https://doi.org/10.24867/GRID-2026-p37

The Invisible Prototype: Ai-Powered Digital Twins for Resilient Graphic Packaging Design

ABSTRACT

Packaging is the first point of visual communication between a brand and its consumers, where graphic design and product security converge. Throughout transit, packaging endures drops, stacking compression, and vibrations that threaten its structural integrity, printed graphics, brand identity, and tamper-evident features. This paper explores how AI-powered digital twins - dynamic virtual replicas - can transform graphic packaging design by realistically simulating logistics conditions before physical production. By integrating AI, designers can predict deformation, print distortion, color degradation, and label damage under handling scenarios - enhancing prepress accuracy and digital image processing - while simultaneously optimizing package geometry, material selection, and visual layouts. The framework enables virtual evaluation of branding elements, authentication graphics, QR codes, and security seals, as well as protective performance, including functional security printing, to ensure packaging remains visually compelling and functionally reliable throughout the supply chain. Acting as an "invisible prototype," the digital twin offers a collaborative platform where graphic designers, prepress operators, engineers, and logistics specialists co-design solutions that drastically reduce physical prototyping and material waste. Rather than replacing creativity, AI augments design decision-making with data-driven insights. This approach effectively bridges graphic aesthetics with engineering performance, positioning intelligent simulation as a key enabler for secure, resilient, and visually impactful packaging systems.

R. Ragavaraja

Department of Printing and Packaging Technology,
College of Engineering Guindy,
Anna University,
Chennai, India

KEY WORDS
Graphic design, packaging design, digital twin, artificial intelligence, visual perception

1. Introduction

Packaging is the first physical encounter between a brand and a consumer. Color, typography, imagery, and form communicate expectations of quality before the contents are ever used. The package is, at the same time, a protective system that contains, preserves, and informs, and is therefore both a marketing tool and a logistics device (Fernandez et al., 2021). In practice, these two roles are engineered separately and in sequence. Structural designers fix the board, geometry, and closure to satisfy stacking and drop requirements, and graphic designers and prepress teams then decorate the result. The hand-off hides a coupling that is rarely examined. A shipper printed with a saturated brand red and a dense QR code can pass every structural test and still arrive with a bulged, scuffed, color-shifted panel and an unreadable authentication code, because none of those outcomes appear in a flat proof or on a press sheet. The designer who approved the artwork never saw the package that the consumer receives.

The consequences extend beyond appearance. Life-cycle studies increasingly count the food lost through inadequate packaging alongside the burden of the packaging itself, although many still rely on rough estimates of

the losses avoided (Williams & Wikström, 2011; Molina-Besch, Wikström & Williams, 2019; Zhang et al., 2026). Right-sizing protection is therefore a sustainability objective in its own right, and a pack that is lighter than it should be may be the more wasteful option once damage is counted. Plastic is the material most frequently assessed in such studies (Hemachandra, Hadjikakou & Pettigrew, 2024). As brand owners move toward fiber-based solutions, a further difficulty arises: fiber-based boards are highly sensitive to humidity and temperature, with stiffness changes of up to roughly 40 percent reported in realistic scenarios (Mann et al., 2026), so their printed surfaces behave differently under load and moisture. Replacing physical iteration with virtual evaluation addresses both pressures at once: fewer sample cartons, inks, and test runs are consumed, and the final design is better matched to the hazards it will meet. This holds only if the virtual evaluation can be trusted.

The digital twin is the natural vehicle for such evaluation. It is now a mainstream concept in manufacturing and engineering, with reviews covering its modeling, twinning, uncertainty quantification, and optimization technologies across many disciplines (Thelen et al., 2022). Surveys distinguish a digital model, a digital shadow, and a digital twin by the direction of data flow between the physical and virtual objects: manual in the first, one-way automatic in the second, and two-way automatic in the third (Yu et al., 2025). In packaging, however, twins have been mostly structural. The printed surface, with its color, image, code, and security layers, has been treated as an afterthought, and the authors' earlier review of augmented and virtual reality in graphic design concluded that immersive prototyping and AR-based authentication are promising but largely qualitative Nirmala, Raja & Kanchana, 2024). Immersive tools show a design; they do not predict how it will age in a truck.

This article, therefore, critically reviews across disciplines what is known about building AI-powered digital twins that predict how structure, print, and security features behave together under logistics loads. It is organized around four review questions (Figure 1). The first asks which logistics stressors damage the printed surface, and through which physical mechanisms. The second asks which twin architectures and AI methods can predict those effects, and how well they are validated. The third asks how security printing and authentication can be evaluated virtually, and the fourth asks what trade-offs with sustainability arise and what evidence is still missing.

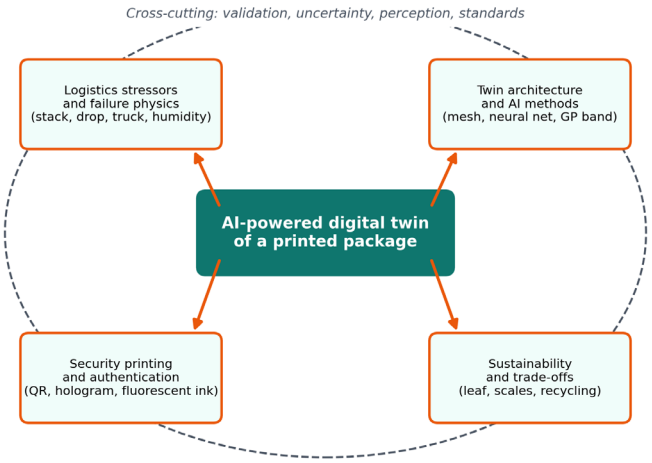
2. Methods

2.1. Scope and review design

This is a structured narrative review. Its scope is the prediction, before physical production, of how a printed package behaves in distribution. A narrative design was chosen over a systematic one because the question sits at the intersection of fields that do not share vocabulary, databases, or evaluation practices. Sources therefore span packaging science, printing and graphic technology, materials chemistry, solid mechanics, computer vision, scientific machine learning, and digital-twin engineering, since no single field covers the coupling. Peer-reviewed journals, conference proceedings, books, and international standards were considered. Items were retained when they (i) addressed structure, print, color, codes, security features, uncertainty, or validation relevant to a coupled prediction, (ii) described their methods transparently, and (iii) where applicable, validated predictions against physical experiments. Standards bodies were treated as primary sources for test conditions and measurement definitions (ASTM International, 2022; ISO, 2024).

2.2. Appraisal criteria

Each body of work was appraised qualitatively against four criteria drawn from verification and validation



» **Figure 1:** *Conceptual map – four questions around the AI-powered digital twin of a printed package*

practice (Jakeman et al., 2026): physical validation, quantification of uncertainty, evidence of generalization beyond the training conditions, and availability of data and documentation sufficient for others to reproduce the result. Physical validation asks whether predictions were compared with measurements of the real artifact. Uncertainty quantification asks whether the prediction carries an honest statement of confidence. Generalization asks whether performance was demonstrated outside the conditions used to build or train the model, and documentation asks whether another group could repeat the work. These four criteria structure the comparative reasoning in Section 3 and the research agenda in Section 4. Because the underlying literature is heterogeneous, no quantitative meta-analysis is attempted.

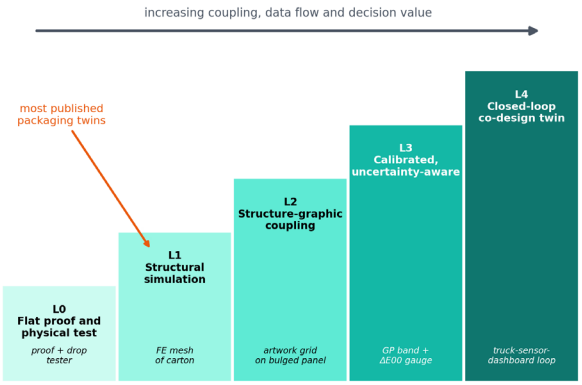
2.3. Coupling taxonomy and maturity ladder

The review uses a taxonomy in which a twin holds three coupled states: a structural state (geometry, material fields, deformation), a graphic state (artwork, ink film, color, code integrity, security features), and a logistics state (loads and climate). The structural state is driven by the logistics state; the graphic state is driven by the structural state and, independently, by the logistics state, for example, through abrasion and light. A twin that holds only the first two states is a conventional structural simulation. A twin that holds all three and propagates changes between them is what this article means by a coupled twin.

By analogy with the data-flow classification of Kritzinger et al. (2018), Table 1 proposes five maturity levels, from conventional practice to a closed-loop twin (Figure 2). The ladder lets readers locate any published system and see what is missing for the next level. Level L1 is where most published packaging twins sit: they predict compression or drop behavior well, but the printed layer is absent from the model.

2.4. Benchmark lens and metrics

To compare case studies on a common basis, a Brand-Integrity Index (BII) aggregates K normalized sub-scores (s_k) in [0, 1] with weights (w_k) set by the brand owner (Equation 1). Typical sub-scores are the color, where (ΔE_{00}) is the CIEDE2000 colour difference (Sharma, Wu & Dalal, 2005) and ΔE_tol is the brand owner's tolerance for that colour; image structure, measured



» **Figure 2:** *Maturity ladder for digital twins of printed packaging, from L0 (flat proof and physical test) to L4 (closed-loop co-design twin)*

Table 1

Proposed maturity ladder for digital twins of printed packaging

Level	Name	Characteristics	Typical evidence
L0	Flat proof and physical test	Artwork approved on a flat proof; structure tested on physical prototypes	ASTM International, 2022
L1	Structural simulation	Offline finite-element or analytical prediction of compression and drop; graphics ignored	Ge, 2022
L2	Structure–graphic coupling	Structural fields warp artwork; color and code change modeled analytically	ISO, 2024; Sharma, Wu & Dalal, 2005
L3	Calibrated, uncertainty-aware twin	AI surrogates, Bayesian calibration, perceptual metrics, and validation envelope	Kapteyn, Pretorius & Willcox, 2021
L4	Closed-loop co-design twin	Field data updates the model; multi-objective search with sustainability and security metrics	Liu et al., 2026

by the structural similarity index or a learned structure-and-texture distance (Wang et al., 2004; Ding et al., 2022); and code readability, graded under ISO (2024). Structural fidelity is expressed as the relative L2 error between predicted and measured fields (Equation 2).

$$\mathrm{BII} = 100 \sum_{k=1}^{K} w_k s_k , \qquad \sum_{k=1}^{K} w_k = 1 \tag{1}$$

$$\varepsilon_u = \frac{|\hat{u}-u|_2}{|u|_2} \tag{2}$$

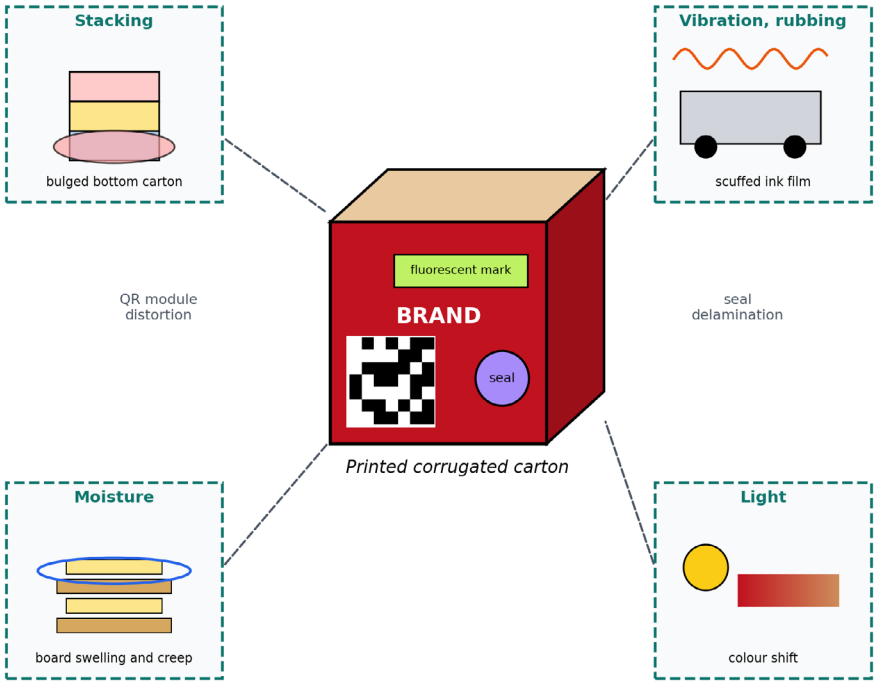
The weights are intentionally left to the brand owner. A luxury cosmetics carton and a bulk industrial shipper will not tolerate the same level of color drift, and a pharmaceutical pack will prioritize weight-code readability above all else. The index is therefore a bookkeeping device that makes assumptions explicit, not a universal quality score.

Six scenarios define what a twin must predict: S1 sustained stacking under humidity, S2 drop and impact, S3 vibration and abrasion, S4 light and climatic aging, S5 security-feature survival, and S6 the composite distribution cycle. Scenarios S1 to S3 are anchored in the shipping-test practice codified by ASTM International (2022); S4 and S5 are added because they act specifically on the printed layer; S6 combines all of them in the order a real shipment would experience. Comparison across methods should follow accepted validation practice (Kersken, Rojas & Strachan, 2025): fixed problems, simple baselines, declared validation envelopes, and reported uncertainty.

3. Results

3.1. Logistics stressors and failure physics of printed packaging

Distribution environments are codified in performance-testing standards that assemble handling shocks, vibration, and compression into representative cycles (ASTM International, 2022). Drops depend on height, orientation, and the fragility of contents, and they concentrate damage in crush zones, corners, and score lines; the shock-absorbing behavior of corrugated board has been studied through drop testing and impact loading (Ge, 2022). Transport vibration energy is typically concentrated at low frequencies, which matters for the printed layer because the same low-frequency motion that fatigues the structure also drives rubbing between adjacent cartons. Compression from stacking often governs corrugated shippers; strength is related to board edge crush, stiffness, and box geometry (Fitas, Schaffrath & Schabel, 2023), and later work has extended prediction to boxes with shifted creases on the flaps (Mrówczyński, Garbowski & Knitter-Piątkowska, 2021), to the optimal design of double-walled packaging (Mrówczyński, Knitter-Piątkowska & Garbowski, 2022), and to ventilated packaging and corrugated medium in the cold chain (Berry et al., 2022; Su et al., 2023). Homogenization of the flute-liner sandwich provides equivalent plate properties for finite-element analysis (Garbowski



» **Figure 3:** *Coupled failure pathways of a printed corrugated carton in distribution*

& Gajewski, 2021; Suarez et al., 2021), and analytical and numerical studies of panel stiffness continue to refine these properties (Suarez et al., 2021). Taken together, these sources show that the structural half of the problem is mature, quantitative, and well standardized.

The printed layer fails differently from the underlying structure (Figure 3). Four pathways dominate. First, panel bulging and score-line rotation distort images and two-dimensional codes. A QR symbology tolerates a degree of damage through built-in error correction, but not unlimited distortion, and print-quality grading penalizes loss of contrast, modulation, and decodability (ISO, 2024). Second, rub and scuff remove pigment and change gloss, so that a saturated solid loses its depth and a fine microtext line loses its edge. Third, color shifts occur through light-induced fading, binder yellowing, and changes in surface optics, all of which can be expressed as a perceptual color difference (Sharma, Wu & Dalal, 2005). Fourth, seals and tamper-evident features are loaded in peel and shear every time a carton is opened, restacked, or re-taped, and they can delaminate or lose their visual response.

Moisture couples all of these pathways. It changes the stiffness that carries the load: numerical homogenization shows that humidity and temperature changes in individual paper layers shift the effective stiffness of 3-ply and 5-ply corrugated board by up to roughly 40 percent (Cornaggia et al., 2023), and failure of the corrugated medium has been examined under simulated cold-chain logistics (Su et al., 2023). A stack that is stable in a dry laboratory can therefore bulge slowly in a humid warehouse; the bulge, in turn, strains the artwork. Moisture, temperature, and coating also change the mechanical properties of coated fiber-based substrates (Nienke, Kwade & Eggerath, 2022), which suggests that ink and coating adhesion, and hence the scuff resistance of the same print, will change as well. The failure of a printed carton is thus a chain, not a single event: climate weakens the structure, the structure deforms the graphics, and handling abrades what the deformation has already stressed. This chain is invisible to a flat proof and to a structural model that ignores the printed layer, and it is exactly what a coupled twin must reproduce.

3.2. Twin architecture and foundations

Digital-twin architectures commonly separate a physical layer, a data and sensing layer, a modeling layer, an intelligence layer, and a service or collaboration layer. This layering follows from the data-flow view of Kritzinger et al. (2018): a twin is distinguished from a mere model by the fact that information moves automatically in both directions between the physical object and its virtual counterpart. For printed packaging, the modeling layer must support structural, graphic, and logistics states, and the sensing layer must align heterogeneous data, such as load histories, spectrophotometric readings, high-resolution scans, and environmental logs, into a common frame. That alignment is more demanding than it sounds, because a bulge measured in millimeters on the board, a color difference measured in ΔE units on the ink, and a decoding grade measured on a letter scale are three different kinds of quantity.

Calibration is central. Bayesian calibration treats the discrepancy between the model and the measurement as an explicit term rather than hiding it within tuned parameters, and predictive twins have been formulated as probabilistic graphical models that update their state as new data arrive (Kapteyn, Pretorius & Willcox, 2021). Credibility also requires separating numerical error, input uncertainty, and model-form error, and stating the range of conditions over which the twin has been validated. In packaging, this means that a twin calibrated on one corrugated grade at moderate humidity must not be silently applied to a different grade at tropical humidity. The validation envelope is a first-class output of the twin, not a footnote.

Immersive augmented and virtual reality interfaces are useful in the collaboration layer, enabling distributed teams to inspect and annotate virtual packages (Nirmala, Raja & Kanchana, 2024), but they present predictions rather than create them. A visually convincing virtual carton is not evidence that the underlying prediction is right. For this reason, the review treats immersive interfaces as a service layer on top of a validated model, not as a substitute for it.

3.3. AI methods for prediction, inspection, rendering, and design search

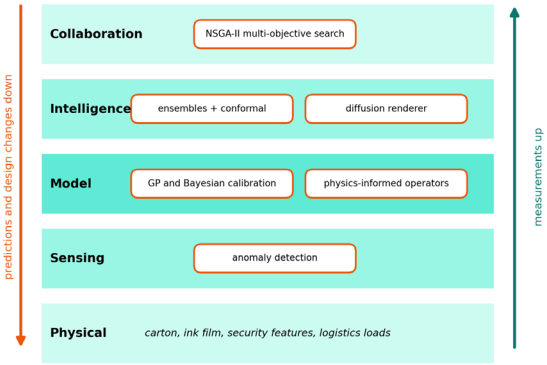
AI methods enter the twin at four points (Figure 4). First, surrogates accelerate expensive simulations. Physics-informed learning embeds governing equations and data together in one model (Würth et al., 2023), and neural operators learn mappings between function spaces, so that one trained network can return a full solution field for a new geometry or load while being orders of magnitude faster than conventional PDE solvers (Kovachki et al., 2023). For packaging, this means that a finite-element run that takes hours could be replaced by a surrogate model evaluated in milliseconds, which makes thousands of virtual design variants affordable. A known caution is that such networks can be difficult to train on stiff or multi-scale problems, and corrugated board, with its thin-layered walls, is exactly that kind of structure; surrogates should therefore be tested against held-out finite-element or experimental fields before being trusted.

Second, uncertainty methods make predictions trustworthy (Abdullah, Hassan & Mustafa, 2024),

and conformal prediction offers distribution-free coverage guarantees on top of any trained model (Zhou et al., 2025). Without such statements, a twin that predicts “ΔE = 2.1” gives a designer no way to know whether the true value could be 1 or 6.

Third, perception models inspect and render. Unsupervised anomaly detection learns what a good print looks like from defect-free examples and flags deviations without requiring a large library of labeled faults, which suits prepress and post-test inspection where defects are rare and varied; unsupervised approaches dominate industrial image anomaly detection because anomalous samples are costly to collect (Liu et al., 2024). Generative models, particularly latent diffusion models, can synthesize highly realistic images (Rombach et al., 2022). Conditioned on contact-pressure and strain fields, they are candidate renderers of scuffing, fading, and distortion, provided every output is labeled as a prediction and not mistaken for a photograph. Fourth, optimization methods search the design space. Bayesian optimization suits costly evaluations such as full finite-element runs (Garnett, 2023), and evolutionary multi-objective algorithms such as NSGA-II return a Pareto front of non-dominated designs, making the trade-offs among protection, graphic fidelity, security, and material use visible rather than collapsing them into a single score.

Taken together, these methods form a pipeline: surrogates predict fields quickly, uncertainty methods bound them, perception models turn them into what a viewer would see, and optimization uses all three to propose better designs. None of the individual components is new. What is new is placing them in a single loop around a printed package and validating the loop through physical distribution trials.



» **Figure 4:** *AI methods mapped onto the layers of the packaging digital twin*

3.4. Graphic fidelity and visual perception

The designer’s question is whether a viewer would notice or care. Perceptually uniform color difference, specifically CIEDE2000 (Sharma, Wu & Dalal, 2005), and the structural similarity index (Wang et al., 2004) quantify color and structure, while learned measures that combine structure and texture similarity track human ratings better on textured content and tolerate texture resampling (Ding et al., 2022). Each metric answers a different question. CIEDE2000 tells the prepress operator whether a brand color has drifted outside tolerance; structural similarity tells whether edges, fine lines, and typography have been preserved; a learned perceptual distance tells whether a textured or photographic panel still looks the same to a person even when a pixel-wise error is large.

A spectral measurement alone does not capture what a consumer sees, because color appearance also depends on illuminant, surround, and observer. A red that matches under the proofing lamp can look different under warehouse lighting or on a retail shelf. Metrics must therefore be treated as proxies and checked against human judgment for the specific viewing condition that matters.

Because a viewer judges a package as a whole rather than as isolated pixels, tolerances should be weighted spatially and semantically: a tight tolerance on the primary brand color and logotype, a looser one on background texture. A global average error would let a small but conspicuous scuff across the logotype pass unnoticed because the rest of the panel is perfect. This is why the Brand-Integrity Index is built from separate sub-scores with brand-owner weights, rather than a single image-difference value. Evidence that links such thresholds to purchase behavior remains thin, and the index leaves weights and tolerances to the brand owner for exactly that reason.

3.5. Security printing and authentication

Security features range from overt devices, microtext, and tamper-evident seals to covert luminescent inks and intrinsic, unclonable randomness. Physical unclonable functions (PUFs) are the most prominent route to the latter: a PUF is a physical object whose random, fingerprint-like structure is hard to replicate, and deep learning models are now used to authenticate it. Sun et al. (2023) built random Au-network PUFs by a fractal-guided film annealing strategy and paired them with a dynamic deep-learning authentication system with an expandable database, reporting no false positives and smartphone readout. Digital layers offer additional possibilities, such as AR-triggered authentication via printed codes, as discussed in the authors’ earlier review (Nirmala et al., 2024).

The shared gap is survival. The example above demonstrates feasibility and high verification accuracy under the conditions tested by its authors, but the literature

rarely quantifies how abrasion, moisture, and handling over a distribution cycle degrade signal retention or verification accuracy. A security feature that reads perfectly on a fresh sample but fails after a pallet journey has no commercial value, and a verification model trained only on pristine labels may reject genuine, scuffed packs. A coupled twin can evaluate this through scenario S5, by degrading the security layer with the same strain, abrasion, and moisture fields that act on the visible print and passing the degraded signal through the verification model. This also exposes a design trade-off: stronger, more complex security features may be harder to print robustly, and the twin is the place where security strength and robustness can be balanced.

3.6. Sustainability and trade-offs

Life-cycle assessment frames the trade-off, and recent reviews stress that the indirect benefit of packaging, the food loss it prevents, must be counted beside its direct burden (Espinoza-Orias & Lundquist, 2025; Hemachandra, Hadjikakou & Pettigrew, 2024). A heavier pack that prevents damage may therefore be environmentally preferable to a lighter pack that fails, whereas excess coating or laminate adds burden and can impair recycling. The answer is not “use less material” but “use the right material for the real hazard”, and the real hazard is exactly what a validated twin predicts. Barrier coatings are central to this discussion: coatings have been evaluated for improving the durability and water-barrier properties of corrugated cardboard (Marinelli et al., 2022). They protect against moisture and so preserve both structure and print, but they can alter surface energy, adhesion, and gloss, and may compromise recyclability. The twin must therefore carry the printed-surface consequences of a coating choice, not only its barrier benefit.

A twin that carries life-cycle indicators beside structural and graphic metrics lets teams see these trade-offs on a Pareto front before committing to production. A designer can see, for instance, that reducing board grade saves material but moves the stacking-bulge distortion of the brand panel beyond tolerance, or that a protective varnish lowers scuff damage while adding a recycling penalty. The twin also removes much of the board, ink, and energy that would otherwise be spent on physical samples. The missing evidence here is quantitative: few studies assign a measured environmental cost to graphic-quality loss, so the exchange rate between a unit of color error and a unit of material burden remains unknown.

3.7. Case studies

Three published bodies of work are analyzed through the maturity ladder and the benchmark scenarios. They are literature-based; no proprietary data were used, and details are given only to the level reported in the sources.

Case study 1: homogenization-based structural modeling of corrugated board under climatic loads. Cornaggia et al. (2023) used numerical homogenization and global sensitivity analysis to map how changes in humidity and temperature within individual paper layers alter the bending, compression, and shear stiffness of 3-ply and 5-ply corrugated board, and reported stiffness changes of up to approximately 40 percent in realistic scenarios. The work builds on homogenization approaches (Garbowski & Gajewski, 2021) and on classical strength theory (Fitas, Schaffrath & Schabel, 2023). This is a representative L1 twin: it addresses a real hazard, humidity, with a quantitative structural model. The printed surface, however, is outside the model, and the relations between climate and material constants are taken from the literature rather than calibrated against new physical tests reported in the study. Under the benchmark, the study would address structural-field error in S1 but provides no evidence for color, code, or security outcomes. Lesson: Structural modeling of climatic effects is a necessary foundation, and the missing step is the strain-to-graphic mapping, which is conceptually small but unvalidated.

Case study 2: AI-authenticated physical unclonable security features. Sun et al. (2023) realized random Au-network PUFs through a fractal-guided film annealing strategy and authenticated them with a dynamic deep-learning system that can expand its database, reporting no false positives and smartphone-based readout. The work shows that a security feature can be verified by a computational model with high accuracy. Under the benchmark, it establishes feasibility and verification accuracy under laboratory conditions (an L2–L3 characteristic for the verification model), but the scenario most relevant to packaging, S5, requires evidence on retention after abrasion, moisture, and handling over a distribution cycle, which the study, framed around tag stability and authentication accuracy, does not address as a packaging scenario. Lesson: the verification model should be inside the twin, so that the trade-off between security strength and robustness can be quantified.

Case study 3: predictive digital twins with calibration and decision-making. Kapteyn, Pretorius and Willcox (2021) formulated predictive digital twins as probabilistic graphical models that assimilate sensor data, update the twin’s state, and support decisions, and demonstrated the approach on an unmanned aerial vehicle. It sits within the digital-twin concept being reviewed across many disciplines (Thelen et al., 2022) and exemplifies L3–L4 behavior: calibrated, uncertainty-aware, and data-driven. Packaging has no comparable exemplar, so the study is a design template rather than a packaging result. The transferable ideas are explicit uncertainty, scalable updating from field data, and decision rules tied to risk. Lesson: a packaging twin should report a validation envelope and propagate uncertainty to graphic consequences, not only to structural ones.

Read side by side, the three cases mark out the territory. The first covers structure but ignores graphics, the second covers security verification but ignores distribution, and the third covers calibration and decisions but ignores packaging. No case covers more than one of the dimensions that a coupled twin needs, and none reaches L3 or L4 for a printed package. The three studies are complementary, which is encouraging: each supplies a component that the others lack. They are also the reason the maturity ladder is useful: it makes the missing step in each case visible at a glance.

4. Discussion

Cross-cutting findings. Four patterns emerge from the synthesis. First, the field is structurally mature but graphically thin: mature structural and homogenization models exist (Cornaggia et al., 2023), yet almost nothing links strain, abrasion, and moisture fields to color, code, and security outcomes. Second, the AI methods needed for the missing link already exist, from physics-informed and operator-learning surrogates to uncertainty-aware prediction and diffusion-based rendering (Liu et al., 2024; Rombach et al., 2022; Zhou et al., 2025), but they have not been benchmarked on packaging problems. Third, security features are evaluated at the laboratory level rather than through a distribution cycle (Sun et al., 2023). Fourth, sustainability is rarely coupled with the performance model, even though packaging burdens and avoided losses offset each other (Hemachandra, Hadjikakou & Pettigrew, 2024).

Why benchmarking matters. Adjacent fields progressed quickly once shared yardsticks existed: industrial anomaly detection is now organized around shared datasets, metrics, and settings (Mao et al., 2025), and scientific machine learning followed with shared benchmark suites for PDE problems (Takamoto et al., 2022). The scenarios, maturity ladder, and Brand-Integrity Index proposed here are a starting point for the same step in packaging. Credibility depends on declared validation envelopes (Kapteyn, Pretorius & Willcox, 2021) and on open, documented data (Gebru et al., 2021), without which two groups cannot tell whether they disagree about the physics or only about the test conditions.

Research agenda. Five priorities follow from the gaps, ranked by how much they unblock the others. The first is open data: there are no public datasets linking logistics histories to graphic outcomes, and a round-robin study across laboratories on a small set of reference artifacts, released with documentation, would give every later method a common test. The second is validated strain-to-graphic mapping: bulge, color, and code grade should be measured together in scenarios S1 and S6 and compared with predictions through the relative (L_2) error, (ΔE _{00}) , and the ISO (2024). The third is security survival: abrasion, moisture, and handling tests with verification models in the loop should quantify signal retention and verification accuracy. The fourth is calibrated generative rendering: diffusion renderers should be compared with real scans, with outputs labeled by confidence, and interval coverage checked. The fifth is to embed life-cycle indicators in the Pareto search, so that graphic quality, protection, security, and environmental burden are optimized together. These priorities emphasize open data, validated coupling, security survival, and calibrated rendering because they decide whether L3–L4 twins can be trusted in industry.

Implications for practice. Converters and brand owners can adopt twins incrementally: digitize material and test data (L0 to L1), couple structure to artwork (L2), introduce calibration and uncertainty (L3), and finally close the loop with field data (L4). Each step delivers value on its own. Even an L2 twin can warn that a particular panel layout puts the QR code over a high-strain zone, and that single warning can save a reprint. Immersive interfaces help teams collaborate but do not replace predictive validity (Nirmala et al., 2024). The organizational implication is equally important: the twin gives graphic designers, prepress operators, packaging engineers, and logistics specialists a shared object to argue over, thereby moving graphic and structural decisions from a sequence of handoffs to a single conversation.

Limitations of this review. The review is narrative, not a meta-analysis, and so each cited work represents a wider family of studies. The literature on coupled packaging twins is sparse, so two of the three case studies are adjacent-domain exemplars rather than packaging twins. Maturity levels and qualitative appraisals are the authors' judgment and should be tested by independent assessors.

5. Conclusions

AI-powered digital twins can turn graphic packaging design from a sequence of hand-offs into an integrated predictive loop in which structure, print, security, and sustainability are evaluated together before production. This review has shown that the necessary foundations already exist: structural models of corrugated packages, homogenized board properties, drop and vibration characterization, Bayesian calibration, physics-informed and operator-learning surrogates, uncertainty quantification, machine-vision inspection, generative rendering, and multi-objective optimization. Security printing has advanced through covert luminescent inks, unclonable physical features, and computational verification. What remains missing is coupling. Few studies connect strain, moisture, and abrasion fields to color shift, code readability, brand-color fidelity, and tamper-evident signal retention across a full distribution cycle. Consequently, the

field remains structurally mature but graphically thin, and security evaluation remains largely laboratory-bound.

The maturity ladder proposed here locates existing work from L0 flat proofing to L4 closed-loop co-design, while the Brand-Integrity Index and six benchmark scenarios provide a shared basis for comparison. Together, they expose the evidence needed for industrial trust: open datasets linking logistics histories to graphic outcomes, validated strain-to-graphic mapping, security-survival testing with verification models in the loop, calibrated generative rendering, and sustainability indicators embedded in Pareto search. The three case studies illustrate both the promise and the gap: structural simulation, AI-authenticated security features, and predictive twins with calibration exist separately, but no packaging exemplar yet couples them.

For practice, the implication is incremental adoption. Converters and brand owners can digitize material and test data, couple structure to artwork, introduce uncertainty-aware calibration, and close the loop with field data. Immersive interfaces can support collaboration, but predictive validity remains the decisive requirement. If the research agenda is pursued openly, the invisible prototype can become a practical decision-support environment in which designers explore bold visual and material solutions while knowing in advance how they will survive stacking, drop, vibration, humidity, and light. The result would be packaging that is more resilient, more secure, less wasteful, and more visually effective, benefiting brands, consumers, and the supply chain. Future work should therefore prioritize transparent benchmarks, independent validation, and cross-disciplinary collaboration between graphic design, prepress, packaging engineering, AI, security printing, and logistics.

6. References

Abdullah, A.A., Hassan, M.M. and Mustafa, Y.T. (2024) 'Leveraging Bayesian deep learning and ensemble methods for uncertainty quantification in image classification: A ranking-based approach', *Heliyon*, 10(2). DOI: 10.1016/j.heliyon.2024.e24188

ASTM International (2022) *ASTM D4169-22: Standard Practice for Performance Testing of Shipping Containers and Systems*. West Conshohocken, PA: ASTM International. DOI: 10.1520/D4169-22

Berry, T.M., Defraeye, T., Shrivastava, C., Ambaw, A., Coetzee, C. and Opara, U.L. (2022) 'Designing ventilated packaging for the fresh produce cold chain', *Food and Bioproducts Processing*, 134, pp. 121–149. DOI: 10.1016/j.fbp.2022.04.005

Cornaggia, A., Gajewski, T., Knitter-Piątkowska, A. and Garbowski, T. (2023) 'Influence of humidity and temperature on mechanical properties of corrugated board – numerical investigation', *BioResources*, 18(4), pp. 7490–7509. DOI: 10.15376/BIORES.18.4.7490-7509

Ding, K., Ma, K., Wang, S. and Simoncelli, E.P. (2022) 'Image quality assessment: Unifying structure and texture similarity', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5), pp. 2567–2581. DOI: 10.1109/TPAMI.2020.3045810

Espinoza-Orias, N. and Lundquist, L. (2025) 'Life cycle assessment of reusable food and beverage packaging systems: A proposal of good practice', *Journal of Cleaner Production*, 499, 145207. DOI: 10.1016/j.jclepro.2025.145207

Fernandez, C.M., Alves, J., Gaspar, P.D. and Lima, T.M. (2021) 'Fostering awareness on environmentally sustainable technological solutions for the post-harvest food supply chain', *Processes*, 9(9), 1611. DOI: 10.3390/PR9091611

Fitas, R., Schaffrath, H.J. and Schabel, S. (2023) 'A review of optimization for corrugated boards', *Sustainability*, 15(21), 15588. DOI: 10.3390/SU152115588

Garbowski, T. and Gajewski, T. (2021) 'Determination of transverse shear stiffness of sandwich panels with a corrugated core by numerical homogenization', *Materials*, 14(8), 1976. DOI: 10.3390/MA14081976

Garnett, R. (2023) *Bayesian Optimization*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108348973

Ge, C. (2022) 'A review and update of the corrugated fibreboard cushioning', *Packaging Technology and Science*, 35(9), pp. 701–715. DOI: 10.1002/PTS.2657

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé III, H. and Crawford, K. (2021) 'Datasheets for datasets', *Communications of the ACM*, 64(12), pp. 86–92. DOI: 10.1145/3458723

Hemachandra, S., Hadjikakou, M. and Pettigrew, S. (2024) 'A scoping review of food packaging life cycle assessments that account for packaging-related food waste', *The International Journal of Life Cycle Assessment*, 29(10), pp. 1899–1915. DOI: 10.1007/S11367-024-02349-Z

ISO (2024) *ISO/IEC 15415:2024: Automatic identification and data capture techniques — Bar code symbol print quality test specification — Two-dimensional symbols*. Geneva: International Organization for Standardization. Available from: https://www.iso.org/standard/76876.html [Accessed 11 October 2026].

Jakeman, J.D., Barba, L.A., Martins, J.R.R.A. and O'Leary-Roseberry, T. (2026) 'Verification and validation for trustworthy scientific machine learning', *Machine Learning: Science and Technology*, 7(2), 025055. DOI: 10.1088/2632-2153/AE59EC

Kapteyn, M.G., Pretorius, J.V.R. and Willcox, K.E. (2021) 'A probabilistic graphical model foundation for enabling predictive digital twins at scale', *Nature Computational Science*, 1(5), pp. 337–347. DOI: 10.1038/s43588-021-00069-0

Kersken, M., Rojas, G. and Strachan, P. (2025) 'Uncertainty of the predictions of different programs and modelling teams based on a detailed empirical validation dataset', *Journal of Building Performance Simulation*, 18(5–6), pp. 631–647. DOI: 10.1080/19401493.2024.2343883

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A. and Anandkumar, A. (2023) 'Neural operator: Learning maps between function spaces with applications to PDEs', *Journal of Machine Learning Research*, 24(89), pp. 1–97.

Kritzinger, W., Karner, M., Traar, G., Henjes, J. and Sihn, W. (2018) 'Digital Twin in manufacturing: A categorical literature review and classification', *IFAC-PapersOnLine*, 51(11), pp. 1016–1022. DOI: 10.1016/J.IFACOL.2018.08.474

Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F., Jin, Y. et al. (2024) 'Deep industrial image anomaly detection: A survey', *Machine Intelligence Research*, 21(1), pp. 104–135. DOI: 10.1007/S11633-023-1459-Z

Liu, X., Wang, H., Han, X., Zan, Y., Zhang, J. and Liu, G. (2026) 'A closed-loop design approach based on the combination of knowledge graph and digital twin: a high-speed train bogie case study', *Advanced Engineering Informatics*, 69. DOI: 10.1016/j.aei.2025.103912

Mann, G.S., Singh, B., Alzahrani, K.A., Khan, A., Kumar, R., Patil, A.Y. and Keshta, B.E. (2026) 'AI-powered innovations in natural fiber packaging for sustainability', *Journal of Natural Fibers*, 23(1). DOI: 10.1080/15440478.2026.2640280

Mao, Y., Chen, Z., Liu, Y., Dong, C. and Song, K. (2025) 'A survey on industrial image anomaly detection: methods, benchmarks and rethinks', *Measurement*, 256, 118377. DOI: 10.1016/J.MEASUREMENT.2025.118377

Marinelli, A., Diamanti, M.V., Lucotti, A., Pedeferri, M.P. and Del Curto, B. (2022) 'Evaluation of coatings to improve the durability and water-barrier properties of corrugated cardboard', *Coatings*, 12(1), 10. DOI: 10.3390/COATINGS12010010

Molina-Besch, K., Wikström, F. and Williams, H. (2019) 'The environmental impact of packaging in food supply chains—does life cycle assessment of food provide the full picture?', *The International Journal of Life Cycle Assessment*, 24(1), pp. 37–50. DOI: 10.1007/s11367-018-1500-6

Mrówczyński, D., Garbowski, T. and Knitter-Piątkowska, A. (2021) 'Estimation of the compressive strength of corrugated board boxes with shifted creases on the flaps', *Materials*, 14(18), 5181. DOI: 10.3390/MA14185181

Mrówczyński, D., Knitter-Piątkowska, A. and Garbowski, T. (2022) 'Optimal design of double-walled corrugated board packaging', *Materials*, 15(6), 2149. DOI: 10.3390/MA15062149

Nienke, T., Kwade, A. and Eggerath, D. (2022) 'Influence of moisture, temperature and bleaching on the mechanical properties of coated fiber-based substrates', *Coatings*, 12(9), 1287. DOI: 10.3390/COATINGS12091287

Nirmala, J.P., Raja, R.R. and Kanchana, M. (2024) 'Transforming graphic design – the impact of augmented and virtual reality on user interaction, security and product packaging optimization'. In: *Proceedings of the International Symposium on Graphic Engineering and Design (GRID 2024)*. Novi Sad: University of Novi Sad, Faculty of Technical Sciences, Department of Graphic Engineering and Design. DOI: 10.24867/GRID-2024-P19

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2022) 'High-resolution image synthesis with latent diffusion models'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685. DOI: 10.1109/CVPR52688.2022.01042

Sharma, G., Wu, W. and Dalal, E.N. (2005) 'The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations', *Color Research and Application*, 30(1), pp. 21–30. DOI: 10.1002/COL.20070

Su, H., Huang, C., Li, C., Zhao, Y., Zheng, D. and Duan, Q. (2023) 'Failure mechanism of the corrugated medium under simulated cold chain logistics', *ACS Omega*, 8(26), pp. 23673–23682. DOI: 10.1021/ACSOMEGA.3C01668

Suarez, B., Muneta, M.L.M., Sanz-Bobi, J.D. and Romero, G. (2021) 'Application of homogenization approaches to the numerical analysis of seating made of multiwall corrugated cardboard', *Composite Structures*, 262, 113642. DOI: 10.1016/j.compstruct.2021.113642

Sun, N., Chen, Z., Wang, Y., Wang, S., Xie, Y. and Liu, Q. (2023) 'Random fractal-enabled physical unclonable functions with dynamic AI authentication', *Nature Communications*, 14(1), 2185. DOI: 10.1038/s41467-023-37588-5

Takamoto, M., Praditia, T., Leiteritz, R., MacKinlay, D., Alesiani, F., Pflüger, D. and Niepert, M. (2022) 'PDEBench: An extensive benchmark for scientific machine learning', *Advances in Neural Information Processing Systems*, 35, pp. 1596–1611. DOI: 10.52202/068431-0117

Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B.D., Todd, M.D., Mahadevan, S., Hu, C. and Hu, Z. (2022) 'A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies', *Structural and Multidisciplinary Optimization*, 65(12), 354. DOI: 10.1007/S00158-022-03425-4

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) 'Image quality assessment: From error visibility to structural similarity', *IEEE Transactions on Image Processing*, 13(4), pp. 600–612. DOI: 10.1109/TIP.2003.819861

Williams, H. and Wikström, F. (2011) 'Environmental impact of packaging and food losses in a life cycle perspective: A comparative analysis of five food items', *Journal of Cleaner Production*, 19(1), pp. 43–48. DOI: 10.1016/j.jclepro.2010.08.008

Würth, T., Krauß, C., Zimmerling, C. and Kärger, L. (2023) 'Physics-informed neural networks for data-free surrogate modelling and engineering optimization – An example from composite manufacturing', *Materials and Design*, 231, 112034. DOI: 10.1016/j.matdes.2023.112034

Yu, W., Liu, G., Zhu, L. and Yu, W. (2025) 'Digital Twin: A literature review of concepts, technologies, and applications', *IEEE Access*, 13, pp. 183075–183095. DOI: 10.1109/ACCESS.2025.3624583

Zhang, L., Zhang, J., Cui, M., Qi, W., Huang, R., Su, R. and Zhang, K. (2026) 'Multifunctional bio-based packaging for perishable foods: Structural design, scalable fabrication, and versatile applications', *Advanced Materials*, 38(33). DOI: 10.1002/ADMA.202521880

Zhou, X., Chen, B., Gui, Y. and Cheng, L. (2025) 'Conformal prediction: A data perspective', *ACM Computing Surveys*, 58(2). DOI: 10.1145/3736575



© 2026 Authors. Published by the University of Novi Sad, Faculty of Technical Sciences, Department of Graphic Engineering and Design. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license 4.0 Serbia (https://creativecommons.org/licenses/by/4.0/deed.en).