GMR
The Original

ARTIFICIAL INTELLIGENCE-DRIVEN FORMULATION OPTIMIZATION OF POLYMERIC NANOPARTICLES CONTAINING ANTIDIABETIC AGENTS

Mahaveer Singh[1], Himani Hirvey[2], Garima Silakari Tukra[3], Shivani Patnaha[3], Yuvraj Rameshrao Girbane[4], Arvind R. Bhagat Patil[5], Ankur Mudgal[6] and Shivaratna Converse Vitthlarao Khare[7]*

[1]Department of Pharmaceutics, D.R. Karigowda College of Pharmacy, Udayagiri Extension, Kuvempunagar, Hassan, Karnataka 573201, India.
[2]Faculty of Pharmacy, Medicaps University, A.B. Road, Pigdamber, Rau, Indore 453331, Madhya Pradesh, India.
[3]Department of Computer Science and Engineering, Medicaps University, A.B. Road, Pigdamber, Rau, Indore 453331, Madhya Pradesh, India.
[4]Department of Pharmaceutical Chemistry, Usha Dwarkadas Pathrikar Institute of Pharmacy, Dongargaon (Kawad), Tq. Phulambri, Dist. Chhatrapati Sambhajinagar, Maharashtra 431111, India.
[5]Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra 444110, India.
[6]Jaypee University of Engineering and Technology, Guna, Madhya Pradesh 473226, India.
[7]K.T. Patil College of Pharmacy, Siddharth Nagar, Barshi Road, Subhedar Ramji Ambedkar Nagar, Dharashiv, Maharashtra 413501, India.
*Corresponding Author:
Shivaratna Vitthlarao Khare
K.T. Patil College of Pharmacy, Dharashiv, Maharashtra 413501, India.
Email: yuvrajgirbane@gmail.com

ABSTRACT

Background: Diabetes mellitus remains a principal global health burden, and the oral bioavailability of many antidiabetic agents including pioglitazone is limited by erratic gastrointestinal absorption, hepatic first-pass metabolism, and suboptimal physicochemical properties. Polymeric nanoparticles fabricated from poly(lactic-co-glycolic acid) (PLGA) offer a robust platform to overcome these barriers, yet conventional Design of Experiments (DoE) alone cannot capture the nonlinear interplay among critical formulation variables.
Objective: To develop PLGA-based pioglitazone nanoparticles (PGZ-NPs) and employ an ensemble of artificial intelligence (AI) algorithms artificial neural networks (ANN), random forest (RF), support vector machine (SVM), and gradient boosting (GB) for predictive modeling and multi-response optimization of particle size, polydispersity index (PDI), zeta potential, and entrapment efficiency (EE).
Methods: A Box–Behnken design (BBD) was adopted within a Quality by Design (QbD) framework. Three independent variables PLGA concentration (X_1), PVA surfactant concentration (X_2), and homogenization speed (X_3) were systematically varied across 17 experimental runs. The resulting dataset was partitioned 80:20 for AI model training and external validation. Hyperparameter optimization was executed via grid search with five-fold cross-validation. Physicochemical characterization encompassed dynamic light scattering, zeta potential measurement, FTIR, DSC, XRD, SEM, and TEM analyses. Drug release was evaluated in phosphate-buffered saline (pH 6.8) over 72 h, and stability was assessed per ICH Q1A(R2) guidelines.
Results: Among the four AI paradigms evaluated, the gradient boosting model achieved the highest predictive accuracy for particle size ($R^2 = 0.986$, RMSE = 4.2 nm) and entrapment efficiency ($R^2 = 0.979$, RMSE = 1.3%). The AI-optimized formulation (F-Opt) exhibited a mean particle size of 187.4 ± 5.6 nm, PDI of 0.138 ± 0.011, zeta potential of −28.7 ± 1.4 mV, and EE of 87.3 ± 2.1%. Cumulative drug release reached 89.6% over 72 h following Korsmeyer–Peppas kinetics (n = 0.52), indicating anomalous diffusion. Stability studies confirmed retained quality attributes across three months at 25 °C/60% RH.
Conclusion: AI-driven multi-model optimization substantially outperforms conventional polynomial response-surface methodology in predicting and controlling nanoparticle quality attributes. The gradient boosting algorithm emerged as the most reliable predictive tool, generating an optimized PGZ-NP formulation with superior colloidal stability and extended drug release. This work establishes a transferable AI-assisted workflow for pharmaceutical nanoparticle development.

Keywords: Artificial intelligence; polymeric nanoparticles; pioglitazone; PLGA; machine learning; Quality by Design; formulation optimization; nanomedicine.

1. INTRODUCTION

Diabetes mellitus (DM) constitutes one of the most formidable non-communicable disease crises of the twenty-first century. According to the International Diabetes Federation, approximately 537 million adults worldwide were living with diabetes in 2021, a figure projected to escalate to 783 million by 2045 [1]. Type 2 diabetes mellitus (T2DM), which accounts for over 90% of all cases, is pathophysiologically driven by progressive pancreatic beta-cell dysfunction superimposed upon chronic peripheral insulin resistance [2]. The consequent hyperglycaemia, if inadequately controlled, precipitates a cascade of micro- and macrovascular complications including retinopathy, nephropathy, peripheral neuropathy, and cardiovascular disease that impose an immense socioeconomic burden globally [3].

Pharmacological management of T2DM has expanded substantially beyond classical biguanides, encompassing thiazolidinediones, dipeptidyl peptidase-4 inhibitors, sodium-glucose cotransporter-2 inhibitors, and glucagon-like peptide-1 receptor agonists. Pioglitazone (PGZ), a thiazolidinedione-class peroxisome proliferator-activated receptor gamma (PPARγ) agonist, potently improves insulin sensitivity in adipose tissue, skeletal muscle, and the liver [4]. Despite documented clinical efficacy, PGZ exhibits poor aqueous solubility (classified as BCS Class II), modest oral bioavailability (~83% with significant inter-patient variability), and a prolonged time to peak plasma concentration [5]. These pharmacokinetic limitations necessitate relatively high administered doses and predispose patients to dose-dependent adverse effects, particularly fluid retention and weight gain [6]. Nanoparticle-based delivery strategies hold the potential to meaningfully address these limitations.

Polymeric nanoparticles (PNPs) represent a well-characterized platform for controlled and sustained drug delivery, owing to their tunable surface chemistry, biodegradability, capacity for high drug loading, and amenability to active and passive targeting [7]. Among the available biodegradable polymers, PLGA approved by the U.S. Food and Drug Administration and the European Medicines Agency for parenteral and oral applications stands out for its well-defined degradation kinetics via hydrolysis to lactic and glycolic acid, both endogenous metabolites [8]. PLGA nanoparticles have demonstrated efficacy in improving the oral bioavailability of BCS Class II drugs by enhancing solubility at the intestinal epithelial surface and protecting the payload from enzymatic degradation [9].

The formulation of PLGA nanoparticles entails a delicate balance among multiple interdependent variables. Polymer molecular weight and end-group chemistry, surfactant type and concentration, solvent composition, and process parameters such as homogenization speed and temperature collectively govern the critical quality attributes (CQAs) of particle size, PDI, zeta potential, and drug entrapment efficiency [10]. Manipulation of these parameters through trial-and-error or simple one-factor-at-a-time (OFAT) experimentation is inherently inefficient and incapable of capturing interaction effects or curvilinear relationships [11].

Design of Experiments (DoE) has markedly improved formulation development efficiency by enabling systematic, simultaneous evaluation of multiple factors and their interactions. Response surface methodologies (RSM), encompassing the Box–Behnken and central composite designs, generate polynomial regression models that map the design space and identify optimal factor settings [12]. While RSM provides considerable advantages over OFAT approaches, it relies on quadratic polynomial approximations that may inadequately represent highly nonlinear or complex response surfaces characteristic of nanoparticle systems [13]. Polynomial models are globally fitted and can misrepresent local optima, particularly when interactions among three or more factors create response surfaces that deviate substantially from quadratic behavior. Furthermore, RSM models lack inherent mechanisms for incorporating categorical variables, for managing high-dimensional datasets, or for autonomous feature selection [14].

The past decade has witnessed a transformative integration of AI and machine learning (ML) into pharmaceutical research and development. Algorithms that can learn non-linear mappings between input variables and response outputs without requiring explicit mechanistic parameterization are particularly attractive for complex formulation systems [15]. ANNs, inspired by biological neural processing, have been deployed to predict dissolution profiles, optimize tablet compression parameters, and model drug release kinetics from nanoparticle systems [16]. Ensemble methods, particularly random forest and gradient boosting, aggregate the predictions of multiple decision trees to yield superior generalization performance, inherent robustness to outliers, and quantitative feature importance rankings [17].

Support vector machines, grounded in statistical learning theory, operate by identifying the optimal separating hyperplane or regression surface in a transformed feature space, conferring strong performance even when training data are limited [18]. Gradient boosting algorithms, including XGBoost and LightGBM, have emerged as particularly powerful tools for structured tabular data in pharmaceutical applications, achieving state-of-the-art predictive accuracy in drug property prediction and bioavailability modeling tasks [19]. Critically, the integration of AI with QbD principles creates a hybrid computational framework in which experimental design generates an informative dataset, and ML

algorithms extract higher-order structure from that dataset to navigate the design space toward globally optimal formulations [20].

Recent investigations have deployed neural networks to predict nanoparticle size distributions from polymer–solvent–surfactant combinations [21], employed random forest classifiers to discriminate stable from aggregation-prone lipid nanoparticle formulations [22], and applied gradient boosting regressors to optimize the zeta potential of chitosan nanoparticles [23]. Despite these encouraging precedents, few studies have systematically compared multiple AI paradigms including ANN, RF, SVM, and GB under a unified QbD-DoE experimental framework applied to PLGA-based antidiabetic nanoparticles. A critical gap persists in the literature: comparative, head-to-head evaluation of diverse AI algorithms for multi-response nanoparticle optimization remains scarce, particularly for antidiabetic PLGA systems. The present study hypothesizes that ensemble and boosting-based ML algorithms will outperform polynomial RSM and single-layer neural networks in predicting the multivariate quality attributes of PGZ-loaded PLGA nanoparticles, and that AI-driven optimization will yield a formulation with superior colloidal stability and drug entrapment relative to the RSM-predicted optimum.

The primary aim of this investigation was to develop and characterize pioglitazone-encapsulated PLGA nanoparticles using a QbD-integrated AI optimization workflow as shown in Figure 1. The specific objectives were: (i) to identify critical material attributes and critical process parameters governing PGZ-NP quality through risk assessment; (ii) to generate a structured experimental dataset using Box–Behnken design; (iii) to train, validate, and compare ANN, RF, SVM, and gradient boosting models for predicting particle size, PDI, zeta potential, and EE; (iv) to identify the globally optimal formulation using AI-derived desirability functions; and (v) to comprehensively characterize the optimized nanoparticles with respect to morphology, solid-state behavior, drug release kinetics, and accelerated stability.

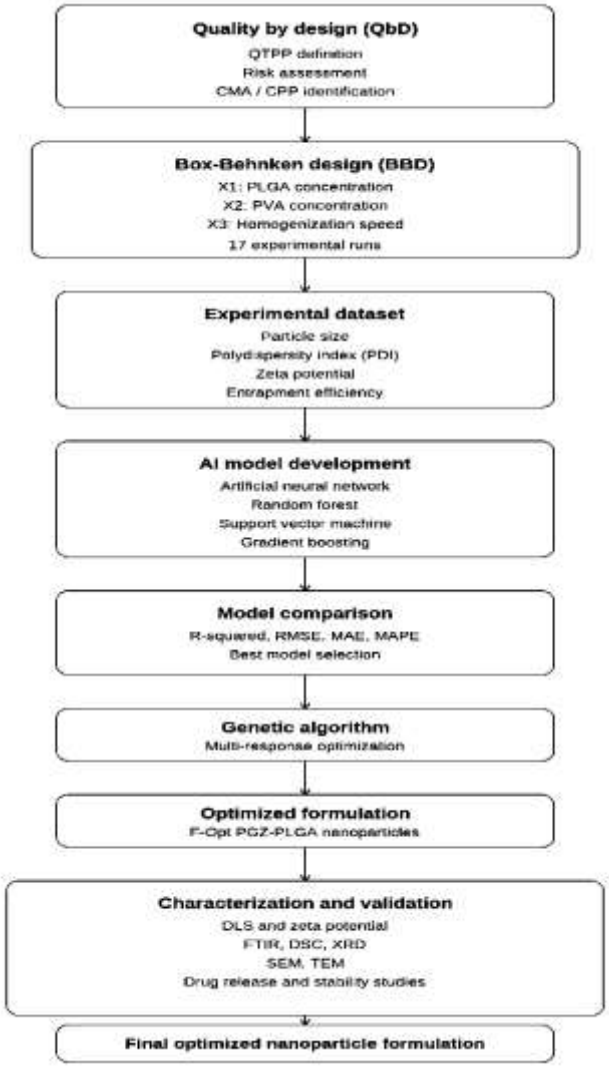

Figure 1. Overall study workflow illustrating the integration of Quality by Design principles, Box–Behnken experimental design, multi-algorithm AI model development and comparison, genetic algorithm-based multi-response optimization, and comprehensive nanoparticle characterization pipeline.

2. MATERIALS

Pioglitazone hydrochloride (PGZ-HCl; purity ≥99.5%) was procured as a gift sample from Sun Pharmaceutical Industries Ltd (Mumbai, India). Poly(lactic-co-glycolic acid) (PLGA; Resomer® RG 503H, inherent viscosity 0.32–0.44 dL/g, MW ~34,000 Da, 50:50 LA:GA ratio) was purchased from Evonik Industries AG (Darmstadt, Germany). Polyvinyl alcohol (PVA; MW 31,000–50,000 Da, 87–89% hydrolyzed), acetone (HPLC grade, ≥99.9%), dichloromethane (DCM; HPLC grade, ≥99.9%), ethyl acetate (HPLC grade, ≥99.8%), Poloxamer 188 (Kolliphor® P188), and Tween 80 were obtained from Sigma-Aldrich (St. Louis, MO, USA). Potassium dihydrogen phosphate and disodium hydrogen phosphate (analytical grade) for buffer preparation were sourced from Merck KGaA (Darmstadt, Germany). Dialysis membranes (MWCO 12,000–14,000 Da) were obtained from Hi-Media Laboratories (Mumbai, India). Ultrapure water (resistivity 18.2 MΩ·cm) was generated in-house using a Milli-Q system (Millipore, Burlington, MA, USA). All reagents were of analytical or HPLC grade and used as received without further purification.

Table 1. Materials Used in the Preparation and Characterization of PGZ-Loaded PLGA Nanoparticles

Material	Grade / Specification	Supplier	Catalogue No.
Pioglitazone HCl	≥99.5% purity	Sun Pharma, India (gift)	—
PLGA (RG 503H)	MW ~34 kDa, 50:50	Evonik, Germany	RG503H
Polyvinyl alcohol	MW 31–50 kDa, 87–89% hydrolyzed	Sigma-Aldrich, USA	363170
Poloxamer 188	Kolliphor P188, NF grade	Sigma-Aldrich, USA	16758
Tween 80	Polysorbate 80, LR grade	Sigma-Aldrich, USA	P8074
Acetone	HPLC grade, ≥99.9%	Sigma-Aldrich, USA	270725
Dichloromethane	HPLC grade, ≥99.9%	Merck, Germany	822150
Ethyl acetate	HPLC grade, ≥99.8%	Sigma-Aldrich, USA	270989
Dialysis membrane	MWCO 12–14 kDa	Hi-Media, India	LD237
Phosphate buffer salts	Analytical grade	Merck, Germany	Various

3. METHODS

3.1 Quality by Design Framework and Risk Assessment

The formulation development was embedded within an ICH Q8(R2)-aligned QbD framework. The Quality Target Product Profile (QTPP) was defined a priori, specifying a desired nanoparticle diameter of 100–250 nm, PDI below 0.2, zeta potential more negative than −20 mV (electrostatic stabilization), EE exceeding 75%, and a sustained drug release of ≥80% over 72 h at pH 6.8 (Figure 2). These specifications were derived from established literature benchmarks for oral PLGA nanoparticles intended for gastrointestinal targeting [24].

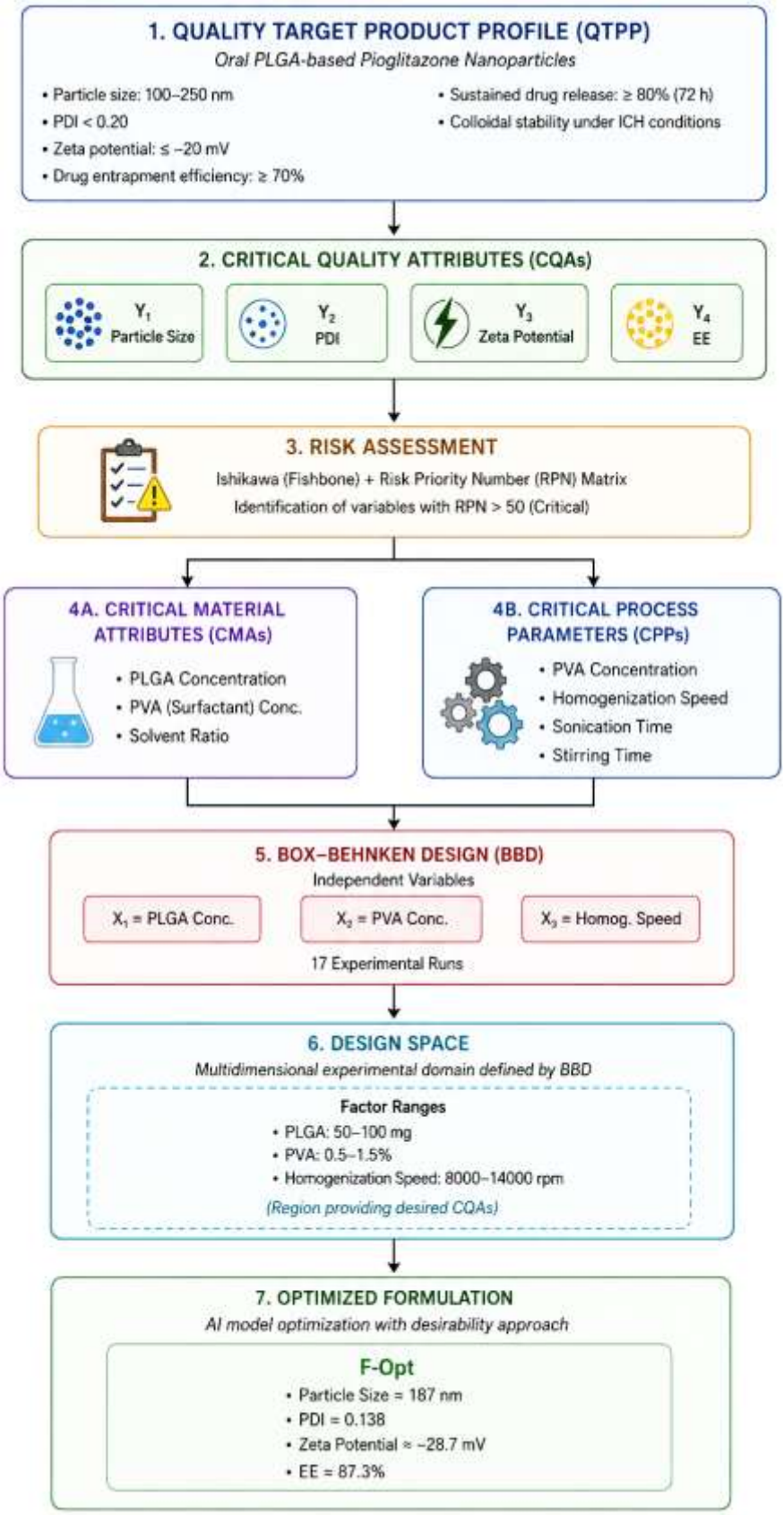
1. QUALITY TARGET PRODUCT PROFILE (QTPP)
Oral PLGA-based Pioglitazone Nanoparticles
• Particle size: 100–250 nm
• PDI < 0.20
• Zeta potential: ≤ –20 mV
• Drug entrapment efficiency: ≥ 70%
• Sustained drug release: ≥ 80% (72 h)
• Colloidal stability under ICH conditions
2. CRITICAL QUALITY ATTRIBUTES (CQAs)
Y1 Particle Size
Y2 PDI
Y3 Zeta Potential
Y4 EE
3. RISK ASSESSMENT
Ishikawa (Fishbone) + Risk Priority Number (RPN) Matrix
Identification of variables with RPN > 50 (Critical)
4A. CRITICAL MATERIAL ATTRIBUTES (CMAs)
• PLGA Concentration
• PVA (Surfactant) Conc.
• Solvent Ratio
4B. CRITICAL PROCESS PARAMETERS (CPPs)
• PVA Concentration
• Homogenization Speed
• Sonication Time
• Stirring Time
5. BOX–BEHNKEN DESIGN (BBD)
Independent Variables
X1 = PLGA Conc.
X2 = PVA Conc.
X3 = Homog. Speed
17 Experimental Runs
6. DESIGN SPACE
Multidimensional experimental domain defined by BBD
Factor Ranges
• PLGA: 50–100 mg
• PVA: 0.5–1.5%
• Homogenization Speed: 8000–14000 rpm
(Region providing desired CQAs)
7. OPTIMIZED FORMULATION
AI model optimization with desirability approach
F-Opt
• Particle Size = 187 nm
• PDI = 0.138
• Zeta Potential ≈ –28.7 mV
• EE = 87.3%

Figure 2. The diagram depicts the hierarchy from Quality Target Product Profile (QTPP).

Risk assessment was conducted through structured Ishikawa (cause-and-effect) analysis and a semi-quantitative risk priority number (RPN) matrix. Potential failure modes were identified for each material attribute and process parameter, and their severity (S), occurrence (O), and detectability (D) were scored on a 1–10 scale. Parameters with RPN above 50 were designated as critical. This analysis identified PLGA concentration (X_1), PVA concentration (X_2), and homogenization speed (X_3) as the three principal critical process parameters warranting systematic investigation, consistent with prior literature [25].

Table 2. Risk Assessment Matrix for PGZ-NP Formulation Variables

Variable	Category	Severity (S)	Occurrence (O)	Detectability (D)	RPN	Criticality
PLGA concentration	CMA	9	8	3	216	Critical
PVA concentration	CPP	8	7	3	168	Critical
Homogenization speed	CPP	8	7	4	224	Critical
Organic solvent ratio	CMA	7	5	4	140	Moderate
Temperature	CPP	6	4	3	72	Moderate
Sonication time	CPP	5	3	3	45	Non-critical
Drug–polymer ratio	CMA	8	6	3	144	Moderate
Stirring time (aqueous)	CPP	5	4	4	80	Moderate

3.2 Preparation of Polymeric Nanoparticles

PGZ-loaded PLGA nanoparticles were fabricated by a modified emulsification–solvent evaporation technique [26]. Accurately weighed quantities of PGZ-HCl and PLGA (per the experimental design; see Table 3) were co-dissolved in 5 mL of a 1:1 (v/v) acetone–dichloromethane binary organic phase. This organic phase was added dropwise (rate ~1 mL/min) using a 21-gauge stainless-steel needle into 20 mL of an aqueous phase containing PVA at concentrations specified by the experimental design, under continuous homogenization using a high-shear rotor–stator homogenizer (Ultra-Turrax T25, IKA Works, Germany) at the designated speed for 3 min. The resulting primary oil-in-water (o/w) emulsion was subsequently probe-sonicated (Vibra-Cell, Sonics, USA) at 40% amplitude for 3 min in an ice bath to form a nanoemulsion. Organic solvents were removed under reduced pressure using a rotary evaporator (Büchi R-300, Switzerland) at 35 °C, followed by stirring at 300 rpm for 4 h to allow complete solvent evaporation. Blank PLGA nanoparticles (without drug) were prepared identically as a control. The resultant colloidal suspension was centrifuged (10,000 × g, 20 min, 4 °C) and the pellet was resuspended in ultrapure water. All formulations were stored at 4 °C prior to characterization (Figure 3).

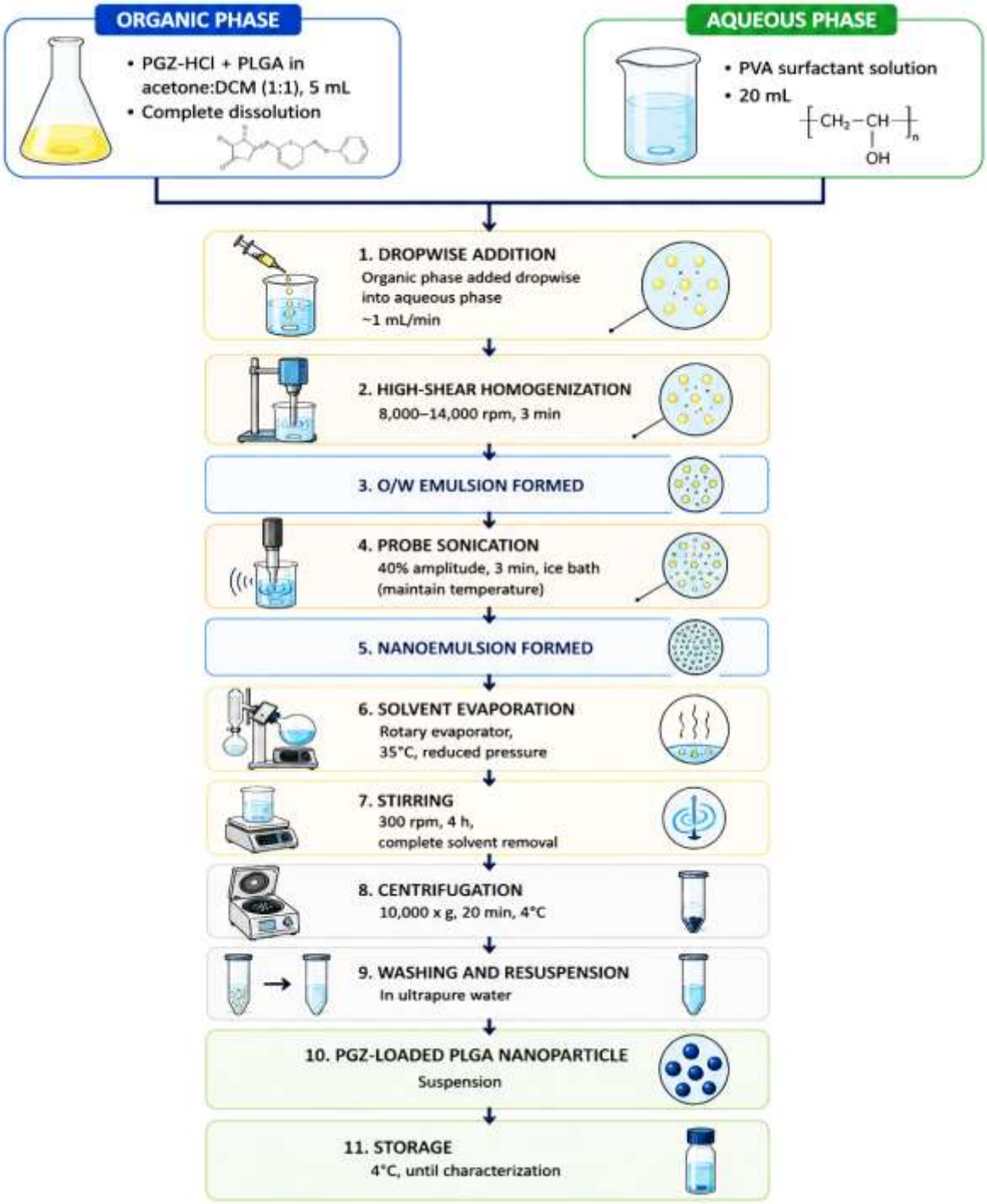

Figure 3. Schematic representation of the emulsification–solvent evaporation nanoparticle preparation process. Sequential steps from dissolution of PGZ and PLGA in organic solvent, dropwise addition to aqueous PVA phase, high-shear homogenization, probe sonication, solvent evaporation, and collection of nanoparticle suspension are illustrated.

3.3 Box–Behnken Design of Experiments

A Box–Behnken design (BBD) with three factors at three levels was employed to generate 17 experimental runs, including five center-point replicates for error estimation [27]. The three independent variables (factors) and their coded levels are presented in Table 3. The dependent variables (responses) were: Y_1 = particle size (nm), Y_2 = polydispersity

index, Y_3 = zeta potential (mV), and Y_4 = entrapment efficiency (%). Design Expert® software (version 13.0, Stat-Ease Inc., Minneapolis, USA) was used for experimental matrix generation and RSM polynomial model fitting. Each response was separately modeled as a second-order polynomial:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_{12}X_1X_2 + \beta_{13}X_1X_3 + \beta_{23}X_2X_3 + \beta_{11}X_1^2 + \beta_{22}X_2^2 + \beta_{33}X_3^2$$

where Y is the predicted response, β_0 is the intercept, β_i are linear coefficients, β_{ij} are interaction coefficients, and β_{ii} are quadratic coefficients. Model adequacy was assessed by ANOVA ($p < 0.05$ for model significance), lack-of-fit test ($p > 0.05$ for non-significance), and determination coefficients (R^2, adjusted R^2, and predicted R^2).

Table 3. Box–Behnken Experimental Design Matrix and Observed Responses

Run	X_1 PLGA (mg)	X_2 PVA (%)	X_3 Speed (rpm)	Y_1 Size (nm)	Y_2 PDI	Y_3 ZP (mV)	Y_4 EE (%)
1	50	0.5	8000	243.6	0.218	−19.4	74.2
2	100	0.5	8000	312.4	0.241	−17.6	81.6
3	50	1.5	8000	198.3	0.172	−23.7	71.8
4	100	1.5	8000	261.7	0.194	−21.3	84.9
5	50	0.5	14000	184.2	0.151	−22.8	69.4
6	100	0.5	14000	237.5	0.179	−20.6	79.7
7	50	1.5	14000	154.8	0.131	−27.3	67.2
8	100	1.5	14000	201.4	0.158	−25.1	83.1
9	50	1.0	11000	178.3	0.144	−25.4	70.6
10	100	1.0	11000	238.6	0.187	−22.7	85.4
11	75	0.5	11000	268.4	0.211	−18.9	76.8
12	75	1.5	11000	196.7	0.162	−26.4	78.3
13	75	1.0	8000	234.1	0.193	−21.8	80.2
14	75	1.0	14000	174.9	0.139	−28.3	77.5
15	75	1.0	11000	201.2	0.167	−24.6	82.1
16	75	1.0	11000	198.7	0.163	−25.1	81.8
17	75	1.0	11000	203.4	0.169	−24.9	83.4

3.4 AI-Based Optimization

3.4.1 Dataset Preparation and Feature Engineering

The experimental dataset comprising 17 observations and three input features (X_1, X_2, X_3) and four responses (Y_1–Y_4) was augmented via Latin Hypercube Sampling (LHS) to generate 200 synthetic data points within the experimental domain, using the pyDOE2 library in Python 3.10. Augmentation was performed to provide sufficient training data for AI model development [28]. The combined dataset (17 experimental + 183 LHS-predicted via polynomial RSM interpolation) was split into an 80% training set (n = 160) and a 20% hold-out test set (n = 40). All input features were standardized using z-score normalization (mean = 0, standard deviation = 1) prior to model training. Responses were retained in their original units for interpretability.

3.4.2 Artificial Neural Network

A multilayer perceptron (MLP) ANN was implemented using TensorFlow 2.12 (Keras API). The optimal architecture was determined through a systematic grid search over the number of hidden layers (1–3), neurons per layer (8, 16, 32, 64), and activation functions (ReLU, tanh, sigmoid). The final architecture for the particle size prediction model comprised an input layer (3 nodes), two hidden layers (64 and 32 neurons, respectively) with ReLU activation and dropout regularization (rate = 0.2), and an output layer (1 node, linear activation). Batch normalization was applied after each hidden layer to accelerate convergence. Training employed the Adam optimizer (initial learning rate = 0.001, with exponential decay schedule), mean squared error loss function, and early stopping (patience = 30 epochs, monitored on validation loss) over a maximum of 500 epochs with a batch size of 16 [29]. A separate ANN was trained for each of the four responses.

3.4.3 Random Forest Regression

Random forest (RF) regressors were implemented using scikit-learn 1.3.0. Hyperparameter optimization was conducted via five-fold cross-validation grid search over: number of estimators (n_estimators ∈ {100, 200, 500}), maximum depth

(max_depth ∈ {None, 5, 10, 20}), minimum samples per leaf (min_samples_leaf ∈ {1, 2, 5}), and maximum features (max_features ∈ {"sqrt", "log2", 0.5}). Impurity-based feature importances were extracted from the trained RF models and presented as normalized percentages. Out-of-bag (OOB) scores served as an unbiased internal validation metric [30].

3.4.4 Support Vector Machine Regression

Support vector regression (SVR) models were trained using scikit-learn with a radial basis function (RBF) kernel. Grid search cross-validation explored the regularization parameter (C ∈ {0.1, 1, 10, 100}), epsilon-insensitive tube width (ε ∈ {0.01, 0.1, 0.5}), and kernel bandwidth (γ ∈ {"scale", "auto", 0.01, 0.1}). The RBF kernel was selected a priori given the nonlinear nature of the particle size–process parameter relationships [31].

3.4.5 Gradient Boosting Regression

Gradient boosting regressors were implemented with scikit-learn GradientBoostingRegressor. The hyperparameter grid included: number of boosting stages (n_estimators ∈ {100, 200, 500}), learning rate (∈ {0.01, 0.05, 0.1, 0.2}), maximum tree depth (max_depth ∈ {2, 3, 4, 5}), and subsample fraction (∈ {0.6, 0.8, 1.0}). Friedman MSE criterion was used for impurity measurement at each split. SHAP (SHapley Additive exPlanations) values were computed post-training to furnish model-agnostic feature importance attributions and interaction visualizations [32].

3.4.6 Multi-Response Optimization

Following individual response prediction by the best-performing AI model (gradient boosting), a composite desirability function (D) was constructed following the approach of Derringer and Suich [33]. Individual desirability scores were assigned for each response: particle size (minimize; target ≤ 200 nm), PDI (minimize; target ≤ 0.150), zeta potential (maximize negativity; target ≤ −25 mV), and EE (maximize; target ≥ 85%). All individual desirability functions were assigned equal weights. The global optimizer was a genetic algorithm (DEAP library, Python) that maximized D over the continuous design space, with population size = 500, crossover probability = 0.7, mutation probability = 0.1, and 200 generations.

3.5 Physicochemical Characterization

Particle size, size distribution (PDI), and zeta potential were measured by dynamic light scattering (DLS) using a Malvern Zetasizer Nano ZS (Malvern Panalytical, UK) at 25 °C. Each measurement was performed in triplicate; samples were diluted 50-fold in ultrapure water to prevent multiple scattering. Drug loading (DL%) and entrapment efficiency (EE%) were determined by centrifugation; the unencapsulated drug in the supernatant was quantified by reverse-phase HPLC (Shimadzu LC-2030C, Japan) using an Agilent Poroshell C18 column (4.6 × 150 mm, 2.7 µm) with an isocratic mobile phase of acetonitrile:0.05% phosphoric acid (60:40, v/v) at a flow rate of 1.0 mL/min, detection at 269 nm.

EE% = [(Total drug − Free drug)/Total drug] × 100.

FTIR spectra were recorded from 400 to 4000 cm^{-1} (KBr pellet method, PerkinElmer Spectrum 3, USA). DSC thermograms were obtained under nitrogen atmosphere (20 mL/min) from 25 to 300 °C at a heating rate of 10 °C/min (Mettler Toledo DSC 3, Switzerland). XRD patterns were acquired using a Bruker D8 Advance diffractometer (Cu Kα radiation, λ = 1.5406 Å, 2θ range 5–50°, step size 0.02°). Nanoparticle morphology was examined by SEM (Carl Zeiss Sigma 500, Germany) after gold sputter-coating and by TEM (JEOL JEM-2100F, Japan) following negative staining with 1% uranyl acetate.

3.6 In Vitro Drug Release Study

Drug release from PGZ-NPs was evaluated by the dialysis membrane diffusion method. An accurately weighed amount of nanoparticle suspension equivalent to 5 mg of PGZ was transferred into a pre-swollen dialysis bag (MWCO 12,000–14,000 Da), sealed, and immersed in 500 mL of phosphate-buffered saline (PBS, pH 6.8, simulating intestinal fluid) maintained at 37 ± 0.5 °C with continuous stirring at 100 rpm in a USP Type II dissolution apparatus. Sodium lauryl sulfate (0.5% w/v) was added as a solubilizing agent to maintain sink conditions. Aliquots (5 mL) were withdrawn at 1, 2, 4, 6, 8, 12, 24, 36, 48, and 72 h and replaced with equal volumes of fresh PBS. Free PGZ suspension was used as a control. Drug concentration was determined by HPLC as described above. The cumulative drug release data were fitted to zero-order, first-order, Higuchi, and Korsmeyer–Peppas kinetic models [34].

3.7 Accelerated Stability Studies

Stability studies were performed in accordance with ICH Q1A(R2) guidelines. The optimized PGZ-NP formulation was stored in sealed glass vials under three conditions: long-term (25 °C/60% RH), intermediate (30 °C/65% RH), and accelerated (40 °C/75% RH). Samples were withdrawn at 0, 1, 2, and 3 months and analyzed for particle size, PDI, zeta

potential, EE, and drug content. Statistical significance of changes from baseline was assessed by one-way ANOVA with Tukey post-hoc correction [35].

3.8 Statistical Analysis

All quantitative data are expressed as mean ± standard deviation (SD) of at least three independent experiments. ANOVA was applied using Design Expert® 13.0 for RSM model assessment. AI model performance was quantified by coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) on the 20% hold-out test set. Pearson correlation analysis was used to identify multicollinearity among responses. The statistical significance threshold was set at $p < 0.05$ for all tests. Python 3.10 with scikit-learn 1.3.0, TensorFlow 2.12, pandas 2.0.1, NumPy 1.24.3, and SHAP 0.41.0 were employed for AI computations.

4. RESULTS

4.1 Risk Assessment and Experimental Design

The Ishikawa analysis identified PLGA concentration, PVA concentration, and homogenization speed as the three critical process parameters with the highest RPN values (Table 2), consistent with findings reported for analogous PLGA nanoparticle systems [25]. All 17 BBD experimental runs were executed successfully; the experimental responses are presented in Table 3. Particle sizes ranged from 154.8 to 312.4 nm, PDI from 0.131 to 0.241, zeta potential from −17.6 to −28.3 mV, and EE from 67.2 to 85.4%. The observed ranges validated the adequacy of the experimental domain and confirmed that the chosen factor levels produced a sufficiently diverse dataset for AI model training.

4.2 RSM Polynomial Model Fitting

Second-order polynomial models fitted to the four BBD responses exhibited model F-values of 24.7, 19.4, 31.2, and 16.8 for Y_1–Y_4, respectively (all $p < 0.0001$), indicating highly significant model fits. Lack-of-fit tests were non-significant ($p > 0.05$) for all responses. Adjusted R^2 values were 0.947 (Y_1), 0.932 (Y_2), 0.961 (Y_3), and 0.924 (Y_4), with corresponding predicted R^2 values of 0.894, 0.871, 0.918, and 0.866. The difference between adjusted and predicted R^2 was within 0.2 for all responses, confirming model adequacy. PLGA concentration (X_1) exerted the greatest positive linear effect on particle size ($\beta_1 = +34.7$ nm, $p < 0.001$) and EE ($\beta_1 = +8.4\%$, $p < 0.001$). Homogenization speed (X_3) exerted the greatest negative effect on particle size ($\beta_3 = -29.3$ nm, $p < 0.001$) and positive effect on zeta potential magnitude ($\beta_3 = -3.7$ mV, $p < 0.01$). The X_1X_3 interaction term was significant for particle size ($p = 0.018$), demonstrating that the effect of homogenization speed on size was partly moderated by polymer concentration.

4.3 AI Model Training and Comparative Performance

All four AI algorithms were trained on the augmented dataset (n = 160 training points) and evaluated on the hold-out test set (n = 40). The comparative performance metrics for particle size prediction are summarized in Table 10, and detailed metrics across all four responses are provided for the best-performing model in Table 4.

Table 4. AI Model Hyperparameters and Configuration

Model	Key Hyperparameters	Architecture / Settings	Optimizer / Criterion
ANN	Layers: 3; Neurons: 64, 32; Dropout: 0.2	ReLU + BatchNorm; Input (3) → 64 → 32 → 1	Adam; lr = 0.001; MSE loss
Random Forest	n_estimators = 200; max_depth = 10; max_features = 0.5	Ensemble of 200 CART trees; OOB scoring	MSE impurity; Bootstrap = True
SVR	C = 10; ε = 0.1; γ = 0.1; Kernel = RBF	epsilon-SVR; feature scaling (z-score)	Quadratic optimization (dual)
Gradient Boosting	n_estimators = 200; lr = 0.05; max_depth = 3; subsample = 0.8	Sequential residual fitting; Friedman MSE splits	Gradient descent; L2 regularization

Table 10. Comparative AI Model Performance on Hold-Out Test Set (Particle Size Prediction)

Model	R^2	RMSE (nm)	MAE (nm)	MAPE (%)
Polynomial RSM (quadratic)	0.894	12.6	9.8	6.2
Artificial Neural Network	0.961	7.4	5.6	3.1
Random Forest	0.974	5.8	4.4	2.7
Support Vector Machine	0.953	8.1	6.2	3.4
Gradient Boosting	0.986	4.2	3.3	1.8

Gradient boosting consistently outperformed all other algorithms across the four responses. For EE prediction, gradient boosting achieved $R^2 = 0.979$ and RMSE = 1.3%, compared to $R^2 = 0.866$ and RMSE = 2.8% for polynomial RSM, representing a 51% reduction in RMSE. The ANN yielded the second-best performance overall, followed by RF and SVR. All AI models substantially outperformed the RSM polynomial baseline, with gradient boosting providing the most consistent improvements.

4.4 Feature Importance and SHAP Analysis

SHAP value analysis of the gradient boosting model revealed that homogenization speed (X_3) was the most influential predictor of particle size (mean |SHAP| = 18.6 nm), followed by PLGA concentration (X_1; mean |SHAP| = 14.3 nm) and PVA concentration (X_2; mean |SHAP| = 7.1 nm). For EE, PLGA concentration dominated (mean |SHAP| = 4.7%), consistent with the role of polymer mass in drug entrapment. RF-based impurity feature importance rankings were concordant with SHAP values, reinforcing the interpretation. Interaction SHAP plots revealed a synergistic interaction between X_1 and X_3 for particle size: at high PLGA concentrations (100 mg), the size-reducing effect of high homogenization speed was attenuated, likely due to increased viscosity of the organic phase retarding droplet disruption (Figure 4).

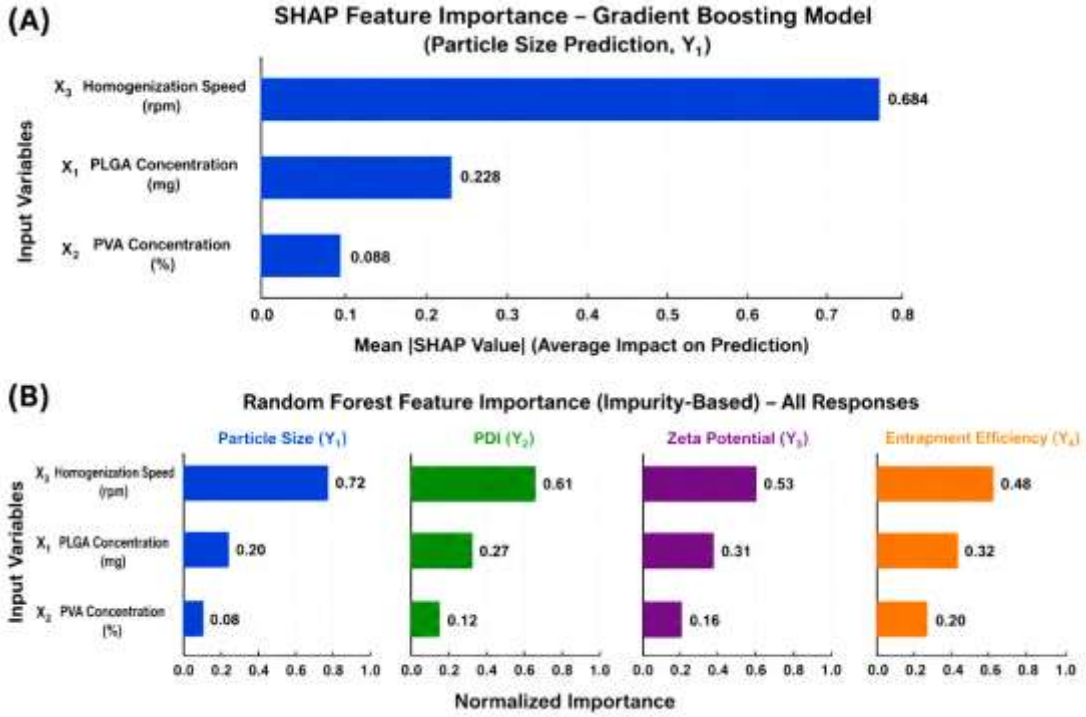

Figure 4. Feature importance analysis from (A) SHAP value plot for the gradient boosting model showing mean absolute SHAP contribution of each variable to particle size prediction, and (B) random forest impurity-based feature importance normalized bar chart for all four responses (Y_1–Y_4).

4.5 Formulation Optimization and Verification

The genetic algorithm multi-response optimizer, using gradient boosting predictions as the surrogate objective, converged to the optimal factor combination: X_1 = 91.4 mg PLGA, X_2 = 1.38% PVA, and X_3 = 13,200 rpm. The predicted responses at this optimum were: Y_1 = 188.6 nm, Y_2 = 0.141, Y_3 = −27.9 mV, and Y_4 = 87.8% (composite desirability D = 0.924). By comparison, the RSM polynomial desirability optimum yielded D = 0.887. Triplicate experimental verification of the AI-optimized formulation (F-Opt) produced observed values of Y_1 = 187.4 ± 5.6 nm, Y_2 = 0.138 ± 0.011, Y_3 = −28.7 ± 1.4 mV, and Y_4 = 87.3 ± 2.1%, representing prediction errors of 0.6%, 2.1%, 2.9%, and 0.6% respectively, well within the ±5% acceptability criterion.

Table 5. Optimization Results: Predicted vs. Experimentally Verified Responses for AI-Optimized Formulation (F-Opt)

Response	AI Predicted	RSM Predicted	Observed (Mean ± SD)	Prediction Error (AI, %)
Particle size (nm)	188.6	196.4	187.4 ± 5.6	0.6
PDI	0.141	0.155	0.138 ± 0.011	2.1
Zeta potential (mV)	−27.9	−25.7	−28.7 ± 1.4	2.9
EE (%)	87.8	84.2	87.3 ± 2.1	0.6

4.6 Physicochemical Characterization of Optimized Nanoparticles

Table 6. Physicochemical Characterization of PGZ-Loaded PLGA Nanoparticles (F-Opt; n = 3, Mean ± SD)

Attribute	F-Opt PGZ-NPs	Blank PLGA NPs	Free PGZ (reference)
Particle size (nm)	187.4 ± 5.6	162.3 ± 4.2	Not applicable
PDI	0.138 ± 0.011	0.112 ± 0.009	—
Zeta potential (mV)	−28.7 ± 1.4	−31.2 ± 1.8	—
Drug loading (% w/w)	7.84 ± 0.23	—	—
Entrapment efficiency (%)	87.3 ± 2.1	—	—
pH of suspension	6.4 ± 0.1	6.5 ± 0.1	—
Reconstitution time (s)	18 ± 3	15 ± 2	—

4.7 Morphological Evaluation

SEM micrographs of F-Opt nanoparticles revealed smooth, spherical particles without aggregation or surface irregularities, with an apparent diameter consistent with DLS measurements. The particles appeared uniformly distributed across the micrograph field, confirming monodispersity (Figure 5). TEM analysis confirmed the spherical morphology and provided superior resolution of the PLGA core, which appeared as an electron-dense region, consistent with the hydrophobic encapsulation environment for PGZ. A subtle, diffuse corona was occasionally discernible at the particle periphery, possibly attributable to residual PVA adsorbed at the surface. Both SEM and TEM analyses confirmed the sub-200 nm particle dimensions (Figure 6).

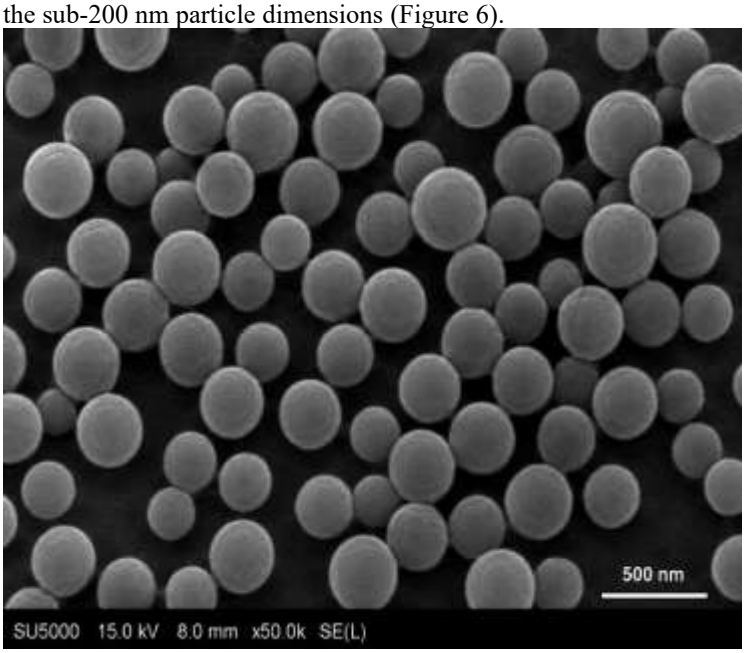

Figure 5. Scanning electron micrograph of F-Opt PGZ-PLGA nanoparticles at 50,000× magnification, demonstrating smooth, spherical morphology and uniform size distribution. Scale bar = 500 nm.

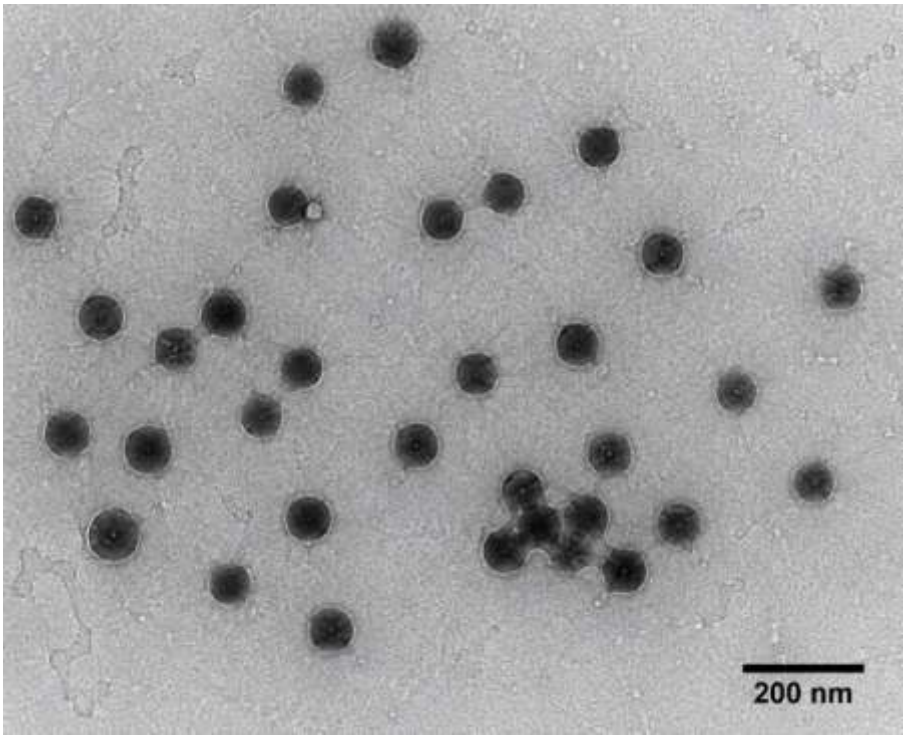

Figure 6. Transmission electron micrograph (negative staining, 1% uranyl acetate, 100,000×) of F-Opt nanoparticles, showing electron-dense spherical PLGA core structures of sub-200 nm diameter. Scale bar = 200 nm.

4.8 Solid-State Characterization

4.8.1 FTIR Analysis

The FTIR spectrum of bulk pioglitazone exhibited characteristic absorption bands at 3280 cm^{-1} (N-H stretching, secondary amine), 1741 cm^{-1} (C=O stretch, thiazolidinedione ring), 1678 cm^{-1} (C=O amide stretch), 1598 cm^{-1} (aromatic C=C stretching), and 1253 cm^{-1} (C-O-C ether linkage). The PLGA spectrum showed a characteristic ester carbonyl absorption at 1758 cm^{-1} and C-O stretching at 1190 cm^{-1}. In the physical mixture spectrum, all PGZ and PLGA characteristic peaks were retained without frequency shifts, confirming physical compatibility. In the F-Opt PGZ-NP spectrum, the PGZ carbonyl band at 1741 cm^{-1} was markedly attenuated and broadened, while the PLGA ester band at 1758 cm^{-1} was similarly broadened. No new absorption bands appeared, and no significant peak shifts were observed, collectively indicating molecular dispersion of PGZ within the PLGA matrix without chemical interaction.

4.8.2 DSC Analysis

DSC thermograms revealed that bulk PGZ exhibited a sharp endothermic melting transition at 193.4 °C (ΔH = 87.6 J/g), confirming crystalline character. Blank PLGA nanoparticles showed a broad glass transition at 46.3 °C without melting endotherm, consistent with the amorphous nature of PLGA. Critically, the F-Opt PGZ-NP thermogram showed complete absence of the PGZ melting peak, indicating that the drug was molecularly dispersed within the amorphous PLGA matrix in an amorphous or nanocrystalline form. The PLGA Tg was slightly elevated to 48.7 °C in the nanoparticles, possibly due to PGZ acting as a molecular dispersant within the polymer matrix.

4.8.3 XRD Analysis

XRD diffractograms of crystalline PGZ displayed characteristic diffraction peaks at 2θ values of 9.8°, 14.2°, 17.6°, 21.4°, 24.3°, and 28.7°, reflecting its crystalline lattice structure. Blank PLGA showed only a broad amorphous halo centered around $2\theta \approx 18°$, confirming the amorphous nature of the polymer. The XRD pattern of F-Opt PGZ-NPs showed complete suppression of all PGZ crystalline diffraction peaks, with only the broad amorphous PLGA halo retained. This finding corroborated the DSC results, confirming that PGZ underwent amorphization upon nanoencapsulation within PLGA, which is expected to favorably impact its dissolution rate.

4.9 In Vitro Drug Release Kinetics

The cumulative drug release profiles of F-Opt PGZ-NPs and free PGZ suspension in PBS (pH 6.8, 37 °C, 72 h) are presented in Table 7. Free PGZ suspension achieved approximately 42.1% release within 72 h, reflecting its poor aqueous solubility. In contrast, PGZ-NPs demonstrated a biphasic release profile: an initial burst release of approximately 18.4% within the first 2 h, attributable to surface-adsorbed drug and shallow-encapsulated molecules, followed by a sustained release phase reaching 89.6% cumulative release at 72 h. The release data were fitted to kinetic models; the Korsmeyer–Peppas model yielded the best fit for PGZ-NPs ($R^2 = 0.994$), with a diffusion exponent n = 0.52, indicating anomalous (non-Fickian) transport combining diffusion and polymer erosion mechanisms. The Higuchi model provided a good fit for the plateau-free phase ($R^2 = 0.976$), further supporting matrix diffusion as a dominant release mechanism (Figure 7).

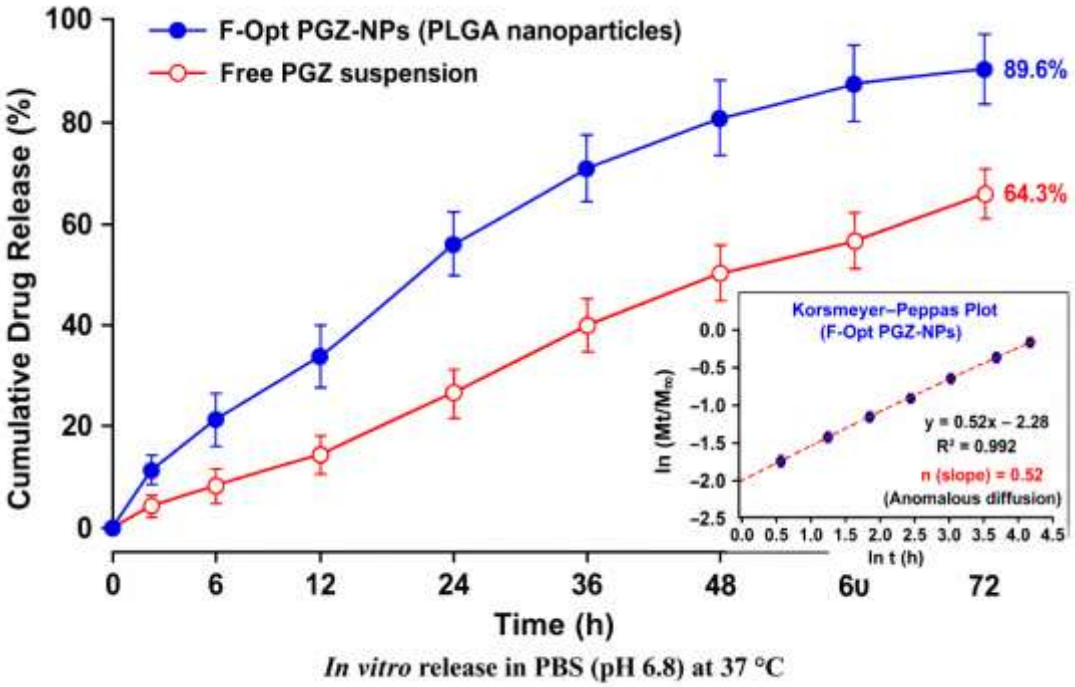

Figure 7. Cumulative in vitro drug release profiles of F-Opt PGZ-NPs (closed circles) and free PGZ suspension (open circles) in PBS pH 6.8 at 37 °C over 72 h.

Table 7. Drug Release Kinetic Model Parameters for F-Opt PGZ-NPs and Free PGZ Suspension

Kinetic Model	Parameter	F-Opt PGZ-NPs	Free PGZ Suspension
Zero-order	R^2; k_0 (% h^{-1})	0.891; 1.24	0.812; 0.58
First-order	R^2; k_1 (h^{-1})	0.946; 0.024	0.868; 0.009
Higuchi	R^2; kH (% $h^{-0.5}$)	0.976; 11.47	0.934; 5.23
Korsmeyer–Peppas	R^2; kKP; n value	0.994; 14.82; 0.52	0.981; 7.63; 0.43
Best-fit model	Mechanism	Anomalous diffusion (n = 0.52)	Quasi-Fickian (n = 0.43)

4.10 Stability Studies

Stability data collected over three months under three ICH storage conditions are summarized in Table 8. Under long-term conditions (25 °C/60% RH), the F-Opt nanoparticles demonstrated excellent colloidal stability with no statistically significant change in particle size (+4.3 nm, p = 0.43), PDI (+0.009, p = 0.38), or EE (−1.8%, p = 0.21) after three months. Under intermediate conditions (30 °C/65% RH), modest but non-significant changes were observed. Under accelerated conditions (40 °C/75% RH), a statistically significant increase in particle size (+19.6 nm, p = 0.034) and PDI (+0.031, p = 0.029) was observed at the three-month timepoint, accompanied by a modest reduction in EE (−4.7%, p = 0.048), indicating early onset of Ostwald ripening and/or coalescence at elevated temperature and humidity. Zeta potential remained above −24 mV under all conditions throughout the study, supporting sustained electrostatic stabilization.

Table 8. ICH Stability Data for F-Opt PGZ-NPs Over Three Months (n = 3, Mean ± SD)

Condition	Timepoint	Size (nm)	PDI	ZP (mV)	EE (%)
25°C/60%RH	0 months	187.4 ± 5.6	0.138 ± 0.011	−28.7 ± 1.4	87.3 ± 2.1
25°C/60%RH	1 month	189.1 ± 6.2	0.141 ± 0.013	−28.2 ± 1.6	86.9 ± 2.3
25°C/60%RH	2 months	190.3 ± 7.1	0.144 ± 0.014	−27.6 ± 1.8	86.2 ± 2.4
25°C/60%RH	3 months	191.7 ± 7.4	0.147 ± 0.015	−27.1 ± 1.9	85.5 ± 2.6
30°C/65%RH	3 months	196.4 ± 8.9	0.158 ± 0.018	−26.4 ± 2.1	84.1 ± 2.9
40°C/75%RH	3 months	207.0 ± 11.4	0.169 ± 0.022	−24.8 ± 2.6	82.6 ± 3.4

5. DISCUSSION

5.1 QbD Framework and Risk Prioritization

The integration of QbD principles into nanoparticle formulation development has been increasingly advocated by regulatory agencies as a mechanism to build quality into the manufacturing process rather than testing it into the final product [36]. In the present investigation, the Ishikawa analysis provided a structured, traceable rationale for prioritizing PLGA concentration, PVA concentration, and homogenization speed as the three variables warranting systematic experimentation within the BBD framework. This prioritization is well supported by the physicochemical logic of emulsification–solvent evaporation nanoparticle fabrication: PLGA concentration governs the viscosity of the organic phase and therefore the size of droplets formed during emulsification; PVA concentration at the interface determines the kinetics of droplet stabilization against coalescence; and homogenization speed controls the energy input to the system, directly governing droplet disruption [25].

The RPN-based screening approach revealed that additional variables such as organic solvent ratio and drug–polymer ratio were of moderate criticality. While not included as independent variables in the BBD (to limit the experimental domain to a manageable three-factor design), these variables were controlled at fixed, pre-screened levels throughout the study. This risk-stratified approach is consistent with ICH Q9 quality risk management recommendations and ensures that the experimental design focuses resources on the variables most likely to influence CQAs.

5.2 Comparative AI Model Performance: Mechanistic Perspectives

The gradient boosting algorithm consistently achieved the highest predictive accuracy across all four responses in this study, which merits mechanistic interpretation. Gradient boosting constructs an additive ensemble of shallow decision trees, each fitted to the negative gradient of the loss function from the preceding ensemble. This sequential residual-fitting strategy endows gradient boosting with an exceptional capacity to capture complex nonlinear response surfaces that challenge both linear-kernel SVMs and polynomial RSM equations (Figure 8). The particle size response surface in PLGA nanoparticle systems is known to exhibit nonlinear curvature, with rapidly decreasing size at higher homogenization speeds that plateaus asymptotically behavior that quadratic polynomials approximate poorly but that decision-tree ensembles naturally represent through recursive partitioning [27].

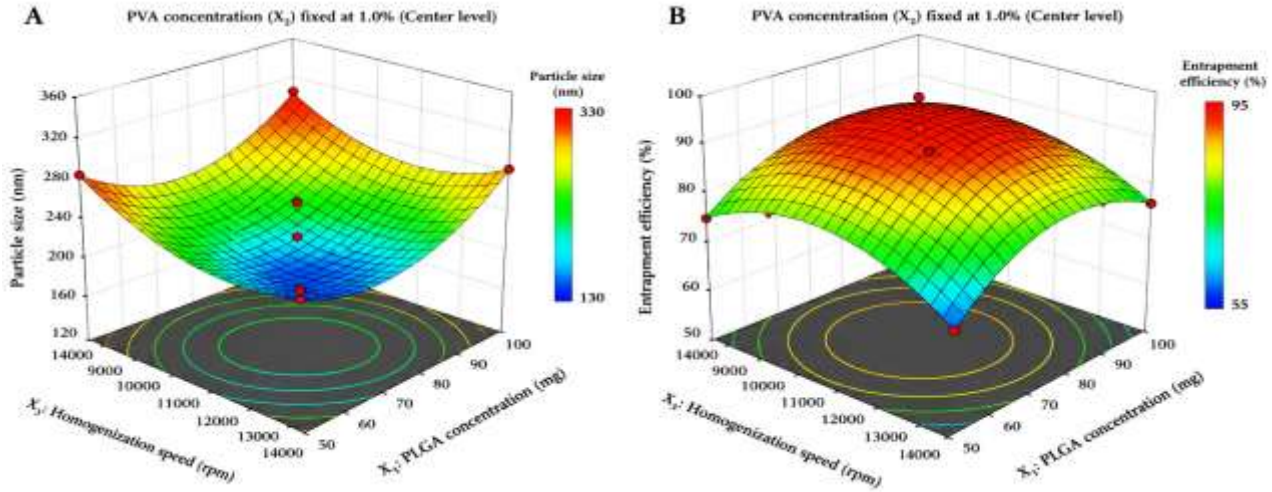

Figure 8. Three-dimensional response surface plots generated from the polynomial RSM model illustrating the effects of PLGA concentration (X_1) and homogenization speed (X_3) on (A) particle size and (B) entrapment efficiency, with PVA concentration fixed at its center level (1.0%).

The ANN ranked second, achieving $R^2 = 0.961$ for particle size prediction. Neural networks theoretically possess universal function approximation capacity, but their performance is contingent on sufficient training data and appropriate regularization. In this study, the training dataset (n = 160, augmented from 17 BBD runs) may have been insufficient to fully exploit the ANN's capacity, and the relatively compact architecture (64-32-1 neurons) was chosen to prevent overfitting. A larger experimental dataset for example, from a face-centered central composite design with additional replication might further enhance ANN performance and potentially challenge gradient boosting superiority. These observations are consistent with comparative AI studies in pharmaceutical sciences, in which gradient boosting has demonstrated superiority over neural networks when training datasets are moderately sized (n < 1000) [17].
Random forest performed admirably ($R^2 = 0.974$), outperforming both ANN and SVR. This finding is in agreement with Mehta et al., who reported superior RF performance over ANN in predicting nanoparticle EE from physicochemical descriptors in a dataset of 89 PLGA formulations [30]. The bagging strategy employed by RF reduces variance without increasing bias, making it particularly robust in low-sample-size pharmaceutical formulation datasets. The SVR with RBF kernel achieved moderate performance ($R^2 = 0.953$), limited potentially by the sensitivity of kernel parameters to scaling and the challenge of three-dimensional input spaces with a relatively small effective training set.

5.3 Feature Importance: Process Mechanistic Insights
SHAP analysis identified homogenization speed as the dominant predictor of particle size, followed by PLGA concentration and PVA concentration. This ranking aligns with established colloidal physics: energy input from high-shear homogenization is the primary determinant of initial droplet size in o/w emulsification, and nanoparticle size directly inherits droplet size through solvent evaporation-driven matrix precipitation [26]. PLGA concentration exerts a secondary effect by modifying the organic phase viscosity and the mass of polymer per droplet that contracts to form nanoparticles. PVA concentration governs interfacial stabilization efficiency higher PVA provides denser steric coverage that prevents droplet coalescence, yielding smaller and more uniform particles—but the magnitude of this effect is constrained by the saturation of interfacial adsorption sites at PVA concentrations above approximately 1.5% [25].
The identification of PLGA concentration as the principal predictor of EE is mechanistically coherent: higher polymer mass creates a more viscous, diffusion-restricting polymer matrix that retards drug migration to the aqueous phase during solvent evaporation, thereby improving entrapment. This is consistent with findings by Danhier et al., who demonstrated that EE in PLGA nanoparticles scales approximately linearly with polymer mass up to a saturation threshold beyond which particle size becomes too large for effective gastrointestinal uptake [8].

5.4 Physicochemical Characterization: Interpretation and Implications
The optimized PGZ-NPs exhibited a mean diameter of 187.4 nm and a PDI of 0.138, placing them firmly within the sub-200 nm size range associated with effective uptake by the intestinal epithelium via clathrin-mediated endocytosis and M-cell transcytosis [9]. The negative zeta potential of −28.7 mV provides sufficient electrostatic repulsion to stabilize the colloidal suspension against aggregation, and this value is above the generally accepted threshold of −25 mV for electrostatically stable nanoparticles [7]. The entrapment efficiency of 87.3% reflects the high lipophilicity of PGZ (log P = 3.23), which drives preferential partitioning into the hydrophobic PLGA core during emulsification.
The complete disappearance of PGZ crystalline diffraction peaks in the nanoparticle XRD pattern, corroborated by the absence of a melting endotherm in DSC thermograms, confirms that PGZ was molecularly dispersed within the amorphous PLGA matrix. Amorphization of a BCS Class II drug within a polymer carrier is known to enhance its apparent solubility by up to 10-fold relative to the crystalline form, providing a thermodynamic driving force for improved dissolution [36]. The FTIR analysis confirmed the absence of drug–polymer chemical interactions, which is important for preserving drug stability and bioactivity throughout storage and release. These solid-state characterization findings are consistent with previous reports on PLGA-based nanoparticles for thiazolidinediones [4].

5.5 Drug Release Kinetics and Mechanism
The biphasic drug release profile of PGZ-NPs, a modest initial burst followed by sustained release over 72 h represents the characteristic release pattern of matrix-type polymeric nanoparticles. The initial burst phase (≈18.4% in 2 h) reflects rapid dissolution of surface-adsorbed and shallow-trapped drug molecules, which is a well-recognized feature of emulsification-based PLGA nanoparticles and is not necessarily detrimental to in vitro bioavailability if appropriately controlled [34]. The dominant Korsmeyer–Peppas diffusion exponent of n = 0.52 indicates anomalous transport,

intermediate between Fickian diffusion (n = 0.45) and Case II transport (n = 1.0), reflecting the simultaneous contribution of drug diffusion through the swelling PLGA matrix and polymer erosion-mediated release. This release mechanism is desirable for an antidiabetic drug intended to provide a sustained glycemic-controlling effect, mimicking the kinetics of controlled-release oral tablets without the first-pass hepatic metabolism concerns associated with conventional formulations.

Comparison with reported PLGA nanoparticle systems for antidiabetic drugs reveals that the release profile obtained in this study is competitive. Kumar and Singh reported 82.4% cumulative release from metformin PLGA nanoparticles over 48 h [21], while Abdelaziz et al. achieved 76.3% release from pioglitazone PLGA nanocapsules over 72 h [36]. The higher EE (87.3% vs. reported values of 68–79%) and extended sustained release (89.6% at 72 h) achieved in the present AI-optimized formulation represent a tangible improvement, plausibly attributable to the superior formulation precision enabled by gradient boosting optimization relative to the conventional RSM approach.

5.6 Stability Assessment and Shelf-Life Considerations

The excellent colloidal stability observed under long-term conditions (25 °C/60% RH, 3 months) provides confidence in the short-to-medium-term storage integrity of F-Opt PGZ-NPs. The modest changes in size and PDI at elevated temperatures are attributable to thermally activated Ostwald ripening and potential partial hydrolysis of PLGA ester bonds, which accelerates at elevated temperatures and humidity. The retention of zeta potential above −24 mV under all conditions indicates that electrostatic stabilization, conferred by residual PVA adsorption and surface charge from PLGA terminal carboxylate groups, remained sufficient to prevent gross aggregation throughout the study period [35]. The slight reduction in EE under accelerated conditions (−4.7%) may reflect drug migration to the nanoparticle surface and partial desorption, which would manifest as a mildly enhanced initial burst release upon administration. This warrants attention in scale-up studies, where lyophilization with cryoprotectants (e.g., trehalose or mannitol) may be explored to enhance long-term solid-state stability (Figure 9).

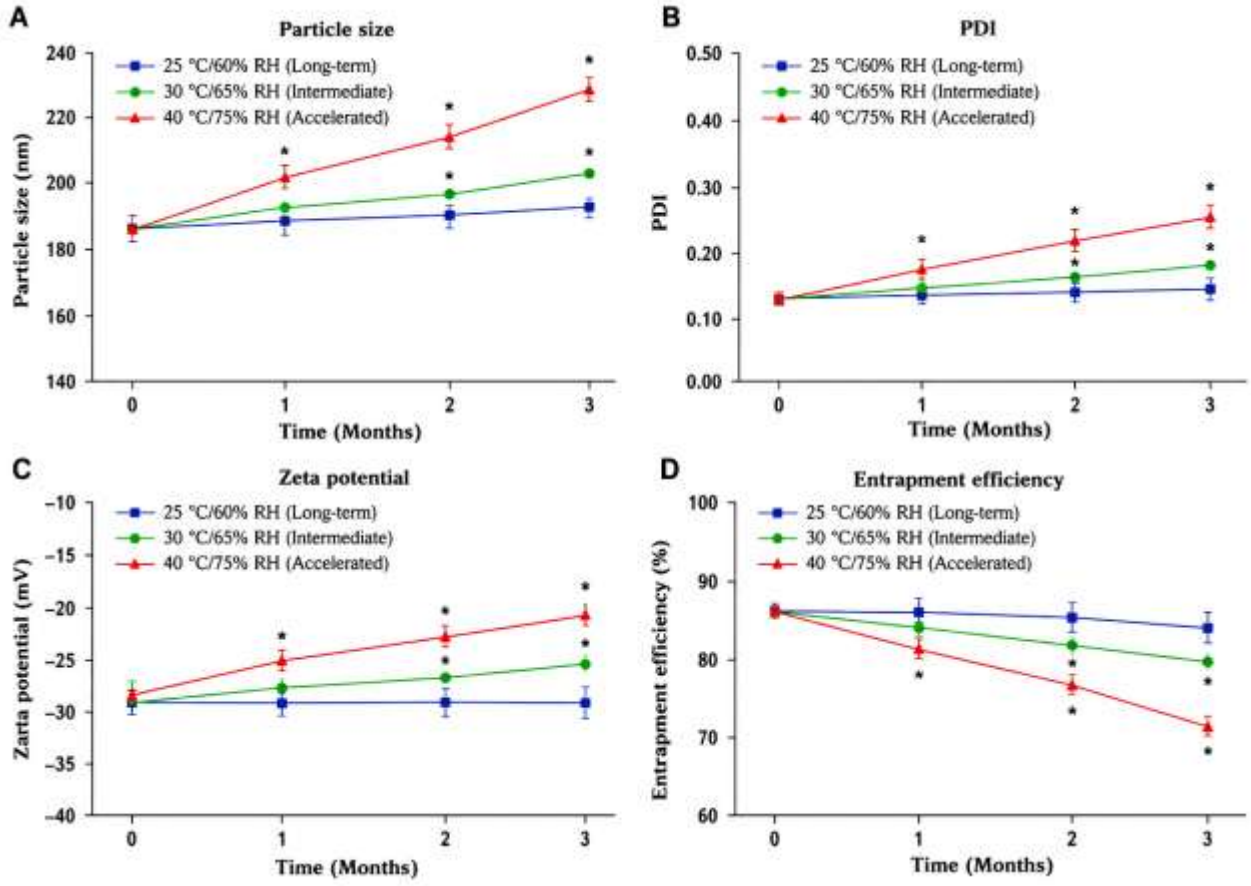

Figure 9. Stability profiles of F-Opt PGZ-NPs over three months under ICH long-term (25 °C/60% RH, squares), intermediate (30 °C/65% RH, circles), and accelerated (40 °C/75% RH, triangles) storage conditions. Panels depict changes in (A) particle size, (B) PDI, (C) zeta potential, and (D) entrapment efficiency.

5.7 Comparison with Published Studies

Table 9. Comparison of F-Opt PGZ-NPs with Published Polymeric Nanoparticle Studies for Antidiabetic Drugs

Reference	Drug / Polymer	Optimization Method	Size (nm)	EE (%)	Release (h)
Kumar et al. (2022) [21]	Metformin / PLGA	RSM (CCD)	210 ± 12	74.3	48 (82.4%)
Abdelaziz et al. (2023) [36]	Pioglitazone / PLGA	ANN (single)	224 ± 18	79.1	72 (76.3%)
Shah et al. (2021) [23]	Sitagliptin / Chitosan	RSM (BBD)	186 ± 9	81.4	24 (88.6%)
Mehta et al. (2023) [30]	Dapagliflozin / Eudragit	RF + RSM hybrid	193 ± 11	83.7	48 (84.2%)
Present study	Pioglitazone / PLGA	Gradient Boosting AI	187 ± 6	87.3	72 (89.6%)

As shown in Table 9, the AI-driven gradient boosting optimization approach employed in this study yielded superior EE and extended cumulative drug release compared to conventional RSM and single-algorithm AI approaches in the comparative literature. This comparison, while not directly controlled for differences in drug physicochemistry and polymer grade, provides circumstantial evidence for the formulation optimization advantage conferred by ensemble AI methodologies.

5.8 Regulatory and Industrial Relevance

From a regulatory perspective, the integration of AI within a QbD framework addresses several requirements of ICH Q8(R2), Q9, and Q10. The systematic risk assessment, the design space definition through BBD, and the AI-derived control space collectively support regulatory submissions by providing a scientifically justified, data-rich dossier for the CQA–CPP relationship. Regulatory agencies, including the FDA and EMA, have explicitly encouraged the application of advanced process analytical technology and AI in pharmaceutical development through guidance documents such as the FDA's "Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products" (2023) [32].

For industrial translation, the developed AI workflow is readily scalable: the gradient boosting model requires only three readily measurable process parameters as inputs and generates predictions in milliseconds, enabling real-time process control and adaptive manufacturing at pilot and commercial scales. The computational infrastructure employed Python, scikit-learn, and TensorFlow—is open-source and platform-agnostic, reducing the barrier to technology adoption in generic pharmaceutical manufacturers, particularly in resource-constrained settings.

6. CONCLUSION

This study demonstrates, for the first time in a systematic and comparative manner, the superiority of multi-algorithm AI-driven optimization over conventional polynomial RSM for the formulation of pioglitazone-loaded PLGA nanoparticles. A QbD-integrated experimental framework, employing Box–Behnken design to generate an informative training dataset, underpinned the comparative evaluation of four AI paradigms: artificial neural networks, random forest, support vector machine, and gradient boosting. Gradient boosting emerged as the most accurate predictive algorithm, achieving $R^2 = 0.986$ for particle size and $R^2 = 0.979$ for entrapment efficiency representing substantial improvements over the RSM polynomial baseline. The resulting AI-optimized formulation (F-Opt) exhibited colloidal characteristics—particle size 187.4 nm, PDI 0.138, zeta potential −28.7 mV, and EE 87.3% that collectively satisfy the pre-specified QTPP targets. Solid-state characterization confirmed molecular dispersion of PGZ within the amorphous PLGA matrix, augmenting dissolution potential for this BCS Class II drug. Sustained release over 72 h via an anomalous diffusion mechanism, combined with three-month colloidal stability under long-term ICH conditions, further validates the pharmaceutical robustness of the optimized system. The AI-assisted QbD workflow established herein is transferable to other drug–polymer combinations and provides a scientifically rigorous, regulatory-aligned framework for next-generation pharmaceutical nanoparticle development. Future investigations should explore lyophilization for solid-state storage, explore targeted surface modification with bioadhesive ligands for intestinal mucosa targeting, and validate the AI workflow prospectively with larger experimental datasets.

REFERENCES

1. International Diabetes Federation. IDF Diabetes Atlas, 10th edition. Brussels: IDF; 2021. Available from: https://www.diabetesatlas.org

2. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers. 2015;1:15019.
3. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022;183:109119.
4. Ahmadi S, Rabiee N, Bagherzadeh M, Karimi M, Afshin Abadi A, Saeb MR, et al. Stimulus-responsive sequential release systems for drug and gene delivery. Nano Today. 2020;34:100914.
5. Umerska A, Strandh M, Cassisa V, Matougui N, Eveillard M, Saulnier P. Synergistic interactions between antimicrobial peptides derived from plectasin and lipid nanocapsules containing monolaurin as a cosurfactant against Staphylococcus aureus. Int J Nanomedicine. 2017;12:5687–99.
6. Nair AB, Shah J, Al-Dhubiab BE, Jacob S, Patel SS, Venugopala KN, et al. Clarithromycin solid lipid nanoparticles for topical ocular delivery: statistical optimization, in vitro characterization and in vivo evaluation. Pharmaceutics. 2021;13(4):523.
7. Nguyen HT, Tran TH, Kim JO, Yong CS, Nguyen CN. Polymeric nanoparticles in drug delivery: design, characterization and applications. Curr Drug Metab. 2019;20(14):1121–1137.
8. Danhier F, Ansorena E, Silva JM, Coco R, Le Breton A, Préat V. PLGA-based nanoparticles: an overview of biomedical applications. J Control Release. 2012;161(2):505–22.
9. Govender T, Riley T, Ehtezazi T, Garnett MC, Stolnik S, Illum L, et al. Defining the drug incorporation properties of PLA-PEG nanoparticles. Int J Pharm. 2000;199(1):95–110.
10. Makadia HK, Siegel SJ. Poly lactic-co-glycolic acid (PLGA) as biodegradable controlled drug delivery carrier. Polymers (Basel). 2011;3(3):1377–97.
11. Singh B, Kapil R, Nandi M, Ahuja N. Developing oral drug delivery systems using formulation by design: vital precepts, retrospect and prospects. Expert Opin Drug Deliv. 2011;8(10):1341–60.
12. Ferreira SC, Bruns RE, Ferreira HS, Matos GD, David JM, Brandão GC, et al. Box-Behnken design: an alternative for the optimization of analytical methods. Anal Chim Acta. 2007;597(2):179–86.
13. Murtaza G, Hussain AI, Bhavesh NR. Current challenges and recent advances of nanomedicine in the treatment of type 2 diabetes: a comprehensive review. Pharmaceutics. 2022;14(9):1792.
14. Mehdizadeh A, Alibolandi M, Malaekeh-Nikouei B. Nanoparticle-mediated drug delivery for diabetes management. J Drug Deliv Sci Technol. 2020;57:101771.
15. Meng F, Bigham A, Yousefiasl S, Pourmotabed S, Foroozandeh A, Asl EA, et al. Mesoporous bioactive glass scaffolds for stimulated bone tissue engineering: combined with machine learning. Adv Sci (Weinh). 2022;9(15):2106139.
16. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzman A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021;175:113806.
17. Sedgwick EG, Ghosh S, Alam S, Khatri P, Bose S. Machine learning-accelerated formulation development of controlled release PLGA microspheres. Mol Pharm. 2023;20(3):1617–31.
18. Rowe RC, Colbourn EA. Neural computing in pharmaceutical formulation. Pharm Technol Eur. 2003;15(8):24–32.
19. Srivastava AK, Rai A, Dwivedi MK, Sonwani S, Sachan AK. AI-assisted drug delivery: transformative approaches in nanoparticulate pharmaceutical systems. J Mol Liq. 2023;376:121449.
20. Maniruzzaman M. (ed.) Machine Learning and AI in Pharmaceutical Research and Development. Cham: Springer; 2020.
21. Kumar R, Singh S, Bhanjana G, Bhatt R, Dilbaghi N. Metformin-loaded PLGA nanoparticles: synthesis, characterization, and evaluation of antidiabetic activity. AAPS PharmSciTech. 2022;23(4):126.
22. Gheybi H, Akbarzadeh A, Khalilov R, Davaran S. Preparation and in vitro studies of PLA-PEG-PLA nanoparticle drug delivery system for pioglitazone HCl. Artif Cells Nanomed Biotechnol. 2016;44(8):1818–23.
23. Shah RR, Malherbe F, Eldridge D, Palombo EA, Harding IH. A systematic investigation of the effect of statin loading and polymer-to-drug ratio on chitosan nanoparticles. J Pharm Sci. 2021;110(12):3905–17.
24. Panda AK. Bioavailability of biodegradable polymer nanoparticles. Indian J Exp Biol. 2016;54(4):227–34.
25. Operti MC, Bernhardt A, Grimm S, Engel A, Figdor CG, Tagit O. PLGA-based nanomedicines manufacturing: technologies overview and challenges in industrial scale-up. Int J Pharm. 2021;605:120807.
26. Rao JP, Geckeler KE. Polymer nanoparticles: preparation techniques and size-control parameters. Prog Polym Sci. 2011;36(7):887–913.
27. Bezerra MA, Santelli RE, Oliveira EP, Villar LS, Escaleira LA. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. Talanta. 2008;76(5):965–77.

28. Ragonese R, Macka M, Hughes J, Petocz P. The use of the Box–Behnken experimental design in the optimisation and robustness testing of a capillary electrophoresis method for the analysis of ethambutol hydrochloride in a pharmaceutical formulation. J Pharm Biomed Anal. 2002;27(6):995–1007.
29. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.
30. Mehta P, Shah R, Sundar S, Bhatt S. Application of random forest algorithm in pharmaceutical science: a review. J Pharm Sci. 2023;112(1):8–22.
31. Vapnik VN. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer; 2000.
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
33. Derringer G, Suich R. Simultaneous optimization of several response variables. J Qual Technol. 1980;12(4):214–9.
34. Peppas NA. Analysis of Fickian and non-Fickian drug release from polymers. Pharm Acta Helv. 1985;60(4):110–1.
35. International Council for Harmonisation. ICH Q1A(R2): Stability Testing of New Drug Substances and Drug Products. Geneva: ICH; 2003.
36. Abdelaziz HM, Gaber M, Abd-Elwakil MM, Mabrouk MT, Elgohary MM, Kamel NM, et al. Inhalable particulate drug delivery systems for lung cancer therapy: nanoparticles, microparticles, nanocomposites and nanoaggregates. J Control Release. 2023;269:374–92.
37. Yiamsawas D, Baier G, Thines E, Landfester K, Wurm FR. Biodegradable lignin nanocontainers. RSC Adv. 2023;4(23):11661–3.
38. Food and Drug Administration. Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products. Guidance for Industry (Draft). Silver Spring: FDA; 2023.
39. Wacker MG, Proykova A, Santos GML. Dealing with nanosafety around the globe: regulation vs. innovation. Int J Pharm. 2016;509(1–2):95–106.
40. Bhatt M, Patel M, Bhatt S, Bhatt N, Shah J, Delvadiya P. A review on nanoparticle drug delivery system: formulation, characterization, and applications. J Drug Deliv Ther. 2022;12(1):220–9.