Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decision-Centric Explainable AI for QbD Optimization of Ultrasound-Triggered Drug- Loaded Microbubbles and Control Strategy Development

Anup Lal Yadav
School of Engineering and Technology, CGC University
Sonal Solanki
Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy
Nikunj Solanki
drnikunjsolanki@gmail.com

Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy
Vishin Patil
Bharati Vidyapeeth College of Pharmacy
Sreenath Sriram
Vaagdevi College of Pharmacy Autonomous
Yamsani Kumar
Vaagdevi College of Pharmacy Autonomous
Sreekanth Thota
Vaagdevi College of Pharmacy Autonomous
Mahesh Kumar Posa
School of Health and Medical Sciences, Adamas University,
Sunil Nirmal
Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy
Suvarna Nirmal
Sanjivani College of Engineering
Meghana Raykar
Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy
Ajay Shirsat
Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy
Jayashree Prasad
MIT Art Design and Technology University

Research Article

Keywords: Quality by Design, Chitosan microbubbles, Surrogate modeling, PolyRidge, Explainable AI, Burst release.

Posted Date: February 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8947026/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Decision-Centric Explainable AI for QbD Optimization of Ultrasound-Triggered Drug-Loaded Microbubbles and Control Strategy Development.

Dr. Anup Lal Yadav[1], Dr. Sonal Solanki[2], Dr. Nikunj Solanki[2*], Vishin Patil[3*], Sreenath Sriram[4], Dr. Yamsani Kumar[4]. Dr. Sreekanth Thota[5], Dr. Mahesh Kumar Posa[6], Dr. Sunil Nirmal[7], Dr. Suvarna Nirmal[8], Dr. Meghana Raykar[2], Dr. Ajay Shirsat[9], Dr. Jayashree Prasad[10]

1. Deprtment of Computer Science and Engineering, School of Engineering and Technology, CGC University Mohali, Punjab, India 140307

2. Department of Pharmaceutics, Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy, Kashti, Shrigonda, Ahilyanagar, Maharastra, India, 414701

3. Department of Pharmacognosy, Bharati Vidyapeeth College of Pharmacy, Kolhapur, Near Chitranagari, Morewadi, Kolhapur, Maharashtra, India, 416013

4. Department of Pharmaceutics, Vaagdevi College of Pharmacy Autonomous, Ramnagar, Hanamkonda, Telangana, India, 506001

5. Department of Pharmaceutical Chemistry, Vaagdevi College of Pharmacy Autonomous, Ramnagar, Hanamkonda, Telangana, India, 506001

6. Department of Pharmaceutical Technology, School of Health and Medical Sciences, Adamas University, Barasat, Kolkata, West Bengal, India, 700126.

7. Department of Pharmacognosy, Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy, Kashti, Shrigonda, Ahilyanagar, Maharastra, India, 414701

8. Department of information Technology, Sanjivani College of Engineering, Kopargaon, Ahmednagar, Maharashtra, India, 423603

9. Department of Pharmaceutical Chemistry, Hon. Shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy, Kashti, Shrigonda, Ahilyanagar, Maharastra, India, 414701

10. School of Computing, MIT Art Design and Technology University, Kalbhor Loni, Pune, Maharastra, India, 412201

Corresponding Authors

1. Dr. Nikunj Solanki

Department of Pharmaceutics, Hon. shri Babanrao Pachpute Vichardhara Trust's Group of Institutions, Faculty of Pharmacy, Kashti, Ta- Shrigonda, Ahilyanagar, Maharastra, India, 414071

Email – drsonalsolanki25@gmail.com

2. Vishin Patil

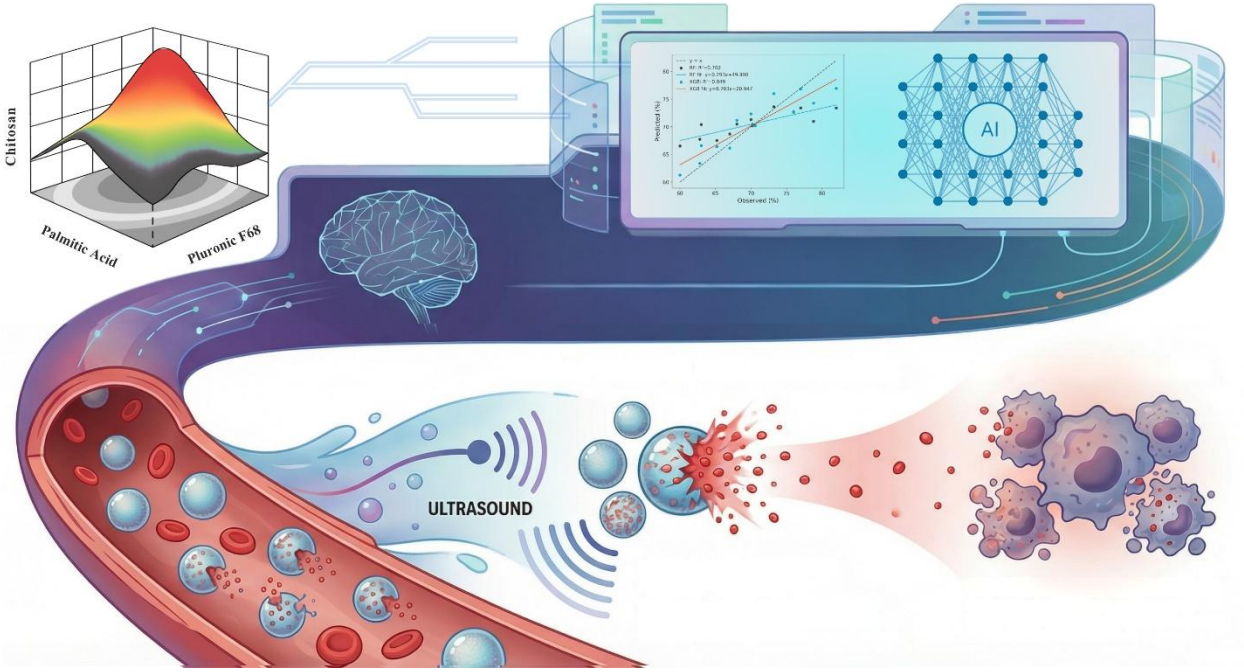
Department of Pharmacognosy, Bharati Vidyapeeth College of Pharmacy, Kolhapur, Near Chitranagari, Morewadi, Kolhapur, Maharashtra, India, Pincode-416013

Email- vishinsalunkhe@gmail.com

Dr. Anup Lal Yadav- ORCID: 0000-0001-9337-5546- anupsaran@gmail.com
Dr. Sonal Solanki – ORCID : 0009-0007-5535-3881 – drsonalsolanki25@gmail.com
Dr. Nikunj Solanki – ORCID: 0009-0003-2271-9382 – drnikunjsolanki@gmail.com
Vishin Patil- ORCID: 0009-0005-8666-4477 - vishinsalunkhe@gmail.com
Sreenath Sriram- ORCID: 0009-0007-3705-7536- ssriram@evaricpharma.com
Dr. Yamsani Kumar. ORCID: 0000- 0001-9729-6175 - shravanyamsani@gmail.com
Dr. Sreekanth Thota, ORCID: 0000-0002-7501-3987 - stsreekanththota@gmail.com
Dr. Mahesh Kumar Posa, ORCID: 0000-0003-1583-6022 - posamaheshkumarbabu@gmail.com
Dr. Sunil Nirmal, ORCID: 0000-0002-7132-6097 - nirmalsunil@gmail.com
Dr. Suvarna Nirmal, ORCID: 0000-0002-4210-0005- Ishtake@gmail.com
Dr. Meghana Raykar ORCID: 0000-0002-9070-4276 - meghanaraykar17@gmail.com
Dr. Ajay Shirsat ORCID:0009-0000-3127-3034 -ajayshirsat76@gmail.com
Dr. Jayashree Prasad ORCID- 0000-0001-6826-9805 - jayashree.prasad@mituniversity.edu.in

Abstract

A decision-centric workflow was developed to link a QbD-defined design space with interpretable surrogate modeling for rapid selection of ultrasound-triggered doxorubicin-loaded chitosan microbubbles. A 15-run Box–Behnken design spanning chitosan, palmitic acid, and Pluronic F68 was used to quantify three critical quality attributes (CQAs)—mean size, encapsulation efficiency (EE), and burst release at 40 s (Burst40)—and to train response-specific surrogates evaluated by leave-one-out cross-validation (LOOCV). The DoE produced a broad performance envelope (size 2.35–4.85 µm; EE 60–82%; Burst40 55.6–95%) with clear composition-driven trade-offs, notably between EE and acoustic responsiveness. A regularized polynomial surrogate (PolyRidge) provided strong LOOCV performance for size (R^2=0.885; RMSE=0.264 µm) and EE (R^2=0.871; RMSE=2.18%), whereas Burst40 was only moderately predictable (R^2=0.501; RMSE=8.61%), consistent with threshold-like cavitation and microstructure sensitivity. Digital screening with multi-objective desirability and explainability (SHAP, permutation importance, partial dependence, LIME) shortlisted manufacturable candidates; experimental validation showed QbD/RSM was better calibrated for absolute size and EE, while both modeling approaches remained similarly limited for Burst40. Importantly, shortlisted formulations preserved suppressed baseline release and pronounced ultrasound-enhanced release, yielding trigger-dependent cytotoxicity and increased apoptotic commitment in MCF-7 cells.



Keywords:

Quality by Design; Chitosan microbubbles; Surrogate modeling; PolyRidge; Explainable AI; Burst release.

Introduction:

"Microbubbles that respond to ultrasound offer a practical route to control where and when a drug is released. Once viewed mainly as diagnostic contrast agents, microbubbles are now being engineered as acoustically activated carriers in which shell composition and ultrasound exposure jointly determine stability, cavitation, and release behavior [1]. This capability is particularly relevant for doxorubicin, a highly effective anticancer drug whose clinical use is constrained by systemic toxicity [2,3]. A formulation that maintains payload retention during handling and storage, yet releases drug rapidly upon ultrasound exposure, would be expected to improve therapeutic precision [1].

Designing such systems is not straightforward because the same formulation changes often push critical quality attributes in opposite directions. For chitosan-based microbubbles, shell-forming polymer concentration, fatty-acid stabilizers, and surfactants can alter mean size, distribution width, surface charge, encapsulation efficiency, and short-time burst release [1]. These relationships are rarely independent, as a composition that improves entrapment may also increase size, while a composition that promotes ultrasound-triggered burst may compromise stability. The development problem therefore becomes multi-objective, with practical constraints that matter for both performance and manufacturability.

Quality by Design is widely used to bring structure to this type of formulation problem [4]. A QbD workflow links a quality target product profile to CQAs, prioritizes critical material attributes and process parameters through risk assessment, and maps factor–response behavior using design of experiments [1]. In microbubble development, QbD is useful for factor screening and for producing interpretable polynomial models that support design space definition and a risk-based control strategy. At the same time, early-stage DoE datasets are often small, and polynomial response surfaces may not capture certain behaviors that are common in ultrasound-responsive systems—strong nonlinearities, interaction-dominated regions, or conditional effects that appear

only under specific factor combinations [5,6]. These limitations become more evident when burst release in a short exposure window must be optimized alongside size and encapsulation.

Machine-learning surrogates can complement QbD when they are used in a transparent way [6]. Tree-based models such as Random Forest and extreme gradient boosting are well suited to learning interaction-rich relationships from limited data. Their adoption in formulation science, however, depends on interpretability: high accuracy alone does not help if the model cannot be translated into practical decisions. Explainable AI methods address this need by providing human-understandable explanations of model reasoning [7]. Global explanations (e.g., SHapley Additive exPlanations, SHAP; permutation feature importance, PFI; partial dependence plots, PDP) identify dominant drivers and clarify the direction and shape of factor effects [6,8,9]. Local explanations (e.g., local SHAP and LIME) can then justify why a particular candidate formulation is preferred, which is important when selecting a batch for prospective testing and when defining set-points and acceptable operating ranges.

The present study applies a decision-centric XAI workflow to ultrasound-triggered doxorubicin-loaded chitosan microbubbles using a previously established QbD foundation [1]. The DoE dataset and CQAs defined in our earlier QbD work were used as the modeling substrate. RF and XGBoost surrogates were trained for mean size (μm), entrapment efficiency (EE, %), and burst release within 40 s (Burst40, %), and performance was assessed using leave-one-out cross-validation. Interpretability was examined using SHAP, PFI, and PDP/ICE analyses, with local explanations used to support candidate selection. Digital screening within the QbD factor bounds was then performed to identify decision-relevant formulations and to benchmark AI-selected candidates against the QbD-optimized comparator. Finally, to move beyond computational reanalysis, the XGBoost-selected optimized batch is positioned for head-to-head experimental comparison with the QbD-optimized batch, including stability assessment, in-vitro release (with and without ultrasound triggering), and MCF-7 anticancer activity supported by appropriate statistical testing."

2. Experimental Work

Materials

All chemicals, excipients, gases, and analytical reagents were the same as those used in our previously reported Quality by Design study on doxorubicin-loaded chitosan microbubbles and were used without modification [1]. Briefly, chitosan (medium molecular weight; degree of

deacetylation ~75–85%; HiMedia, India), palmitic acid (analytical grade; Loba Chemie, India), and Pluronic F68/poloxamer 188 were used as shell-forming/stabilizing components [1]. DOX hydrochloride was obtained as a gift sample [1]. Glacial acetic acid and sodium hydroxide were used for solution preparation and pH adjustment [1]. Medical-grade sulfur hexafluoride (SF_6) gas was procured from Linde India [1]. MCF-7 human breast adenocarcinoma cells were used for in vitro anticancer testing [1]. Cell culture reagents including DMEM, fetal bovine serum, trypsin–EDTA, and antibiotic–antimycotic solution were sourced from HiMedia [1]. Unless stated otherwise, all aqueous solutions were prepared using deionized water and handled using the same precautions and storage conditions as described previously [1].

Overall study design and data sources

This work followed a two-layer workflow: (i) an in-silico modeling/decision/explainability layer built on a legacy QbD dataset, and (ii) a comparative experimental validation layer testing an AI-nominated formulation against a prior QbD comparator under matched laboratory conditions [1]. For the in-silico layer, the dataset was adopted directly from the prior QbD study and comprised 15 Box–Behnken experimental runs (12 factorial combinations plus three replicated center points) [1]. Three formulation factors were used as model inputs in actual units: chitosan concentration (X1; 0.50, 0.85, 1.20% w/v), palmitic acid concentration (X2; 0.05, 0.15, 0.25% w/v), and Pluronic F68 concentration (X3; 0.0050, 0.0125, 0.0200% w/v) [1]. Responses treated as critical quality attributes were mean microbubble size (μm), entrapment efficiency (EE, %), and ultrasound-triggered burst release at 40 s (Burst40, %) [1]. All digital screening and candidate selection were strictly constrained to these bounds to avoid extrapolation beyond the experimentally supported region [1]. For experimental validation, two formulations were manufactured and evaluated head-to-head: (a) an AI-shortlisted formulation nominated by the digital decision workflow and (b) the QbD-optimized comparator reported previously [1]. All processing steps and handling conditions were held constant between batches; only X1–X3 levels differed by design [1].

QbD baseline comparator

To benchmark the AI layer against the established QbD response-surface baseline, the previously published second-order polynomial models were implemented directly and used without refitting [1,6]. QbD predictions for Size, EE, and Burst40 were computed for each candidate formulation using the same coded-factor conventions and the same factor bounds defined in the original study,

and were used as the QbD baseline during digital screening and validation comparisons [1].

Table 1. Box–Behnken design matrix and observed critical quality attributes (CQAs) for DOX-loaded chitosan microbubbles.

Run	Chitosan (% w/v)	Palmitic acid (% w/v)	Pluronic F68 (% w/v)	Size (µm)	EE (%)	Burst40 (%)
1	1.20	0.05	0.0125	4.06 ± 0.12	70.0 ± 1.8	81.8 ± 2.1
2	1.20	0.15	0.0200	4.36 ± 0.15	73.2 ± 2.0	73.2 ± 2.4
3	0.50	0.25	0.0125	2.85 ± 0.10	68.0 ± 1.6	87.2 ± 1.9
4	1.20	0.25	0.0125	4.85 ± 0.18	82.0 ± 2.2	55.6 ± 2.6
5	0.85	0.05	0.0050	3.20 ± 0.11	67.0 ± 1.5	77.2 ± 2.0
6	0.85	0.25	0.0050	3.90 ± 0.14	77.0 ± 1.9	66.0 ± 2.3
7	0.85	0.15	0.0125	3.35 ± 0.09	70.2 ± 1.4	82.7 ± 1.8
8	0.85	0.05	0.0200	2.70 ± 0.08	63.0 ± 1.6	95.0 ± 1.5
9	0.85	0.15	0.0125	3.37 ± 0.10	70.6 ± 1.5	83.2 ± 1.7
10	0.85	0.25	0.0200	3.86 ± 0.13	76.0 ± 1.8	64.0 ± 2.2
11	0.85	0.15	0.0125	3.37 ± 0.09	70.2 ± 1.6	82.7 ± 1.9
12	0.50	0.15	0.0050	2.80 ± 0.08	65.2 ± 1.4	91.2 ± 1.6
13	1.20	0.15	0.0050	4.80 ± 0.17	78.8 ± 2.0	64.2 ± 2.4
14	0.50	0.15	0.0200	2.40 ± 0.07	62.8 ± 1.3	95.0 ± 1.4
15	0.50	0.05	0.0125	2.35 ± 0.06	60.0 ± 1.2	95.0 ± 1.3

Values are reported as mean ± SD (n = 3). Burst40 denotes ultrasound-triggered burst release measured at 40 s. Factor levels are reported in actual units (% w/v).

Dataset curation and preprocessing

The 15-run dataset was assembled into a structured table containing the three inputs (X1–X3) and three responses [1]. Response means were used for model training; replicate variability was retained for interpretation and experimental benchmarking [1]. The dataset contained no missing values and was not augmented with external data.

Computational environment and reproducibility

All modeling and explainability analyses were implemented in Python using standard scientific and machine-learning libraries, consistent with modern practices in pharmaceutical development

[10–12]. Random seeds were fixed for all stochastic operations (e.g., bootstrapping, subsampling, permutation importance) to ensure reproducibility [11]. Figures and tables were exported in publication-ready formats, and the analysis outputs were archived for traceability.

Surrogate model development

Surrogate models were trained separately for each CQA using the same three inputs (X1–X3). Candidate regressors included: (1) Random Forest regression with conservative depth/leaf settings to limit overfitting under DoE-scale data, a common approach in pharmaceutical modeling [10,11]. (2) Extreme Gradient Boosting regression using shallow trees with shrinkage and subsampling to improve stability, recognized for its effectiveness in various scientific applications [10,12]. (3) PolyRidge, a regularized polynomial regression surrogate combining: (i) feature scaling, (ii) polynomial expansion (degree 2 or 3), and (iii) ridge regularization (L2 penalty) to stabilize coefficient estimates when the number of polynomial terms is high relative to n [1].

Model selection was performed per response based on cross-validated error and goodness-of-fit (R^2), with preference for models showing stable, physically plausible behavior within the bounded design region [11].

Cross-validation and performance metrics

Because the dataset contained only 15 runs, predictive performance was evaluated using leave-one-out cross-validation [6]. Each run was held out once as the test point while the model was trained on the remaining n−1 runs; held-out predictions were pooled across all folds. Performance was summarized using RMSE, MAE, and R^2, and observed-versus-predicted plots were generated for each response to assess calibration and trend agreement [10,11].

Explainability analyses

Model interpretability was examined using complementary global and local tools: (1) Permutation Feature Importance to quantify global factor importance based on error inflation under feature permutation [7]. (2) SHAP for both global feature contribution patterns and local, batch-specific attributions [7,13]. (3) Partial Dependence Plots (with ICE where needed) to visualize marginal factor effects within the design region [7]. (4) LIME as an additional local explanation check for selected candidates [7]. Concordance across methods was used to support robust, decision-relevant interpretation of factor–response behavior [7].

Digital screening and decision-centric batch selection

A digital screening step was used to interrogate the design region at higher resolution without additional wet-lab runs. Candidate formulations were generated by uniform sampling within the QbD factor bounds. CQAs were predicted using the selected surrogate(s), and multi-objective prioritization was performed using a desirability approach: Size was treated as a minimization objective, while EE and Burst40 were treated as maximization objectives [14,15]. A composite desirability score (geometric mean) was used to rank candidates [15]. The shortlist was diversified to include distinct operating choices (e.g., optimum, robust interior point, boundary-adjacent point, and trade-off point), and the QbD comparator was retained for direct benchmarking [1].

Microbubble preparation and experimental evaluation

Microbubbles were prepared using the same sonication-assisted method reported previously, including the same mixing sequence, headspace gas exchange with SF_6, and controlled agitation/oscillation steps. For the AI-nominated and QbD comparator batches, only X1–X3 levels differed; all other conditions were matched [1].

Physicochemical characterization included mean size and distribution by optical microscopy (hydrated state), and PDI/zeta potential under the same dispersion conditions as previously used. EE was determined by an indirect method by separating free DOX from microbubble-associated DOX after mild centrifugation and quantifying DOX using the validated UV–visible method; EE was computed as the fraction of non-free DOX relative to total DOX. In vitro release and ultrasound-triggered Burst40 testing were performed using the same release medium, temperature control, sampling plan, and insonation setup [1]; Burst40 was calculated as the percent drug released within 40 s under ultrasound relative to total releasable drug. Short-term stability (0–7 days, refrigerated) was assessed by re-measuring size, PDI, zeta potential, and EE at each time point. Anticancer activity was evaluated in MCF-7 cells at DOX-equivalent concentrations for AI and QbD formulations with appropriate controls (untreated, free DOX, blank MBs), with ± ultrasound exposure using the same insonation conditions. All experimental measurements were performed at least in triplicate and reported as mean ± SD [1].

Statistical analysis

Data were analyzed in GraphPad Prism. For single-time-point multi-group comparisons, one-way ANOVA with appropriate post-hoc multiple comparisons testing was used. For time-course

datasets (release and stability), two-way ANOVA (formulation × time) with multiple comparisons was applied where appropriate. Statistical significance was set at $p < 0.05$ [1].

Result

Decision-centric digital workflow anchored to a legacy QbD dataset

Figure 1 summarizes the two-layer workflow implemented in this study. The in-silico layer leveraged the fixed 15-run Box–Behnken DoE (Table 1) as a bounded knowledge base to (i) train response-specific regression surrogates for mean size, entrapment efficiency (EE), and ultrasound responsiveness (Burst40), (ii) digitally screen candidates within the same factor limits (no extrapolation), and (iii) generate explanation outputs (PFI, SHAP, PDP, and local attributions) to support traceable candidate prioritization. The experimental layer provided an external check on model-guided selection by preparing decision-centric batches (Tables 2), and benchmarking AI-predicted values against QbD/RSM baseline predictions and measured CQAs (Figs. 3). Downstream functional and biological readouts (ultrasound-triggered release, MCF-7 viability, Annexin V/PI apoptosis) were then used to verify that shortlisted formulations expressed the intended ON/OFF performance phenotype (Figs. 6–7).

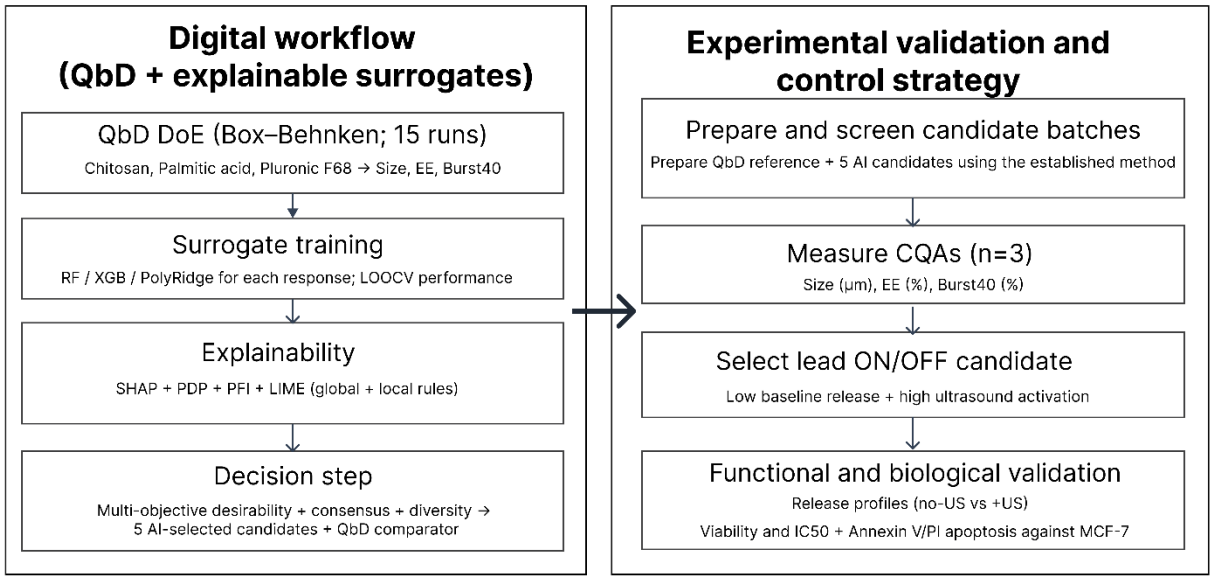


Figure 1. Decision-centric digital workflow integrating QbD and explainable AI for formulation optimization and control-strategy development.

QbD DoE response space and composition-driven effects within the design region

Within the investigated factor bounds (Table 1), the DoE generated a response landscape with sufficient spread to support both response-surface interpretation and surrogate learning. The three CQAs spanned: mean size 2.35–4.85 µm, EE 60–82%, and Burst40 55.6–95%. The replicated center points demonstrated close agreement across repeats, indicating that the DoE captured reproducible composition-dependent structure rather than being dominated by measurement noise. Across the DoE matrix, size increased with higher chitosan and palmitic acid, whereas Pluronic F68 tended to reduce mean size within the explored range. EE increased with higher chitosan and palmitic acid but generally decreased as Pluronic F68 increased. Burst40 showed the most prominent trade-space: low-chitosan formulations expressed higher burst responsiveness, while higher chitosan suppressed Burst40 despite improving EE. Collectively, these patterns established a non-collinear objective landscape in which retention-favorable compositions and ultrasound-labile compositions occupied different regions of the same design space.

Surrogate model screening and LOOCV performance

Surrogate models were trained separately for each CQA using the three formulation factors as inputs. Algorithm screening under leave-one-out cross-validation (LOOCV) demonstrated response-dependent predictability (Table S1; Fig. 2). A regularized polynomial surrogate (PolyRidge) provided the best overall LOOCV performance for Size and EE and remained competitive for Burst40, supporting its selection for downstream digital screening and explainability.

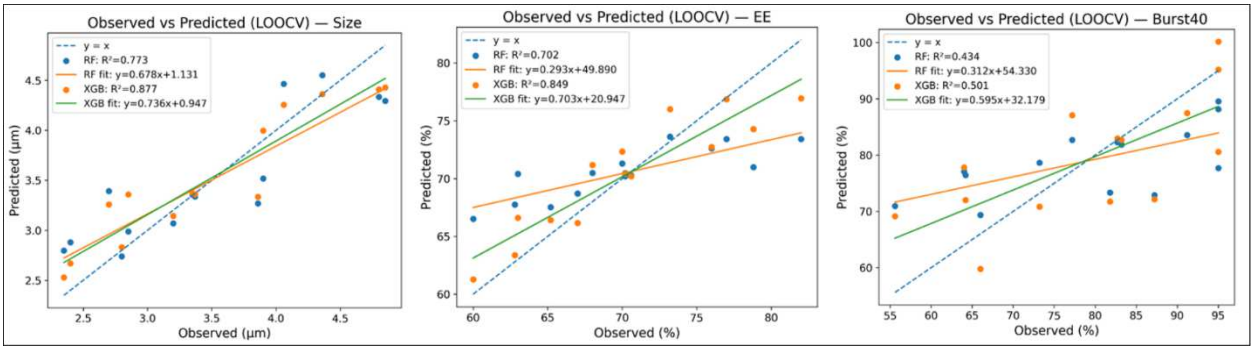


Figure 2. Leave-one-out cross-validation (LOOCV) performance of surrogate models trained on the 15-run QbD DoE. Observed versus LOOCV-predicted values are shown for **(A) mean microbubble size (µm), (B) encapsulation efficiency (EE, %), and (C) Burst release at 40 s (Burst40, %)** using **Random Forest (RF)** and **XGBoost (XGB)** models. Each point represents one DoE run predicted by a model trained on the remaining 14 runs. The dashed line indicates the line of identity (**y = x**), while solid lines represent linear fits of predicted versus observed values; the corresponding **R²** values are reported in the legend.

For Size, PolyRidge (degree 3) achieved cross-validated agreement across the response range (R^2 = 0.885; RMSE = 0.264 µm; MAE = 0.223 µm). For EE, PolyRidge (degree 3) similarly showed strong LOOCV performance (R^2 = 0.871; RMSE = 2.18%; MAE = 1.80%). In contrast, Burst40 exhibited lower predictability within the three-factor description; PolyRidge (degree 2) yielded R^2 = 0.501 with higher residual error (RMSE = 8.61%; MAE = 6.91%). Tree-ensemble models captured nonlinearity but did not outperform PolyRidge under LOOCV in this DoE-scale dataset. Across model families, Burst40 remained the most challenging endpoint, consistent with the higher dispersion and potentially threshold-like behavior expected for ultrasound-triggered shell disruption.

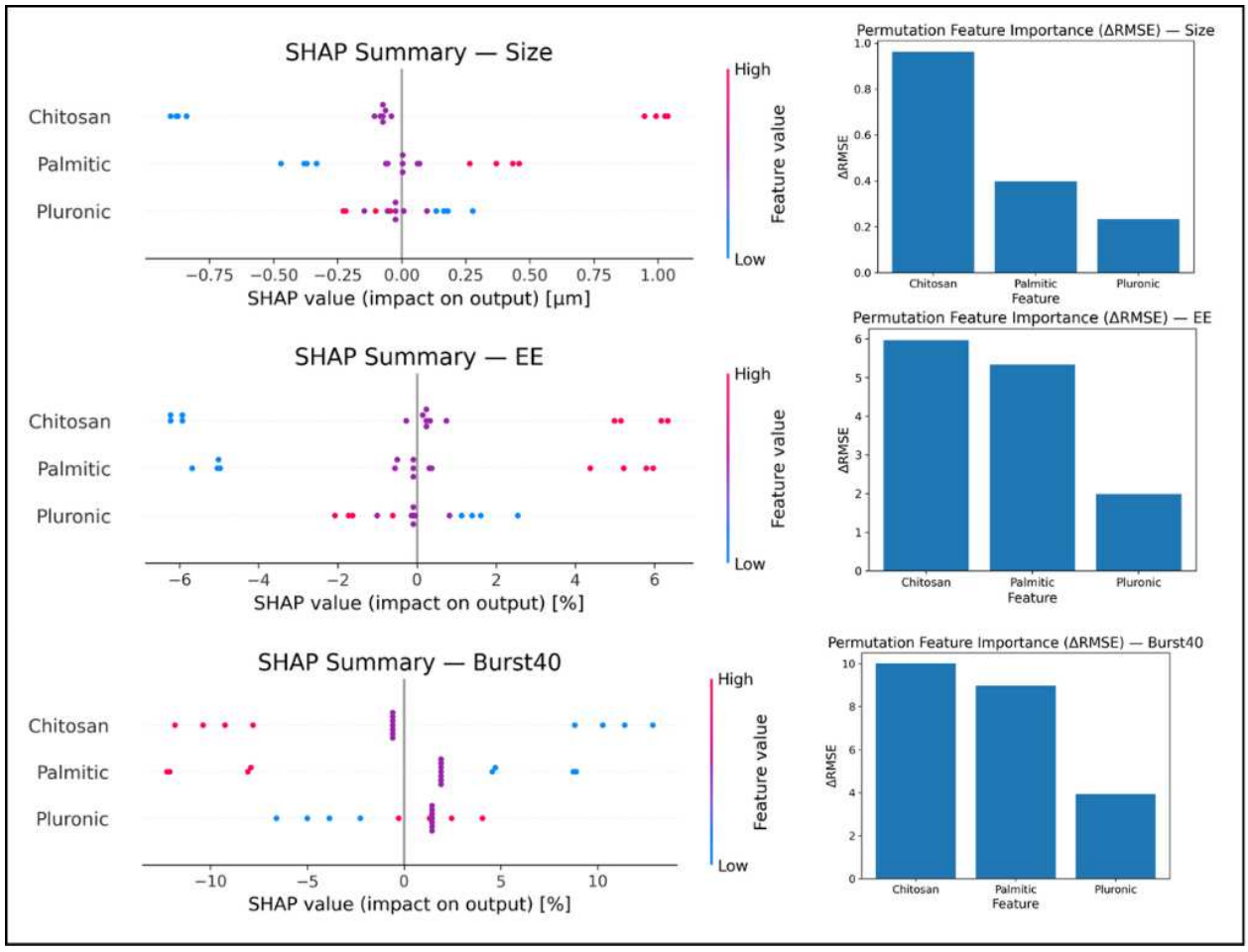


Figure 3A. Model interpretability for CQAs using SHAP and permutation importance. Left: SHAP summary plots for **Size (µm)**, **EE (%)**, and **Burst40 (%)** showing the direction and magnitude of each factor's effect on predictions (red = high feature value, blue = low; positive SHAP increases the predicted response). Right: permutation feature importance (ΔRMSE) ranking factor influence for each CQA.

Explainability outputs: global factor importance and consistency across methods

Explainability analysis converged on a consistent global importance structure (Fig. 3A & 3B). Permutation feature importance ranked chitosan as the dominant driver across all three CQAs, with palmitic acid contributing substantively—particularly for drug-related attributes—and Pluronic F68 acting as a secondary modulator within the tested bounds. SHAP global summaries and partial dependence trends were directionally concordant with PFI, supporting internal consistency between the learned surrogate structure and the empirical DoE landscape. Local attributions provided formulation-specific rationales consistent with global trends, enabling traceable interpretation of why particular candidates were shortlisted.

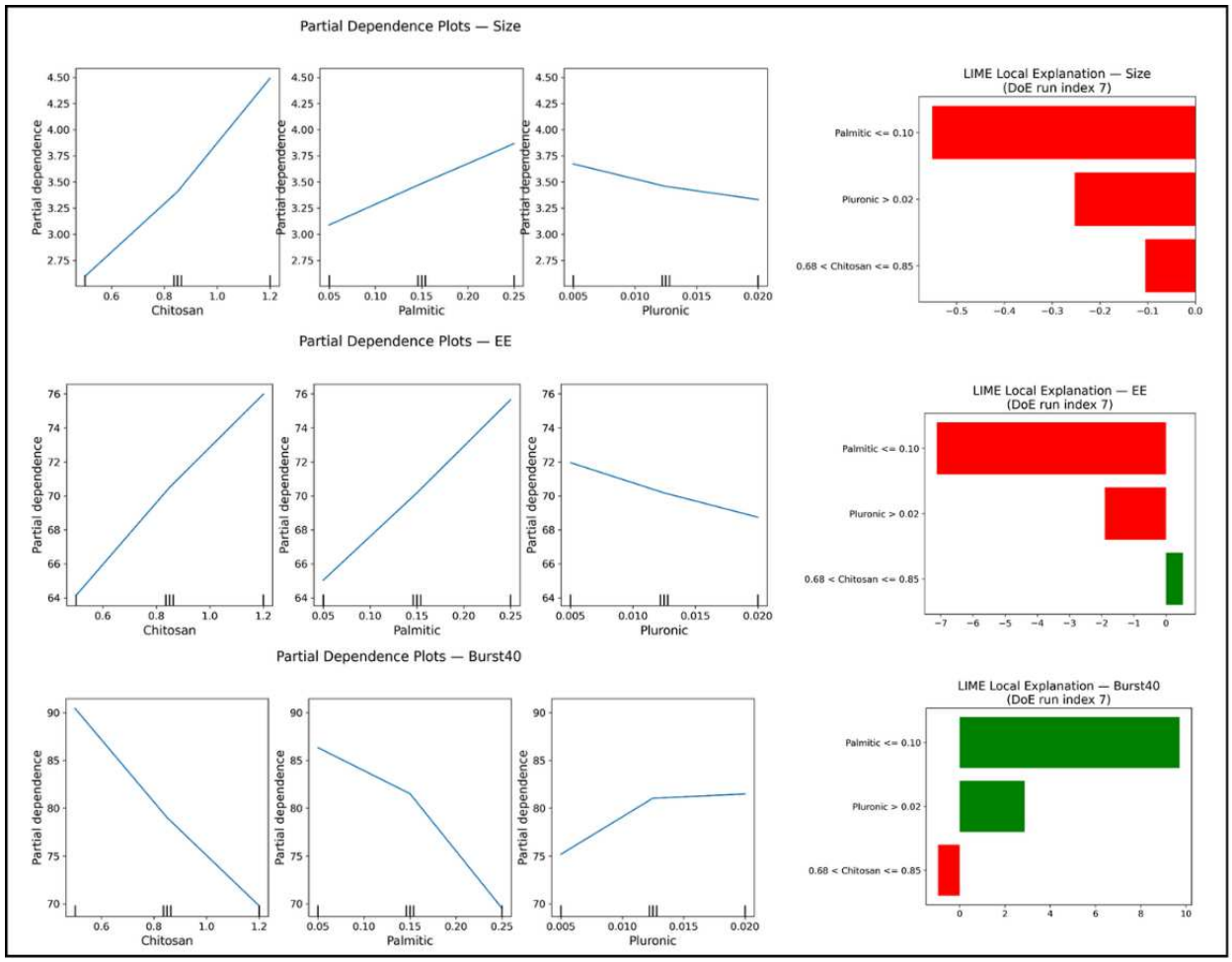


Figure 3B. Global and local interpretation of AI-surrogate predictions using Partial Dependence Plots (PDP) and LIME analysis. Left panels: PDPs showing the average effect of **chitosan, palmitic acid, and Pluronic F68** on predicted **Size (µm), EE (%), and Burst40 (%)** across the design space. Right panels: LIME local explanations for a representative validation formulation (DoE run index 7), indicating the direction and magnitude of each factor's contribution (green = positive effect, red = negative effect) to the predicted responses.

External benchmarking: AI predictions versus QbD/RSM baseline against experimentally prepared decision batches

To evaluate out-of-DoE generalization in a development-relevant setting, five AI-shortlisted batches were prepared and characterized in triplicate and compared against both AI surrogate predictions and the QbD/RSM polynomial baseline (Tables 2; Fig.4). Experimentally, the AI-shortlisted set clustered tightly for size and EE (mean size 3.70 ± 0.12 to 4.30 ± 0.15 µm; EE 74.0 ± 1.8 to 76.0 ± 1.9%), with moderate Burst40 values (69.0 ± 2.0 to 72.0 ± 2.1%). The QbD comparator expressed a distinct trade-off profile: smaller size and lower EE but markedly higher Burst40 (Tables 2).

Table 2. External validation of shortlisted batches: experimental CQAs versus AI surrogate and QbD predictions

Batch	Chitosan (% w/v)	Palmitic acid (% w/v)	Pluronic F68 (% w/v)	Size (µm)	AI Size (µm)	QbD Size (µm)	|Error| AI Size	|Error| QbD Size
AI 1	0.85	0.2	0.013	3.70 ± 0.12	3.45	3.55	0.25	0.15
AI 2	1.15	0.16	0.012	4.30 ± 0.15	3.97	4.37	0.33	0.07
AI 3	1	0.17	0.013	4.05 ± 0.13	3.54	3.88	0.51	0.17
AI 4	1.05	0.17	0.012	4.15 ± 0.14	3.65	4.06	0.5	0.09
AI 5	0.9	0.19	0.012	3.80 ± 0.12	3.46	3.67	0.34	0.13
QbD	0.75	0.19	0.019	3.20 ± 0.11	3.33	3.19	0.13	0.01
Batch	**Chitosan (% w/v)**	**Palmitic acid (% w/v)**	**Pluronic F68 (% w/v)**	**EE (%)**	**AI EE (%)**	**QbD EE (%)**	**|Error| AI EE**	**|Error| QbD EE**
AI 1	0.85	0.2	0.013	74.0 ± 1.8	71.1	73	2.9	1
AI 2	1.15	0.16	0.012	76.0 ± 1.9	73.2	75.9	2.8	0.1
AI 3	1	0.17	0.013	74.5 ± 1.7	71.2	73.9	3.3	0.6
AI 4	1.05	0.17	0.012	75.5 ± 1.8	71.8	74.9	3.7	0.6
AI 5	0.9	0.19	0.012	74.8 ± 1.7	71.1	73.5	3.7	1.3
QbD	0.75	0.19	0.019	69.5 ± 1.6	70.2	69.9	0.7	0.4
Batch	**Chitosan (% w/v)**	**Palmitic acid (% w/v)**	**Pluronic F68 (% w/v)**	**Burst40 (%)**	**AI Burst40 (%)**	**QbD Burst40 (%)**	**|Error| AI Burst40**	**|Error| QbD Burst40**
AI 1	0.85	0.2	0.013	72.0 ± 2.1	77.3	77.1	5.3	5.1
AI 2	1.15	0.16	0.012	69.0 ± 2.0	72.9	72	3.9	3
AI 3	1	0.17	0.013	71.0 ± 1.9	76.2	75.8	5.2	4.8
AI 4	1.05	0.17	0.012	69.5 ± 1.8	74.3	73.7	4.8	4.2
AI 5	0.9	0.19	0.012	70.5 ± 2.0	76.6	76.3	6.1	5.8
QbD	0.75	0.19	0.019	86.0 ± 1.7	80.9	81	5.1	5

Experimental values are mean ± SD (n = 3). |Error| denotes absolute error relative to the experimental mean.

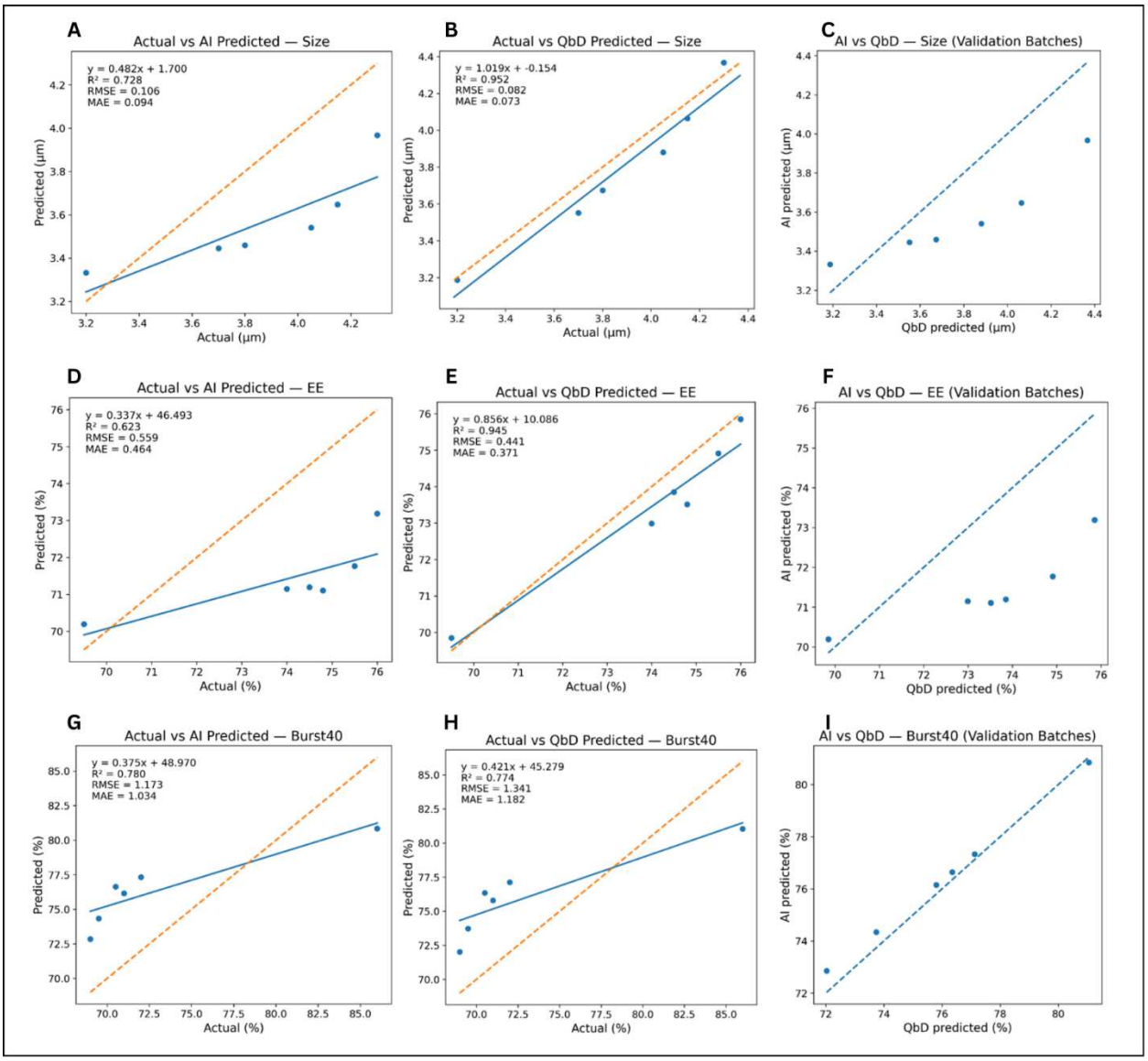


Figure 4. External validation and agreement analysis between AI-surrogate and QbD/RSM predictions for shortlisted DOX-loaded chitosan microbubble batches.

Panels **A–C** show microbubble size (µm), **D–F** entrapment efficiency (EE, %), and **G–I** ultrasound-triggered burst release at 40 s (Burst40, %). Panels **A, D, G** present experimentally measured values plotted against AI-surrogate predictions; **B, E, H** show experimental values versus QbD/RSM predictions; and **C, F, I** compare AI-surrogate predictions directly with QbD/RSM predictions for the same validation batches. Each point represents one experimentally prepared batch (n = 6, including the QbD comparator). The dashed line denotes the line of identity (y = x), while the solid line represents the linear regression fit (y = mx + c). Corresponding goodness-of-fit statistics (R^2, RMSE, and MAE) are reported within each panel. The plots demonstrate strong agreement for size and EE predictions and acceptable concordance for Burst40, confirming the predictive reliability of both modelling approaches within the validated design space.

When predictions were benchmarked against measurements, QbD/RSM displayed stronger absolute calibration for Size and EE across the five AI-shortlisted batches, whereas both approaches showed comparable residual error for Burst40. Averaged across the AI-shortlisted set, mean absolute errors were 0.388 μm (AI) versus 0.120 μm (QbD/RSM) for Size, and 3.28% (AI) versus 0.735% (QbD/RSM) for EE; Burst40 errors were similar in magnitude (5.06% AI versus 4.61% QbD/RSM). Linear calibration diagnostics (Fig. 4) reflected the same hierarchy: QbD/RSM retained near-identity scaling for Size and EE, whereas AI predictions exhibited weaker calibration despite trend tracking within the shortlisted region.

Decision-centric candidate selection outcome

Candidate selection integrated multi-CQA desirability within the bounded QbD factor space and required that shortlisted candidates were experimentally feasible and explanation-consistent. The resulting shortlist occupied a compositionally coherent region characterized by tight experimental dispersion in Size and EE and moderate Burst40 values (Tables 2). Inclusion of the QbD comparator provided an anchored benchmark for interpreting whether AI-guided selection preserved the expected EE–Burst40 trade-space while maintaining manufacturability and reproducibility.

Ultrasound-triggered release: ON/OFF separation in vitro

Release profiles compared free doxorubicin, QbD-optimized microbubbles, and AI-optimized microbubbles under matched baseline and ultrasound conditions (Fig. 5). Under no-ultrasound conditions, both microbubble formulations demonstrated suppressed baseline release relative to free drug. By ~6 h, cumulative release from the no-US microbubble groups remained ~34–38%, whereas free DOX exceeded ~80% over the same interval. Upon ultrasound exposure, both microbubble formulations transitioned to a triggered release phenotype, with cumulative release approaching ~88–90% by ~6 h and clearly separating from their corresponding no-US trajectories. The +US release profiles of AI-optimized and QbD-optimized formulations were closely aligned, indicating convergence on a functionally similar ON/OFF release behavior.

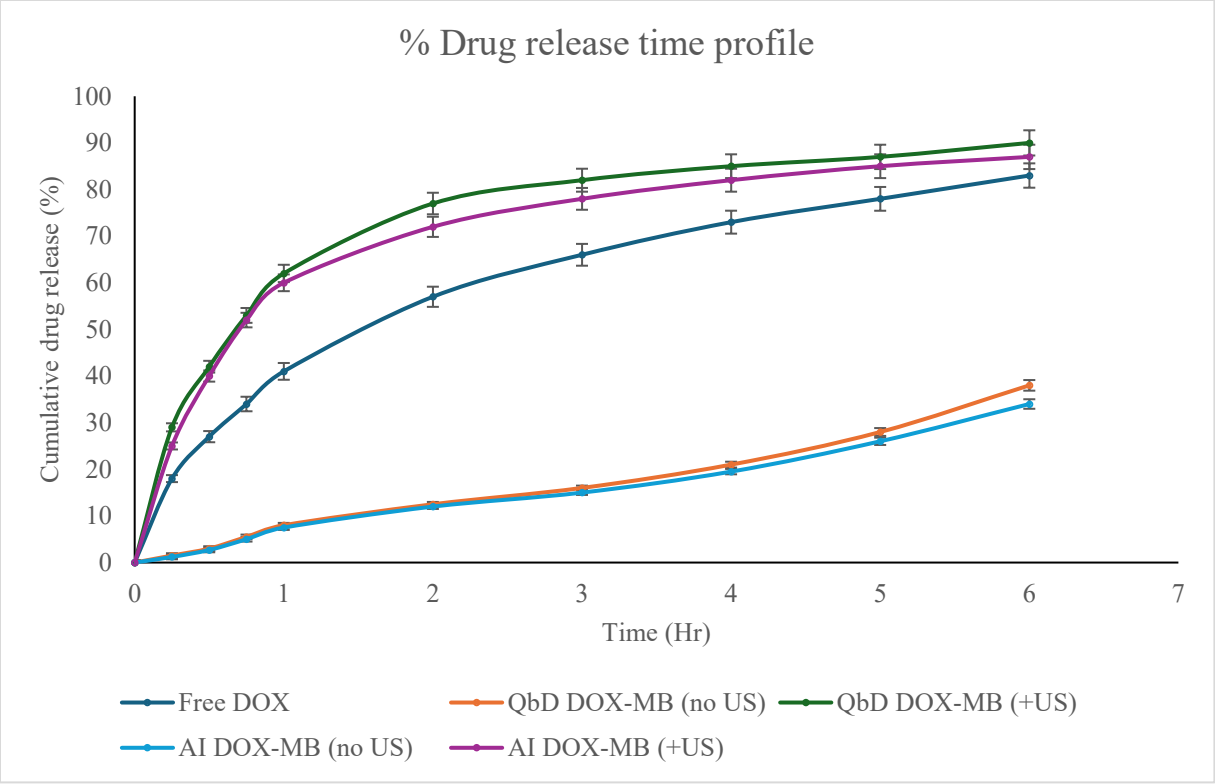


Figure 5. In vitro ultrasound-triggered release profiles of doxorubicin from optimized microbubble formulations. Cumulative drug release (%) as a function of time for free DOX, QbD-optimized DOX-microbubbles, and AI-optimized DOX-microbubbles under baseline (no ultrasound, no-US) and ultrasound-activated (+US) conditions

Trigger-dependent anticancer activity and apoptosis signaling in MCF-7

In MCF-7 viability experiments (Fig. 6H), both optimized microbubble formulations demonstrated ultrasound-dependent enhancement of cytotoxicity. At matched nominal doxorubicin concentrations, microbubble-treated cells showed higher residual viability in the absence of ultrasound, while ultrasound exposure shifted the dose–response toward lower viability, consistent with increased effective exposure upon triggering.

Annexin V/PI profiling (Fig. 6A–F) supported a trigger-dependent shift in cell-death phenotype. Untreated controls remained predominantly viable. Free DOX increased Annexin-positive populations, consistent with apoptosis induction. In contrast, DOX-microbubbles without ultrasound produced a smaller redistribution from the viable quadrant, whereas ultrasound activation increased Annexin-positive fractions and elevated total apoptosis (early + late) relative to the corresponding non-insonated microbubble condition. Apoptosis profiles under ultrasound

were comparable between the AI-optimized and QbD-optimized formulations, supporting that both candidates generated sufficient triggered exposure to engage apoptosis-associated cell death.

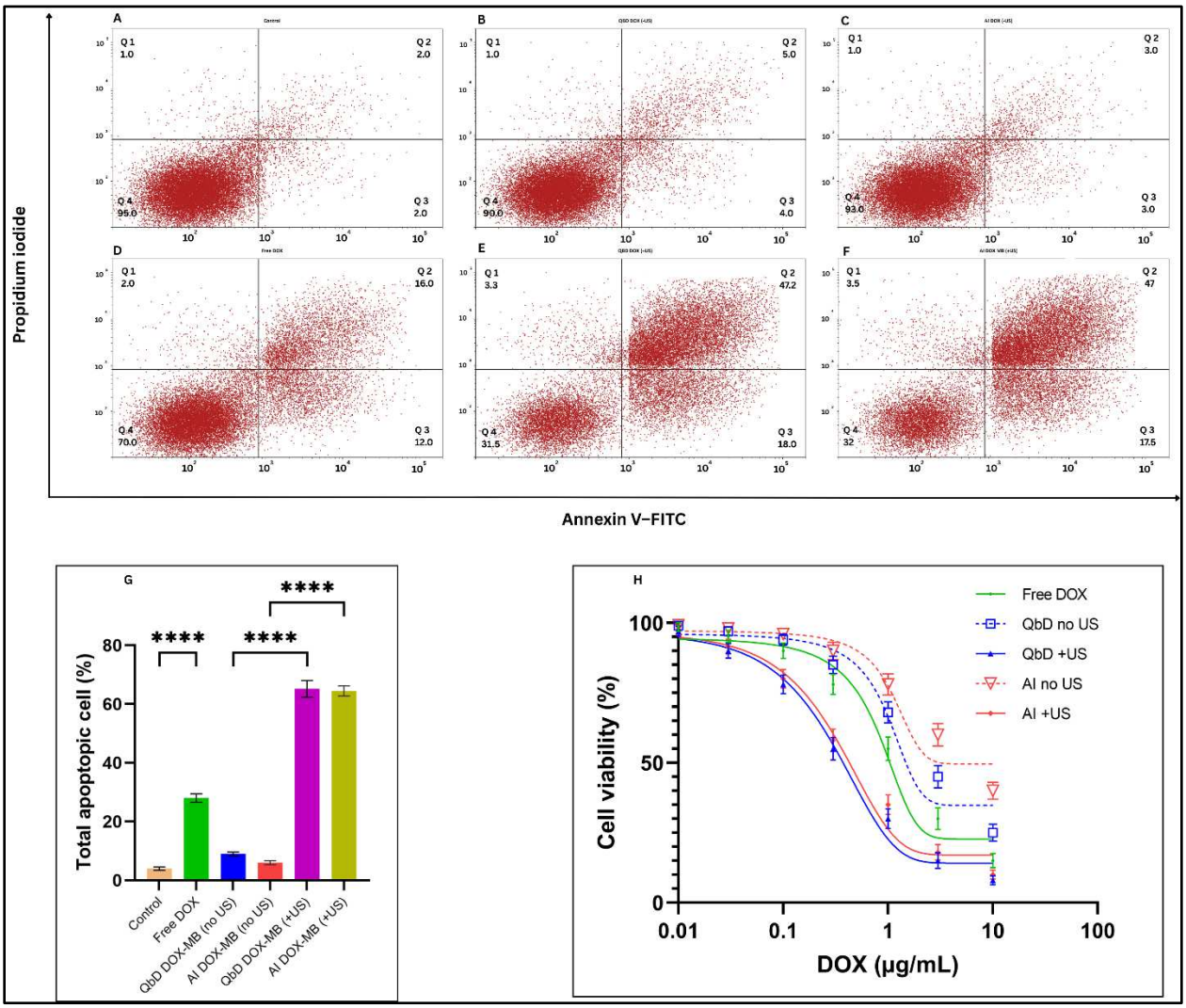


Figure 6. Ultrasound activation amplifies the cytotoxic and pro-apoptotic response of doxorubicin-loaded chitosan microbubbles in MCF-7 cells. Annexin V–FITC/propidium iodide (PI) flow-cytometry dot plots (A–F) show quadrant-gated distributions, where Q4 (Annexin−/PI−) represents viable cells, Q3 (Annexin+/PI−) early apoptosis, Q2 (Annexin+/PI+) late apoptosis/secondary necrosis, and Q1 (Annexin−/PI+) primary necrosis; quadrant percentages are reported on each plot. (A) Untreated control. (B) QbD-optimized DOX-MB, no ultrasound (−US). (C) AI-optimized DOX-MB, −US. (D) Free DOX. (E) QbD-optimized DOX-MB with ultrasound (+US). (F) AI-optimized DOX-MB, +US. (G) Quantification of total apoptosis (Q2+Q3) demonstrating a marked increase upon ultrasound activation for both optimized microbubble formulations; bars represent mean ± SD (n = 3), with statistical significance indicated (****). (H) Dose–response viability curves (DOX, µg/mL; log scale) comparing Free DOX and DOX-MB formulations under −US and +US conditions, showing an ultrasound-dependent leftward shift (enhanced potency) for microbubble groups; points represent mean ± SD (n = 3).

4. Discussion

The Box–Behnken DoE delineated a clearly bounded experimental region and generated a wide CQA envelope (size 2.35–4.85 μm; EE 60–82%; Burst40 55.6–95%), confirming that chitosan, palmitic acid, and Pluronic F68 jointly create a meaningful formulation trade-space that must be balanced against the intended therapeutic objective [1]. The response ranges also make the central development tension explicit [1,16]: compositions enriched in chitosan and palmitic acid produced the highest EE (up to 82%) yet the lowest ultrasound responsiveness (Burst40 down to 55.6%), whereas low-chitosan conditions shifted toward the opposite phenotype (Burst40 up to 95% with reduced EE, typically 60–68%). This antagonism is consistent with stronger drug–matrix association and a more cohesive shell improving retention while simultaneously increasing resistance to ultrasound-driven disruption. Importantly, the replicated center points were tightly clustered, indicating that the intermediate region represents a reproducible operating window rather than an artifact of experimental variability, and therefore a practical "balanced" space for development decisions [1].

Mechanistically, chitosan emerged as the dominant lever across all CQAs, which aligns with its role as the primary shell-forming polymer: increasing chitosan plausibly thickens and stiffens the interfacial layer, enlarges the microbubble population, strengthens drug retention, and elevates the cavitation threshold—collectively suppressing Burst40 while improving EE. The increasing contribution of palmitic acid from size to drug-related CQAs highlights the functional importance of the lipid component in tuning shell mechanics and acoustic susceptibility, thereby modulating both encapsulation and ultrasound-triggered release [1,17–25]. Taken together, these findings support a rational selection logic within the design space: retention-prioritized use cases favor the high-chitosan/high-palmitic acid region (accepting lower Burst40), whereas trigger-prioritized use cases favor the low-chitosan region (accepting reduced EE) [1].

The LOOCV analysis showed clear response-dependent differences in predictability: PolyRidge (degree 3) modeled Size and EE with strong cross-validated agreement (R^2 = 0.885 and 0.871), whereas Burst40 remained only moderately predictable (R^2 = 0.501) despite an analogous modeling strategy [1]. This separation is consistent with the underlying physics of the CQAs. Mean size is largely set during formation by composition-governed interfacial and rheological

conditions—polymer concentration–driven viscosity, interfacial tension, and breakup/coalescence kinetics—which change smoothly across the design region and therefore map well onto low-order polynomial structure [26–28]. The low dispersion at the center points further supported that the Size signal was dominated by reproducible formulation effects rather than stochastic noise, explaining the low RMSE (0.264 μm) and the near-linear observed–predicted agreement [1,29].

Encapsulation efficiency, while more mechanistically layered, also behaved as a predominantly continuous response within the studied bounds. EE reflected a coupled partitioning/association process governed by drug–polymer interactions and lipid-assisted retention, together with kinetic trapping during shell formation [30–32]. Because these contributions evolve gradually with composition (rather than switching abruptly), the response surface remained sufficiently smooth for stable learning, yielding RMSE = 2.18% under LOOCV and consistent trend capture across the DoE [1,33,34].

Burst40, in contrast, represented an ultrasound-activated failure outcome that is intrinsically closer to a threshold phenomenon than a smooth interpolation problem. Triggered release depends on cavitation-driven oscillation and collapse, where the transition from stable oscillation to disruptive inertial events occurs once acoustic forcing exceeds a composition-dependent threshold governed by shell thickness, viscoelastic damping, and local heterogeneity [38–43]. In this regime, small compositional shifts can produce disproportionately large functional changes near the threshold, while producing minimal change far from it, yielding heterogeneous sensitivity that is difficult for polynomial models to approximate reliably. The shell itself adds irreducible variability: defect density, local thickness variations, and microscale compositional non-uniformity can create preferential rupture sites that amplify stochasticity under insonation [44–47]. Finally, the measured Burst40 likely integrated multiple mechanisms—catastrophic rupture, ultrasound-enhanced diffusion through partially intact shells, and microstreaming/radiation-force–driven transport—whose relative contributions can shift across the design space, reducing additivity and further limiting predictability from bulk composition alone [36,48–50].

Burst40 was consistently the least modelable endpoint in the present three-factor description (R^2 = 0.501; RMSE = 8.61%), indicating that ultrasound-triggered release retains a larger irreducible component of variability than Size or EE under DoE-scale learning [1]. This behavior is mechanistically plausible because Burst40 is governed by cavitation-mediated shell failure rather than a purely composition-driven equilibrium response. Acoustic cavitation involves bubble

oscillation and collapse, with a pronounced threshold separating stable oscillation from disruptive inertial events [38,39]. When the applied acoustic conditions sit near a formulation-dependent cavitation threshold, modest compositional shifts can switch the dominant response regime, producing a nonlinear—sometimes effectively discontinuous—composition–response relationship that is difficult for low-order polynomials to approximate smoothly [1,35,38,40]. In addition, microbubble shells are mechanically heterogeneous composites; local thickness variation, defect density, and microstructural inhomogeneity can create preferential rupture sites, increasing stochasticity in triggered release even when bulk composition is nominally matched [44,45]. Finally, Burst40 likely aggregates multiple concurrent mechanisms (catastrophic rupture for labile shells, ultrasound-enhanced diffusion for more intact shells, and microstreaming/radiation-force–assisted transport), whose relative contribution shifts across the design region, further limiting predictability from composition alone [48,49]. Collectively, these results justify a risk-stratified use of surrogates within the design space: Size and EE support calibrated interpolation and decision support, whereas Burst40 should be interpreted more cautiously and is a prime candidate for expansion with larger datasets and/or mechanistically informed descriptors that explicitly capture shell mechanics and cavitation physics [1,51–55].

Across models, the algorithm comparison reinforced an important practical point for DoE-scale datasets: regularized polynomial structure remained competitive—and in this case superior—to tree ensembles under LOOCV for all three CQAs. For Size and EE, PolyRidge (degree 3) delivered the strongest cross-validated performance (R^2 = 0.885 and 0.871), while for Burst40 the simpler PolyRidge (degree 2) was optimal, indicating response-specific complexity control rather than "one-model-fits-all" tuning [1]. This outcome is consistent with the data regime: tree ensembles need enough observations to learn stable split rules; with n = 15, bootstrap resamples contain limited unique information and boosting can inadvertently amplify noise when successive trees chase residuals that reflect sampling idiosyncrasies rather than signal [61]. In contrast, L2-regularized polynomials impose a controlled bias–variance trade-off and stabilize coefficient estimates by shrinking weak and correlated terms, which is advantageous when the effective feature space approaches the sample size. Importantly, the Box–Behnken design is inherently optimized for polynomial estimation; its near-orthogonal structure supports efficient identification of main effects, interactions, and curvature with minimal runs, thereby aligning the experimental design with the inductive bias of polynomial surrogates [62–64]. For formulation development in

the common DoE range (≈10–20 runs), these results argue for an evidence-led strategy: begin with regularized polynomial surrogates that exploit the design structure, then benchmark against ensembles rather than assuming ensemble superiority by default [51,52].

The external validation step directly tested whether the AI layer delivered incremental decision value beyond the established QbD polynomial baseline when confronted with newly manufactured "decision batches." In this head-to-head benchmark, the QbD/RSM equations showed clearly stronger absolute calibration for the composition-dominated CQAs. Across the five AI-shortlisted batches, mean absolute error was markedly lower for QbD/RSM than for the AI surrogates—0.120 μm versus 0.388 μm for Size (≈3.2-fold) and 0.735% versus 3.28% for EE (≈4.5-fold). In contrast, Burst40 errors were similar for both approaches (4.61% vs 5.06%), reinforcing that ultrasound-triggered release remains the most difficult endpoint to predict robustly from a three-factor composition space alone [1,43].

Calibration plots clarified why the AI surrogates underperformed externally despite acceptable internal cross-validation: the dominant failure mode was not loss of rank-ordering, but scale compression and systematic bias. QbD/RSM retained near-identity scaling for Size and EE (slopes close to 1), indicating that predicted ranges expanded and contracted in step with the measured response. The AI surrogates, by comparison, tracked local trends within the shortlisted region but did so with weaker calibration—effectively dampening the predicted dynamic range—so that differences between batches were underrepresented in absolute units.

Several features of the modeling regime plausibly explain this gap. First, DoE-scale data impose a hard ceiling on how much flexibility can be supported without paying a generalization penalty. A third-order polynomial ridge surrogate introduces ~20 terms with only 15 runs; even with L2 shrinkage, this feature-to-sample ratio makes the learned surface sensitive to idiosyncrasies of the training set and increases the risk that internal performance looks better than true prospective behavior. Second, LOOCV can be optimistic when models are moderately complex relative to n because each fold is trained on 14/15 observations—nearly the full dataset—so the model effectively "sees" almost everything in every fold; this can understate the error that emerges when confronted with genuinely new combinations and experimental noise [1]. Third, the QbD/RSM equations were estimated within the Box–Behnken structure itself, which is deliberately constructed to support stable estimation of polynomial main effects, interactions, and curvature with minimal confounding; the near-orthogonality of the design improves coefficient

identifiability and tends to translate into better calibrated interpolation within the investigated bounds [62–64].

Taken together, these results position the QbD/RSM models as the more reliable quantitative baseline for calibrated prediction of Size and EE within the validated design region, while also underscoring a practical lesson for AI-enabled DoE: external validation should be treated as the decisive filter, and calibration—not only cross-validated R^2—should determine whether a surrogate is fit for quantitative decision-making or best reserved for exploratory screening and interpretability support [1].

Release testing showed a clear trigger dependence: in the absence of insonation, the microbubbles retained doxorubicin and released it only gradually, whereas ultrasound exposure shifted the system into a rapid, activation-driven release mode. This behavior is consistent with the physical basis of microbubble-triggered delivery, where drug liberation is governed less by simple diffusion and more by cavitation-mediated shell disruption under an applied acoustic field.

Mechanistically, ultrasound-responsive release from coated microbubbles is dominated by acoustic cavitation—bubble oscillation and, when the applied pressure exceeds a formulation-specific threshold, collapse events that generate strong local mechanical stresses capable of rupturing or fragmenting the shell. Below this threshold, microbubbles primarily undergo stable oscillations that generate microstreaming and radiation forces; these can accelerate transport but typically do not cause catastrophic shell failure. Once the threshold is crossed, inertial cavitation becomes more likely, producing violent collapse, shock waves, and shear forces that promote rapid shell disruption and drug expulsion [1,65,66].

The key point from a formulation-development perspective is that the cavitation threshold is not fixed; it is tuned by microbubble architecture—size distribution, shell thickness, and viscoelastic properties—which in turn are governed by composition. Increasing chitosan generally increases shell thickness and stiffness, raising the acoustic pressure needed to achieve disruptive oscillations. Palmitic acid can shift this balance by modulating packing, cohesion, and flexibility within the composite shell, thereby altering how readily the coating buckles, cracks, or fails during insonation. As a result, small composition changes can reposition a formulation relative to the activation threshold, which explains why ultrasound-triggered endpoints often show higher variability and lower statistical “modelability” than composition-dominated CQAs such as Size and EE [1,67].

Functionally, this trigger dependence is developmentally desirable: low passive release helps preserve drug association during handling and exposure outside the insonated region, while ultrasound provides a controllable “release switch” at the target site. At the same time, the same threshold physics that enables spatial control also makes Burst40 difficult to predict from bulk composition alone, because microstructural heterogeneity (local weak points, coating nonuniformity, and defect distributions) can dominate rupture behavior under ultrasound. These observations support the rationale for moving toward mechanistically informed descriptors (e.g., shell viscoelasticity proxies) or hybrid models that couple cavitation likelihood with empirical release kinetics when the goal is more reliable prediction of ultrasound-triggered performance.

In MCF-7 assays, both optimized doxorubicin–microbubble formulations exhibited a clear ultrasound-dependent gain in anticancer activity. At matched nominal doxorubicin doses, cells treated with microbubbles under non-insonated conditions retained higher viability, whereas ultrasound exposure shifted the response toward lower viability—consistent with an increase in effective intracellular exposure at the time of triggering [1]. This pattern is mechanistically coherent with the release behavior: intact shells under baseline conditions limit drug availability in the pericellular environment, while insonation promotes rapid liberation of associated drug and can simultaneously enhance cellular entry through cavitation-driven membrane perturbation (sonoporation), together producing a measurable potency shift [1].

Flow-cytometric Annexin V/PI profiling provided orthogonal support that the viability reduction reflected apoptotic commitment rather than non-specific membrane destruction. Untreated controls remained largely viable, and free doxorubicin increased Annexin-positive fractions as expected for a DNA-damage–driven cytotoxic agent. Microbubbles without ultrasound produced a smaller redistribution away from the viable quadrant, consistent with limited drug exposure. With ultrasound, both microbubble formulations showed an increased Annexin-positive population and a higher total apoptotic fraction (early + late apoptosis) relative to the corresponding non-ultrasound condition [1]. Because Annexin V+/PI− (early apoptosis) and Annexin V+/PI+ (late apoptosis) are characteristic of programmed cell death, the shift toward these quadrants indicates that the triggered formulations predominantly amplified doxorubicin’s intended pharmacology rather than inducing primarily necrotic injury [1,68,69].

Importantly, the apoptosis profiles under ultrasound were comparable for the AI-optimized and QbD-optimized batches, indicating convergence at the biologically relevant endpoint despite

differences in modeling route and absolute CQA calibration [1]. Collectively, the paired viability and apoptosis results support the translational premise of this platform: the formulation suppresses baseline exposure, while ultrasound concentrates drug action in space and time, enhancing tumor-cell killing where insonation is applied and reducing exposure elsewhere—an effect that underpins the selectivity advantage of ultrasound-responsive delivery systems [1,3,70-72].

The integration of machine-learning surrogates into a QbD development program has direct consequences for both control-strategy design and model-risk governance [6,73]. Here, the Box–Behnken DoE defined a bounded design region with experimentally supported CQA ranges, and the digital workflow was intentionally constrained to this space; consequently, the AI layer functioned as a within-design-space decision aid rather than a tool for extrapolative exploration. This distinction is important from a development perspective because it preserves the primacy of the DoE as the evidence base, while allowing computational models to identify practically useful subregions ("preferred operating zones") without extending beyond conditions supported by experimentation. Within this framework, the role of the control strategy is to translate design-space knowledge into planned controls that assure consistent performance, spanning raw-material attributes, process parameters, in-process monitoring, and end-product specifications.

A key outcome of the head-to-head benchmarking is that model performance can be mapped directly onto specification tightness and intended model use. The QbD/RSM equations showed strong external calibration for Size and EE, supporting relatively tighter acceptance targets (e.g., Size 3.0–4.5 μm; EE ≥70%) and routine, calibrated interpolation within the established region [1]. By contrast, Burst40 remained only moderately predictable, with residual errors of similar magnitude across approaches; accordingly, broader acceptance bands (e.g., 60–90%) are more defensible at the current evidence level to accommodate both measurement variability and model uncertainty [1]. In practical terms, this implies that QbD/RSM polynomials are suitable candidates for integration into process control and decision support where predictions must be quantitatively reliable, whereas the present AI surrogates—given their poorer external calibration—should not be positioned for high-stakes uses such as real-time release testing without additional calibration/validation work [1,6]. Importantly, the explainability layer still contributes to control-strategy development by highlighting subregions that consistently align with specific CQA priorities; these can be framed as proven acceptable ranges within the broader design space, enabling manufacturing flexibility while maintaining predictable performance [1,6].

External validation further enabled explicit risk stratification of model use. For Size and EE, QbD/RSM predictions showed low mean absolute errors (0.120 μm and 0.735%) with near-identity calibration slopes, indicating low model risk for control-strategy applications such as specification justification and routine batch decision support [1,74,75]. For Burst40, both QbD/RSM and AI predictions showed moderate residual error (mean |error| 4.61–5.06%) with imperfect calibration; this places Burst40 in a moderate-risk category, where model outputs are best treated as approximate guides and critical decisions should be confirmed experimentally (e.g., triggered-release checks for lot disposition or change evaluation) [1]. For Size and EE, the AI surrogates exhibited higher mean absolute errors (0.388 μm and 3.28%) with clear calibration bias, indicating high model risk for manufacturing-facing decisions; in this state, these surrogates should be limited to exploratory screening and explainability-supported hypothesis generation unless recalibrated and prospectively revalidated [1,6,76].

On this basis, model applicability boundaries can be stated transparently. Within boundaries (low risk): predictions are restricted to factor settings inside the DoE design space, with Size and EE estimated using QbD/RSM equations for uses such as specification setting, process control, and routine batch decision support [74,75,78,79]. At boundaries (moderate risk): near the edges of the design space, Burst40 predictions from any model should be interpreted cautiously and paired with experimental verification when used for formulation selection or change assessment [80]. Outside boundaries (high risk): beyond the validated factor region, neither QbD/RSM nor AI predictions should be used to justify manufacturing decisions without a new DoE and a renewed validation package; AI outputs in particular should be treated as research-only unless recalibrated with additional data [78,81]. Finally, if the workflow is to support regulatory-facing claims, the same evidence trail used here should be formalized into a concise documentation set—Model Development, Validation (including calibration diagnostics), Explainability outputs linked to formulation rationale, and a risk assessment that states the intended use and limitations—thereby strengthening the credibility of model-informed decisions and facilitating regulatory review [82].

The present study establishes a decision-centric workflow that couples a QbD-defined experimental design space with interpretable surrogate modeling to strengthen formulation selection for ultrasound-triggered, doxorubicin-loaded chitosan microbubbles [1,6]. The 15-run Box–Behnken DoE created a bounded, mechanistically meaningful knowledge space in which key formulation trade-offs were experimentally resolved [1,62]. Within this space, regularized

polynomial surrogates evaluated under LOOCV provided high-fidelity learning for composition-dominated CQAs (Size and EE; R^2 = 0.885 and 0.871), whereas Burst40 was inherently less predictable (R^2 = 0.501), consistent with the threshold-like and microstructure-sensitive nature of cavitation-mediated shell disruption [1,40]. External validation further clarified model utility: QbD/RSM retained superior calibration for absolute prediction of Size and EE (mean errors 0.120 μm and 0.735% versus AI 0.388 μm and 3.28%), while both approaches showed comparable residual limitations for Burst40 (mean errors ~4.6–5.1%) [1,43].

Importantly, despite these calibration differences, the AI-nominated candidates maintained the intended functional phenotype. The AI-optimized and QbD-optimized formulations both demonstrated distinct trigger-governed release behavior in vitro (suppressed baseline release ~35% versus ~88–90% under ultrasound) and a corresponding, ultrasound-dependent enhancement of cytotoxicity in MCF-7 cells with increased apoptotic commitment [1]. Beyond prediction, the explainability layer provided practical development value by converting model behavior into auditable formulation levers [6,83]. Convergent SHAP, PFI, PDP, and LIME readouts consistently implicated chitosan and palmitic acid as primary control variables, with Pluronic F68 acting predominantly as a stabilizing modifier within the explored ranges [1,6].

Collectively, these results support a defensible QbD narrative in which the DoE/RSM framework anchors the validated design space and quantitative baseline, while interpretable AI adds complementary decision support—improving traceability of choices, enabling consensus-based prioritization, and reducing selection bias when advancing candidates to experimental verification [1,73].

Conclusion

This study introduces a decision-centric workflow that effectively integrates a Quality by Design-defined experimental design space with interpretable surrogate modeling. This approach significantly strengthens and accelerates the selection of formulations for ultrasound-triggered, doxorubicin-loaded chitosan microbubbles.

The 15-run Box–Behnken Design of Experiments established a bounded and mechanistically relevant knowledge space, allowing for the experimental determination of key formulation trade-offs. Within this framework, regularized polynomial surrogates, trained using Leave-One-Out Cross-Validation, demonstrated high-fidelity prediction for composition-dominated Critical

Quality Attributes such as Size and Encapsulation Efficiency. However, predictability for Burst40 was more modest, reflecting the inherently threshold-like and microstructure-sensitive nature of acoustic disruption.

External validation revealed that while QbD/Response Surface Methodology models maintained superior calibration for absolute predictions of Size and EE, both the QbD/RSM and AI-based approaches encountered comparable limitations in predicting Burst40. Crucially, the AI-shortlisted candidates successfully maintained the intended functional phenotype, exhibiting clear ON/OFF release behavior and trigger-dependent cytotoxicity in MCF-7 cells with enhanced apoptotic commitment.

The integrated explainability layer provided practical value by translating complex model behaviors into auditable design rules. Convergent interpretations from methods like SHAP, PFI, PDP, and LIME consistently highlighted chitosan and palmitic acid as primary drivers, with Pluronic F68 serving as a key stabilizing modifier. From the perspective of ICH Q8–Q11 guidelines, this combined workflow offers a robust development narrative. QbD establishes the foundational design space and quantitative baseline, while interpretable AI complements this by enhancing decision transparency, facilitating consensus-based prioritization, reducing selection bias, and strengthening the traceability of formulation choices.

In conclusion, these findings advocate for an integrated approach rather than positioning AI as a replacement for QbD design-space establishment. Explainable AI surrogates augment QbD by improving decision transparency and accelerating the progression toward functional and biological validation. Future refinements should prioritize enhancing the prediction of ultrasound-functional endpoints through the incorporation of detailed descriptors of shell microstructure and process history, ultimately strengthening the control strategy development for clinically relevant, trigger-responsive microbubble formulations.

References

1. Solanki S, Thota S, Solanki N, Dhakar R, Aphale S, Kaul-Ghanekar R, et al. Ultrasound-responsive doxorubicin microbubbles engineered by QbD: enhanced in vitro anti-breast-cancer efficacy with attenuated cardiac cell toxicity. 2026. https://doi.org/10.21203/rs.3.rs-8565614/v1.

2. Tinkov S, Coester C, Serba S, Geis NA, Katus HA, Winter G, et al. New doxorubicin-loaded phospholipid microbubbles for targeted tumor therapy: in-vivo characterization. J Control Release. 2010;148:368. https://doi.org/10.1016/j.jconrel.2010.09.004.

3. Yu FTH, Chen X, Wang J, Qin B, Villanueva FS. Low intensity ultrasound mediated liposomal doxorubicin delivery using polymer microbubbles. Mol Pharm. 2015;13:55. https://doi.org/10.1021/acs.molpharmaceut.5b00421.

4. Lambert EG, Janjic JM. Quality by design approach identifies critical parameters driving oxygen delivery performance in vitro for perfluorocarbon based artificial oxygen carriers. Sci Rep. 2021;11. https://doi.org/10.1038/s41598-021-84076-1.

5. Rodriguez-Granrose D, Jones AK, Loftus H, Tandeski T, Heaton W, Foley KT, et al. Design of experiment (DOE) applied to artificial neural network architecture enables rapid bioprocess improvement. Bioprocess Biosyst Eng. 2021;44:1301. https://doi.org/10.1007/s00449-021-02529-3.

6. Solanki S, Kumari G, Solanki N, Jindal D, Usha Kiranmai G, Chakorkar S, et al. From QbD to explainable AI: interpretable random forest surrogates for design space understanding of voriconazole–β-cyclodextrin inclusion complexes. 2026.

7. König C, Vellido A. Understanding predictions of drug profiles using explainable machine learning models. BioData Min. 2024;17. https://doi.org/10.1186/s13040-024-00378-w.

8. Molnar C, Freiesleben T, König G, Herbinger J, Reisinger T, Casalicchio G, et al. Relating the partial dependence plot and permutation feature importance to the data generating process. Commun Comput Inf Sci. 2023. p. 456. https://doi.org/10.1007/978-3-031-44064-9_24.

9. Jorayev P, Soritz S, Sung S, Jeraal MI, Russo D, Barthelme A, et al. Machine learning-driven optimization of continuous-flow photoredox amine synthesis. Org Process Res Dev. 2025;29:1411. https://doi.org/10.1021/acs.oprd.4c00533.

10. Đuriš J, Kurćubić I, Ibrić S. Review of machine learning algorithms' application in pharmaceutical technology. Arh Farm. 2021;71:302. https://doi.org/10.5937/arhfarm71-32499.

11. Murray JD, Lange JJ, Bennett-Lenane H, Holm R, Kuentz M, O'Dwyer PJ, et al. Advancing algorithmic drug product development: recommendations for machine learning approaches in drug formulation. Eur J Pharm Sci. 2023;106562. https://doi.org/10.1016/j.ejps.2023.106562.

12. Jiang J, Ma X, Ouyang D, Williams RO. Emerging artificial intelligence (AI) technologies used in the development of solid dosage forms. Pharmaceutics. 2022;2257. https://doi.org/10.3390/pharmaceutics14112257.

13. Bobadilla AVP, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci. 2024. https://doi.org/10.1111/cts.70056.

14. Singh G, Pai RS, Devi VK. Optimization of pellets containing solid dispersion prepared by extrusion/spheronization using central composite design and desirability function. J Young Pharm. 2012;4:146. https://doi.org/10.4103/0975-1483.100020.

15. Kumari PVK, Likhitha G, Mustaq SJH. Approach to optimize the self-microemulsifying drug delivery system for azilsartan medoxomil using Box–Behnken design and desirability function. Int J Appl Pharm. 2024;92. https://doi.org/10.22159/ijap.2024v16i2.49519.

16. Lu C, Zhao Y, Gao HS, Tian J, Zhou ZC, Zhao GT, et al. Comparing encapsulation efficiency and ultrasound-triggered release for protein between phospholipid-based microbubbles and liposomes. J Microencapsul. 2009;27. https://doi.org/10.1080/02652040903051996.

17. Lytra A, Sboros V, Giannakopoulos AE, Pelekasis N. Modeling atomic force microscopy and shell mechanical properties estimation of coated microbubbles. Soft Matter. 2020;16:4661. https://doi.org/10.1039/d0sm00300j.

18. Garg S, Thomas AA, Borden MA. The effect of lipid monolayer in-plane rigidity on in vivo microbubble circulation persistence. Biomaterials. 2013;34:6862. https://doi.org/10.1016/j.biomaterials.2013.05.053.

19. Wu SY, Chen C, Tung Y, Olumolade O, Konofagou EE. Effects of the microbubble shell physicochemical properties on ultrasound-mediated drug delivery to the brain. J Control Release. 2015;212:30. https://doi.org/10.1016/j.jconrel.2015.06.007.

20. Zalloum IO, Jafari SA, Paknahad A, Kolios MC, Tsai S, Karshafian R. Controlled tempering of lipid concentration and microbubble shrinkage as a possible mechanism for fine-tuning microbubble size and shell properties. Preprint at http://arxiv.org/abs/2307.13836 (2023).

21. Lin CW, Fan C, Yeh C. The impact of surface drug distribution on the acoustic behavior of DOX-loaded microbubbles. Pharmaceutics. 2021;13:2080. https://doi.org/10.3390/pharmaceutics13122080.

22. Borden MA, Kruse DE, Caskey CF, Zhao S, Dayton PA, Ferrara KW. Influence of lipid shell physicochemical properties on ultrasound-induced microbubble destruction. IEEE Trans Ultrason Ferroelectr Freq Control. 2005;52:1992. https://doi.org/10.1109/tuffc.2005.1561668.
23. Zalloum IO, Sojahrood AJ, Paknahad A, Kolios MC, Tsai S, Karshafian R. Controlled tempering of lipid concentration and microbubble shrinkage as a possible mechanism for fine-tuning microbubble size and shell properties. Langmuir. 2023;39:17622. https://doi.org/10.1021/acs.langmuir.3c01599.
24. Lum JS, Stobbe DM, Borden MA, Murray TW. Photoacoustic technique to measure temperature effects on microbubble viscoelastic properties. Appl Phys Lett. 2018;112. https://doi.org/10.1063/1.5005548.
25. Lum JS, Dove JD, Murray TW, Borden MA. Single microbubble measurements of lipid monolayer viscoelastic properties for small-amplitude oscillations. Langmuir. 2016;32:9410. https://doi.org/10.1021/acs.langmuir.6b01882.
26. Polte J. Fundamental growth principles of colloidal metal nanoparticles – a new perspective. CrystEngComm. 2015;17:6809. https://doi.org/10.1039/c5ce01014d.
27. Wuithschick M, Witte S, Kettemann F, Rademann K, Polte J. Illustrating the formation of metal nanoparticles with a growth concept based on colloidal stability. Phys Chem Chem Phys. 2015;17:19895. https://doi.org/10.1039/c5cp02219c.
28. Privman V, Goia DV, Park J, Matijević E. Mechanism of formation of monodispersed colloids by aggregation of nanosize precursors. J Colloid Interface Sci. 1999;213:36. https://doi.org/10.1006/jcis.1999.6106.
29. Gopaiah K, Rao GK, Kalyani G, Kumar GR, Gopaiah M. Asian J Pharm. 2018;12. https://doi.org/10.22377/ajp.v12i04.2840.
30. Dhakar RC. From formulation variables to drug entrapment efficiency of microspheres: a technical review. J Drug Deliv Ther. 2012. https://doi.org/10.22270/jddt.v2i6.160.
31. Hernot S, Klibanov AL. Microbubbles in ultrasound-triggered drug and gene delivery. Adv Drug Deliv Rev. 2008. p. 1153. https://doi.org/10.1016/j.addr.2008.03.005.
32. Yeo Y, Park K. Control of encapsulation efficiency and initial burst in polymeric microparticle systems. Arch Pharm Res. 2004. https://doi.org/10.1007/bf02980037.

33. Rai A, Mhatre S, Chandler CJ, Opere CA, Singh S. Application of quality by design in the development of hydrogen sulfide donor loaded polymeric microparticles. AAPS PharmSciTech. 2024;25. https://doi.org/10.1208/s12249-024-02840-8.

34. Liu M, Dasgupta A, Koczera P, Schipper S, Rommel D, Shi Y, et al. Drug loading in poly(butyl cyanoacrylate)-based polymeric microbubbles. Mol Pharm. 2020;17:2840. https://doi.org/10.1021/acs.molpharmaceut.0c00242.

35. Fateh ST, Moradi L, Kohan E, Hamblin MR, Dezfuli AS. Comprehensive review on ultrasound-responsive theranostic nanomaterials: mechanisms, structures and medical applications. Beilstein J Nanotechnol. 2021. p. 808. https://doi.org/10.3762/bjnano.12.64.

36. Liu HL, Fan C, Ting CY, Yeh C. Combining microbubbles and ultrasound for drug delivery to brain tumors: current progress and overview. Theranostics. 2014. p. 432. https://doi.org/10.7150/thno.8074.

37. Lajoinie G, Luan Y, Gelderblom E, Dollet B, Mastik F, Dewitte H, et al. Non-spherical oscillations drive the ultrasound-mediated release from targeted microbubbles. Commun Phys. 2018;1. https://doi.org/10.1038/s42005-018-0020-9.

38. Husseini GA, Rosa MAD de la, Richardson ES, Christensen DA, Pitt WG. The role of cavitation in acoustically activated drug delivery. J Control Release. 2005;107:253. https://doi.org/10.1016/j.jconrel.2005.06.015.

39. Stride E, Coussios C. Nucleation, mapping and control of cavitation for drug delivery. Nat Rev Phys. 2019. p. 495. https://doi.org/10.1038/s42254-019-0074-y.

40. Ho YJ, Chang HC, Lin CW, Fan C, Lin Y, Wei K, et al. Oscillatory behavior of microbubbles impacts efficacy of cellular drug delivery. J Control Release. 2021;333:316. https://doi.org/10.1016/j.jconrel.2021.03.044.

41. Wrenn SP, Dicker S, Small E, Dan N, Mleczko MJ, Schmitz G, et al. Bursting bubbles and bilayers. Theranostics. 2012. p. 1140. https://doi.org/10.7150/thno.4305.

42. Somaglino L, Mousnier L, Giron A, Urbach W, Tsapis N, Taulier N. In vitro evaluation of polymeric nanoparticles with a fluorine core for drug delivery triggered by focused ultrasound. Colloids Surf B Biointerfaces. 2021;200:111561. https://doi.org/10.1016/j.colsurfb.2021.111561.

43. Chuang CF, Lin CW, Yeh C. Ultrasound-triggered drug release and cytotoxicity of microbubbles with diverse drug attributes. Ultrason Sonochem. 2024;112:107182. https://doi.org/10.1016/j.ultsonch.2024.107182.

44. Nguyen QN, Kanagawa T. Nonlinear ultrasound in liquid containing multiple coated microbubbles: effect of buckling and rupture of viscoelastic shell on ultrasound propagation. Nonlinear Dyn. 2023;111:10859. https://doi.org/10.1007/s11071-023-08228-w.

45. Anderton N, Carlson CS, Matsumoto R, Shimizu R, Poortinga AT, Kudo N, et al. First-cycle oscillation excursions of Pickering-stabilised microbubbles subjected to a high-amplitude ultrasound pulse. Curr Dir Biomed Eng. 2022;8:30. https://doi.org/10.1515/cdbme-2022-1009.

46. Postema M, Schmitz G. Ultrasonic bubbles in medicine: influence of the shell. Ultrason Sonochem. 2007;14:438. https://doi.org/10.1016/j.ultsonch.2006.09.013.

47. Kooiman K, Rooij T van, Qin B, Mastik F, Vos HJ, Versluis M, et al. Focal areas of increased lipid concentration on the coating of microbubbles during short tone-burst ultrasound insonification. PLoS One. 2017;12. https://doi.org/10.1371/journal.pone.0180747.

48. Tsao NH, Hall EAH. Model for microcapsule drug release with ultrasound-activated enhancement. Langmuir. 2017;33:12960. https://doi.org/10.1021/acs.langmuir.7b02954.

49. Lajoinie G, Roovers S, Cock ID, Luan Y, Gelderblom E, Dollet B, et al. Ultrasound-driven microbubbles for local drug delivery. Preprint at https://hal.archives-ouvertes.fr/hal-03240367 (2020).

50. Abstracts from the International Society for Therapeutic Ultrasound Conference 2017. J Ther Ultrasound. 2018;6. https://doi.org/10.1186/s40349-018-0110-x.

51. Tsochatzidi A, Cenci F, Aroniada M, Papageorgiou LG. A novel surrogate-based optimisation framework for pharmaceutical process systems. Int J Pharm. 2025;682:125861. https://doi.org/10.1016/j.ijpharm.2025.125861.

52. Williams B, Cremaschi S. Surrogate model selection for design space approximation and surrogate-based optimization. Comput Aided Chem Eng. 2019. p. 353. https://doi.org/10.1016/b978-0-12-818597-1.50056-4.

53. Yang S, Xing-ming H, Zhu J, Zheng B, Bi W, Wang X, et al. Aspects and implementation of pharmaceutical quality by design from conceptual frameworks to industrial applications. Pharmaceutics. 2025;17:623. https://doi.org/10.3390/pharmaceutics17050623.

54. Clegg LE, Gabhann FM. Molecular mechanism matters: benefits of mechanistic computational models for drug development. Pharmacol Res. 2015. p. 149. https://doi.org/10.1016/j.phrs.2015.06.002.

55. Matsunami K, Vandeputte T, Jiménez AAB, Peeters M, Ghijs M, Hauwermeiren DV, et al. Validation of model-based design of experiments for continuous wet granulation and drying. Int J Pharm. 2023;646:123493. https://doi.org/10.1016/j.ijpharm.2023.123493.

56. Jacob E, Perrillat-Mercerot A, Palgen JL, L'Hostis A, Ceres N, Boissel J, et al. Empirical methods for the validation of time-to-event mathematical models taking into account uncertainty and variability: application to EGFR+ lung adenocarcinoma. BMC Bioinformatics. 2023;24. https://doi.org/10.1186/s12859-023-05430-w.

57. Cucurull-Sanchez L. An industry perspective on current QSP trends in drug development. J Pharmacokinet Pharmacodyn. 2024. https://doi.org/10.1007/s10928-024-09905-y.

58. Schmidt BJ, Papin JA, Musante CJ. Mechanistic systems modeling to guide drug discovery and development. Drug Discov Today. 2012. p. 116. https://doi.org/10.1016/j.drudis.2012.09.003.

59. Joshi A, Ramanujan S, Jin JY. The convergence of pharmacometrics and quantitative systems pharmacology in pharmaceutical research and development. Eur J Pharm Sci. 2023;182:106380. https://doi.org/10.1016/j.ejps.2023.106380.

60. White A, Tolman M, Thames HD, Withers HR, Mason KA, Transtrum MK. The limitations of model-based experimental design and parameter estimation in sloppy systems. PLoS Comput Biol. 2016;12. https://doi.org/10.1371/journal.pcbi.1005227.

61. Velarde G, Sudhir A, Deshmane S, Deshmunkh A, Sharma K, Joshi V. Evaluating XGBoost for balanced and imbalanced data: application to fraud detection. Preprint at http://arxiv.org/abs/2303.15218 (2023).

62. Uçurum M, Özdemir A, Teke Ç, Serencam H, İpek M. Optimization of adsorption parameters for ultra-fine calcite using a Box–Behnken experimental design. Open Chem. 2018;16:992. https://doi.org/10.1515/chem-2018-0114.

63. Nguyen N, Borkowski JJ, Vuong MP. Perspective chapter: cyclic generation of Box–Behnken designs and new second-order designs. In: IntechOpen eBooks. IntechOpen; 2022. https://doi.org/10.5772/intechopen.107178.

64. Roche A. Local optimization of black-box functions with high or infinite-dimensional inputs: application to nuclear safety. Comput Stat. 2017;33:467. https://doi.org/10.1007/s00180-017-0751-1.

65. Chowdhury SM, Abou-Elkacem L, Lee T, Dahl J, Lutz AM. Ultrasound and microbubble mediated therapeutic delivery: underlying mechanisms and future outlook. J Control Release. 2020. p. 75. https://doi.org/10.1016/j.jconrel.2020.06.008.

66. Al-Jawadi S, Thakur S. Ultrasound-responsive lipid microbubbles for drug delivery: a review of preparation techniques to optimise formulation size, stability and drug loading. Int J Pharm. 2020. p. 119559. https://doi.org/10.1016/j.ijpharm.2020.119559.

67. Gupta I, Su X, Jonnalagadda US, Das D, Pramanik M, Kwan J. Investigating the acoustic response and contrast enhancement of drug-loadable PLGA microparticles with various shapes and morphologies. Ultrasound Med Biol. 2021;47:1844. https://doi.org/10.1016/j.ultrasmedbio.2021.02.012.

68. Sergeeva TF, Shirmanova MV, Zagaynova EV, Lukyanov KA. Modern research techniques of apoptotic cell death (review). Sovrem Tekhnol Med. 2015;7:172. https://doi.org/10.17691/stm2015.7.3.21.

69. Baig S, Seevasant I, Mohamad J, Mukheem A, Huri HZ, Kamarul T. Potential of apoptotic pathway-targeted cancer therapeutic research: where do we stand? Cell Death Dis. 2016;7. https://doi.org/10.1038/cddis.2015.275.

70. Wilson MG, Parikh A, Dara A, Beaver AS, Kubanek J. Targeted drug release from stable and safe ultrasound-sensitive nanocarriers. Front Mol Biosci. 2024;11. https://doi.org/10.3389/fmolb.2024.1408767.

71. Lentacker I, Geers B, Demeester J, De Smedt SC, Sanders NN. Design and evaluation of doxorubicin-containing microbubbles for ultrasound-triggered doxorubicin delivery: cytotoxicity and mechanisms involved. Mol Ther. 2009;18:101. https://doi.org/10.1038/mt.2009.160.

72. Goertz DE, Todorova M, Mortazavi O, Agache V, Chen B, Karshafian R, et al. Antitumor effects of combining docetaxel (Taxotere) with the antivascular action of ultrasound stimulated microbubbles. PLoS One. 2012;7. https://doi.org/10.1371/journal.pone.0052307.

73. Zhu Z. Intelligent information management enables quality-by-design in pharmaceutical production. Sci Rep. 2025;15:44201. https://doi.org/10.1038/s41598-025-27879-w.

74. Marschall L, Taylor CS, Zahel T, Kunzelmann M, Wiedenmann A, Presser B, et al. Specification-driven acceptance criteria for validation of biopharmaceutical processes. Front Bioeng Biotechnol. 2022;10:1010583. https://doi.org/10.3389/fbioe.2022.1010583.

75. Dong Y, Liu Z, Li C, Pinter E, Potts AR, Tadey T, et al. Perspectives in modeling and model validation during analytical quality by design chromatographic method evaluation: a case study. AAPS Open. 2021;7. https://doi.org/10.1186/s41120-021-00037-y.

76. Cross JM, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health. 2024;3. https://doi.org/10.1371/journal.pdig.0000651.

77. Bano G. Pharmaceutical development and manufacturing in a quality by design perspective: methodologies for design space description. Preprint at http://hdl.handle.net/11577/3427191 (2019).

78. Bano G, Facco P, Ierapetritou M, Bezzo F, Barolo M. Design space maintenance by online model adaptation in pharmaceutical manufacturing. Comput Chem Eng. 2019;127:254. https://doi.org/10.1016/j.compchemeng.2019.05.019.

79. Bhatia H, Read EK, Agarabi C, Brorson K, Lute S, Yoon S. A design space exploration for control of critical quality attributes of mAb. Int J Pharm. 2016;512:242. https://doi.org/10.1016/j.ijpharm.2016.08.046.

80. Briskot T, Hiltmann D, Rischawy F, Studts J, Kluters S. Qualification of mechanistic models in biopharmaceutical process development. J Pharm Sci. 2024;114:1095. https://doi.org/10.1016/j.xphs.2024.11.021.

81. Richardson S, Irurzun-Arana I, Nowojewski A, Zhou D, Leander J, Tang W, et al. A machine learning approach to population pharmacokinetic modelling automation. 2024. https://doi.org/10.21203/rs.3.rs-4534358/v1.

82. Yu H, Ceross A, Bergmann J. AI for regulatory affairs: balancing accuracy, interpretability, and computational cost in medical device classification. Preprint at http://arxiv.org/abs/2505.18695 (2025).

83. Madathil AP, Luo X, Liu Q, Walker C, Madarkar R, Cai Y, et al. Intrinsic and post-hoc XAI approaches for fingerprint identification and response prediction in smart manufacturing processes. J Intell Manuf. 2024;35:4159. https://doi.org/10.1007/s10845-023-02266-2.

Authorship: NS and SS conceived and designed the study, developed the integrated QbD–XAI framework, supervised experimental and computational work, and led manuscript drafting and revision. JP and ALY developed the AI/ML modeling pipeline and performed surrogate training, validation, and statistical analyses. VP conducted formulation development, data curation, and external validation experiments. SuN (Suvarna Nirmal) contributed to AI generation, implementation of explainability analyses (SHAP, PFI, PDP, LIME), and computational support. SnN (Sunil Nirmal) supported manuscript writing, editing, and literature synthesis. ST contributed to scientific writing, mechanistic interpretation, and critical review of the manuscript. YK, SSr, MKP, MR, and AS assisted in experimental execution, analytical interpretation, and manuscript refinement. All authors reviewed and approved the final manuscript.

Funding:

No external funding was received for this work.

Ethics declarations:

Clinical trial number: not applicable.

Data Availability Statement: No new datasets were generated or analyzed in this study beyond the previously published QbD dataset referenced in the manuscript. The computational scripts and modeling programs used for surrogate development, validation, and explainability analyses will be made available from the corresponding author upon reasonable request.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementrypaper.docx