ORIGINAL ARTICLE

lajss.org© Latin American Journal of Solids and Structures

www.lajss.org

# Simulation-Admissibility-Aware Surrogate Optimization of Electric-Vehicle Side-Pole Crashworthiness

**Lijun Cao[a,b], Zhenjiang Wu[c], Zehui Huang[a,b*], Hongbin Tang[a,b], Jianjiao Deng[a,b], Shibin Wang[a,b], Ledan Liu[a,b], Qianxi Cheng[a,b], Xingbo Zhou[a,b]**

[a] FAW Global R&D Center, Changchun 130013, China. Email: caolijun@faw.com.cn, huangzehui1@faw.com.cn, tanghongbin@faw.com.cn, dengjianjiao@faw.com.cn, wangshibin@faw.com.cn, liuledan@faw.com.cn, chengqianxi@faw.com.cn, zhouxingbo@faw.com.cn

[b] National Key Laboratory of Advanced Vehicle Integration and Control, Changchun 130013, China.

[c] China Automotive Technology and Research Center Co., Ltd., Tianjin 300000, China. Email: wuzhenjiang@catarc.ac.cn

* Corresponding author

https://doi.org/10.1590/1679-7825/e9231

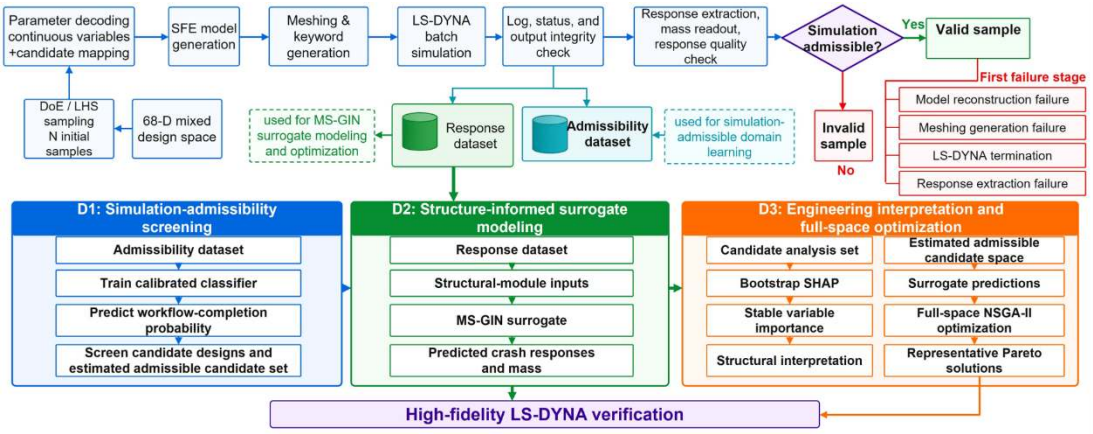
## Abstract

Crashworthiness surrogates are usually trained on successful finite element (FE) simulations, although automated full-vehicle workflows may include mixed-variable designs that fail during reconstruction, meshing, solution, or response extraction. This study proposes a simulation-admissibility-aware surrogate-assisted framework for electric-vehicle side-pole crashworthiness design. A battery-electric SUV was parameterized with 32 continuous geometry and 36 discrete material/thickness variables. Of 600 SFE–LS-DYNA designs, 528 were admissible. A workflow-admissibility classifier was evaluated by repeated stratified nested cross-validation. In the admissible domain, the structure-informed MS-GIN surrogate achieved a mean $R^2$ of 0.91 and NRMSE of 6.5%. Bootstrap SHAP identified influential variables in the B-pillar and lower body. NSGA-II optimization with FE verification yielded a compromise design reducing body mass, B-pillar chest intrusion, and battery-neighborhood intrusion by 3.6%, 9.5%, and 14.6%, while satisfying the other constraints. The framework captures admissibility risk, structural sensitivity, and lightweighting–crashworthiness trade-offs in mixed-variable crash design.

## Keywords

electric vehicle; side-pole impact; simulation-admissible domain; structure-informed surrogate model; multi-objective optimization.

## Graphical abstract




Received August 11, 2026. In revised form August 27, 2026. Accepted September 14, 2026. Available online September 16, 2026.
https://doi.org/10.1590/1679-7825/e9231

Latin American Journal of Solids and Structures. ISSN 1679-7825. Copyright © **2026**. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1 INTRODUCTION

Electric-vehicle side-pole crashworthiness is governed by a highly localized lateral load. The body structure must simultaneously protect the occupants and traction battery while maintaining structural lightweighting (Ikpe et al., 2017; Mortazavi Moghaddam et al., 2023; Setiawan and Salim, 2017). Because the available lateral deformation space is limited, the B-pillar, sill, upper rail, floor, seat cross-member, and battery-adjacent structures act as a coupled load-transfer system during impact (Chombo et al., 2021; Harrison et al., 2020; Long et al., 2025). Their geometry, material, and thickness therefore need to be considered jointly rather than optimized independently at the component level.

Full-vehicle crash finite-element (FE) analyses are computationally expensive, making direct exploration of high-dimensional mixed design spaces impractical. Surrogate-assisted optimization has therefore been widely used in crashworthiness design, employing response surfaces, Kriging, support vector regression, neural networks, and multi-objective evolutionary algorithms to approximate crash responses and identify Pareto trade-offs (Fang et al., 2005; Fang et al., 2017; Gao et al., 2019; Lian et al., 2025; Lualdi et al., 2023; Qiu et al., 2018). Most existing implementations, however, construct response surrogates using only successfully completed simulations and implicitly assume that candidate designs can be evaluated through the prescribed numerical workflow (Chen et al., 2024; Liu et al., 2018a). This assumption may not hold in automated full-vehicle optimization involving coupled geometry, material, and thickness changes.

Some candidate designs may fail during parametric-model reconstruction, mesh or keyword generation, numerical solution, or response extraction (Mazumdar et al., 2024; Müller and Day, 2019). These outcomes can be treated as hidden or unrelaxable constraints in simulation-based optimization (Audet et al., 2020; Le Digabel and Wild, 2024). Failed evaluations may be handled retrospectively through penalty objective or constraint values, whereas classification or evaluability prediction provides an established alternative for using observed success/failure information to guide subsequent search (Amini et al., 2025; Niehoff et al., 2024). Importantly, failures arising at different stages of the workflow do not necessarily share a common physical mechanism; their common relevance here is whether the prescribed automated FE workflow can be completed reliably. Accordingly, simulation admissibility is treated in this study as a workflow-level numerical constraint rather than as a unified physical failure boundary.

Reliable response prediction within the numerically admissible domain presents a second challenge. Vehicle-body design variables are mixed, with geometric variables being continuous and material/thickness selections being discrete. Discrete selections should therefore be represented through their relevant mechanical properties rather than arbitrary categorical indices (Garrido-Merchán and Hernández-Lobato, 2020; Moustapha et al., 2022; Yang et al., 2022; Zhu et al., 2024). In addition, side-pole responses arise from interactions among connected structural regions, whereas conventional flat surrogate models do not explicitly encode such modular structural relationships (Franco et al., 2023; Li et al., 2023; Li et al., 2025; Parisi et al., 2024). Organizing design information by structural region and representing engineering adjacency can therefore provide a useful structural prior for analyzing inter-region interactions when the available training data are limited (Choi and Kumar, 2024; Pfaff et al., 2021; Zhao et al., 2024). Such a representation should be assessed against established surrogate models rather than assumed to provide an inherent predictive advantage.

A further challenge is to identify which design variables and structural regions consistently influence the competing crashworthiness responses. Feature-importance estimates obtained from a single fitted model or a single response can be sensitive to the sampled data and may not reflect the trade-offs among multiple outputs (Lundberg and Lee, 2017; Lundberg et al., 2020; Hancock et al., 2025; Kelodjou et al., 2024). Stability analysis across resampled datasets provides a more robust basis for identifying consistently influential variables and interpreting their structural significance (Alaiz-Rodríguez and Parnell, 2020; Nogueira et al., 2018). In the present study, variable-importance analysis is therefore used for engineering interpretation rather than for restricting the optimization space. Multi-objective search is subsequently performed in the complete mixed-variable design space so that the attainable Pareto trade-offs are not constrained by the importance ranking (Blank and Deb, 2020; Deb et al., 2002; Sharma and Kumar, 2022).

Accordingly, this study addresses three engineering questions in mixed-variable electric-vehicle side-pole design: (1) how numerical inadmissibility arises and is distributed across the stages of an automated full-vehicle FE workflow; (2) which structural regions and design variables exert consistent influence on occupant- and battery-related crash responses; and (3) what trade-offs among structural mass, occupant-compartment intrusion, and battery-neighborhood intrusion can be achieved and verified using high-fidelity FE analysis. A calibrated simulation-admissibility classifier, a structure-informed surrogate model, bootstrap-based SHAP analysis, and NSGA-II are employed as established computational tools to address these questions. The resulting framework is used to characterize workflow-level numerical admissibility, structural design sensitivities, inter-region interactions, and FE-verified crashworthiness trade-offs within the investigated side-pole design domain.

## 2 SIDE-POLE CRASHWORTHINESS PROBLEM AND AUTOMATED DATASET CONSTRUCTION

### 2.1 Finite Element Model and Side-Pole Crash Scenario

A full-vehicle FE model of a battery-electric sport utility vehicle was developed for the side-pole impact analysis (Figure 1). The impact configuration followed the 2024 C-NCAP side-pole protocol. The vehicle impacted a 254 mm rigid pole at 32 km/h and an angle of 75°. The analysis covered the first 120 ms after impact (China Automotive Technology and Research Center, 2024).

The body structure was discretized with shell elements at a nominal mesh size of 8 mm. The model comprised approximately 845,000 nodes and 830,000 elements. Steel and aluminum-alloy components were represented with the LS-DYNA *MAT_024 material model. The material cards were based on the corresponding stress-strain curves and included strain-rate effects. Spot-weld connection elements represented the spot welds, whereas solid elements represented the structural adhesive layers. The friction coefficients for body self-contact and body-to-pole contact were 0.1 and 0.2, respectively.

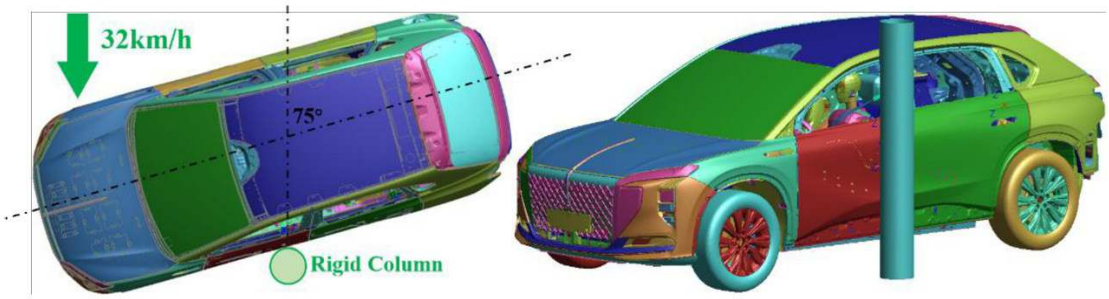



**Figure 1** Boundary conditions for the vehicle side-pole impact simulation.

The baseline model and all retained samples were required to satisfy the same quality criteria. Each simulation had to terminate normally and produce complete key-response files. Runs with negative volume, severe element distortion, or abnormal contact penetration were rejected. Both the hourglass-to-internal-energy ratio and the mass added by mass scaling were limited to 5% (Tang et al., 2016).

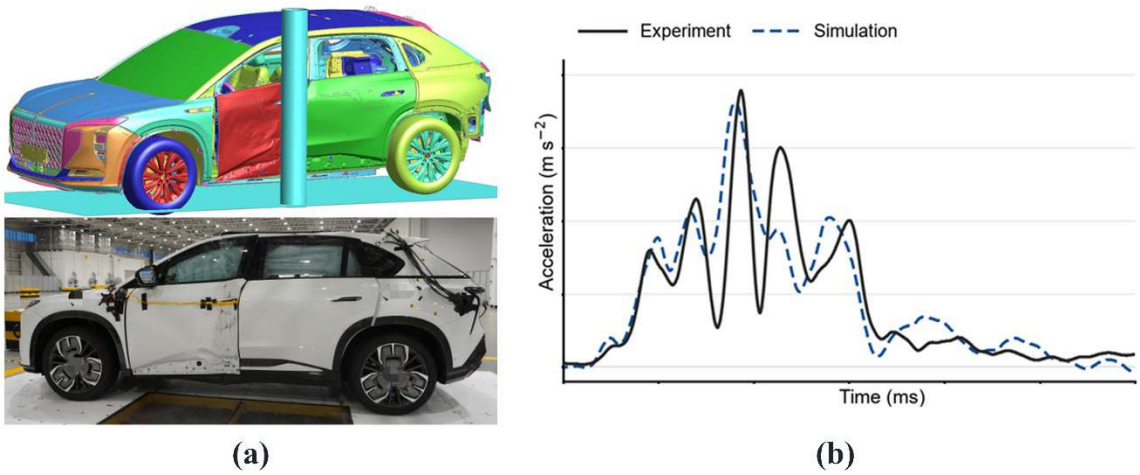



**Figure 2** Validation of the baseline side-pole FE model against the physical test: (a) deformation comparison; (b) acceleration time-history comparison.

The calibrated FE model reproduced the principal side-structure deformation observed in the physical test (Figure 2a). The simulated acceleration captured the dominant impact pulse and its multi-peak form (Figure 2b). Local differences remained during the early and late response phases. This agreement provided the basis for the subsequent parametric simulations and optimization.

### 2.2 Mixed Design Variables, Material-Property Representation, and Structural-Module Partitioning

The parametric design covered the B-pillar, sill, floor and battery-adjacent structures, upper rail, front body, seat cross-member, and dog-leg region. Together, these regions form the main load-transfer paths and govern lateral deformation during a side-pole impact. The SFE parameter set comprised continuous geometry/position variables and discrete material/thickness variables:

$$\mathbf{x}^{\mathrm{SFE}} = \left[\mathbf{x}^{\mathrm{geo}}, \mathbf{x}^{\mathrm{idx}}\right] \tag{1}$$

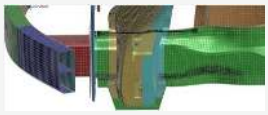
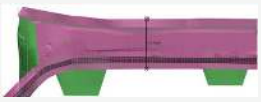
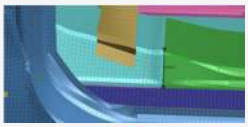
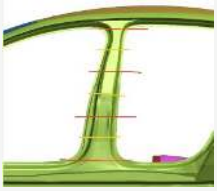
Where $\mathbf{x}^{\mathrm{geo}}$ contains 32 continuous geometric/position variables and x_idx contains 36 discrete material/thickness candidate indices, forming a 68-dimensional mixed design space (Moustapha et al., 2022; Yang et al., 2022). Continuous variables were sampled within their prescribed bounds and written to the SFE parameter file (Table 1). Each discrete index was decoded through a predefined material-thickness mapping to generate the corresponding material card and panel thickness (Garrido-Merchán and Hernández-Lobato, 2020; Zhu et al., 2024).

The $j$th discrete variable was represented in the surrogate model by:

$$\mathbf{p}_j = \left[t_j, \rho_j, E_j, \sigma_{y,j}, \sigma_{u,j}\right] \tag{2}$$

where $t_j$, $\rho_j$, $E_j$, $\sigma_{y,j}$, and $\sigma_{u,j}$ denote panel thickness, density, elastic modulus, yield strength, and ultimate tensile strength, respectively.

**Table 1** Selected geometric and positional input variables.

| Order | Input variable | Initial value | Range | Illustration |
|---|---|---|---|---|
| 1 | VIR/G/LON-F/X | 277 | [-50,50] |  |
| 2 | VIR/G/LON-R/Z | 158.4 | [-10,10] |  |
| 3 | VIR/G/SHOTGUN-F/Z | 32.2 | [-10,10] |  |
| 4 | VIR/G/TUNNEL/Z | 50.5 | [-5,15] |  |
| 5 | VIR/G/ROCKER/Z | 87.5 | [-5,5] |  |
| 6 | VIR/G/ROOF-SIDE-RAIL/Z | 116.3 | [-10,10] |  |
| 7 | VIR/G/BPILLAR-Lower/X | 248.2 | [-10,10] |  |
| 8 | VIR/G/BPILLAR-Lower/Y | 81.2 | [0,10] | |
| 9 | VIR/G/BPILLAR-Middle/X | 207.6 | [-10,10] | |
| 10 | VIR/G/BPILLAR-Middle/Y | 75.4 | [0,10] | |
| 11 | VIR/G/BPILLAR-Upper/X | 159.8 | [-10,10] | |
| 12 | VIR/G/BPILLAR-Upper/Y | 32.2 | [0,10] | |
| … | … | … | … | … |
| 32 | VIR/G/R-Roof-Side-Rail/Y | 88 | [0,10] |  |

For the surrogate model, the design variables were assigned to six structural modules: B-pillar, lower body and battery neighborhood, upper rail, front body, seat cross-member, and dog-leg:

$$\mathcal{X} = \{\mathbf{z}_{BP}, \mathbf{z}_{LB}, \mathbf{z}_{UR}, \mathbf{z}_{FB}, \mathbf{z}_{SC}, \mathbf{z}_{DL}\} \tag{3}$$

Here, $\mathbf{z}_k$ denotes the input features of structural module k. The terms $\mathbf{x}_k^g$ and $\mathbf{p}_k$ denote the geometric/position variables and material-property features within that module, respectively. The complete variable-to-module assignment is given in Table 2.

**Table 2** Mixed design variables and their model representations.

| Module symbol | Structural module | Included regions |
|---|---|---|
| BP | B-pillar module | Upper, middle, and lower B-pillar; B-pillar reinforcement |
| LB | Lower-body and battery-neighborhood module | Sill, floor, battery-adjacent channels, and lower-body cross-members |
| UR | Upper-rail module | Upper rail, roof rail, and roof-side-rail region |
| FB | Front-body module | Front longitudinal member, shotgun, front cross-member, and front-door neighborhood |
| SC | Seat-cross-member module | Front-seat cross-member and mounting-bracket region |
| DL | Dog-leg module | Dog-leg region and local transition zones |

For each simulation-admissible sample, B-pillar, front-door, and battery-neighborhood intrusions were measured from a common reference position or the protected-region envelope. Let $u_r(t)$ be the lateral position of measurement point *r* relative to that reference. The direction coefficient s_r assigns a positive sign to inward motion toward the occupant or battery region. Maximum intrusion was defined as:

$$d_r = \max_{t\in[0,T]} s_r\, u_r(t) \tag{4}$$

The outputs comprised B-pillar and front-door intrusions, local deformation of the seat cross-member and sill, battery-neighborhood structural intrusion, and body-structure mass. Measurement locations and symbols are listed in Table 3. B-pillar chest intrusion, battery-neighborhood structural intrusion, and body-structure mass were optimization objectives. The remaining responses were constrained not to exceed their baseline values.

**Table 3** Side-pole impact response definitions.

| Response group | Measurement position | Symbol | Output head |
|---|---|---|---|
| B-pillar intrusion displacement | Head, shoulder, chest, abdomen, pelvis | $d_{BP,k}(k = 1,\ldots,5)$ | BP head |
| Front-door intrusion displacement | Shoulder, chest, abdomen, pelvis | $d_{FD,k}(k = 1,2,3,4)$ | FD head |
| Front-seat rear cross-member bracket displacement | Rear mounting points | $d_{SC}$ | Local head |
| Sill outer panel deformation | Sill outer panel | $d_{Sill}$ | Local head |
| Battery-neighborhood structural intrusion | Battery-neighborhood structural channels | $d_{Bat}$ | Battery-neighborhood head |
| Body structure mass | -- | $m$ | Mass head |

### 2.3 Automated Dataset Construction and Simulation-Admissibility Labeling

An automated SFE–LS-DYNA workflow was used to construct two linked datasets for simulation-admissibility classification and crash-response surrogate modeling (Figure 3) (Fang et al., 2017; Lualdi et al., 2023). The workflow comprised variable sampling, parameter decoding, model reconstruction, mesh and keyword generation, LS-DYNA analysis, response extraction, and numerical-quality checking (Chen et al., 2024; Liu et al., 2018a).

Latin hypercube sampling was applied to the continuous geometry/position variables within their prescribed ranges (Viana, 2016). For each discrete material/thickness variable, stratified samples in [0,1] were mapped to the available component-specific candidates. Continuous values were written directly to the SFE parameter files, whereas discrete indices were decoded through the material-thickness mappings into panel thicknesses, material cards, and FE assignment data (Liu et al., 2018b).

Each run was checked for successful model reconstruction, mesh and keyword generation, solver termination, output completeness, response extractability, and numerical quality. A sample was classified as simulation-admissible only when all of the following conditions were satisfied:

1. The model, mesh, and keyword files were generated successfully;
2. LS-DYNA terminated normally and all required output files were complete and readable;
3. All required responses were extractable and contained neither NaN nor infinite values; and
4. all numerical-quality criteria specified in Section 2.1 were satisfied, including the checks for negative volume, severe element distortion, abnormal contact penetration, hourglass-to-internal-energy ratio, and added mass.

A sample was classified as simulation-inadmissible when failure of one or more of the above checks was reproducible under the prescribed modeling and solver settings. For diagnostic purposes, the stage at which inadmissibility was first detected was recorded separately as SFE reconstruction, mesh/keyword generation, LS-DYNA solution, or response extraction/numerical-quality checking. SFE-reconstruction and mesh/keyword-generation failures were grouped as pre-solver model-generation failures; abnormal LS-DYNA termination was treated as a solver-stage numerical failure; and failures identified during response extraction or post-solution numerical-quality checking were treated as post-solver failures. These stage labels indicate where inadmissibility was first detected in the automated workflow and are not intended to represent a common physical or numerical failure mechanism. Runs associated with transient computational interruptions were rerun under the same modeling and solver settings, and cases that subsequently completed normally were excluded from the simulation-inadmissible class.

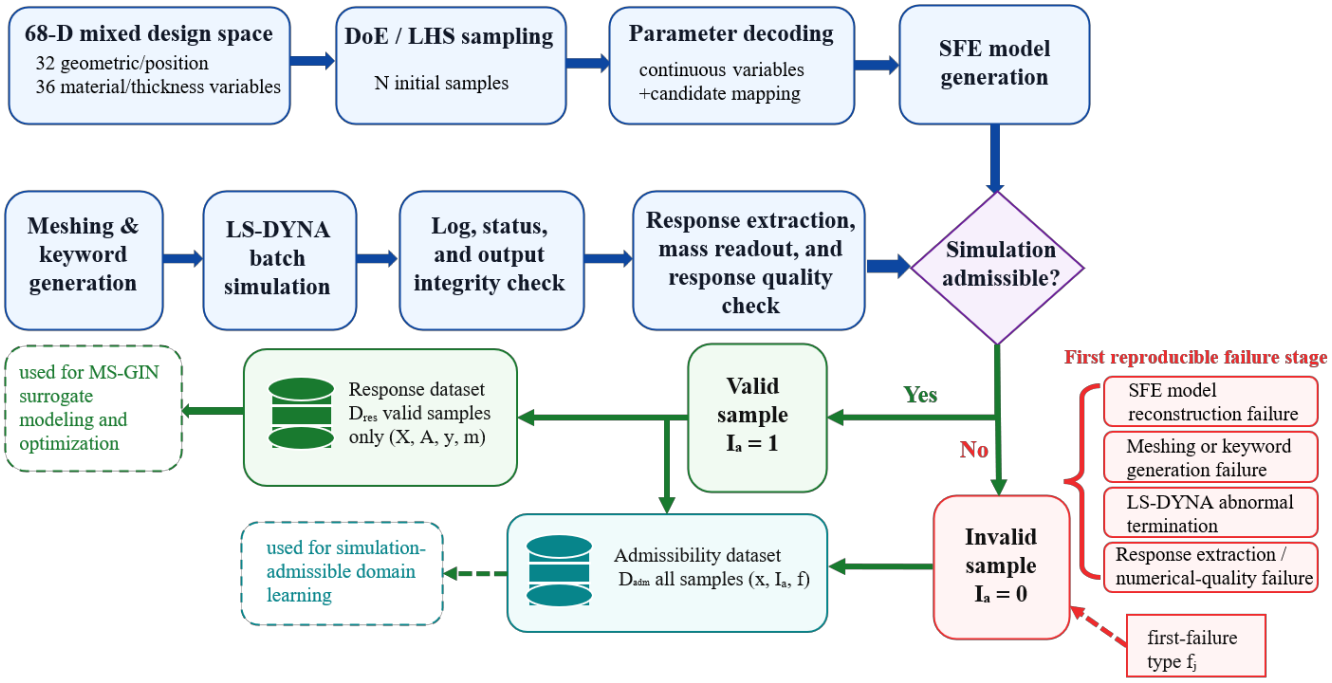



**Figure 3** Automated mixed-variable dataset construction and simulation-admissibility labeling workflow.

These checks yielded the simulation-admissibility dataset:

$$\mathcal{D}_{\text{adm}} = \{(\mathbf{x}_i, I_{a,i}, f_i)\}_{i=1}^{N} \tag{5}$$

where $x_i$ is the $i$th raw mixed-variable design, $I_{a,i}$ is the binary workflow-completion label, and $f_i$ records the first reproducible workflow stage at which inadmissibility was detected. The first-failure-stage label $f_i$ was retained only for stage-wise diagnostic analysis and was not used as a multiclass prediction target. The dataset contains both simulation-admissible and simulation-inadmissible samples and was used for classifier development. Samples that passed all checks formed the response dataset:

$$\mathcal{D}_{res} = \{(\mathcal{X}_i, \mathbf{A}, \mathbf{y}_i, m_i) \mid I_{a,i} = 1\} \tag{6}$$

where $\mathcal{X}_i$ is the modular input of the $i$th simulation-admissible sample, $A$ is the fixed reference-assembly structural adjacency shared by all samples, $y_i$ contains the intrusion and local-deformation responses, and $m_i$ is the body-structure mass. All data objects were linked by sample ID. The response dataset was used for surrogate-model development and subsequent optimization, whereas simulation-inadmissible samples were excluded from response regression.

## 3 HIDDEN-CONSTRAINT-AWARE SURROGATE OPTIMIZATION FRAMEWORK

### 3.1 Framework Overview and Problem Formulation

The datasets described in Section 2 were used to construct the hidden-constraint-aware surrogate-optimization framework shown in Figure 4 (Chen et al., 2024; Fang et al., 2017; Liu et al., 2018a). The framework comprises simulation-admissibility screening, structure-informed multi-response surrogate modeling, bootstrap-based variable-importance analysis, full-space multi-objective optimization, and high-fidelity FE verification. During surrogate optimization, candidates with an estimated workflow-completion probability below the prescribed screening threshold are excluded from response-based ranking. For the retained candidates, MS-GIN predicts the crash responses and body-structure mass. Bootstrap SHAP analysis is used separately to identify variables with stable influence for structural interpretation, whereas NSGA-II searches the complete mixed-variable design space for Pareto solutions (Khatouri et al., 2022; Li et al., 2021). Representative Pareto solutions are subsequently recalculated using the high-fidelity FE model to verify simulation admissibility, constraint satisfaction, and the predicted crashworthiness trade-offs.

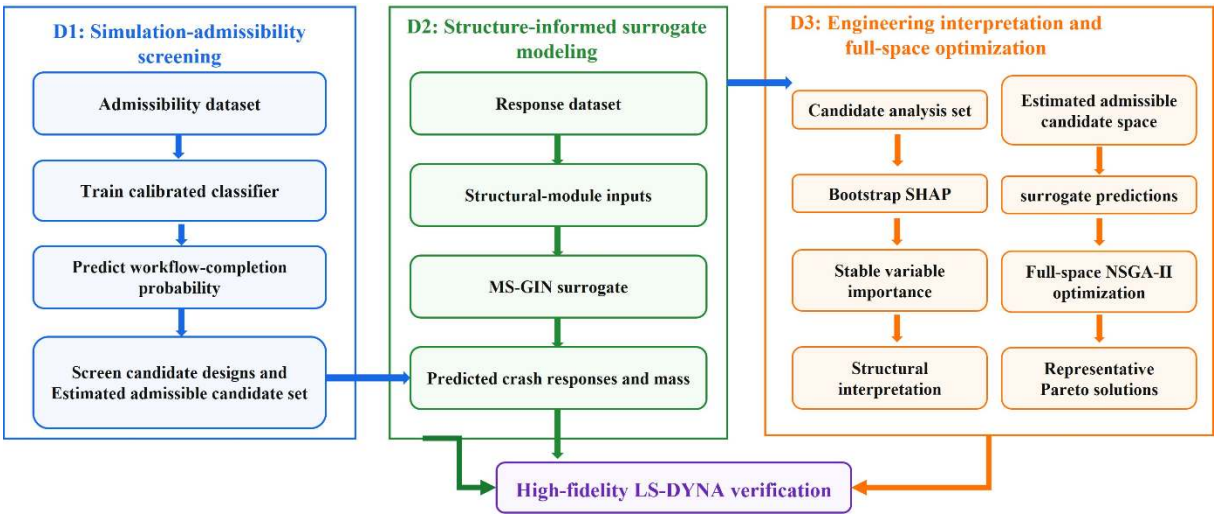



**Figure 4** Hidden-constraint-aware MS-GIN surrogate-optimization framework.

The mixed design space is defined as:

$$\Omega = \Omega_g \times \Omega_m \tag{7}$$

where $\Omega_g$ and $\Omega_m$ denote the continuous geometry/position space and discrete material/thickness candidate space, respectively. The simulation-admissible domain is defined as

$$\Omega_a = \{\mathbf{x} \in \Omega \mid I_a(\mathbf{x}) = 1\} \tag{8}$$

Because the workflow-completion label is unavailable for unevaluated candidates, an estimated simulation-admissible domain is defined using the calibrated workflow-completion probability introduced in Section 3.2. The multi-objective side-pole crashworthiness problem is defined as:

$$\min_{\mathbf{x}\in\Omega} \mathbf{F}(\mathbf{x}) = \left[m(\mathbf{x}), d_{BP,chest}(\mathbf{x}), d_{bat}(\mathbf{x})\right] \tag{9}$$

subject to:

$$\begin{aligned} & p_{adm}(\mathbf{x}) \geq \tau \\ & \mathbf{g}(\mathbf{x}) \leq \mathbf{g}_{base} \\ \mathbf{x}_g^L \leq \mathbf{x}_g \leq \mathbf{x}_g^U, \quad & \mathbf{x}_m \in \Omega_m \end{aligned} \tag{10}$$

where $m(x)$, $d_{\mathrm{BP,chest}}(x)$, and $d_{\mathrm{bat}}(x)$denote body-structure mass, B-pillar chest intrusion, and battery-neighborhood structural intrusion, respectively. The vector $g(x)$contains the remaining intrusion and local-deformation responses, and $g_{\mathrm{base}}$contains their corresponding baseline values. Equation (10) requires each constrained crash response to remain at or below its baseline value. Because the response surrogate is trained only on simulation-admissible samples, the calibrated workflow-completion probability is imposed as an additional screening constraint during surrogate-based optimization. Bootstrap-based variable-importance analysis and multi-objective optimization serve different purposes: the former supports structural interpretation, whereas the latter is conducted in the complete mixed-variable design space. The variable-importance ranking is therefore not used to restrict the design space during NSGA-II optimization.

### 3.2 Simulation-Admissibility Classification and Screening

An XGBoost classifier was trained on the simulation-admissibility dataset $D_{\mathrm{adm}}$(Chen and Guestrin, 2016). The binary target was the workflow-completion label defined in Section 2.3. Rather than representing a common physical failure mechanism, the classifier was used only to estimate whether a candidate could complete model generation, numerical solution, response extraction, and numerical-quality checking under the prescribed SFE–LS-DYNA workflow.

For an input design $\boldsymbol{x}$, XGBoost first produced a raw classification score $s(\boldsymbol{x})$. The raw score was subsequently calibrated using Platt scaling (Platt, 1999) to obtain the workflow-completion probability:

$$\begin{aligned} s(\mathbf{x}) &= C_{\mathrm{XGB}}\big(\phi(\mathbf{x})\big) \\ p_{\mathrm{adm}}(\mathbf{x}) &= \frac{1}{1+\exp[-(as(\mathbf{x})+b)]} \end{aligned} \tag{11}$$

where $C_{\mathrm{XGB}}$ denotes the XGBoost classifier, $a$and $b$are the Platt-calibration parameters, and $p_{\mathrm{adm}}(\boldsymbol{x}) \in [0,1]$ is the calibrated probability that the prescribed numerical workflow will be completed successfully. The classifier input was

$$\phi(\mathbf{x}) = \left[\tilde{\mathbf{X}}_g, \mathbf{z}_1, \ldots, \mathbf{z}_{n_d}\right] \tag{12}$$

where $\tilde{\boldsymbol{x}}_g$contains the normalized continuous geometry/position variables and $\boldsymbol{z}_j$is the material-property representation of the $j$th discrete material/thickness variable defined in Equation (2).

To account for class imbalance, simulation-admissible samples were assigned a training weight of 1, whereas simulation-inadmissible samples were assigned

$$w_{inad} = \frac{N_{adm}}{N_{inad}} \tag{13}$$

where $N_{\mathrm{adm}}$and $N_{\mathrm{inad}}$denote the numbers of admissible and inadmissible samples in the corresponding training fold. The class weights were recalculated independently within each training fold.

Ensemble size was treated as the principal model-complexity hyperparameter, with candidate values of 50, 100, 200, and 300 boosting trees evaluated within the inner cross-validation. The remaining XGBoost settings were fixed a priori to limit model flexibility: maximum tree depth = 3, learning rate = 0.03, minimum child weight = 3, subsample ratio = 0.85, feature-sampling ratio = 0.80, and L1/L2 regularization coefficients = 0.10 and 2.00, respectively. These fixed settings were not selected using the outer evaluation folds.

Classifier performance was evaluated using repeated stratified five-fold nested cross-validation. The complete nested procedure was repeated ten times using different stratified partitions. In each repeat, one outer fold was held out exclusively for performance evaluation. Ensemble-size selection, class-weight calculation, probability calibration, and screening-threshold determination were all performed using only the corresponding outer-training data to avoid selection bias from reuse of the outer evaluation folds (Cawley and Talbot, 2010).

Within each outer-training set, stratified inner cross-validation was used to generate out-of-fold raw XGBoost scores. The Platt-calibration parameters $a$ and $b$ were fitted using these inner out-of-fold scores and their corresponding workflow-completion labels. The resulting calibrated out-of-fold probabilities were subsequently used for threshold determination. The outer evaluation fold was not involved in model selection, probability calibration, or threshold selection.

Performance was assessed using ROC-AUC, PR-AUC for the simulation-inadmissible class, Brier score, and expected calibration error. Because simulation-inadmissible designs formed the minority class, PR-AUC was reported together with ROC-AUC (Saito and Rehmsmeier, 2015). For ROC and precision–recall analyses, simulation-inadmissible samples were treated as the positive class using

$$p_{inad}(\mathbf{x}) = 1 - p_{adm}(\mathbf{x}) \tag{14}$$

as the classification score.

For a screening threshold $\tau$, candidates satisfying $p_{\mathrm{adm}}(\boldsymbol{x}) \geq \tau$were retained for subsequent response prediction and optimization. Screening behavior was characterized by the inadmissible-sample rejection rate, admissible-sample recall, candidate-retention rate, and accepted-set validity. Accepted-set validity was defined as

$$V_{acc}(\tau) = \frac{N_{adm,acc}(\tau)}{N_{acc}(\tau)} \tag{15}$$

where $N_{\mathrm{adm,acc}}$is the number of truly simulation-admissible samples retained at threshold $\tau$, and $N_{\mathrm{acc}}$is the total number of retained samples. The corresponding accepted-set risk was

$$R_{risk}(\tau) = 1 - V_{acc}(\tau) \tag{16}$$

A 5% accepted-set risk limit was prescribed as a conservative screening criterion. Within each outer-training set, the threshold was selected from the inner out-of-fold calibrated probabilities by maximizing candidate retention while requiring $R_{\mathrm{risk}} \leq 0.05$:

$$\tau^* = \arg\max_{\tau} R_{ret}(\tau) \quad \text{subject to} \quad R_{risk}(\tau) \leq 0.05 \tag{17}$$

where $R_{\mathrm{ret}}$ is the fraction of all candidates retained at the corresponding threshold. The 5% limit was used as an empirical workflow-screening criterion to reduce contamination by numerically inadmissible candidates while retaining broad coverage of the candidate space; it is not interpreted as a transferable probabilistic guarantee outside the investigated design domain.

The calibrated workflow-completion probability defines the estimated simulation-admissible domain as

$$\widehat{\Omega}_{\mathrm{adm}} = \{\mathbf{x} \in \Omega : p_{\mathrm{adm}}(\mathbf{x}) \geq \tau\} \tag{18}$$

After completion of the repeated nested assessment, the deployment classifier complexity was fixed according to the cross-validated model-complexity analysis. Stratified full-data out-of-fold predictions were then used to fit the final Platt-calibration mapping and to determine a single deployment threshold using the same 5% accepted-set risk criterion. The XGBoost classifier was subsequently refitted using all labeled samples for candidate screening.

The classifier target was exclusively the binary workflow-completion label defined in Section 2.3. The first-failure-stage label was retained only for stage-wise diagnostic analysis and was not used as a multiclass prediction target. Accordingly, the estimated boundary represents workflow-level simulation admissibility under the prescribed SFE–LS-DYNA modeling and solver settings rather than a unified physical failure boundary.

### 3.3 MS-GIN Structure-Informed Gated-Interaction Surrogate Model

The surrogate received 68 design-variable inputs: 32 continuous geometric/position inputs and 36 discrete material/thickness inputs represented according to Equation (2). These 68 inputs were partitioned among the six predefined structural modules according to their associated body regions and were not replicated across modules. MS-GIN was trained using only simulation-admissible samples to predict B-pillar and front-door intrusions, local structural deformations, battery-neighborhood structural intrusion, and body-structure mass (Franco et al., 2023; Li et al., 2025; Parisi et al., 2024).

Using the module partition defined in Equation (3), the input associated with module $k$was mapped to a common hidden representation:

$$\mathbf{h}_k^{(0)} = \mathrm{E}_k(\mathbf{x}_k) \tag{19}$$

where $\boldsymbol{x}_k$ is the input feature vector of module $k$, $E_k(\cdot)$ is the fully connected encoder for that module, and $\boldsymbol{h}_k^{(0)} \in \mathbb{R}^{d_h}$ is the corresponding hidden representation. All six encoders used the same architecture but had independent trainable parameters.

A symmetric engineering-adjacency matrix $A$was defined from the direct structural interfaces among the six modules in the reference body assembly. An undirected edge was introduced only when two modules shared a direct structural connection in the undeformed reference configuration. The resulting base graph contained BP–LB, BP–UR, BP–DL, LB–FB, LB–SC, LB–DL, and UR–FB. Each undirected edge was represented by two directed message channels.

The prescribed adjacency was used as a static structural prior rather than as a complete time-dependent description of the load path during impact. In particular, BP and SC do not share a direct structural interface in the reference assembly, although inward B-pillar deformation may generate direct BP–SC contact during the side-pole event. The influence of explicitly introducing this deformation-induced interaction was therefore examined separately through a topology-sensitivity analysis.

For an ordered adjacent pair $(j, i)$satisfying $A_{ij} = 1$, the message from module $j$ to module $i$ was calculated as

$$\begin{aligned}\mathbf{q}_{j\to i} &= \left[\mathbf{h}_i^{(0)} \| \mathbf{h}_j^{(0)}\right] \\ \mathbf{g}_{j\to i} &= \sigma\left(W_g \mathbf{q}_{j\to i} + b_g\right) \\ \mathbf{e}_{j\to i} &= \mathbf{g}_{j\to i} \odot \mathrm{MLP}_e\left(\mathbf{q}_{j\to i}\right)\end{aligned} \tag{20}$$

where ‖denotes feature concatenation, $\odot$ denotes element-wise multiplication, $\sigma(\cdot)$ is the sigmoid function, $W_g$ and $b_g$ are the trainable gating parameters, $MLP_e$ is the edge-message multilayer perceptron, and $\boldsymbol{e}_{j\to i}$ is the gated message transmitted from module $j$to module $i$. The learned gate values quantify information exchange within the prescribed surrogate graph and are not interpreted as physical force fractions, energy-transfer ratios, or direct measures of load-path intensity.

The updated representation of module $i$ was

$$\mathbf{h}_i^{(1)} = \mathrm{MLP}_u\left(\mathbf{h}_i^{(0)} \| \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \mathbf{e}_{j\to i}\right) \tag{21}$$

where $\mathcal{N}(i)$ is the set of modules adjacent to module $i$, | $\mathcal{N}(i)$ | is the corresponding neighborhood size, and $MLP_u$ is the update multilayer perceptron. Messages in the two directions of an undirected engineering edge were calculated independently and were not constrained to be equal.

The six updated module representations were pooled into a global crash representation:

$$\mathbf{h} = \mathrm{MLP}_s\left(\sum_{k\in\mathcal{K}} \mathbf{h}_k^{(1)}\right) \tag{22}$$

where $\mathcal{K} = \{\mathrm{BP}, \mathrm{LB}, \mathrm{UR}, \mathrm{FB}, \mathrm{SC}, \mathrm{DL}\}$is the six-module set, and $MLP_s$maps the pooled module representation to the global latent vector $\boldsymbol{h}$.

Five output heads were used to predict the response groups:

$$\hat{\mathbf{y}}_r = \mathrm{Head}_r(\mathbf{h}), \quad r \in \{\mathrm{BP}, \mathrm{FD}, \mathrm{Local}, \mathrm{Battery}, \mathrm{Mass}\} \tag{23}$$

where $\hat{\boldsymbol{y}}_{\mathrm{BP}}$ and $\hat{\boldsymbol{y}}_{\mathrm{FD}}$ denote the B-pillar and front-door intrusion vectors, respectively; $\hat{\boldsymbol{y}}_{\mathrm{Local}}$ contains the seat-cross-member and sill local-deformation responses; $\hat{\boldsymbol{y}}_{\mathrm{Battery}}$ is the battery-neighborhood intrusion; and $\hat{m}$ is the body-structure mass. $Head_{\mathrm{BP}}$, $Head_{\mathrm{FD}}$, $Head_{\mathrm{Local}}$, $Head_{\mathrm{Battery}}$, and $Head_{\mathrm{Mass}}$ are the corresponding output multilayer perceptrons, and $n_r$ denotes the number of responses assigned to output head $r$.

Training minimized grouped mean-squared errors for the standardized responses. A separate loss was calculated for each response group. Weighted aggregation limited the influence of differences in group size and numerical scale. The response-group weights matched those used in the training implementation.

Model capacity was selected using only the development data. Candidate hidden dimensions of 8, 16, 32, and 64 were evaluated by cross-validation, and the smallest configuration within one standard error of the minimum cross-validated NRMSE was retained. This procedure selected $d_h = 16$, yielding 6,797 trainable parameters. A single gated-interaction stage was used together with dropout of 0.10 (Srivastava et al., 2014), AdamW weight decay, and early stopping.

The model was optimized using AdamW (Loshchilov and Hutter, 2019) with an initial learning rate of $1.0 \times 10^{-3}$, weight decay of $1.0 \times 10^{-4}$, batch size of 32, and a maximum of 500 epochs. The learning rate was halved after 20 epochs without improvement in validation loss, and training was stopped after 40 such epochs. Input and output standardization, architecture selection, and training decisions were based only on the corresponding development/training data.

The 528 simulation-admissible samples were partitioned into a 422-sample development set and a fixed 106-sample independent test set (Vabalas et al., 2019). The development set was used for preprocessing, architecture selection, hyperparameter selection, and training, whereas the fixed test set was reserved for final out-of-sample evaluation. Predictive performance was evaluated using $R^2$, RMSE, MAE, and NRMSE, with NRMSE defined as RMSE divided by the observed range of the corresponding response (Khatouri et al., 2022; Li et al., 2021).

For comparative evaluation, Kriging regression with an automatic-relevance-determination kernel, XGBoost regression, LightGBM regression, and radial-basis-function regression were trained using the same 422-sample development set and evaluated on the same fixed 106-sample test set as MS-GIN. Preprocessing and hyperparameter selection for each model were confined to the development data, and identical $R^2$, RMSE, MAE, and NRMSE definitions were used for all models.

A fully connected interaction model was additionally evaluated using the same six-module representation and training protocol but without restricting inter-module message exchange to the prescribed engineering-adjacency graph. This model served as a topology control for evaluating whether the structural-adjacency restriction materially affected predictive performance. For the BP–SC topology-sensitivity analysis, a second graph was constructed by adding a permanent BP–SC edge to the base adjacency while retaining the same input representation, hidden dimension, development/test partition, and training protocol. The graph configurations were compared under matched training conditions.

### 3.4 Variable-Importance Analysis and Full-Space Multi-Objective Optimization

Bootstrap DeepSHAP was used to identify design variables whose influence on the predicted crash responses remained stable across resampled development data. The resulting importance rankings were used exclusively for structural interpretation and design insight and were not used to define the optimization variables or their search ranges. All prescribed mixed design variables therefore remained active during the subsequent NSGA-II optimization.

For variable-importance analysis, 10,000 candidate designs were sampled from the original mixed design space $\Omega$. The calibrated simulation-admissibility classifier was used to retain the estimated simulation-admissible candidates. DeepSHAP was then applied to the trained MS-GIN model to quantify the contribution of each input feature to each response (Hancock et al., 2025; Lundberg and Lee, 2017; Lundberg et al., 2020). One hundred background samples were drawn from the development data, and 1,000 explanation samples were drawn from the estimated simulation-admissible candidate set. For reporting at the original design-variable level, the attributions associated with the material-property representation of each discrete material/thickness input were aggregated to one importance value for that input.

For each response, the mean absolute SHAP values were normalized across the design variables before multi-response aggregation. The combined importance of variable $j$ was calculated using Equation (24), where O and C denote the objective- and constrained-response sets, respectively; O contains body-structure mass, B-pillar chest intrusion, and battery-neighborhood structural intrusion, C contains the remaining ten crash responses, and the barred-phi term denotes the response-normalized mean absolute SHAP value for input j and response r. The weighting coefficient was set to $\beta = 0.5$, assigning equal aggregate weight to the objective and constrained-response groups and preventing the larger constrained-response group from dominating the combined ranking.

$$I_j = \frac{\beta}{|O|}\sum_{r\in O}\bar{\phi}_{j,r} + \frac{1-\beta}{|C|}\sum_{r\in C}\bar{\phi}_{j,r} \tag{24}$$

Selection stability was assessed using $B = 100$bootstrap resamples of the development data (Hancock et al., 2025; Kelodjou et al., 2024). Model fitting and importance estimation were repeated for each resample, while the fixed independent test set remained excluded from the bootstrap procedure. Within each resample, the variables were ranked by combined importance and the smallest set accounting for at least 85% of the cumulative importance was identified (Alaiz-Rodríguez and Parnell, 2020; Nogueira et al., 2018). The selection frequency of variable $j$ was defined as

$$s_j = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(j \in \mathcal{S}^{(b)}\right) \tag{25}$$

and the bootstrap-stable influential-variable set was defined as

$$S_{inf} = \left\{j : s_j \geq s_{min}\right\} \tag{26}$$

where $S^{(b)}$ is the minimum cumulative-importance set identified in bootstrap resample $b$, and $s_{\min}$is the prescribed stability threshold. The stable influential-variable set was used only to summarize robust design sensitivities; variables outside $S_{\inf}$remained active during multi-objective optimization.

The crashworthiness optimization was performed directly in the complete mixed design space using NSGA-II with Deb's constraint-domination rule (Blank and Deb, 2020; Deb et al., 2002; Sharma and Kumar, 2022):

$$\begin{aligned} &\min_{\mathbf{x}\in\Omega} \mathbf{F}(\mathbf{x}) = \left[\hat{m}(\mathbf{x}), \hat{d}_{BP,chest}(\mathbf{x}), \hat{d}_{bat}(\mathbf{x})\right] \\ &\text{s.t.} \quad p_{adm}(\mathbf{x}) \geq \tau \\ &\quad \hat{g}_r(\mathbf{x}) + \Delta_r \leq g_{r,base}, \quad r = 1, \ldots, R \\ &\quad \mathbf{x}_g^L \leq \mathbf{x}_g \leq \mathbf{x}_g^U \\ &\quad \mathbf{x}_m \in \Omega_m \end{aligned} \tag{27}$$

where $\hat{m}$, $\hat{d}_{\mathrm{BP,chest}}$, and $\hat{d}_{\mathrm{bat}}$are the surrogate-predicted body-structure mass, B-pillar chest intrusion, and battery-neighborhood structural intrusion, respectively; $p_{\mathrm{adm}}$is the calibrated workflow-completion probability; $\tau$is the deployment screening threshold defined in Section 3.2; $\hat{g}_r$denotes the $r$th constrained crash response; and $g_{r,\mathrm{base}}$is its baseline value.

To reduce the effect of surrogate underestimation on constraint screening, one-sided prediction errors were calculated from development-set out-of-fold predictions:

$$e_{i,r}^{+} = \max\left[0, g_{i,r}^{\mathrm{FE}} - \hat{g}_{i,r}^{\mathrm{OOF}}\right] \tag{28}$$

and the empirical margin for response $r$was defined as

$$\Delta_r = Q_{1-\alpha/R}\left(\left\{e_{i,r}^{+}\right\}\right) \tag{29}$$

The optimization included $R = 10$constrained responses and used an overall empirical error-coverage level of $\alpha = 0.10$; consequently, $\Delta_r$corresponded to the 99th percentile of the one-sided underestimation errors for each constrained response. These margins are empirical corrections derived from development-set errors and do not constitute a formal feasibility guarantee for unseen optimization candidates. Final feasibility was therefore assessed using high-fidelity FE analysis.

The full-space NSGA-II search used a population of 200 for 250 generations. The crossover probability was 0.9, and the per-variable mutation probability was set to the reciprocal of the total number of design variables. Ten independent runs with different random initializations were performed. The nondominated solutions from the ten runs were pooled and reranked by nondominance to obtain the final candidate Pareto set.

The final nondominated set was used to identify the three single-objective extreme solutions and an equal-weight compromise solution. For compromise selection, the three objective values were normalized over the final nondominated set, and the solution with the minimum Euclidean distance to the ideal point was selected as $P^*$. The representative Pareto solutions were subsequently recalculated using the high-fidelity SFE–LS-DYNA model to verify workflow admissibility, constraint satisfaction, and the predicted crashworthiness trade-offs.

## 4 RESULTS AND DISCUSSION

### 4.1 Dataset Characteristics and Simulation-Admissibility Screening

Failures attributable to resource scheduling, file transfer, or other transient software interruptions were excluded when rerunning under the prescribed settings completed normally. The analysis therefore comprised 600 LHS designs, including 528 simulation-admissible and 72 reproducibly simulation-inadmissible samples, as shown in Figure 5, corresponding to an admissibility rate of 88.0%. Among the 72 inadmissible samples, 12 (16.7%) failed during SFE reconstruction, 16 (22.2%) during mesh or keyword generation, 31 (43.1%) during the LS-DYNA solution, and 13 (18.1%) during response extraction or post-solution numerical-quality checking. The distribution shows that simulation inadmissibility arose at multiple stages of the automated workflow, with the largest proportion occurring during the LS-DYNA solution stage.

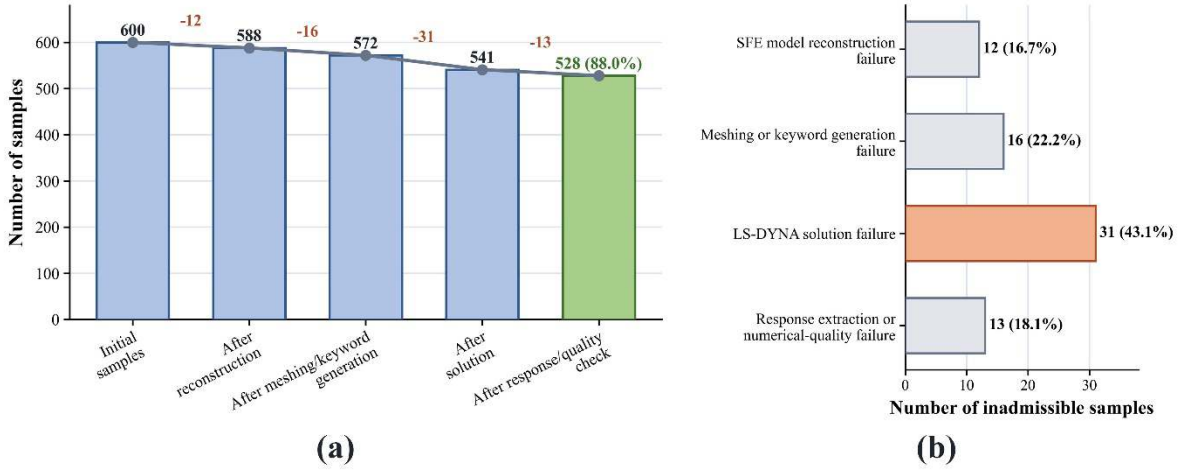



**Figure 5** Dataset construction and simulation-inadmissibility statistics: (a) stage-wise sample retention; (b) first reproducible workflow stages at which simulation inadmissibility was detected.

The response distributions of the 528 simulation-admissible samples are summarized in Figure 6. The intrusion and local-deformation responses exhibited substantial variation, but their distributions differed across structural regions. B-pillar chest intrusion, front-door chest intrusion, and sill local deformation were centered near or above their baseline levels, whereas the median seat-cross-member deformation and battery-neighborhood intrusion were below baseline. Sill deformation exhibited a comparatively broad distribution. Body-structure mass also varied on both sides of the 444.85 kg baseline. The sampled design domain therefore contained both improved and degraded configurations and provided sufficient variation for evaluating lightweighting–crashworthiness trade-offs.

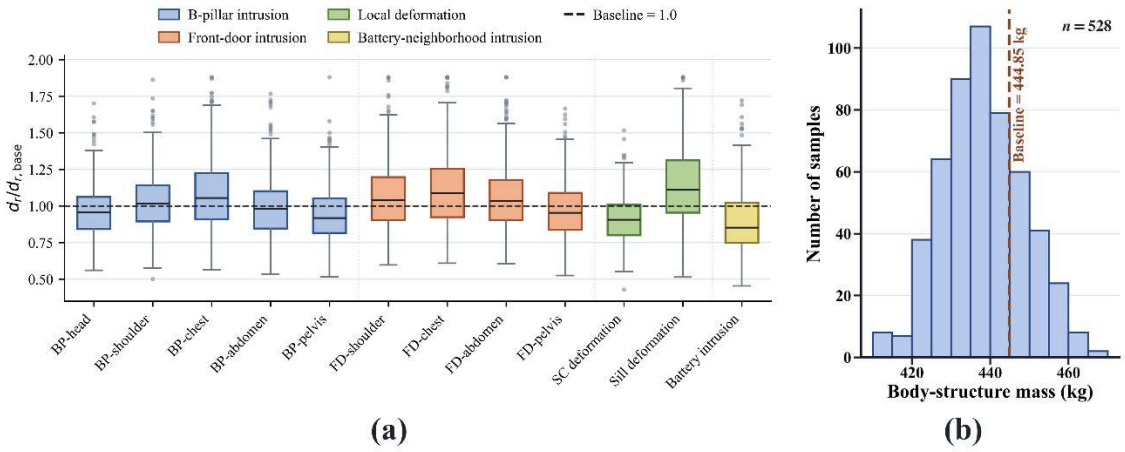



**Figure 6** Response distributions of the 528 simulation-admissible samples: (a) normalized maximum intrusion and local-deformation responses; (b) body-structure-mass distribution.

Classifier performance under repeated stratified five-fold nested cross-validation is summarized in Table 4 and Figure 7. Across the ten repeated out-of-fold evaluations, the ROC-AUC for distinguishing simulation-inadmissible designs was $0.90 \pm 0.02$. Because only 12% of the samples were simulation-inadmissible, PR-AUC provides a more informative measure of minority-class detection. The PR-AUC for the simulation-inadmissible class was $0.67 \pm 0.04$, substantially above the class prevalence of 0.12. The limited dispersion across repeated partitions indicates that the observed discrimination was not dominated by a particular five-fold split.

**Table 4** Repeated nested-cross-validated performance of the simulation-admissibility screening model.

| Metric | Value | Metric | Value | Metric | Value |
|---|---|---|---|---|---|
| ROC-AUC | 0.90±0.02 | PR-AUC (inadmissible) | 0.67±0.04 | Brier score | 0.060±0.006 |
| ECE | 0.048±0.008 | Inadmissible-sample rejection rate | 0.70±0.05 | Admissible-sample recall | 0.91±0.02 |
| Empirical candidate-retention rate | 0.84±0.02 | Accepted-set validity | 0.95±0.01 | Selected probability threshold, $\tau$ | 0.67±0.03 |

Probability calibration was also stable across repeated partitions. The Brier score was $0.060 \pm 0.006$, and the expected calibration error was $0.048 \pm 0.008$(Figure 7c). The repeat-level reliability curves remained close to the ideal calibration line over most of the probability range, although moderate deviations persisted in several probability intervals. These results support the use of the calibrated probability as a workflow-screening quantity rather than as a deterministic representation of the simulation-failure boundary.

The screening-specific metrics quantify the corresponding engineering trade-off. At the selected operating point, the classifier rejected $0.70 \pm 0.05$of the simulation-inadmissible designs while retaining $0.91 \pm 0.02$of the simulation-admissible designs. Consequently, the empirical candidate-retention rate was $0.84 \pm 0.02$, and the accepted-set validity was $0.95 \pm 0.01$. Across the repeated nested evaluations, the selected screening threshold remained concentrated around $\tau = 0.67 \pm 0.03$(Figure 7d). The corresponding accepted-set inadmissible fraction was $0.05 \pm 0.01$, compared with 0.12 in the unscreened dataset, while approximately 84% of the candidates remained available for subsequent surrogate-based evaluation.

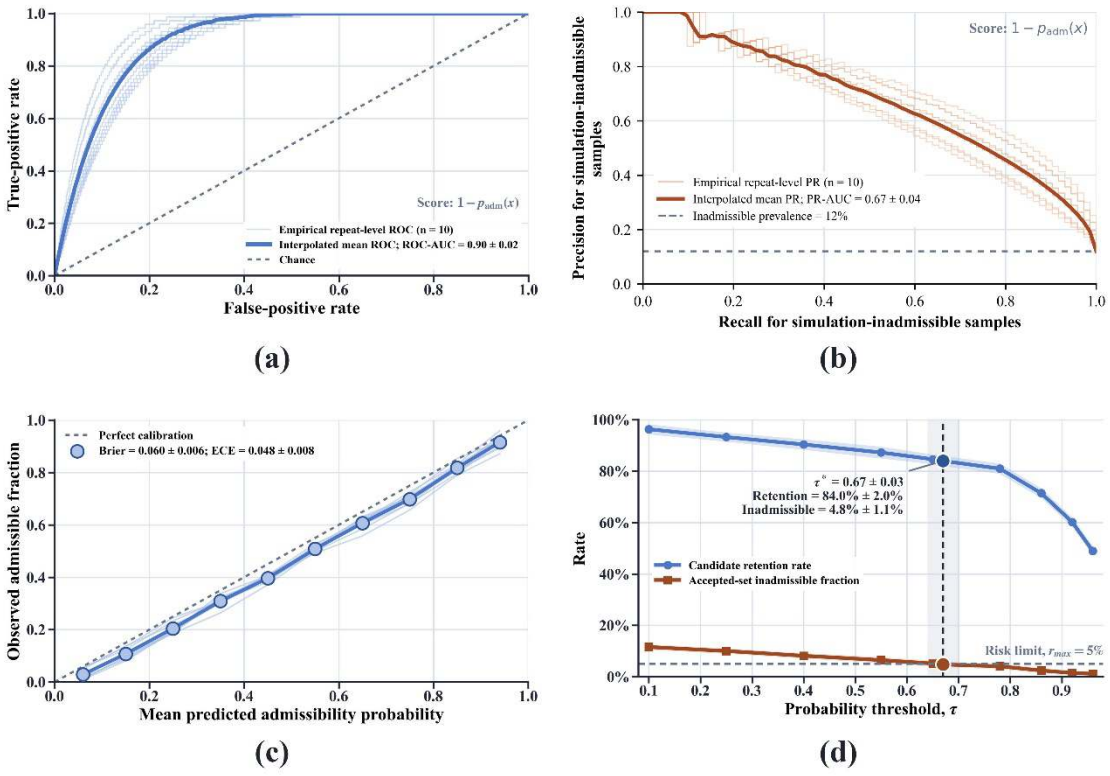



**Figure 7** Performance of the simulation-admissibility classifier: (a) pooled ROC curve for the simulation-inadmissible class; (b) pooled precision-recall curve for the simulation-inadmissible class; (c) reliability diagram; (d) threshold scan for candidate filtering.

Model-complexity sensitivity was examined using 50, 100, 200, and 300 boosting trees (Figure 8). Increasing the ensemble from 50 to 100 trees improved both ROC-AUC and minority-class PR-AUC, whereas further increases to 200 or 300 trees produced only marginal changes relative to the variation across repeated data partitions. A similar plateau was observed for accepted-set validity, admissible-sample recall, and inadmissible-sample rejection. The inner-validation procedure therefore selected the 100-tree configuration as the lower-complexity model within the stable performance region.

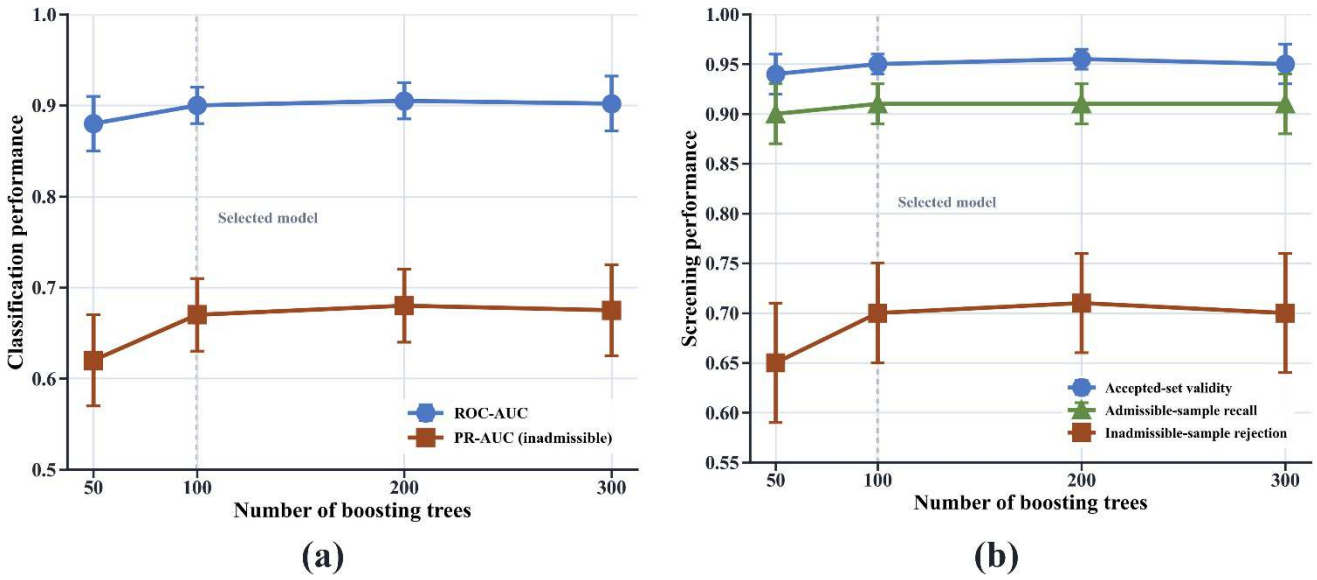



**Figure 8** Sensitivity of simulation-admissibility screening to XGBoost model complexity: (a) ROC-AUC and PR-AUC for the simulation-inadmissible class; (b) accepted-set validity, admissible-sample recall, and inadmissible-sample rejection rate.

The repeated validation and model-complexity analyses reduced the dependence of classifier performance on a particular data partition or ensemble size. However, only 72 samples were simulation-inadmissible, and the inadmissible-sample rejection rate still showed relatively large variation across repeated partitions. The classifier is therefore used to screen workflow admissibility within the present design domain and under the adopted SFE–LS-DYNA settings. The resulting decision boundary is specific to this numerical workflow and dataset. High-fidelity FE analyses are retained for final verification of the representative optimized designs.

### 4.2 Surrogate Accuracy and Comparative Evaluation

The 528 simulation-admissible samples were divided into 422 development samples and a fixed 106-sample independent test set. Figure 9 compares MS-GIN predictions with the corresponding FE results for six representative responses. The $R^2$ values for the three optimization objectives—B-pillar chest intrusion, battery-neighborhood structural intrusion, and body-structure mass—were 0.95, 0.93, and 0.95, respectively, indicating that the surrogate captured most of the observed variation in these responses on the fixed hold-out set. Front-door chest intrusion and seat-cross-member local deformation achieved $R^2$values of 0.88 and 0.89, respectively, and exhibited greater scatter. These responses were therefore more difficult to approximate than the three optimization objectives. Several responses also retained visible bias near the limits of their sampled ranges, motivating the one-sided surrogate-underestimation margins used during constrained optimization.

Figure 10 summarizes prediction performance for all 13 outputs on the fixed independent test set. All responses achieved $R^2 \geq 0.87$ and NRMSE below 10%. Across the 13 responses, the mean $R^2$was 0.91 and the mean NRMSE was 6.5%. These results indicate that MS-GIN provided acceptable hold-out accuracy across the investigated responses, although prediction difficulty varied among response types, with local-deformation responses generally exhibiting greater scatter than body-structure mass and the principal intrusion responses.

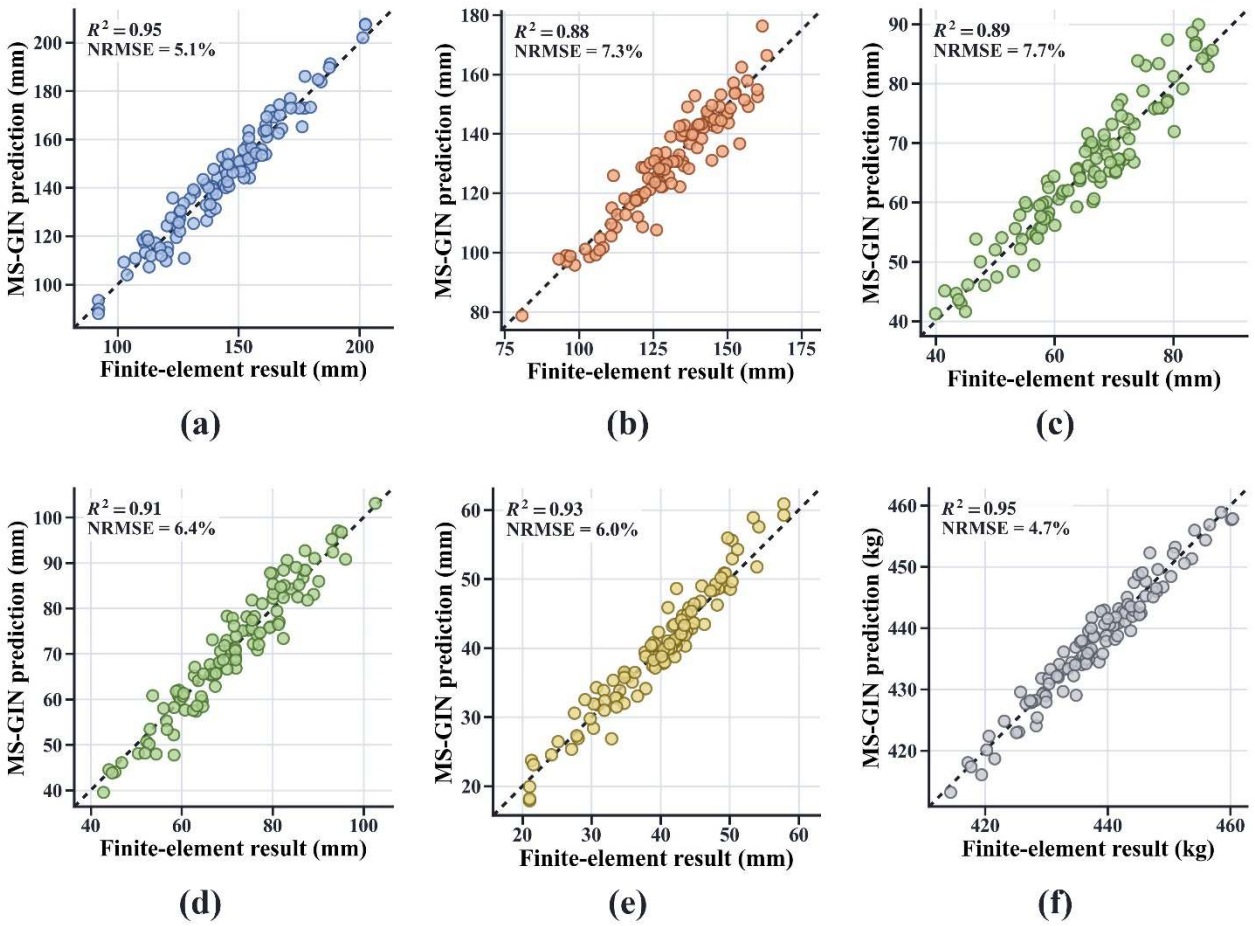



**Figure 9** MS-GIN prediction performance on the fixed independent test set: (a) B-pillar chest intrusion; (b) front-door chest intrusion; (c) seat-cross-member local deformation; (d) sill local deformation; (e) battery-neighborhood structural intrusion; (f) body-structure mass.

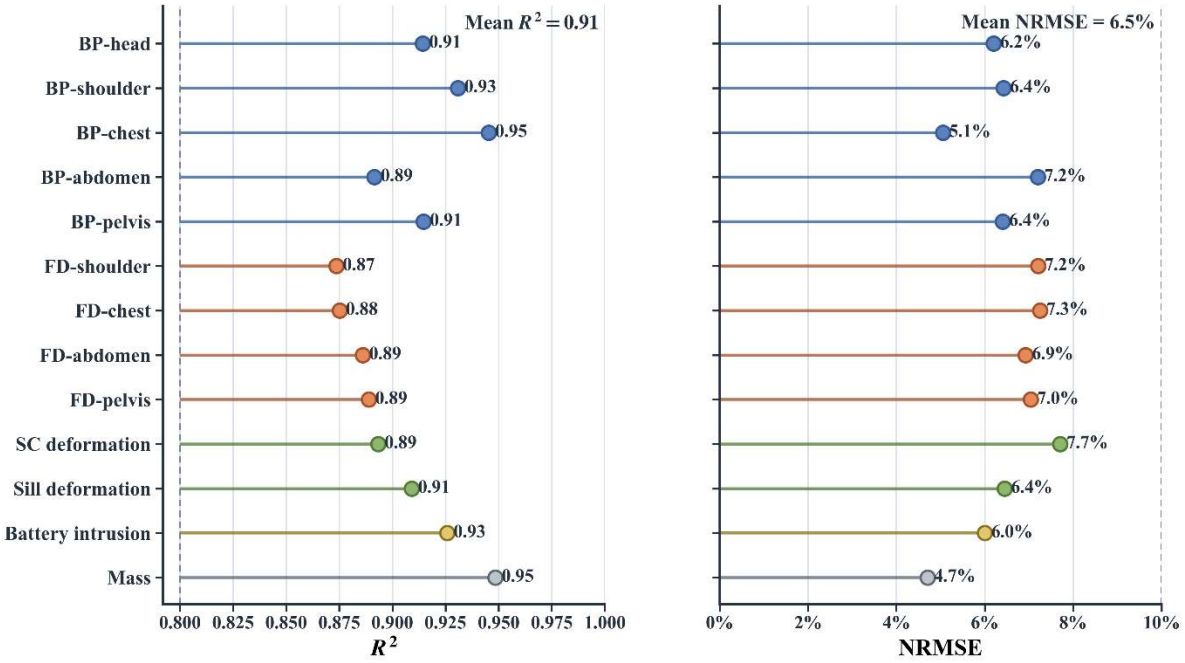



**Figure 10** Response-wise MS-GIN performance on the fixed independent test set.

Table 5 compares MS-GIN with conventional engineering surrogate models and an unrestricted-interaction topology control using the same fixed independent test set. The conventional models achieved predictive accuracy comparable to MS-GIN. MS-GIN yielded a mean $R^2$ of 0.91 and a mean NRMSE of 6.5%, whereas Kriging yielded a mean $R^2$ of 0.89 and a slightly lower mean NRMSE of 6.3%. XGBoost, LightGBM, and RBF regression yielded mean $R^2$ values of 0.88, 0.87, and 0.89 and mean NRMSE values of 6.9%, 7.0%, and 7.2%, respectively. The differences among MS-GIN and the strongest conventional surrogates were therefore modest and do not support a decisive predictive advantage for the graph model.

The fully connected interaction model achieved similar hold-out accuracy, with a mean $R^2$of 0.88 and a mean NRMSE of 6.7%. The small difference from MS-GIN indicates that the prescribed engineering-adjacency topology had limited influence on overall predictive accuracy. Its main role is to provide an explicit representation of structural-module connectivity for subsequent inter-region interaction analysis.

**Table 5** Comparison of MS-GIN with conventional surrogate models and an unrestricted-interaction topology control.

| Model | $R^2(d_{BP,chest})$ | $R^2(d_{Bat})$ | $R^2(m)$ | Mean $R^2$ over all responses | Mean NRMSE |
|---|---|---|---|---|---|
| MS-GIN | 0.95 | 0.93 | 0.95 | 0.91 | 6.5% |
| Fully connected interaction | 0.93 | 0.90 | 0.95 | 0.88 | 6.7% |
| Kriging | 0.93 | 0.91 | 0.95 | 0.89 | 6.3% |
| XGBoost | 0.94 | 0.89 | 0.96 | 0.88 | 6.9% |
| LGBM | 0.95 | 0.88 | 0.95 | 0.87 | 7.0% |
| RBF | 0.93 | 0.89 | 0.94 | 0.89 | 7.2% |

Figure 11 shows the reference-assembly adjacency and learned gated interactions among the six structural modules. BP connects the upper rail, lower body, and dog-leg regions, while LB further links the front body, seat cross-member, and dog-leg. The dashed BP–SC edge denotes the additional topology examined in the sensitivity analysis and is excluded from the base-graph gate matrix. The learned coefficients are direction dependent, with relatively strong interactions along BP–LB and LB–SC, indicating pronounced information exchange among the B-pillar, lower-body, and seat-cross-member modules. The gate coefficients represent learned inter-module information weighting within MS-GIN.

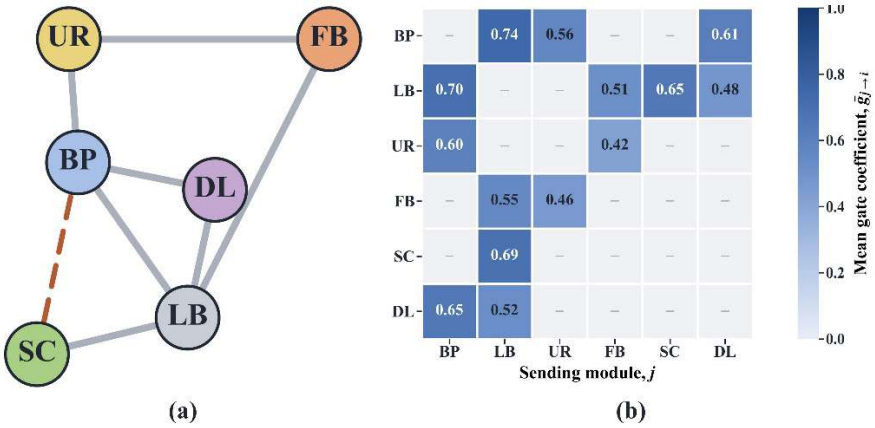



**Figure 11** Reference-assembly engineering adjacency and learned gated interactions: (a) base structural adjacency, with the dashed BP–SC edge indicating the additional interaction examined in the topology-sensitivity analysis; (b) learned directed gate coefficients for the base adjacency.

### 4.3 Stable Variable Importance and Structural Interpretation

Figure 12 summarizes the bootstrap-averaged combined importance and stability-selection frequency of the design variables. Using a stability threshold of $s_{\min} = 0.70$, 18 bootstrap-stable influential variables were identified, as listed in Table 6. Together, these variables accounted for 88.0% of the combined importance across the original design space. The influential-variable set covered all six structural modules, although 11 of the 18 variables belonged to the BP and LB modules. These variables were identified solely to summarize stable design sensitivities and structural influence; they did not define the subsequent optimization space, and variables outside this set remained active during full-space NSGA-II optimization.

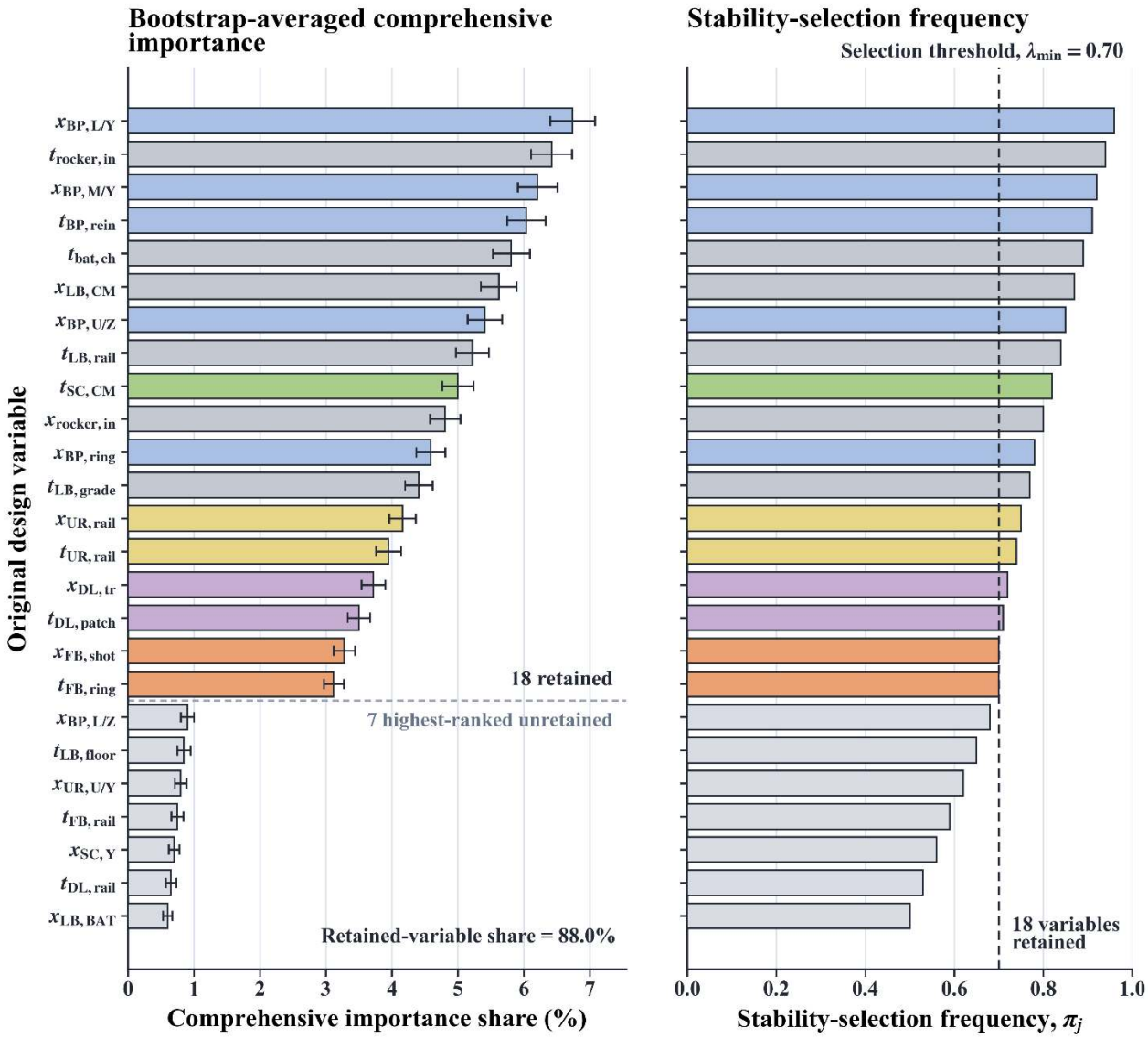



**Figure 12** Bootstrap-stable influential-variable analysis within the estimated simulation-admissible domain.

The four variables with the highest stability were the B-pillar lower lateral position, sill inner-panel material/thickness configuration, B-pillar middle lateral position, and B-pillar reinforcement material/thickness configuration, with selection frequencies of 0.96, 0.94, 0.92, and 0.91, respectively. Together, these four variables contributed 25.4% of the combined importance.

The importance pattern shows a concentration of stable influence in the B-pillar and lower-body regions. B-pillar geometric variables were associated primarily with B-pillar and front-door intrusions at multiple heights, indicating that changes in B-pillar position and local geometry strongly affect deformation of the occupant-compartment side structure within the investigated design domain. In contrast, material/thickness variables in the sill and lower-body modules were associated simultaneously with battery-neighborhood intrusion, sill deformation, and body-structure mass, reflecting the competing effects of local structural resistance and lightweighting.

Although the BP and LB modules contained most of the highly ranked variables, bootstrap-stable influential variables were identified in all six structural modules. This distributed pattern indicates that the investigated side-pole crashworthiness response cannot be attributed to a single local component. Occupant-compartment intrusion, battery-proximal deformation, and structural mass are associated with coordinated design changes across multiple connected structural regions.

DeepSHAP quantified the relative influence of the design variables within the sampled design domain. The resulting ranking was used to support structural interpretation, and all design variables were retained in the subsequent multi-objective optimization.

**Table 6** Bootstrap-stable influential design variables identified from the 68-variable space

| Variable | Module | Type | Stability frequency $\pi_j$ | Normalized importance $I_j$ | Importance share | Dominant responses |
|---|---|---|---|---|---|---|
| $x_{BP,L/Y}$ | BP | G | 0.96 | 1.000 | 6.74% | $d_{BP,chest}$, $d_{BP,pelvis}$ |
| $t_{rocker,in}$ | LB | M/T | 0.94 | 0.952 | 6.42% | $d_{bat}$, $d_{sill}$, $m$ |
| $x_{BP,M/Y}$ | BP | G | 0.92 | 0.921 | 6.21% | $d_{BP,chest}$, $d_{FD,chest}$ |
| $t_{BP,rein}$ | BP | M/T | 0.91 | 0.895 | 6.04% | $d_{BP,chest}$, $d_{BP,shoulder}$, $m$ |
| $t_{sill,outer}$ | LB | M/T | 0.89 | 0.861 | 5.81% | $d_{sill}$, $d_{bat}$, $m$ |
| $x_{ROCKER/Z}$ | LB | G | 0.87 | 0.833 | 5.62% | $d_{sill}$, $d_{bat}$ |
| $x_{BP,M/X}$ | BP | G | 0.85 | 0.802 | 5.41% | $d_{BP,chest}$, $d_{BP,abdomen}$ |
| $t_{bat,ch}$ | LB | M/T | 0.84 | 0.774 | 5.22% | $d_{bat}$, $m$ |
| $t_{seat,cross}$ | SC | M/T | 0.82 | 0.742 | 5.02% | $d_{SC}$, $d_{sill}$, $m$ |
| $x_{TUNNEL/Z}$ | LB | G | 0.80 | 0.713 | 4.83% | $d_{bat}$, $d_{SC}$ |
| $x_{BP,U/Y}$ | BP | G | 0.78 | 0.681 | 4.59% | $d_{BP,head}$, $d_{BP,shoulder}$ |
| $t_{floor,cross}$ | LB | M/T | 0.77 | 0.654 | 4.41% | $d_{bat}$, $d_{SC}$, $m$ |
| $x_{RRSR/Y}$ | UR | G | 0.75 | 0.617 | 4.16% | $d_{BP,head}$, $d_{FD,shoulder}$ |
| $t_{roof,rail}$ | UR | M/T | 0.74 | 0.586 | 3.95% | $d_{BP,head}$, $d_{BP,shoulder}$, $m$ |
| $x_{DL/Y}$ | DL | G | 0.72 | 0.552 | 3.72% | $d_{sill}$, $d_{FD,pelvis}$, $d_{bat}$ |
| $t_{dogleg}$ | DL | M/T | 0.71 | 0.519 | 3.50% | $d_{sill}$, $d_{FD,pelvis}$, $m$ |
| $x_{SHOTGUN-F/Z}$ | FB | G | 0.70 | 0.486 | 3.28% | $d_{FD,shoulder}$, $d_{FD,chest}$ |
| $t_{frontrail}$ | FB | M/T | 0.70 | 0.462 | 3.12% | $d_{FD,chest}$, $m$ |

**Note:** G denotes geometry; M/T denotes material/thickness.

### 4.4 Optimization Performance and High-Fidelity FE Validation

NSGA-II was applied directly to the complete 68-variable mixed design space without restricting the search according to the bootstrap SHAP ranking. Figure 13a shows the trade-offs among body-structure mass, B-pillar chest intrusion, and battery-neighborhood intrusion. The pooled archive contained 159 nondominated solutions, indicating that improvements in the three objectives cannot generally be achieved independently within the investigated design domain. The solutions were further ranked according to their equal-weight normalized distance to the ideal point (Figure 13b), and the minimum-distance solution was selected as the compromise design $P^*$. Figure 13c shows that its predicted mass, B-pillar chest intrusion, and battery-neighborhood intrusion were 0.957, 0.941, and 0.900 times their respective baseline values.

Table 7 summarizes the three single-objective extreme solutions and the selected compromise design. The minimum-mass solution reduced body-structure mass by 5.7%, whereas the minimum-B-pillar and minimum-battery solutions reduced their corresponding intrusion objectives by 13.2% and 19.3%, respectively. For $P^*$, the predicted body-structure mass, B-pillar chest intrusion, and battery-neighborhood intrusion were 425.92 kg, 136.50 mm, and 37.80 mm, corresponding to reductions of 4.3%, 5.9%, and 10.0%. Its simulation-admissibility probability was 0.84, exceeding the deployment threshold of 0.67. Compared with the three extreme solutions, $P^*$ therefore provides a more balanced compromise among structural lightweighting, occupant-compartment deformation, and battery-neighborhood protection rather than maximizing any single objective.

The selected $P^*$ was subsequently recalculated using the high-fidelity SFE–LS-DYNA model. As shown in Figure 14a, the FE-verified body-structure mass, B-pillar chest intrusion, and battery-neighborhood intrusion were 428.88 kg, 131.23 mm, and 35.86 mm, corresponding to approximately 0.964, 0.905, and 0.854 times their baseline values. The surrogate-to-FE relative errors for the three objectives were approximately −0.7%, 4.0%, and 5.4%, respectively. Figure 14b further shows that the ten non-objective responses remained below their corresponding baseline limits after FE recalculation, with response ratios of approximately 0.87–0.95. The empirical prediction margins were used only during surrogate-based screening; final constraint satisfaction was determined from the high-fidelity FE results.

Table 8 provides the response-wise comparison between the surrogate predictions and LS-DYNA results. Most responses showed relatively small verification discrepancies, whereas the largest relative error occurred for the seat-cross-member local deformation. Its surrogate prediction was 1.66 mm and the FE result was 1.45 mm, giving an absolute difference of only 0.21 mm but a relative error of approximately 14.5% because of the small magnitude of this response. Compared with the global intrusion responses, the seat-cross-member deformation is more localized and exhibits stronger sensitivity to local contact and deformation-mode changes, which increases the difficulty of surrogate interpolation. Final satisfaction of this local constraint was determined using the high-fidelity FE result.

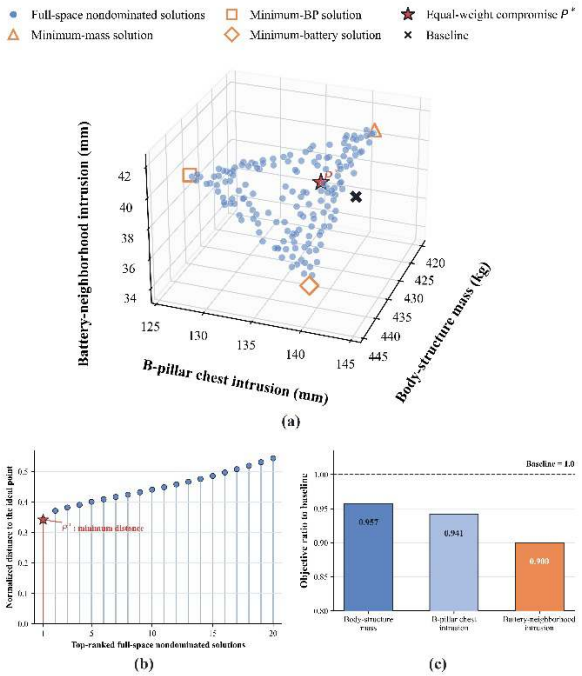



**Figure 13** Multi-objective optimization and compromise-design selection: (a) pooled nondominated solutions obtained from the 68-variable NSGA-II searches; (b) normalized ideal-point ranking; (c) surrogate-predicted objectives of $P^*$ relative to baseline.

**Table 7** Surrogate-predicted objectives of representative Pareto solutions.

| Design | $m(kg)$ | Change | $d_{BP,chest}(mm)$ | Change | $d_{Bat}(mm)$ | Change | $p_{adm}$ |
|---|---|---|---|---|---|---|---|
| Baseline | 444.85 | - | 145.00 | - | 42.00 | - | - |
| $Min_{mass}$ | 419.50 | -5.7% | 140.53 | -3.1% | 40.41 | -3.8% | 0.82 |
| $Min_{BP}$ | 436.81 | -1.8% | 125.82 | -13.2% | 39.63 | -5.7% | 0.84 |
| $Min_{Bat}$ | 438.18 | -1.5% | 138.68 | -4.3% | 33.87 | -19.3% | 0.86 |
| $P^*$ | 425.92 | -4.3% | 136.50 | -5.9% | 37.80 | -10.0% | 0.84 |

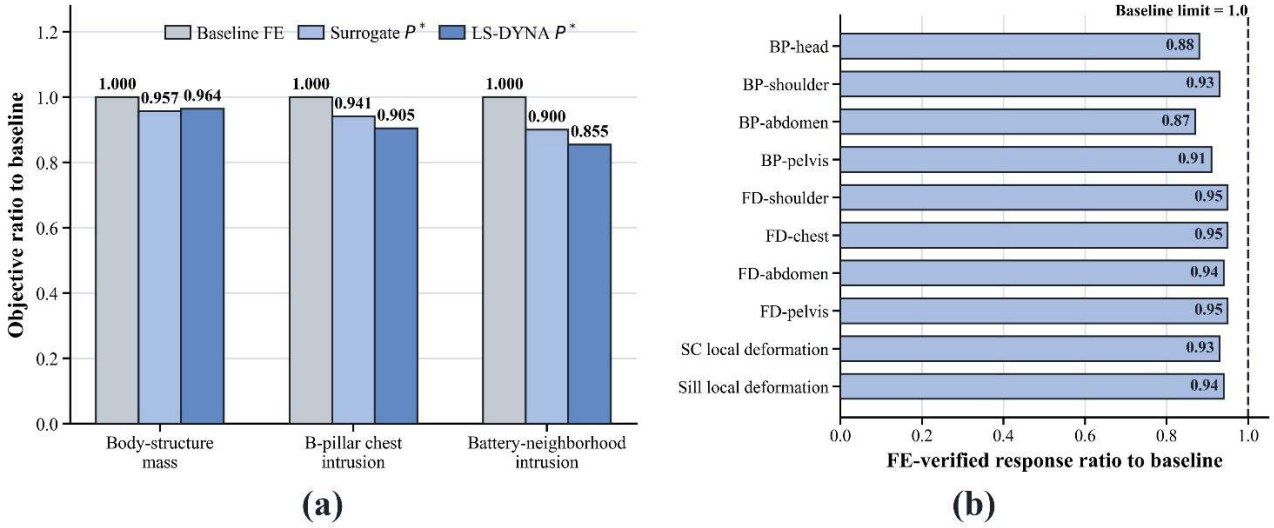



**Figure 14** FE-validated improvements and relative-constraint satisfaction for the selected ideal-point solution: (a) surrogate prediction and LS-DYNA verification of P*; (b) FE-verified non-objective constraints.

**Table 8** Relative errors between surrogate predictions and LS-DYNA results for the selected compromise design.

| | $d_{BP1}$ | $d_{BP2}$ | $d_{BP3}$ | $d_{BP4}$ | $d_{BP5}$ | $d_{FD1}$ | $d_{FD2}$ |
|---|---|---|---|---|---|---|---|
| Prediction | 19.91 | 12.85 | 136.50 | 44.22 | 51.37 | 82.01 | 90.89 |
| FE result | 18.42 | 13.51 | 131.23 | 42.19 | 53.58 | 85.23 | 88.10 |
| Relative error | 8.2% | −4.8% | 4.0% | 4.7% | −4.1% | −3.8% | 3.2% |
| | $d_{FD3}$ | $d_{FD4}$ | $d_{SC}$ | $d_{Sill}$ | $d_{BAT}$ | $m$ | |
| Prediction | 103.42 | 125.88 | 1.66 | 198.91 | 37.80 | 425.92 | |
| FE result | 107.47 | 129.13 | 1.45 | 189.68 | 35.86 | 428.88 | |
| Relative error | −3.8% | −2.5% | 14.5% | 4.9% | 5.4% | −0.7% | |

High-fidelity FE verification confirmed improvements in all three optimization objectives. The parameterized body-structure mass decreased from 444.85 to 428.88 kg, corresponding to a reduction of 15.97 kg (3.6%). B-pillar chest intrusion and battery-neighborhood intrusion were reduced by 9.5% and 14.6%, respectively. As shown in Figure 15, the battery-neighborhood deformation decreased from 42.00 to 35.86 mm, with an absolute reduction of 6.14 mm. The optimized design showed lower inward deformation and greater local clearance in the battery-adjacent region. The selected design provides a balanced improvement across the three objectives.

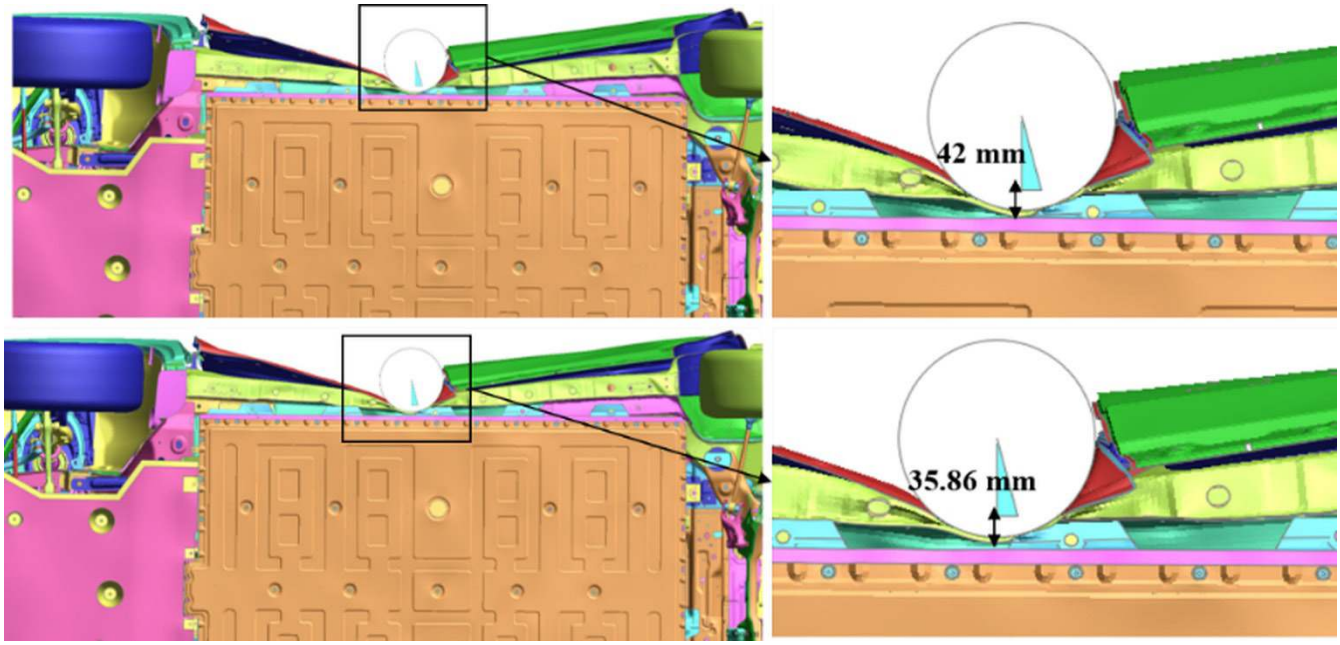



**Figure 15** Comparison of battery-compartment intrusion before and after optimization: baseline design (top) and optimized design (bottom).

## 5 CONCLUSIONS

This study evaluated a simulation-admissibility-aware surrogate-assisted workflow for mixed-variable electric-vehicle side-pole crashworthiness design. The classification, graph-based surrogate, SHAP analysis, and evolutionary optimization methods were used as established computational tools to investigate workflow-level numerical admissibility, structural-response relationships, and lightweighting–crashworthiness trade-offs. The main findings are summarized as follows:

- The automated SFE–LS-DYNA workflow evaluated 600 designs in the 68-variable mixed design space, including 528 simulation-admissible and 72 reproducibly inadmissible samples. Failures occurred at model reconstruction, mesh/keyword generation, LS-DYNA solution, and post-solution stages. Repeated stratified nested cross-validation yielded a ROC-AUC of 0.90 ± 0.02, a minority-class PR-AUC of 0.67 ± 0.04, an accepted-set validity of 0.95 ± 0.01, and an admissible-sample recall of 0.91 ± 0.02, while retaining 0.84 ± 0.02 of the candidates.

- MS-GIN achieved a mean $R^2$of 0.91 and a mean NRMSE of 6.5% across the 13 responses on the independent test set, with predictive accuracy comparable to Kriging, XGBoost, LightGBM, RBF regression, and the fully connected interaction model. Topology sensitivity showed that adding a permanent BP–SC edge produced only minor changes in prediction performance. The learned gated interactions were relatively strong along BP–LB and LB–SC, highlighting the coupling among the B-pillar, lower-body, and seat-cross-member modules.
- Bootstrap SHAP identified 18 influential variables accounting for 88.0% of the combined importance, with the strongest contributions concentrated in the B-pillar and lower-body regions. NSGA-II identified Pareto trade-offs among body-structure mass, B-pillar chest intrusion, and battery-neighborhood intrusion. High-fidelity FE verification of the selected compromise design reduced the three objectives by 3.6%, 9.5%, and 14.6%, respectively, while all remaining response constraints were satisfied.

## Acknowledgments

This research was funded by the MAJOR SCIENCE AND TECHNOLOGY SPECIAL PROJECT OF JILIN PROVINCE AND CHANGCHUN CITY, grant number 20240301010ZD.

**Author's Contributions:** Conceptualization, Lijun Cao; methodology, Lijun Cao; software, Lijun Cao; validation, Zhenjiang Wu and Hongbin Tang; formal analysis, Jianjiao Deng; investigation, Shibin Wang; resources, Zehui Huang; data curation, Ledan Liu; writing—original draft preparation, Lijun Cao; writing—review and editing, Zehui Huang; visualization, Qianxi Cheng; project administration, Xingbo Zhou; funding acquisition, Zehui Huang.

**Data availability statement:** Research data is only available upon request.

**Editor:** Rogério José Marczak.

## References

Alaiz-Rodríguez, R., Parnell, A.C. (2020). An information theoretic approach to quantify the stability of feature selection and ranking algorithms. Knowl.-Based Syst. 195:105745.

Amini, S., Van Nieuwenhuyse, I., Morales-Hernández, A. (2025). Constrained Bayesian optimization: A review. IEEE Access 13:1581–1593.

Audet, C., Caporossi, G., Jacquet, S. (2020). Binary, unrelaxable and hidden constraints in blackbox optimization. Oper. Res. Lett. 48:467–471.

Blank, J., Deb, K. (2020). pymoo: Multi-objective optimization in Python. IEEE Access 8:89497–89509.

Cawley, G.C., Talbot, N.L.C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11:2079-2107.

Chen, H., Zhang, Z., Li, W., Liu, Q., Sun, K., Fan, D., Cui, W. (2024). Ensemble of surrogates in black-box-type engineering optimization: Recent advances and applications. Expert Syst. Appl. 248:123427.

Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

China Automotive Technology and Research Center Co., Ltd. (2024). C-NCAP Management Regulation (2024 Edition), CATARC (Tianjin, China).

Choi, Y., Kumar, K. (2024). Graph neural network-based surrogate model for granular flows. Comput. Geotech. 166:106015.

Chombo, P.V., Laoonual, Y., Wongwises, S. (2021). Lessons from the electric vehicle crashworthiness leading to battery fire. Energies 14:4802.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6:182–197.

Fang, H., Rais-Rohani, M., Liu, Z., Horstemeyer, M.F. (2005). A comparative study of metamodeling methods for multiobjective crashworthiness optimization. Comput. Struct. 83:2121–2136.

Fang, J., Sun, G., Qiu, N., Kim, N.H., Li, Q. (2017). On design optimization for structural crashworthiness and its state of the art. Struct. Multidiscip. Optim. 55:1091–1119.

Franco, N.R., Fresca, S., Tombari, F., Manzoni, A. (2023). Deep learning-based surrogate models for parametrized PDEs: Handling geometric variability through graph neural networks. Chaos 33:123121.

Gao, F.L., Bai, Y.C., Lin, C., Kim, I.Y. (2019). A time-space Kriging-based sequential metamodeling approach for multi-objective crashworthiness optimization. Appl. Math. Model. 69:378–404.

Garrido-Merchán, E.C., Hernández-Lobato, D. (2020). Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes. Neurocomputing 380:20–35.

Hancock, J.T., Khoshgoftaar, T.M., Liang, Q. (2025). A problem-agnostic approach to feature selection and analysis using SHAP. J. Big Data 12:12.

Harrison, A., Christensen, J., Bastien, C., Kanarachos, S. (2020). Crashworthy structures for future vehicle architecture of autonomous pods and heavy quadricycles on public roads: A review. Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 234:3–16.

Ikpe, A.E., Owunna, I.B., Satope, P. (2017). Design optimization of a B-pillar for crashworthiness of vehicle side impact. J. Mech. Eng. Sci. 11:2693–2710.

Kelodjou, G., Rozé, L., Masson, V., Galárraga, L., Gaudel, R., Tchuente, M., Termier, A. (2024). Shaping up SHAP: Enhancing stability through layer-wise neighbor selection. Proc. AAAI Conf. Artif. Intell. 38:13094–13103.

Khatouri, H., Benamara, T., Breitkopf, P., Demange, J. (2022). Metamodeling techniques for CPU-intensive simulation-based design optimization: A survey. Adv. Model. Simul. Eng. Sci. 9:1.

Le Digabel, S., Wild, S.M. (2024). A taxonomy of constraints in black-box simulation-based optimization. Optim. Eng. 25:1125–1143.

Li, Q., Wang, Z., Li, L., Hao, H., Chen, W., Shao, Y. (2023). Machine learning prediction of structural dynamic responses using graph neural networks. Comput. Struct. 289:107188.

Li, X., Yang, Q., Wang, Y., Han, X., Cao, Y., Fan, L., Ma, J. (2021). Development of surrogate models in reliability-based design optimization: A review. Math. Biosci. Eng. 18:6386–6409.

Li, Y., Wang, C., Hou, W. (2025). A universal surrogate modeling method based on heterogeneous graph neural network for nonlinear analysis. Comput. Methods Appl. Mech. Eng. 437:117793.

Lian, F., Wang, D., Zhang, Z., Chen, H., Xu, W. (2025). Multi-objective optimization of lightweight and crashworthiness of automotive front-end structures based on stacked regression surrogate. Eng. Appl. Artif. Intell. 155:111138.

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

Liu, H., Xu, S., Wang, X., Yang, S., Meng, J. (2018b). A multi-response adaptive sampling approach for global metamodeling. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 232:3–16.

Long, F.X., van Stein, N., Frenzel, M., Krause, P., Gitterle, M., Bäck, T. (2025). Surrogate-based automated hyperparameter optimization for expensive automotive crashworthiness optimization. Struct. Multidiscip. Optim. 68:68.

Loshchilov, I., Hutter, F. (2019). Decoupled weight decay regularization. In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019; arXiv:1711.05101.

Lualdi, P., Sturm, R., Siefkes, T. (2023). A multi-fidelity successive response surface method for crashworthiness optimization problems. Appl. Sci. 13:11452.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2:56–67.

Lundberg, S.M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.

Mazumdar, A., Burkotová, J., Krátký, T., Chugh, T., Miettinen, K. (2024). Handling simulation failures of a computationally expensive multiobjective optimization problem in pump design. Eng. Appl. Artif. Intell. 136:108897.

Mortazavi Moghaddam, A., Kheradpisheh, A., Asgari, M. (2023). An integrated energy absorbing module for battery protection of electric vehicle under lateral pole impact. Int. J. Crashworthiness 28:321–333.

Moustapha, M., Galimshina, A., Habert, G., Sudret, B. (2022). Multi-objective robust optimization using adaptive surrogate models for problems with mixed continuous-categorical parameters. Struct. Multidiscip. Optim. 65:357.

Müller, J., Day, M. (2019). Surrogate optimization of computationally expensive black-box problems with hidden constraints. INFORMS J. Comput. 31:689–702.

Niehoff, M., Bestle, D., Kupijai, P. (2024). Model-based design optimization taking into account design viability via classification. Eng. Model. Anal. Simul. 1:1–12.

Nogueira, S., Sechidis, K., Brown, G. (2018). On the stability of feature selection algorithms. J. Mach. Learn. Res. 18(174):1–54.

Parisi, F., Ruggieri, S., Lovreglio, R., Fanti, M.P., Uva, G. (2024). On the use of mechanics-informed models to structural engineering systems: Application of graph neural networks for structural analysis. Structures 59:105712.

Platt, J.C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Advances in Large Margin Classifiers; Smola, A.J., Bartlett, P.L., Schölkopf, B., Schuurmans, D., Eds.; MIT Press: Cambridge, MA, USA; pp. 61–74.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., Battaglia, P.W. (2021). Learning mesh-based simulation with graph networks. In Proceedings of the 9th International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021; arXiv:2010.03409.

Qiu, N., Gao, Y., Fang, J., Sun, G., Li, Q., Kim, N.H. (2018). Crashworthiness optimization with uncertainty from surrogate model and numerical error. Thin-Walled Struct. 129:457–472.

Saito, T., Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 10:e0118432.

Setiawan, R., Salim, M.R. (2017). Crashworthiness design for an electric city car against side pole impact. J. Eng. Technol. Sci. 49:587–603.

Sharma, S., Kumar, V. (2022). A comprehensive review on multi-objective optimization techniques: Past, present and future. Arch. Comput. Methods Eng. 29:5605–5633.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15:1929–1958.

Tang, Z., Liu, F.J., Guo, S.H., Chang, J., Zhang, J.J. (2016). Evaluation of coupled finite element/meshfree method for a robust full-scale crashworthiness simulation of railway vehicles. Adv. Mech. Eng. 8:1–13.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J. (2019). Machine learning algorithm validation with a limited sample size. PLoS ONE 14:e0224365.

Viana, F.A.C. (2016). A tutorial on Latin hypercube design of experiments. Qual. Reliab. Eng. Int. 32:1975–1985.

Yang, J., Wu, Z., Wang, W., Zhang, W., Zhao, H., Sun, J. (2022). A surrogate-based optimization method for mixed-variable aircraft design. Eng. Optim. 54:113–133.

Zhao, Y., Li, H., Zhou, H., Attar, H.R., Pfaff, T., Li, N. (2024). A review of graph neural network applications in mechanics-related domains. arXiv:2407.11060.

Zhu, M., Mroz, A., Gui, L., Jelfs, K.E., Bemporad, A., del Rio-Chanona, E.A., Lee, Y.S. (2024). Discrete and mixed-variable experimental design with surrogate-based approach. Digit. Discov. 3:2589–2606.