



# Machine Learning Across the Amorphous Solid Dispersion Design Space for Predicting Miscibility, Stability, Drug Release, and Formulation Failure

Michael Roberts<sup>1\*</sup>, Sarah Thompson<sup>2</sup>, James Anderson<sup>3</sup>, Robert Smith<sup>4</sup>

<sup>1</sup>Department of ML for Amorphous Solid Dispersions, Faculty of Pharmacy, University of Glasgow, Glasgow, United Kingdom.

<sup>2</sup>Department of Miscibility and Stability Prediction, Faculty of Pharmacy, University of Auckland, Auckland, New Zealand.

<sup>3</sup>Department of Drug Release and Formulation Performance, Faculty of Pharmacy, University of Manchester, Manchester, United Kingdom.

<sup>4</sup>Department of Formulation Failure Prediction, Faculty of Pharmacy, University of Edinburgh, Edinburgh, United Kingdom.

## ABSTRACT

Amorphous solid dispersions (ASDs) are frequently treated in predictive modeling as formulations whose success can be represented by a single endpoint, such as miscibility, glass-transition temperature, physical stability, or dissolution. This abstraction is increasingly inadequate because an ASD is a dynamic material system whose behavior changes with composition, molecular interactions, manufacturing history, temperature, humidity, aging, aqueous exposure, and the timescale on which failure is defined. This state-of-the-art review examines how machine learning can operate across that multidimensional design space while remaining grounded in the physicochemical evidence that defines meaningful prediction targets. Current studies demonstrate useful task-specific prediction of properties including polymer-dependent thermal behavior, drug-polymer compatibility, phase behavior, ASD formation, physical stability, and dissolution-related performance. However, these successes largely arise from distinct datasets, endpoint definitions, representations, and validation regimes. The central synthesis proposed here is therefore a failure-aware view of ASD modeling in which intermediate material states, manufacturing context, performance objectives, and failure outcomes remain analytically separable before being integrated. Such a framework requires models to distinguish thermodynamic compatibility from kinetic persistence, storage stability from aqueous release, process feasibility from long-term robustness, and internal predictive performance from transfer to new chemical, polymeric, manufacturing, or laboratory domains. Machine learning can strengthen ASD development when it organizes rather than erases these distinctions. Its principal current limitations are sparse negative evidence, inconsistent labels, restricted external validation, heterogeneous process histories, and incomplete representation of time- and condition-dependent formulation failure.

**Key Words:** Amorphous solid dispersion, Machine learning, Miscibility, Physical stability, Dissolution, Formulation failure

eIJPPR 2026; 16(4):40-49

**HOW TO CITE THIS ARTICLE:** Roberts M, Thompson S, Anderson J, Smith R. Machine Learning Across the Amorphous Solid Dispersion Design Space for Predicting Miscibility, Stability, Drug Release, and Formulation Failure. Int J Pharm Phytopharmacol Res. 2026;16(4):40-9. <https://doi.org/10.51847/uN3XDxLimx>

## INTRODUCTION

Amorphous solid dispersions are established formulation strategies for improving the performance of poorly water-

soluble drugs, but their successful implementation does not define a single physicochemical recipe. Approved products span different drug chemistries, polymer systems, drug loadings, and manufacturing approaches, illustrating that

**Corresponding author:** Michael Roberts

**Address:** Department of ML for Amorphous Solid Dispersions, Faculty of Pharmacy, University of Glasgow, Glasgow, United Kingdom.

**E-mail:** ✉ [michael.roberts@glasgow.ac.uk](mailto:michael.roberts@glasgow.ac.uk)

**Received:** 30 March 2026; **Revised:** 16 July 2026; **Accepted:** 19 July 2026

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



workable ASDs occupy multiple regions of formulation space rather than one universal composition–process optimum [1]. That heterogeneity matters computationally. A formulation cannot be represented adequately by molecular identity alone when its measurable state is also conditioned by processing, storage, and subsequent exposure to an aqueous environment.

Contemporary ASD development accordingly combines thermal and solid-state characterization, thermodynamic reasoning, dissolution analysis, molecular simulation, process understanding, and increasingly machine-learning methods [2]. Machine learning enters this landscape most usefully when its prediction targets correspond to experimentally meaningful formulation states. A model trained to classify ASD formation, for example, answers a different question from one predicting shelf stability, dissolution kinetics, or downstream absorption. Treating these outcomes as interchangeable forms of “success” risks generating high apparent predictive performance while weakening the scientific meaning of the prediction.

Polymer selection makes this problem particularly clear because polymers simultaneously influence molecular mixing, intermolecular interactions, molecular mobility, crystallization resistance, moisture response, and drug release [3]. A polymer that is favorable for one of these functions need not optimize the others. The central premise of this review is therefore that the ASD design problem should be treated as a dynamic, multi-state and multi-objective predictive problem. This formulation is proposed as an organizing synthesis rather than a validated universal modeling architecture. It requires distinguishing variables that describe the material, variables that describe its history and environment, intermediate physicochemical states, measured performance endpoints, and explicit modes of failure.

#### *Review scope and state-of-the-art approach*

This review critically evaluates machine learning in ASD development through the scientific structure of the formulation problem rather than through an algorithm-by-algorithm catalogue. The relevant evidence includes direct ASD machine-learning studies, experimental work that defines defensible prediction targets, thermodynamic and molecular studies that test candidate descriptors, and formulation studies that expose conditions under which apparently favorable material properties do not translate into stability or release. This scope reflects the contemporary movement toward combining computation with physicochemical characterization rather than treating predictive modeling as an independent replacement for formulation science [2].

The review therefore asks three related questions. First, what should be predicted? Miscibility, glass transition,

crystallization propensity, storage stability, dissolution, and formulation failure are related but non-equivalent outcomes. Second, what information should be represented? Molecular structure, polymer chemistry, composition, intermolecular interactions, processing history, environmental exposure, and analytical measurements may each carry task-dependent information. Third, what constitutes convincing validation? Random hold-out performance within one compiled dataset is not equivalent to transfer across unseen APIs, polymer families, manufacturing routes, storage regimes, or laboratories.

The analysis deliberately separates experimentally supported relationships from broader computational proposals. Where evidence shows that a variable is informative within a defined material or process domain, the claim is kept within that domain. Where this review integrates such findings into a broader modeling architecture, that integration is identified as proposed. The resulting perspective is intended to clarify where machine learning currently adds value, where mechanistic formulation knowledge constrains interpretation, and where apparently accurate models remain vulnerable to hidden formulation failure.

#### *Mapping the variables that define asd design space*

ASD design space is not simply the Cartesian product of drug identity and polymer identity. Manufacturing reviews show that hot-melt extrusion, spray drying, solvent-based routes, thermal history, shear, solvent removal, and post-processing conditions can generate materially different states from otherwise similar compositions [4]. Process history should therefore be represented as part of the formulation state rather than retained only as descriptive metadata. A model that ignores how an amorphous dispersion was created may inadvertently attribute process-conditioned behavior to molecular structure.

Intermediate properties can nevertheless be learnable within carefully defined domains. In a structured polymer-design study, formulation and polymer features were used to predict glass-transition behavior across a controlled library, demonstrating that material descriptors can encode useful structure–property relationships [5]. Such results support targeted machine learning but do not establish that the predicted intermediate property is equivalent to long-term ASD success. Thermal state is one layer of the design problem, not its endpoint.

Direct comparison of hot-melt extrusion and spray drying further illustrates why manufacturing route cannot be reduced to an incidental categorical label. Within the investigated system, preparation route influenced recrystallization, stability, and dissolution behavior [6]. The interpretation should remain system-specific, but the

modeling implication is broader: formulations with identical nominal drug, polymer, and loading may occupy different effective regions of design space because their processing histories generate different microstructures and kinetic states.

For the present review, a variable is therefore considered evidence-eligible for predictive modeling when it has a

defensible experimental, physicochemical, or formulation role in defining the state or performance being predicted.

**Table 1** organizes these variable classes while preserving the distinction between intrinsic composition, experimentally conditioned state, and contextual exposure.

**Table 1.** Evidence-eligibility map for variables entering the ASD predictive design space

| Variable family                             | Examples of eligible information                                                           | Analytical role in prediction                             | Interpretation boundary                                                        |
|---------------------------------------------|--------------------------------------------------------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------------------------------|
| <b>Drug and polymer identity</b>            | Molecular structure, polymer chemistry, ionization-relevant features                       | Defines the chemical system being modeled                 | Identity alone does not define the realized material state                     |
| <b>Composition</b>                          | Drug loading, polymer fraction, additional formulation components                          | Positions the formulation within composition space        | Nominal composition does not prove molecular-scale homogeneity                 |
| <b>Interaction and miscibility evidence</b> | Experimental or physically grounded indicators of drug–polymer association                 | Describes compatibility and phase-behavior tendencies     | A favorable interaction metric is not equivalent to long-term stability        |
| <b>Thermal and mobility state</b>           | Glass transition, thermal history, mobility-sensitive measurements                         | Describes kinetic constraints on molecular rearrangement  | Intermediate-state descriptors are not complete failure labels                 |
| <b>Manufacturing history</b>                | Route, temperature–composition conditions, solvent removal, thermal or mechanical exposure | Conditions the microstructure actually produced           | Process categories are not interchangeable across equipment or formulations    |
| <b>Storage environment</b>                  | Temperature, humidity, aging time                                                          | Defines the conditions under which stability is evaluated | Stability labels without condition and time metadata are incomplete            |
| <b>Performance state</b>                    | Solid-state persistence, dissolution or release behavior                                   | Provides task-specific endpoints                          | Different performance endpoints should not be collapsed into one success label |

Note: The categorization is proposed for this review. Evidence eligibility indicates a defensible role in formulation-state representation; it does not imply that every variable is necessary, causal, or universally predictive for every ASD modeling task.

### *Drug–polymer miscibility prediction*

Drug–polymer miscibility is a foundational ASD design question, but it is also a difficult machine-learning target because “miscibility” can denote different experimental or thermodynamic constructs. Drug–polymer miscibility is an operationally and methodologically conditioned property that cannot safely be reduced to a single regular-solution parameter or one analytical surrogate across chemically diverse ASDs [7-9]. Thermodynamic treatments remain valuable, particularly for organizing composition and temperature effects, but assumptions developed for simpler mixtures can become inadequate when specific drug–polymer interactions and nonideal behavior dominate [7]. Recent hybrid approaches illustrate a more defensible direction. Empirical measurements can be combined with physically motivated modeling so that predictions are constrained by both observed formulation behavior and thermodynamic structure rather than by a single calculated affinity measure [8]. Orthogonal analytical evidence is equally important because apparently similar dispersions can reveal different degrees or scales of mixing depending on preparation history and the experimental modality used to probe them [9]. For machine learning, the implication is

that miscibility labels require explicit operational definitions.

Computational descriptors also require falsification rather than acceptance because they are chemically plausible. Molecular-dynamics analysis has shown that solvation-free-energy-based reasoning can be misleading when used as a direct surrogate for ASD compatibility [10]. The appropriate conclusion is not that molecular simulation lacks value, but that the connection between a calculated quantity and the formulation endpoint must be demonstrated.

Interaction-informed machine learning offers a complementary strategy by incorporating physicochemical information into models of complex pharmaceutical phase behavior [11]. Such work supports physics-informed representation within the chemistry covered by the training evidence. It does not yet establish broad transfer across unrelated drugs, polymers, compositions, or processing regimes. Miscibility prediction should therefore remain a bounded task with clearly specified state variables, measurement definitions, and applicability domains.

### *Glass transition and molecular mobility*

Glass-transition temperature is among the most frequently invoked indicators of ASD physical stability because molecular rearrangement generally becomes more restricted as mobility decreases. Yet the relationship between thermal state and long-term physical stability is not one-to-one. Long-duration studies of PVP- and PVPVA-based dispersions show that physical stability emerges from the interaction of thermodynamic driving forces and kinetic constraints rather than from glass transition alone [12]. For predictive modeling,  $T_g$  is therefore better interpreted as an intermediate material-state descriptor than as a universal stability label.

This distinction becomes especially important when machine learning is optimized against a readily measurable property. A model may predict  $T_g$  accurately while remaining uninformed about whether a particular formulation will crystallize under a specified humidity, separate into amorphous phases, or retain adequate release. Prediction of an intermediate property can be scientifically valuable, but its validation should be aligned with that specific target rather than expanded rhetorically into a claim about overall formulation robustness.

Polymer choice further exposes the limits of single-objective reasoning. In lumefantrine ASDs, polymers that altered molecular interactions and solid-state stabilization did not produce a simple, universally aligned ranking of physical stability and dissolution performance [13]. The evidence supports the narrower but important conclusion that stability and release can be competing formulation objectives. A polymer may restrict molecular mobility or inhibit crystallization while simultaneously changing hydration, drug-polymer association, or release behavior. A machine-learning architecture that predicts only an average “good formulation” score can obscure this structure. The proposed alternative is to retain thermal state, mobility-related evidence, physical stability, and performance as separate but connectable targets. Their relationships may then be learned or tested explicitly rather than being assumed to move monotonically together. This separation also creates the conceptual basis for later failure-aware modeling, where stability must be indexed to composition, process, environmental condition, observation time, and the specific failure event being measured.

#### *Crystallization propensity and physical stability*

Crystallization is a formulation failure only after its context is specified: the material, composition, storage condition, observation time, analytical sensitivity, and consequence for performance all matter. High-throughput stability experiments show that measured formulation attributes can support statistical prediction of crystallization behavior within a defined experimental domain [14]. Yet storage

studies using structurally similar drugs demonstrate that apparently modest chemical differences can interact with environmental conditions to produce divergent physical-stability behavior [15]. Residual crystallinity further complicates binary labels because its significance depends on amount, distribution, detectability, and subsequent evolution rather than on presence alone [16]. These findings support representing crystallization propensity as a condition-indexed outcome rather than a context-free stable/unstable class.

A failure-aware model should therefore preserve the distinction between an initially amorphous state, latent crystallization risk, detectable crystallinity, and performance-relevant failure. That distinction is proposed here as a modeling requirement, not as an established universal failure taxonomy. Benchmark design should therefore preserve when a sample was evaluated, under which storage condition, and whether the detected solid-state change remained analytical or altered formulation performance. Without these qualifiers, different failure states may collapse into one class.

#### *Moisture sensitivity and storage conditions*

Water introduces a further dynamic variable because it can change both molecular mobility and phase behavior. Experimental and thermodynamic analysis has shown that water sorption, crystallization, and amorphous phase separation can influence one another rather than behaving as independent perturbations [17]. Moisture uptake by commonly used polymeric excipients can also depress glass-transition temperature, with the magnitude dependent on polymer chemistry and relative humidity [18].

Consequently, “humidity” should not be encoded merely as a generic environmental penalty. Its predictive meaning depends on polymer hygroscopicity, composition, prior state, temperature, and exposure time. For machine learning, storage conditions are therefore part of the input state defining the target observation. Models trained on stability labels stripped of temperature, humidity, and aging duration risk learning an average of incompatible experimental regimes rather than a reproducible material relationship.

#### *Dissolution, supersaturation, and drug release*

The transition from storage to dissolution changes the physical problem. Ternary phase-diagram analysis demonstrates that ASD release can depend on evolving drug-polymer-water phase behavior after hydration, so performance cannot be inferred solely from the initial dry-state miscibility of the dispersion [19]. Drug-polymer interactions may also change role across states. In HPMCAS-based systems, interactions associated with

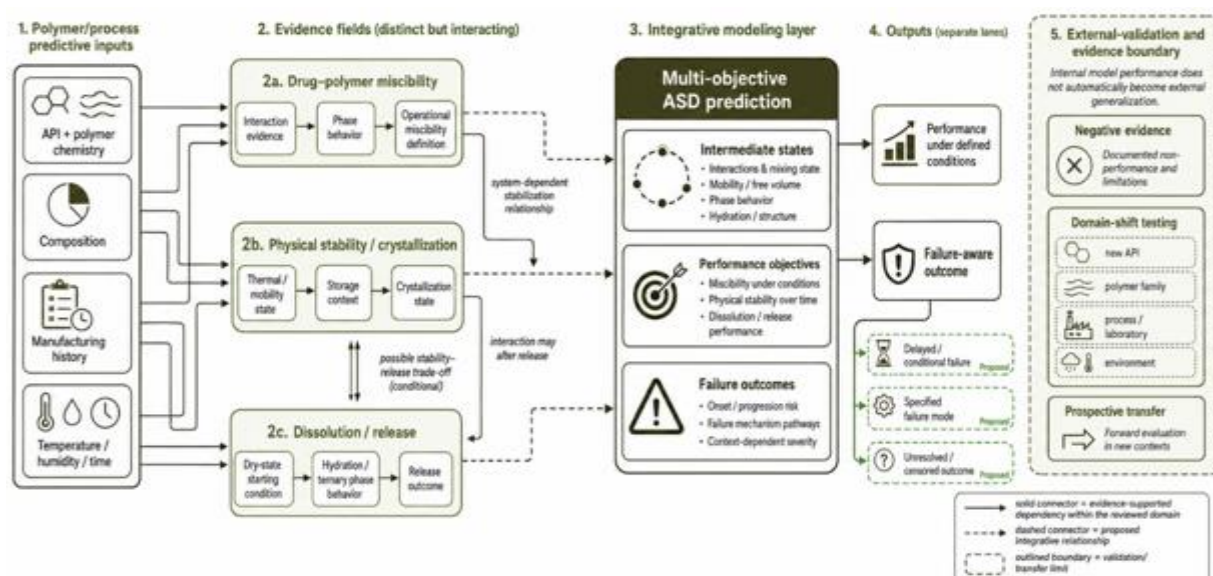
stabilization can coexist with constraints on drug release, demonstrating that “stronger interaction” is not a monotonic formulation advantage [20]. Machine-learning representations intended to predict release should therefore distinguish dry-state properties from hydration-dependent states and measured dissolution outcomes. Supersaturation generation, maintenance, phase separation, precipitation, and polymer-controlled release can occupy different temporal stages even when summarized experimentally by one dissolution curve. The proposed implication is that release models should preserve state transitions where the available data permit them rather than forcing all behavior into one static compatibility descriptor. A formulation may be physically persistent during storage yet perform poorly on hydration, or may sacrifice some dry-state stability while generating superior release. These are competing outcomes, not contradictory data.

### Machine learning for polymer selection

Polymer selection provides a natural test case for combining molecular, experimental, and thermodynamic representations. Liquid-state NMR measurements have been used to distinguish drug–polymer pairs in a predictive

preformulation setting, showing that experimentally measured interaction information can complement structure-derived variables [21]. Machine learning incorporating thermodynamic descriptors has likewise been applied to dissolution-related prediction, supporting the use of physically interpretable inputs when they are linked to the actual endpoint [22]. At a larger formulation-data scale, high-throughput screening coupled with machine learning has generated labeled binary and ternary ASD outcomes, demonstrating how experimental data generation and predictive modeling can be designed together [23].

These approaches should not be treated as interchangeable ranking systems. NMR-based information requires measurement; thermodynamic descriptors depend on model assumptions and parameterization; high-throughput labels remain tied to the experimental route and outcome definition. The synthesis developed here is shown in **Figure 1**: the key point is not to funnel all variables toward one success score, but to preserve distinct material-state and performance layers until a defined multi-objective prediction is made and then challenged at an external-validation boundary.



**Figure 1.** Failure-aware integration of the ASD machine-learning design space

**Table 2** compares the reviewed predictive evidence according to task, representation, target definition, and

transfer boundary rather than treating heterogeneous model metrics as directly comparable.

**Table 2.** Predictive evidence domains relevant to polymer and ASD formulation selection

| Prediction domain               | Representation emphasized                                      | Ground truth or target                | Main analytical value                            | Principal boundary                                     |
|---------------------------------|----------------------------------------------------------------|---------------------------------------|--------------------------------------------------|--------------------------------------------------------|
| Thermal property                | Polymer/formulation descriptors                                | Measured glass transition             | Learns an intermediate material property         | Does not establish overall formulation success         |
| Compatibility or phase behavior | Interaction, thermodynamic, simulation, or analytical features | Operational miscibility/phase outcome | Connects molecular information to material state | Target definition and chemistry remain domain-specific |

| ASD formation                   | Molecular, formulation, and high-throughput descriptors | Experimentally assigned amorphous outcome | Expands labeled formulation space             | Label depends on preparation and measurement protocol       |
|---------------------------------|---------------------------------------------------------|-------------------------------------------|-----------------------------------------------|-------------------------------------------------------------|
| Physical stability              | Molecular/formulation descriptors plus stability labels | Defined stable/unstable outcome           | Demonstrates endpoint-specific classification | Observation time and storage context constrain transfer     |
| Dissolution-related performance | Thermodynamic and formulation descriptors               | Measured release behavior                 | Extends ML beyond dry-state formation         | Media, hydration state, and process history remain relevant |

Note: Categories summarize the reviewed evidence and do not imply direct comparability of model performance across tasks.

#### *Process parameters as predictive inputs*

Process-aware machine learning has already been applied to hot-melt-extruded ASDs using molecular, formulation, and manufacturing information to classify defined formation and chemical-stability outcomes [24]. Such work makes clear that a route label alone is an impoverished process representation. Temperature, composition, residence-related exposure, and other manufacturing conditions determine whether the intended amorphous state can be generated without degradation or residual crystallinity.

Temperature–composition phase diagrams provide a mechanistic example of this principle by defining processing regions associated with residual crystallinity during hot-melt extrusion [25]. They do not predict every downstream quality attribute, but they show why process variables can have physically structured relationships rather than acting as arbitrary metadata. A proposed next step for machine learning is therefore to encode process history at the resolution justified by the data—conditions, trajectories, or physically motivated summaries—while preserving the distinction between process feasibility, immediate product state, storage robustness, and dissolution performance.

#### *Multi-objective models across stability and performance*

ASD-specific machine learning has demonstrated that physical stability can be treated as a learnable classification problem within a compiled formulation dataset [26]. That achievement is important, but stability classification represents only one objective. The selected ASD machine-learning evidence still partitions predictive success into separate tasks such as ASD formation, manufacturing-associated amorphization or chemical stability, and physical stability [23, 24, 26]. Their labels, experimental origins, and validation designs differ, so their reported model metrics cannot be interpreted as a common benchmark.

A multi-objective model should consequently avoid replacing this heterogeneity with a single latent “good ASD” target unless that target is experimentally defined. The framework proposed here retains intermediate-state predictions and performance outcomes as separate heads or linked tasks: miscibility, process feasibility, physical

persistence, release, and specified failure outcomes can share information without being assumed equivalent. Multi-objective learning would be scientifically useful only if trade-offs remain visible. A candidate ranked highly for storage stability but poorly for release should not disappear inside an averaged score.

#### *Why models predicting average performance miss formulation failure*

Failure is especially poorly represented by average performance because it can be delayed, conditional, and asymmetric. Long-term analysis across multiple preparation approaches and storage conditions showed that crystallization behavior depended on factors including process, humidity, temperature, drug characteristics, and interactions over the observation period [27]. Such evidence supports treating time and environment as components of the outcome definition rather than nuisance variation.

This review therefore proposes a failure-aware target structure in which formulations can remain successful, fail under specified conditions, fail after a measurable delay, or remain unresolved within the observation window. The categories are conceptual and require prospective operationalization. Their purpose is to prevent a model from learning that an unobserved failure is equivalent to durable success. Failure-aware modeling should also distinguish mechanism where evidence permits: residual crystallinity, amorphous phase separation, storage crystallization, release limitation, and processing-induced chemical instability are not interchangeable negative labels.

#### *External validation and formulation transferability*

Integration of artificial intelligence with physiologically based pharmacokinetic modeling has begun to connect ASD formulation information with in vitro and downstream in vivo behavior in defined systems [28]. This is an important expansion beyond isolated material-property prediction, but it does not establish general transferability across new drugs, polymers, processes, laboratories, or clinical settings.

External validation should therefore be decomposed according to the shift being tested. Chemical transfer asks

whether a model generalizes to new APIs; formulation transfer challenges polymer or composition space; process transfer changes manufacturing history; measurement transfer changes laboratory or analytical protocol; and downstream transfer asks whether a formulation model remains informative for biopharmaceutic behavior. A compatibility surrogate that appears plausible in one computational regime can fail under direct scrutiny [10], while HME-focused learning remains tied to its manufacturing domain [24]. Prospective validation across such shifts is consequently a stronger test than a random split of records generated under closely related conditions. This hierarchy is proposed as an evaluation framework, not a validated scoring standard.

#### *Data standardization and negative formulation evidence*

Standardization is necessary because an ASD label is meaningful only with its provenance. High-throughput formation labels arise from a defined preparation and characterization workflow [23], whereas process-oriented datasets may encode hot-melt-extrusion-specific outcomes

[24]. Pooling such records without preserving route, composition, analytical definition, and experimental conditions can create apparent sample size while weakening construct validity.

Negative evidence deserves equal attention. Long-term observations may include delayed crystallization or formulations that remain unresolved at the end of monitoring [27]. Computational work can also reveal that a plausible descriptor fails to track the intended endpoint [10]. Both forms of negative evidence are informative because they define where representations or formulations break down. Standardization should also distinguish directly measured variables, calculated descriptors, and retrospectively assigned labels. That provenance helps expose leakage or circular prediction.

**Table 3** summarizes the evidence structure proposed for failure-aware datasets. It is not a reporting standard; it identifies information that the reviewed literature indicates should remain recoverable when heterogeneous ASD records are combined.

**Table 3.** Evidence structure for failure-aware ASD datasets

| Evidence component                    | Information to preserve                                  | Why it matters                                 | Failure if omitted                             |
|---------------------------------------|----------------------------------------------------------|------------------------------------------------|------------------------------------------------|
| <b>Material identity</b>              | API, polymer, composition, additives                     | Defines chemical and formulation domain        | Hidden chemistry shift                         |
| <b>Process provenance</b>             | Preparation route and relevant conditions                | Connects nominal composition to realized state | Process effects misassigned to chemistry       |
| <b>Observation context</b>            | Temperature, humidity, medium, aging time                | Defines conditions of stability or release     | Incompatible labels pooled                     |
| <b>Endpoint definition</b>            | Measurement, threshold logic, observation window         | Specifies what the model predicts              | Ambiguous success/failure                      |
| <b>Negative or unresolved outcome</b> | Failed, delayed, censored, or indeterminate observations | Preserves boundary cases                       | Survivorship and label bias                    |
| <b>Validation provenance</b>          | Held-out chemistry, process, laboratory, or condition    | Identifies the tested generalization           | Internal accuracy mistaken for transferability |

Note: This evidence structure is proposed for model development and comparison; it is not presented as a validated regulatory or community standard.

#### *Next-generation failure-aware asd modeling*

The next generation of ASD models should move from isolated endpoints toward structured integration without erasing endpoint identity. Analytical representations such as interaction-sensitive NMR measurements can complement molecular descriptors [21], while thermodynamic descriptors provide a route for embedding physically interpretable information into learning models [22]. High-throughput experimentation can expand labeled formulation space and, in principle, be redirected toward regions where current models are uncertain or contradictory [23]. Active selection of such experiments remains a research direction rather than an established ASD workflow.

A proposed architecture would therefore combine composition and molecular descriptors with process history, state-specific analytical measurements, and environmental context; produce distinct predictions for intermediate states and performance outcomes; and retain a dedicated failure representation conditioned on time and exposure. Long-term failure observations support the need for temporal context [27], whereas integrated AI/PBPK work indicates that downstream fate can be modeled as a separate level rather than assumed from in vitro formulation success [28]. Prospective benchmarks should test these levels under deliberately withheld chemistry, polymer, process, laboratory, and environmental domains. Such benchmarks would make failure prediction falsifiable

rather than merely improve interpolation within familiar formulation neighborhoods.

#### *Review limitations*

The direct ASD machine-learning literature remains smaller than the broader physicochemical evidence base, and available studies address different targets with heterogeneous labels and validation schemes. Historical stability classification demonstrates feasibility but cannot define expected performance in newly sampled formulation domains [26]. Molecular-simulation studies additionally show that chemically plausible surrogate variables may not preserve their intended relationship to compatibility [10], while high-throughput formation datasets remain conditioned by their preparation and labeling procedures [23].

This review therefore does not support quantitative cross-study ranking of algorithms, a universal feature set, or a validated composite definition of ASD success. It also cannot establish that the proposed failure-aware architecture improves prospective formulation selection. Downstream integration remains encouraging but system-bounded [28]. The framework should be treated as a structured research agenda whose value depends on prospective testing, transparent negative evidence, and domain-specific external validation.

## CONCLUSION

Machine learning is beginning to operate across several scientifically meaningful parts of ASD development, but current predictive tasks remain more mature as separate problems than as one integrated formulation model. Miscibility, thermal state, crystallization resistance, moisture response, manufacturing history, dissolution, and downstream performance describe connected but non-equivalent aspects of a dynamic material system.

The principal contribution of this review is a proposed failure-aware organization of that design space. It preserves intermediate physicochemical states, process and environmental context, performance objectives, and failure outcomes as analytically distinct before integration. This structure prevents a favorable prediction in one domain from being interpreted automatically as overall formulation success.

Progress now depends less on adding increasingly complex algorithms to poorly harmonized datasets than on improving the meaning of the data supplied to those algorithms. Context-rich labels, negative and unresolved formulations, explicit observation windows, physically defensible representations, and validation across genuine domain shifts are central requirements. The proposed framework is not a validated decision rule; it defines a

testable direction for developing ASD models whose predictions remain interpretable when formulations move beyond the conditions under which they were learned.

**Acknowledgments:** None

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## REFERENCES

- [1] Moseson DE, Tran TB, Karunakaran B, Ambardekar R, Hiew TN. Trends in amorphous solid dispersion drug products approved by the U.S. Food and Drug Administration between 2012 and 2023. *Int J Pharm X*. 2024;7:100259. doi:10.1016/j.ijpx.2024.100259
- [2] Bhatta HP, Han HK, Maharjan R, Jeong SH. Recent techniques to improve amorphous dispersion performance with quality design, physicochemical monitoring, molecular simulation, and machine learning. *Pharmaceutics*. 2025;17(10):1249. doi:10.3390/pharmaceutics17101249
- [3] Alzyoud A, El Sayed M, Rawas-Qalaji MM, Alhalaweh A. Polymer selection in amorphous solid dispersions: From stabilization mechanisms to marketed drug products. *J Control Release*. 2026;395:115014. doi:10.1016/j.jconrel.2026.115014
- [4] Bhujbal SV, Mitra B, Jain U, Gong Y, Agrawal A, Karki S, et al. Pharmaceutical amorphous solid dispersion: A review of manufacturing strategies. *Acta Pharm Sin B*. 2021;11(8):2505-36. doi:10.1016/j.apsb.2021.05.014
- [5] Di Mare EJ, Punia A, Lamm MS, Rhodes TA, Gormley AJ. Data-driven design of novel polymer excipients for pharmaceutical amorphous solid dispersions. *Bioconj Chem*. 2024;35(9):1363-72. doi:10.1021/acs.bioconjchem.4c00294
- [6] Martynek D, Ridvan L, Sívén M, Šoš M. Stability and recrystallization of amorphous solid dispersions prepared by hot-melt extrusion and spray drying. *Int J Pharm*. 2025;672:125331. doi:10.1016/j.ijpharm.2025.125331
- [7] Anderson BD. Predicting solubility/miscibility in amorphous dispersions: It is time to move beyond regular solution theories. *J Pharm Sci*. 2018;107(1):24-33. doi:10.1016/j.xphs.2017.09.030
- [8] de Castro M, Cordeiro AS, Li M, Lübbert C, McColl C, Khurana J, et al. Advancing amorphous solid dispersions through empirical and hybrid modeling of drug-polymer solubility and miscibility: A case study

- using ibuprofen. *Int J Pharm X*. 2025;10:100373. doi:10.1016/j.ijpx.2025.100373
- [9] Song S, Zhang T, Xu J, Munson EJ, Siegel RA. Probing miscibility of spray-dried and melt-quenched amorphous dispersions using rheology and solid-state NMR spectroscopy. *Mol Pharm*. 2026;23(3):1564-71. doi:10.1021/acs.molpharmaceut.5c01127
- [10] Higginbotham T, Meier K, Ramirez J, Garaizar A. Predicting drug-polymer compatibility in amorphous solid dispersions by MD simulation: On the trap of solvation free energies. *Mol Pharm*. 2025;22(2):760-70. doi:10.1021/acs.molpharmaceut.4c00810
- [11] Ge K, Huang Y, Ji Y. Machine learning with active pharmaceutical ingredient/polymer interaction mechanism: Prediction for complex phase behaviors of pharmaceuticals and formulations. *Chin J Chem Eng*. 2024;66(2):263-72. doi:10.1016/j.cjche.2023.09.006
- [12] Lehmkeper K, Kyeremateng SO, Heinzerling O, Degenhardt M, Sadowski G. Long-term physical stability of PVP- and PVPVA-amorphous solid dispersions. *Mol Pharm*. 2017;14(1):157-71. doi:10.1021/acs.molpharmaceut.6b00763
- [13] Hiew TN, Zemlyanov DY, Taylor LS. Balancing solid-state stability and dissolution performance of lumefantrine amorphous solid dispersions: The role of polymer choice and drug-polymer interactions. *Mol Pharm*. 2022;19(2):392-413. doi:10.1021/acs.molpharmaceut.1c00481
- [14] Ghazi NF, Burley JC, Dryden IL, Roberts CJ. High-throughput microarray approaches for predicting the stability of drug-polymer solid dispersions. *Mol Pharm*. 2025;22(1):343-62. doi:10.1021/acs.molpharmaceut.4c00955
- [15] Jarrells TW, Yuan X, Munson EJ. Impact of storage conditions on the physical stability of amorphous solid dispersions containing two structurally similar drugs. *Mol Pharm*. 2025;22(6):2927-38. doi:10.1021/acs.molpharmaceut.4c01279
- [16] Moseson DE, Taylor LS. Crystallinity: A complex critical quality attribute of amorphous solid dispersions. *Mol Pharm*. 2023;20(10):4802-25. doi:10.1021/acs.molpharmaceut.3c00526
- [17] Luebbert C, Wessner M, Sadowski G. Mutual impact of phase separation/crystallization and water sorption in amorphous solid dispersions. *Mol Pharm*. 2018;15(2):669-78. doi:10.1021/acs.molpharmaceut.7b01076
- [18] Patel NG, Serajuddin ATM. Moisture sorption by polymeric excipients commonly used in amorphous solid dispersion and its effect on glass transition temperature: I. Polyvinylpyrrolidone and related copolymers. *Int J Pharm*. 2022;616:121532. doi:10.1016/j.ijpharm.2022.121532
- [19] Deac A, Luebbert C, Qi Q, Courtney RM, Indulkar AS, Gao Y, et al. Dissolution mechanisms of amorphous solid dispersions: Application of ternary phase diagrams to explain release behavior. *Mol Pharm*. 2024;21(4):1900-18. doi:10.1021/acs.molpharmaceut.3c01179
- [20] Bapat P, Paul S, Tseng YC, Taylor LS. Interplay of drug-polymer interactions and release performance for HPMCAS-based amorphous solid dispersions. *Mol Pharm*. 2024;21(3):1466-78. doi:10.1021/acs.molpharmaceut.3c01106
- [21] Coutinho AL, Hom K, Polli JE. Prediction of successful amorphous solid dispersion pairs through liquid state nuclear magnetic resonance. *Mol Pharm*. 2024;21(12):6153-65. doi:10.1021/acs.molpharmaceut.4c00359
- [22] Ge K, Shen J, Chen H, Wang J, Cui Y, Shen C, et al. Advancing amorphous solid dispersions design: Insights into dissolution kinetics via thermodynamic descriptor and machine learning. *Mol Pharm*. 2025;22(7):4259-68. doi:10.1021/acs.molpharmaceut.5c00518
- [23] Lu T, Wu Y, Xiong P, Zhong H, Ding Y, Li H, et al. Combining high-throughput screening and machine learning to predict the formation of both binary and ternary amorphous solid dispersion formulations for early drug discovery and development. *Pharm Res*. 2025;42(4):697-709. doi:10.1007/s11095-025-03853-z
- [24] Jiang J, Lu A, Ma X, Ouyang D, Williams RO 3rd. The applications of machine learning to predict the forming of chemically stable amorphous solid dispersions prepared by hot-melt extrusion. *Int J Pharm X*. 2023;5:100164. doi:10.1016/j.ijpx.2023.100164
- [25] Moseson DE, Taylor LS. The application of temperature-composition phase diagrams for hot melt extrusion processing of amorphous solid dispersions to prevent residual crystallinity. *Int J Pharm*. 2018;553(1-2):454-66. doi:10.1016/j.ijpharm.2018.10.055
- [26] Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, et al. Predicting physical stability of solid dispersions by machine learning techniques. *J Control Release*. 2019;311-312:16-25. doi:10.1016/j.jconrel.2019.08.030
- [27] Wu J, Van den Mooter G. Statistical analysis of long-term physical stability testing of amorphous solid dispersions. *Int J Pharm*. 2025;681:125844. doi:10.1016/j.ijpharm.2025.125844

- [28] Zhu J, Xiong P, Wang W, Lu T, Ouyang D. Integrating artificial intelligence and physiologically based pharmacokinetic modeling to predict in vitro and in vivo fate of amorphous solid dispersions. J Control Release. 2025;386:114123. doi:10.1016/j.jconrel.2025.114123