Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6, 355-373
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i6.13148

A systematic literature review of shallow and deep prior modeling methods for hyperspectral-multispectral image fusion

Wenchao Cai[1]*, Wei Jen Chew[2], Pei Xuan Ku[3]
[1,2,3]School of Engineering, Faculty of Innovation and Technology, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia.
caiwenchao019@sina.com (W.C.) WeiJen.Chew@taylors.edu.my (W.J.C.) PeiXuan.Ku@taylors.edu.my (P.X.K.).

Abstract: This review examines the evolution of prior modeling in hyperspectral–multispectral (HS–MS) image fusion and compares how shallow, deep, and hybrid priors address reconstruction accuracy, interpretability, and real-world deployment challenges. A PRISMA-guided review was conducted on 62 journal studies retrieved from Scopus, Taylor & Francis Online, and Wiley Online Library. The selected studies were classified according to methodological paradigm, historical trajectory, prior type, and reported limitations. The evidence shows a clear transition from Bayesian, matrix-factorization, sparse, and tensor-based priors to CNN-, GAN-, and Transformer-based deep models, followed by model-driven and physics-guided hybrid frameworks. Shallow priors remain effective for spectral fidelity, limited data, and interpretability, whereas deep priors enhance nonlinear feature learning and spatial detail recovery. Hybrid priors increasingly combine these strengths but remain constrained by sensor uncertainty, spectral variability, and domain shift. No single paradigm fully resolves the practical challenges of HS–MS fusion. Generalizable hybrid prior modeling represents the most promising direction. Future research should embed sensor physics, uncertainty awareness, and real-scene robustness into learning-based fusion to improve cross-sensor applicability and operational deployment.

Keywords: *Deep prior, Hyperspectral-multispectral image fusion, Model-driven deep learning, Remote sensing, Shallow prior, Systematic literature review.*

1. Introduction

Hyperspectral-multispectral (HS-MS) image fusion, often referred to in the adjacent literature as hyperspectral super-resolution or hypersharpening, addresses a persistent sensing trade-off in remote sensing systems: sensors with high spectral resolution typically operate at low spatial resolution, whereas sensors with high spatial resolution typically provide only a small number of broad spectral bands. Fusion seeks to reconstruct a latent high-resolution hyperspectral cube by combining these complementary observations. The problem is important because many downstream remote sensing tasks, including material identification, precision agriculture, land-cover mapping, environmental monitoring, and urban analysis, require both fine spectral discrimination and spatial detail [1]. As a result, HS-MS fusion has become a central topic at the intersection of signal processing, inverse problems, and machine learning.

The technical challenge of HS-MS fusion goes beyond a simple image-merging task. In practical settings, the observed low-resolution hyperspectral image (HSI) and high-resolution multispectral image (MSI) are linked to the latent target through sensor-dependent degradations, including spectral response functions (SRFs), point-spread functions (PSFs), blurring, downsampling, noise, and potential cross-image variability. Consequently, the fusion problem is ill-posed: many high-resolution hyperspectral cubes can plausibly explain the same pair of observations. Prior modeling is therefore not a secondary design choice; it is the central mechanism that makes the inverse problem solvable. Whether the prior is written explicitly as a Bayesian regularizer, a matrix/tensor factorization

History: Received: 3 April 2026; Revised: 8 June 2026; Accepted: 12 June 2026; Published: 24 June 2026
* Correspondence: caiwenchao019@sina.com

constraint, a sparsity assumption, or learned implicitly through deep neural networks, the fusion result is fundamentally determined by the structure imposed on the latent image and on the degradation process.

Early research in the field was grounded in model-based formulations. Bayesian and wavelet-domain methods exploited statistical assumptions to stabilize inference [2]. Coupled nonnegative matrix factorization (CNMF) and related unmixing approaches framed fusion as shared endmember-abundance estimation under physically motivated observation models [3]. Convex and Bayesian matrix-factorization variants improved estimation stability, while optimization-based and Poissonian formulations further enriched the shallow-prior repertoire [4, 5]. These methods established much of the field's conceptual vocabulary: spectral subspaces, low-rank structure, abundance consistency, sparse coefficients, and regularized inverse reconstruction. More importantly, they set the benchmark for spectral faithfulness and interpretability, two criteria that remain decisive in remote sensing applications.

As the field matured, it became increasingly clear that handcrafted shallow priors, while principled, do not always capture the full complexity of HS-MS fusion. Real remote sensing imagery often exhibits nonlinear mixtures, local texture variability, cross-scene heterogeneity, unknown or mismatched sensor responses, and scale interactions that challenge rigid analytical assumptions. This limitation helped motivate a transition toward deep prior modeling. Deep hyperspectral sharpening demonstrated that convolutional neural networks (CNNs) could learn fusion mappings directly from data and outperform traditional methods on benchmark settings [6]. Subsequent work expanded the design space with tensor-aware neural networks, unsupervised variational models, adversarial learning, and detail-injection architectures [7-13]. The recent emergence of Transformer-based fusion and other foundation-like sequence modeling strategies has extended deep fusion from local convolutional reasoning to broader spectral-spatial context aggregation [14-17].

Yet the apparent shift from shallow to deep priors is not adequately described as a simple replacement of old methods by new ones. The most consequential recent trend is methodological integration. Deep networks have inherited, adapted, and re-embedded concepts that originated in shallow prior modeling: spectral subspaces, unmixing, matrix decomposition, iterative optimization, and physics-based degradation consistency. Deep unfolding networks are explicit evidence of this convergence, translating iterative optimization steps into trainable stages while preserving interpretable structure [13]. Physics-guided adversarial fusion and related hybrid designs likewise demonstrate that learning-based models increasingly require model-driven constraints to improve reliability beyond synthetic benchmarks [18, 19]. In other words, the key contemporary question is not whether shallow or deep priors are superior in the abstract, but how these paradigms can be combined to balance performance, interpretability, and robustness.

This shift creates a need for a more conceptually ambitious review than a simple catalog of methods. Prior reviews have documented general image-fusion techniques or broad hyperspectral enhancement strategies [1], but the literature now warrants a focused synthesis centered on prior modeling itself. Such a synthesis should explain how shallow and deep priors can be taxonomized, how the field has evolved from explicit handcrafted constraints to implicit learned priors, and why hybridization has become central. It should also identify the unresolved bottlenecks that continue to limit operational HS-MS fusion, especially spectral fidelity, blind or semi-blind degradation estimation, and generalization across sensors and scenes.

Accordingly, this review addresses three research questions. RQ1 asks how existing shallow- and deep-prior modeling methods for HS-MS image fusion can be systematically classified and what evolutionary trajectory they exhibit. RQ2 asks what comparative advantages and limitations define shallow versus deep prior modeling, and how these paradigms are increasingly integrated to improve fusion performance, interpretability, and robustness. RQ3 asks which critical bottlenecks continue to constrain current prior modeling, particularly regarding spectral fidelity and generalization, and which trajectories appear most promising for real-world remote sensing scenarios. These questions position the review at the methodological core of the field rather than at the level of application overview.

The contribution of this article is fourfold. First, it builds a taxonomy of prior-modeling paradigms that distinguishes not only shallow and deep methods, but also the growing class of hybrid and model-driven deep approaches. Second, it reconstructs the field's methodological evolution from Bayesian and factorization-based priors to Transformers, diffusion-like designs, and physics-guided learning. Third, it synthesizes the comparative logic of the literature, clarifying where shallow priors remain advantageous and where deep priors create genuine advances. Fourth, it distills the bottlenecks that recur across the included studies and translates them into a concrete future agenda for generalizable, real-scene fusion systems.

The remainder of the article is organized as follows. Section 2 describes the PRISMA-guided review methodology, including the database strategy, screening process, and inclusion/exclusion criteria. Section 3 reports the findings for RQ1-RQ3, supported by structured tables. Section 4 discusses the implications of the results for theory, algorithm design, and practical deployment. Section 5 concludes the review and outlines the next research frontier in prior modeling for HS-MS fusion.

2. Methodology

2.1. Review Design

This study adopted a systematic literature review (SLR) design and followed the PRISMA 2020 reporting guidelines to ensure transparent, reproducible identification, screening, eligibility assessment, and inclusion of studies [20]. Although PRISMA was originally formulated for intervention-oriented systematic reviews, its structured identification and reporting workflow is increasingly used in engineering and computer-science evidence synthesis when a field has become too large, heterogeneous, or fast-moving for narrative review alone. That rationale applies directly to HS-MS image fusion: the literature contains a mixture of optimization-based, probabilistic, factorization-based, and learning-based methods distributed across multiple publishing venues and described with inconsistent terminology.

The review did not aim to perform a meta-analysis because the included studies differed substantially in datasets, sensor assumptions, degradation models, evaluation protocols, and reported metrics. Instead, the design emphasized qualitative synthesis supported by structured extraction. This was especially suitable because the review questions were not limited to "which method performs best," but focused on method taxonomy, paradigm evolution, integration mechanisms, and bottlenecks. The unit of analysis was the peer-reviewed journal article. The final corpus was analyzed comparatively rather than study-by-study so that the synthesis could identify cross-paper patterns, recurring tensions, and emerging trajectories.

2.2. Data Sources and Search Strategy

Literature was searched in three sources: Scopus, Taylor & Francis Online, and Wiley Online Library. Scopus served as the primary indexing database because it supports advanced document search and field-specific searches, such as TITLE-ABS-KEY, allowing the review to target topic-relevant records while maintaining broad recall Elsevier [21]. Wiley Online Library [22] was included because it provides advanced search options suitable for complex scholarly retrieval Wiley Online Library [22]. Taylor & Francis [23] Online was included as an additional publisher platform with searchable journal content relevant to image processing and remote sensing [23]. The combination was not intended to treat publisher platforms as equivalent to abstracting-and-indexing databases; rather, it was designed to complement Scopus with broader publisher-level coverage under the access conditions available to the study.

A key methodological decision concerned the search terminology. Initial testing showed that explicitly searching for the phrase "prior modeling" in Scopus dramatically under-retrieved relevant literature, as many foundational and high-impact HS-MS fusion papers do not label their methods with that term. Instead, they describe model families through problem-specific keywords such as "matrix factorization," "sparse representation," "low-rank," "tensor," "detail injection," "deep network," or "transformer." To reduce the risk of missing relevant studies, the search strategy therefore treated HS-MS fusion as the primary retrieval target. It relied on subsequent title/abstract and full-text screening to determine whether a study genuinely contributed to shallow, deep, or hybrid prior modeling.

The Scopus search that generated the main record set was performed in TITLE-ABS-KEY using a task-centered Boolean expression for hyperspectral and multispectral image fusion. The query retrieved 312 records. Additional searches in Taylor & Francis Online and Wiley Online Library yielded 47 and 14 records, respectively. Across the three sources, 373 records were identified before deduplication. Nineteen duplicates were removed before screening, resulting in 354 unique records. The study reported the final source counts and screening stages in a PRISMA-style flow summary, reproduced in Table I.

2.3. Eligibility Criteria

Inclusion and exclusion criteria were defined before full-text reading to ensure consistent screening logic. Studies were included if they met all of the following criteria. First, the study had to focus on HS-MS image fusion, hyperspectral super-resolution using MSI guidance, or a directly equivalent remote-sensing fusion setting. Second, the article had to propose, analyze, or evaluate a method whose core contribution could be interpreted as prior modeling, including explicit shallow priors, implicit deep priors, or hybrid/model-driven combinations. Third, the study had to provide sufficient methodological detail to support classification and synthesis. Fourth, the record had to be a journal article or review article written in English.

Studies were excluded if they met any of the following conditions. First, papers focused on other image-fusion tasks, such as pansharpening outside the HS-MS setting, infrared-visible fusion, medical image fusion, or generic multimodal fusion, were excluded. Second, conference papers, editorials, notes, book chapters, and other non-peer-reviewed or non-journal formats were excluded from the final synthesis. Third, studies that emphasized only application outcomes, without substantive fusion-method modeling, were excluded. Fourth, records whose full texts could not be retrieved were excluded at the report-retrieval stage. Finally, studies that appeared relevant at the title/abstract level but proved not to address HS-MS prior modeling at the full-text level were excluded during eligibility assessment.

2.4. Screening Procedure

The screening process unfolded in four stages. In the identification stage, the three databases and platforms yielded 373 records. In the deduplication stage, 19 duplicate records were removed, leaving 354 unique items for screening. During title and abstract screening, 72 records were excluded for one of three reasons: they addressed non-HS-MS image-fusion tasks (n = 38), were conference/editorial/non-peer-reviewed publications (n = 21), or focused only on application outcomes without modeling (n = 13). This left 282 reports to be retrieved.

A notable feature of the review is that 209 reports could not be retrieved in full text. Because this number is unusually high, it deserves explicit acknowledgment as a review limitation rather than silent absorption into the screening chain. The inaccessible reports were primarily records indexed through searchable platforms but unavailable through the access pathways available to this project. Seventy-three full-text reports were successfully retrieved and assessed for eligibility. Of these, 11 were excluded after full-text reading: five lacked sufficient methodological detail for prior-modeling analysis, three turned out to be primarily application-oriented without substantive methodological contribution, and three were judged to concern non-HS-MS image fusion despite appearing initially relevant. The final qualitative synthesis, therefore, comprised 62 journal studies.

The review treats this final pool as an analytically coherent, though not exhaustive, representation of the field. The inaccessible reports mean that the study should not be interpreted as the final bibliographic census of all HS-MS fusion research. However, because the 62 included papers span 2009-

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 355-373, 2026
DOI: 10.55214/2576-8484.v10i6.13148

2026 and cover the major shallow, deep, and hybrid paradigms, the corpus is sufficiently broad to support the review's conceptual and comparative aims.

2.5. Data Extraction and Synthesis Strategy

A structured extraction sheet was developed to support synthesis around the three research questions. For each full-text study, the extraction recorded the screening decision, the triggering exclusion rule where relevant, and dedicated extraction notes for RQ1, RQ2, and RQ3. The RQ1 field summarizes the paper's contribution to the classification and evolution of prior modeling. The RQ2 field captured the stated or inferable advantages, limitations, and integration mechanisms of the method relative to shallow and deep priors. The RQ3 field recorded bottlenecks, open problems, and future directions discussed or implied by the study. This design moved the review beyond bibliographic listing toward question-driven analytical coding.

The synthesis followed a comparative thematic logic. First, the 62 included studies were grouped into broad methodological families and then re-read across those families to identify the field's evolutionary trajectory. Second, papers were compared based on their assumptions, strengths, and failure modes rather than solely on reported numerical metrics. Third, recurring bottlenecks were aggregated across papers to identify high-frequency limitations such as spectral distortion, generalization failure, sensor mismatch, or incomplete physical modeling. This approach made it possible to derive not only what methods exist, but how the field thinks about its own limitations and future evolution.

2.6. Trustworthiness and Limitations of the Review Process

Like any SLR, the present study is shaped by its retrieval boundaries. The reliance on a single major indexing database and two publisher platforms may have reduced coverage compared with reviews that combine multiple large abstracting databases. The high number of unretrieved reports is another limitation. Nevertheless, the study's internal trustworthiness is supported by four features. First, the review questions were defined before synthesis. Second, the eligibility criteria were explicit and consistent. Third, the extraction form was aligned directly with the review questions, preventing the synthesis from drifting into an unsystematic narrative summary. Fourth, the final corpus includes foundational, transitional, and recent frontier studies across the major branches of HS-MS fusion. These features do not eliminate selection bias, but they make the review logic explicit and auditable.

Figure 1 summarizes the search sources and PRISMA counts used in the remainder of the article.

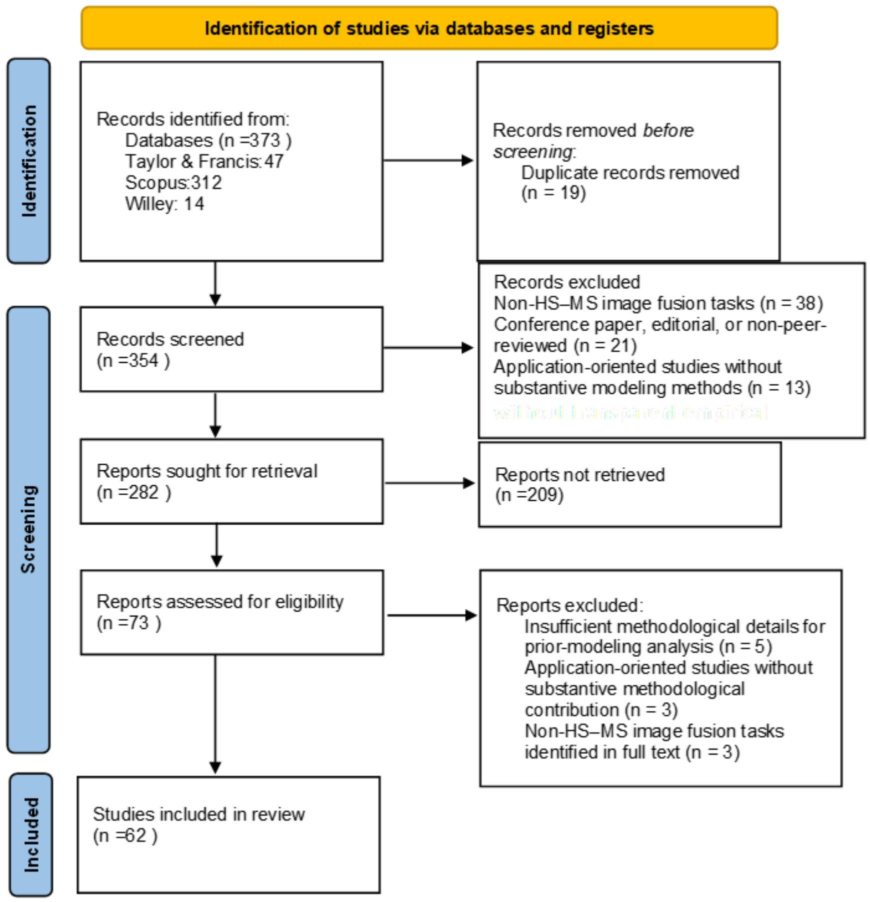

Figure 1.
Search sources and PRISMA screening summary.

3. Findings

3.1. Corpus Profile and Publication Trend

The final pool of 62 studies covers the period from 2009 to 2026. The temporal distribution itself is informative because it reflects the field's shift from foundational probabilistic and factorization methods to a recent surge of deep and hybrid architectures. Only a small number of studies appear before 2017, and these are largely shallow prior works that establish the basic inverse-problem formulations of the field [2-5]. A stronger expansion begins around 2018, when deep hyperspectral sharpening and related learning-based models start to reframe the problem [6]. The corpus then accelerates after 2020, with substantial growth in both optimization-aware tensor methods and learning-based fusion networks. The period from 2021 onward is especially important: it contains the densest concentration of papers in the final corpus and also the greatest methodological diversity, including GAN-based, Transformer-based, self-supervised, and deep-unfolding approaches [11-19, 24, 25].

This temporal pattern suggests that HS-MS fusion has entered a mature yet unsettled phase. Mature, because the field now has established benchmarks, recurring evaluation metrics, and a recognizable set of canonical method families. Unsettled, because there is still no consensus that one paradigm dominates across all realistic settings. Instead, the literature increasingly fragments by assumptions: some papers prioritize physical interpretability and sensor realism, some maximize benchmark performance through data-driven learning, and others attempt to connect both worlds through hybrid models. The final pool, therefore, supports the claim that the history of HS-MS fusion is not just a chronology of algorithms; it is a shifting debate over what kinds of priors are most trustworthy for an ill-posed imaging problem.

3.2. Findings for RQ1: Taxonomy and Evolutionary Trajectory of Prior Modeling

Table 1.
Taxonomy of prior-modeling paradigms in the final pool.

Paradigm	Main subfamilies	Representative studies	Typical strengths	Typical limitations
Shallow prior	Bayesian/statistical; CNMF and unmixing; sparse/low-rank; tensor decomposition; variational/transform-domain	Zhang, et al. [2]; Yokoya, et al. [3]; Lin, et al. [4]; Zou and Xia [5], Dian, et al. [7], Zhang, et al. [8], Guo, et al. [16], Long and Peng [17], Lin, et al. [19], Benhalouche, et al. [24], Wang, et al. [26]; Priya and Rajkumar [27] and Xu, et al. [28]	High interpretability; strong spectral grounding; lower dependence on large paired training sets	Rigid assumptions; sensitivity to model mismatch; heavier iterative computation in blind/complex settings
Deep prior	CNN and residual learning; detail injection; unsupervised variational networks; GANs; attention and Transformers	Dian, et al. [6], Wang, et al. [10]; Luo, et al. [11]; Lu, et al. [12], Yu, et al. [14], Zhu, et al. [15], Xiao, et al. [18] and Zheng, et al. [29]	Strong nonlinear representation; effective spatial-detail recovery; end-to-end learning	Weak interpretability; higher data dependence; risk of spectral distortion and domain shift
Hybrid / model-driven deep	Deep unfolding; physics-guided adversarial learning; decomposition-aware and response-learning networks.	Zhang, et al. [13], Xiao, et al. [18], Borsoi, et al. [25] and Zheng, et al. [29]	Balances structure and flexibility; better potential robustness and physical consistency	Still maturing; can be architecturally complex; real-scene generalization remains unresolved.

RQ1 asked how existing shallow- and deep-prior modeling methods can be systematically classified and what evolutionary trajectory they exhibit. The evidence from the 62 included studies indicates that a useful taxonomy should distinguish three major paradigms rather than only two: shallow prior modeling, deep prior modeling, and hybrid/model-driven deep prior modeling. The third category is essential because it captures a large and growing share of recent work that cannot be described adequately as either purely handcrafted or purely data-driven.

3.2.1. Shallow Prior Modeling

Shallow prior modeling refers to methods in which the prior structure is manually designed, mathematically explicit, and embedded directly into the optimization problem or inference model. Within this broad family, the reviewed literature reveals at least five important subgroups.

The first subgroup is Bayesian and statistical priors. Early wavelet-domain Bayesian fusion modeled the relationship between HSI and MSI using probability distributions over transformed coefficients and was particularly attentive to noise robustness and spectral preservation [2]. Bayesian double matrix factorization extended this tradition by combining latent factorization with probabilistic parameter estimation and uncertainty-aware inference [4]. These studies illustrate the classical advantage of shallow priors: they explain what is assumed, why it is assumed, and how the assumption affects reconstruction.

DOI: 10.55214/2576-8484.v10i6.13148

The second subgroup is matrix-factorization and spectral-unmixing priors. CNMF remains one of the most influential foundations in the field because it explicitly couples hyperspectral and multispectral observations via a shared endmember-abundance structure within a linear mixture model [3]. Later variants extended this line in multiple directions: linear-quadratic mixing to reflect nonlinear interactions, convex CNMF to stabilize optimization, and blind or semi-blind factorization to relax the need for known sensor responses [17, 26, 27]. The key conceptual contribution of this subgroup is that it frames fusion as a structured latent-variable problem grounded in a physical spectral interpretation.

The third subgroup is sparse and low-rank priors. These methods assume that the target image or its transformed representation is sparse, low-dimensional, or both. The intuition is that high-dimensional hyperspectral data often lie near lower-dimensional manifolds or subspaces. At the same time, local patches or coefficients can be represented efficiently in dictionaries or structured-sparse bases. Sparse and low-rank formulations became especially attractive because they balance expressiveness with computational tractability. They also preserve a desirable property of shallow priors: the reconstruction mechanism remains interpretable in terms of coefficients, subspaces, or rank structure.

The fourth subgroup is tensor-based priors. Tensor methods extend matrix modeling by treating hyperspectral cubes as higher-order objects whose spectral-spatial organization should be preserved rather than flattened. Graph-regularized low-rank tensor decomposition, nonlocal sparse tensor factorization, nonlocal coupled tensor CP decomposition, and low-rank Hankel tensor representation all exemplify this move from matrix to tensor reasoning [7, 8, 16, 28]. Their importance lies in recognizing that fusion depends not only on spectral subspaces but also on multidimensional dependencies and self-similarity. Tensor priors, therefore, represent an internal evolution within shallow prior modeling: they remain handcrafted and optimization-based, but are structurally richer than early factorization methods.

The fifth subgroup includes transform-domain and variational regularization methods, such as wavelet, curvelet, Poissonian, and spatial enhancement models [2, 5, 24]. These methods are often less dominant in current frontier literature, but they remain relevant because they foreground physically grounded degradations, noise models, and explicit regularization. In many respects, they define the benchmark against which newer deep methods must justify their claims to robustness.

Taken together, these shallow families show that "shallow" does not mean simplistic. On the contrary, the shallow literature evolved substantially from early statistical priors to sophisticated tensor, graph-regularized, and variability-aware models. What unites them is not low methodological sophistication, but the fact that their prior assumptions are explicit, handcrafted, and interpretable.

3.2.2. Deep Prior Modeling

Deep prior modeling refers to methods in which the regularizing structure is learned implicitly or semi-implicitly through neural architectures rather than being specified fully in closed form. In the reviewed literature, deep prior modeling unfolds through several stages.

The first stage is CNN-based fusion. Deep Hyperspectral Image Sharpening is a canonical transition point because it demonstrates that residual learning and convolutional feature extraction can learn a mapping from paired HSI/MSI observations to a fused output with strong spatial detail recovery [6]. This stage is characterized by encoder-decoder architectures, residual blocks, local receptive fields, and supervised training under assumptions of synthetic degradation. Deep priors here are implicit: the network's weights encode the regularization, but it is not written as a transparent mathematical penalty.

The second stage is architectural specialization. Later CNN methods no longer treat the network as a generic predictor but tailor architecture to fusion-specific needs, such as detail injection, modality-specific streams, coupled branches, multi-level fusion, and local-awareness mechanisms [9, 11, 12, 29]. Unsupervised variational and self-supervised formulations also emerge during this period, reflecting an important shift away from reliance on fully paired training labels [11]. These works indicate that deep

DOI: 10.55214/2576-8484.v10i6.13148

prior modeling in HS-MS fusion is not monolithic; it already encompasses supervised, unsupervised, and self-supervised branches, each with different assumptions about data availability and reconstruction consistency.

The third stage is adversarial and generative modeling. GAN-based methods introduce adversarial objectives to improve texture realism and spatial sharpness, while physics-based GANs attempt to temper purely generative freedom with sensor-consistency constraints [12, 19]. Adversarial models reveal both the promise and the risk of deep priors. They can produce visually and metrically strong results, but they also heighten concerns about spectral hallucination and interpretability.

The fourth stage is Transformer and sequence-style modeling. Recent studies using Swin-like modules, sparse mix-attention, or hierarchical spectral-spatial Transformers demonstrate the field's attempt to capture longer-range dependencies and more global contextual interactions than classical CNNs allow [14, 15, 25]. These methods are especially important because they mark an evolution in what the deep prior is expected to model. Instead of only local texture injection, recent networks aim to learn cross-scale spectral-spatial structure, modality interaction, and global dependency. This significantly broadens the representational role of deep priors.

3.2.3. Hybrid and Model-Driven Deep Prior Modeling

The third major category, hybrid/model-driven deep prior modeling, is the clearest sign of the field's current direction. These methods integrate explicit shallow structures into trainable deep architectures. Deep unfolding is the most visible example: the iterative logic of optimization is unrolled into network stages, allowing the model to retain interpretable computational structure while learning parameters from data [13]. Physics-guided GANs, matrix-decomposition-informed networks, and other architecture families that insert degradation consistency or factorization logic into neural models belong in the same broad family [18, 19].

This hybridization matters for two reasons. First, it shows that deep methods increasingly require explicit structural anchors to remain reliable as benchmark assumptions weaken. Second, it reveals that shallow priors are not disappearing; they are being absorbed, translated, and repurposed within deep models. From a taxonomy perspective, hybrid approaches are not transitional anomalies. They are a distinct paradigm, and their rise is one of the review's most important findings.

3.2.4. Evolutionary Trajectory

Across the 62 studies, the field's evolution can be summarized as a three-step trajectory. The first phase is the era of explicit handcrafted priors, dominated by Bayesian, matrix-factorization, sparse, low-rank, and variational methods [2-5]. The second phase is the rise of data-driven implicit priors, beginning with CNN-based hyperspectral sharpening and expanding to unsupervised variational, adversarial, and attention-based fusion [6, 11, 12]. The third phase is the current emergence of physics-guided and model-driven hybrids that seek to reconcile the representational power of deep learning with the interpretability and physical grounding of shallow models [13, 18, 19].

This evolutionary trajectory is not linear in the sense that shallow methods become obsolete once deep methods appear. Rather, the literature suggests cyclical return and methodological inheritance. Problems discovered in deep models, such as spectral inconsistency, limited transferability, and sensor mismatch, drive renewed interest in explicit priors. Conversely, the limitations of handcrafted models in highly nonlinear settings motivate the need for learning-based flexibility. The result is a field that has moved from opposition between paradigms toward increasingly strategic integration.

3.3. Findings for RQ2: Comparative Advantages, Limitations, and Integration of Shallow and Deep Priors

Table 2.
Comparative synthesis of shallow, deep, and hybrid priors.

Criterion	Shallow prior	Deep prior	Hybrid / model-driven deep
Prior formulation	Explicit, handcrafted constraints and observation models	Implicit or semi-implicit learned regularization	Partly explicit, partly learned
Training data requirement	Low to moderate	Moderate to high	Moderate
Interpretability	High	Low	Medium to high
Spectral fidelity	Often strong if assumptions hold	Can be excellent, but more variable	Potentially strongest when physics is retained
Generalization risk	Model mismatch	Domain shift/dataset bias	Reduced but not eliminated
Computational profile	Iterative optimization; slower inference possible	High training cost, often fast inference	Moderate to high in both training and design complexity
Best-fit scenarios	Limited data, high interpretability demand, and strict physical consistency	Abundant data, complex nonlinear structure, high-detail reconstruction	Operational settings needing both performance and trustworthiness

RQ2 asked how shallow and deep priors compare and how the literature integrates them. The answer from the final pool is clear: the shallow-versus-deep contrast remains analytically useful, but the highest-value recent work is defined by attempts to overcome the limitations of each paradigm through hybridization.

3.3.1. Strengths of Shallow Priors

The most consistently reported strength of shallow priors is interpretability. In Bayesian, unmixing, low-rank, and tensor models, the reconstruction can usually be explained in terms of physically meaningful latent variables, regularization terms, or clearly stated assumptions. This matters in remote sensing because the fused product is often used for scientific or operational interpretation, not just for downstream machine learning. When spectral fidelity is critical, an interpretable model offers not only transparency but diagnostic power: one can understand whether an error arises from mixing assumptions, degradation mismatch, or regularization bias [2-5, 7, 8].

A second strength is robustness under limited data. Many shallow models do not require large paired training corpora. This gives them a practical advantage in settings where matched HR-HSI ground truth is unavailable or synthetic supervision is unconvincing. A third strength is their close alignment with sensor physics. Factorization methods that model endmembers and abundances, or blind-fusion methods that estimate SRF/PSF, often handle inverse-problem structure more explicitly than purely data-driven alternatives [3, 17]. A fourth strength is controllability: optimization objectives can be modified to include domain constraints, noise models, or application-specific regularizers without retraining an entire system.

3.3.2. Limitations of Shallow Priors

The literature is equally clear about the limitations of shallow priors. The first is representational rigidity. Many handcrafted priors rely on assumptions such as linear mixing, global low rank, or fixed local similarity. These can be too restrictive for real scenes containing complex boundaries, nonlinear interactions, or heterogeneous material variability. The second limitation is sensitivity to model mismatch. If the assumed subspace, rank, degradation operator, or mixture law deviates from the real imaging process, the method may preserve spectral consistency only by sacrificing spatial detail or vice versa. The third limitation is scalability. Sophisticated tensor- and optimization-based methods can be computationally intensive, especially when blind or semi-blind estimation is involved. In short, shallow priors are often trustworthy when their assumptions hold, but they can degrade sharply when those assumptions are violated.

3.3.3. Strengths of Deep Priors

Deep priors are consistently valued for expressive power. CNNs, GANs, and Transformers can model highly nonlinear spectral-spatial relations and can inject high-frequency detail more effectively than many handcrafted formulations [6, 10-15, 25]. This advantage is particularly evident on benchmark datasets where degradation operators are known or synthetically controlled. Deep methods are also adaptable in a design sense: attention modules, multibranch networks, modality-specific encoders, and residual learning enable targeting of local texture, global context, and cross-modal interaction within a single trainable system.

Another strength of deep priors is their capacity to collapse multiple stages of inference into end-to-end learning. Instead of estimating subspaces, abundances, blur, and reconstruction separately, a deep architecture can jointly optimize these interactions during training. This creates strong empirical performance and often lower inference-time latency after training. Transformer-based designs extend this benefit by modeling longer-range contextual relationships that CNNs may miss [14, 15, 25]. The best deep models, therefore, do not merely emulate shallow priors; they learn representations that would be difficult to specify manually.

3.3.4. Limitations of Deep Priors

The most common limitation of deep priors is weak interpretability. Even when a network performs well, it is often difficult to explain why a specific fusion decision was made, how spectral distortions were introduced, or how sensor mismatches are handled internally. A second limitation is dependence on training data and training protocol. Many deep methods are evaluated under synthetic supervision that assumes known degradations and accurate alignment, but these assumptions can be much cleaner than real acquisition conditions. A third limitation is domain shift. Deep networks trained on one dataset, sensor pair, or degradation model may generalize poorly to another. This is especially problematic in operational remote sensing, where acquisition conditions vary widely.

A fourth limitation concerns spectral credibility. Spatial sharpness gained by a powerful network may come at the cost of subtle spectral distortion, band inconsistency, or hallucinated detail. In applications such as material identification, these errors are more damaging than small losses in visual sharpness. The literature, therefore, repeatedly warns against equating higher benchmark scores with genuine operational trustworthiness.

3.3.5. Integration Mechanisms

The review found three major mechanisms by which the literature integrates shallow and deep priors. The first is structural integration, in which explicit optimization steps or decompositions are embedded into neural architectures. Deep unfolding is the clearest example [13]. The second is physical integration, in which neural networks are trained or constrained using degradation consistency, sensor models, or reconstruction equations, as seen in physics-guided adversarial and transformer-based work [18, 19]. The third is representational integration, in which concepts such as unmixing, matrix decomposition, sparse coding, or low-rank structure are translated into deep modules or auxiliary losses even if the final architecture remains recognizably neural.

These integration mechanisms are not peripheral innovations. They address the central weakness of a strict shallow/deep dichotomy. Shallow priors alone may be too rigid; deep priors alone may be too opaque and dataset-dependent. Hybrid methods attempt to combine the explicit inductive bias of the former with the nonlinear approximation power of the latter. The literature increasingly treats this combination not as an optional enhancement, but as the natural next stage in the field's evolution.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 6: 355-373, 2026
DOI: 10.55214/2576-8484.v10i6.13148

3.4. Findings for RQ3: Critical Bottlenecks and Promising Future Trajectories

Table 3.
Major bottlenecks and promising trajectories in HS-MS prior modeling.

Bottleneck	Why it persists	Most affected paradigms	Promising response
Spectral fidelity	Spatial enhancement can distort fine spectral signatures, especially near edges and mixed pixels.	All paradigms, especially unconstrained deep models	Spectral-aware losses; physics-guided supervision; variability-aware priors
Generalization	Benchmark-trained models and fixed assumptions transfer poorly across sensors/scenes.	Deep and hybrid models, but also shallow models under mismatch	Cross-sensor evaluation; self-supervision; robust domain adaptation
Sensor uncertainty	PSF/SRF is often unknown or inaccurate in real deployments	Shallow blind models and deep models trained on fixed degradations	Blind or semi-blind fusion; joint degradation estimation
Spectral variability	Endmembers and signatures vary over time, season, and local context	Classical factorization and fixed-subspace models	Variability-aware factorization and tensor models; adaptive hybrid priors
Computational deployability	Sophisticated optimization and large networks increase cost	Tensor methods and Transformer-style models	Lightweight architectures; efficient attention; scalable optimization
Uncertainty awareness	Most studies output point estimates without confidence information	All paradigms	Confidence maps; probabilistic outputs; uncertainty-aware evaluation

RQ3 asked what bottlenecks continue to constrain prior modeling in HS-MS fusion and what trajectories appear most promising. The final pool shows that the field's bottlenecks are highly concentrated rather than diffuse. Two problems dominate: spectral fidelity and generalization.

3.4.1. Spectral Fidelity as a Persistent Bottleneck

Spectral fidelity remains the most fundamental bottleneck because the entire point of hyperspectral fusion is to preserve fine spectral information while adding spatial detail. Many studies explicitly or implicitly acknowledge that improving spatial sharpness can introduce spectral distortion, especially around edges, mixed pixels, or high-frequency structures. This problem is not confined to deep methods, but it becomes especially acute when learning-based models prioritize visual sharpness or adversarial realism [12, 19]. Transformer and attention-based methods may improve long-range dependency modeling, yet they do not automatically guarantee spectral faithfulness if the loss functions and degradation constraints are misaligned.

Shallow models often appear stronger on this dimension because their assumptions are tied directly to spectral mixing, abundance structure, or probabilistic observation consistency [3-5]. However, even shallow methods suffer when the assumed mixing model is oversimplified or when spectral variability is ignored. This explains the emergence of variability-aware factorization and tensor methods that explicitly account for seasonal or inter-image spectral change [9, 16, 28]. The literature, therefore, treats spectral fidelity not as a solved issue but as the core criterion against which any new prior model must still be judged.

3.4.2. Generalization and Real-World Robustness

The second dominant bottleneck is generalization. Many included studies achieve promising results under synthetic settings where the PSF and SRF are known, alignment is ideal, and training/test distributions are closely matched. Yet real remote sensing deployments rarely satisfy these conditions. Sensor responses may be unknown, approximated, or time-varying. Scenes may differ by season, atmosphere, land-cover composition, or acquisition angle. Degradation may be nonuniform. As a result, a model that performs impressively on a benchmark can fail to transfer to real data.

This concern appears in both shallow and deep literatures, but in different ways. In shallow work, the issue appears as model mismatch: assumptions about mixing, rank, or degradation may not hold. In deep work, the issue appears as domain shift: the learned mapping may not survive transfer across sensors or scenes. Blind and semi-blind fusion methods, variability-aware factorization, self-supervised learning, and physics-guided training all emerge partly in response to this problem [9, 11, 17, 18].

3.4.3. Sensor Uncertainty and Degradation Mismatch

A related bottleneck is uncertainty in sensor characterization. Many fusion methods assume known PSF and SRF, but this information is not always available or accurate. Blind matrix-factorization methods and Bayesian formulations attempt to estimate some of these operators jointly with the latent image [4, 17]. Nonetheless, the literature indicates that robust blind fusion remains difficult, especially when combined with spectral variability or strong nonlinear structure. Sensor mismatch is therefore not merely a nuisance parameter issue; it is a foundational challenge for prior modeling because the prior and the degradation model interact directly during inversion.

3.4.4. Computational Complexity and Practical Deployability

Another bottleneck concerns computational efficiency and deployability. Sophisticated tensor decompositions and blind optimization can be computationally expensive, while deep Transformer-style models may require significant training resources and large memory footprints. The literature increasingly recognizes that future methods must not only improve accuracy but also remain deployable under practical constraints, including limited computational budgets and incomplete training data. This is one reason why lightweight networks, efficient attention, patch-wise learning, and compact hybrids are receiving growing attention [10, 14, 15].

3.4.5. Promising Research Trajectories

The review identifies six especially promising trajectories. The first is model-driven deep unfolding, which offers an interpretable route to learning-enhanced optimization [13]. The second is physics-guided deep learning, where degradation consistency and sensor structure are retained inside trainable models [18, 19]. The third is unsupervised and self-supervised fusion, which addresses the mismatch between benchmark supervision and real-world data availability [11]. The fourth is blind or semi-blind fusion capable of estimating unknown sensor responses jointly with reconstruction [17]. The fifth is generalizable cross-sensor learning, where models are designed or trained explicitly for transfer rather than only for in-domain performance. The sixth is uncertainty-aware fusion, an emerging but still underdeveloped direction that would quantify confidence in reconstructed spectra and spatial detail.

Taken together, these trajectories suggest that the future of HS-MS fusion is less about inventing ever more elaborate black-box architectures and more about making learned priors trustworthy under real imaging conditions. This is a crucial shift in emphasis. The literature is increasingly aware that the next breakthrough will likely come not from marginal gains on synthetic benchmarks, but from models that remain stable when physical assumptions are incomplete and distributions shift.

3.4.6. Additional Discussion on Taxonomy Usefulness

A further implication of the taxonomy developed in this review is methodological clarity for future authors. A recurring weakness in the current literature is that papers often position themselves only against a narrow subset of recent competitors while leaving their deeper prior assumptions underspecified. For example, a method may be introduced as a "new deep fusion network" even when its novelty lies mainly in replacing one attention block with another, without explaining whether the model is fundamentally attempting to learn a degradation-consistent inverse mapping, inject spatial detail, estimate latent abundance-like structure, or approximate an unfolding procedure. By organizing the literature in terms of prior families rather than superficial architecture labels, the taxonomy proposed here offers a more diagnostic vocabulary. It helps distinguish whether a method is truly changing the prior, merely changing the optimizer, or only modifying the feature extractor. This is important for cumulative research because conceptual progress cannot be evaluated if contributions are described only at the implementation level.

The taxonomy is also useful for benchmarking. Comparing a Bayesian factorization method, an adversarial network, and a Transformer only by aggregate metrics can obscure what each class is trying to optimize. A shallow method may intentionally sacrifice local sharpness to preserve spectral plausibility; an adversarial method may privilege texture realism; a hybrid unfolding model may prioritize reconstruction consistency under known degradation operators. Once methods are grouped by prior logic, their strengths and weaknesses become more interpretable. This suggests that future benchmark design should not compare all methods on a single undifferentiated leaderboard. Instead, evaluation could be stratified by scenario: known versus unknown degradations, abundance-sensitive versus texture-sensitive tasks, and in-domain versus cross-sensor transfer settings. Such scenario-aware comparison would better reflect the decision structure faced by practitioners.

Finally, the taxonomy clarifies the relationship between engineering novelty and scientific utility. In remote sensing, the value of a fused image depends on whether it preserves information needed for downstream interpretation, not only whether it improves image quality indices. A model can be technically innovative while remaining scientifically risky if it introduces spectral artifacts that are hard to diagnose. Conversely, a method may appear conservative in architectural terms yet remain highly valuable because it preserves trustworthy spectra under challenging conditions. This distinction is especially important as the field moves toward increasingly complex learned models. The review therefore recommends that future studies state explicitly what kind of prior they adopt, what inverse-problem assumptions they relax, and what types of downstream use the resulting fused image can legitimately support.

4. Discussion

The findings reveal a field in methodological transition. On the surface, the literature might appear to tell a familiar story in which shallow model-based algorithms dominate an early era, are overtaken by deep learning, and are now being displaced again by larger Transformer-style architectures. That interpretation is too simplistic. What the evidence actually shows is a progressive renegotiation of what a "useful prior" must look like in HS-MS fusion. Early work prioritized explicitness and physical legibility. Deep work prioritized representational flexibility and end-to-end learning. Recent hybrid work suggests that the field now values priors that are simultaneously expressive, physically anchored, and transferable. The significance of this evolution reaches beyond HS-MS fusion itself; it reflects a broader shift in computational imaging from handcrafted regularization to learned inverse modeling and then toward structured hybrids that recover the strengths of both.

One implication is conceptual. In much of the earlier literature, priors were viewed as constraints added to an optimization problem. In more recent learning-based work, priors are often treated as emergent properties of data-trained architectures. The review shows that these are not mutually exclusive conceptions. A deep network can be interpreted as an implicit prior, but implicitness alone does not guarantee suitability for remote sensing. The operational demands of the domain - spectral credibility, physical plausibility, and cross-sensor robustness, continually pressure the field to reintroduce explicit structure. This is why model-driven deep learning has become more than a technical niche. It is an epistemic compromise: it acknowledges that purely handcrafted priors are insufficiently expressive and that purely black-box priors are insufficiently trustworthy.

A second implication concerns evaluation culture. Much of the HS-MS fusion literature still relies on benchmark-centric comparison, often using synthetic degradation processes and established metrics such as PSNR, SAM, ERGAS, and SSIM. These metrics are useful, but the review suggests they are not enough. A model can rank highly under synthetic protocols and still be fragile when the assumed PSF or SRF is inaccurate, when spatial misregistration exists, or when spectral variability departs from training conditions. The deep-learning literature in particular risks overstating progress when benchmark conditions are treated as surrogates for operational reality. The shallow literature is not immune to this issue either; optimization methods may also be validated under idealized assumptions. However, deep models are especially vulnerable because their performance depends not only on physical modeling but on the representativeness of the training distribution. The field therefore needs more evaluation protocols that test transfer, robustness to unknown degradations, and preservation of physically meaningful spectra.

This observation helps explain why spectral fidelity remains the domain's most persistent bottleneck. In many image-fusion problems, visually plausible reconstruction may be a sufficient goal. In HS-MS fusion, it is not. The output is often used as a scientific measurement proxy, not only as a display image. Subtle spectral errors may compromise material discrimination, unmixing, change detection, or downstream environmental inference. The review therefore reinforces an important methodological principle: in HS-MS fusion, spatial enhancement is not an autonomous objective; it is subordinate to preserving spectrally meaningful information. This principle is easier to articulate in shallow priors than in deep ones, which partly explains the continued relevance of explicit model-based methods despite the empirical gains of deep learning.

The field's movement toward variability-aware methods is also revealing. Classical factorization and subspace models often assumed relatively stable spectral signatures, but later work had to address seasonal variability, inter-image mismatch, and class-wise spectral spread [10, 16]. This suggests that one hidden limitation of early prior design was not simply "lack of complexity" but insufficient realism about the scene. Real scenes are not static collections of fixed endmembers. They contain changing illumination, seasonal differences, and subpixel heterogeneity. Deep methods can absorb some of this complexity implicitly, but they do not necessarily understand it in a transferable way. Variability-aware shallow and hybrid methods may therefore be especially important because they inject realism into the prior without abandoning interpretability.

Another major discussion point concerns the relation between blind fusion and prior modeling. Blind or semi-blind methods show that the inverse problem is not only about reconstructing the latent image but also about inferring the observation process itself [4, 17]. This is highly consequential in practice because full sensor characterization is often unavailable. Yet blind fusion remains underdeveloped relative to supervised deep reconstruction. The likely reason is that blind estimation multiplies uncertainty: the method must jointly infer both latent content and degradation operators. This is difficult for shallow and deep models alike. Still, the review suggests that blind modeling may become one of the most strategically important directions in the next generation of HS-MS fusion, especially for deployment across heterogeneous or partially documented sensors.

The emergence of Transformers deserves separate attention. Transformer-based methods appear in the literature as responses to limitations of CNN locality and as attempts to model long-range spectral-spatial interactions [14, 15, 25]. Their rise is meaningful, but the review also suggests caution. The main question is not whether Transformers can improve scores - they often do - but whether they solve the fundamental problems that matter most in HS-MS fusion. Global attention can help capture context, yet spectral fidelity, degradation consistency, and cross-sensor generalization still depend on the integration of physical constraints and domain-relevant inductive bias. In that sense, Transformer adoption may improve representational scope without resolving the field's deepest validation problem. The same caution applies to diffusion-like and other recent generative models: expressiveness is valuable, but uncontrolled expressiveness can be dangerous when the reconstructed output is interpreted as a physically meaningful signal.

The strong showing of hybrid methods in the final pool should therefore be interpreted as more than a trend. It may represent the field's most credible answer to the twin bottlenecks of spectral fidelity and generalization. Deep unfolding is especially revealing because it does not merely combine shallow and deep ideas superficially. It preserves an optimization logic that experts can interpret, while still allowing data-driven parameter learning [13]. Physics-guided adversarial and transformer-based methods follow the same basic principle from a different direction: they maintain the representational benefits of deep networks but tether them to reconstruction equations, consistency losses, or known sensor interactions [18, 19]. The review suggests that this is not just good engineering practice; it is a methodological necessity in a domain where purely learned priors remain difficult to trust outside controlled conditions.

From a practical standpoint, the review implies that algorithm choice should depend on deployment constraints rather than on benchmark rankings alone. If the use case requires strict interpretability, strong spectral credibility, limited training data, or operation under incomplete supervision, shallow or hybrid methods may be preferable. If the use case provides abundant training data and the main challenge is recovering complex spatial detail under relatively stable degradation assumptions, deep priors may be more attractive. In other words, the field would benefit from replacing the question "Which method is best?" with the question "Which prior family is most appropriate under which observational regime?" That shift would make future comparison studies more meaningful and less dominated by leaderboard logic.

The review also highlights a gap in uncertainty modeling. Many studies report accuracy metrics but do not quantify confidence in the fused output. This is problematic because decision-makers may assume that the reconstructed image is equally reliable across all bands and locations, even though errors may concentrate in regions of strong mixing, sharp edges, or poorly represented materials. Uncertainty-aware fusion could help address this gap by estimating confidence intervals, spectral credibility maps, or degradation-sensitive risk measures. Such tools would be especially valuable in scientific monitoring contexts where false certainty is costly. Although only a limited number of reviewed studies moved in this direction, the need becomes obvious once the field is viewed through the lens of real-world reliability rather than benchmark optimization.

Methodologically, the review itself also invites reflection. The final corpus of 62 studies was sufficient for analytical synthesis, but the high number of unretrieved reports means that coverage was necessarily shaped by access conditions. This limitation should not invalidate the findings, but it should encourage care in interpreting them. The review is strongest in identifying broad methodological patterns and conceptual tensions, not in claiming exhaustive quantitative dominance of one method family over all others. Future reviews with expanded database access could refine the corpus further and perhaps test some of the trends observed here statistically, such as the growth rate of hybrid methods or the changing emphasis on blind fusion and self-supervision.

Overall, the discussion points to a central conclusion: the history of HS-MS fusion before modeling is not a march from shallow to deep, but a dialectic between explicit structure and learned flexibility. The field now appears to recognize that neither pole alone is sufficient. The next stage of progress will likely come from priors that are physically faithful enough to be trusted, flexible enough to model complex remote sensing variability, and generalizable enough to survive deployment beyond synthetic benchmarks.

4.1. Review Limitations and Future Review Work

The present review also suggests priorities for future evidence synthesis. First, future reviews should attempt broader database triangulation and, where possible, more complete full-text retrieval so that access-related attrition can be reduced. Second, the field would benefit from supplementary datasets that code not only method families but also degradation assumptions, training regimes, dataset types, and transfer settings. Third, as hybrid and self-supervised methods continue to grow, future reviews may need to separate "architecture novelty" from "prior novelty" more systematically. These steps would allow the literature to move from broad conceptual mapping toward more granular evidence statements, such as which classes of priors are most robust under blind settings or which evaluation designs are most likely to overestimate practical utility. In this sense, the present study should be seen as both a synthesis of the current field and a call for more rigorous meta-research about how HS-MS fusion is validated.

5. Conclusion

This article reviewed 62 journal studies on shallow and deep prior modeling for hyperspectral-multispectral image fusion using a PRISMA-guided screening process across Scopus, Taylor & Francis Online, and Wiley Online Library. The review was organized around three research questions concerning taxonomy and evolution, comparative advantages and limitations, and the field's critical bottlenecks and future trajectories.

Three main conclusions emerge. First, the literature is best understood through a three-part taxonomy rather than a simple shallow-versus-deep split. Shallow prior modeling includes Bayesian/statistical, matrix-factorization, sparse/low-rank, tensor, and variational families. Deep prior modeling includes CNN, adversarial, attention-based, and Transformer-style architectures. A third category - hybrid or model-driven deep prior modeling - has become increasingly important because it embeds explicit physical or optimization structure into trainable networks. Second, the field has evolved from explicit handcrafted priors to implicit learned priors and then toward hybridization. This evolution reflects not only technological progress but a deeper methodological negotiation over how much of the inverse problem should be specified analytically and how much should be learned from data. Third, the dominant unresolved bottlenecks remain spectral fidelity and generalization. These are amplified by sensor uncertainty, spectral variability, and the gap between synthetic evaluation and real-scene deployment.

The review argues that future progress in HS-MS fusion will depend less on ever larger black-box models and more on trustworthy prior design. The most promising research directions are model-driven deep unfolding, physics-guided learning, unsupervised and self-supervised fusion, blind or semi-blind degradation estimation, cross-sensor generalization, and uncertainty-aware reconstruction. In short, the next generation of HS-MS fusion methods should be judged not only by reconstruction scores but by whether they can produce spectrally credible, interpretable, and transferable fused imagery under real-world remote sensing conditions.

Taken together, these observations indicate that the future value of HS-MS fusion research will be determined not only by algorithmic sophistication but by methodological honesty about assumptions, uncertainty, and deployment conditions.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] A. Mookambiga and V. Gomathi, "Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery," *Multidimensional Systems and Signal Processing*, vol. 27, pp. 863-889, 2016. https://doi.org/10.1007/s11045-016-0415-2

[2] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834-3843, 2009. https://doi.org/10.1109/TGRS.2009.2017737
[3] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528-537, 2011. https://doi.org/10.1109/TGRS.2011.2161320
[4] B. Lin, X. Tao, M. Xu, L. Dong, and J. Lu, "Bayesian hyperspectral and multispectral image fusions via double matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5666-5678, 2017. https://doi.org/10.1109/TGRS.2017.2711640
[5] C. Zou and Y. Xia, "Poissonian hyperspectral image superresolution using alternating direction optimization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4464-4479, 2016. https://doi.org/10.1109/JSTARS.2016.2585158
[6] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5345-5355, 2018. https://doi.org/10.1109/TNNLS.2018.2798162
[7] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Transactions on Cybernetics*, vol. 50, no. 10, pp. 4469-4480, 2019. https://doi.org/10.1109/TCYB.2019.2951572
[8] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1030-1040, 2018. https://doi.org/10.1109/JSTARS.2017.2785411
[9] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Transactions on Image Processing*, vol. 29, pp. 116-127, 2019. https://doi.org/10.1109/TIP.2019.2928895
[10] Z. Wang, B. Chen, R. Lu, H. Zhang, H. Liu, and P. K. Varshney, "FusionNet: An unsupervised convolutional variational network for hyperspectral and multispectral image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 7565-7577, 2020. https://doi.org/10.1109/TIP.2020.3004261
[11] X. Luo, J. Yin, X. Luo, and X. Jia, "A novel adversarial based hyperspectral and multispectral image fusion," *Remote Sensing*, vol. 11, no. 5, p. 492, 2019. https://doi.org/10.3390/rs11050492
[12] X. Lu, D. Yang, F. Jia, and Y. Zhao, "Coupled convolutional neural network-based detail injection method for hyperspectral and multispectral image fusion," *Applied Sciences*, vol. 11, no. 1, p. 288, 2020. https://doi.org/10.3390/app11010288
[13] B. Zhang, X. Cao, and D. Meng, "A deep unfolding network for multispectral and hyperspectral image fusion," *Remote Sensing*, vol. 16, no. 21, p. 3979, 2024. https://doi.org/10.3390/rs16213979
[14] S. Yu, X. Zhang, and H. Song, "Sparse mix-attention transformer for multispectral image and hyperspectral image fusion," *Remote Sensing*, vol. 16, no. 1, p. 144, 2023. https://doi.org/10.3390/rs16010144
[15] T. Zhu, Q. Liu, and L. Zhang, "Hierarchical spectral–spatial transformer for hyperspectral and multispectral image fusion," *Remote Sensing*, vol. 16, no. 22, p. 4127, 2024. https://doi.org/10.3390/rs16224127
[16] S. Guo, X. a. Chen, H. Jia, Z. Han, Z. Duan, and Y. Tang, "Fusing hyperspectral and multispectral images via low-rank hankel tensor representation," *Remote Sensing*, vol. 14, no. 18, p. 4470, 2022. https://doi.org/10.3390/rs14184470
[17] J. Long and Y. Peng, "Blind fusion of hyperspectral multispectral images based on matrix factorization," *Remote Sensing*, vol. 13, no. 21, p. 4219, 2021. https://doi.org/10.3390/rs13214219
[18] J. Xiao, J. Li, Q. Yuan, M. Jiang, and L. Zhang, "Physics-based GAN with iterative refinement unit for hyperspectral and multispectral image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6827-6841, 2021. https://doi.org/10.1109/JSTARS.2021.3075727
[19] C.-H. Lin, F. Ma, C.-Y. Chi, and C.-H. Hsieh, "A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1652-1667, 2017. https://doi.org/10.1109/TGRS.2017.2766080
[20] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *bmj*, vol. 372, p. n71, 2021. https://doi.org/10.1136/bmj.n71
[21] Elsevier, "How can I best use the Advanced search? - Scopus," 2024. https://service.elsevier.com/app/answers/detail/a_id/11365/supporthub/scopus/. [Accessed Apr. 14, 2026]
[22] Wiley Online Library, "Using advanced search," 2020. https://wolsupport.wiley.com/s/article/using-advanced-search. [Accessed Apr. 14, 2026]
[23] Taylor & Francis, "Searching our site," n.d. https://help.tandfonline.com/s/article/Searching-our-site. [Accessed Apr. 14, 2026]
[24] F. Z. Benhalouche, M. S. Karoui, Y. Deville, and A. Ouamri, "Hyperspectral and multispectral data fusion based on linear-quadratic nonnegative matrix factorization," *Journal of Applied Remote Sensing*, vol. 11, no. 2, pp. 025008-025008, 2017. https://doi.org/10.1117/1.JRS.11.025008

[25] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. C. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 702-717, 2021. https://doi.org/10.1109/JSTSP.2021.3054338

[26] Y. Wang, Q. Zhu, Y. Shi, M. Song, and C. Yu, "A spatial-enhanced LSE-SFIM algorithm for hyperspectral and multispectral images fusion," *Remote Sensing*, vol. 13, no. 24, p. 4967, 2021. https://doi.org/10.3390/rs13244967

[27] K. Priya and K. Rajkumar, "Multiplicative iterative nonlinear constrained coupled non-negative matrix factorization (MINC-CNMF) for hyperspectral and multispectral image fusion," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021. https://doi.org/10.14569/IJACSA.2021.0120660

[28] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 348-362, 2019. https://doi.org/10.1109/TGRS.2019.2936486

[29] K. Zheng *et al.*, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2487-2502, 2020. https://doi.org/10.1109/TGRS.2020.3006534