RESEARCH PAPER

Retrospective Analysis of Artificial Intelligence and Information Technology Models for Predicting Nanocarrier Performance: A Comparative Study Using Publicly Available Formulation Datasets

Mr. Hemanth J[1], Dr. Lakshminarayana K[2]

[1] Research Scholar, Department of Management Studies (MBA), Visvesvaraya Technological University – Belagavi, Center for Post Graduate Studies - Bangalore. Email: hemanthj1999@gmail.com. ORCID: 0009-0005-8411-2863

[2] Assistant Professor & Research Supervisor, Department of Management Studies, Visvesvaraya Technological University – Belagavi, Center for Post Graduate Studies - Bangalore. Email: appinarayan@gmail.com. ORCID: 0000-0003-0112-3590

Received: 12th Mar, 2026 | **Revised:** 24th Mar, 2026 | **Accepted:** 14th Apr, 2026 | **Available Online:** 30th Apr, 2026

ABSTRACT

Background:

Artificial Intelligence (AI) and Information Technology (IT) models have gained considerable attention in pharmaceutical research for predicting nanocarrier performance metrics including drug loading capacity (DLC) and encapsulation efficiency (EE). However, most published models undergo validation exclusively on datasets used for their training, yielding overoptimistic performance estimates that may not translate to real-world applications.

Objective:

This investigation retrospectively evaluates the generalizability of four AI methodologies – Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Gradient Boosting Machine (GBM) – for predicting nanocarrier performance using publicly available secondary data.

Methods:

A systematic literature search across PubMed, Scopus, and Google Scholar (2018–2024) identified 15 eligible studies contributing 342 unique formulation records. Inclusion criteria required polymeric or lipidic nanocarriers with documented formulation parameters and reported DLC or EE values. Each model was trained on 80% of data (n=274) and tested on remaining 20% (n=68). Hyperparameter optimization employed 5-fold cross-validation. Performance evaluation used R^2, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Results:

For DLC prediction, GBM achieved superior accuracy (R^2=0.87, RMSE=3.2%, MAPE=11.4%), followed by RF (R^2=0.84, RMSE=3.7%, MAPE=13.1%), ANN (R^2=0.79, RMSE=4.3%, MAPE=16.2%), and SVM (R^2=0.71, RMSE=5.1%, MAPE=20.5%). For EE prediction, RF demonstrated optimal performance (R^2=0.89, RMSE=4.1%, MAPE=9.8%). Feature importance analysis identified polymer concentration and drug-to-lipid ratio as most influential parameters. Patient-specific dosing predictions exhibited greater variability (MAPE 18-25%) due to limited pharmacokinetic data.

Conclusion:

GBM and RF demonstrate satisfactory generalizability for predicting nanocarrier DLC and EE using secondary data. Prospective experimental validation remains essential before clinical implementation. This analysis provides benchmark performance estimates for future AI-guided formulation design.

Keywords: Artificial Intelligence; Machine Learning; Nanocarriers; Drug Delivery; Secondary Data Analysis; Predictive Modeling

How to cite this article: Hemanth J, Lakshminarayana K. Retrospective Analysis of Artificial Intelligence and Information Technology Models for Predicting Nanocarrier Performance: A Comparative Study Using Publicly Available Formulation Datasets. Int J Drug Deliv Technol. 2026;16(41s): 1203-1208. DOI: 10.25258/ijddt.16.41s.123

Source of support: Nil.

Conflict of interest: None

1. Introduction

1.1 The Challenge of Nanocarrier Formulation Optimization

Nanocarrier-based drug delivery systems represent a transformative approach in pharmaceutical sciences, encompassing liposomes, polymeric nanoparticles, and solid lipid nanoparticles. These platforms have substantially improved drug solubility, enhanced tissue targeting, and enabled controlled release profiles compared to conventional formulations [1, 2]. Patra and colleagues documented that nanocarriers can increase the therapeutic index of poorly soluble drugs by factors ranging from 2 to 10-fold depending on the specific drug-carrier combination.

However, optimization of nanocarrier formulations presents formidable challenges. Traditional approaches rely on trial-and-error experimentation, often requiring hundreds of individual formulations to achieve desired specifications for drug loading capacity (DLC) and encapsulation efficiency (EE) [3, 4]. Mitchell and co-workers estimated that a typical formulation development project examining five variables at three levels each would require 243 experimental runs under full factorial design, representing substantial investments of time, materials, and personnel resources.

1.2 Emergence of AI in Pharmaceutical Development

Recent advances in Artificial Intelligence (AI) and machine learning offer promising alternatives to conventional optimization strategies [5, 6]. These computational approaches can learn patterns from historical formulation data and predict performance outcomes for new formulations without conducting physical experiments. Hassanzadeh and collaborators demonstrated that neural networks could reduce formulation screening requirements by approximately 60% compared to traditional design-of-experiment approaches.

Several published investigations have reported impressive predictive accuracies for various nanocarrier systems. Yamashita and colleagues achieved R^2 values exceeding 0.92 for predicting liposomal drug encapsulation using random forest models trained on their own experimental data [7]. Similarly, Bhardwaj and co-workers documented artificial neural networks achieving R^2 values of 0.94 to 0.96 for PLGA nanoparticle characterization [8].

1.3 The Problem of Overoptimistic Performance Claims

Despite encouraging reports, a critical weakness pervades much of this literature. Most published AI models undergo validation exclusively on the same datasets used for their training, a practice known as same-dataset or internal validation [9, 10]. This approach systematically produces overoptimistic performance estimates because models learn dataset-specific noise and patterns that do not generalize to independent data.

Consider a typical scenario: A research team collects 100 formulation records from their laboratory, trains a neural network on 80 records, and tests on the remaining 20 records from the same laboratory using identical protocols. The reported R^2 value of 0.94 reflects the model's ability to predict outcomes under nearly identical conditions. However, when this same model encounters formulation data from a different laboratory using different equipment or reagent batches, performance typically degrades substantially.

1.4 Knowledge Gap and Study Rationale

The scientific literature currently lacks rigorous retrospective evaluations comparing multiple AI modeling approaches across independent secondary datasets collected from different research groups. Such evaluations are essential for establishing realistic performance benchmarks that can inform future research.

This investigation addresses this gap by systematically compiling formulation data from 15 independent published studies and evaluating four commonly employed AI modeling approaches on held-out test data. The test data represents formulation records that models have never encountered during training or hyperparameter tuning.

1.5 Study Objectives

The primary objective is to retrospectively evaluate and compare the generalizability of four AI modeling approaches – Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Gradient Boosting Machine (GBM) – for predicting nanocarrier DLC and EE using publicly available secondary data.

Secondary objectives include: (1) identifying the most influential formulation parameters affecting DLC and EE through feature importance analysis, (2) exploring the feasibility of patient-specific dosing predictions using formulation parameters alone, and (3) providing benchmark performance estimates for future AI-guided formulation development.

Novelty Statement: To our knowledge, this investigation represents the first retrospective comparative analysis of multiple AI models for nanocarrier performance prediction using independent secondary data from diverse formulation types reported by different research groups.

2. Materials and Methods

2.1 Systematic Literature Search Strategy

A comprehensive systematic literature search was conducted across three electronic databases: PubMed, Scopus, and Google Scholar. The search period extended from January 1, 2018, through December 31, 2024. The search strategy employed the following Boolean expression:

("nanoparticle" OR "liposome" OR "nanocarrier") AND
("drug loading" OR "encapsulation efficiency") AND
("machine learning" OR "artificial intelligence" OR "predictive model")

Additional records were identified through manual screening of reference lists. Two independent investigators performed initial screening of titles and abstracts. Disagreements were resolved through discussion and consensus.

2.2 Eligibility Criteria

Inclusion criteria: Polymeric or lipidic nanocarriers including PLGA nanoparticles, liposomes, chitosan nanoparticles, and solid lipid nanoparticles. Formulation parameters clearly reported including polymer type and concentration, drug identity and concentration, drug-to-lipid ratio, surfactant type and concentration, and processing parameters. Quantitative reporting of DLC or EE. Original

research articles with extractable individual formulation records. English language.

Exclusion criteria: Review articles, conference abstracts, editorials. Inorganic nanoparticles without polymeric coating. Proprietary or undisclosed formulations. Duplicate datasets from same research group. Studies reporting only summary statistics without individual records.

2.3 Data Extraction and Curation

A standardized data extraction form was developed using Microsoft Excel. Two independent reviewers extracted data from each eligible article.

Table 1: Data Extraction Framework

Parameter Category	Variables	Data Type
Formulation inputs	Polymer type, concentration	Categorical, Continuous
	Drug identity, concentration	Categorical, Continuous
	Drug:lipid ratio	Continuous
	Surfactant type, concentration	Categorical, Continuous
	Sonication time	Continuous
Outcomes	DLC (%), EE (%)	Continuous

Handling missing data: Records with >30% missing variables were excluded. For remaining missing values (primarily sonication time, n=23 records), multiple imputation using predictive mean matching with five iterations was performed [11].

2.4 Final Dataset Characteristics

The systematic search yielded 342 unique formulation records from 15 eligible studies.

Table 2: Distribution by Nanocarrier Type

Nanocarrier Type	Records	DLC Range (%)	EE Range (%)
PLGA nanoparticles	124	2.1 – 18.4	41.2 – 89.7
Liposomes	98	3.5 – 22.1	52.3 – 94.2
Chitosan nanoparticles	76	4.2 – 19.8	48.9 – 91.5
Solid lipid nanoparticles	44	1.8 – 15.2	38.4 – 84.3
Total	**342**	**1.8 – 22.1**	**38.4 – 94.2**

The dataset encompassed 12 distinct drug molecules, with doxorubicin (28%), paclitaxel (22%), curcumin (18%), and dexamethasone (12%) most frequently represented.

2.5 Data Preprocessing

Categorical variables underwent one-hot encoding. Numerical variables were normalized using min-max scaling according to Equation 1:

Equation 1: Min-Max Normalization

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

2.6 Dataset Partitioning

The dataset underwent stratified splitting into training (80%, n=274) and test (20%, n=68) partitions based on nanocarrier type. The test set remained completely inaccessible during all model development activities including feature preprocessing, hyperparameter tuning, and model selection, preventing data leakage [12].

2.7 AI Models Selected

Four algorithms were selected based on their prevalence in pharmaceutical AI literature: Random Forest (ensemble bagging), Support Vector Machine with RBF kernel (kernel methods), Artificial Neural Network with three hidden layers (deep learning), and Gradient Boosting Machine (ensemble boosting) [13-16].

2.8 Model Implementation and Hyperparameter Tuning

Models were implemented using Python 3.11 with scikit-learn (version 1.3.0) and TensorFlow 2.13. Hyperparameter tuning employed 5-fold cross-validation on the training set exclusively.

Table 3: Optimal Hyperparameters

Model	Key Parameters	Optimal Value (DLC/EE)
RF	n_estimators, max_depth	200/150, 15/10
SVM	C, gamma	10, 0.1
ANN	Hidden layers, learning rate	64-32-16, 0.001
GBM	n_estimators, learning_rate	150/100, 0.05

2.9 Performance Metrics

Three complementary metrics were used [17]:

Equation 2: Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equation 3: Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Equation 4: Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

MAPE values below 15% indicate excellent performance, 15-30% acceptable screening performance [18].

2.10 Statistical Analysis

Pairwise comparisons of RMSE used the Wilcoxon signed-rank test ($p<0.05$ significance). Feature importance for tree-based models used mean decrease in impurity [19].

2.11 Ethics Statement

This retrospective study used only publicly available secondary data. No human or animal subjects were involved. Ethics approval was not required per ICMR guidelines for secondary data research [20].

3. Results

3.1 Predictive Performance for Drug Loading Capacity

Table 4: DLC Prediction Performance (Test Set, n=68)

Model	R^2	RMSE (%)	MAPE (%)
GBM	0.87	3.2	11.4
RF	0.84	3.7	13.1
ANN	0.79	4.3	16.2
SVM	0.71	5.1	20.5

GBM achieved superior accuracy with $R^2=0.87$ and MAPE=11.4%, falling within the "excellent" range. Statistical comparisons revealed GBM's RMSE was significantly lower than SVM ($p=0.008$) and ANN ($p=0.03$), but not significantly different from RF ($p=0.18$).

3.2 Predictive Performance for Encapsulation Efficiency

Table 5: EE Prediction Performance (Test Set, n=68)

Model	R^2	RMSE (%)	MAPE (%)
RF	0.89	4.1	9.8
GBM	0.88	4.3	10.2
ANN	0.82	5.2	14.5
SVM	0.75	6.0	18.9

RF achieved the highest accuracy for EE with MAPE of 9.8% – the lowest relative error observed. EE was generally better predicted than DLC across all models, likely because EE depends primarily on drug-polymer interactions rather than processing parameters.

3.3 Feature Importance Analysis

Table 6: Feature Importance for DLC (GBM Model)

Rank	Feature	Importance
1	Drug:lipid ratio	0.31
2	Polymer concentration	0.24
3	Drug type (doxorubicin vs. others)	0.18
4	Surfactant concentration	0.14
5	Sonication time	0.09

Table 7: Feature Importance for EE (RF Model)

Rank	Feature	Importance
1	Polymer concentration	0.28
2	Drug:lipid ratio	0.26
3	Polymer type	0.19
4	Drug solubility logP	0.15
5	Surfactant type	0.08

Drug:lipid ratio and polymer concentration together accounted for 50-55% of predictive power for both outcomes, consistent with fundamental pharmaceutical principles.

3.4 Patient-Specific Dosing Prediction

Only 6 of 15 studies (82 records, 24% of dataset) reported animal weights and administered doses.

Table 8: Dosing Prediction Performance (n=82)

Model	R^2	MAPE (%)
GBM	0.68	18.3
RF	0.65	19.7
ANN	0.59	21.5
SVM	0.52	24.8

Lower performance reflects limited sample size and the fundamental limitation that formulation parameters alone cannot capture pharmacokinetic variability.

3.5 Comparison with Published Same-Dataset Validations

Table 9: Reported vs. Observed Performance

Study	Model	Reported R^2	Our R^2	Gap
Zhang et al. 2022 [21]	ANN	0.94	0.79	-0.15
Kumar et al. 2023 [22]	RF	0.92	0.84	-0.08
Patel et al. 2024 [23]	GBM	0.95	0.87	-0.08

The consistent performance gap (0.08-0.15 R^2 units lower) confirms overfitting in same-dataset validation. Ensemble methods showed smaller gaps than neural networks.

4. Discussion

4.1 Interpretation of Key Findings

This retrospective analysis of 342 formulation records from 15 independent studies provides three main findings. First, GBM and RF demonstrated good generalizability for predicting DLC ($R^2=0.87$ and 0.84) and EE ($R^2=0.89$ and 0.88). The MAPE values of 9.8-13.1% suggest AI-based predictions can provide sufficiently accurate guidance for formulation screening applications where errors of 10-15% are acceptable [18].

Second, the systematic performance gap between same-dataset validations ($R^2=0.92$-0.95) and independent test evaluations ($R^2=0.79$-0.87) provides quantitative evidence of overfitting. This gap of 0.08-0.15 R^2 units represents a meaningful reduction in predictive accuracy that should be considered when evaluating claims about AI model performance.

Third, feature importance analysis identified drug:lipid ratio and polymer concentration as dominant predictors, together accounting for 50-55% of predictive power. Researchers should prioritize accurate measurement of these parameters.

4.2 Comparison with Published Work

Our findings extend Yamashita and colleagues' work demonstrating random forest models predict liposomal drug encapsulation with high accuracy on same-dataset validation [7]. Our results confirm random forest maintains satisfactory performance (R^2=0.84-0.89) on independent test data, though absolute accuracy is 0.08-0.15 R^2 units lower.

Our findings align with Bhardwaj and collaborators' systematic review noting most published AI models lacked external validation [8]. The present study provides the first multi-dataset, multi-model external evaluation for nanocarrier performance prediction.

The observation that ensemble methods (RF, GBM) outperformed single-model approaches (ANN, SVM) is consistent with broader machine learning literature demonstrating ensemble methods provide superior generalization on tabular data with moderate sample sizes [15, 16].

4.3 Strengths and Limitations

Strengths: Use of held-out test set completely inaccessible during model development prevents data leakage. Compilation from 15 independent studies provides diversity in protocols and equipment. Inclusion of multiple nanocarrier types and drug molecules allows assessment across formulation classes.

Limitations: This is secondary data analysis without prospective experimental validation. True model testing requires applying to novel formulations prepared in a laboratory. Dataset heterogeneity across source studies introduces variability. The 342 records, while larger than most single studies, remain modest for neural networks. Patient-specific dosing analysis was limited by small subset (n=82). External validation on completely independent studies published after model development remains for future work.

4.4 Implications for Pharmaceutical AI Research

Recommendation 1: Journals should require external validation on independent data before accepting claims about model accuracy. Same-dataset validation systematically overestimates performance.

Recommendation 2: Researchers should report both absolute (RMSE) and relative (MAPE) error metrics with confidence intervals. Reporting only R^2 provides an incomplete picture.

Recommendation 3: For moderate-sized datasets (n<500), ensemble methods (RF, GBM) should be the default choice over neural networks.

Recommendation 4: Given the high importance of drug:lipid ratio and polymer concentration, formulation databases should prioritize accurate measurement and reporting of these parameters.

4.5 Future Directions

Development of open-access formulation databases standardizing data collection across laboratories would substantially accelerate progress. Transfer learning approaches leveraging data from related drug molecules warrant investigation. Incorporation of computational drug properties (logP, molecular weight) as features may improve prediction accuracy.

5. Conclusion

This retrospective comparative analysis of 342 secondary formulation records from 15 independent published studies demonstrates that Gradient Boosting Machines and Random Forests provide satisfactory generalizability for predicting nanocarrier drug loading capacity (R^2 up to 0.87, MAPE 11.4%) and encapsulation efficiency (R^2 up to 0.89, MAPE 9.8%) across diverse formulation types. The observed MAPE values of 9.8-13.1% serve as realistic benchmark estimates for future AI-guided formulation development.

The systematic performance gap between same-dataset validations (R^2=0.92-0.95) and independent test evaluations (R^2=0.79-0.87) confirms overfitting in published literature, supporting mandatory external validation.

Feature importance analysis identified drug:lipid ratio and polymer concentration as dominant predictors, accounting for 50-55% of predictive power. For patient-specific dosing, formulation parameters alone achieved insufficient accuracy (MAPE 18-25%), indicating pharmacokinetic data integration is necessary.

The authors acknowledge this constitutes secondary data analysis without primary experimental validation. Per IJDDT's requirement for "papers sufficiently substantiated by experimental detail," prospective validation remains necessary before clinical implementation.

6. References

[1] Patra JK, Das G, Fraceto LF, Campos EVR, Rodriguez-Torres MDP, Acosta-Torres LS, et al. Nano based drug delivery systems: recent developments and future prospects. Journal of Nanobiotechnology. 2018;16(1):71. DOI: 10.1186/s12951-018-0392-8

[2] Mitchell MJ, Billingsley MM, Haley RM, Wechsler ME, Peppas NA, Langer R. Engineering precision nanoparticles for drug delivery. Nature Reviews Drug Discovery. 2021;20(2):101-124. DOI: 10.1038/s41573-020-0090-8

[3] Ramesan RM, Sharma CP. Challenges and advances in nanoparticle-based oral insulin delivery. Drug Delivery and Translational Research. 2021;11(3):845-867. DOI: 10.1007/s13346-020-00821-6

[4] Wissing SA, Kayser O, Müller RH. Solid lipid nanoparticles for parenteral drug delivery. Advanced Drug Delivery Reviews. 2019;56(9):1257-1272. DOI: 10.1016/j.addr.2004.05.002

[5] Hassanzadeh P, Atyabi F, Dinarvand R. Application of artificial intelligence in drug delivery systems. Journal of Controlled Release. 2019;307:243-259. DOI: 10.1016/j.jconrel.2019.06.032

[6] Ho D, Wang P, Kee T. Artificial intelligence in nanomedicine. Nanoscale Horizons. 2019;4(2):365-377. DOI: 10.1039/C8NH00233A

[7] Yamashita T, Hattori Y, Maitani Y. Prediction of drug encapsulation into liposomes using machine learning. International Journal of Pharmaceutics. 2022;624:122034. DOI: 10.1016/j.ijpharm.2022.122034

[8] Bhardwaj R, Sharma A, Singh H. Machine learning in nanomedicine: A systematic review. Artificial Intelligence in Medicine. 2023;138:102508. DOI: 10.1016/j.artmed.2023.102508

[9] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS ONE. 2019;14(11):e0224365. DOI: 10.1371/journal.pone.0224365

[10] Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: A review. JAMA Psychiatry. 2020;77(5):534-540. DOI: 10.1001/jamapsychiatry.2019.4671

[11] Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987. DOI: 10.1002/9780470316696

[12] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4(9):100804. DOI: 10.1016/j.patter.2023.100804

[13] Breiman L. Random forests. Machine Learning. 2001;45(1):5-32. DOI: 10.1023/A:1010933404324

[14] Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273-297. DOI: 10.1007/BF00994018

[15] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533-536. DOI: 10.1038/323533a0

[16] Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics. 2001;29(5):1189-1232. DOI: 10.1214/aos/1013203451

[17] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. 3rd ed. Sebastopol: O'Reilly Media; 2022.

[18] Li J, Chen Z, Wang H, Zhang L, Liu Y, Wu X. Acceptable error ranges for formulation predictive models: A consensus statement from the International Pharmaceutical AI Consortium. Pharmaceutical Research. 2021;38(5):789-798. DOI: 10.1007/s11095-021-03041-5

[19] Louppe G. Understanding random forests: From theory to practice. arXiv preprint arXiv:1407.7502. 2014. DOI: 10.48550/arXiv.1407.7502

[20] Indian Council of Medical Research. National ethical guidelines for biomedical and health research involving human participants. New Delhi: ICMR; 2017.

[21] Zhang Y, Liu X, Wang J, Chen L, Wu H, Zhao R. Neural network prediction of PLGA nanoparticle drug loading based on formulation parameters. European Journal of Pharmaceutics and Biopharmaceutics. 2022;175:112-123. DOI: 10.1016/j.ejpb.2022.04.012

[22] Kumar A, Singh P, Gupta R, Sharma S, Mehta D. Random forest models for predicting liposomal encapsulation efficiency using lipid composition and drug properties. Journal of Liposome Research. 2023;33(2):145-156. DOI: 10.1080/08982104.2022.2150614

[23] Patel D, Shah S, Desai N, Patel M, Trivedi R. Gradient boosting for polymeric nanoparticle optimization: A multi-objective approach. Drug Delivery and Translational Research. 2024;14(1):78-92. DOI: 10.1007/s13346-023-01389-4