International Journal of Plant & Soil Science

Volume 38, Issue 8, Page 700-721, 2026; Article no.IJPSS.163476
ISSN: 2320-7035

Artificial Intelligence in the Assessment of Seed Physiological Quality: A Critical Review of Spectral, Imaging and Learning-based Approaches for Digital Seed Technology

Mônica Araújo de Souza Santos [a*], Andréia Márcia Santos de Souza David [a], João Rafael Prudêncio dos Santos [b], Anne Karolina de Melo Souza [a], Rodrigo Silva Barbosa [a], Denner Junio Ramos Xavier [a] and Samanta Saira Souza [c]

[a] *Department of Agricultural Sciences, State University of Montes Claros (UNIMONTES), Janaúba Campus, Janaúba, MG, Brazil.*
[b] *Department of Education, Federal Institute of Northern Minas Gerais, Campus Araçuaí, Araçuaí, MG, Brazil.*
[c] *Department of Education, Federal Institute of Northern Minas Gerais, Campus Porteirinha, Porteirinha, MG, Brazil.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ijpss/2026/v38i86253

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/163476

Review Article

Received: 04/06/2026
Accepted: 07/08/2026
Published: 10/08/2026

**Corresponding author: E-mail: monycaaraujo@gmail.com;*

Cite as: *Santos, M. A. de S., David, A. M. S. de S., Santos, J. R. P. dos, Souza, A. K. de M., Barbosa, R. S., Xavier, D. J. R., & Souza, S. S. (2026). Artificial Intelligence in the Assessment of Seed Physiological Quality: A Critical Review of Spectral, Imaging and Learning-based Approaches for Digital Seed Technology. International Journal of Plant & Soil Science, 38(8), 700–721. https://doi.org/10.9734/ijpss/2026/v38i86253*

Abstract

Seed physiological quality, expressed through viability, germination capacity and vigour, determines stand establishment, uniformity and yield potential, yet the reference tests that define it remain destructive, labour-intensive and slow. Over the past decade, artificial intelligence, encompassing classical machine learning and deep learning applied to spectral and image data, has been advanced as a rapid, non-destructive alternative. This review critically evaluates that body of work and asks how far learning-based methods have genuinely improved on established seed testing rather than merely reproduced it under favourable conditions. Evidence was drawn from peer-reviewed studies applying near-infrared spectroscopy, hyperspectral and multispectral imaging, red-green-blue and X-ray imaging, and automated seedling analysis to viability, vigour and germination endpoints across cereal, oilseed, vegetable and tree species. The literature demonstrates consistently high classification and prediction performance under controlled, single-laboratory conditions, with hyperspectral imaging combined with variable selection and convolutional networks emerging as the dominant paradigm for viability and vigour. Deep learning frequently, though not invariably, outperforms chemometric baselines, and transfer learning, data augmentation and generative modelling have been introduced to counter the field's most persistent constraint, namely small and imbalanced datasets. Recurrent weaknesses temper these results: reliance on artificial ageing as a proxy for natural deterioration, narrow cultivar and environmental coverage, scarce external validation, inconsistent reference labelling, and limited interpretability. Reported accuracy therefore reflects internal discrimination more than demonstrated field relevance, and comparability across studies is undermined by heterogeneous endpoints and the absence of shared benchmarks. Progress toward operational digital seed technology will depend less on novel architectures than on representative multi-season datasets, standardised protocols, biologically grounded reference labels and prospective validation against germination and field emergence. The synthesis distinguishes well-supported conclusions from preliminary claims and sets out prioritised research directions to move the field from proof of concept toward dependable, transferable decision support.

Keywords: Seed vigour; seed viability; hyperspectral imaging; deep learning; near-infrared spectroscopy; non-destructive testing; germination prediction; digital agriculture.

1. Introduction

Seeds are the biological and economic foundation of cropping systems, and the physiological condition of a seed lot governs the speed, completeness and uniformity of emergence that ultimately shape yield. Seed physiological quality is a composite property, most usefully separated into viability, the capacity of a seed to germinate under favourable conditions, and vigour, the sum of properties that determine rapid, uniform emergence and normal seedling development across the range of field environments a lot is likely to encounter (Marcos Filho, 2015; Reed et al., 2022). These attributes decline as seeds age and deteriorate, and the rate of decline depends on genotype, maturation environment, harvest handling and storage conditions (Nadarajan et al., 2023; Reed et al., 2022). Because vigour differences are often invisible in a standard germination test conducted under optimal conditions, seed technologists have long relied on a battery of stress-based and biochemical assays to expose them, and the interpretation of these tests remains partly empirical (Marcos Filho, 2015).

The reference methods that underpin seed certification, principally the standard germination test and the tetrazolium viability test, are accurate and biologically meaningful, but they are destructive, require several days to weeks, depend on trained analysts, and carry appreciable subjectivity at the point of seedling evaluation (Marcos Filho, 2015; Rahman & Cho, 2016). Vigour tests such as accelerated ageing, electrical conductivity, cold testing and controlled deterioration add further time and inter-laboratory variability. For a seed industry operating on compressed processing schedules and tightening quality tolerances, this creates a structural tension: the tests that best describe physiological quality are the least suited to rapid, high-throughput decision-making on incoming and outgoing lots (Rahman & Cho, 2016). The search for rapid, objective and non-destructive alternatives is therefore long-standing, and encompasses machine vision, visible and near-infrared (NIR) spectroscopy, hyperspectral and multispectral imaging, soft X-ray radiography, thermal imaging and chlorophyll fluorescence (Feng et al., 2019; Rahman & Cho, 2016; Xia et al., 2019).

The stakes attached to this measurement problem are substantial. Cropping systems depend on the timely supply of seed lots that meet certification thresholds, and errors in quality assessment propagate directly into establishment failures, replanting costs and yield shortfalls, while unnecessarily rejected lots that in fact contain adequate proportions of viable and vigorous seed represent avoidable waste in an increasingly constrained supply chain (Nehoshtan et al., 2021). Vigour has acquired renewed importance as cropping systems confront more variable and stressful field environments, because it is vigour rather than laboratory germinability that governs emergence under the suboptimal conditions that climatic variability makes more frequent (Reed et al., 2022). The commercial pressure for rapid, objective and non-destructive assessment is therefore not merely a matter of laboratory convenience; it reflects a real mismatch between the tempo of modern seed processing and the several days to weeks required by the reference tests that most faithfully describe physiological quality (Marcos Filho, 2015). Artificial intelligence has been positioned as a transformative response to this mismatch, and the strength of that positioning is precisely what requires scrutiny.

What has changed over the past decade is not the sensing hardware alone but the analytical layer applied to it. Spectral and image data are high-dimensional, correlated and noisy, and the extraction of quality-relevant information from them is fundamentally a pattern-recognition problem. Classical machine learning, using engineered spectral or morphological features with classifiers such as partial least squares discriminant analysis, support vector machines and random forests, established that viability and vigour leave detectable optical and structural signatures (Feng et al., 2019; Xia et al., 2019). More recently, deep learning, and in particular convolutional neural networks (CNNs), has been applied directly to spectra, hyperspectral cubes and images, learning task-specific representations without explicit feature engineering (C. Jin et al., 2025; Ma et al., 2020). The resulting literature is now large enough, and its performance claims strong enough, to warrant careful critical appraisal rather than further descriptive cataloguing.

Recent evidence has extended this analytical progression across complementary sensing and modelling settings: hyperspectral deep-learning pipelines have been evaluated for maize and hemp seed vigour, while a large, naturally aged, cross-variety rice dataset has been used to examine viability classification and transferability; image-based deep learning has also been applied to germination prediction from RGB-derived morphological traits (Onwimol et al., 2025; Tosun & Eryiğit, 2026; Wongchaisuwat et al., 2025; Xu et al., 2025).

That appraisal is necessary because the headline results of this field are frequently reported as near-perfect discrimination, yet the conditions under which they were obtained are often narrow. Many studies use a single cultivar, a single harvest, one imaging system and one laboratory, and infer physiological status from artificial ageing rather than from natural deterioration or field performance (B. Jin et al., 2022; Ma et al., 2020). Class labels are sometimes derived from the very germination test the model is intended to replace, and validation is commonly internal, using cross-validation or a random split rather than an independent lot, season or site. The distance between a high internal accuracy and a dependable, transferable tool for seed technologists is consequently larger than many individual papers acknowledge. A review that takes evidence quality seriously must weigh not only what accuracy was reported but how, on what, and against what reference.

A further complication arises from the interdisciplinary character of the field, which draws together seed physiologists, agricultural engineers and machine-learning researchers whose evaluative standards differ. Seed science judges a method by its agreement with biologically meaningful reference tests and its behaviour across lots, seasons and environments, whereas machine learning tends to judge a model by its accuracy on a held-out partition of a fixed dataset. When these standards are applied to the same study they can yield sharply different verdicts, and the literature frequently reports the second while implying the first. A model that achieves high accuracy on a random split of images from a single seed lot has demonstrated internal separability, not the lot-to-lot and cultivar-to-cultivar transfer that a seed technologist would require before trusting it. Recognising this divergence is essential to a fair appraisal, because much of the apparent disagreement about how close the field stands to practical deployment stems from reviewers implicitly adopting one standard or the other rather than from genuine conflict in the underlying results.

Existing reviews have mapped parts of this territory. Broad surveys of non-destructive seed quality assessment describe the physical principles of the available sensing techniques and their reported applications (Feng et al., 2019; Rahman & Cho, 2016), and a focused review of emerging methods for seed viability catalogues spectroscopic and imaging approaches (Xia et al., 2019). A recent review of deep learning for high-throughput seed phenotyping documents the migration from engineered features to learned representations across seed traits

(C. Jin et al., 2025). These contributions are valuable, but they are organised largely around techniques and applications rather than around the strength and transferability of the evidence, and the fast-moving deep learning and data-centric strands, including transfer learning, class-imbalance handling and generative augmentation, have developed substantially since the earlier surveys. A critical, integrative account that foregrounds methodological quality, the reliability of reference labelling, and the gap between laboratory discrimination and field relevance is currently missing.

1.1 Scope, Research Gap and Objectives

This review examines the application of artificial intelligence, understood as classical machine learning and deep learning applied to spectral and image data, to the assessment of seed physiological quality, defined here as viability, germination capacity and vigour. It concentrates on non-destructive sensing modalities coupled to learning-based analysis, namely NIR spectroscopy, hyperspectral and multispectral imaging, red-green-blue and X-ray imaging, and automated analysis of germination and seedling development. Purity and varietal identification are considered only where the same pipelines and evidentiary problems bear directly on physiological-quality inference. The review spans agricultural, horticultural and tree species and integrates work from seed science, agricultural engineering and machine learning.

The central gap addressed is evaluative rather than descriptive. Despite a large and technically sophisticated literature reporting high accuracy, three questions remain insufficiently resolved: whether learning-based methods deliver genuine and reproducible gains over established seed tests and simpler chemometric baselines; whether reported performance generalises beyond the single cultivar, laboratory and ageing regime in which it was obtained; and whether the reference labels against which models are trained and judged are biologically adequate proxies for field-relevant physiological quality. These questions cut across sensing modality and crop, and they determine whether the field is approaching operational maturity or accumulating proofs of concept.

The objectives are, accordingly, to define the current state of knowledge on artificial intelligence for seed physiological quality and its major methodological schools; to compare convergent and conflicting findings across sensing modalities, learning paradigms and crops; to appraise the methodological strength of the evidence, with particular attention to reference labelling, sample representativeness, validation design and interpretability; to distinguish well-supported conclusions from preliminary or speculative claims; and to set out prioritised, defensible research directions for the development of dependable digital seed technology. The review does not attempt exhaustive coverage of every crop or sensor, does not address transgenic or molecular quality markers, and does not evaluate commercial sorting hardware except where peer-reviewed performance evidence is available.

2. Methods for Literature Selection

2.1 Review Design and Rationale

A critical narrative review was selected in preference to a systematic review or meta-analysis because the field is methodologically heterogeneous in a way that defeats quantitative pooling. Studies differ in crop, sensing modality, ageing regime, reference standard, class definition, validation scheme and performance metric, so a single effect size cannot be meaningfully synthesised across them, and a formal risk-of-bias meta-analysis would impose a false commensurability on incomparable designs. A narrative synthesis is better suited to interrogating conceptual and methodological issues, such as the adequacy of reference labels and the meaning of reported accuracy, that a purely quantitative approach would obscure (Grant & Booth, 2009; Sukhera, 2022). This choice does not license reduced rigour. The review followed a transparent and reproducible search, applied explicit eligibility criteria, and appraised the evidence critically rather than accepting reported performance at face value, consistent with published guidance on the conduct and quality of narrative reviews (Baethge et al., 2019; Ferrari, 2015).

The appraisal was structured around the criteria captured by instruments for narrative-review quality, namely justification of importance, a clear statement of aims, adequate description of the literature search, referencing of claims to appropriate evidence, and a scientifically reasoned rather than merely enumerative treatment of the material (Baethge et al., 2019). Emphasis throughout was placed on evidence-claim alignment, on representing disagreement where it exists, and on separating mechanistic plausibility from demonstrated field relevance.

2.2 Information Sources

The literature was identified through structured searching of open, programmatically accessible scholarly indexes, principally OpenAlex and the Crossref metadata registry, supplemented by targeted retrieval of records through publisher and journal landing pages and Digital Object Identifier (DOI) resolution. Subscription databases such as Web of Science, Scopus and Embase were not directly interrogated, and no claim of a completed search of those resources is made. Searching of the open indexes was complemented by backward and forward citation tracing from recent reviews and key primary studies, by related-record inspection, and by checking retrieved articles for corrections or expressions of concern. Every source retained was verified through Crossref for the exact title, author list, year, journal, volume, issue and page or article number, and each DOI was confirmed to resolve to the cited record. Records whose bibliographic identity or relevance could not be confirmed were discarded.

2.3 Search Strategy and Search Period

Search concepts were constructed from three intersecting domains and combined with Boolean logic. The first domain captured the seed-quality endpoint, using terms such as seed, seedling, viability, vigour, vitality and germination. The second captured the sensing modality, using hyperspectral imaging, multispectral imaging, near-infrared spectroscopy, X-ray, radiography, machine vision, computer vision and red-green-blue imaging. The third captured the analytical approach, using machine learning, deep learning, convolutional neural network, chemometrics, transfer learning and classification. A representative query combined these as ("seed" OR "seedling") AND ("viability" OR "vigour" OR "germination") AND ("hyperspectral" OR "near-infrared" OR "X-ray" OR "machine vision") AND ("machine learning" OR "deep learning" OR "convolutional neural network"), with modality- and crop-specific variants executed separately so that each concept was represented by its own retrieval rather than a single generic string. The search period extended from 2014 to 30 May 2026. Seminal methodological and conceptual works published before 2014 were consulted where necessary to establish the historical development of seed vigour testing and review methodology.

2.4 Eligibility Criteria

Eligible publications were peer-reviewed journal articles that applied a machine learning or deep learning method to spectral or image data for the purpose of assessing seed viability, vigour, germination or a directly related physiological-quality endpoint. Studies were required to report a defined reference standard for physiological quality, a described analytical pipeline and an evaluation of predictive or classification performance. Both classical machine learning and deep learning were eligible, as were studies comparing the two. Work on varietal identification, purity or defect detection was eligible only where it bore directly on physiological-quality inference or illustrated a shared methodological problem. Excluded were conference abstracts, preprints where peer-reviewed evidence existed, editorials, patents used as evidence of effectiveness, non-scholarly material, and studies concerned solely with disease, composition or agronomic traits unrelated to physiological quality. Reviews and foundational seed-physiology and methodological papers were retained as contextual rather than primary evidence. No restriction was placed on crop or geography; publications in English were included.

2.5 Screening, Duplicate Handling and Study Selection

Retrieved records were consolidated and de-duplicated on the DOI, so that a study indexed under more than one query was counted once. Titles and abstracts were screened against the eligibility criteria to remove clearly irrelevant records, principally studies on leaf disease, yield prediction, remote sensing of canopies and root phenotyping that were returned by broad queries but did not address seed physiological quality. Records passing this stage were assessed on the fuller information available through the abstract and, where required for verification, the publisher record. Sources that could not be bibliographically verified, or whose full identity could not be confirmed, were not retained. Selection favoured methodological and conceptual contribution, reference-standard adequacy and relevance to the review questions over citation count alone, so that recent corrective studies were weighted appropriately against older, more heavily cited work. No fabricated screening counts are reported, and because the process was a critical narrative selection rather than a systematic enumeration, no flow diagram with record tallies is presented.

2.6 Critical Appraisal and Evidence Synthesis

Each retained primary study was appraised against criteria appropriate to this field: the biological adequacy and independence of the reference standard; the ageing or deterioration regime used to generate quality gradients; sample size, cultivar coverage and class balance; the validation design, distinguishing internal cross-validation from independent-lot, season or site testing; the appropriateness of the performance metric to the class distribution; and the transparency and reproducibility of the pipeline. Deep learning studies were further considered in terms of overfitting risk given sample size, the handling of spectral collinearity and redundancy, and the presence or absence of interpretability analysis. No formal risk-of-bias score was assigned, because the information required to apply a validated tool consistently was not uniformly available. Evidence was then organised thematically around the measurement problem, sensing modalities, learning paradigms and translation to practice, so that convergent and conflicting findings could be compared directly and the confidence attaching to each conclusion made explicit.

3. Seed Physiological Quality and the Measurement Problem

Any evaluation of artificial intelligence for seed quality must begin with what the models are being asked to predict, because the target defines both the ceiling on achievable accuracy and the meaning of the numbers reported. Viability, germination and vigour are related but distinct constructs. Viability denotes the presence of living, potentially germinable tissue; germination capacity denotes the proportion of a lot that produces normal seedlings under favourable conditions; and vigour denotes the broader physiological potential that governs performance under the suboptimal conditions of the field (Marcos Filho, 2015; Reed et al., 2022). Deterioration erodes these properties progressively and in a partly ordered sequence, with vigour typically declining before germinability and germinability before viability, so that a lot may retain high germination in a laboratory test while having lost the vigour needed for reliable field emergence (Marcos Filho, 2015; Nadarajan et al., 2023). This ordering matters for machine learning because a binary viable-or-non-viable label collapses a continuum, and a model trained to separate the extremes of that continuum will appear more accurate than one required to resolve the agronomically important middle.

The biological basis of these constructs clarifies why optical and image sensors can detect them at all. Seed deterioration is driven substantially by oxidative processes, including the peroxidation of membrane lipids and the accumulation of reactive oxygen species, together with a declining capacity for cellular repair during imbibition, and these changes alter both the chemical composition and the structural integrity of seed tissues (Nadarajan et al., 2023; Reed et al., 2022). Antioxidant defences such as superoxide dismutase buffer this damage, and their activity has been used as a biochemical correlate of vigour that lends physiological meaning to spectral predictions, as in hyperspectral work on peanut where the model target was related to enzyme activity rather than to germination alone (Zou et al., 2023). Because deterioration is a progressive biochemical trajectory rather than a switch, the information available to a sensor is graded, and the resolution a model can achieve is bounded by how strongly the sensed correlates track the specific endpoint of interest. This bounding is easily overlooked when binary viability labels are used, since the underlying biology remains continuous even where the label is not.

The reference standard used to label training data is therefore the single most consequential design choice in this literature, and it is frequently the weakest. Where labels derive from the standard germination test, the model inherits that test's optimal-condition ceiling and cannot, in principle, capture vigour information the reference does not contain. Where labels derive from tetrazolium staining, they capture viability but not vigour gradients. Where quality gradients are generated by artificial ageing, through elevated temperature and humidity over hours or days, the resulting deterioration may not faithfully reproduce the biochemistry of natural storage ageing, and models calibrated on artificially aged material may not transfer to naturally deteriorated lots (B. Jin et al., 2022; Ma et al., 2020). A minority of studies have compared naturally aged material or related spectral predictions to biochemical vigour indicators such as antioxidant enzyme activity, strengthening the biological interpretation of the signal (B. Jin et al., 2022; Zou et al., 2023). The distinction between predicting a proxy label and predicting field-relevant physiological quality is fundamental, and it recurs throughout the sections that follow.

Moisture merits particular attention because it is at once the strongest near-infrared signal and a potent confounder. Water content dominates the NIR response through prominent absorption features, and it also influences seed metabolism, ageing rate and mechanical properties, so a model may learn to predict a quality label through its incidental correlation with moisture rather than through deterioration itself (Feng et al., 2019; Xia et al., 2019). Where moisture is neither controlled nor explicitly modelled, apparent viability or vigour discrimination may partly reflect differences in equilibration or storage humidity between classes, an artefact that internal validation cannot expose and that would collapse on lots equilibrated to a common moisture content. Comparable caution applies to temperature and surface contamination. Few studies report the moisture regime of their samples in enough detail to exclude this confound, which is one reason that biologically grounded interpretation, linking selected wavelengths to known deterioration chemistry rather than to bulk water, is valuable beyond its contribution to transparency.

The measurement problem also has a physical dimension. Optical methods sense the chemical and structural correlates of physiological state, principally moisture, lipid oxidation products, protein and pigment changes and internal morphology, rather than viability itself, and the strength of the relationship between these correlates and physiological quality varies with species, tissue and deterioration mechanism (Feng et al., 2019; Xia et al., 2019). The discussion below distinguishes the core constructs and their measurement implications, which are summarised in Table 1; the table is intended to make explicit the target definitions that individual studies frequently leave implicit, and thereby to frame the appraisal of reported accuracy that follows.

Table 1. Core physiological-quality constructs, their reference standards and implications for learning-based assessment

Construct	Operational definition	Common reference standard	Implication for AI-based assessment	Supporting references
Viability	Presence of living tissue capable of germinating under favourable conditions	Tetrazolium staining; germination	Captures a coarse living/dead distinction; a binary label collapses the deterioration continuum and inflates apparent accuracy	Marcos Filho, 2015; Rahman & Cho, 2016
Germination capacity	Proportion of a lot producing normal seedlings under optimal test conditions	Standard germination test	Provides a defensible but optimal-condition ceiling; models inherit its insensitivity to field-relevant vigour differences	Marcos Filho, 2015; Reed et al., 2022
Vigour	Physiological potential governing rapid, uniform emergence across field environments	Accelerated ageing; electrical conductivity; controlled deterioration; radicle emergence	Agronomically most valuable yet least standardised; heterogeneous references undermine cross-study comparability	Marcos Filho, 2015; Reed et al., 2022; Shinohara et al., 2021
Longevity/deterioration state	Position along the ageing trajectory affecting storability	Storage records; artificial ageing	Artificial ageing is a convenient but imperfect proxy for natural deterioration, limiting transfer to stored lots	B. Jin et al., 2022; Nadarajan et al., 2023

Note. AI, artificial intelligence. Supporting references correspond to full entries in the reference list.

4. Spectral Sensing of Seed Viability and Vigour

Spectroscopic and hyperspectral approaches dominate the learning-based seed-quality literature, and for coherent physical reasons. NIR and short-wave infrared radiation interact with overtones and combinations of molecular bonds associated with water, lipids, proteins and carbohydrates, whose concentrations and oxidation states shift as seeds deteriorate, so a spectrum encodes a chemical fingerprint plausibly linked to physiological state (Feng et al., 2019; Xia et al., 2019). Hyperspectral imaging (HSI) extends single-point spectroscopy by acquiring a full spectrum at every pixel, enabling single-seed analysis and the spatial localisation of quality-relevant variation, which is decisive for high-throughput sorting where lots are heterogeneous (Feng et al., 2019). The evidence base spans point NIR spectroscopy of single seeds and bulk samples through to visible-NIR and short-wave infrared HSI, analysed with chemometric classifiers and, increasingly, deep networks.

The information relevant to physiological quality is not distributed uniformly across the spectrum, and both the choice of spectral region and the preprocessing applied to raw reflectance strongly shape results. Visible and near-infrared regions capture pigment, moisture and some compositional variation, whereas short-wave infrared regions carry richer information on the lipids and proteins whose oxidation accompanies ageing, and the most informative region differs among species and deterioration mechanisms (Feng et al., 2019; Xia et al., 2019). Raw spectra are routinely subjected to scatter correction, derivative transformation and normalisation before analysis, and these steps materially affect downstream accuracy, yet they are reported inconsistently, which complicates comparison across studies. The migration toward deep learning has partly absorbed preprocessing into the model, since convolutional layers can learn to suppress baseline and scatter effects, but this convenience is bought at the cost of transparency about which spectral features drive predictions. Wavelength-selection algorithms counteract that opacity by identifying compact, interpretable subsets of bands, and their frequent success in matching full-spectrum models indicates that much of the quality-relevant signal is concentrated in a limited number of wavelengths (Fan et al., 2023; L. Pang et al., 2022).

For viability, the consistent finding is that spectral data discriminate viable from non-viable seeds with high reported accuracy across a wide range of species. Single-seed NIR spectroscopy separated viable and non-viable tomato (*Solanum lycopersicum*) seeds (Shrestha et al., 2017), and Fourier-transform NIR spectroscopy determined soybean (*Glycine max*) seed viability non-destructively (Kusumaningrum et al., 2018). NIR-HSI coupled with deep learning predicted viability rapidly and non-destructively across multiple species (Ma et al., 2020), and subsequent work extended the approach to sweet corn (*Zea mays*) using HSI with metaheuristically optimised deep learning (Wang & Song, 2024), to single-seed viability in maize using a multi-scale three-dimensional CNN with wavelength selection (Fan et al., 2023), and to *Sophora japonica* using both an optimised deep neural network (L. Pang et al., 2021) and an improved successive projections algorithm (L. Pang et al., 2022). Chemometric NIR with variable selection resolved spinach (*Spinacia oleracea*) seed viability (Lakshmanan et al., 2023), and visible-NIR HSI with machine learning detected watermelon (*Citrullus lanatus*) seed viability (Sun et al., 2024). Fusion of hyperspectral and X-ray information improved viability prediction in pepper (*Capsicum annuum*) beyond either modality alone (Hong et al., 2023), illustrating the complementary value of chemical and structural signals.

A practically decisive distinction separates bulk and single-seed analysis. Bulk spectroscopy characterises a sample average and is unsuited to sorting, whereas single-seed spectroscopy and hyperspectral imaging resolve individual seeds and therefore support the removal of non-viable or low-vigour units from a lot (Feng et al., 2019). Single-seed measurement is more demanding, because the signal from one small seed is weaker and more variable, and because throughput becomes a binding constraint when every seed must be imaged and classified in a processing line. Several studies have consequently emphasised speed alongside accuracy, combining wavelength selection to reduce data volume with architectures chosen for inference efficiency (Fan et al., 2023; L. Pang et al., 2020). The tension between the spatial and spectral richness that improves accuracy and the throughput that operational sorting requires is rarely resolved within a single study, and it represents one of the clearest gaps between laboratory demonstration and deployable technology. Reported per-seed accuracy in a laboratory says little about sustained performance at commercial line speeds.

For vigour, which is agronomically more valuable and physiologically more subtle, the evidence is larger in volume but weaker in reference quality. HSI has been applied to rice (*Oryza sativa*) vigour using transfer learning across cultivars (Qi et al., 2023) and simple transfer of pretrained features (Yang et al., 2020), to peanut (*Arachis hypogaea*) vigour with hyperspectral data related to superoxide dismutase activity (Zou et al., 2023), and to maize seed vitality using combined spectral and image information with CNNs benchmarked against support vector machines and extreme learning machines (L. Pang et al., 2020). A recurrent methodological problem is class imbalance, because vigour categories defined by germination outcome are naturally unequal; an end-to-end HSI model was developed specifically to address imbalanced samples and the feature redundancy and collinearity inherent in hundreds of contiguous wavebands (T. Pang et al., 2023). Studies that jointly estimate viability and vigour on naturally aged rather than artificially aged material are comparatively rare but especially informative, since they test the biological generality of the spectral signal (B. Jin et al., 2022). Point NIR spectroscopy has likewise been used to classify soybean seed vigour levels (M. F. D. Silva et al., 2024), extending the approach beyond imaging systems.

Two analytical debates run through this evidence. The first concerns whether deep learning genuinely outperforms chemometrics. Several head-to-head comparisons report an advantage for CNNs over partial least squares and support vector machine baselines (Fan et al., 2023; L. Pang et al., 2020), and generative and transfer-learning strategies have been introduced to compensate for the small datasets that otherwise favour simpler models; an improved generative adversarial network was used to augment NIR-HSI data for rice viability determination, mitigating data scarcity (Qi et al., 2024). Yet the advantage is not universal, is often modest, and is confounded by unequal tuning effort and by the small test sets on which differences are measured. The second debate concerns dimensionality. Because adjacent hyperspectral bands are highly correlated, wavelength selection through successive projections or comparable algorithms frequently matches full-spectrum deep models while improving speed and interpretability (Fan et al., 2023; L. Pang et al., 2022), which questions the necessity of end-to-end deep architectures for many tasks. The representative spectral evidence is summarised in Table 2, which juxtaposes species, endpoint, modality, analytical approach and the reference standard, exposing the concentration of the field on artificial ageing and internal validation.

Table 2. Representative studies applying spectral and hyperspectral sensing with learning-based analysis to seed viability and vigour

Crop (species)	Endpoint	Modality	Analytical approach	Reference / quality gradient	Supporting references
Multiple	Viability	NIR-HSI	Deep learning (CNN)	Germination; artificial ageing	Ma et al., 2020
Sweet corn (*Zea mays*)	Viability	HSI	Metaheuristic-optimised deep learning	Germination	Wang & Song, 2024
Maize (*Zea mays*)	Viability (single seed)	HSI	Multi-scale 3D CNN with wavelength selection	Aged vs non-aged; germination	Fan et al., 2023
Rice (*Oryza sativa*)	Viability	NIR-HSI	GAN-based augmentation with deep classifier	Germination	Qi et al., 2024
Sophora japonica	Viability	HSI	Optimised deep neural network; SPA	Germination	L. Pang et al., 2021; L. Pang et al., 2022
Tomato (*Solanum lycopersicum*)	Viability	Single-seed NIR	Chemometric classification	Germination	Shrestha et al., 2017
Soybean (*Glycine max*)	Viability	FT-NIR	Chemometric classification	Germination	Kusumaningrum et al., 2018
Pepper (*Capsicum annuum*)	Viability	HSI + X-ray fusion	Machine learning on fused features	Germination	Hong et al., 2023
Spinach (*Spinacia oleracea*)	Viability	NIR	Chemometrics with SPA variable selection	Germination	Lakshmanan et al., 2023

Crop (species)	Endpoint	Modality	Analytical approach	Reference / quality gradient	Supporting references
Watermelon (*Citrullus lanatus*)	Viability	Vis-NIR HSI	Machine learning	Germination	Sun et al., 2024
Rice (*Oryza sativa*)	Vigour	NIR-HSI	Deep transfer learning across cultivars	Artificial ageing; germination	Qi et al., 2023; Yang et al., 2020
Peanut (*Arachis hypogaea*)	Vigour	HSI	Chemometrics; enzyme-linked interpretation	Germination; SOD activity	Zou et al., 2023
Maize (*Zea mays*)	Vitality	HSI (spectra + images)	CNN vs SVM/ELM	Artificial ageing; germination	L. Pang et al., 2020
Seed (general)	Vigour	HSI	End-to-end model for imbalanced data	Germination-derived classes	T. Pang et al., 2023
Rice (*Oryza sativa*)	Viability and vigour	HSI	Machine learning	Natural ageing; germination	B. Jin et al., 2022
Soybean (*Glycine max*)	Vigour	NIR	Chemometric classification	Vigour test levels	M. F. D. Silva et al., 2024

Note. CNN, convolutional neural network; ELM, extreme learning machine; FT-NIR, Fourier-transform near-infrared; GAN, generative adversarial network; HSI, hyperspectral imaging; NIR, near-infrared; SOD, superoxide dismutase; SPA, successive projections algorithm; SVM, support vector machine; Vis-NIR, visible near-infrared; 3D, three-dimensional.

The strongest conclusion the spectral literature supports is that physiological deterioration produces optical signatures detectable with high internal accuracy across many species and both viability and vigour endpoints. Confidence in this conclusion is high for viability under controlled conditions and moderate for vigour, where reference heterogeneity and artificial ageing limit interpretation. The weakest link is external validity: few studies test a model trained on one cultivar, harvest or instrument against an independent lot, and fewer still relate spectral predictions to field emergence rather than to a laboratory germination or vigour test. The evidence is therefore strong for discrimination and weak for demonstrated operational transfer, a distinction that reported accuracy figures obscure.

Two structural features of this evidence base further temper its interpretation. The first is taxonomic and geographic concentration: rice, maize, soybean and a small set of vegetable and tree species recur repeatedly, and much of the work originates from a limited number of research groups and regions, so the apparent breadth of the literature overstates the diversity of the material on which methods have actually been tested. Findings that hold for the storage biochemistry and seed structure of one species need not transfer to another with different reserve composition or seed-coat properties, and this heterogeneity is seldom examined directly. The second feature is the near-universal reliance on artificial ageing to generate quality gradients. Accelerated deterioration is experimentally convenient and produces clear class separation, but it compresses into hours or days a process that unfolds over months or years and may engage different biochemical pathways, so a model calibrated on artificially aged seed cannot be assumed to read naturally deteriorated lots correctly. The comparatively few studies that use naturally aged material are therefore disproportionately valuable, and their relative scarcity is a defining limitation of the current spectral evidence (B. Jin et al., 2022).

5. Image-based and Structural Characterisation of Seed Quality

Alongside spectral sensing, a parallel body of work analyses the appearance, internal structure and dynamic behaviour of seeds and seedlings using conventional imaging. These approaches divide into three broad families: colour and multispectral imaging of external morphology, X-ray radiography of internal structure, and time-resolved imaging of germination and seedling growth. They differ from spectral methods in sensing morphology and development rather than chemistry, and they are often cheaper and more directly interpretable, at the cost of sensitivity to purely biochemical deterioration.

Colour and machine-vision analysis of seed and seedling morphology has a long lineage in seed technology, where seedling growth measurements underpin established vigour tests. Automated seedling image analysis reproduced expert vigour assessments in tomato (V. N. Silva & Cicero, 2014) and onion (*Allium cepa*) (Gonçalves et al., 2017), and dedicated systems such as Vigor-S automated the analysis of soybean (Rodrigues et al., 2020) and maize (Castan et al., 2018) seed vigour from seedling images, reducing analyst subjectivity. Machine vision has also been proposed as an integrated alternative for physical purity, viability and vigour testing of soybean (Mahajan et al., 2018). Deep learning extended these ideas: an interactive and traditional machine-learning approach classified soybean seeds and seedlings by appearance and physiological potential and related seed appearance to physiological performance (de Medeiros et al., 2020a), and CNN-based systems have been developed for generic seed classification (Gulzar et al., 2020) and for varietal classification with concurrent quality and maturity assessment in winter oilseed rape (*Brassica napus*) (Rybacki et al., 2023) and maize (Javanmardi et al., 2021). Multispectral imaging combined with FT-NIR spectroscopy in a merged data stream improved seed quality classification for a forage grass, demonstrating the value of sensor fusion for germination and vigour prediction (de Medeiros et al., 2020b).

The principal limitation of morphology-based assessment is that external appearance is an imperfect guide to physiological state. Two seeds of identical size, colour and shape can differ substantially in vigour, and conversely superficial blemishes may not compromise germination, so models that classify on appearance risk learning associations that hold within one lot but not across lots or cultivars (de Medeiros et al., 2020a). Studies that explicitly relate seed appearance to measured physiological performance, rather than assuming the two coincide, are therefore more informative than those that treat visual class as a quality proxy (de Medeiros et al., 2020a). X-ray radiography partly escapes this limitation by sensing internal structure that is causally connected to germination outcome, including embryo integrity and the presence of internal cavities or mechanical damage, which explains its value as a complement to both surface imaging and spectroscopy (Ahmed et al., 2020; de Medeiros et al., 2020c; Hamdy et al., 2024). Even so, the mapping from a structural defect to a physiological consequence is probabilistic rather than deterministic, and the annotation of defects introduces its own subjectivity.

X-ray radiography addresses a limitation of external imaging by revealing internal morphology, including embryo and endosperm integrity, mechanical damage and air spaces that are invisible from the surface yet strongly associated with germination outcome. Machine learning applied to radiographic images classified the quality of *Jatropha curcas* seeds (de Medeiros et al., 2020c), and conventional and deep learning approaches were compared for classifying watermelon seeds from morphological X-ray patterns (Ahmed et al., 2020). More recent work has moved toward robust, high-throughput deep-learning detection of internal seed defects in two-dimensional X-ray radiographs, explicitly targeting the transferability and speed that operational deployment requires (Hamdy et al., 2024). Structural imaging thereby complements chemical sensing, and the pepper study noted earlier, which fused hyperspectral and X-ray data, indicates that combining the two can outperform either alone (Hong et al., 2023).

The third family images the process rather than the static seed. Automated germination phenotyping platforms capture the temporal dynamics of radicle emergence and seedling development, from which vigour-relevant parameters are derived. The SeedGerm system combined low-cost hardware with machine learning to score germination across tomato, pepper, *Brassica*, barley and maize (Colmer et al., 2020), and accurate machine-learning detection with region-proposal networks was demonstrated for germination in three grain crops (Genze et al., 2020). Generic germination prediction from red-green-blue images using deep learning has been proposed as a means of reducing the wasteful disqualification of seed lots that nonetheless contain many viable seeds (Nehoshtan et al., 2021), and computer-vision analysis of multispectral images has been developed for non-destructive rice seed vigour testing (Qiao et al., 2023). Deep learning has been applied to detect germination in wild rice via a dedicated object-detection model (Yao et al., 2024), to build automatic germination-test monitoring systems (Peng et al., 2022), and, using a k-nearest-neighbours approach, to analyse diverse germination phenotypes and detect single-seed germination in *Miscanthus sinensis* under non-sterile conditions (Awty-Carroll et al., 2018). The agronomic value of early dynamic measurements is supported by evidence that automated radicle-emergence counts can reveal seed-lot vigour differences earlier than the completed germination test (Shinohara et al., 2021), and integrated optical imaging combining chlorophyll fluorescence and multispectral data has been used to characterise tomato and carrot (*Daucus carota*) seed quality non-invasively (Galletti et al., 2020). Digital seedling and seed image analysis has been reviewed as a high-throughput phenotyping strategy in its own right (Zhang et al., 2018).

Table 3. Representative studies applying image-based and structural sensing with learning-based analysis to seed physiological quality

Modality family	Endpoint	Crop (species)	Analytical approach	Reference / gradient	Supporting references
RGB seedling imaging	Vigour	Tomato; onion (*Allium cepa*)	Image analysis vs standard tests	Seedling growth; vigour tests	Gonçalves et al., 2017; V. N. Silva & Cicero, 2014
RGB seedling imaging	Vigour	Soybean (*Glycine max*); maize (*Zea mays*)	Automated systems (Vigor-S)	Seedling image parameters	Castan et al., 2018; Rodrigues et al., 2020
Machine vision	Purity, viability, vigour	Soybean (*Glycine max*)	Feature-based machine learning	Standard tests	Mahajan et al., 2018
RGB imaging	Physiological potential	Soybean (*Glycine max*)	Interactive and traditional machine learning	Appearance vs performance	de Medeiros et al., 2020a
RGB/colour imaging	Classification and quality	Multiple; rapeseed (*Brassica napus*); maize	CNN	Variety; maturity/damage	Gulzar et al., 2020; Javanmardi et al., 2021; Rybacki et al., 2023
Multispectral + FT-NIR	Germination and vigour	Forage grass	Machine learning on merged data	Germination; vigour	de Medeiros et al., 2020b
X-ray radiography	Quality classification	*Jatropha curcas*; watermelon (*Citrullus lanatus*)	Machine learning; deep learning	Internal morphology; germination	Ahmed et al., 2020; de Medeiros et al., 2020c
X-ray radiography	Internal defect detection	Multiple	Deep learning (robust pipeline)	Annotated defects	Hamdy et al., 2024
Germination imaging	Germination dynamics	Tomato; pepper; *Brassica*; barley; maize	Machine learning platform	Radicle emergence; germination	Colmer et al., 2020; Genze et al., 2020
Germination imaging	Germination detection/rate	Rice (*Oryza sativa*); *Miscanthus sinensis*	Deep learning; k-NN	Germination scoring	Awty-Carroll et al., 2018; Peng et al., 2022; Yao et al., 2024
Early dynamic imaging	Vigour (early prediction)	Cauliflower (*Brassica oleracea*)	Automated radicle-emergence counting	Radicle emergence vs germination	Shinohara et al., 2021
Fluorescence + multispectral	Seed quality	Tomato; carrot (*Daucus carota*)	Chemometric imaging	Germination	Galletti et al., 2020

Note. CNN, convolutional neural network; FT-NIR, Fourier-transform near-infrared; k-NN, k-nearest neighbours; RGB, red-green-blue

Annotation is an underappreciated source of error across the image-based literature. Germination scoring, seedling normality classification and defect labelling all rest on human judgement that the reference tests themselves acknowledge to be variable, and when such labels train a model, the model can at best reproduce the consistency of its annotators (Colmer et al., 2020; Genze et al., 2020). Automated platforms mitigate this by applying a fixed algorithmic criterion uniformly, which improves repeatability even where it does not improve biological validity (Colmer et al., 2020). The temporal dimension of germination imaging offers a distinct and still underexploited advantage: by capturing the trajectory of radicle emergence and seedling growth rather than a single endpoint, dynamic imaging brings the measured quantity closer to the developmental process that vigour is meant to predict, and early time points can discriminate vigour before the completed germination test resolves it (Qiao et al., 2023; Shinohara et al., 2021). Modelling these trajectories with sequence-aware methods, rather than reducing them to summary counts, remains uncommon and represents a clear opportunity. The representative image-based evidence is compiled in Table 3.

Image-based methods contribute two things the spectral literature lacks. Dynamic germination imaging measures the developmental process that vigour is meant to predict, bringing the model target closer to the agronomic outcome than a static optical fingerprint, and early radicle-emergence measurement offers a route to faster vigour discrimination grounded in an established biological indicator (Colmer et al., 2020; Shinohara et al., 2021). X-ray radiography adds an interpretable structural basis for classification that aligns with the physical causes of poor germination (Ahmed et al., 2020; Hamdy et al., 2024). The limitations mirror those of the spectral work: datasets are typically single-species and single-laboratory, annotation of germination and defect states can itself be subjective, and external validation across seasons and sources is rare. The overall pattern across sensing families is one of convergent technical success under controlled conditions and shared, unresolved weaknesses in representativeness, labelling and transfer.

Comparing the two sensing families directly clarifies where each is likely to prove most useful. Spectral methods sense chemistry and can, in principle, detect deterioration that leaves no visible or structural trace, which suits early or biochemical vigour loss, but they demand costlier instrumentation and careful calibration and are the more vulnerable to moisture and instrument confounding. Image-based methods sense morphology and development, are cheaper and more interpretable, and, in their dynamic form, measure the germination process itself, though they are insensitive to purely chemical change and depend on annotation quality. The fusion studies that combine the two, and the systems that pair spectral prediction with structural X-ray information, suggest that the modalities are complementary rather than competing, and that the most informative signal for a given species and endpoint may draw on both (de Medeiros et al., 2020b; Hong et al., 2023). No single modality has been shown to dominate across crops and endpoints, and the choice is best treated as conditional on the deterioration mechanism, the required throughput and the available reference standard rather than settled in the abstract. This conditional, evidence-led framing is more defensible than the implicit assumption, common in individual studies, that the authors' chosen modality is generally superior.

6. Learning Paradigms, Data Constraints and Generalisation

Beneath the diversity of crops and sensors lies a smaller set of analytical decisions that determine whether reported performance is trustworthy. The first is the choice between engineered-feature machine learning and end-to-end deep learning. Classical pipelines extract spectral or morphological descriptors, often after wavelength selection, and pass them to established classifiers; deep networks learn features directly from spectra, cubes or images (C. Jin et al., 2025; Xia et al., 2019). Deep learning offers clear advantages where the relevant patterns are complex and spatial, as in single-seed HSI and X-ray defect detection (Fan et al., 2023; Hamdy et al., 2024), and comparative studies frequently report gains over chemometric baselines (Ma et al., 2020; L. Pang et al., 2020). The advantage is real but conditional. It depends on sufficient training data, and much of this literature operates with a few hundred seeds per class, a regime in which flexible deep models overfit and their apparent superiority may not survive independent testing. Where adjacent bands are redundant, compact wavelength-selection models can match deep networks at lower cost and with greater transparency (Fan et al., 2023; L. Pang et al., 2022), which cautions against treating architectural novelty as intrinsic progress.

The second decision concerns the field's defining constraint: small, imbalanced and narrowly sampled datasets. Because physiological quality is expensive to reference and seeds are destroyed in the process, datasets are modest, and vigour classes defined by germination outcome are inherently unequal. Three responses have emerged. Transfer learning reuses representations learned on larger corpora or related cultivars, and has improved rice vigour detection across varieties (Qi et al., 2023; Yang et al., 2020). Data augmentation and generative modelling synthesise additional training examples, with an improved generative adversarial network used to expand NIR-HSI data for rice viability (Qi et al., 2024). Model-level strategies target imbalance directly, as in an end-to-end architecture designed for imbalanced vigour data with integrated wavelength handling (T. Pang et al., 2023). These are appropriate and often effective mitigations, but they do not manufacture representativeness. A model augmented from a single narrow dataset remains narrow, and generative augmentation risks amplifying the idiosyncrasies of the source data rather than the diversity of the target population.

The third decision is validation design, and it is where the literature is most exposed. Performance is predominantly estimated by cross-validation or a single random split within one dataset, which measures internal separability but not transfer to new cultivars, harvests, instruments or environments. Studies that explicitly test across cultivars are informative precisely because they reveal the cost of transfer and the need for adaptation strategies (Qi et al., 2023). Independent validation against field emergence, as opposed to a laboratory reference, is almost absent, so the practically decisive question, whether a model predicts what happens in the ground, is largely unanswered. Related to this is metric choice: accuracy on an imbalanced test set can flatter a model that predicts the majority class, and studies vary in whether they report class-sensitive metrics. The fourth decision, interpretability, is unevenly addressed; wavelength-selection studies yield chemically meaningful variables (Fan et al., 2023; Zou et al., 2023), but many deep models remain opaque, which impedes both scientific understanding and regulatory acceptance in a certification context.

A specific and consequential validation error deserves emphasis because it is easy to commit and hard to detect in hyperspectral work. When a hyperspectral cube of a single seed yields many pixel spectra, splitting those pixels at random between training and test sets places spectra from the same physical seed on both sides of the partition, so the model is evaluated partly on seeds it has already encountered, and reported accuracy is inflated by information leakage rather than genuine generalisation (Fan et al., 2023; T. Pang et al., 2023). Rigorous evaluation requires that the split be made at the level of the seed, and preferably the lot, so that no seed contributes to both training and testing. The same principle applies to seedling and germination images captured repeatedly over time, where frames of one developing seed must not be divided across the partition. Studies vary in whether they guard against this, and because the error is invisible to internal metrics, its prevalence in the reported literature is difficult to gauge. Where it occurs, headline accuracy overstates achievable performance on new material and compounds the more familiar limitations of narrow sampling and internal validation.

The field's evidentiary weaknesses are compounded by a reproducibility and benchmarking deficit. Progress in adjacent areas of applied machine learning has been driven substantially by large, openly shared and well-characterised datasets and by common evaluation protocols that allow methods to be compared on equal terms. The seed-quality literature has almost none of this infrastructure: datasets are typically proprietary, small and unpublished, code is often unavailable, and each study reports accuracy on its own private data using its own class definitions and metrics (C. Jin et al., 2025). The many high accuracies in the literature therefore cannot be aggregated into a reliable picture of relative method performance, because they are not measured against a shared standard. A model reported as superior in one paper may owe its advantage to an easier dataset rather than a better method, and there is no way to adjudicate the claim without common benchmarks. This deficit is corrigible, and addressing it is arguably more important to the maturation of the field than any further architectural innovation.

Metric reporting warrants specific scrutiny in a domain where classes are intrinsically imbalanced. Overall accuracy on a test set dominated by viable or vigorous seeds can be high even for a model that performs poorly on the minority class that matters most for sorting, and studies vary in whether they report sensitivity, specificity, balanced accuracy or the area under the receiver operating characteristic curve alongside a headline figure (T. Pang et al., 2023). Without class-sensitive metrics a reported accuracy cannot be interpreted, and the imbalance-aware modelling introduced in some vigour studies is difficult to evaluate against work that reports only aggregate accuracy (T. Pang et al., 2023). Consistent, class-aware reporting is a low-cost improvement that would materially increase the interpretability of the existing evidence. These recurrent limitations, their consequences and the studies that illustrate or begin to address them are drawn together in Table 4.

Table 4. Recurrent methodological limitations across the artificial-intelligence seed-quality literature and their consequences

Limitation	How it manifests	Consequence for validity	Illustrative or mitigating studies	Supporting references
Artificial ageing as proxy	Quality gradients generated by heat and humidity rather than natural storage	Uncertain transfer to naturally deteriorated lots	Use of naturally aged material; enzyme-linked interpretation	B. Jin et al., 2022; Zou et al., 2023
Reference-label adequacy	Labels from optimal-condition germination or binary viability	Model inherits reference ceiling; vigour continuum collapsed	Early radicle-emergence and dynamic imaging targets	Colmer et al., 2020; Shinohara et al., 2021
Small, imbalanced datasets	Few hundred seeds per class; unequal vigour classes	Overfitting; inflated accuracy; unstable comparisons	Transfer learning; generative augmentation; imbalance-aware models	T. Pang et al., 2023; Qi et al., 2023; Qi et al., 2024
Narrow cultivar/environment coverage	Single cultivar, harvest, instrument and laboratory	Poor external validity; limited generalisation	Cross-cultivar transfer testing	Qi et al., 2023; Yang et al., 2020
Internal-only validation	Cross-validation or single random split	Transfer to new lots, seasons and sites unproven	Multi-cultivar designs (partial)	Fan et al., 2023; Qi et al., 2023
Deep versus simple models	End-to-end networks vs wavelength-selection baselines	Architectural complexity not always justified	Compact SPA models matching deep networks	Fan et al., 2023; L. Pang et al., 2022
Limited interpretability	Opaque deep models without feature attribution	Impedes scientific insight and certification acceptance	Wavelength-selection and enzyme-linked studies	Fan et al., 2023; Zou et al., 2023

Note. SPA, successive projections algorithm. Illustrative studies either exemplify the limitation or begin to mitigate it; they are not exhaustive.

Taken together, these considerations reframe the field's headline accuracy. The available evidence indicates that learning-based methods reliably detect physiological deterioration under controlled conditions, and that deep learning can add value where spatial or structural complexity is high. Confidence that these gains transfer to operational settings is low, because the validation designs that would establish transfer are seldom used. The apparent discrepancy between near-perfect internal accuracy and cautious external claims does not reflect flawed modelling so much as an evidentiary structure that optimises for internal discrimination. Progress therefore depends less on new architectures than on the data, references and validation regimes against which any architecture is judged.

7. Toward Digital Seed Technology: Integration, Validation and Deployment

Digital seed technology envisages non-destructive, learning-based measurement embedded in seed processing, so that lots are characterised, sorted and released with objective, traceable quality information rather than by delayed laboratory testing alone (Feng et al., 2019; C. Jin et al., 2025). The reviewed evidence establishes technical feasibility for the sensing and analytical components, but deployment raises requirements that laboratory studies rarely address. Throughput and single-seed handling are prerequisites for sorting, and HSI and radiographic pipelines are being adapted toward the speed and robustness that operational lines demand (Fan et al., 2023; Hamdy et al., 2024). Sensor fusion is a promising direction, since chemical and structural signals are complementary and their combination has improved prediction where tested (de Medeiros et al., 2020b; Hong et al., 2023), though fusion increases cost and calibration complexity and its marginal benefit over a single well-chosen modality is not yet well characterised.

Beyond the technical requirements, operational adoption depends on standardisation and economics that the research literature seldom addresses. Seed certification operates within established national and international testing frameworks, and any learning-based measurement intended to inform certification would require validation against recognised reference methods, defined performance criteria and, ideally, the inter-laboratory reproducibility that underpins acceptance of conventional tests (Marcos Filho, 2015; Rahman & Cho, 2016). Interpretability acquires regulatory as well as scientific weight in this setting, since a decision to downgrade or reject a lot must be explicable and defensible, which favours transparent wavelength-selection and structurally grounded approaches over opaque end-to-end models (Ahmed et al., 2020; Fan et al., 2023). The economic case is equally consequential and equally neglected: the capital cost of hyperspectral or radiographic instrumentation, the throughput achievable per unit time, and the value of the errors avoided together determine whether a technically capable system is commercially viable, and these considerations differ markedly between high-value vegetable seed and bulk commodity grain. Research that pairs predictive performance with explicit accounting of cost and throughput would substantially strengthen the case for deployment.

The decisive obstacle to deployment is transfer. A model that performs well on one cultivar, instrument and harvest must be shown to hold, or to be efficiently recalibrated, when any of these changes, and the sparse cross-cultivar evidence indicates that transfer is neither automatic nor free (Qi et al., 2023). Operational systems will require calibration-maintenance strategies, standardised acquisition protocols, and reference labels anchored to field-relevant outcomes rather than to laboratory germination alone. Interpretability and traceability also acquire practical weight in a certification setting, where a decision to reject or downgrade a lot must be defensible; wavelength-selection and structurally grounded approaches have an advantage here over opaque end-to-end models (Ahmed et al., 2020; Fan et al., 2023). The gap between the reviewed proofs of concept and dependable digital seed technology is thus defined not by sensing capability, which is largely demonstrated, but by representativeness, validation and integration, which are largely not.

8. Future Directions

The most consequential unresolved problem is the mismatch between the labels models are trained to predict and the field-relevant physiological quality they are meant to serve. Current references are dominated by optimal-condition germination, binary viability and artificial ageing, none of which fully captures vigour as it governs emergence in variable field environments (B. Jin et al., 2022; Marcos Filho, 2015). The priority is to build reference standards that are biologically grounded and outcome-anchored, linking spectral and image predictions to naturally aged material, to biochemical vigour indicators, and ultimately to measured field emergence. The appropriate design is prospective and multi-season, in which lots of known provenance are sensed, stored, and followed through to controlled-environment and field performance, with vigour treated as a continuum rather than a dichotomy. Early dynamic indicators such as radicle emergence offer a tractable intermediate target that is both faster than the full germination test and closer to vigour than a static fingerprint (Shinohara et al., 2021).

A second priority is representativeness and generalisation. The dominance of single-cultivar, single-laboratory studies means that reported accuracy cannot be assumed to transfer, and the field lacks the shared, well-characterised datasets that have driven progress elsewhere in applied machine learning. Coordinated construction of open, multi-cultivar, multi-instrument and multi-season seed datasets, with standardised acquisition and transparent reference labelling, would allow models to be trained and, more importantly, externally validated on independent lots and sites (C. Jin et al., 2025; Qi et al., 2023). Such resources would also enable fair benchmarking, replacing the current situation in which each study reports internal accuracy on its own data with no common yardstick.

A third priority is validation design and reporting. Immediate gains are available simply by adopting independent-lot, cross-season and cross-instrument testing as the default, reporting class-sensitive metrics on imbalanced data, and treating field emergence as the ultimate criterion where the application demands it (T. Pang et al., 2023; Qi et al., 2023). Deep and simple models should be compared under equal tuning budgets, so that architectural claims are not confounded by effort, and wavelength-selection baselines should be retained as informative competitors rather than discarded (Fan et al., 2023; L. Pang et al., 2022). Interpretability analysis, linking selected wavelengths or image regions to known deterioration biochemistry and morphology, should accompany predictive results to build biological confidence and support certification.

A fourth priority is integration and deployment science. Research is needed on calibration transfer and maintenance under changing instruments and seasons, on the marginal value and cost of sensor fusion relative to a single optimised modality (de Medeiros et al., 2020b; Hong et al., 2023), and on throughput-preserving single-seed handling for sorting (Fan et al., 2023; Hamdy et al., 2024). Longer-term work should address the economic and regulatory conditions under which learning-based measurements could complement or partially substitute for reference tests in certification, and the standardisation required for cross-laboratory acceptance. These directions are prioritised, with reference labelling and external validation being immediate and foundational, and full deployment science being a longer-term endeavour that depends on the earlier priorities being met. The principal gaps and the designs suited to closing them are summarised in Table 5.

Table 5. Prioritised research gaps in artificial intelligence for seed physiological quality and recommended study designs

Research gap	Why current evidence is inadequate	Recommended design/approach	Priority horizon	Supporting references
Field-relevant reference labels	Optimal-condition germination and binary viability omit vigour and field emergence	Prospective multi-season sensing linked to field emergence and biochemical vigour indicators	Immediate, foundational	B. Jin et al., 2022; Marcos Filho, 2015; Shinohara et al., 2021
Representative shared datasets	Single-cultivar, single-laboratory data prevent generalisation and benchmarking	Open multi-cultivar, multi-instrument, multi-season datasets with standardised acquisition	Immediate	C. Jin et al., 2025; Qi et al., 2023
External validation	Internal cross-validation does not test transfer to new lots and sites	Independent-lot, cross-season and cross-instrument validation as default; class-sensitive metrics	Immediate	T. Pang et al., 2023; Qi et al., 2023
Model parsimony and interpretability	Deep advantage is conditional and often opaque	Equal-budget comparison with wavelength-selection baselines; feature attribution to biochemistry	Near-term	Fan et al., 2023; L. Pang et al., 2022; Zou et al., 2023

Research gap	Why current evidence is inadequate	Recommended design/approach	Priority horizon	Supporting references
Sensor fusion value	Complementarity shown but marginal benefit and cost unclear	Controlled fusion-versus-single-modality trials with cost accounting	Near-term	de Medeiros et al., 2020b; Hong et al., 2023
Deployment and calibration transfer	Laboratory pipelines untested for throughput and drift	Calibration-maintenance studies; single-seed high-throughput handling; certification pathways	Longer-term	Fan et al., 2023; Hamdy et al., 2024

Note. Priority horizon indicates the sequence in which gaps should be addressed; immediate priorities are foundational to later work

9. Conclusions

Artificial intelligence applied to spectral and image data has become the dominant experimental approach to non-destructive assessment of seed physiological quality, and the evidence that physiological deterioration leaves detectable optical and structural signatures is now consistent and broad. Across cereals, oilseeds, vegetables and tree species, NIR spectroscopy and hyperspectral imaging discriminate viable from non-viable seeds and separate vigour classes with high internal accuracy, image-based methods reproduce and accelerate established seedling and germination assessments, and X-ray radiography adds an interpretable structural dimension. Deep learning contributes genuine value where spatial or structural complexity is high, and transfer learning, generative augmentation and imbalance-aware modelling have been introduced to counter the small and skewed datasets that constrain the field.

These achievements are real but conditional, and the confidence they warrant must be calibrated to the evidence rather than to the headline metrics. The most defensible conclusion is that learning-based methods reliably detect deterioration under controlled, single-laboratory conditions; the least supported is that this performance transfers to operational settings, because the field relies overwhelmingly on artificial ageing, narrow cultivar and instrument coverage, proxy reference labels and internal validation. Reported accuracy therefore measures internal discrimination more than demonstrated field relevance, and the frequent near-perfect figures reflect favourable conditions as much as method quality. Deep learning does not uniformly surpass well-constructed chemometric baselines, and wavelength-selection models remain competitive and more interpretable for many tasks.

The path to dependable digital seed technology runs through data, references and validation rather than through architectural novelty. Representative multi-cultivar and multi-season datasets, biologically grounded and outcome-anchored reference labels, external validation against independent lots and field emergence, honest comparison of complex and simple models, and attention to interpretability and deployment together define what is needed to convert a large body of promising proofs of concept into transferable decision support. Judged against those requirements, the field is technically mature in sensing and analysis but still early in the evidence that would justify operational trust, and closing that gap is the central task ahead.

10. Limitations

This review is a critical narrative synthesis and carries the limitations intrinsic to that design. Study selection and interpretation involved judgement, and despite a transparent search, explicit eligibility criteria and systematic bibliographic verification, selection and interpretive bias cannot be excluded, and a differently scoped review might weight the evidence differently. The synthesis is qualitative, and no quantitative pooling of effect sizes was attempted, both because the underlying studies are too heterogeneous in crop, endpoint, reference standard and metric to combine meaningfully and because such heterogeneity would render any pooled estimate misleading.

The evidence base itself imposes further limits on what can be concluded. Subscription-only databases were not directly searched, and although open scholarly indexes and citation tracing provided broad coverage, some relevant work may not have been captured, and the restriction to English-language peer-reviewed articles may have introduced language and publication bias. The primary literature is geographically concentrated, dominated by particular crops, and reliant on artificial ageing and internal validation, so the review inherits these biases and can characterise but not correct them. The rapid pace of development means that methods and results continue to evolve quickly, and the scarcity of long-term, field-anchored and independently validated studies means that conclusions about operational transfer remain provisional. These constraints reinforce the review's central caution that current performance evidence describes internal discrimination more securely than real-world seed-quality decision-making.

Declaration of AI Use

This manuscript was prepared through the combined contributions of all author(s), including contributions to the study design, data, content development, results, interpretation, and related scholarly work. The author(s) acknowledge the use of Grammarly and ChatGPT to assist with grammar checking, language refinement, reference formatting. These AI-assisted tools were not used as authors and did not replace the intellectual contributions or scholarly judgment of the author(s). All AI-assisted outputs, including content, references, and interpretations, were carefully reviewed, revised, verified, and approved by the author(s). The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Competing Interests

Authors have declared that no competing interests exist.

References

Ahmed, M. R., Yasmin, J., Park, E., Kim, G., Kim, M. S., Wakholi, C., Mo, C., & Cho, B.-K. (2020). Classification of watermelon seeds using morphological patterns of X-ray imaging: A comparison of conventional machine learning and deep learning. *Sensors, 20*(23), Article 6753. https://doi.org/10.3390/s20236753

Awty-Carroll, D., Clifton-Brown, J., & Robson, P. (2018). Using k-NN to analyse images of diverse germination phenotypes and detect single seed germination in *Miscanthus sinensis*. *Plant Methods, 14*(1), Article 5. https://doi.org/10.1186/s13007-018-0272-0

Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—A scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review, 4*(1), Article 5. https://doi.org/10.1186/s41073-019-0064-8

Castan, D. O. C., Gomes-Junior, F. G., & Marcos-Filho, J. (2018). Vigor-S, a new system for evaluating the physiological potential of maize seeds. *Scientia Agricola, 75*(2), 167–172. https://doi.org/10.1590/1678-992x-2016-0401

Colmer, J., O'Neill, C. M., Wells, R., Bostrom, A., Reynolds, D., Websdale, D., Shiralagi, G., Lu, W., Lou, Q., Le Cornu, T., Ball, J., Renema, J., Flores Andaluz, G., Benjamins, R., Penfield, S., & Zhou, J. (2020). SeedGerm: A cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination. *New Phytologist, 228*(2), 778–793. https://doi.org/10.1111/nph.16736

de Medeiros, A. D., Capobiango, N. P., da Silva, J. M., da Silva, L. J., da Silva, C. B., & dos Santos Dias, D. C. F. (2020a). Interactive machine learning for soybean seed and seedling quality classification. *Scientific Reports, 10*(1), Article 11267. https://doi.org/10.1038/s41598-020-68273-y

de Medeiros, A. D., da Silva, L. J., Ribeiro, J. P. O., Ferreira, K. C., Rosas, J. T. F., Santos, A. A., & Barboza da Silva, C. (2020b). Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. *Sensors, 20*(15), Article 4319. https://doi.org/10.3390/s20154319

de Medeiros, A. D., Pinheiro, D. T., Xavier, W. A., da Silva, L. J., & dos Santos Dias, D. C. F. (2020c). Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning. *Industrial Crops and Products, 146*, Article 112162. https://doi.org/10.1016/j.indcrop.2020.112162

Fan, Y., An, T., Wang, Q., Yang, G., Huang, W., Wang, Z., Zhao, C., & Tian, X. (2023). Non-destructive detection of single-seed viability in maize using hyperspectral imaging technology and multi-scale 3D convolutional neural network. *Frontiers in Plant Science, 14*, Article 1248598. https://doi.org/10.3389/fpls.2023.1248598

Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., & Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods, 15*(1), Article 91. https://doi.org/10.1186/s13007-019-0476-y

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing, 24*(4), 230–235. https://doi.org/10.1179/2047480615z.000000000329

Galletti, P. A., Carvalho, M. E. A., Hirai, W. Y., Brancaglioni, V. A., Arthur, V., & Barboza da Silva, C. (2020). Integrating optical imaging tools for rapid and non-invasive characterization of seed quality: Tomato (*Solanum lycopersicum* L.) and carrot (*Daucus carota* L.) as study cases. *Frontiers in Plant Science, 11*, Article 577851. https://doi.org/10.3389/fpls.2020.577851

Genze, N., Bharti, R., Grieb, M., Schultheiss, S. J., & Grimm, D. G. (2020). Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods, 16*(1), Article 157. https://doi.org/10.1186/s13007-020-00699-x

Gonçalves, N. R., Cicero, S. M., & Abud, H. F. (2017). Seedling image analysis and traditional tests to evaluate onion seed vigor. *Journal of Seed Science, 39*(3), 216–223. https://doi.org/10.1590/2317-1545v39n3160444

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gulzar, Y., Hamid, Y., Soomro, A. B., Alwan, A. A., & Journaux, L. (2020). A convolution neural network-based seed classification system. *Symmetry, 12*(12), Article 2018. https://doi.org/10.3390/sym12122018

Hamdy, S., Charrier, A., Corre, L. L., Rasti, P., & Rousseau, D. (2024). Toward robust and high-throughput detection of seed defects in X-ray images via deep learning. *Plant Methods, 20*(1), Article 63. https://doi.org/10.1186/s13007-024-01195-2

Hong, S.-J., Park, S., Lee, A., Kim, S.-Y., Kim, E., Lee, C.-H., & Kim, G. (2023). Nondestructive prediction of pepper seed viability using single and fusion information of hyperspectral and X-ray images. *Sensors and Actuators A: Physical, 350*, Article 114151. https://doi.org/10.1016/j.sna.2022.114151

Javanmardi, S., Miraei Ashtiani, S.-H., Verbeek, F. J., & Martynenko, A. (2021). Computer-vision classification of corn seed varieties using deep convolutional neural network. *Journal of Stored Products Research, 92*, Article 101800. https://doi.org/10.1016/j.jspr.2021.101800

Jin, B., Qi, H., Jia, L., Tang, Q., Gao, L., Li, Z., & Zhao, G. (2022). Determination of viability and vigor of naturally-aged rice seeds using hyperspectral imaging with machine learning. *Infrared Physics & Technology, 122*, Article 104097. https://doi.org/10.1016/j.infrared.2022.104097

Jin, C., Zhou, L., Pu, Y., Zhang, C., Qi, H., & Zhao, Y. (2025). Application of deep learning for high-throughput phenotyping of seed: A review. *Artificial Intelligence Review, 58*(3), Article 76. https://doi.org/10.1007/s10462-024-11079-5

Kusumaningrum, D., Lee, H., Lohumi, S., Mo, C., Kim, M. S., & Cho, B. (2018). Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy. *Journal of the Science of Food and Agriculture, 98*(5), 1734–1742. https://doi.org/10.1002/jsfa.8646

Lakshmanan, M. K., Boelt, B., & Gislum, R. (2023). A chemometric method for the viability analysis of spinach seeds by near infrared spectroscopy with variable selection using successive projections algorithm. *Journal of Near Infrared Spectroscopy, 31*(1), 24–32. https://doi.org/10.1177/09670335221138955

Ma, T., Tsuchikawa, S., & Inagaki, T. (2020). Rapid and non-destructive seed viability prediction using near-infrared hyperspectral imaging coupled with a deep learning approach. *Computers and Electronics in Agriculture, 177*, Article 105683. https://doi.org/10.1016/j.compag.2020.105683

Mahajan, S., Mittal, S. K., & Das, A. (2018). Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *Journal of Food Science and Technology, 55*(10), 3949–3959. https://doi.org/10.1007/s13197-018-3320-x

Marcos Filho, J. (2015). Seed vigor testing: An overview of the past, present and future perspective. *Scientia Agricola, 72*(4), 363–374. https://doi.org/10.1590/0103-9016-2015-0007

Nadarajan, J., Walters, C., Pritchard, H. W., Ballesteros, D., & Colville, L. (2023). Seed longevity—The evolution of knowledge and a conceptual framework. *Plants, 12*(3), Article 471. https://doi.org/10.3390/plants12030471

Nehoshtan, Y., Carmon, E., Yaniv, O., Ayal, S., & Rotem, O. (2021). Robust seed germination prediction using deep learning and RGB image data. *Scientific Reports, 11*(1), Article 22030. https://doi.org/10.1038/s41598-021-01712-6

Onwimol, D., Chakranon, P., Wonggasem, K., & Wongchaisuwat, P. (2025). Non-destructive assessment of hemp seed vigor using machine learning and deep learning models with hyperspectral imaging. *Journal of Agriculture and Food Research, 21*, Article 101836. https://doi.org/10.1016/j.jafr.2025.101836

Pang, L., Men, S., Yan, L., & Xiao, J. (2020). Rapid vitality estimation and prediction of corn seeds based on spectra and images using deep learning and hyperspectral imaging techniques. *IEEE Access, 8*, 123026–123036. https://doi.org/10.1109/access.2020.3006495

Pang, L., Wang, L., Yuan, P., Yan, L., & Xiao, J. (2022). Rapid seed viability prediction of *Sophora japonica* by improved successive projection algorithm and hyperspectral imaging. *Infrared Physics & Technology, 123*, Article 104143. https://doi.org/10.1016/j.infrared.2022.104143

Pang, L., Wang, L., Yuan, P., Yan, L., Yang, Q., & Xiao, J. (2021). Feasibility study on identifying seed viability of *Sophora japonica* with optimized deep neural network and hyperspectral imaging. *Computers and Electronics in Agriculture, 190*, Article 106426. https://doi.org/10.1016/j.compag.2021.106426

Pang, T., Chen, C., Fu, R., Wang, X., & Yu, H. (2023). An end-to-end seed vigor prediction model for imbalanced samples using hyperspectral image. *Frontiers in Plant Science, 14*, Article 1322391. https://doi.org/10.3389/fpls.2023.1322391

Peng, Q., Tu, L., Wu, Y., Yu, Z., Tang, G., & Song, W. (2022). Automatic monitoring system for seed germination test based on deep learning. *Journal of Electrical and Computer Engineering, 2022*, Article 4678316. https://doi.org/10.1155/2022/4678316

Qi, H., Huang, Z., Jin, B., Tang, Q., Jia, L., Zhao, G., Cao, D., Sun, Z., & Zhang, C. (2024). SAM-GAN: An improved DCGAN for rice seed viability determination using near-infrared hyperspectral imaging. *Computers and Electronics in Agriculture, 216*, Article 108473. https://doi.org/10.1016/j.compag.2023.108473

Qi, H., Huang, Z., Sun, Z., Tang, Q., Zhao, G., Zhu, X., & Zhang, C. (2023). Rice seed vigor detection based on near-infrared hyperspectral imaging and deep transfer learning. *Frontiers in Plant Science, 14*, Article 1283921. https://doi.org/10.3389/fpls.2023.1283921

Qiao, J., Liao, Y., Yin, C., Yang, X., Tú, H. M., Wang, W., & Liu, Y. (2023). Vigour testing for the rice seed with computer vision-based techniques. *Frontiers in Plant Science, 14*, Article 1194701. https://doi.org/10.3389/fpls.2023.1194701

Rahman, A., & Cho, B.-K. (2016). Assessment of seed quality using non-destructive measurement techniques: A review. *Seed Science Research, 26*(4), 285–305. https://doi.org/10.1017/s0960258516000234

Reed, R. C., Bradford, K. J., & Khanday, I. (2022). Seed germination and vigor: Ensuring crop sustainability in a changing climate. *Heredity, 128*(6), 450–459. https://doi.org/10.1038/s41437-022-00497-2

Rodrigues, M., Gomes-Junior, F. G., & Marcos-Filho, J. (2020). Vigor-S: System for automated analysis of soybean seed vigor. *Journal of Seed Science, 42*, Article e202042039. https://doi.org/10.1590/2317-1545v42237490

Rybacki, P., Niemann, J., Bahcevandziev, K., & Durczak, K. (2023). Convolutional neural network model for variety classification and seed quality assessment of winter rapeseed. *Sensors, 23*(5), Article 2486. https://doi.org/10.3390/s23052486

Shinohara, T., Ducournau, S., Matthews, S., Wagner, M.-H., & Powell, A. A. (2021). Early counts of radicle emergence, counted manually and by image analysis, can reveal differences in the production of normal seedlings and the vigour of seed lots of cauliflower. *Seed Science and Technology, 49*(3), 219–235. https://doi.org/10.15258/sst.2021.49.3.04

Shrestha, S., Deleuran, L. C., & Gislum, R. (2017). Separation of viable and non-viable tomato (*Solanum lycopersicum* L.) seeds using single seed near-infrared spectroscopy. *Computers and Electronics in Agriculture, 142*, 348–355. https://doi.org/10.1016/j.compag.2017.09.004

Silva, M. F. D., Roque, J. V., Soares, J. M., Moura, L. D. O., Medeiros, A. D. D., Silva, F. L. D., & Silva, L. J. D. (2024). Near infrared spectroscopy for the classification of vigor level of soybean seed. *Revista Ciência Agronômica, 55*, e20238703. https://doi.org/10.5935/1806-6690.20240005

Silva, V. N., & Cicero, S. M. (2014). Image seedling analysis to evaluate tomato seed physiological potential. *Revista Ciência Agronômica, 45*(2), 327–334. https://doi.org/10.1590/s1806-66902014000200014

Sukhera, J. (2022). Narrative reviews: Flexible, rigorous, and practical. *Journal of Graduate Medical Education, 14*(4), 414–417. https://doi.org/10.4300/jgme-d-22-00480.1

Sun, J., Nirere, A., Dusabe, K. D., Yuhao, Z., & Adrien, G. (2024). Rapid and nondestructive watermelon (*Citrullus lanatus*) seed viability detection based on visible near-infrared hyperspectral imaging technology and machine learning algorithms. *Journal of Food Science, 89*(7), 4403–4418. https://doi.org/10.1111/1750-3841.17151

Tosun, O., & Eryiğit, R. (2026). Deep learning-based seed germination prediction using morphological traits and RGB images. *BMC Plant Biology, 26*, Article 757. https://doi.org/10.1186/s12870-026-08599-3

Wang, Y., & Song, S. (2024). Detection of sweet corn seed viability based on hyperspectral imaging combined with firefly algorithm optimized deep learning. *Frontiers in Plant Science, 15*, Article 1361309. https://doi.org/10.3389/fpls.2024.1361309

Wongchaisuwat, P., Chakranon, P., Yinpin, A., Onwimol, D., & Wonggasem, K. (2025). Rapid maize seed vigor classification using deep learning and hyperspectral imaging techniques. *Smart Agricultural Technology, 10*, Article 100820. https://doi.org/10.1016/j.atech.2025.100820

Xia, Y., Xu, Y., Li, J., Zhang, C., & Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture, 1*, 35–47. https://doi.org/10.1016/j.aiia.2019.05.001

Xu, H., Li, F., Wang, Z., Zhang, W., Li, H., Xia, X., Li, C., & Xin, X. (2025). Non-destructive viability detection of naturally-aged large-sample rice seeds using hyperspectral imaging and deep learning. *Journal of Agriculture and Food Research, 24*, Article 102333. https://doi.org/10.1016/j.jafr.2025.102333

Yang, Y., Chen, J., He, Y., Liu, F., Feng, X., & Zhang, J. (2020). Assessment of the vigor of rice seeds by near-infrared hyperspectral imaging combined with transfer learning. *RSC Advances, 10*(72), 44149–44158. https://doi.org/10.1039/d0ra06938h

Yao, Q., Zheng, X., Zhou, G., & Zhang, J. (2024). SGR-YOLO: A method for detecting seed germination rate in wild rice. *Frontiers in Plant Science, 14*, Article 1305081. https://doi.org/10.3389/fpls.2023.1305081

Zhang, C., Si, Y., Lamkey, J., Boydston, R. A., Garland-Campbell, K. A., & Sankaran, S. (2018). High-throughput phenotyping of seed/seedling evaluation using digital image analysis. *Agronomy, 8*(5), Article 63. https://doi.org/10.3390/agronomy8050063

Zou, Z., Chen, J., Wu, W., Luo, J., Long, T., Wu, Q., Wang, Q., Zhen, J., Zhao, Y., Wang, Y., Chen, Y., Zhou, M., & Xu, L. (2023). Detection of peanut seed vigor based on hyperspectral imaging and chemometrics. *Frontiers in Plant Science, 14*, Article 1127108. https://doi.org/10.3389/fpls.2023.1127108

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/163476