Research Article

Computational Modeling and Smart Optimization For Ballistic Protection: A Systematic Review of Materials, Structural Response, Energy Systems, Control, and Predictive Maintenance

Rashmi Pandey[1], Kantimahanti Padmaja[2]

[1*]Asst. Professor, Department of Computer Application, ITM College,Gwalior Gwalior, Madhya Pradesh, Email: rashmi.pandey@itmgoi.in
[2]Asst Professor, CSE, Geethanjali College of Engineering and Technology, Medak, Hyderabad, Telangana .

Abstract

Computational modeling and intelligent optimization can be used to enhance ballistic protection, but results need to be carefully interpreted based on the evidence of validation. The applications of this systematic review to protective materials, structural response, model calibration, uncertainty assessment and post-impact damage detection are discussed. Searches of Europe PMC and OpenAlex, with an existing exploratory search register, were performed on 30 October 2026 for publications first available from January 2015 to October 2026. Using PRISMA 2020 reporting, 341 different records were screened, 16 full reports were assessed and 14 studies were included. Structured narrative synthesis was conducted; meta-analysis was not performed because applications, reference data, validation procedures, and outcome measures were heterogeneous. The approaches included were finite element or molecular simulations in conjunction with neural networks, Gaussian processes, evolutionary optimization, inverse identification, and uncertainty quantification. The mean indentation calibration error was decreased from 17.27% to 1.74%, the damage-segmentation F1-scores were in the range of 87.39–92.74%, and the computational screening workflow only required approximately 5.16% of the estimated exhaustive finite element (EFE) runtime. These results are different computational tasks and cannot provide a common measure of protective reliability. Independent physical validation, uncertainty characterization and reproducibility was limited. Independent physical benchmarks, separating calibration from evaluation, explicit model-discrepancy assessment, and accessible data and code are necessary for future progress, and must be considered as part of the steps involved.

Keywords: Ballistic protection; computational modeling; machine learning; surrogate modeling; intelligent optimization; model calibration; uncertainty quantification; systematic review.

1. Introduction

Ballistic protection uses materials and structures to reduce load transmission, deformation, and penetration during rapid impacts. These effects arise from interactions among contact, stress propagation, damage initiation, fracture, and energy dissipation [1]. Thus, protective performance is influenced by the constitutive behaviour, material arrangement, loading, and test and simulation boundary conditions [2]. Experimental results on protective laminates provide reference measurements, while numerical models can investigate configurations that are difficult to test physically [3].

However, modeling these processes is difficult. Simplified material descriptions, discretization options, contact formulations, and failure criteria can affect the predicted responses. Discretisation errors should be made explicit and checked for refinement [4]. Thus, computational efficiency must be considered alongside physical fidelity. Furthermore, agreement on a specific configuration does not imply applicability to other materials, geometries, or loading conditions. Experimental comparison is still necessary to determine which parts of a model are supported by measurements [5]. A key challenge in protective engineering is developing efficient computational methods that minimize investigative effort while maintaining a defensible link between modeled behavior, observed structural response, and the safety questions that drive testing across experimental and numerical conditions. Related research on protective structures is the finite element analysis of TPMS ceramic lattices [6]. Complementary experimental and numerical evaluations of energy-absorbing protection are shown for blast-loaded foam concrete [7].

Physics-based finite element models capture structural deformation and damage, while molecular simulations focus on material responses at a smaller scale. These methods can be complemented with machine learning surrogates, which approximate relations that can be learned from computational or experimental datasets, and can speed repeated evaluations [8]. Molecular simulations can also provide training data for probabilistic models, which multiobjective optimization then combines with [9]. Another contribution is made by algorithmic parameter

identification, which fits a constitutive description to reference measurements [10]. Uncertainty-aware approaches consider how incomplete parameter information affects estimates of protective response [11]. Existing impact damage can be detected and delineated using learning-based image analysis [12].

The above functions require different interpretations. Prediction estimates an outcome for given inputs, calibration optimizes parameters to match reference observations [13], and optimization determines inputs and/or parameters that meet specified objectives and constraints [14]. Multiobjective techniques can retain conflicting solutions instead of forcing a single optimum [15]. A method can also integrate all three functions, but each function must be evaluated separately. Accurate reproduction of simulation outputs shows surrogate agreement with the numerical value [16]. Similarly, fitting experimental curves indicates calibration performance. Neither result shows correct prediction performance on an independent physical test performed independently of the training data [17].

Existing reviews already examine the integration of machine learning with computational mechanics and the challenges of validating ballistic-response models. Pawlik et al. classify computational-mechanics applications into those related to modeling, surrogate prediction, inverse analysis and uncertainty quantification [43]. In [44] Wei et al. address the issue of multiscale modeling of three-dimensional woven ballistic composites, and make a clear distinction between calibration and independent predictive validation. The distinction is not new, however, as per the present review. The task of its is to develop a uniform comparison method between the different applications and reference data origins for prediction, calibration, optimization, uncertainty assessment, and damage detection in the context of protective-response.

All the studies are analyzed based on their computational purpose, their reference dataset, their evaluation procedure, and their conclusion in the field of engineering. This distinction is between the level of agreement with fitted observations, the prediction of withheld numeric outputs, experimental confirmation of a physical model, and predictions of independent physical measurements. It also identifies direct protective-response evidence from supporting material-scale and test-calibration investigations [18]. The review is intended to critically and comprehensively collate computational and intelligent approaches to ballistic protection and related testing, highlighting validation independence, uncertainty, computational efficiency and reproducibility. The results apply only to materials, scales, configuration, and loading studied. The below research questions guide synthesis:

- **RQ1:** Which computational and algorithmic approaches support protection and related testing?
- **RQ2:** What prediction targets, optimization objectives, and calibration methods are reported?
- **RQ3:** How do studies address numerical verification, experimental validation, and uncertainty?
- **RQ4:** What evidence supports improvements in accuracy, efficiency, or reliability?
- **RQ5:** Which limitations restrict reproducibility and transferability?
- **RQ6:** Which proposed application domains have direct, adjacent, or no included evidence?

Together, these questions guide the comparison of computational contributions without assuming that every application domain has equivalent evidence. In particular, the review distinguishes demonstrated damage detection from prospective maintenance, thermal-property modeling from energy-system evaluation, and calibration agreement from independently validated prediction. This approach links reported methodological advances to the strength and limits of the evidence supporting their practical use in protective engineering.

2. Systematic Review Methodology

2.1 Review design and protocol amendment

This systematic review uses PRISMA 2020 reporting and structured narrative synthesis [19]. Protocol version 1 focused on computational modeling in 6 proposed domains. After the pilot discovery and awareness of candidate discoveries, version 2 added a requirement for intelligent prediction, algorithmic optimization, inverse identification, or optimal uncertainty assessment. Modeling-only studies were thus no longer primary evidence but were contextual. It was a substantive amendment and not registered prospectively. The additional protocol is documented chronologically, with justification and consequences. Eligibility has been modified post-discovery; thus, the selection may have been biased against methods that were new to them and broader modeling conclusions cannot be derived. The revised criteria therefore need to be read in the context of the amendment history, and the limitations of the resulting evidence set.

2.2 Information sources and search strategy

We conducted searches on 30 October 2026, covering publications first available from 1 January 2015 through 29 October 2026. Europe PMC was searched through seven broad title/abstract queries and one focused query; OpenAlex supplied a complementary focused title search. An existing 33-record exploratory register provided another identification source.. Domain-specific queries additionally explored fluid dynamics, energy-related modeling, control, and maintenance. Exact executed strings, source-specific syntax, filters, reported counts, and complete response exports are provided in the supplement, consistent with transparent search reporting [20].

An existing exploratory register containing 33 records provided another route for identification. Crossref supported bibliographic verification and was not counted as an independent discovery database. Publisher platforms and public repositories supported full-text retrieval. We distinguished exported query occurrences from unique records because overlapping searches repeatedly returned the same publications. Preliminary OpenAlex syntax trials without complete exports were excluded from selection denominators. Search coverage was limited to the documented sources and interfaces; exhaustive coverage of every engineering database or systematic citation chasing was not claimed.

2.3 Eligibility criteria

Requirements included clear linkages to protective materials, protective structures, impact-test support and post-impact damage assessment. If a computational contribution was made, it was required to be predictive learning, an explicit algorithmic search or decision procedure, inverse parameter identification, or optimal uncertainty assessment. No algorithm was proposed for this optimization, and reporting an improved configuration or the term “optimization” alone was not enough. The multicriteria methods in contrast, were not excluded just because they did not employ machine learning, but had to be assessed in terms of their actual decision rules.

Peer-reviewed primary journal articles and substantive full conference reports were eligible when complete English texts were legally accessible within the publication window. Public author manuscripts associated with verified journal publications may represent eligible studies, with uncertainty retained regarding differences from the published version. Reviews, isolated datasets, unverified preprints, unrelated ballistic-transport applications, and generic engineering investigations without a protective connection were excluded.

Yarn-scale mechanics, tissue-simulant calibration, and high-rate compression studies were classified as adjacent evidence. Their inclusion supported methodological comparison without implying direct validation of complete protective systems. Applicability boundaries were retained during extraction and synthesis. English abstracts alone did not satisfy the full-text language requirement. Contextual and methodological references are not included in the 14-study synthesis; the publication-window, language, and access criteria apply to the primary evidence set. The inclusion and exclusion criteria governing the primary evidence set are summarized in Table 1.

Table 1: Inclusion and exclusion criteria for the primary evidence set.

Element	Inclusion requirement	Exclusion or retrieval disposition
Application	Protective materials, protective structures, relevant impact-test support, or damage assessment	Applications without an explicit protective-engineering connection
Computational contribution	Intelligent prediction, algorithmic optimization, inverse identification, or optimal uncertainty assessment	Modeling-only parameter sweeps without a qualifying algorithmic contribution
Publication	Peer-reviewed primary research or a substantive full conference report	Reviews, isolated datasets, unverified preprints, or conference abstracts without full reports
Access	Legally accessible complete report; associated public author manuscript permitted with a version caveat	Unavailable reports recorded as not retrieved, separately from full-text exclusions
Language	English full text	Non-English full text, even if an English abstract is available
Publication window	First publication from 1 January 2015 through 29 October 2026	Publications outside the specified window

2.4 Deduplication and study selection

Initial screening and source assessment were conducted using an AI-assisted workflow, and without independent duplicate human review or consensus adjudication. The following processing script encoded fixed retrieval and inclusion sets and set several exclusion categories using keyword rules. These outputs will not prove that all records received a decision based on criteria. The initial screening results are therefore to be re-evaluated before the study set and selection totals are deemed as final. Decisions should be traceable to the appropriate eligibility criterion and to the appropriate record/report in support of that criterion.

Potentially eligible reports were sought through public publisher and repository routes. Complete retrieved reports underwent eligibility assessment, with a primary reason recorded for each full-text exclusion. Landing pages and abstracts alone were insufficient for inclusion. Reports that could not be obtained were classified as not retrieved rather than excluded following full-text assessment. The original 23 retrieval decisions, including seven unavailable reports and two full-text exclusions, require reassessment alongside the corrected screening register.

2.5 Data extraction

The study identifiers were structured to allow the bibliographic information, application, computational methods, data origin, sample size, prediction/optimization target, comparator, outcome definition, validation procedure, uncertainty analysis, computational cost, reproduction, and limitations to be extracted. Record of source location, tables and report locations were made for checking of individual findings. Distinct studies of material scale were differentiated from direct protective-response investigations, as were test-calibration studies.

Calibration results, held-out simulation evaluations, and experimental comparisons were recorded separately. Metric labels and comparison conditions were preserved rather than converted into a common performance measure. Public author-manuscript use and publication-date discrepancies were documented where relevant. Information absent from accessible reports was treated as unreported rather than inferred. Extraction was AI-assisted, without independent duplicate extraction, and the resulting matrix supplied the basis for structured study-level comparison and narrative synthesis.

2.6 Study appraisal

Studies were assessed using a review-specific engineering appraisal covering six domains: objective clarity; assumptions and methods; numerical or algorithmic verification; independent empirical validation; uncertainty or sensitivity; and reproducibility. Each domain received a Yes, Partial, no, or Not applicable rating, accompanied by an explanatory judgment and source location. Partial ratings indicated limited or incomplete supporting evidence, while applicability decisions depended on the method being evaluated.

Experimental observations used for parameter fitting did not automatically establish independent predictive validation. Likewise, mesh descriptions were distinguished from demonstrated discretization checks, and sensitivity analysis was distinguished from comprehensive uncertainty propagation. The appraisal was not MMAT, a validated risk-of-bias instrument, or a numerical certainty score. Ratings were not summed into a total quality ranking. No independent appraisal reviewer participated, and judgments informed cautious interpretation of contributions. Supporting explanations and source locations for the study-appraisal ratings are provided in Table S5.

2.7 Synthesis approach

All 14 included studies entered structured narrative synthesis. The results were classified by computational function and compared based on the origin of the reference data, the independence of the evaluation, physical size and applicability. The agreement in calibration, withheld NP, baseline physical-model testing and independent physical prediction were understood individually. Appraisal judgments provided levels of strength and scope of each conclusion, but not numerical study weights. Differing data and missing information were explicitly kept. Prediction errors, segmentation F1-scores, calibration metrics, failure-probability bounds and runtime were not pooled, since they measure different quantities under different conditions. A meta-analysis and a formal reporting-bias test were not done, nor was a formal certainty grading.

3. Results and Thematic Synthesis

3.1 Study selection

Table 2 presents the reconciled totals for identification, screening, report retrieval, and final inclusion.

Table 2: Reconciled identification, screening, retrieval, and inclusion totals

Selection stage	Number
Exported occurrences, including the prior register	853
Duplicate occurrences and overlaps removed	512
Unique records screened	341
Title/abstract exclusions	318
Reports sought	23
Reports not retrieved	7
Full reports assessed	16
Full-text exclusions	2
Included studies	14

Figure 1 presents the reconciled identification, screening, retrieval, and inclusion process for the 14-study synthesis.

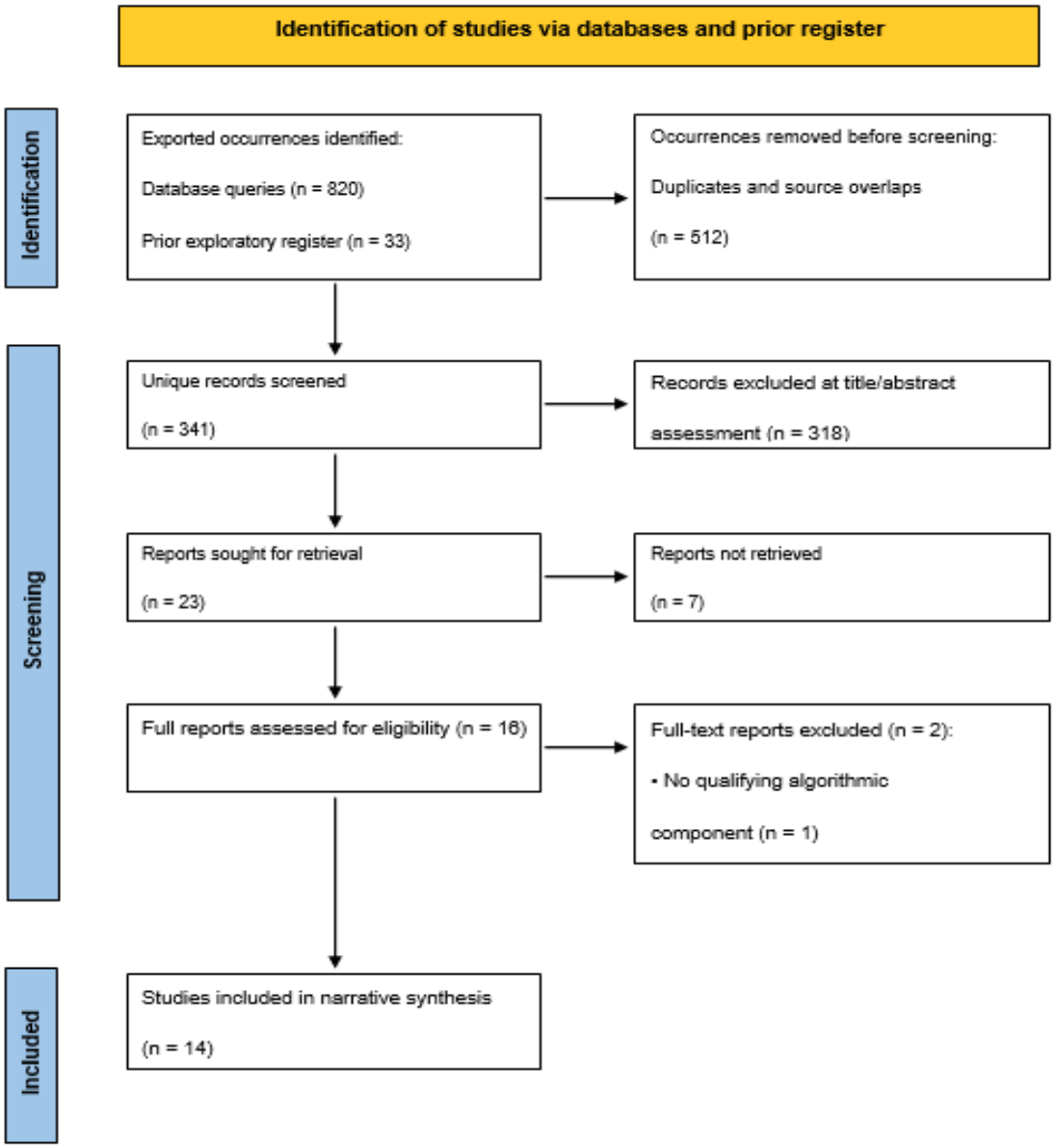


Figure 1. PRISMA 2020 flow diagram of study identification, screening, retrieval, eligibility assessment, and inclusion for the review of computational modeling and intelligent methods in ballistic protection and related testing [19].

3.2 Characteristics of the included evidence

T studies in this collection were published from 2018 to 2026 and cover a diverse range of protective and supportive applications. Protective-response investigations include spacecraft protection, metallic materials, ceramic protection, and composite laminates. Other topics discussed include backing-clay calibration, constitutive identification, tissue simulants, high-rate compression, yarn mechanics, and post-impact damage detection. Physical scales range from molecular simulations and textile interactions to structural panels, protection configurations, and thoracic test models.

In finite element analysis, the principal response model, or the training reference for a surrogate, is usually provided. Other methods include molecular dynamics, Gaussian processes, evolutionary optimization, gradient-based identification, and convolutional networks. Data sources include numerical campaigns, experimental reference curves, historical impact databases, and thermographic measurements. Validation then involves fitting reference observations and/or comparing modeled responses with experiments on held-out observations. A summary of these characteristics and a differentiation from methodological contributions that are adjacent to the direct protection evidence is given in Table 3. This ensures that supporting tests or nanoscale results do not become evidence of full validation of the protective system, but that their computational contribution can be measured within the context of their specific use. Table 3 summarizes the characteristics, reference-data origins, and validation approaches of all 14 included studies.

Table 3: Characteristics, data origins, and validation approaches of the 14 included studies

Study	Application and scale	Reference data and sample size	Validation approach	Evidence directness and applicability
S01 Artero-Guerrero et al. 2018 [21]	Composite laminate panels; structural scale	40 FE layups: 37 training, 3 tests; 4,096 configurations explored	Experiments support baseline FE; ANN tested on withheld FE cases	Direct protective-response modeling; surrogate evaluation uses FE references, with experimental support for the baseline model.
S02 Lei et al. 2021 [8]	Unidirectional fiber composite; microstructure to plate response	185 FE cases: 175 training, 10 tests	Internal FE holdout; no independent physical prediction test	Direct protective-response modeling; numerical holdout performance does not establish independent physical prediction.
S03 Peinado et al. 2021 [3]	UHMWPE laminates; structural scale	Layer-position task: 50 training + 6 tests; weight-ratio task: 25 + 7	Selected impact experiments support FE; ANN tests use FE outputs	Direct protective-response modeling; experimental FE comparisons and ANN-to-FE evaluations are separate validation checks.
S04 Wei et al. 2021 [12]	Curved CFRP specimens; post-impact imaging	8 specimens: 4 damaged, 4 undamaged; image halves; 75/12.5/12.5 split	Internal thermographic/manual labels; no external specimen collection	Direct post-impact damage assessment; does not validate remaining life or maintenance decisions.
S05 Sun 2022 [11]	AZ31B magnesium plate; structural impact	Constitutive and failure parameters in specified intervals	Numerical uncertainty calculations; no new independent test campaign	Direct protective uncertainty assessment; conclusions are conditional on parameter bounds and model assumptions.

S06 Kim et al. 2023 [25]	Backing clay; drop-test support	125 numerical parameter combinations; 3 reference drop cases	Indentation calibration and numerical re-evaluation	Adjacent backing-material calibration; does not validate complete protective assemblies or injury outcomes.
S07 Zelepugin et al. 2023 [26]	Copper specimens; high-rate material response	Measured and published Taylor-impact profiles; >900 numerical runs	Fits evaluated against impact profiles used in identification	Adjacent constitutive identification; fitted impact profiles do not establish prediction under independent conditions.
S08 Gupta et al. 2024 [9]	High-entropy alloys; nanoscale response	64 training + 16 validation MD samples	Numerical reference comparisons; candidate optima rechecked using MD	Adjacent nanoscale material exploration; numerical results do not establish macroscopic protective performance.
S09 Gómez-Garraza et al. 2024 [10]	SEBS and gelatin; material-curve scale	Synthetic curves and published experimental stress–strain curves	Known-parameter synthetic checks; fitting of experimental references	Adjacent test-material calibration; curve-fitting errors are not injury-prediction errors.
S10 Ryan et al. 2025 [18]	Whipple shields; structural protection	1,135 historical shielding experiments; 80/20 split stated for GP/SVC workflow	Historical experimental holdouts; physical-consistency checks	Direct shielding-response prediction using historical experiments; comparator matching and extrapolation limits remain relevant.
S11 Mutu 2025 [24]	Silicon-carbide protection; structural scale	FE-generated dataset; 70/30 training/test; mesh size among features	Internal numerical test set; literature/manufacturer comparisons	Direct protective-response modeling; internal numerical testing does not establish transfer to new physical tests.
S12 Bu et al. 2026 [31]	Biphasic composites; high-rate compression and thermal response	12,002 FE cases; 80/10/10 training/validation/test	Selected SHPB experiments support material model; CNN holdouts are FE	Adjacent high-rate compression and thermal co-design; does not certify complete ballistic or energy systems.
S13 Bakbak et al. 2026 [32]	Nano-B4C/aramid; yarn-interface scale	Measured uncoated and 0.3 wt.% states reused in fitting/evaluation	Agreement with fitted reference compositions; no independent concentration test	Adjacent yarn-interface mechanics; untested-concentration responses are model-generated.
S14 Wu et al. 2026 [27]	Physical/digital thorax; impact-test scale	2 investigated conditions, 3 repeats each; one effective parameter	Matched global displacement; three mesh levels; fitting/evaluation overlap	Adjacent impact-test calibration; global displacement agreement does not validate local tissue or injury outcomes.

3.3 Finite element–surrogate prediction of protective response

Four studies show the physics-to-machine learning linkage for protective-response prediction. Artero-Guerrero et al. used experimentally validated finite element modeling and an artificial neural network to investigate laminate stacking sequences [21]. They developed 40 layups: 37 for learning and 3 for testing. They observed about 3.7% and 7% learning and testing errors, respectively. The trained network explored 4096 symmetric configurations, but it did not produce design predictions equivalent to physical tests of all configurations.

To represent composite microstructures, Lei et al. adopted two-point correlation features and selected simulated absorbed energy as the prediction target [8]. They used 185 finite element cases, with 175 for training and 10 for testing. Gradient boosting regression was compared to the support vector and random forest approaches [22] [23]. It is reported to have an average error of 6.94% and a maximum error of 12.69% at the selected setting. These values reflect the degree of agreement. They don't compare the accuracy with independent physical impact measurements, particularly because of the specifics of the impact problem that were simplified in the underlying model.

Peinado et al. made the physical-model assessment and surrogate assessment more explicitly separated [3]. Their finite element model reproduced the investigated experimental ballistic limits with errors within 3.6%. The mean errors for held-out numerical cases were 0.55% for layer-position comparisons and 1.36% for weight-ratio comparisons in the two separate neural-network evaluations. These small surrogate errors are thus at a different validation level than the experiment-to-experiment comparison.

Mutu compared linear regression, ElasticNet, and multilayer perceptron models using explicit finite element outputs for ceramic protection [24]. With the reported training-test split of 70/30, the test results for residual velocity were R = 0.9884, MAE = 36.6704 and RMSE = 43.5008. The R label is kept since its meaning is not meant to be silently replaced with R^2. In the four studies, the perceived size of error is less important than the source of the data for the reference. S01 only had three withheld layups, and S02 only had 10 numerical cases. S03 reported experimental agreement of the FE baseline and ANN agreement with FE outputs separately [3]. These evaluations are used to numerically approximate the situation in the sampled configurations, with varying experimental support for the different models. They do not prove that optimized designs or unseen impact conditions are going to be predicted accurately on independent physical tests. The prediction target, comparator, validation evidence and applicability limit is thus maintained in Table 4 for each study. Table 4 compares computational targets and reported outcomes while distinguishing calibration agreement, numerical prediction, and physical validation.

Table 4: Computational targets, reported outcomes, validation evidence, and applicability limits across the included studies

Study	Computational target and reference	Reported result	Validation evidence and supported conclusion	Main applicability limit
S01 [21]	Ballistic-limit prediction; FE layup references	Learning error 3.7%; test error approximately 7%; predicted improvement up to 41%	Supports ANN approximation and numerical design exploration	Only 3 test layups; predicted best configurations not all physically tested
S02 [8]	Absorbed-energy prediction; FE references	GBR average error 6.94%; maximum error 12.69%	Agreement with numerical energy-absorption values	Small FE test set and simplified physical model; no independent impact validation
S03 [3]	Ballistic-limit baseline and ANN response prediction	FE-to-experiment error ≤3.6%; ANN-to-FE mean errors 0.55% and 1.36%	Separate experimental support for FE and numerical support for ANN	ANN errors are not physical-test errors; thermal/rate simplifications remain
S04 [12]	MWIR/LWIR damage-region segmentation	F1: MWIR 92.74%; LWIR 87.39% on the reported damaged test specimen	Supports internal agreement with thermographic/manual labels	8-specimen base; related image portions; no external validation or remaining-life outcome
S05 [11]	Uncertainty contributions and failure-probability bounds	Conservative bounds and interval-specific parameter rankings	Conditional uncertainty assessment under stated assumptions	Bounds are not observed failure frequencies; model discrepancy not fully established

S06 [25]	Inverse calibration of backing-clay indentation	Mean relative indentation error: 17.27% → 1.74%	Improved agreement for the reference drop-test conditions	Does not validate full protective assemblies, new conditions, or injury prediction
S07 [26]	Constitutive parameters fitted to copper impact profiles	Average study-defined solution-quality improvement approximately 10%; some fits worsen	Demonstrates calibration changes under the selected objective	Not a universal 10% accuracy gain; independent predictive holdout not established
S08 [9]	MD response prediction and competing material objectives	Reported response errors approximately within ±10%; optima rechecked by MD	Simulation-based surrogate and optimization support	Nanoscale numerical results do not establish macroscopic armor performance
S09 [10]	Visco-hyperelastic fitting of reference curves	Final RMSAPE: SEBS 7.484%, gelatin 19.69%; timings 132 and 13 min	Curve-fitting agreement; synthetic tests support the identification procedure	Not injury-prediction error; source contains inconsistent generation/Prony counts
S10 [18]	Shield-limit classification using historical experiments	EQNet test accuracy 79.4%; BLE bulk accuracy 76.5%	Modest reported improvement and physical-consistency evidence	Bulk and test accuracy are not necessarily matched comparison quantities; no new extreme-condition tests
S11 [24]	Residual-velocity prediction; FE-generated test data	Source-labeled test R 0.9884; MAE 36.6704; RMSE 43.5008	Internal numerical prediction performance	Retain R terminology; do not convert to R^2. No matched independent physical test demonstrated
S12 [31]	FE response prediction and workflow screening	Mean absolute relative errors approximately 3.9%; R^2 >0.98; workflow approximately 5.16% of exhaustive-FE time estimate	Numerical holdout accuracy; study-specific cost estimate	SHPB/thermal co-design is adjacent evidence; CPU/GPU and design-space assumptions limit cost comparison
S13 [32]	Stage-wise yarn pull-out curves; reference compositions	Reported R^2 generally >0.98 in crimp-extension and rupture stages	Agreement with measured compositions used for model development	Untested-concentration curves are model-generated; no independent concentration holdout
S14 [27]	Physical/digital displacement consistency calibration	NMAE: 4.8% → 3.4% and 12.5% → 4.9% for the two conditions	Improved global displacement agreement within investigated conditions	Not independent validation of local tissue, contact-force, energy, or injury outcomes

3.4 Algorithm-assisted calibration and inverse identification

A series of different relationships between model parameters and observed responses were considered in calibration studies. Kim et al. used numerical simulation, neural networks, and inverse tracking to determine backing-clay indentation parameters [25]. They trained on 125 parameter combinations and used 3 controlled drop-impact cases to limit parameter selection. The mean relative indentation error reported decreased from 17.27% to 1.74% with calibration. This improvement reflects agreement with the reference indentation condition and does not demonstrate the accuracy of a complete protective assembly.

Zelepugin et al. used Nesterov gradient descent to determine modified Johnson–Cook parameters for copper profiles after impact deformation [26]. Over 900 numerical calculations were reported. The average solution quality was enhanced by around 10%, but not consistently for all of the experimental fits. The study therefore concluded that the quality function should not be regarded as a universal prediction-accuracy percentage.

Gómez-Garraza et al. optimized the parameters for visco-hyperelastic descriptions of synthetic reference curves and experimental ones [10]. The final RMSAPE reported were 7.484% for SEBS and 19.69% for gelatin, the calibration times were 132 and 13 minutes, respectively. The results represent various types of material fitting problems and are not to be interpreted as a direct comparison of predictive accuracy. Identification behavior is tested by synthetic tests with known parameters while agreement with experimental curves for fitting is used to determine calibration. Reported generation and Prony-series counts should not be assumed to be consistent, but rather should be viewed as a reporting problem.

Wu et al. developed a physical digital thoracic response relationship using sequential response-surface optimization [27]. Under one condition, Displacement NMAE was lowered from 4.8% to 3.4%, and under the other condition, from 12.5% to 4.9%. Numerical assessment was possible only with mesh checks, whereas the experimental comparison was restricted to the global response under two conditions. Across these studies, improved agreement concerns the specified calibration observations. Transferability depends on parameter identifiability, objective-function choice, reference-data coverage, and separation of fitting from evaluation. The reported improvements do not establish prediction accuracy for new materials, untested loading conditions, local tissue responses, or injury outcomes.

3.5 Probabilistic modeling and uncertainty-aware assessment

Uncertainty was introduced into the reviewed methods in various mathematical and evidential ways. Sun [11] studied ambiguity in constitutive and failure parameters within specified ranges. Their framework included finite element response calculations, genetic-algorithm searches, concentration-of-measure inequalities, and optimal uncertainty quantification [28]. The outputs provided conservative failure-probability bounds under the given information and assumptions. The parameter importance rankings indicated which parameters matter in that context, not for any protective-material or loading situation.

Gupta et al. used molecular dynamics in conjunction with Gaussian-process surrogates and multiobjective optimization [9]. The surrogate used 64 samples for training and 16 for validation; reported prediction errors were about +/-10% for the modeled responses. Candidate optima were checked by additional molecular calculations. Macroscopic experimental validation information and a full quantification of physical-model discrepancy were not provided by Gaussian-process variability or by repeated computational comparisons.

To evaluate models, Ryan et al. used past Whipple-shield experiments and added ballistic-limit equations [18]. EQNet test accuracy was reported as 79.4%, while an equation-based bulk accuracy was 76.5%. Because these figures are not established here as evaluations on an identical test subset under matched conditions, their difference should not be interpreted as a controlled performance gain This continuity provides information, but it is not the same as experimental confirmation on unseen data. In all three studies, sensitivity is characterized by response variation [29], probabilistic prediction by model uncertainty, and failure bounds by restricted events. On their own, they do not capture the differences between the computational representation and the physical system; thus, they should not be understood as different forms of reliability. Variance-based sensitivity indices are based on the variability of model output due to uncertain input [30].

3.6 Physics-guided material prediction and coupled-property modeling

Physical information was used to facilitate learning in two ways: data generation and the structure of the model. Bu et al. used a three-dimensional convolutional neural network to simulate biphasic composites using an experiment-informed finite element method [31]. The dataset consisted of 12,002 data points, split into training (80%), validation (10%), and testing (10%) sets. The reported mean absolute relative errors for the tests were around 3.9%, and the R-squares were > 0.98. The framework also included a dynamic response analysis, in addition to thermal property

screening. It is an experimental investigation, not certification of complete projectile-resistant systems, as in high-rate compression. Whereas the other one, high-rate compression, is experimental, it is not the total composite system that is being investigated, but only selected configurations.

Bakbak et al. added mechanically meaningful stages to a physics-based modelling of yarn pull-out behaviour [32]. Admissible response shapes were preserved during the stage transitions and continuity constraints. The R-squared (R2) values were generally > 0.98 for crimp-extension and rupture areas. But reference compositions were measured for training, calibration, and validation, both coated and uncoated. The model was used to produce responses at other concentrations. Therefore, if one of the predictions of the concentration-dependent equations is validated through agreement with the measured references, this does not automatically validate the other predictions. Both are examples of ways to incorporate physical knowledge, but they apply only to the scale and response they were studied on. The adjacent material evidence is not evidence of certified performance, durability, or safety of an assembled protective system in independent ballistic testing, but rather evidence of high-rate compression and yarn pull-out.

3.7 Learning-based post-impact damage assessment

Wei et al. used deep learning to segment post-impact damage in curved CFCC samples examined by infrared thermography [12]. There were eight specimens, four of which were impacted and four undamaged. The mid-wave and long-wave infrared measurements were processed separately, and thermographic analysis and manual marking provided damage labels. Next, U-Net models [33] were used to make spatial damage predictions.

The F1-scores of the segmentation of the damaged test specimen reported were 92.74% for MWIR and 87.39% for LWIR. These measure level of agreement of the damage labels in the study. There were eight specimens in the evidence base and splitting into left and right parts did not increase specimen independence while increasing image samples. As a result, image level partitioning can leave some characteristics of the specimens in subsets that are different from the rest of the image. Performance should be viewed in this limited sample set as external specimen validation was reported as unknown. Damage segmentation is not a means to determine remaining useful life, future deterioration or a validated predictive-maintenance decision.

Furthermore, label boundaries are uncertain, as thermal-contrast boundaries and manually labeled areas may not match the actual damage boundaries. The study noted false-positive pixels around damaged areas and constraints arising from limited training samples. The study did not include external validation at an alternate facility or through additional sample collection. The contribution helps to detect and locate impact damage. It provides no further indication of remaining life, future degradation projections, or a proven relationship between identified damage and the timing of maintenance and/or replacement of the protective system.

3.8 Computational efficiency and reproducibility

There was some variation in the reporting of computational efficiency. Bu et al. presented a workflow that consists of finite element data generation, network training, and large-scale inference [31]. The total time amounted to about 5.16% of the estimated time for exhaustive finite element screening. This comparison assumes a specific design space and various CPU and GPU resources and should not be considered a hardware-independent speedup.

Gómez-Garraza et al. measured calibration times of 132 minutes for SEBS and 13 minutes for gelatin in their computing environment [10]. These timings reflect parameter fitting, not simulation of the entire protective system. Reproducibility also varied, with some studies providing code and/or data resources and others providing only limited reporting or information upon request. Availability statements were differentiated from verified executionThe reported reduction in computational runtime demonstrates a workflow-specific efficiency advantage; however, electrical energy consumption was not directly measured, and corresponding energy savings cannot be quantified from the available information [34].

3.9 Coverage of the original six domains

Materials and structural response have greater representation in the included studies, alongside supporting calibration and uncertainty methods. Thermal co-design seems to be featured in one of the neighbouring high-rate material investigations, and one imaging investigation offers a condition-monitoring forerunner. No included primary studies provide protective computational fluid dynamics, independently validated closed-loop control, or predictive maintenance under the revised eligibility rules. It is not the same as impact energy dissipation or as damage detection for maintenance prediction. The results reported here are for the included subset of counties. They do not show that the literature is not consulted, which is beyond the scope and limitations of the review.

4. Discussion

The accompanying evidence shows that intelligent techniques can speed up predictions of protective response, parameter identification, and exploration of response design. However, the usefulness of the reference data depends on its fidelity, representativeness, and coverage. These dependencies explain why we reach application-specific conclusions from the reviewed results, rather than a universal ranking of algorithms.

For RQ1 and RQ2, finite element–surrogate workflows, molecular simulations, evolutionary searches, inverse identification, probabilistic methods, and image segmentation address complementary purposes. Their targets range from response quantities and material parameters to damage regions and probability bounds. These purposes require different comparators and evaluation designs. The central contribution of the synthesis is consequently an evidence classification linking each reported improvement to the question actually evaluated.

In RQ3, evidence is differentiated into numerical verification, experimental comparison, and uncertainty assessment, but these pieces are not always combined. The prediction ability is provided by experiment-informed surrogate workflows, and calibration studies are mostly used to determine the agreement with the reference observations [3]. Probabilistic methods are used to deal with uncertainties, but they do not necessarily explain all the model discrepancies in physics [11].

Reported errors and timings are not interchangeable performance measures, but RQ4 provides computational benefits and improved agreement within certain workflows. Restricted data coverage, limited independent testing, limited reproducibility, and model assumptions are listed as constraints on transferability in RQ5. The evidence in RQ6 is largely comprised of materials and structural response, with associated thermal and monitoring response, but without any demonstration of control and maintenance capabilities.

These findings can be combined to validate the standards of matching to computational purpose. The real question is whether a method provides reliable evidence for the intended protective-engineering decision, rather than whether the reported accuracy is superior to that of a different method that was not related to the decision in question, but that was tested under different conditions.

There are four obvious types of relations that need to be explained. An agreement between calibration and reference observations is considered. Out-of-sample prediction using simulation indicates that the surrogate approximation to the numerical prediction is not provided during the fitting process. Experimental confirmation of a base physical model helps validate it for specific conditions. Evaluation of independent physical prediction needs to be done against measurements that have not been used for model development and selection [35]. These connections create different kinds of credibility tests; they should not be treated as the same validation claims. Figure 2 distinguishes the computational workflow and validation checks, explicitly separating reused development data from independent physical test data.

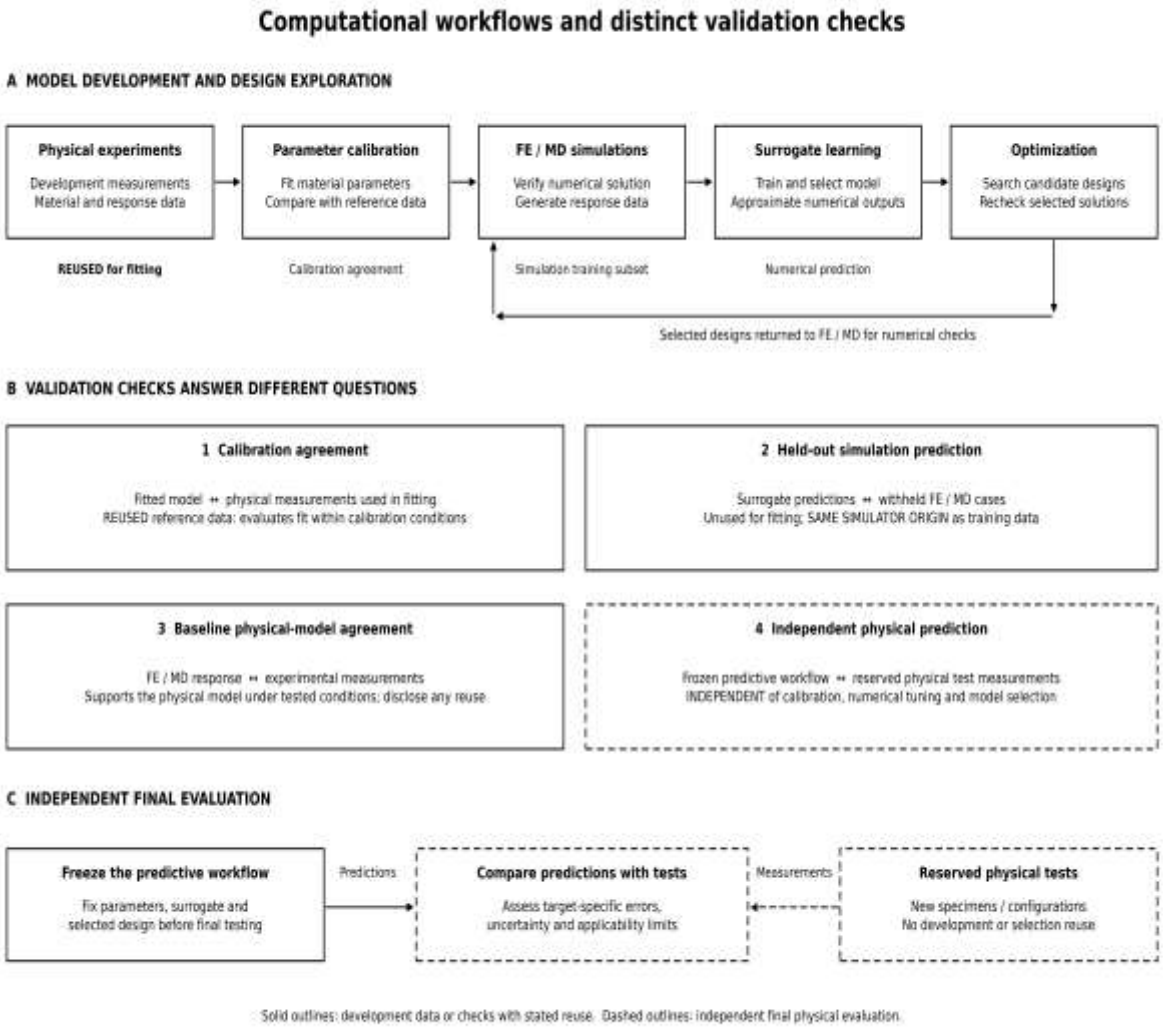


Figure 2. Computational workflows and distinct validation checks, distinguishing reused calibration and training data from independent physical test data

Peinado et al. show why the distinction matters: experimental comparison validates the finite-element baseline, and numerical references (smaller neural-network errors) relate to that distinction. On the other hand, Bakbak et al. use a set of reference concentrations that are measured for calibration and validation and subsequently predict concentrations in other ranges that have not been tested; predictions in these regions remain untested [32].

If specimens share similar material properties, geometries, or origins, small holdout regions can yield unreliable performance estimates. Image parts may be correlated, or numerical sets may be small, with little information relative to the number of samples [36]. Additionally, the transferability is limited by simplified constitutive behavior, contact assumptions, and limited loading ranges. Therefore, generalizing requires evidence relevant to the deployment conditions. Efficient screening may be warranted by agreement in a computational domain. Still, for safety-related decisions, the physical validity, uncertainties, and consequences of predictions beyond the range of demonstrated applicability must be examined.

These limitations that have been discussed suggest multiple priorities for reporting and evaluation. Research should document the entire validation process, specifying the measurements used for constitutive calibration, the simulations used for learning data, and the measurements used for final evaluation that are not part of the

constitutive calibration process. Model fitting, hyperparameter selection, and final testing must be done explicitly [37]. For grouped data splits, if multiple instances observe the same specimen, configuration, or source study, the splits should preserve these relationships rather than spreading related data indiscriminately [38].

If surrogate training follows numerical verification, do it after numerical verification to prevent learning algorithms from simply replicating unresolved discretization errors. The report of computational costs should contain information on the assumptions used in the comparisons, the training and inference processes, and the hardware used, especially if acceleration is being used as a contribution [31]. Reported comparisons would be easier to inspect and reproduce if they were referenced to accessible datasets, definitions of the parameters used, and versions of the software and implementation information. These recommendations reflect observed evidence gaps and are not meant to imply that a single evaluation protocol applies to all physical scales, prediction targets, and protective applications.

This synthesis has a few limitations. Discovery relied on two public indexes that don't cover all engineering literature. The difference in retrieval breadth was also achieved using the 'Europe PMC' title search and abstract retrieval and OpenAlex title retrieval. Requirements for English-language and accessible full-text may introduce language and access bias, and seven sought reports were unavailable. They may be excluded from the assessment, which could affect the evidence distribution [39].

The substantive protocol amendment came after the pilot discovery and awareness of candidate findings, which can lead to selection bias that cannot be removed in retrospect. There was no independent human duplicate screening, extraction, or appraisal of AI-assisted work, which could result in errors or differing judgments. The engineering appraisal was review-specific and is not comparable to a validated risk-of-bias instrument or formal certainty grading. Lastly, there was no scientific basis for pooling results quantitatively because of differences in scale, reference standard, evaluation conditions, and metric definition. Conclusions are thus general statements about the evidence considered and the lack of comprehensiveness and/or portability of computational performance.

5. Research Gaps and Future Directions

Future studies should focus on physical test datasets used only for independent testing. Measurements should be independent of material property fitting, finite element tuning, surrogate training, and algorithm selection. Use shared benchmarks to document material provenance, specimen preparation, geometry, boundary conditions, instrumentation, measurement uncertainty, and response definitions [40]. Matching evaluation conditions would allow proper comparison between computational methods, rather than differences in experimental design. Benchmark protocols should clearly separate baseline model validation from the use of optimised configurations, and should also mandate that any re-used observations should be reported. These practices would strengthen evidence for reported computational enhancements and yield credible predictions of untouched protective responses.

Holdouts of specimens and configurations should be included in evaluation whenever specimens and/or assumptions have physical commonality or numerical commonality. If there were many different test results across independent splits, initializations, and relevant operating conditions, this would help clarify performance variability [41]. Instead of using prediction intervals as a full description of error, uncertainty analysis should measure physical-model discrepancy in addition to measurement and surrogate uncertainty. All datasets and executable code should document the preprocessing operations, parameter definitions, software versions, and computational settings. Using consistent definitions, units, denominators, and comparison conditions would improve interpretability. It would also be beneficial to report failure to predict and sensitivity to method choice to strengthen the envelope of trust in which methods operate.

Protective fluid dynamics, coupled thermal problems, control, and prospective maintenance remain directions for further research and expanded review, rather than demonstrated capabilities of the included set. Relevant fluid–structure applications, numerical verification [42] and independent measurements of predicted responses are all required for Protective CFD. In thermal studies, the temperature-dependent properties and heat-transfer results are determined by experiments. For control studies, closed-loop stability, disturbance-response, and operational testing are necessary. Studies of maintenance need to be conducted longitudinally, that is, provide information on the condition, degradation, remaining life, and the maintenance decisions that have been made. These requirements do not infer that such research is not carried out elsewhere in the literature but that this evidence was not found in the retrieved literature.

6. Conclusion

This review categorizes the various computational and intelligent approaches to ballistic protection and related testing based on the purpose, reference data, validation process, and applicability. This 14-study synthesis is original.This is the original synthesis of 14 studies that includes surrogate prediction, inverse identification, optimization, uncertainty assessment, and damage segmentation. The strongest findings are related to application-specific numerical approximation and the better level of agreement, within the explored calibration conditions. There are inconsistencies in experimental support and in physical prediction independence. Transferability limited by small evaluation sets, shared data origins, model simplifications, and incomplete uncertainty reporting. The review is further constrained by search coverage and the post-pilot eligibility amendment, the lack of independent duplicate assessment, and the lack of duplicate assessment prior to the review. The number of final selections and the conclusions to draw depend on the methodological reassessment.

Priorities include criterion-based screening, verified extraction, independent physical benchmarks, appropriate grouped evaluation, explicit model-discrepancy assessment, and accessible implementation details. Applications should be directly relevant to underrepresented domains and domain-specific validation is needed. When the uncertainty, evaluation independence and physical assumption of the reference data apply to the decision which is being supported, computational and intelligent methods can support protective engineering.

References

[1] S. S. Morye, P. J. Hine, R. A. Duckett, D. J. Carr, and I. M. Ward, "Modelling of the energy absorption by polymer composites upon ballistic impact," Composites Science and Technology, vol. 60, no. 14, pp. 2631–2642, 2000. https://doi.org/10.1016/s0266-3538(00)00139-1

[2] H. L. Gower, D. S. Cronin, and A. Plumtree, "Ballistic impact response of laminated composite panels," International Journal of Impact Engineering, vol. 35, no. 9, pp. 1000–1008, 2008. https://doi.org/10.1016/j.ijimpeng.2007.07.007

[3] J. Peinado, L. Jiao-Wang, Á. Olmedo, and C. Santiuste, "Use of Artificial Neural Networks to Optimize Stacking Sequence in UHMWPE Protections," Polymers, vol. 13, no. 7, Art. no. 1012, 2021. https://doi.org/10.3390/polym13071012

[4] P. J. Roache, "Perspective: A Method for Uniform Reporting of Grid Refinement Studies," Journal of Fluids Engineering, vol. 116, no. 3, pp. 405–413, 1994. https://doi.org/10.1115/1.2910291

[5] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," Computer Methods in Applied Mechanics and Engineering, vol. 200, no. 25-28, pp. 2131–2144, 2011. https://doi.org/10.1016/j.cma.2011.03.016

[6] Z. Y. Xie, C. H. Yang, and W. N. Ma, "Numerical study on anti-penetration performance of TPMS structure," Journal of Ballistics, vol. 37, no. 1, pp. 104–112, 2025. (in Chinese; English title). https://doi.org/10.12115/ddxb.2023.11008

[7] Y. X. Gao, X. D. Zu, Z. X. Huang, and C. C. Huang, "Dynamic response and energy absorption characteristics of foam concrete under blast loads," Journal of Ballistics, vol. 37, no. 1, pp. 85–92, 2025. (in Chinese; English title). https://doi.org/10.12115/ddxb.2023.11012

[8] X. D. Lei, X. Q. Wu, Z. Zhang, K. L. Xiao, Y. W. Wang, and C. G. Huang, "A machine learning model for predicting the ballistic impact resistance of unidirectional fiber-reinforced composite plate," Scientific Reports, vol. 11, no. 1, Art. no. 6503, 2021. https://doi.org/10.1038/s41598-021-85963-3

[9] K. K. Gupta, S. Barman, S. Dey, S. Naskar, and T. Mukhopadhyay, "On exploiting nonparametric kernel-based probabilistic machine learning over the large compositional space of high entropy alloys for optimal nanoscale ballistics," Scientific Reports, vol. 14, no. 1, Art. no. 16795, 2024. https://doi.org/10.1038/s41598-024-62759-9

[10] S. Gómez-Garraza, R. de Santos, D. Infante-García, and M. Marco, "Visco-hyperelastic material model fitting to experimental stress–strain curves using a genetic algorithm and its application to soft tissue simulants," Scientific Reports, vol. 14, no. 1, Art. no. 18026, 2024. https://doi.org/10.1038/s41598-024-67603-8

[11] X. Sun, "Uncertainty Quantification of Material Properties in Ballistic Impact of Magnesium Alloys," Materials, vol. 15, no. 19, Art. no. 6961, 2022. https://doi.org/10.3390/ma15196961

[12] Z. Wei, H. Fernandes, H. G. Herrmann, J. R. Tarpani, and A. Osman, “A Deep Learning Method for the Impact Damage Segmentation of Curve-Shaped CFRP Specimens Inspected by Infrared Thermography,” Sensors, vol. 21, no. 2, Art. no. 395, 2021. https://doi.org/10.3390/s21020395
[13] M. C. Kennedy and A. O'Hagan, “Bayesian Calibration of Computer Models,” Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 63, no. 3, pp. 425–464, 2001. https://doi.org/10.1111/1467-9868.00294
[14] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient Global Optimization of Expensive Black-Box Functions,” Journal of Global Optimization, vol. 13, no. 4, pp. 455–492, 1998. https://doi.org/10.1023/a:1008306431147
[15] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002. https://doi.org/10.1109/4235.996017
[16] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, “Design and Analysis of Computer Experiments,” Statistical Science, vol. 4, no. 4, pp. 409–423, 1989. https://doi.org/10.1214/ss/1177012413
[17] J. Brynjarsdóttir and A. O’Hagan, “Learning about physical parameters: the importance of model discrepancy,” Inverse Problems, vol. 30, no. 11, Art. no. 114007, 2014. https://doi.org/10.1088/0266-5611/30/11/114007
[18] S. Ryan, H. Le, J. Berk, A. A. Kumar, and S. Venkatesh, “Physics-informed machine learning for predicting the ballistic limit of whipple shields,” International Journal of Impact Engineering, vol. 203, Art. no. 105364, 2025. https://doi.org/10.1016/j.ijimpeng.2025.105364
[19] M. J. Page et al., “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” BMJ, vol. 372, Art. no. n71, 2021. https://doi.org/10.1136/bmj.n71
[20] M. L. Rethlefsen et al., “PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews,” Systematic Reviews, vol. 10, no. 1, Art. no. 39, 2021. https://doi.org/10.1186/s13643-020-01542-z
[21] J. A. Artero-Guerrero, J. Pernas-Sánchez, J. Martín-Montal, D. Varas, and J. López-Puente, “The influence of laminate stacking sequence on ballistic limit using a combined Experimental/FEM/Artificial Neural Networks (ANN) methodology,” Composite Structures, vol. 183, pp. 299–308, 2018. https://doi.org/10.1016/j.compstruct.2017.03.068
[22] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001. https://doi.org/10.1214/aos/1013203451
[23] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/a:1010933404324
[24] H. B. Mutu, “Ballistic Performance Analysis of Silicon Carbide Ceramic Body Armor Using Finite Element Method and Machine Learning Algorithms,” Savunma Bilimleri Dergisi, vol. 21, no. 2, pp. 267–290, 2025. https://doi.org/10.17134/khosbd.1731217
[25] Y. Kim, Y. A. Kim, S. H. Park, and Y. Kim, “Hybrid numerical modeling of ballistic clay under low-speed impact using artificial neural network,” Journal of Mechanical Science and Technology, vol. 37, no. 2, pp. 865–873, 2023. https://doi.org/10.1007/s12206-023-0128-z
[26] S. A. Zelepugin, R. O. Cherepanov, and N. V. Pakhnutova, “Optimization of Johnson–Cook Constitutive Model Parameters Using the Nesterov Gradient-Descent Method,” Materials, vol. 16, no. 15, Art. no. 5452, 2023. https://doi.org/10.3390/ma16155452
[27] J. Wu, S. Wang, J. Song, J. Wu, B. Chen, and Z. Cai, “Verification and optimization of impact-response consistency in human thoracic targets for blunt ballistic impact testing,” Frontiers in Bioengineering and Biotechnology, vol. 14, Art. no. 1929295, 2026. https://doi.org/10.3389/fbioe.2026.1929295
[28] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz, “Optimal Uncertainty Quantification,” SIAM Review, vol. 55, no. 2, pp. 271–345, 2013. https://doi.org/10.1137/10080782x
[29] I. M. Sobol’, “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates,” Mathematics and Computers in Simulation, vol. 55, no. 1-3, pp. 271–280, 2001. https://doi.org/10.1016/s0378-4754(00)00270-6
[30] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, “Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index,” Computer Physics Communications, vol. 181, no. 2, pp. 259–270, 2010. https://doi.org/10.1016/j.cpc.2009.09.018

[31] L. Bu et al., “Machine learning-driven optimization of high-velocity impact resistance for three-dimensional biphasic composites,” Nature Communications, vol. 17, no. 1, Art. no. 9819, 2026. https://doi.org/10.1038/s41467-026-76632-y
[32] H. E. Bakbak, A. Onan, E. Bilisik, and K. Bilisik, “Predicting Interfacial Pull-Out Performance of Nano-B4C/Aramid Material with Stage-Wise Physics-Guided Machine Learning,” Polymers, vol. 18, no. 13, Art. no. 1618, 2026. https://doi.org/10.3390/polym18131618
[33] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015. https://doi.org/10.1007/978-3-319-24574-4_28
[34] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau, “Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning,” Journal of Machine Learning Research, vol. 21, no. 248, pp. 1–43, 2020. https://www.jmlr.org/papers/v21/20-312.html
[35] S. Varma and R. Simon, “Bias in error estimation when using cross-validation for model selection,” BMC Bioinformatics, vol. 7, no. 1, Art. no. 91, 2006. https://doi.org/10.1186/1471-2105-7-91
[36] S. Kapoor and A. Narayanan, “Leakage and the reproducibility crisis in machine-learning-based science,” Patterns, vol. 4, no. 9, Art. no. 100804, 2023. https://doi.org/10.1016/j.patter.2023.100804
[37] G. C. Cawley and N. L. C. Talbot, “On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation,” Journal of Machine Learning Research, vol. 11, no. 70, pp. 2079–2107, 2010. https://www.jmlr.org/papers/v11/cawley10a.html
[38] D. R. Roberts et al., “Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure,” Ecography, vol. 40, no. 8, pp. 913–929, 2017. https://doi.org/10.1111/ecog.02881
[39] J. Pineau et al., “Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program),” Journal of Machine Learning Research, vol. 22, no. 164, pp. 1–20, 2021. https://www.jmlr.org/papers/v22/20-303.html
[40] T. Gebru et al., “Datasheets for datasets,” Communications of the ACM, vol. 64, no. 12, pp. 86–92, 2021. https://doi.org/10.1145/3458723
[41] X. Bouthillier et al., “Accounting for Variance in Machine Learning Benchmarks,” Proceedings of Machine Learning and Systems, vol. 3, pp. 747–769, 2021. https://proceedings.mlsys.org/paper_files/paper/2021/hash/0184b0cd3cfb185989f858a1d9f5c1eb-Abstract.html
[42] I. B. Celik, U. Ghia, P. J. Roache, C. J. Freitas, H. Coleman, and P. E. Raad, “Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications,” Journal of Fluids Engineering, vol. 130, no. 7, Art. no. 078001, 2008. https://doi.org/10.1115/1.2960953
[43] Ł. Pawlik, J. Ł. Wilk-Jakubowski, D. Frej, and G. Wilk-Jakubowski, “Applications of Computational Mechanics Methods Combined with Machine Learning and Neural Networks: A Systematic Review (2015–2025),” Applied Sciences, vol. 15, no. 19, Art. no. 10816, 2025. doi: 10.3390/app151910816.
[44] Q. Wei, X. Liu, L. Zhu, H. Zhang, M. Wang, and X. Wu, “Three-dimensional woven fabric composites under ballistic impact: A systematic review of multiscale modeling strategies from fiber to structural scale,” Journal of Composite Materials, published online Sep. 19, 2026. doi: 10.1177/00219983261480233.

Supplementary material

Tables S1–S5 are supplied as separate CSV files in the accompanying supplementary archive.

Table S1. Complete executed search strategies, filters, dates, and query-occurrence counts.
Source: executed search logs; reporting follows PRISMA-S [20].

Table S2. Protocol history and substantive amendment from version 1 to version 2.
Source: review protocol, screening records, or cited study reports as identified within the table.

Table S3. Record-level screening, retrieval, and eligibility decisions for all 341 unique records.
Source: review protocol, screening records, or cited study reports as identified within the table.

Table S4. Complete study-level extraction with source locations, validation details, and applicability limits.
Source: review protocol, screening records, or cited study reports as identified within the table.

Table S5. Detailed study-appraisal ratings, supporting explanations, and source locations.
Source: review protocol, screening records, or cited study reports as identified within the table.