Learning-Based Material Grading Through Multimodal Integration of Visual Features and Physical Properties

Britt-Mari Ostling
Department of Computer Science, University of North Texas, Denton, TX, USA.
brittmari.ostling@unt.edu

Joris van Waas
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
joris.vanwaas@ucf.edu

Abstract

Material grading relies on evidence from surface appearance and measured physical properties, but these evidence streams are usually processed separately. This article develops a methodological framework for learning-based material grading that integrates image-derived visual features with calibrated physical property vectors through a gated multimodal encoder. The framework formalizes grading as a noisy mapping from visual and physical observations to ordinal or categorical grades, extracts visual embeddings with convolutional or transformer encoders, standardizes physical measurements, and fuses modality-specific representations with a sample-adaptive gate. A composite objective combines classification or ordinal loss, a physics-informed consistency term that aligns predicted derived properties with measured quantities such as density-velocity relationships, and parameter regularization. Predictive distributions are calibrated and interpreted through entropy and cost-sensitive decision rules, allowing grades and abstention thresholds to reflect asymmetric error costs. The article also specifies evaluation logic, leakage controls, and failure modes, including spurious visual cues, sensor drift, missing modalities, and physics-model misspecification. Rather than reporting experimental outcomes, the analysis clarifies how multimodal architecture, physical inductive bias, uncertainty calibration, and decision theory interact in material grading. The framework provides a defensible basis for designing and auditing learning-based grading systems where visual evidence is informative but insufficient on its own.

Keywords

Multimodal learning; material grading; visual feature extraction; physical property calibration; uncertainty-aware decision making; physics-informed regularization

1. Introduction

Material grading is a decision problem in which heterogeneous evidence must be converted into a categorical or ordinal judgment about quality, structural suitability, or process state. In timber, engineered wood, composites, and other heterogeneous materials, surface appearance carries information about knots, fiber deviation, cracks, discoloration, and machining defects, whereas physical measurements carry information about density, moisture content, ultrasonic velocity, stiffness, and other bulk properties. Neither evidence stream is sufficient in isolation. Visual inspection can identify localized defects but may not reveal

internal density variation or moisture gradients. Physical testing can summarize bulk behavior but often loses spatial information and can be sensitive to sensor placement, specimen geometry, and environmental conditions. A learning system that grades materials therefore needs more than a classifier trained on images or a regression model trained on sensor values. It needs a principled treatment of multimodal evidence, measurement calibration, physical consistency, and decision costs.

Deep learning has made such integration technically plausible. General-purpose neural models can learn hierarchical representations from raw observations [1], and deep architectures have repeatedly improved image classification and recognition tasks when sufficient labeled data and appropriate inductive biases are available [2]. Convolutional networks exploit local spatial structure and weight sharing [3], while residual connections allow very deep networks to be optimized more reliably [4]. Transformer architectures extend attention-based representation learning to image patches and can capture long-range dependencies that convolutional receptive fields may not capture efficiently [5]. Large-scale pretraining and vision-language alignment have further improved transferability, especially when labeled domain data are limited [6]. These developments support a central premise of the present article: visual evidence in material grading can be encoded as a dense representation rather than reduced to handcrafted defect descriptors.

That premise is necessary but not sufficient. Material grading involves physical quantities whose meaning is constrained by measurement theory and material mechanics. A visual encoder may learn that dark streaks correlate with decay or that wavy grain correlates with strength loss, but it does not by itself know that ultrasonic velocity and density jointly determine a dynamic modulus, or that moisture content changes the interpretation of stiffness. Multimodal learning provides a vocabulary for combining heterogeneous sources, including early fusion, late fusion, and intermediate fusion [7]. Fusion strategies differ in how much modality-specific structure they preserve and how they model interactions [8]. For material grading, naive concatenation is a weak default because the two modalities have different noise structures, different dimensionalities, and different failure modes. A visual feature may be high-dimensional and spatially structured, while a physical property vector may be low-dimensional but measured with calibration error and batch effects. The integration mechanism should therefore allow the model to weight modalities at the sample level and should expose when one modality is unreliable.

The specific gap addressed here is the absence of a coherent learning framework that treats visual features and physical properties as complementary evidence while preserving physical interpretability and decision relevance. Many visual grading studies optimize classification accuracy without modeling measurement uncertainty or the cost of false acceptance. Many sensor-based grading studies use physical properties directly but ignore spatial defect patterns that are visible in images. Fusion studies often report improved predictive performance without specifying how the fused representation should behave under missing modalities, sensor drift, or distribution shift. These omissions matter in material grading because a false acceptance of a low-grade structural element can be more costly than a false rejection, and because physical measurements are often available only for a subset of specimens. A credible framework must therefore connect representation learning, calibration, physics-informed regularization, and cost-sensitive decision making.

This article develops such a framework. The contribution is methodological and analytical rather than experimental. It formulates material grading as a noisy mapping from

visual and physical observations to grades, specifies a gated multimodal encoder that learns sample-adaptive fusion weights, introduces a composite objective with a physics-informed consistency term, and derives a cost-sensitive decision rule from calibrated predictive distributions. The analysis also identifies failure modes that are specific to multimodal material grading, including spurious visual correlations, calibration drift in physical sensors, missing modality behavior, and misspecification of the physics constraint. The framework is intended to be implementation-ready at the level of architecture and objective design while remaining honest about the evidence required for empirical validation. The sections that follow define the grading problem, develop the multimodal learning framework, analyze uncertainty and decision logic, assess methodological risks, and discuss implications for material grading practice.

2. Material Grading Problem and Learning Formulation

Material grading can be viewed as a supervised learning problem in which a specimen is assigned a grade from a finite ordered or unordered set. In structural timber, grades may reflect strength class, defect severity, or suitability for a particular end use. In engineered composites, grades may reflect fiber orientation, consolidation quality, or void content. In recycled materials, grades may reflect contamination level or process consistency. The grade label is not merely a description of appearance; it is a decision variable that summarizes the expected performance of the material under defined conditions. This distinction matters because the learning target should be aligned with the decision that the grade will support. A classifier trained to reproduce historical labels may learn labeling conventions, inspector-specific thresholds, or batch effects rather than the physical evidence that justifies a grade.

Let x denote the visual observation of a specimen, such as a grayscale or color image, a multi-view image set, or an image stack from different illumination conditions. Let p denote a vector of measured physical properties, such as density, moisture content, ultrasonic velocity, electrical resistance, or stiffness measured by a non-destructive method. Let y denote the grade. The grading process can be written as a noisy mapping from observations to grades:

$$y = f(x, p) + \varepsilon$$

Here f is the unknown grading function, and ε represents observation noise, labeling inconsistency, and unmodeled material variation. If grades are categorical, y belongs to a finite set and f can be interpreted through class probabilities. If grades are ordinal, y can be treated as a latent continuous score with thresholds, or as an ordered label with a cumulative-link model. The formulation is deliberately general because material grading systems differ in whether the output is a class, a strength estimate, or a decision about acceptance. The noise term is not a nuisance to be ignored. It encodes the fact that two specimens with similar visual and physical observations may receive different grades because of hidden defects, measurement error, or human judgment. A useful learning framework should reduce this noise where possible and represent the remaining uncertainty explicitly.

Visual evidence in material grading is spatially structured. Knots, checks, splits, resin pockets, fungal staining, fiber deviation, and surface roughness produce local patterns that may be visible at different scales. A global image embedding can capture overall texture and color, but grading decisions often depend on the location and severity of a defect relative to the specimen geometry. Physical evidence is different. Density, moisture content, and ultrasonic velocity are scalar or low-dimensional quantities that summarize bulk behavior. They are often measured at a few locations and may not capture localized defects. The two

modalities therefore provide complementary information: vision provides spatial and surface evidence, while physical measurements provide bulk and mechanical evidence. Fusion is useful when the modalities are conditionally independent given the grade, but it can also introduce redundancy and correlated noise. The framework must therefore learn when to rely on each modality rather than assuming that more inputs always improve prediction.

Multimodal learning distinguishes between early fusion, which combines raw or low-level features before a joint model, late fusion, which combines modality-specific predictions, and intermediate fusion, which combines learned representations [7]. Early fusion can capture cross-modal interactions but is sensitive to dimensionality and noise. Late fusion is robust to missing modalities but cannot learn fine-grained interactions. Intermediate fusion offers a compromise by preserving modality-specific encoders and combining their representations in a learned space [8]. For material grading, intermediate fusion is attractive because visual and physical encoders have different architectures and different data requirements. A convolutional or transformer encoder can process images, while a multilayer perceptron or calibration layer can process physical properties. The fused representation can then support both grade prediction and physical consistency checks. The next section develops this architecture in detail.

3. Multimodal Representation and Fusion Framework

3.1 Visual Feature Extraction

The visual encoder maps an image or image set into a dense representation that summarizes appearance, texture, and defect patterns. Convolutional networks are a natural choice when the training set is moderate and local texture is important. They impose translation equivariance and local connectivity, which are useful inductive biases for surface inspection [3]. Residual connections can improve optimization when the encoder is deep [4]. Transformer-based encoders are an alternative when long-range spatial relationships or multi-view attention are important, but they often require more data or pretraining to outperform convolutional models on small domain-specific datasets [5]. Large-scale pretraining can improve transfer, especially when the target dataset is limited, but the pretraining distribution may differ substantially from industrial material images [6]. The visual encoder should therefore be selected with attention to data scale, defect scale, and the need for interpretability.

Let φ_θ denote the visual encoder with parameters θ. The visual embedding is

$$z_v = \varphi_\theta(x)$$

where z_v is a vector in a learned visual feature space. The encoder may produce a global pooled feature or a set of local features. Local features are useful when defect location matters, but they increase the dimensionality of the fusion problem. A practical compromise is to use a global feature for grade prediction and a pooled local feature map for auxiliary defect localization or attention-based interpretability. The visual embedding should not be treated as a complete description of material state. It captures what is visible under the imaging conditions, including illumination, surface preparation, and camera calibration. If those conditions vary across batches, the encoder may learn batch-specific shortcuts. Data augmentation, color normalization, and domain-adversarial training can reduce this risk, but they do not eliminate it. The fusion mechanism should therefore be able to downweight visual evidence when it is inconsistent with physical measurements.

3.2 Physical Property Encoding and Calibration

Physical properties require different treatment. A raw vector p may contain quantities with different units and scales, such as density in kilograms per cubic meter, moisture content as a percentage, and ultrasonic velocity in meters per second. Standardization is necessary before these values enter a neural network, but standardization alone does not address calibration drift or measurement error. A sensor may be biased relative to a reference instrument, or its response may change with temperature and humidity. The physical encoder should therefore include a calibration transformation that maps raw measurements into a comparable feature space. Let μ_p and σ_p denote location and scale parameters obtained from reference measurements or calibration specimens. The calibrated physical vector is $\tilde{p} = (p - \mu_p) \oslash \sigma_p$, where $\oslash$ denotes elementwise division. The physical embedding is

$$z_p = \psi_\eta(\tilde{p}), \tilde{p} = (p - \mu_p) \oslash \sigma_p$$

where ψ_η is a physical encoder with parameters η. The encoder can be a linear projection, a multilayer perceptron, or a small network with monotonic constraints if domain knowledge supports them. Monotonicity is not always appropriate. For example, higher density often correlates with higher strength in wood, but excessive density can also indicate abnormal growth or moisture-related defects. A monotonic constraint should be used only when the underlying physical relationship is well established and stable across the grading population. Otherwise, the model should be free to learn non-monotonic relationships while being regularized toward physical consistency.

The physical vector may also contain derived quantities. A dynamic modulus can be estimated from density and ultrasonic velocity using the relation $s = \rho v^2$, where ρ is density and v is ultrasonic velocity. This relation is a simplified model of wave propagation in a elastic medium and is not exact for anisotropic and viscoelastic materials. It is nevertheless a useful consistency check because it links two measured quantities to a property that is relevant to grading. The framework can include a physics-informed head that predicts s from the fused representation and penalizes disagreement with the measured or derived value. This is not a substitute for experimental validation. It is an inductive bias that discourages the model from learning fused representations that are physically implausible.

3.3 Gated Fusion and Architecture

The fusion mechanism should allow the model to weight modalities differently for different specimens. A specimen with a clear surface defect may be graded largely from visual evidence, while a specimen with ambiguous appearance but strong physical measurements may be graded largely from physical evidence. A fixed concatenation cannot express this sample-specific reliability. A gated fusion layer computes a gate vector that controls the contribution of each modality. Let W_v and W_p be projection matrices that map z_v and z_p into a common dimension. Let $[\cdot;\cdot]$ denote concatenation, σ the logistic sigmoid, and W_g, b_g the gate parameters. The gate is

$$g = \sigma\left(W_g\left[W_v z_v; W_p z_p\right] + b_g\right)$$

The fused representation is

$$h = g \odot W_v z_v + (1 - g) \odot W_p z_p$$

where $\odot$ denotes elementwise multiplication and 1 is a vector of ones. The gate is learned jointly with the encoders and the prediction head. It can be interpreted as a soft reliability estimate, but it should not be overinterpreted. A high gate value does not prove that visual evidence is correct; it only indicates that the model has learned to rely more on the

visual projection for the current input. The gate can nevertheless be useful for auditing. If the gate is consistently saturated toward one modality across a batch, the model may be ignoring the other modality or exploiting a shortcut. If the gate is highly variable for specimens with similar grades, the model may be unstable. Monitoring gate statistics during development can reveal these patterns.

Gated fusion has practical advantages over more complex tensor fusion methods. It has relatively few parameters, is compatible with missing modalities if the missing modality is represented by a learned token or a zero vector with a corresponding gate adjustment, and can be regularized toward balanced modality use. It also supports a natural extension to more than two modalities by using a softmax gate over modality-specific projections. The main limitation is that the gate is computed from the same features that are being fused, so it can be confounded by correlated noise. If both modalities are corrupted by the same batch effect, the gate may not detect the corruption. A separate reliability estimator based on sensor diagnostics or image quality metrics can be added as an auxiliary input to the gate. Such diagnostics might include image blur, exposure, sensor temperature, or the variance of repeated physical measurements. These variables should be used cautiously because they can also become shortcuts if they correlate with grade in the training set.

3.4 Objective Function and Physics-Informed Regularization

The model is trained with a composite objective that balances grade prediction, physical consistency, and parameter regularization. Let L_cls denote the classification or ordinal loss. For categorical grades, L_cls can be the cross-entropy between the observed grade and the predicted class probability. For ordinal grades, a cumulative-link loss or an ordinal regression loss is more appropriate because it respects the ordering of grades. Let L_phys denote a physical consistency term. If the physical head predicts a derived quantity s from the fused representation, then L_phys can be the mean squared error between the predicted and measured derived quantity. Let Θ denote all trainable parameters, and let λ_p and λ_r be nonnegative weights. The objective is

$$L = L_cls + \lambda_p\, L_phys + \lambda_r \|\Theta\|^2$$

The regularization term controls model complexity and reduces overfitting, especially when the visual encoder is large and the labeled dataset is small. The physics weight λ_p controls the trade-off between predictive fit and physical consistency. If λ_p is too small, the model may ignore the physical constraint and learn visually plausible but physically inconsistent representations. If λ_p is too large, the model may sacrifice grade accuracy to satisfy an approximate physical relation that does not hold for all specimens. There is no universal value for λ_p. It should be selected through validation on held-out data and through sensitivity analysis across material batches. The same applies to λ_r. The objective should be interpreted as a design choice that encodes a preference for physically consistent solutions, not as a guarantee of physical validity.

The physical consistency term requires care. A derived quantity such as $s = \rho v^2$ is based on simplifying assumptions about wave propagation and material homogeneity. In anisotropic materials, the relation between ultrasonic velocity, density, and elastic modulus depends on the direction of measurement relative to the grain or fiber orientation. If the physical vector p includes direction-specific velocities, the consistency term should be formulated per direction or replaced by a more general constitutive model. If the material is viscoelastic, the dynamic modulus may depend on frequency and temperature. Using a single scalar relation can

therefore introduce systematic bias. The framework should allow the physical head to be disabled or replaced by a material-specific model when the simplified relation is inappropriate. This is a substantive boundary condition, not a technical afterthought. Physics-informed regularization is useful when the physical model is approximately correct and harmful when it is confidently wrong.

4. Uncertainty, Calibration, and Grading Decisions

4.1 Predictive Distributions and Calibration

A grade prediction should be accompanied by a measure of confidence. The fused representation h can be passed through a linear output layer to produce logits for K grades. The predictive distribution is

$$P(y = k \mid x, p) = softmax(W_o\, h + b_o)_k$$

where W_o and b_o are output parameters and k indexes the grade classes. The softmax output is not automatically calibrated. Modern neural networks can be overconfident, especially when they are trained on small or imbalanced datasets [10]. Calibration is important in material grading because decisions often depend on thresholds. A model that predicts a grade with high confidence but is poorly calibrated can produce unreliable acceptance decisions. Temperature scaling, Dirichlet calibration, or isotonic regression can be applied to the logits or probabilities using a held-out calibration set. The calibration procedure should be evaluated with reliability diagrams and proper scoring rules, not only with accuracy. If the calibration set is not representative of the deployment population, calibration may not transfer. Batch effects, sensor drift, and changes in material source can shift the relationship between logits and true grade probabilities.

Uncertainty can be decomposed into aleatoric and epistemic components [9]. Aleatoric uncertainty arises from inherent noise in the observations and labels, such as ambiguous visual evidence or measurement error. Epistemic uncertainty arises from limited knowledge about the model parameters, such as regions of the input space with few training examples. Both are relevant to material grading. A specimen with an unusual defect pattern may have high epistemic uncertainty even if the image is clear. A specimen with noisy physical measurements may have high aleatoric uncertainty even if the model is well trained. Predictive entropy can be used as a simple summary of total uncertainty, but it conflates these components. For decision making, it may be more useful to estimate the probability that the true grade exceeds a threshold, or the probability that the predicted grade is wrong. These quantities can be derived from the predictive distribution and, where appropriate, from an ensemble or Bayesian approximation. The choice of uncertainty method should be matched to the decision rule and the amount of data available for calibration.

4.2 Cost-Sensitive Decision Rule

Material grading is not a neutral classification task. The cost of assigning a high grade to a low-quality specimen may differ substantially from the cost of assigning a low grade to a high-quality specimen. In structural applications, false acceptance can compromise safety, while false rejection can increase waste and cost. The decision rule should therefore incorporate an explicit cost matrix. Let C(c, k) denote the cost of making decision c when the true grade is k. The expected cost of decision c is the sum over grades of the predictive probability times the cost. The optimal decision is

$$c* = argmin_c \sum_k P(y = k \mid x, p)\, C(c, k)$$

This rule reduces to the standard maximum-probability decision when the cost matrix is symmetric, but it generalizes to asymmetric and ordinal costs. It can also support an abstention decision if the cost of deferring to a human inspector is included as an additional decision option. The cost matrix should be defined with domain experts and should reflect the actual consequences of grading errors, not merely the historical error rates of an existing process. If the cost matrix is uncertain, sensitivity analysis should be performed to determine how the optimal decision changes with different cost assumptions. A decision rule that is optimal under one cost matrix may be unacceptable under another. The framework should therefore separate grade prediction from grade decision. The model outputs calibrated probabilities; the decision layer maps those probabilities and the cost matrix to an action.

The decision rule also provides a principled way to set thresholds. Instead of choosing a threshold that maximizes F1 score or accuracy, the threshold is chosen to minimize expected cost. This is especially relevant for ordinal grades, where the cost of confusing adjacent grades may be lower than the cost of confusing distant grades. A cost matrix can encode this structure, and the decision rule can be solved by enumerating the possible decisions. If K is large, the enumeration is still inexpensive because K is typically small in material grading. The main computational cost is in obtaining calibrated probabilities, not in the decision step. This separation also makes the system easier to audit: the probability model can be evaluated independently of the cost assumptions, and the decision rule can be reviewed independently of the model architecture.

5. Methodological Assessment and Failure Modes

5.1 Data Requirements and Evaluation Logic

A multimodal grading framework requires paired observations: images and physical measurements collected from the same specimens, together with reliable grade labels. The pairing is a critical design requirement. If images and physical measurements are collected from different specimens or different locations on the same specimen, the model may learn spurious associations. The dataset should include multiple material batches, imaging conditions, and sensor calibration states to support robustness assessment. Splitting the data randomly at the specimen level can overestimate performance because specimens from the same batch may share visual and physical characteristics. Grouped splits by batch, source, or acquisition session are more appropriate for estimating generalization. If the physical measurements are expensive, the dataset may have many images but few paired physical vectors. The framework should be evaluated under different missingness patterns to determine whether the gate and physical encoder degrade gracefully.

Evaluation should not rely on accuracy alone. Material grading often has imbalanced classes and asymmetric costs. Macro-averaged recall, balanced accuracy, ordinal error, and expected decision cost are more informative than raw accuracy. Calibration should be assessed with expected calibration error, reliability diagrams, and proper scoring rules. Uncertainty estimates should be evaluated for their ability to identify misclassified or high-cost cases. If the model supports abstention, the coverage-risk curve should be reported. These metrics should be computed on held-out batches and, where possible, on an external validation set collected under different conditions. The evaluation protocol should be specified before model development to avoid selective reporting. Because the present article is methodological, these evaluation criteria are design requirements rather than reported outcomes. They indicate what would be needed to establish empirical support for the framework.

5.2 Failure Modes

Several failure modes are specific to multimodal material grading. The first is spurious visual correlation. A model may learn that a particular lighting condition, camera, or background correlates with grade because high-grade specimens were photographed in a different session. This is a form of shortcut learning. It can be mitigated by standardizing acquisition, augmenting images with nuisance transformations, and testing on batches with different acquisition conditions. The second is sensor drift. Physical sensors can change response over time, and a model trained on one calibration state may fail when the sensor is recalibrated or replaced. Regular calibration checks and drift-aware features are necessary. If the physical encoder includes calibration parameters, those parameters should be monitored and updated using reference specimens. The third is missing modality behavior. If physical measurements are unavailable at inference time, the model must either impute them, rely on visual evidence alone, or abstain. The gate can be extended to handle missing modalities by learning a default embedding and a reliability penalty, but this requires explicit training with missingness.

The fourth failure mode is physics-model misspecification. A physics-informed loss can improve generalization when the physical relation is approximately correct, but it can also bias the model when the relation is wrong for the material or measurement configuration. The dynamic modulus relation $s = \rho\, v^2$ is illustrative, not universal. For anisotropic materials, direction-specific measurements and constitutive models are needed. For moisture-sensitive materials, the relation may depend on moisture content. The physical consistency term should therefore be material-specific and should be validated against independent measurements. The fifth failure mode is class imbalance and label noise. Low grades may be rare, and historical labels may be inconsistent. Cost-sensitive decision rules can partly address imbalance, but they do not fix noisy labels. Label audits and inter-rater reliability studies are important. The sixth failure mode is over-reliance on the gate as an interpretability tool. The gate indicates learned reliance, not causal importance. A high gate value for visual evidence does not prove that visual evidence caused the prediction. Attribution methods and ablation studies are needed to support interpretability claims.

Real-time defect recognition and optimized decision making for structural timber jointing illustrate how perception and decision logic can be coupled in a production setting [13]. That connection is relevant here because material grading systems often fail not at the level of classification but at the level of decision integration. A defect detector may identify a localized feature, yet the grading decision depends on whether that feature exceeds a threshold, whether the specimen can be reworked, and what the cost of misclassification is. The framework developed in this article treats decision integration as a separate layer with an explicit cost matrix, which makes these considerations visible rather than burying them in a classifier threshold. The required reference is used only for this general point about coupling perception with optimized decision making; it is not treated as evidence for the performance of the present framework.

5.3 Physical Property Interpretation

Physical properties in material grading are not interchangeable. Density, moisture content, ultrasonic velocity, and stiffness have different measurement scales, different error structures, and different relationships to grade. Wood properties are anisotropic and hygroscopic, and their interpretation depends on grain direction and moisture content [14]. Acoustic methods can provide useful information about stiffness and internal structure, but the relation between

acoustic velocity and mechanical properties depends on the material and measurement configuration [15]. Ultrasonic techniques have been used to assess wood properties, but they require careful calibration and interpretation [16]. Acoustical properties of wood also vary with anatomical structure and are relevant to grading in musical instruments and other applications where vibration behavior matters [17]. These domain considerations imply that the physical encoder should not be a generic multilayer perceptron applied to raw numbers without calibration. It should be designed with knowledge of the measurement physics and the material's structure.

The same caution applies to visual features. Surface appearance can reveal defects, but it can also reflect surface preparation, lighting, and camera characteristics. A visual encoder trained on one set of imaging conditions may not generalize to another. Pretraining can help, but it can also introduce biases from natural images that are irrelevant to material surfaces [6]. The framework should therefore include acquisition standardization and domain randomization where possible. If multiple imaging modalities are available, such as color, near-infrared, or polarization images, the visual encoder can be extended to process them as separate channels or as additional modalities. The fusion gate can then learn which imaging modality is most informative for which specimen. This extension is conceptually straightforward but increases data requirements and model complexity.

6. Technical Discussion and Implications

The framework has several implications for the design of learning-based material grading systems. First, multimodal fusion should be treated as a reliability allocation problem, not merely a feature combination problem. The gate allows the model to allocate influence between visual and physical evidence, but that allocation must be trained and audited. A model that achieves high training accuracy by ignoring one modality may fail when the other modality is corrupted. Regularization toward balanced modality use, modality dropout during training, and missing-modality validation can reduce this risk. Second, physical consistency should be encoded selectively. Physics-informed regularization is most useful when the physical relation is well established and the measurement configuration matches the assumptions of the relation [11], [12]. It is least useful when the relation is approximate, anisotropic, or poorly calibrated. The framework should therefore allow multiple physical heads or material-specific consistency terms rather than forcing all materials through a single equation.

Third, uncertainty and calibration should be part of the grading pipeline, not an optional add-on. A grade without a calibrated confidence estimate is difficult to audit. A confidence estimate without a decision rule is difficult to use. The cost-sensitive decision rule connects the two and makes the consequences of grading errors explicit. This is particularly important in structural materials, where false acceptance can have safety implications. The cost matrix should be documented and reviewed, and the decision rule should be tested under different cost assumptions. If the decision rule is sensitive to small changes in the cost matrix, the system may need additional evidence or human review. Fourth, the framework should be evaluated with external validation and batch-level splits. Random splits can hide overfitting to acquisition conditions or material sources. External validation is the strongest evidence that a grading model will generalize.

The framework also connects to broader work in materials informatics and machine learning for materials science. Machine learning has been used to accelerate materials discovery, property prediction, and process optimization [18], [19]. The integration of domain

knowledge with data-driven models is increasingly recognized as important for engineering systems [20]. Material grading differs from materials discovery in that the target is often a decision rather than a property, and the data are often observational rather than experimentally designed. Nevertheless, the same principles apply: domain knowledge should constrain the model where appropriate, uncertainty should be represented, and evaluation should reflect the intended use. The framework developed here is a specific instantiation of these principles for multimodal grading. It does not claim that gated fusion or physics-informed regularization will always improve performance. It claims that these design choices address identifiable failure modes and should be evaluated explicitly.

There are also practical deployment considerations. Image acquisition, physical measurement, and decision logic may operate at different speeds. A real-time inspection system may need to make a decision before all physical measurements are available. The framework can support staged decisions by using visual evidence for an initial screening and physical evidence for a final decision. The gate can be conditioned on the available modalities at each stage. This is a form of decision-time multimodal learning. It requires training with modality availability patterns that match deployment. If the model is trained only with complete modalities, it may not learn how to behave when one modality is missing. Another practical consideration is maintenance. The visual encoder may need retraining when lighting or camera hardware changes, and the physical encoder may need recalibration when sensors drift. A versioned calibration and model update process is necessary. The framework should include monitoring for input distribution shift, prediction drift, and gate saturation.

The main limitation of the present analysis is that it is methodological. It does not report empirical performance, and it does not establish that the proposed architecture will outperform simpler alternatives on a specific material. That limitation is inherent in the article's scope. The framework is intended to make the design space explicit and to specify the evidence that would be required for validation. A future empirical study would need paired multimodal data, external validation, and a cost matrix defined with domain experts. It would also need ablations that isolate the contribution of gated fusion, physics-informed regularization, and cost-sensitive decision making. Without such ablations, it is difficult to know which component is responsible for any observed improvement. The framework provides a structure for those ablations.

7. Conclusion

This article has developed a methodological framework for learning-based material grading through multimodal integration of visual features and physical properties. The framework formulates grading as a noisy mapping from visual and physical observations to grades, encodes visual evidence with convolutional or transformer encoders, calibrates physical property vectors, and fuses modality-specific representations with a sample-adaptive gate. A composite objective combines grade prediction, physics-informed consistency, and regularization. Predictive probabilities are calibrated and translated into decisions through a cost-sensitive rule that reflects asymmetric grading costs. The analysis specifies data requirements, evaluation logic, and failure modes, including spurious visual correlations, sensor drift, missing modalities, physics-model misspecification, and overinterpretation of the gate. The framework does not assert experimental superiority. It provides a coherent design and audit structure for systems in which visual evidence is informative but insufficient on its own. The central implication is that material grading should be treated as an evidence-integration and decision problem, not merely as an image classification or sensor regression

problem. When visual and physical evidence are combined with calibrated uncertainty and explicit costs, the resulting system is more likely to be robust, interpretable, and aligned with the consequences of grading errors.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (Vol. 25, pp. 1097–1105).

4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).

5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations.

6. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (pp. 8748–8763).

7. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443.

8. Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6), 345–379.

9. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems (Vol. 30, pp. 5574–5584).

10. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (pp. 1321–1330).

11. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.

12. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422–440.

13. Luo, D., Gattas, J. M., & Tan, P. S. S. (2020, July). Real-time defect recognition and optimized decision making for structural timber jointing. In The International Conference on Computational Design and Robotic Fabrication (pp. 36-45). Singapore: Springer Singapore.

14. Ross, R. J. (Ed.). (2010). Wood handbook: Wood as an engineering material (Centennial ed.). USDA Forest Service, Forest Products Laboratory.

15. Bucur, V. (2006). Acoustics of wood (2nd ed.). Springer.

16. Beall, F. C. (2002). Overview of the use of ultrasonic technologies in research on wood properties. Wood Science and Technology, 36(3), 197–212.

17. Brémaud, I. (2012). Acoustical properties of wood in string instruments soundboards and tuned idiophones: Biological and cultural diversity. Journal of the Acoustical Society of America, 131(1), 807–818.

18. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: Recent applications and prospects. npj Computational Materials, 3(1), 54.

19. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 559(7715), 547–555.

20. Dimiduk, D. M., Holm, E. A., & Niezgoda, S. R. (2018). Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. Integrating Materials and Manufacturing Innovation, 7(3), 157–172.