International Journal of Medical and Pharmaceutical Sciences
Open access peer-reviewed Journal
Volume 2, Issue 11, Page 11-18.
https://doi.org/10.5281/zenodo.23938132

Review Article

Application of Artificial Intelligence in Stability Prediction and Shelf-Life Estimation of Pharmaceutical Formulations

Tanmay Giradkar*[1], Rohan Rathod[1], Shejal Kadu[2], Reshma Rajput[2], Purva Dod[3], Janhavi Choukhnade[4], Vivek Daberao[5]

[1]*Kamla Nehru College of Pharmacy, Butibori, (RTMNU)*
[2]*Shri Sant Gajanan Maharaj College of Pharmacy, Bhuldana*
[3]*Dr. Rajendra Gode, Institute of Pharmacy, Amravati*
[4]*Vidyaniketan College of Pharmacy, Anjangaon Surji, Dist. Amravati, Maharashtra 444705.*
[5]*Dr. Rajendra Gode, College of Pharmacy, Malkapur*

Stability testing determines whether a pharmaceutical product keeps its identity, strength, quality and purity over its labeled shelf life, but conventional studies are slow, expensive and material-intensive, and they offer limited insight into why a formulation degrades. Artificial intelligence (AI), and machine learning (ML) in particular, is increasingly used to predict degradation pathways, physical stability, and time to specification failure from formulation, process, structural and early stability data. This review summarizes the scientific and regulatory background of stability assessment; the main AI methods applied (tree ensembles, kernel methods, Gaussian processes, deep and graph neural networks, and hybrid mechanistic-ML models); data requirements and feature engineering; applications to small molecules, amorphous solid dispersions, biologics, nucleic acids and drug-delivery systems; and an end-to-end workflow for AI-assisted shelf-life estimation. We discuss validation, uncertainty quantification, explainability and regulatory expectations, and outline the challenges that must be solved before AI-derived predictions can support labeled shelf-life claims. Current evidence suggests that AI is best positioned today as a decision-support tool for formulation screening, study design and risk assessment, complementing rather than replacing real-time stability data.

Keywords: artificial intelligence; machine learning; stability prediction; shelf life; accelerated stability assessment; forced degradation; quality by design; explainable AI.

INTRODUCTION

Shelf life is the period during which a drug product is expected to remain within approved specifications when stored under labeled conditions. It is established from stability studies designed under the International Council for Harmonisation (ICH) Q1A(R2) guideline and evaluated statistically under ICH Q1E [1,2]. Long-term studies commonly run for 12 months or more before submission and continue for the proposed shelf life, so stability data often arrive late in development, when reformulation is costly. Quality by design (QbD) and quality risk management ask developers to understand how formulation and process variables affect critical quality attributes, including stability [3-5]. Accelerated approaches such as the Accelerated Stability Assessment Program (ASAP) use short studies at elevated temperature and humidity, fitted with humidity-corrected Arrhenius models, to rank formulations and forecast long-term behavior of many solid dosage forms [6-8]. These methods work well when degradation follows a single dominant, well-behaved mechanism, but they are less reliable for complex, multi-mechanism or non-Arrhenius behavior. In parallel, AI has become established across drug discovery, formulation and

delivery [9-11]. The same techniques can learn structure-stability relationships from historical and experimental data, potentially shortening development timelines and reducing the number of experiments. This review examines how AI is being applied to stability prediction and shelf-life estimation, what evidence supports it, and what limits its regulatory use today.

2. Conventional Stability Science and Its Limits

Chemical degradation in pharmaceuticals is classically described by reaction kinetics, with temperature dependence following the Arrhenius equation and additional dependence on humidity, pH, oxygen, light and excipient interactions [12,13]. Forced degradation (stress testing) under heat, humidity, acid, base, oxidation and light identifies likely degradation pathways and supports development of stability-indicating methods [14,15]. Biologics add further layers of complexity, including aggregation, deamidation, oxidation, fragmentation and interfacial damage [16,17]. The ICH framework prescribes standard storage conditions, summarized in Table 1.

Table 1. Standard stability study conditions for the general case under ICH Q1A(R2) [1].

Study	Storage condition	Minimum duration at submission
Long-term	25 °C ± 2 °C / 60% RH ± 5% (or 30 °C / 65% RH)	12 months
Intermediate	30 °C ± 2 °C / 65% RH ± 5%	6 months
Accelerated	40 °C ± 2 °C / 75% RH ± 5%	6 months

Conventional approaches face several recurring limitations:

- **Time and cost.** Real-time data take months to years and consume drug substance, which is scarce early in development.
- **Extrapolation risk.** Linear or Arrhenius extrapolation can fail when mechanisms change with temperature, or when physical changes such as crystallization or phase separation intervene.
- **Limited generalization.** Each study characterizes one formulation; knowledge from earlier products is rarely reused quantitatively.
- **Multivariate complexity.** Interactions among drug, excipients, moisture, packaging and process are hard to capture with a few factors.

3. Artificial Intelligence Methods Used for Stability Prediction

3.1 Classical machine learning

Tree ensembles such as random forests and gradient-boosted trees handle mixed numeric and categorical inputs, tolerate missing data and perform well on the small to medium tabular datasets typical of formulation work [18,20]. Support vector machines are effective for classification tasks such as stable versus unstable, particularly with limited samples [19]. Gaussian processes provide predictions with built-in uncertainty estimates and suit data-scarce settings and sequential experimental design [21].

3.2 Deep learning

Deep neural networks learn hierarchical representations from large datasets [22]. Recurrent architectures such as long short-term memory networks model stability time series, for example degradant growth across pull points [23]. Graph neural networks operate directly on molecular graphs and learn degradation-relevant substructures without hand-crafted descriptors [24,25]. Deep learning typically requires larger datasets than are available for a single product, so it is most valuable when trained on pooled or public data, or pretrained and fine-tuned.

3.3 Hybrid and physics-informed models

Physics-informed neural networks and neural ordinary differential equations embed kinetic laws into the learning problem, so predictions respect mass balance and known rate behavior while the network learns the unknown terms [26,27]. For stability, this suggests learning how rate constants depend on

formulation variables while keeping an Arrhenius or humidity-modified kinetic backbone, an approach that extends ASAP-style modeling.

3.4 Uncertainty, optimization and transfer learning

Shelf-life decisions depend on how confident a prediction is, not only its value. Bayesian deep learning and dropout-based approximations provide uncertainty estimates for neural models [28], and Bayesian optimization uses them to choose the next informative experiment [29]. Transfer learning reuses knowledge from related compounds or data-rich tasks to improve prediction on small datasets [30].

3.5 Explainability

Model-agnostic explanation tools such as SHAP and LIME attribute predictions to input features and help scientists check whether a model relies on chemically sensible drivers such as moisture content, reactive functional groups or excipient impurities [31,32]. Others argue that for high-stakes decisions inherently interpretable models should be preferred over post hoc explanations of black boxes [33].

Table 2. Common AI approaches for stability prediction.

Approach	Typical stability use	Strength	Limitation	Ref.
Random forest, gradient boosting	Degradant level or stability class from formulation variables	Robust on small tabular data; feature importance	Limited extrapolation outside training range	[18,20]
Support vector machine	Stable versus unstable classification	Good with few samples	Sensitive to kernel and scaling; no native uncertainty	[19]
Gaussian process	Small-data regression; experiment selection	Calibrated uncertainty	Scales poorly to very large datasets	[21]
Deep and recurrent networks	Time-series degradation profiles	Learns complex nonlinear patterns	Data-hungry; opaque	[22,23]
Graph neural networks	Molecule-level degradation liability	No manual descriptors	Needs large, diverse training sets	[24,25]
Physics-informed / neural ODE	Kinetics with learned parameters	Respects known kinetics; better extrapolation	More complex to build and validate	[26,27]

4. Data Sources and Feature Engineering

Model quality is limited by data quality. Typical inputs fall into five groups:

1. **Molecular descriptors and structure.** Fingerprints, quantum-chemical descriptors, functional-group flags and learned embeddings [25].

2. **Formulation composition.** Excipient types and levels, pH, buffer, water content, drug loading and polymer grade.

3. **Process and packaging.** Manufacturing route, particle size, residual solvent, container-closure and desiccant.

4. **Stress and stability data.** Forced degradation outcomes and time-temperature-humidity pulls, including ASAP-style studies [14,15].

5. **Analytical and spectroscopic data.** Near-infrared, Raman and chromatographic data, often preprocessed to remove scatter and baseline effects [34], and process analytical technology streams [35].

Stability data are frequently scattered across reports, instruments and sites. Applying the FAIR principles (findable, accessible, interoperable, reusable) to stability data, with harmonized units, metadata and ontologies, is a prerequisite for pooled modeling [36]. Because degradation endpoints are often censored (below the limit of quantitation) and unbalanced

across conditions, careful handling of censoring, replicate structure and batch effects is essential.

5. Applications

5.1 Chemical degradation of small molecules

Knowledge-based software such as Zeneth predicts likely degradants from structure and conditions, and benchmarking against forced-degradation outcomes has shown useful but imperfect agreement [37]. Data-driven reaction-outcome models, trained on large reaction corpora, offer a complementary route to anticipating transformations such as oxidation and hydrolysis [38]. Combining these tools with experimental stress data can prioritize which degradants to monitor.

5.2 Physical stability and amorphous solid dispersions

Crystallization of amorphous drugs undermines solubility advantages. ML models trained on solid-dispersion datasets have been used to predict physical stability from drug properties, polymer type, drug loading and storage conditions [39]. Classification of crystallization tendency from molecular properties and thermal behavior provides a basis for feature selection [40,41].

5.3 Biologics and developability

For therapeutic antibodies, developability assessment uses biophysical and sequence-derived features to flag aggregation- and viscosity-prone candidates early. Landscape analyses of clinical-stage antibodies defined property ranges that inform such models [42], and computational guidelines operationalize them for screening [43]. Sequence-based tools predict solubility and aggregation propensity, linking molecular design to stability outcomes [44].

5.4 Nucleic-acid products

RNA instability limits the storage of mRNA vaccines and therapeutics. A crowdsourced dataset and deep learning models predicted degradation along RNA sequences, showing that sequence and structure can be optimized for stability [45]. Similar approaches may inform design of lipid-nanoparticle formulations.

5.5 Formulation design and delivery systems

ML models are being used to design formulations that meet performance and stability targets. Examples include models that accelerate the design of polymeric long-acting injectables [46] and deep-learning models that predict in vitro properties of oral and other formulations [47]. In these settings, stability-related endpoints become one of several objectives optimized jointly.

5.6 From degradation rate to shelf-life estimate

Shelf-life assignment under ICH Q1E relies on regression of attribute versus time and the time at which the one-sided 95% confidence limit of the mean intersects the specification limit [2]. AI can enter at three points: predicting initial rate parameters from formulation data before real-time results exist, pooling batches or related formulations to narrow intervals, and flagging atypical batches. Any AI-derived estimate should be reported with an uncertainty interval and confirmed by real-time data before it supports a labeled claim.

Table 3. Representative AI applications in stability and related areas.

Application	Typical inputs	Typical output	Ref.
Degradant prediction	Structure, conditions	Likely degradation products	[37,38]
Accelerated shelf-life modeling	Short-term temperature and humidity data	Degradation rate and time to limit	[6-8]
Amorphous dispersion physical stability	Drug, polymer, loading, storage	Stable or crystallizing	[39-41]
Antibody developability	Sequence and biophysical properties	Aggregation or viscosity risk	[42-44]
mRNA degradation	Sequence and structure	Per-nucleotide degradation	[45]
Formulation design	Composition and process variables	Predicted performance and stability	[46,47]

6. An End-to-End Workflow for AI-Assisted Shelf-Life Estimation

A practical workflow links data, models and decisions so that each prediction is traceable:

1. **Define the decision.** For example, rank candidate formulations, or estimate time to a degradant limit for a given pack.
2. **Assemble and curate data.** Combine historical stability, forced-degradation and ASAP data with harmonized metadata.
3. **Engineer features.** Use chemically meaningful descriptors, composition and process variables, and handle values below quantitation explicitly.
4. **Train with appropriate validation.** Hold out whole formulations or batches, not random rows, to test generalization to new products.
5. **Quantify uncertainty.** Report prediction intervals and flag out-of-domain inputs.
6. **Verify against real-time data.** Compare predictions with ongoing studies and update the model.
7. **Document and govern.** Record data lineage, model version, assumptions and limits of use.

7. Validation and Regulatory Considerations

Current guidelines do not exclude model-based approaches: ICH Q8 and Q9 encourage science- and risk-based development, and QbD explicitly values mechanistic understanding [3-5]. For regulatory use, the U.S. Food and Drug Administration has proposed a risk-based credibility framework for AI used to support regulatory decisions on drugs and biologics, built around the model's context of use [48], and the European Medicines Agency has published a reflection paper on AI across the medicinal product lifecycle [49]. For shelf-life claims, the established expectation remains real-time data evaluated under ICH Q1E; AI predictions are therefore most defensible when used to support study design, justify bracketing or matrixing, prioritize risks, and provide supportive evidence, with clear context of use, validation and human oversight.

Key validation elements include independent test sets from different formulations or batches, prospective confirmation, evaluation of applicability domain, sensitivity analysis, and change control for model updates.

8. Challenges

- **Data scarcity and heterogeneity.** Stability datasets are small, proprietary and inconsistently reported, limiting model generality.
- **Extrapolation and mechanism shift.** ML models can fail when conditions or mechanisms differ from training data, especially at elevated temperature.
- **Bias in historical data.** Failed or discontinued formulations are underreported.
- **Interpretability.** Scientists and regulators need chemically plausible explanations.
- **Integration and governance.** Models must be versioned, validated and integrated with laboratory information systems.
- **Regulatory acceptance.** Standards for validating AI as evidence in stability claims are still emerging.

9. FUTURE PERSPECTIVES

Likely developments include federated learning across companies to pool data without sharing proprietary formulations, hybrid models that combine kinetics with learned corrections, active-learning loops that choose the next stability experiment, digital twins of storage and supply-chain conditions, and tighter coupling of in-line sensing with predictive models. Standardized benchmark datasets and shared reporting formats would accelerate progress and permit fair comparison between methods.

CONCLUSION

AI offers a promising route to faster, more informative stability assessment by learning from

structure, formulation and early stability data. Current successes are strongest in screening, risk ranking, and physical and biologic stability prediction, with hybrid mechanistic-ML approaches particularly attractive for shelf-life estimation. Realizing regulatory-grade use requires high-quality pooled data, rigorous validation, uncertainty quantification, interpretability and prospective confirmation against real-time data. Used in this way, AI can shorten development timelines and reduce experimental burden while retaining the evidentiary standard on which shelf-life claims depend.

REFERENCES

1. International Council for Harmonisation. ICH Q1A(R2): Stability testing of new drug substances and products. 2003.
2. International Council for Harmonisation. ICH Q1E: Evaluation of stability data. 2003.
3. International Council for Harmonisation. ICH Q8(R2): Pharmaceutical development. 2009.
4. Yu LX. Pharmaceutical quality by design: product and process development, understanding, and control. Pharm Res. 2008;25(4):781-791.
5. International Council for Harmonisation. ICH Q9(R1): Quality risk management. 2023.
6. Waterman KC, Adami RC. Accelerated aging: prediction of chemical stability of pharmaceuticals. Int J Pharm. 2005;293(1-2):101-125.
7. Waterman KC, Carella AJ, Gumkowski MJ, et al. Improved protocol and data analysis for accelerated shelf-life estimation of solid dosage forms. Pharm Res. 2007;24(4):780-790.
8. Waterman KC. The application of the Accelerated Stability Assessment Program (ASAP) to quality by design (QbD) for drug product stability. AAPS PharmSciTech. 2011;12(3):932-937.
9. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477.
10. Hassanzadeh P, Atyabi F, Dinarvand R. The significance of artificial intelligence in drug delivery system design. Adv Drug Deliv Rev. 2019;151-152:169-190.
11. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021; 175:113806.
12. Connors KA, Amidon GL, Stella VJ. Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists. 2nd ed. New York: Wiley; 1986.
13. Yoshioka S, Stella VJ. Stability of Drugs and Dosage Forms. New York: Kluwer Academic/Plenum; 2000.
14. Blessy M, Patel RD, Prajapati PN, Agrawal YK. Development of forced degradation and stability indicating studies of drugs: a review. J Pharm Anal. 2014;4(3):159-165.
15. Baertschi SW, Alsante KM, Reed RA, eds. Pharmaceutical Stress Testing: Predicting Drug Degradation. 2nd ed. Boca Raton: CRC Press; 2011.
16. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: an update. Pharm Res. 2010;27(4):544-575.
17. Hawe A, Wiggenhorn M, van de Weert M, Garbe JHO, Mahler HC, Jiskoot W. Forced degradation of therapeutic proteins. J Pharm Sci. 2012;101(3):895-913.
18. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
19. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-297.
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:785-794.
21. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press; 2006.
22. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
23. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-1780.
24. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning. 2017:1263-1272.
25. Yang K, Swanson K, Jin W, et al. Analyzing learned molecular representations for property

prediction. J Chem Inf Model. 2019;59(8):3370-3388.
26. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys. 2019; 378:686-707.
27. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural ordinary differential equations. In: Advances in Neural Information Processing Systems 31. 2018.
28. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning. 2016:1050-1059.
29. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems 25. 2012.
30. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016; 3:9.
31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30. 2017.
32. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:1135-1144.
33. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-215.
34. Rinnån Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem. 2009;28(10):1201-1222.
35. U.S. Food and Drug Administration. Guidance for industry: PAT, a framework for innovative pharmaceutical development, manufacturing, and quality assurance. 2004.
36. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016; 3:160018.
37. Kleinman MH, Baertschi SW, Alsante KM, et al. In silico prediction of pharmaceutical degradation pathways: a benchmarking study using Zeneth. Mol Pharm. 2014;11(11):4179-4188.
38. Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of organic reaction outcomes using machine learning. ACS Cent Sci. 2017;3(5):434-443.
39. Han R, Xiong H, Ye Z, et al. Predicting physical stability of solid dispersions by machine learning techniques. J Control Release. 2019;311-312:16-25.
40. Baird JA, Van Eerdenbrugh B, Taylor LS. A classification system to assess the crystallization tendency of organic molecules from undercooled melts. J Pharm Sci. 2010;99(9):3787-3806.
41. Kawakami K. Crystallization tendency of pharmaceutical glasses: relevance to compound properties, impact of formulation process, and implications for design of amorphous solid dispersions. Pharmaceutics. 2019;11(5):202.
42. Jain T, Sun T, Durand S, et al. Biophysical properties of the clinical-stage antibody landscape. Proc Natl Acad Sci USA. 2017;114(5):944-949.
43. Raybould MIJ, Marks C, Krawczyk K, et al. Five computational developability guidelines for therapeutic antibody profiling. Proc Natl Acad Sci USA. 2019;116(10):4025-4030.
44. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol. 2015;427(2):478-490.
45. Wayment-Steele HK, Kladwang W, Strom AI, et al. Deep learning models for predicting RNA degradation via dual crowdsourcing. Nat Mach Intell. 2022;4(12):1174-1184.
46. Bannigan P, Bao Z, Hickman RJ, et al. Machine learning models to accelerate the design of polymeric long-acting injectables. Nat Commun. 2023; 14:35.
47. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for in vitro prediction of pharmaceutical formulations. Acta Pharm Sin B. 2019;9(1):177-185.
48. U.S. Food and Drug Administration. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products (draft guidance). 2025.
49. European Medicines Agency. Reflection paper on the use of artificial intelligence in the medicinal

product lifecycle.
EMA/CHMP/CVMP/83833/2023. 2024.

Cite: Tanmay Giradkar*, Rohan Rathod, Shejal Kadu, Reshma Rajput, Purva Dod, Janhavi Choukhnade, Vivek Daberao, Application of Artificial Intelligence in Stability Prediction and Shelf-Life Estimation of Pharmaceutical Formulations, Int. J. Med. Pharm. Sci., 2026, 2 (11), 11-18. https://doi.org/10.5281/zenodo.23938132