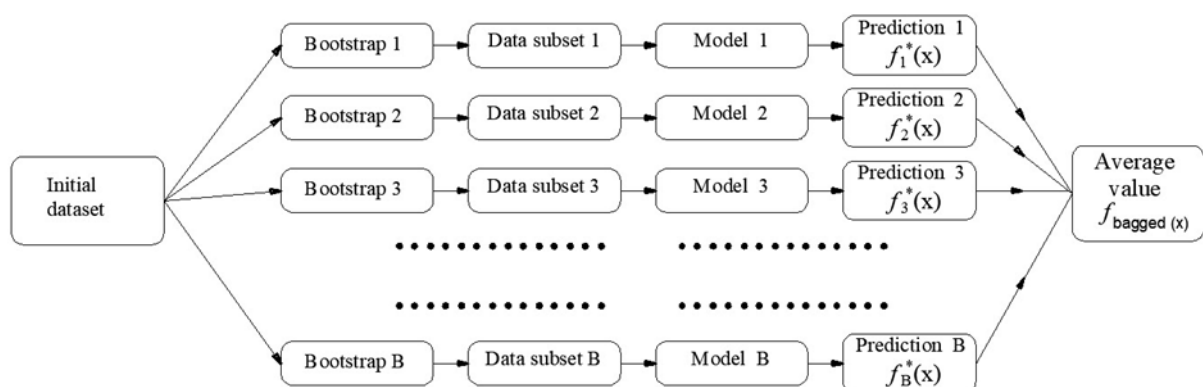


Figure 5. Schematic representation of the bagging-based ensemble regression framework [19].

Conceptually, Random Forest models can be interpreted as an ensemble of simple local regression models whose aggregated output yields a smooth and robust approximation of a complex input–output relationship.

2.2.4. Gradient Boosted Tree Ensembles

In contrast to bagging-based ensembles, Gradient Boosted Trees rely on sequential model construction [21–23]. Each regression tree is trained to approximate the residuals of the ensemble obtained in the previous iteration, thereby incrementally improving predictive accuracy (Figure 6). This process corresponds to a stage-wise minimization of the loss function.

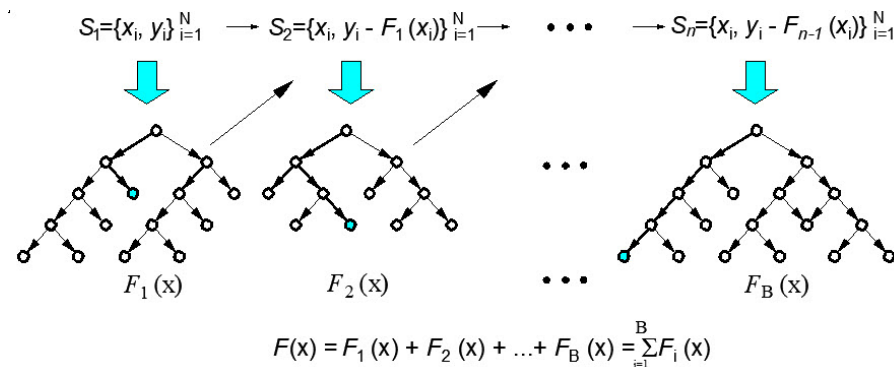


Figure 6. Gradient boosting ensemble model [19].

By employing shallow trees and a small learning rate, the model retains a local approximation character while controlling excessive complexity. Gradient boosting can thus be interpreted as an additive model composed of many weak, locally valid regressors whose cumulative effect captures complex nonlinear behavior in a controlled manner.

The final prediction for TreeBagger, Random Forest, and Gradient Boosted trees is obtained by averaging the outputs of all trees in the ensemble. This aggregation substantially reduces model variance while preserving the ability to capture localized nonlinear dependencies.

2.3. Pseudo-Linear Symbolic Regression Using Multigene Genetic Programming (MGGP)

Multigene Genetic Programming was adopted as a symbolic regression framework to derive an explicit analytical mapping between the predictors and the target response, without prespecifying the functional form. In MGGP, each candidate model consists of multiple expression trees (“genes”), where terminal nodes are formed by input variables and numerical constants, while internal nodes represent operators from a predefined function set. The initial population is generated randomly, ensuring a broad exploration of feasible model structures [24–26].

Each model in the population is evaluated using a fitness function, and selection is biased toward higher-performing individuals while preserving diversity through probabilistic reproduction. New candidate models (offspring) are generated using crossover, mutation, and direct copying, and this process is repeated over successive generations (Figure 7).

MGGP produces a pseudo-linear model: the final predictor is a linear combination of nonlinear gene outputs (4). This structure enables nonlinear feature construction at the gene level while retaining stable coefficient estimation through linear regression at the model level.

$$\hat{y} = b_0 + b_1 t_1 + b_2 t_2 + \dots + b_G t_G \quad (4)$$

where b_0 is the bias coefficient, b_i denote gene-specific scaling parameters, and t_i are the outputs of individual trees (genes). Accordingly, the gene response matrix is given by (5):

$$\mathbf{G} = [1, t_1, t_2, \dots, t_G]. \quad (5)$$

The vector $\mathbf{b}$ is determined via least-squares estimation according to (6):

$$\mathbf{b} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y}. \quad (6)$$

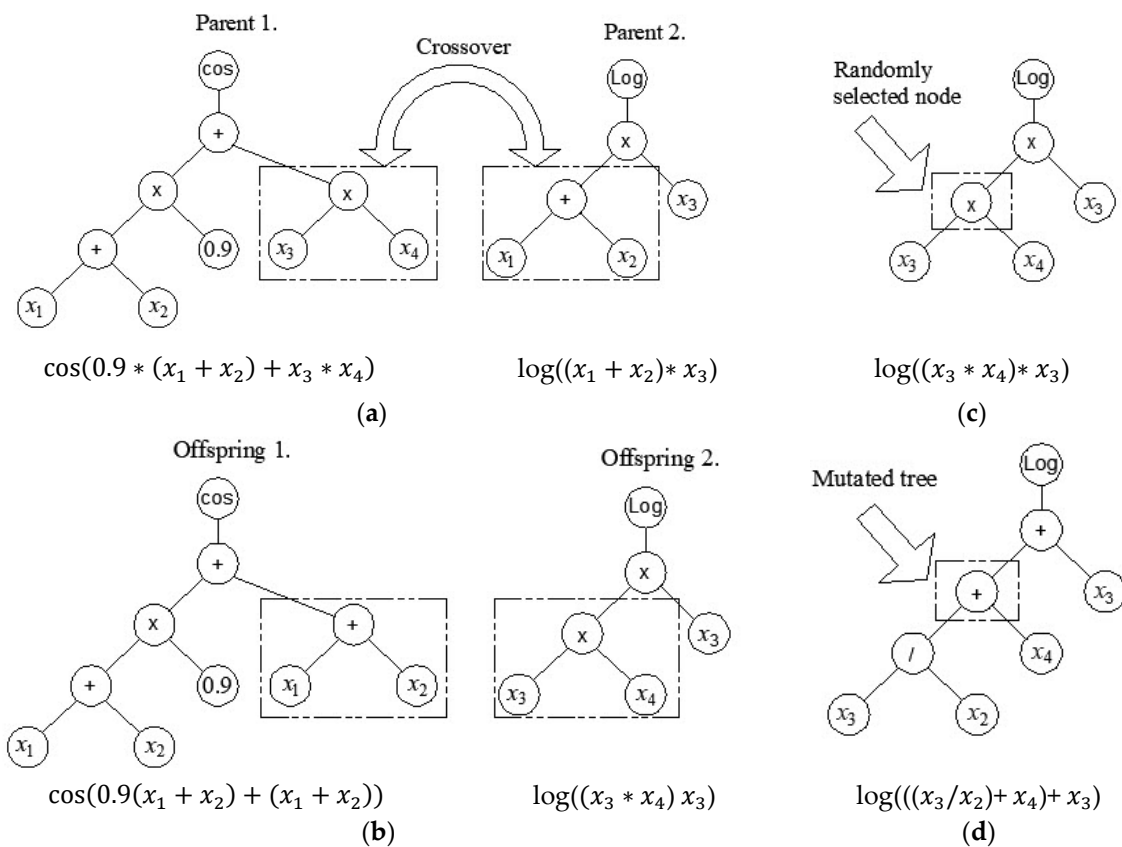


Figure 7. Crossover and mutation mechanisms in MGGP models: crossover point selection (a), subtree exchange (b), mutation point selection (c), and subtree mutation (d) [19].

This formulation allows the model to preserve a linear aggregation framework while incorporating complex nonlinear components. Consequently, MGGP models can be classified as pseudo-linear, offering high predictive capability together with an explicit mathematical form suitable for engineering interpretation and practical use.

3. Dataset

For model development, it is necessary to construct structured datasets in which the input variables and target responses are explicitly defined in matrix and vector form, respectively. Accordingly, the input data are organized into the matrix (7):

$$\mathbf{X} = [x_{ij}] \in \mathbb{R}^{N \times 3}, \quad (7)$$

where N denotes the number of tested specimens, each row corresponds to one specimen, and each column represents a distinct material descriptor.

The first column, x_{i1} , corresponds to the granulometric composition (GC), defined as the percentage share (0–100%) of the fine rubber fraction (0–2 mm). The coarse rubber fraction (2–3.5 mm) is not introduced as an independent variable because it is fully complementary to GC. For example, a value of $GC = 35\%$ directly implies a coarse fraction content of 65%. This formulation avoids redundancy in the input space and prevents multicollinearity while fully preserving the granulometric information of the rubber mixture.

The second column, x_{i2} , represents the panel density (SC), expressed in kg/m^3 , which reflects the degree of compaction and governs mass-related airborne sound insulation mechanisms.

The third column, x_{i3} , corresponds to the polyurethane adhesive dosage (PU_ADH), expressed in g/m^2 , describing the binder content that controls inter-particle bonding, stiffness, and microstructural connectivity.

Together, these three variables comprehensively characterize the granulometric, physical, and binder-related properties of the composite panels and constitute the predictor matrix used in all regression and machine learning models developed in this study. The corresponding output data are defined by the target vector (8):

$$R_w \in \mathbb{R}^N \quad (8)$$

where each element represents the experimentally measured weighted sound reduction index R_w for the associated specimen.

The dataset (X, R_w) was evaluated under 100 repeated random splits (Figure 8) into training and test subsets (70/30). For each split, only the training subset was augmented using a regression-oriented SMOTE procedure, while the test subset was kept untouched to prevent data leakage [27–29].

Training predictors were z-score standardized and grouped using k-means clustering [16] to encourage local interpolation. Synthetic samples were generated by linear interpolation between a randomly selected training instance and one of its k nearest neighbors. The augmented training size was capped at $N_{\text{target}} = \min(120.5N_{\text{train}})$. To preserve physical plausibility, synthetic R_w values were constrained within an extended training-based range ($\pm 10\%$ of the observed training span).

A grid search was conducted over $\text{numClusters} \in \{1, 2, 3\}$ and $k\text{NN} \in \{3, 5, 7\}$, jointly with the split selection. Here, numClusters denotes the number of clusters used in the k-means partitioning of the training data, while $k\text{NN}$ denotes the number of nearest neighbors considered when generating synthetic samples in the regression-oriented SMOTE procedure [29]. Representativeness was quantified using Train–Synthetic (TS) and Train–Test (TT) comparisons for both R_w and each input feature, applying: two-sample Kolmogorov–Smirnov test (distributional similarity), two-sample t -test (mean consistency), Levene’s test (variance homogeneity), and Cohen’s d (effect size) [30–33].

An aggregate score favoring high p -values and low $|d|$ was minimized (Figure 8) to select the optimal configuration. In the construction of the aggregate score, different penalty formulations were adopted to reflect the distinct interpretation of statistical tests and effect size metrics. For hypothesis-testing criteria, p -values were transformed using terms of the form $1 - p$, ensuring that configurations associated with higher p -values—indicative of weaker evidence against distributional similarity—incur smaller penalties. This transformation provides a continuous and monotonic mapping from statistical significance to a comparable contribution within the aggregate score.

Conversely, for effect size measures such as Cohen’s d , the absolute value was incorporated directly, as practical equivalence is governed by the magnitude of the standardized difference rather than its sign. By combining these complementary penalty structures, the scoring function simultaneously discourages statistically significant discrepancies and practically relevant deviations, thereby enforcing consistency at both inferential and applied levels. This design allows the heuristic to capture nuanced differences across variables while maintaining a coherent global optimization criterion.

The optimal configuration was: split = 62, numClusters = 2, kNN = 3, yielding 19 training and 8 test samples and an augmented training set of 95 samples (19 original + 76 synthetic). As summarized in Table 2, high p -values and very small effect

sizes indicate a strong agreement between training and synthetic data (TS) and satisfactory consistency between training and test data (TT), supporting multivariate representativeness for model development and independent validation.

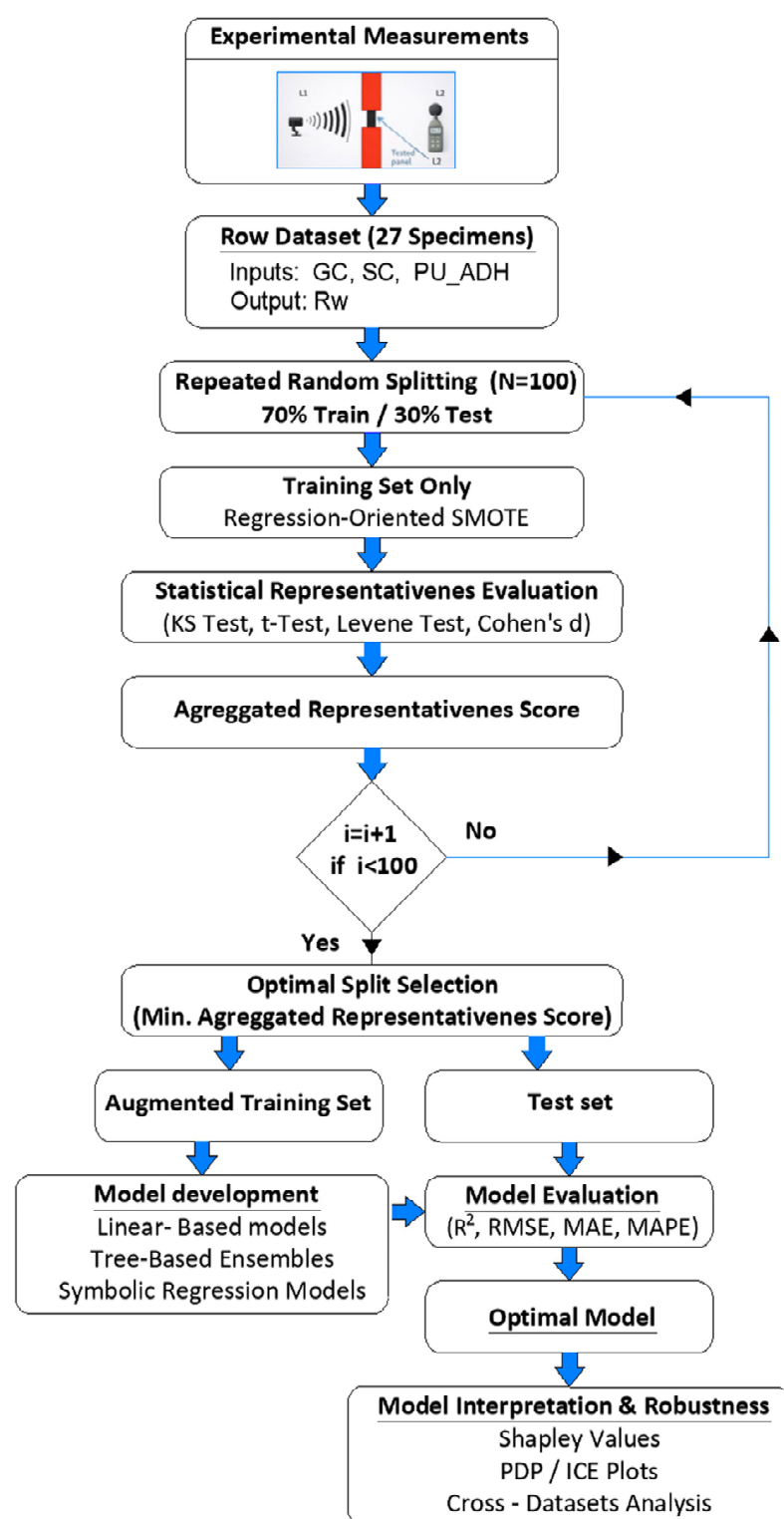


Figure 8. Workflow of the data processing, resampling, model development, and evaluation framework for predicting the airborne sound insulation index.

Table 2. Statistics for the optimal configuration (split = 62, numClusters = 2, kNN = 3).

Comparison	Variable	KS Test p	t -Test p	Levene p	Cohen's d
Train vs. Synthetic (TS)	R_w	0.9393	0.9755	0.6325	0.0079
Train vs. Test (TT)	R_w	0.9514	0.9564	0.6956	0.0233
Train vs. Synthetic (TS)	Inputs X (mean over features)	0.9015	0.9702	0.6492	0.0096 *
Train vs. Test (TT)	Inputs X (mean over features)	0.9532	0.6843	0.4445	0.1783 *

* For X, Cohen's d is reported as mean $|d|$ across input features.

Table 2 confirms that the selected split and SMOTE configuration yield statistically consistent datasets. Table 3 presents the descriptive statistics of the input variables and the output R_w for the training, SMOTE-augmented training, and test datasets.

Table 3. Descriptive statistics of input variables and output R_w for the training, training augmented, and test datasets.

Training set (n = 19)				
Variable	Min	Max	Mean	Std
Granulometric composition	0.0	100.0	46.1	44.6
Density	585.0	1100.0	816.7	174.3
PU_ADH	296.0	1340.0	667.9	343.5
R_w	6.3	33.5	17.2	7.4
Training augmented (SMOTE) set (n = 95 = 19 + 76)				
Variable	Min	Max	Mean	Std
Granulometric composition	0.00	100.00	46.19	41.22
Density	585.00	1100.00	834.31	158.90
PU_ADH	296.00	1340.00	649.55	316.76
R_w	6.3	33.5	17.7	6.1
Test set (n = 8)				
Variable	Min	Max	Mean	Std
Granulometric composition	0.00	100.00	42.50	38.64
Density	600.00	916.00	766.38	132.36
PU_ADH	296.00	1340.00	616.38	387.13
R_w	7.2	36.5	17.4	9.1

Table 2 shows that, for R_w , Train–Synthetic comparisons exhibit a high agreement (KS $p = 0.9393$, t -test $p = 0.9755$, Levene $p = 0.6325$, $d = 0.0079$), indicating that the data augmentation procedure did not alter the target distribution. Train–Test consistency for R_w is also strong (KS $p = 0.9514$, t -test $p = 0.9564$, Levene $p = 0.6956$, $d = 0.0233$), supporting the representativeness of the hold-out test set. Across the input variables, Train–Synthetic differences are negligible (mean $|d| = 0.0096$), while Train–Test comparisons show a small effect (mean $|d| = 0.1783$), which is plausible given the limited size of the test dataset.

Table 3 demonstrates that the augmented training set closely preserves the statistical properties of the original training data, confirming that the SMOTE procedure maintains the underlying data distributions without introducing distortion. The test dataset exhibits comparable central tendencies with slightly higher variability, which is expected due to its smaller sample size. Overall, the consistency observed across the training, augmented training, and test datasets supports their suitability for model development and independent validation.

4. Results

4.1. Linear Regression Models and Linear-Based Formulations

Six candidate models were considered (Table 4). For the exponential and log–log models, the response was fitted in logarithmic space and mapped back using the exponential function. For log-input models, a small constant $\varepsilon = 10^{-6}$ was added where needed to avoid undefined logarithms.

Table 4. Regression model families and functional forms.

Model	Functional Form
Linear	$\hat{R}_w = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$
Exponential	$\ln(\hat{R}_w) = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3$
Interaction	$\hat{R}_w = b_0 + \sum b_i X_i + \sum b_{ij} X_i X_j$
Quadratic (2nd order)	$\hat{R}_w = b_0 + \sum b_i X_i + \sum b_{ij} X_i X_j + \sum b_{ii} X_i^2$
Log-linear	$\hat{R}_w = c_0 + \sum c_i \ln(X_i + \varepsilon)$
Log–log	$\ln(\hat{R}_w) = d_0 + \sum d_i \ln(X_i + \varepsilon)$

All coefficients were estimated using Ordinary Least Squares (OLS). Tables 5 and 6 report the fitted parameters. For interaction and quadratic models, the following basis ordering was used:

1. Interaction model: $[1, X_1, X_2, X_3, X_1 X_2, X_1 X_3, X_2 X_3]$.
2. Quadratic model: $[1, X_1, X_2, X_3, X_1 X_2, X_1 X_3, X_2 X_3, X_1^2, X_2^2, X_3^2]$.

Table 5. Estimated coefficients for primary and log-based models.

Parameter	Linear	Exponential ($\ln R_w$)	Log-Linear ($\ln X$)	Log-Log ($\ln R_w, \ln X$)
Intercept	−13.349120	0.881634	−181.174212	−9.695069
$X_1 / \ln(X_1)$	0.047630	0.002819	0.136474	0.008573
$X_2 / \ln(X_2)$	0.027944	0.001822	24.960448	1.634241
$X_3 / \ln(X_3)$	0.008462	0.000422	4.933264	0.241707

Table 6. Estimated coefficients for interaction and quadratic models (explicit terms).

Term	Interaction Model	Quadratic Model
Intercept	20.066029	−34.871199
X_1	0.148594	0.063052
X_2	−0.014918	0.105695
X_3	−0.050384	−0.028955
$X_1 X_2$	-6.1916×10^{-5}	3.4669×10^{-6}
$X_1 X_3$	-9.2818×10^{-5}	-5.1753×10^{-5}
$X_2 X_3$	7.7741×10^{-5}	6.5718×10^{-5}
X_1^2	—	-1.7589×10^{-5}
X_2^2	—	-6.7252×10^{-5}
X_3^2	—	-8.7535×10^{-6}

Models were evaluated on the independent test set using R^2 , RMSE, MAE, and MAPE (Table 7). The quadratic model achieved the best predictive performance and was selected as the optimal linear-based formulation.

Table 7. Test set performance comparison.

Model	R^2	RMSE	MAE	MAPE (%)
Quadratic	0.9422	2.0401	1.4034	9.0621
Interaction	0.8287	3.5116	2.6455	23.856
Linear	0.6383	5.1029	3.5124	17.808
Log-linear	0.6380	5.1046	2.9027	13.891
Exponential	0.5899	5.4332	3.6559	19.104
Log-log	0.5773	5.5162	3.1906	15.510

The quadratic model provided the highest predictive accuracy ($R^2 = 0.942$) and the lowest error levels (RMSE = 2.040, MAE = 1.403, MAPE = 9.06%), indicating that the R_w response is governed by nonlinear curvature and coupled effects among predictors. Log-transformed variants did not outperform the polynomial expansion, implying that logarithmic scaling alone is insufficient to capture the observed nonlinearities.

The optimal regression architecture (Quadratic) was additionally evaluated to verify robustness and to ensure that the superior predictive performance was not an artifact of a particular split or augmentation step. The model was therefore assessed on four datasets (Table 8): (i) the real (non-augmented) training subset, (ii) the SMOTE-augmented training subset, (iii) the independent test subset, and (iv) all original data (full dataset without synthetic samples). Across all cases, the model maintained high accuracy and low error, confirming stable generalization (Table 8).

Table 8. Robustness evaluation of the optimal model (Quadratic).

Model	Dataset	R^2	RMSE	MAE	MAPE (%)
Quadratic	Train (real)	0.9704	1.2418	1.0686	7.1590
Quadratic	Train (SMOTE)	0.9772	0.91386	0.71964	4.3919
Quadratic	Test	0.9422	2.0401	1.4034	9.0621
Quadratic	All original	0.9601	1.5226	1.1678	7.7229

4.2. TreeBagger and Random Forest Ensemble Models

To improve the predictive accuracy of the weighted sound reduction index R_w , TreeBagger and Random Forest (RF) regression models were developed based on the bootstrap aggregation (bagging) framework. Both models were trained on the augmented training dataset in order to enhance robustness under limited data availability and to reduce sensitivity to small-sample effects.

Since TreeBagger and Random Forest models share the same methodological foundation, they were analyzed jointly. The key distinction between the two approaches lies in the treatment of input variables at each split: the TreeBagger model considers all available input variables when determining candidate splits, whereas the Random Forest model introduces additional randomness by selecting only a random subset of input variables at each split. This mechanism reduces correlation among individual trees and improves generalization performance.

Because the predictive performance of both models strongly depends on tree complexity and the degree of randomness introduced during training, a systematic grid search was conducted over two key hyperparameters:

- **Minimum leaf size (L):** $L = 1, 2, \dots, 10$. This parameter controls the minimum number of observations in each terminal node and therefore regulates tree depth and model complexity.
- **Number of predictors sampled at each split ($mtry$):** $mtry = 1, 2, 3$. This parameter defines how many input variables are randomly considered for candidate splits, affecting both diversity among trees and bias–variance trade-offs.

For all configurations, the number of trees was fixed at 500 to ensure stable ensemble performance. During the grid search, candidate models were assessed on the test set, and the configuration yielding the best generalization performance was selected.

The grid-search landscape reveals a consistent dependence of error metrics on ensemble structure.

In general, increasing L (i.e., enforcing larger terminal nodes and shallower trees) increased RMSE, indicating reduced predictive fidelity (Figure 9). This behavior suggests that, for the present dataset, deeper trees (smaller leaf sizes) are required to capture interactions between the predictors (granulometric composition, density, PU_ADH) and the target response R_w .

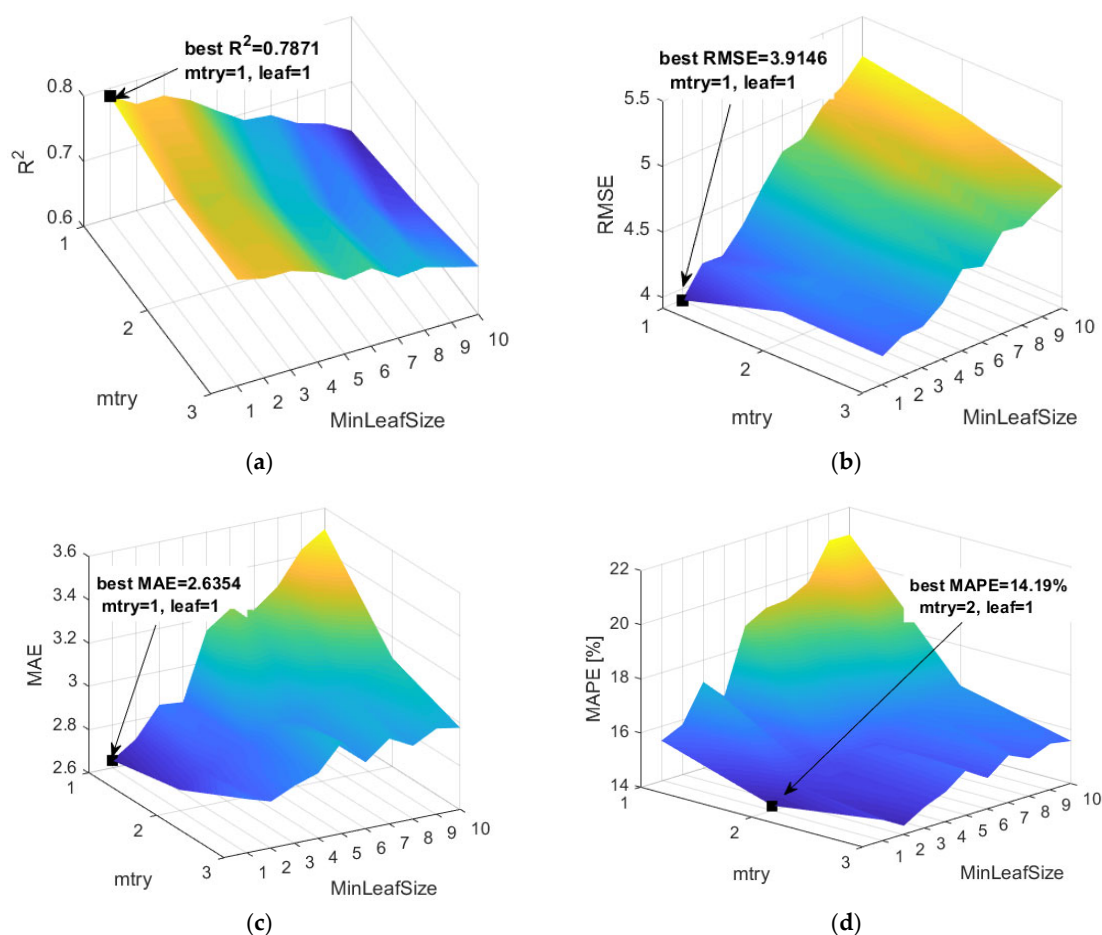


Figure 9. Grid-search results for TreeBagger and Random Forest models evaluated on the independent test datasets. (a) R^2 , (b) RMSE, (c) MAE, and (d) MAPE, as functions of the minimum leaf size (L) and the number of randomly selected variables at each split ($mtry$).

After selecting the optimal hyperparameter combination ($mtry = 1$, $leaf = 1$), the final RF model was evaluated on four datasets: the original real training subset, the augmented training subset, the unseen test subset, and the combined dataset (Table 9).

Table 9. Performance of the optimized RF model across datasets.

Dataset	R^2	RMSE	MAE	MAPE (%)
Train (real)	0.9399	1.7711	1.2620	8.635
Train (SMOTE)	0.9706	1.0419	0.7257	5.018
Test	0.7871	3.9146	2.6354	15.480
All (Original)	0.8838	2.5977	1.6689	10.663

4.3. Gradient Boosted Regression Tree Ensemble Model

To model the complex nonlinear dependency between the material descriptors and the sound reduction index (R_w), a Gradient Boosted Regression Tree (GBRT) ensemble was employed. Specifically, the LBoost (least-squares boosting) algorithm was adopted, which minimizes the mean squared error (MSE) by sequentially adding regression trees that approximate the residuals of the current ensemble. In this additive framework, each boosting iteration refines the predictor by correcting the errors remaining after the previous trees, thereby enabling the model to capture high-order nonlinearities and interaction effects without explicit feature engineering.

The predictive performance of LBoost is strongly governed by its hyperparameters, which jointly control the bias–variance trade-off and the effective complexity of the ensemble. Therefore, a multi-dimensional grid search was performed over three key parameters:

Learning rate (ν): $\nu \in \{0.001, 0.01, 0.1, 0.25, 0.5, 0.75, 1.0\}$.

The learning rate scales the contribution of each newly added tree (shrinkage). Smaller values generally improve stability and generalization but require more boosting iterations to reach optimal performance.

Maximum number of splits (MaxNumSplits): $[1, 2, 4, 8, 16]$.

This parameter controls the complexity of each regression tree (weak learner), with larger values permitting deeper, more expressive trees capable of capturing intricate non-linear patterns.

Number of learning cycles (T): $[50, 100, 200, 300, 400, 500]$.

This represents the total number of sequential trees in the ensemble, i.e., the boosting iterations.

The grid search revealed marked interactions between the learning rate and tree complexity. Performance trends were visualized using 3D surface plots in the (ν , MaxNumSplits) plane at the tree count corresponding to the globally optimal RMSE ($T = 500$).

- **Learning-rate sensitivity:** Very small learning rates ($\nu = 0.001$) produced a severely degraded test performance, indicating underfitting: the shrinkage was too strong for the ensemble to converge to an adequate solution within the predefined number of learning cycles.
- **Effect of tree complexity:** Increasing MaxNumSplits consistently improved predictive performance within the effective learning-rate regime (notably $\nu \approx 0.01$), implying that deeper base learners were necessary to capture the nonlinear interactions embedded in the dataset.

The best-performing configurations are summarized below (Table 10).

Table 10. Optimal hyperparameters and best test set performance by metric.

Metric	Best Value	Optimal Configuration
R^2	0.8332	$T = 500$, MaxNumSplits = 16, $\nu = 0.01$
RMSE	3.4649	$T = 500$, MaxNumSplits = 16, $\nu = 0.01$
MAE	2.5170	$T = 50$, MaxNumSplits = 2, $\nu = 0.25$
MAPE	13.02%	$T = 50$, MaxNumSplits = 2, $\nu = 0.25$

The RMSE-optimal model ($\nu = 0.01$, $T = 500$, MaxNumSplits = 16) follows the classical boosting principle of a small learning rate with many iterations, which typically yields stable generalization and reduces the impact of occasional large residuals (Figure 10).

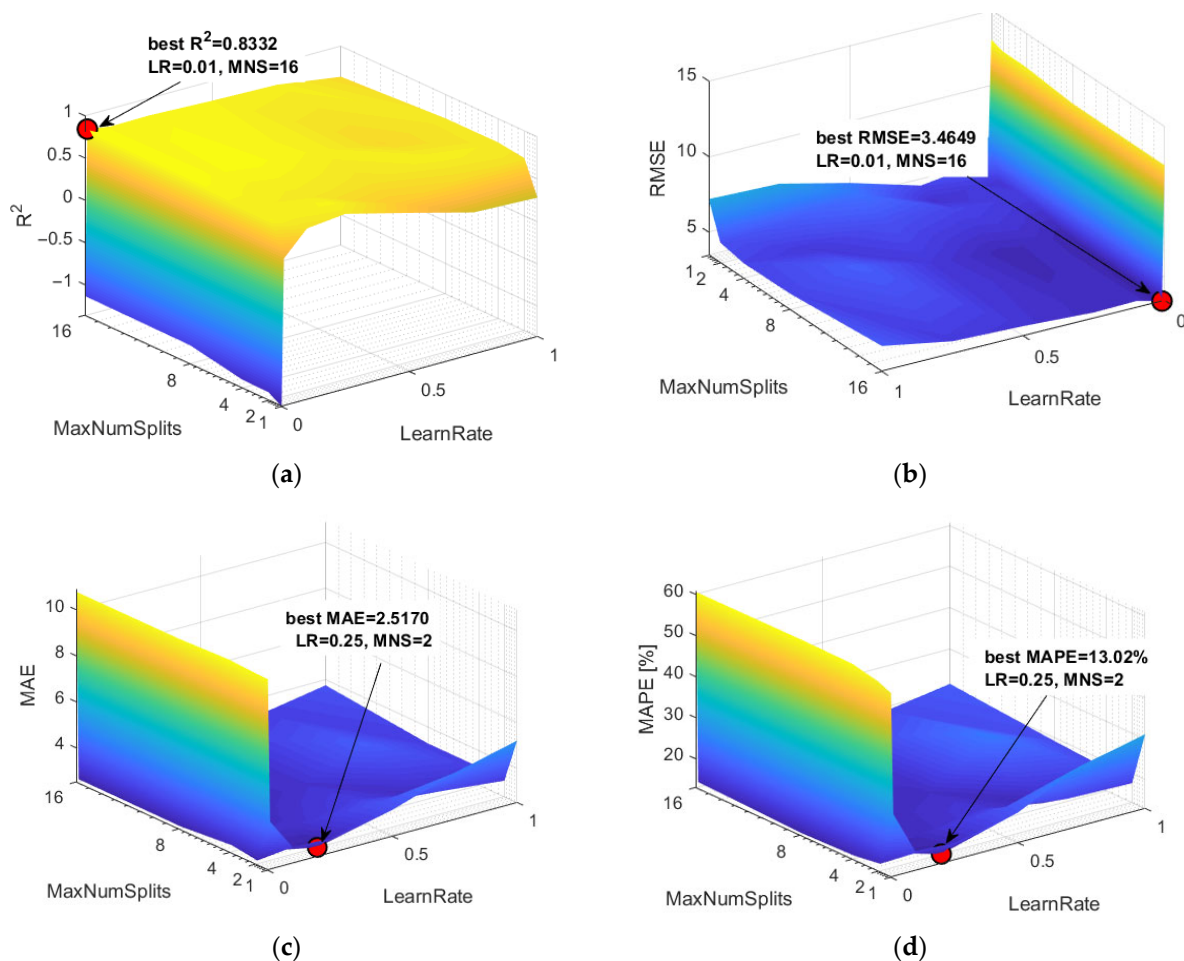


Figure 10. Grid-search results for Gradient Boosted Regression Tree models evaluated on the independent test dataset showing the variation of (a) R^2 , (b) RMSE, (c) MAE, and (d) MAPE as functions of the maximum number of splits (MaxNumSplits) and the learning rate (ν).

In contrast, MAE and MAPE favored a more aggressive learning rate with shallower trees ($\nu = 0.25$, $T = 50$, MaxNumSplits = 2), suggesting that this setting better reduces the typical (average absolute/relative) deviation, even if it is not optimal for penalizing larger errors (Figure 10). Table 11 presents a performance comparison of LSBoost models across training, training augmented, test, and combined datasets.

Table 11. Performance comparison of LSBoost models across training, training augmented, test, and combined datasets.

Model	Selection	Dataset	Num. Trees	Max Num. Splits	Learn Rate	R ²	RMSE	MAE	MAPE (%)
LSBoost	Best By RMSE	Train (real)	500	16	0.01	0.9763	1.1122	0.5309	2.8279
LSBoost	Best By RMSE	Train (SMOTE)	500	16	0.01	0.9912	0.5706	0.2712	1.5412
LSBoost	Best By RMSE	Test	500	16	0.01	0.8332	3.4649	2.6068	14.174
LSBoost	Best By RMSE	All	500	16	0.01	0.9237	2.1042	1.1460	6.1897
LSBoost	Best By MAE	Train (real)	50	2	0.25	0.9238	1.9937	1.3381	9.7375
LSBoost	Best By MAE	Train (SMOTE)	50	2	0.25	0.9504	1.3536	0.9367	6.4473
LSBoost	Best By MAE	Test	50	2	0.25	0.7746	4.0283	2.5170	13.02
LSBoost	Best By MAE	All	50	2	0.25	0.8690	2.7577	1.6874	10.71

The MAE-optimized LSBoost model yields smaller average absolute deviations across datasets, whereas the RMSE-optimized model generalizes better on the test set and reduces large-error risk, as reflected by a lower RMSE and higher R² (Table 11).

Overall, both selected LSBoost models demonstrate substantially improved predictive capability relative to traditional linear regression formulations, highlighting the importance of accounting for nonlinearities and interaction effects when modeling acoustic insulation indicators from material–property descriptors.

4.4. Symbolic Regression Modeling via Genetic Programming (GPTIPS)

To obtain an explicit analytical model for predicting the Sound Reduction Index R_w from the available material descriptors, multigene symbolic regression was performed using GPTIPS 2 [24–26]. Symbolic regression was selected because it can simultaneously capture strongly nonlinear relationships and variable interactions and return an interpretable closed-form expression, which is suitable for reporting and re-use in engineering practice.

Prior to model training, the input variables were rescaled using a min–max normalization procedure to map each predictor into the interval [0, 1]. Importantly, the normalization parameters were computed exclusively from the augmented training set, which served as the reference set in order to avoid any information leakage from the test set. The column-wise bounds were computed, and the normalization was applied as (9):

$$\tilde{x}_{ij} = \frac{x_{ij} - x_j^{\min}}{(x_j^{\max} - x_j^{\min}) + \varepsilon}, \quad (9)$$

where $x_j^{\min}$ and $x_j^{\max}$ denote the minimum and maximum values of the j-th predictor column, respectively, and $\varepsilon = 10^{-12}$ was added to the denominator to ensure numerical stability in the unlikely event that range is equal to zero. This normalization step is particularly important in symbolic regression when exponential, logarithmic, and power operators are allowed because it reduces the risk of extreme intermediate values and improves the stability of the evolutionary search. This design ensures that the genetic programming process optimizes models on the augmented training data, while model performance is verified on the same independent test set implemented to all models to quantify generalization.

All symbolic regression runs were conducted with reproducible stochastic behavior by fixing the random stream using `rng('default')`. The evolutionary search used a population size of 100 individuals and proceeded for 100 generations, repeated over 10 independent runs. Multiple runs were employed to reduce sensitivity to random initialization and to

increase the likelihood of locating high-quality models in the highly multimodal search space typical of symbolic regression.

A multigene representation was adopted with a maximum of five genes. In this framework, each gene corresponds to a symbolic expression tree, and the final predictor is formed as a weighted combination of these gene outputs. This structure increases modeling flexibility while retaining interpretability, as each gene may represent a distinct physical effect or interaction term.

To limit model complexity and mitigate overfitting, the maximum tree depth was constrained to 3. This restriction acts as an explicit structural regularization mechanism, ensuring that evolved expressions remain compact and scientifically interpretable.

Selection was performed using tournament selection with tournament size 2, providing moderate selection pressure and preserving diversity. Additionally, Pareto-based selection pressure was introduced through (`gp.selection.tournament.p_pareto` = 0.7), which promotes a trade-off between model accuracy and complexity and supports the discovery of parsimonious solutions. Elitism was enabled by retaining the best 5% of individuals to prevent the loss of top solutions between generations.

The function set was selected to enable both physically plausible transformations and nonlinear interactions commonly observed in material–acoustic phenomena. The allowed operators included: times, minus, plus, `rdivide` (restricted division), square, `add3` (sum of three terms), `exp`, `log`, `mult3` (multiplicative combination of three terms), `sqrt`, cube, and power.

The grid search shows strong underfitting on the test set in the simplest setting ($g = 1$, $d = 1$), where $R^2 = 0.44$ and $RMSE = 6.37$. Increasing depth and/or the number of genes generally improves accuracy by capturing stronger nonlinearities (e.g., for $g = 3$, $d = 3$: $R^2 = 0.91$, $RMSE = 2.54$).

The best overall result is achieved at $g = 4$, $d = 3$ ($R^2 = 0.993$, $RMSE = 0.69$, $MAE = 0.55$, $MAPE = 3.68\%$), indicating a sufficient model capacity to represent the underlying relationship. Moving to $g = 5$ does not further improve the optimum, suggesting diminishing returns beyond $g = 4$ at depth 3 (Figure 11).

The Pareto front shows the trade-off between model complexity (x-axis) and training error $1 - R^2$ (Figure 12). Each point represents one GP-derived equation; the Pareto-optimal solutions forming the front are highlighted in green, meaning no other model is simultaneously simpler and more accurate on the training set. The red point marks the selected optimal training set solution. Models near the “knee” of the green front provide the best accuracy–interpretability balance and were therefore considered primary candidates for final selection and subsequent test set validation.

Figure 13 lists representative Pareto-optimal symbolic regression models obtained by GPTIPS on the training set, reporting each model’s ID, goodness-of-fit (R^2), and expression complexity together with the explicit analytical form. Subsequently, only seven models exhibiting comparable predictive accuracy (highlighted in red) were subjected to further analysis.

Among the high- R^2 Pareto candidates, ID = 103 and ID = 166 provide the strongest evidence of robust generalization because they simultaneously achieve a very high test accuracy ($R^2_{\text{test}} \approx 0.993\text{--}0.994$) and the lowest test errors ($RMSE \approx 0.65\text{--}0.69$, $MAPE \approx 3.5\text{--}3.7\%$). In contrast, models such as ID = 567/815/820/925/928 show a clear train $\rightarrow$ test performance drop (higher $RMSE/MAE/MAPE$ on test), indicating reduced transferability to unseen samples.

Robustness is defined as the consistency of R^2 , $RMSE$, MAE , and $MAPE$ values across different datasets (Figure 14).

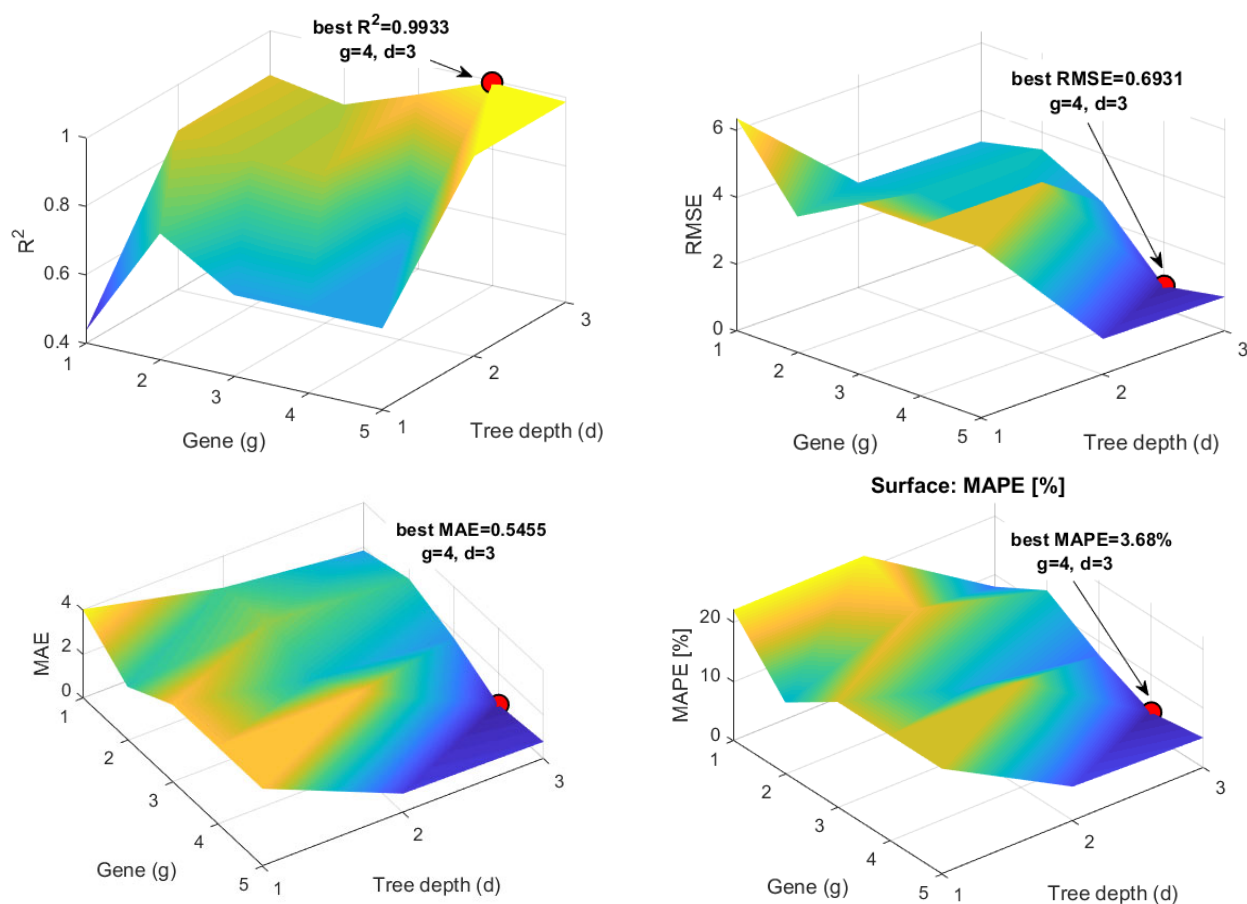


Figure 11. Influence of the number of genes and tree depth on the predictive performance of MGGP models evaluated on the independent test dataset.

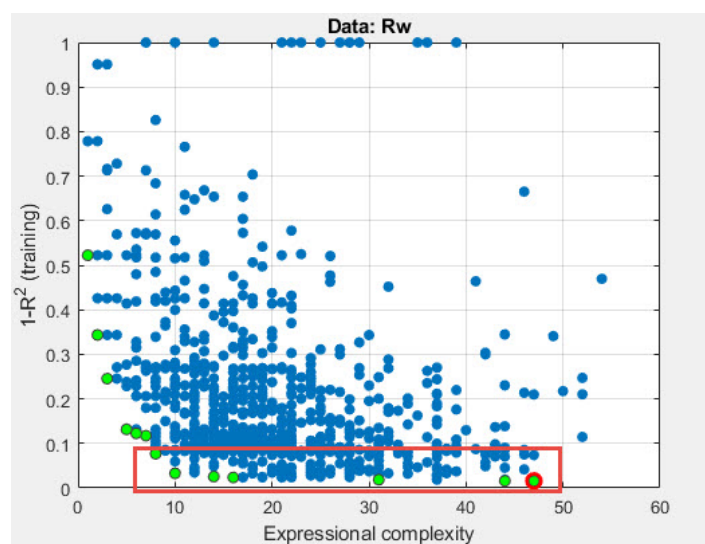


Figure 12. Pareto front of MGGP models illustrating the trade-off between model accuracy and complexity.

Model ID	Goodness of fit (R^2)	Model complexity	Model
103	0.984	44	53.8 SG + 53.8 PU_ADH SG ² + 3.87 GC SG ^{3/2} - 49.0 SG ^{3/2} - 17.1 GC PU_ADH ³ SG + 5.12
166	0.984	47	53.7 SG + 53.7 PU_ADH SG ² + 3.82 GC SG ^{3/2} - 48.7 SG ^{3/2} - 17.1 GC PU_ADH ⁴ SG + 5.11
567	0.981	31	5.94 GC + 40.5 SG - 20.0 exp(SG) - 0.683 exp(GC + 2.0 PU_ADH) + 34.6 PU_ADH SG + 27.4
815	0.976	16	16.8 exp(SG) + 25.1 PU_ADH SG - 21.4 SG ³ - 5.1 PU_ADH ⁴ GC - 5.51
820	0.975	14	46.4 SG - 22.3 exp(SG) + 25.3 PU_ADH SG - 5.12 PU_ADH ⁴ GC + 33.0
925	0.967	10	3.53 GC + 22.4 SG + 19.8 PU_ADH SG - 13.0 SG ³ + 5.79
928	0.967	10	3.53 GC + 22.4 SG + 19.8 PU_ADH SG - 13.0 SG ³ + 5.79
817	0.923	8	8.37 PU_ADH + 32.5 SG + 1.9 exp(GC) - 17.4 SG ³ + 0.285
987	0.883	7	4.54 GC + 8.13 SG + 23.7 PU_ADH SG + 8.34
72	0.878	6	8.04 PU_ADH + 35.8 SG - 19.6 SG ³ + 2.67
991	0.878	6	8.04 PU_ADH + 35.8 SG - 19.6 SG ³ + 2.67
672	0.868	5	8.18 PU_ADH + 47.9 SG - 30.4 SG ² + 1.68
239	0.755	3	4.76 GC + 8.83 PU_ADH + 14.4 SG + 5.5
920	0.657	2	8.8 PU_ADH + 15.9 SG + 6.97
227	0.478	1	13.6 SG + 11.1

Figure 13. Screenshot from the GPTIPS software showing Pareto-optimal MGGP models, including goodness of fit, model complexity, and analytical expressions.

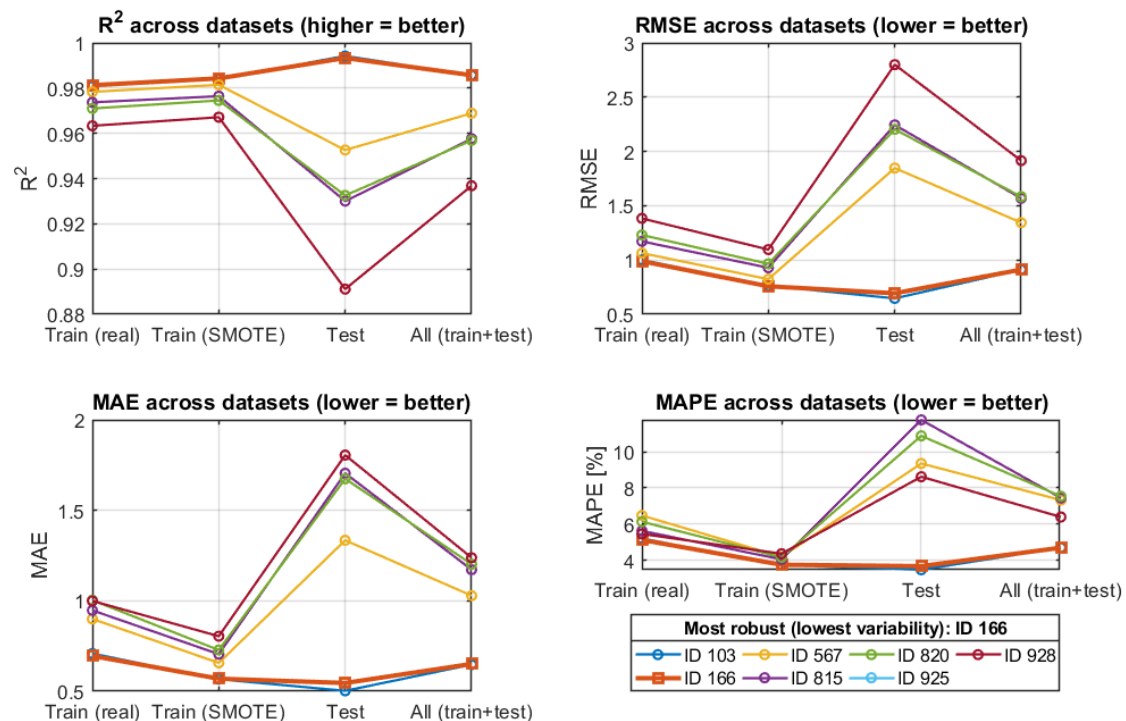


Figure 14. Robustness comparison of MGGP models across different datasets.

Large discrepancies (e.g., excellent training performance but degraded test performance) indicate dataset sensitivity, suggesting potential overfitting and/or dependence on the underlying data distribution. Conversely, similar values across datasets indicate a stable performance and consistent generalization, i.e., higher robustness (Figure 14).

For each candidate model i , four values of the same performance metric are available (10), evaluated on different datasets (Train–real, Train–SMOTE, Test, and All), e.g.,

$$R^2_{i, \text{train_real}}, R^2_{i, \text{train_smote}}, R^2_{i, \text{test}}, R^2_{i, \text{all}}, \quad (10)$$

and analogously for RMSE, MAE, and MAPE.

For each metric, robustness is operationalized as the standard deviation across the four datasets. For RMSE (analogously for MAE and MAPE), for model i (11),

$$s_{\text{RMSE}, i} = \text{std.}(\text{RMSE}_{i, \text{train_real}}, \text{RMSE}_{i, \text{train_smote}}, \text{RMSE}_{i, \text{test}}, \text{RMSE}_{i, \text{all}}) \quad (11)$$

A smaller $s_{RMSE,i}$ indicates that RMSE remains similar across datasets, implying a more stable model. Likewise, for (12) and (13),

$$s_{MAE,i} = std.(MAE_{i,train_real}, MAE_{i,train_smote}, MAE_{i,test}, MAE_{all}). \quad (12)$$

$$s_{MAPE,i} = std.(MAPE_{i,train_real}, MAPE_{i,train_smote}, MAPE_{i,test}, MAPE_{all}). \quad (13)$$

RMSE, MAE, and MAPE are error measures (lower is better), whereas R^2 is a goodness-of-fit measure (higher is better). To treat all criteria consistently in a “lower-is-better” form, R^2 is transformed into an error-like quantity (14):

$$e_{R^2} = 1 - R^2. \quad (14)$$

Robustness with respect to R^2 then could be quantified as (15):

$$s_{R^2,i} = std.(1 - R^2_{i,train_real}, 1 - R^2_{i,train_smote}, 1 - R^2_{i,test}, 1 - R^2_{all}). \quad (15)$$

Because the metrics have different numerical scales, the standard deviations for metrics i are normalized by the mean dispersion across all analyzed models (16):

$$\hat{s}_{m,i} = \frac{S_{m,i}}{mean(S_{m,i})}, m \in \{R^2, RMSE, MAE, MAPE\}. \quad (16)$$

A single robustness score is then defined as the sum (17):

$$rob_i = \hat{s}_{R^2,i} + \hat{s}_{RMSE,i} + \hat{s}_{MAE,i} + \hat{s}_{MAPE,i}. \quad (17)$$

Lower rob_i values indicate that all metrics remain more consistent across datasets, i.e., a more robust model. The most robust model (Figure 14) is selected as (18):

$$i^* = \operatorname{argmin}(rob_i). \quad (18)$$

Figure 15 illustrates the robustness of Pareto-optimal MGGP models by quantifying performance variability across the training, augmented training, and independent test datasets. Models with minimal cross-dataset error variation are identified as robust, indicating stable generalization under limited experimental data. This analysis complements accuracy-based Pareto selection and supports the identification of MGGP models that are both accurate and reliable for the practical prediction of airborne sound insulation.

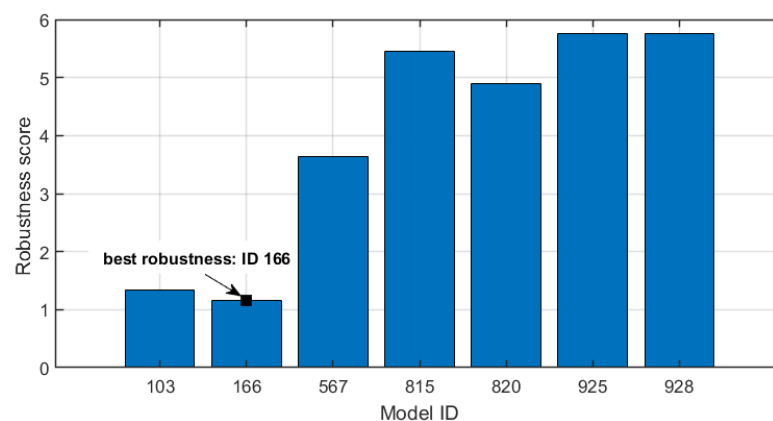


Figure 15. Quantitative robustness assessment of Pareto-optimal MGGP models based on cross-dataset performance variability.

A mean-based Shapley value analysis [34,35] was applied to assess the global importance of input variables in the optimal (Model ID 166) MGGP symbolic regression model. The analysis was performed on the complete normalized dataset using the mean of all inputs as the baseline reference. Shapley values were computed by averaging the marginal contribution of each variable across all possible combinations of predictors, ensuring a fair attribution of both individual and interaction effects (Figure 16).

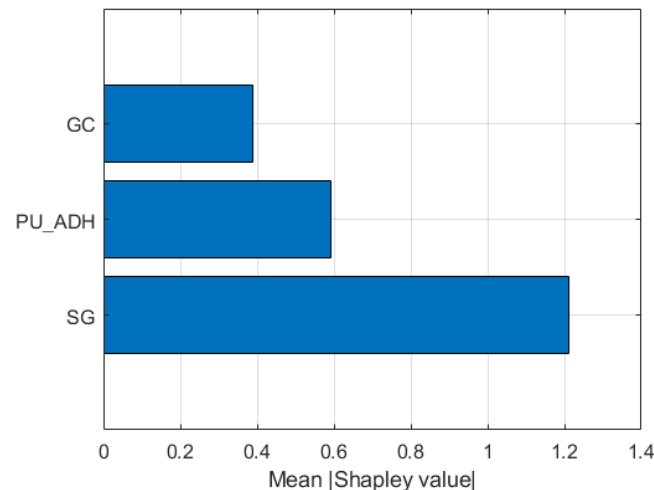


Figure 16. Mean-based Shapley feature importance for the selected MGGP model evaluated on all available data.

The results show (Figure 16) that SG is the most influential variable, accounting for approximately 55% of the total absolute Shapley importance, followed by PU_ADH with about 27%, while GC exhibits the smallest contribution (approximately 18%). Overall, the analysis confirms that the model predictions are primarily driven by SG, with secondary influence from PU_ADH and a weaker but positive contribution from GC.

Partial Dependence Plot (PDP) and Individual Conditional Expectation (ICE) plots (Figure 17) were produced to interpret the MGGP symbolic model using the entire normalized dataset (inputs scaled to [0, 1]).

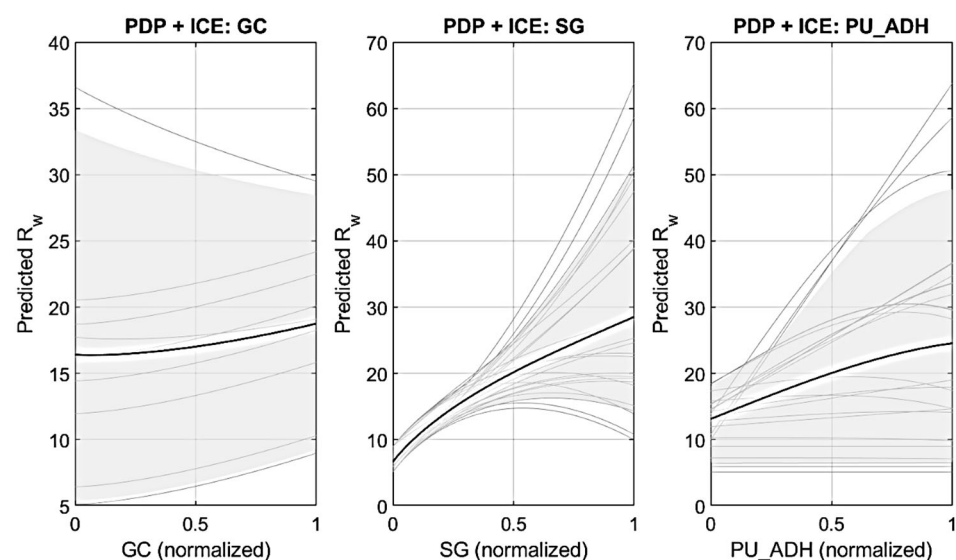


Figure 17. PDP (bold) and ICE (gray) for GC, SG and PU_ADH using all normalized data [0, 1] the shaded area denotes the 10–90% prediction band.

For each feature (GC, SG, PU_ADH), the feature was varied over a uniform grid in [0, 1] while all other inputs were kept at their observed values, and predictions were recomputed. ICE curves show the sample-wise response $\hat{R}_w$ to the feature change, while the PDP (bold line) is the mean of ICE curves, representing the average marginal effect. The shaded band corresponds to the 10–90% prediction range, summarizing response heterogeneity and interaction effects.

GC shows a relatively weak average influence (small PDP change). SG and PU_ADH show stronger positive trends, with increasing ICE spread at higher values, indicating that their impact on R_w is nonlinear and interaction-dependent (effect varies across samples).

The final equation of the optimal refined model (Model ID 166), expressed in a simplified form, is given as follows (19):

$$R_w = 53.7 \times SG + 53.7 \times SG^2 \times PU_{ADH} + 3.82 \times GC^{\frac{3}{2}} - 48.7 \times SG^{\frac{3}{2}} - 17.1 \times GC \times SG \times PU_{ADH}^4 + 5.11 \quad (19)$$

where Equations (20)–(22) define the normalized input variables as

$$GC = GC_{true} / (100 + \varepsilon), \quad (20)$$

$$SG = (SG_{true} - 585) / (515 + \varepsilon), \quad (21)$$

$$PU_{ADH} = (PU_{ADH}_{true} - 296) / (1044 + \varepsilon). \quad (22)$$

Here, GC_{true} , SG_{true} , and PU_{ADH}_{true} denote the original (non-normalized) input variables. Specifically, GC_{true} represents the granulometric composition expressed as the percentage of the 0–2 mm fraction, SG_{true} denotes the material density (kg/m^3), and PU_{ADH}_{true} is the polyurethane adhesive dosage (gr/mm^2). Value $\varepsilon = 10^{-12}$ was introduced to avoid a zero denominator. The structure of the MGGP model is presented in Figure 18.

The symbolic regression equation was then evaluated using these normalized variables, which allows the direct use of laboratory-scale measurements while preserving the model structure learned on the normalized domain [0, 1].

The comparative performance summary presented in Table 12 clearly highlights substantial differences in predictive accuracy among the investigated modeling approaches. Classical regression formulations (linear, interaction, logarithmic, and exponential models) exhibit a limited predictive capability, reflected by relatively low coefficients of determination and high error levels. Although the quadratic regression model improves accuracy by introducing curvature effects, its performance remains constrained by the fixed functional structure and limited ability to capture complex interactions among material parameters.

Tree-based ensemble methods (Random Forest and Gradient Boosted Trees) achieve notably better predictive accuracy by effectively modeling localized nonlinearities and variable interactions. However, despite their improved performance relative to regression-based models, their prediction errors remain higher than those of the symbolic regression approach, and their black-box nature limits interpretability and direct engineering applicability.

The multigene symbolic regression model (GPTIPS, ID 166) clearly outperforms all alternative models, achieving the highest coefficient of determination and the lowest RMSE, MAE, and MAPE values (Table 12). This result confirms that the symbolic model provides the most accurate and stable representation of the underlying relationship between material descriptors and the sound reduction index. Importantly, this superior accuracy is achieved while retaining an explicit analytical formulation, offering a unique combination of predictive performance and interpretability that is not attainable with conventional regression or ensemble learning techniques.

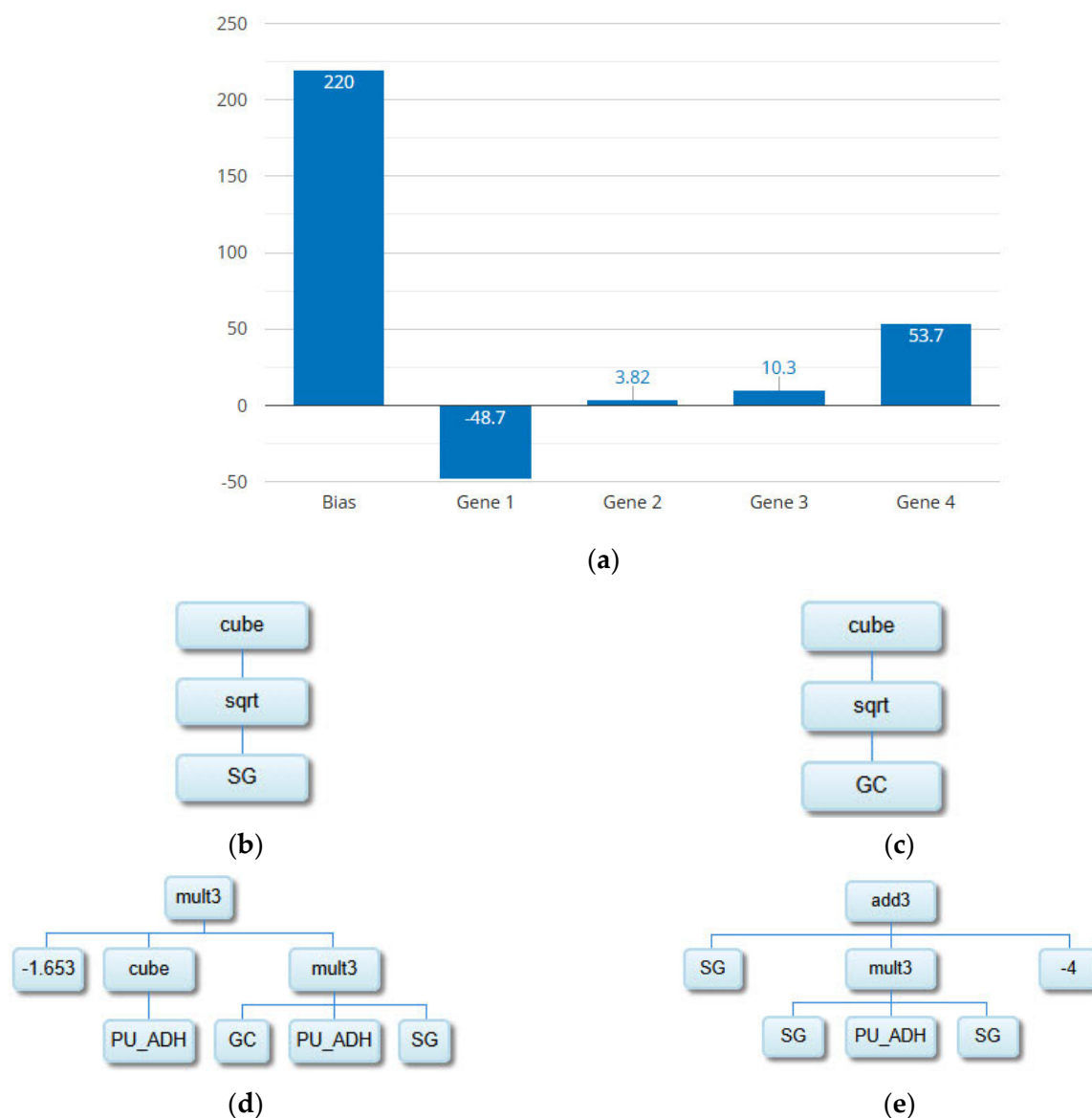


Figure 18. Internal structure of the optimal MGGP model obtained using GPTIPS: (a) weights of individual genes and bias term; (b–e) symbolic expression trees of Genes 1–4, respectively.

Table 12. Comparative performance of conventional regression and machine learning models on the independent test dataset.

Model	R ²	RMSE	MAE	MAPE (%)
Quadratic	0.9422	2.0401	1.4034	9.0621
Interaction	0.8287	3.5116	2.6455	23.856
Linear	0.6383	5.1029	3.5124	17.808
Log-linear	0.6380	5.1046	2.9027	13.891
Exponential	0.5899	5.4332	3.6559	19.104
Log–log	0.5773	5.5162	3.1906	15.510
Random Forest	0.7871	3.9146	2.6354	15.480
LS Boost best by RMSE	0.8332	3.4649	2.6068	14.174
LS Boost best by MAE	0.7746	4.0283	2.5170	13.02
GP ID 166	0.9933	0.6931	0.5455	3.6814

The residual statistics (Table 13) indicate stable and unbiased predictive behavior across all datasets. The mean residuals for the training, augmented training, and test sets are close to zero (-0.051 , -3.3×10^{-5} , and 0.050 , respectively), confirming the absence of systematic overestimation or underestimation of the weighted airborne sound reduction index R_w (Figure 19). A clear reduction in residual dispersion is observed from the original training set to the augmented training and test sets. The RMSE decreases from 1.001 for the training data to 0.764 for the augmented training data and further to 0.646 for the independent test data, indicating improved numerical stability and generalization following regression-oriented data augmentation, without evidence of overfitting. The Kolmogorov–Smirnov test confirms residual normality for all datasets ($p = 0.2597$, 0.5935 , and 0.8829 for training, augmented training, and test sets, respectively), suggesting that the prediction errors are predominantly random. Overall, the residual analysis demonstrates that the proposed framework yields statistically consistent and robust predictions under limited experimental data conditions.

Table 13. Residual statistics and normality assessment for the proposed model.

Dataset	Mean Residual	Standard Deviation	RMSE	KS p -Value
Training	-0.051	1.027	1.001	0.2597
Training + SMOTE	-3.3×10^{-5}	0.768	0.764	0.5935
Test	0.050	0.689	0.646	0.8829

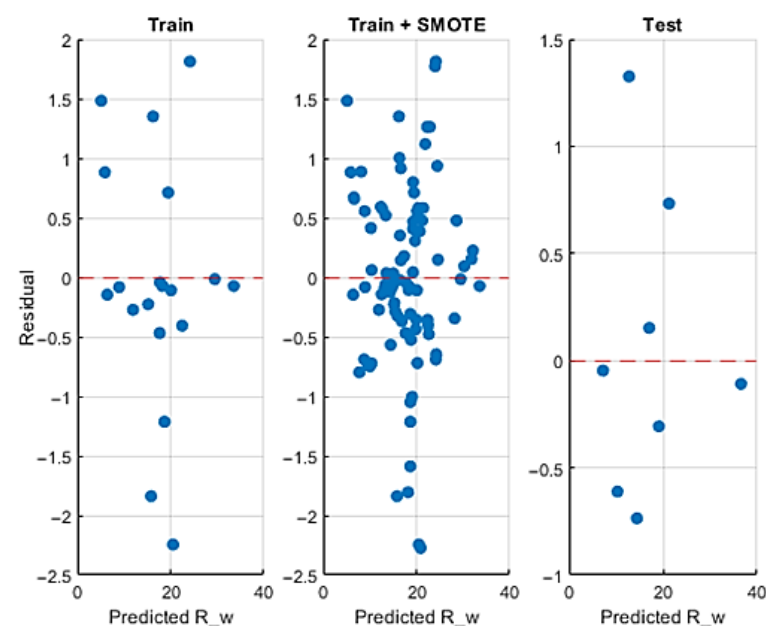
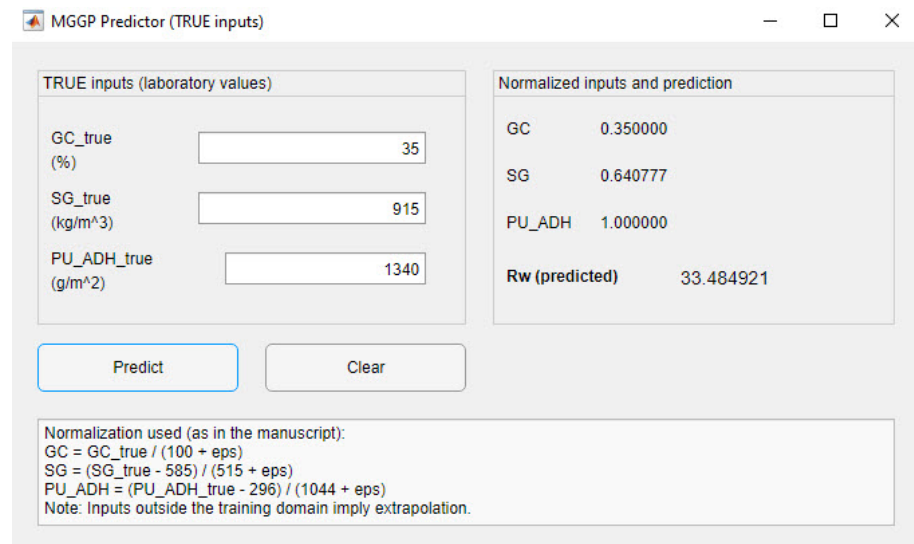


Figure 19. Residuals versus predicted values for the training, augmented training (SMOTE), and independent test datasets.

To support practical deployment, a MATLAB-based graphical user interface (GUI) was implemented to enable the direct input of laboratory-measured parameters and real-time evaluation of the proposed MGGP model (Figure 20).



MGGP Predictor (TRUE inputs)

TRUE inputs (laboratory values)

GC_true (%)

SG_true (kg/m³)

PU_ADH_true (g/m²)

Normalized inputs and prediction

GC 0.350000

SG 0.640777

PU_ADH 1.000000

Rw (predicted) 33.484921

Normalization used (as in the manuscript):
 GC = GC_true / (100 + eps)
 SG = (SG_true - 585) / (515 + eps)
 PU_ADH = (PU_ADH_true - 296) / (1044 + eps)
 Note: Inputs outside the training domain imply extrapolation.

Figure 20. MATLAB-based GUI for real-time prediction of the weighted airborne sound reduction index R_w using the MGGP model.

5. Conclusions

This study developed a comprehensive data-driven framework for predicting the weighted airborne sound reduction index R_w of recycled rubber–polyurethane composite panels. Through a systematic comparison of classical regression models, tree-based ensemble methods, and symbolic regression, it was demonstrated that acoustic insulation performance is governed by pronounced nonlinear relationships and interaction effects among the governing material descriptors.

While quadratic polynomial regression outperformed simple linear formulations by capturing basic curvature effects, ensemble learning approaches—namely Random Forest and Gradient Boosted Trees—achieved a higher predictive accuracy by modeling localized nonlinearities. However, the inherently black-box nature of these ensemble models limits their direct applicability in engineering-oriented design and interpretation.

The highest predictive accuracy and robustness were achieved using multigene symbolic regression. Importantly, despite the limited size of the original experimental dataset, the selected symbolic regression model exhibited stable and consistent predictive performance across multiple control datasets, including the real training set, the SMOTE-augmented training set, the independent test set, and the full original dataset. The close agreement of performance metrics across these datasets confirms that the model’s accuracy is not driven by a specific data split or augmentation step but reflects a genuine generalization capability. In addition, dataset representativeness was rigorously verified using multiple statistical criteria, including distributional similarity, mean consistency, variance homogeneity, and effect size measures, ensuring that both training–synthetic and training–test subsets remain statistically consistent.

The influence of input variables was quantitatively assessed using explainable machine learning techniques. Shapley value analysis identified material density as the dominant contributor to R_w (approximately 55%), followed by polyurethane adhesive dosage (about 27%), while granulometric composition exhibited a weaker but non-negligible influence (approximately 18%). These findings were further supported by PDP and ICE analyses, which revealed strong nonlinear and interaction-dependent effects for density and binder content.

Overall, the proposed symbolic regression model constitutes a practical and interpretable predictive tool for the preliminary design and optimization of recycled rubber acoustic panels. By reducing the reliance on extensive experimental testing and enabling in-

formed material–property optimization, the framework offers clear benefits for sustainable material development and early-stage engineering decision-making.

Despite the strong predictive performance, robustness, and interpretability achieved, several limitations should be acknowledged. The original experimental dataset is relatively small, which is a common constraint in laboratory-based acoustic testing. Nevertheless, robustness and statistical representativeness were explicitly validated through repeated data splitting, cross-dataset evaluation, and multiple statistical hypothesis tests, supporting the reliability of the proposed model within the investigated experimental domain. Furthermore, only a limited set of material descriptors was considered. While granulometric composition, density, and adhesive dosage captured the dominant effects governing R_w , additional microstructural and processing-related parameters—such as porosity, pore connectivity, compaction pressure, or curing conditions—may further enhance model generality and physical fidelity.

Future research will therefore focus on expanding the experimental database, validating the proposed framework on larger and industrial-scale panels, and extending the methodology toward frequency-resolved acoustic prediction. The integration of physics-informed constraints and hybrid modeling strategies also represents a promising direction for improving robustness and transferability under limited-data conditions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app16052410/s1>.

Author Contributions: Conceptualization, M.K., A.C., B.B. and P.P.; Methodology, M.K., B.B. and P.P.; Software, M.K.; Validation, M.K., A.C., B.B. and P.P.; Formal analysis, M.K.; Investigation, M.K. and A.C.; Resources, M.K., A.C. and B.B.; Data curation, M.K. and A.C.; Writing—original draft, M.K., A.C., B.B. and P.P.; Writing—review & editing, M.K., B.B. and P.P.; Visualization, M.K., B.B. and P.P.; Supervision, M.K., B.B. and P.P.; Project administration, M.K., A.C., B.B. and P.P.; Funding acquisition, M.K. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The experimental dataset supporting the findings of this study is available in the Supplementary Materials as an Excel file.

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

References

1. Hashamfirooz, M.; Dehghani, M.H.; Khanizadeh, M.; Aghaei, M.; Bashardoost, P.; Hassanvand, M.S.; Hassanabadi, M.; Momeniha, F. A systematic review of the environmental and health effects of waste tires recycling. *Heliyon* **2025**, *11*, e41909. [\[CrossRef\]](#)
2. Dabic-Miletic, S.; Simic, V.; Karagoz, S. End-of-life tire management: A critical review. *Environ. Sci. Pollut. Res. Int.* **2021**, *28*, 68053–68070. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Karaagaç, B.; Kalkan, M.E.; Deniz, V. End of life tyre management: Turkey case. *J. Mater. Cycles Waste Manag.* **2017**, *19*, 577–584. [\[CrossRef\]](#)
4. Kolendo, G.; Voronova, V.; Bumanis, G.; Korjamins, A.; Bajare, D. Life cycle assessment of end-of-life tire disposal methods and potential integration of recycled crumb rubber in cement composites. *Appl. Sci.* **2024**, *14*, 11667. [\[CrossRef\]](#)
5. Crnoja, A.; Kos, Ž. Današnja potreba recikliranja automobilskih guma i primjena reciklata u građevinarstvu. In *Matrib 2022: 22nd International Conference on Materials, Tribology and Recycling*; Bušić, M., Leder Horina, J., Tropša, V., Eds.; University North: Varaždin, Croatia, 2022; pp. 52–61.
6. Smirnova, O.M.; Menéndez Pidal de Navascués, I.; Mikhailevskii, V.R.; Kolosov, O.I.; Skolota, N.S. Sound-absorbing composites with rubber crumb from used tires. *Appl. Sci.* **2021**, *11*, 7347. [\[CrossRef\]](#)

7. ISO 10534-2; Acoustics—Determination of Acoustic Properties in Impedance Tubes—Part 2: Two-Microphone Technique for Normal Sound Absorption Coefficient and Normal Surface Impedance. ISO: Geneva, Switzerland, 2023.
8. Colom, X.; Girbau, J.; Marin, M.; Formela, K.; Saeb, M.R.; Carrillo, F.; Cañavate, J. New approach in the reuse of modified ground tire rubber as thermal and acoustic insulation to be used in civil engineering. *J. Mater. Cycles Waste Manag.* **2023**, *25*, 3557–3566. [\[CrossRef\]](#)
9. Badida, M.; Andrejiova, M.; Pinosova, M.; Moravec, M. Acoustic analysis of soundproofing materials using recycled rubber from automobiles. *Materials* **2025**, *18*, 3144. [\[CrossRef\]](#)
10. Algaifi, H.A.; Syamsir, A.; Baharom, S.; Alrshoudi, F.; Qaid, A.; Al-Fakih, A.M.; Mhaya, A.M.; Salah, H.A. Assessment of acoustic and mechanical properties in modified rubberized concrete. *Case Stud. Constr. Mater.* **2024**, *20*, e03063. [\[CrossRef\]](#)
11. Wang, R.; Yao, D.; Zhang, J.; Xiao, X.; Xu, Z. Identification of key factors influencing sound insulation performance of high-speed train composite floor based on machine learning. *Acoustics* **2024**, *6*, 1–17. [\[CrossRef\]](#)
12. Yao, D.; Zhang, J.; Lei, J.; Zhao, Z.; Zhang, Y.; Zhao, Y.; Pang, J.; Li, J. A comprehensive review of acoustic metamaterials: Applications and challenges for lightweight noise control in large-scale transportation. *Mater. Des.* **2025**, *260*, 115002. [\[CrossRef\]](#)
13. McCarthy, R.A.; Zhang, Y.; Verburg, S.A.; Jenkins, W.F.; Gerstoft, P. Machine Learning in Acoustics: A Review and Open-Source Repository. *npj Acoust.* **2025**, *1*, 18. [\[CrossRef\]](#)
14. Wilk-Jakubowski, J.L.; Kuchcinski, A.; Pawlik, L.; Wilk-Jakubowski, G. Advanced Sound Insulating Materials: An Analysis of Material Types and Properties. *Appl. Sci.* **2025**, *15*, 6156. [\[CrossRef\]](#)
15. EN ISO 10140-2:2010; Acoustics—Laboratory Measurement of Sound Insulation of Building Elements—Part 2: Measurement of Airborne Sound Insulation. International Organization for Standardization: Geneva, Switzerland, 2010.
16. Cimbola, Z.; Crnoja, A.; Barišić, I.; Netinger Grubeša, I. Evaluation of the applicability of waste rubber in insulation panels with regard to its grain size and panel thickness. *Materials* **2024**, *17*, 5251. [\[CrossRef\]](#) [\[PubMed\]](#)
17. EN ISO 717-1:2013; Acoustics—Rating of Sound Insulation in Buildings and of Building Elements—Part 1: Airborne Sound Insulation. International Organization for Standardization: Geneva, Switzerland, 2013.
18. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2009.
19. Kovačević, M.; Hadzima-Nyarko, M.; Petronijević, P.; Vasiljević, T.; Radomirović, M. Comparative analysis of machine learning models for predicting interfacial bond strength of fiber-reinforced polymer–concrete. *Computation* **2025**, *13*, 17. [\[CrossRef\]](#)
20. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
21. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
22. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
23. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [\[CrossRef\]](#)
24. Searson, D.P. GPTIPS 2: An open-source software platform for symbolic data mining. In *Handbook of Genetic Programming Applications*; Gandomi, A.H., Alavi, A.H., Ryan, C., Eds.; Springer: Cham, Switzerland, 2015; pp. 551–573. [\[CrossRef\]](#)
25. Searson, D.P.; Leahy, D.E.; Willis, M.J. GPTIPS: An open source genetic programming toolbox for multigene symbolic regression. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2010)*, Hong Kong, China, 17–19 March 2010; pp. 77–80.
26. Searson, D.P.; Willis, M.J.; Montague, G.A. Co-evolution of nonlinear PLS model components. *J. Chemom.* **2007**, *21*, 592–603. [\[CrossRef\]](#)
27. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
28. Torgo, L.; Ribeiro, R.P.; Pfahringer, B.; Branco, P. SMOTE for regression. In *Progress in Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 378–389.
29. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [\[CrossRef\]](#)
30. Massey, F.J. The Kolmogorov–Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [\[CrossRef\]](#)
31. Welch, B.L. The generalization of “Student’s” problem when several different population variances are involved. *Biometrika* **1947**, *34*, 28–35. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Levene, H. Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*; Olkin, I., Ed.; Stanford University Press: Stanford, CA, USA, 1960; pp. 278–292.
33. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.

34. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games II*; Kuhn, H.W., Tucker, A.W., Eds.; Princeton University Press: Princeton, NJ, USA, 1953; pp. 307–317.
35. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.