

Role Of Artificial Intelligence And Machine Learning In Nanoparticle Design And Optimization: A Review

Sonali Aher^{1*}, Nitin Jain², Usha Jain², Shraddha Kamankar¹

¹Department of Pharmaceutical Quality Assurance, Rashtrasant Janardhan Swami College of Pharmacy, Kokamthan, Tal- Kopargaon, Dist. Ahilyanagar, Maharashtra, 423601, India

²Department of Pharmaceutical Chemistry, Rashtrasant Janardhan Swami College of Pharmacy, Kokamthan, Tal- Kopargaon, Dist. Ahilyanagar, Maharashtra, 423601, India

ABSTRACT

Artificial Intelligence (AI) and Machine Learning (ML) are being introduced in combination with the nanoparticle-based drug delivery systems this is one of the major breakthroughs in nanomedicine. These computational approaches can provide data-driven optimization of critical parameters of the formulation such as particle size, morphology, and surface charge, as well as drug loading and stability, thereby decreasing the trial-and-error experimental process and expediting the formulation development process. Complex nonlinear mechanisms between formulation variables and important quality attributes are accurately model using advanced algorithms like deep learning and genetic optimization, thus enabling accurate control of the design of nanoparticles and drug release behaviour. Moreover, autonomous analysis algorithms are used to complement nanoparticle characterization by automated analysis of spectral and image data, which increases the accuracy and reproducibility. Although the issues of data availability and regulatory acceptance remain, the combination of AI/ML with the Quality by Design (QbD) concept and experimental design plans can become a rapid way to develop nanomedicine and help to translate personalized and effective drug delivery systems into clinical practice.

Keywords: Artificial Intelligence; Machine Learning (ML); Nanoparticle Drug Delivery Systems; Nanomedicine; Formulation Optimization.

INTRODUCTION

The successful development of nanotechnology in the past few years, especially the development of novel nanomaterials, has provided new ideas and promising methods for the diagnosis and treatment of many dangerous diseases. Due to their unique physicochemical and biological properties, nanomaterials are widely used in drug delivery systems (DDS)². When compared to traditional DDS, nano-DDS has great success in improving the pharmacokinetic and pharmacodynamic properties of encapsulated pharmaceuticals, and achieves targeted drug delivery and controlled drug release as a result of their distinctive size, shape and material characteristics. The successful targeting of therapeutic drugs, as well as their accumulation preponderantly at the desired site, is called targeted medication delivery. Nanoparticles can enhance efficacy and safety by improving the transmembrane

transport process, prolonging circulation time, and stability and solubility of encapsulated payload¹⁻³.

However, this strategy has limitations. Firstly, it is not possible to understand if some variables have more leverage than others, to perform when increasing the number of considered parameters and their different levels. Many interrelated formulation and process variables, such as polymer or lipid content, surfactant level, solvent ratio, stirring speed, temperature, and injection rate, are involved in the formulation of nanoparticles. Conventional trial-and-error methods are ineffective because small changes in these parameters can cause non-linear changes in important quality attributes such as particle size, polydispersity index, and drug encapsulation efficiency⁴.

Recent developments in ML and AI offer new approaches for nanomedicine development.

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

By analysing vast, high-dimensional datasets and deriving patterns, nanomedicine design and formulation can be accelerated. Andam-driven methods are gaining increasing importance in the prediction of preclinical outcomes such as biodistribution, toxicity and efficacy. Recent research has focused on the importance of machine learning models in improving the efficiency and effectiveness of these systems. Researchers can use AI approaches together with computational chemistry to optimize the design of nanoparticles, thereby ensuring that they can be effective in drug delivery but also safe for clinical use⁵⁻⁶.

Despite these developments, the reliability and utility of AI-generated results can be compromised by data scarcity, algorithmic bias, and regulatory issues⁷. In particular, data scarcity is still a fundamental obstacle as high-quality and well-annotated datasets are limited in size and consistency in nanomedicine. Predictive models have been used to design self-assembling nanomedicines based on drug pairs in order to overcome challenges related to toxicity and instability. AI has also been applied to the clinical translation of nanomedicines to overcome barriers such as manufacturing scalability, regulatory barriers and reproducibility. AI-driven frameworks aid of nanomedicines from laboratory to clinical establishment that serves as a new avenue in designing stable and efficient nanocarriers for targeted delivery⁸⁻¹⁰.

Fundamentals of AI and Machine Learning

AI/ML is especially suitable for nanomedicine due to its capacity to model nonlinear relationships, optimize multidimensional design spaces and integrate various types of experimental and simulation data, as well as data obtained from literature²⁰. These strengths are needed to address complex formulation challenges, such as relating physicochemical properties of Nanoparticles to in vivo performance or predicting protein corona formation under different biological conditions¹¹⁻¹².

Artificial Intelligence (AI)

The complex optimization of the nanomedicines entails the optimization of interdependent factors such as particle size, shape, surface groups, and drug cargo. Traditionally, researchers work with expensive and

time-consuming trial-and-error processes with limited scalability where one parameter is varied at a time and complexity of capturing complex structure-function relationships. Such approaches are often not successful in finding the best designs because of their inability to efficiently explore the enormous chemical design space or their inability to detect nonlinear dependencies between formulation variables and biological outcomes¹³.

Machine Learning (ML)

Unlike conventional programming which is where specific instructions are provided for each task, machine learning algorithms take a data-driven approach to look for patterns and improve their performance over time with no particular programming. This power to transform makes it possible to use ML in different areas such as healthcare and medical services, finance, robotics, and nanotechnology¹⁴.

Advances in ML have provided new opportunities for understanding and optimizing the stability of Nanoparticles and endosomal escape. By analysing a large dataset of physicochemical descriptors from size, zeta potential and shape, to chemical composition, using a ML model, it is possible to detect patterns that experiments may fail. Based on this, these predictive models can help in the rational design of nanoparticles with increased colloidal stability, low aggregation and enhanced endosomal escape¹⁵.

Deep Learning (DL)

These capabilities have been extended through deep learning methods which have the ability to model intricate and nonlinear relationships that describe the behaviour of nanomaterials in biological systems. Convolutional neural networks (CNNs) and recurrent neural network (RNN) architecture can be applied to high-dimensional data sets (TEM/imaging-derived nanoparticle morphology and/or time series pharmacokinetic data) for enhanced accuracy in the predictive mapping of design parameter (i.e. polymer composition, ligand density, surface charge, etc) to biodistribution, cellular uptake, and therapeutic efficacy¹⁶.

Types of Machine Learning

Supervised learning:

Supervised learning is a method of predicting particular outcomes using the labelled data from experimental results and has been used more often in the pharmaceutical formulation studies, instead of using multiple experiments of trial and error, researchers are now using supervised learning models for relating the formulation and process parameters to the final product characteristics. Wang et al. (2021) used supervised machine learning to predict the size of a nanoparticle with parameters such as polymer concentration (1-3%), surfactant concentration (0.5-2%), homogenization speed (10,000-20,000 rpm) etc. with the conventional method, particle size optimization needs 8-10 experiments that cost 2-3 weeks and results in particle sizes of 170-200 nm with an error range of 20-30 nm. while a random forest based supervised learning machine learning model predicts a particle size with reduced error¹⁷.

Unsupervised Learning:

Xu et al. (2020) used an unsupervised machine learning method to classify shapes of nanostructured lipid carrier (NLC) particles from Transmission Electron Microscopy (TEM) images. particle contours were extracted and parameterized using Hu moments, followed by hierarchical agglomerative clustering without prior labelling. more than 300 NLC particles were automatically classified in spherical (72%), elongated (18%) and irregular (10%) clusters with good cluster separation. the AI-based approach facilitated fast and reproducible particle characterization, enabling researchers to analyse large datasets efficiently¹⁸.

Reinforcement Learning:

Hybrid Deep Reinforcement Learning is now used to optimize nanoparticle formulation by using predictive models in conjunction with a reinforcement learning agent that adjusts formulation parameters incrementally (for example, polymer concentration $\pm$ 0.2%, surfactant concentration $\pm$ 0.1%), predicting consequences. For example, a study optimizing PLGA nanoparticles for drug delivery, a hybrid DRL was used to optimize formulation with the agent recommending a formulation of 2.2% polymer, 1%

surfactant, 8:2 lipid ratio which resulted in a particle size of $\sim$ 120 nm, PDI 0.18 and encapsulation efficiency of the studies of Nanoparticles using Machine Learning Algorithms¹⁹.

Machine Learning Algorithms in Nanoparticle Studies

Support Vector Machine (SVM).

Support Vector Machine is machine learning method for classification of nanoparticles as toxic or non-toxic. For example, CLIO Iron oxide nanoparticles 10-15nm, pseudo-caged iron oxide nanoparticles 20-25nm, monocrySTALLINE iron oxide nanoparticles 15-20nm, carbon dot quantum dots 5-10nm were tested in monocytes, hepatocytes, endothelial and smooth muscle cells. on SVM model CLIO and pseudo caged iron oxide nanoparticles showed toxicity for higher doses i.e. [50-100 μ g/mL]²⁰.

Random Forest (RF): The RF analysis of gold and iron oxide nanoparticles identified the surface chemistry and zeta potential as the most influencing parameters ($\sim$ 50%) for the formation of protein corona, followed by the particle size ($\sim$ 25-30%) Practically, nanoparticles with a positive charge (+15 to + 25mV) showed a higher adsorption of opsonin proteins, which resulted in a higher uptake by macrophages, and PEGylation decreased the protein adsorption of NPs by $\sim$ 35 - 45%, which improved the stability of circulation²¹.

Artificial Neural Networks : (ANN) were practically used to predict the particle size of Tio₂ doped chitosan nanoparticles using the formulation parameters such as concentration of Chitosan (0.5-2%), stirring speed (500-1500rpm) etc. ANN models such as (MLP Multilayer Perceptron) and RBF (RBF Radial Basis Function) were used to accurately predict particle size in the range of 90-180 nm, of which RBF network showed good result with R² value > 0.95 demonstrated high predictive accuracy²².

Convolutional Neural Networks (CNN): Convolutional Neural Networks (CNNs) were practically applied to analysed TEM images of metal nanoparticles to morphology and size classification. Our CNN model showed high classification accuracy ($\sim$ 95%) in the case of spherical (80-120nm), rod-shaped and the aggregated nanoparticles (>200nm)

while manual analysis showed higher variability. CNN based image analysis enhanced the size measurement error by <5% and analysis time from several hours to a few minutes in our experiments²³.

Principal Component Analysis (PCA): practically applied with nanoparticle datasets with physicochemical descriptors such as particle size, surface charge, hydrophobicity and surface chemistry. 3 principal components (PC1,PC2,PC3) were extracted with PCA, reducing more than 40 descriptors to 3 principal components, representing ~85% of the total variance. Principal component (PC1) in the case of experimental data was mostly explained by surface charge and hydrophobicity and it explained ~50-55% variance. principal component (PC1) and surface charge and hydrophobicity were strongly correlated with protein adsorption²⁴.

6) Genetic Algorithms (GA)

Genetic Algorithms(GA) have been widely used in optimization of crucial formulation and process variables such as lipid/polymer concentration, surfactants percentage, drug loading, homogenization speed and sonication time in nanoparticle systems. Santos et al. (2018) used a GA coupled with artificial neural network(ANN) to optimize nanostructured lipid carriers (NLCs) loaded with a poorly water-soluble drug solubility. The GA showed good performance in exploring the formulation space by varying lipid ratio and surfactants concentration.²⁶ The optimized NLCs showed mean particle size ~160-180nm²⁵.

Comparison: Conventional VS AI/ML-Based Nanoparticle Design

Conventional Nanoparticle Design

Conventional nanoparticle design mostly takes a trial-and-error approach, where formulation variables are systematically varied and each batch is experimentally tested to achieve the desired physicochemical properties. Haloperidol loaded PLGA nanoparticles by varying polymer concentration (50-200 mg, drug-to-polymer ratios (1:5-1:15), PVA concentration (0.5-2%w/v), stirring speed, and solvent ratio were optimized to achieve desired nanoparticle characteristics ²⁶.

AI/ML-Based Nanoparticle Design

Artificial neural networks (ANN) for optimizing polymeric nanoparticles. The model was trained based on a dataset of about 90 formulations, which included effects of varying polymer concentration, type of surfactant and drug loading. The ANN was able to predict the particle size accurately with high R² value > 0.9, demonstrating high predicting accuracy of model. Using virtual screening, the number of experimental batches has been reduced from 25-30 to less than 10, saving lots of time and raw materials. The predicted formulation achieved the target parameters, particles size ²⁷.

2. Particle Size and Size Distribution

Conventional Approach:

In conventional batch nanoprecipitation, PLGA nanoparticles were prepared with particle sizes from ~132 nm to 281 nm but the particle size and PDI showed wide variation depending on the concentration of polymer and PVA stabilizer. poor optimization conditions sometimes produced very large unstable particles (~1483-4793 nm, PDI ~0.9), reflecting the variability from batch to batch and the limitation of trial-and-error method ²⁸.

AI/ML Approach:

AI/ML-based methods take as input features synthesis variables, polymer type, polymer concentration, stabilizer type/concentration, anti-solvent/concentration, particle size and zeta potential. E.g. Gaussian process regression (GPR) achieved high accuracy (R² ~0.94, MSE ~87.5 nm²), predicted the best synthesis conditions and achieved the same particle size and uniformity before actual experimentation ²⁹.

3. Shape and Morphology

Conventional Approach:

SEM or TEM are the commonly used methods to assess the form and morphology of nanoparticles after production. The limitations of shape optimization result from the fact that trials need to be repeated and synthesis procedures need to be modified. Example ;PLGA Nanoparticle preparation via emulsification-solvent evaporation and nanoprecipitation techniques

was performed and characterized using SEM and DLS. SEM micrographs showed regular spherical shape with smooth surfaces for dried PLGA nanoparticles.¹⁴ Under certain conditions (10mg mL⁻¹ PLGA in organic phase, 5% PVA, organic solvent fraction 0.167, agitation 400 rpm, 75% sonication, and centrifugation at 20,000 rpm were used for nanoparticle preparation and characterization³⁰.

AI/ML Approach:

Convolutional Neural Networks (CNNs), in particular, are deep learning models that automatically classify nanoparticles shape and surface morphology from microscope pictures. Rational control over form is made possible by ML-guided synthesis, which improves cellular absorption, circulation duration and biodistribution. Examples were used to analysed 1480 nanoparticles from SEM images, which were automatically segmented in shapes and measured in size. The mean diameter was $\sim 213 \pm 14$ nm, circularity $\sim 0.97 \pm 0.03$, which was very similar to manual SEM measurement (~ 230 nm). This approach has the benefit of being fast and high throughput, with less labour compared to the ordinary way³¹.

4. Drug Loading and Encapsulation Efficiency

Conventional Approach:

Different drug to carrier ratios is used in experimentation in order to measure the drug loading and encapsulation efficiency. It often needs a high amount of screening to ensure a high encapsulation efficiency Example: PLGA nanoparticle research in prostate cancer treatment, capecitabine loaded PLGA NPs prepared by solvent evaporation exhibited high encapsulation efficiency ($88.4 \pm 0.17\%$) and loading ($16.98 \pm 0.7\%$) and average particle size of particle $< 144.5 \pm 2.5$ nm and zeta potential = 14.8 ± 0.5 mV³².

AI/ML Approach: Drug-carrier compatibility & Interaction energy using ML + molecular dynamics simulations. By finding formulations that have high drug loading and encapsulation efficiency before synthesis makes these predictions possible in order to improve the therapeutic efficacy. Example: AI/ML approaches rely on utilization of historical formulation data and predictive models (i.e. Random Forest) for predicting value of EE and DL before

synthesis. In one study with 300+ formulations, the model was also able to achieve R^2 0.96 for EE and R^2 0.93 for DL which allowed them to identify optimal drug carrier combinations and processing conditions without running many trials in the lab.³³

5. Drug Release Kinetics

Conventional Approach:

In vitro dissolution or diffusion studies are used to calculate release profiles and kinetic models are then used to fit curves. This procedure consumes much time and often requires reformulation. Example: ketoprofen entrapped PLGA nanoparticles were used to perform an in vitro test after 120 hours of release at $\sim 96.8\%$ (KN1), 56.5% (KN2) and 38.3% (KN3). KN1 followed first-order kinetics ($R^2 = 0.94$) and KN2 and KN3 followed the Korsmeyer-Peppas model ($R^2 = 0.98-0.99$), which implies diffusion-controlled release³⁴.

AI/ML Approach:

Supervised machine learning models are applied to drug release behaviour - with machine learning models used to predict release behaviour of drugs in various pH, temperature and enzymatic conditions. AI is allowing for manipulation of regulated or sustained release profiles which are appropriate for specific therapeutic applications. Example: Gaussian process regression GPR was used to predict the release of Doxorubicin under varying pH and particle sizes. Experimental data showed that there is a higher release is in acidic or alkaline state and also the model was able to simulate the trends with $R^2 \sim 0.61$ and MSE 0.00007 which reduce the need to perform repeated dissolution experiments³⁵.

4. Role of AI/ML in Nanoparticle Design

4.1 Predictive Modelling

4.1.1 Prediction of Particle Size from Formulation Variables

Prediction of drug loading (DL) and encapsulation efficiency (EE%) of nanoparticles prepared with PLGA using microfluidic platforms can be efficiently realized by machine learning (ML) models, especially ensemble models such as Random Forest (RF).

Dataset and Features: A complete dataset of over 300 PLGA nanoparticle formulations were identified in the literature and comprising 25 features in terms of material properties, solution properties and microfluidic operation parameters. The most significant features in the prediction of DL and EE were defined using RF and LASSO feature selection methods.

Feature Selection: RF and LASSO feature selection techniques were used to find the most important features for DL and EE prediction. For DL, some features of importance were drug type and concentration, nanoparticle size, aqueous and organic flow rates and solvent ratios. For EE the drug type and concentration, the nanoparticle size and the type of chip were critical.

Modelling Approach: Three regression models were tested i.e. Random Forest (RF), Gradient Boosting (GB) and Support Vector Regression (SVR). RF performed better as compared to others with R^2 values of around 0.96 for DL and 0.93 for EE on test sets, and thus showed high predictive accuracy.

Interdependence Analysis: DL and EE were found to have little impact on each other's prediction, which means that they can be optimized separately.

Local Explanation: Model interpretability using LIME indicated that factors, such as flow ratio, had a negative effect on both DL and EE and the drug and PLGA concentration positively affected EE.

Practical Implications: ML models have the ability to predict DL and EE efficiently with the ability to reduce the need for heavy experimental trials.

Limitations: The predictive power depends on the quality and size of the dataset (about 300 data points), and future work could enhance models by incorporating more data and advanced ML architectures³⁶.

Example: In studies on PLGA and hybrid nanoparticles, Random Forest, Gradient Boosting, SVR and XG boosting were applied to predict drug loading (DL) and encapsulation efficiency(EE%). Experimental DL ranged 42-78% and EE 50-88% in different formulations. AI/ML models were able to

predict DL=65-72% ($R^2=0.087-0.91$, RMSE=3.5-4.5%) and EE=75-79% ($R^2=0.80-0.85$)³⁷.

4.1.2 Stability Prediction of Nanoparticles

Stability is one of the biggest issues in the field of nanoparticle development, as physical and chemical instability can cause aggregation, drug leakage or degradation during storage. Machine learning is a promising method for stability prediction from historical data of formulation, physicochemical properties and storage conditions. Although direct ML-based stability prediction studies in nanoparticles are still emerging, the use of deep learning and predictive analytics in biological and pharmaceutical systems to predict long-term behaviour and degradation trends has been successfully implemented. These approaches can be extended to nanoparticle systems in predicting shelf-life and storage stability.

Example: liposomal nanoparticles, where a Support vector Machine (SVM) classified formulations as stable or unstable, Stable formulations had PDI < 0.220 and zeta potential ≥ 25 mV, while unstable ones had PDI > 0.300 and zeta potential < +15mV. SVM has a high accuracy of about 92% (cross-validated: $88 \pm 5\%$) with main factors known to be the particle size, concentration of the surfactant and the composition of the polymer³⁸.

4.1.3 Quality by Design (QBD) Integrated with Machine Learning

Quality by Design (QBD) is a methodical approach in the pharmaceutical development. Quality Target Product Profile (QTPP) and identifying Critical Quality Attributes (CQAs), Critical Material Attributes (CMAs) and Critical Process Parameters (CPPs). However, these approaches may not be able to model complex and nonlinear relationships in systems like nanoparticles. To overcome these limitations, machine learning methods are incorporated into the QBD framework. ML models, such as artificial neural networks, analysed large datasets on experimental and process analytical technology and provide more precise prediction of CQAs. ML-integrated QBD approaches enhance the design space identification, reduce experimental trials, and save time and cost by improving formulation robustness through efficient

prediction of parameters such as particle size, encapsulation efficiency, and stability .

Example: In an ML- integrated QBD study, resveratrol-loaded polymeric nanoparticles were optimized based on CMAs: polyacrylic acid(0.1% w/v), gelatine (0.5%w/v), and Poloxamer 407(1.11%w/v)and CPPs: sonication frequency (21.43 Hz) and time (5.02 min). The ANN model was applied to the prediction of CQAs with high degree of accuracy,

obtaining a particle size of 72.85 ± 10.21 nm, PDI of 0.30 ± 0.09 , zeta potential of -18.11 ± 8.56 mV and encapsulation efficiency of 79.25 ± 3.22 %. ML-QBD reduced experimental trials, improved efficiency, and produced robust and stable nanoparticles³⁹.

4.2 Feature Selection

4.2.1 Identification of Critical Formulation Parameters Using AI

Feature selection plays an important role in the identification of the most influential formulation and process parameters controlling critical quality attributes (CQAs) such as particle size, polydispersity index, encapsulation efficiency, drug loading, and stability of nanoparticles. Though the classical statistical methods such as factorial design and response surface methodology are useful in preliminary screening, they have been proven to be inadequate when dealing with nonlinear, and high-dimensional formulation datasets. In contrast to this, AI-based models like artificial neural networks, Random Forest and support vector machines allow data-driven identification of key formulation variables through learning of nonlinear relationships between input and output. These models give feature importance rankings and sensitivity analysis which reduces the experimental burden support better model interpretation and for Quality by Design (QBD)-oriented nanoparticle formulation development.

Example: In a study on the chitosan coated polymeric nanoparticles, Random Forest was used to identify the critical formulation parameters. polymer to drug ratio (40%), chitosan concentration's (30%), and stirring speed (20%) were found to be the most influential. under optimized condition, particle size of 132 ± 4

nm, EE $85 \pm 2\%$, 12h drug release $78 \pm 3\%$ were obtained experimentally⁴⁰.

4.2.2 Dimensionality Reduction Techniques (PCA, LDA) and AI/ML in Nanoparticle Research

Nanoparticle research often yields large, complicated data sets from experiments such as spectroscopy, microscopy or simulation. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis(LDA), reduce the data to a small number of key components that retain the most important information. PCA is used to find patterns, reducing the data sets to some key components, making it easier to visualize them, and more focus on separating different classes; such as nanoparticle types, size groups or surface modification. PCA and LDA with neural networks for analysis of UV-VIS spectra of gold nanoparticles and effective prediction of particle sizes other studies have applied PCA the structural features of silver and copper nanoparticles for efficient classification by using clustering methods by decreasing the data complexity, these methods can make AI/ML models faster, more reliable, and interpretable for effective characterisation and design of nanoparticle

Example: Spectroscopic nanoparticle data were subjected to LDA and PCA. Using the FTIR and UV-Vis's spectra, it was possible to correctly classify the nanoparticle surface functionalization and composition with a classification accuracy of 90-92% with LDA. In parallel, PCA reduced the spectral variables, more than 200, to 8-10 principal components, explaining 95% of the total variance. Following PCA, the accuracy of clustering was enhanced to 93% as compared to 82% and the size prediction RMSE was minimized to 7 nm compared to 12 nm, and the computational cost was also lower than 30-35%⁴¹.

AI/ML-assisted process optimization in nanoparticle design

Selecting optimal Nanoparticle Preparation Method

Choosing the right nanoparticle preparation method - such as NLCs, SLNs, liposomes, nanoemulsions,or polymeric nanoparticles is often a difficult task due to the many formulation and process factors that affect

the final product. Machine learning helps simplify this problem by learning from existing experimental and literature data. ML models can predict more suitable preparation methods based on desired properties like particle size, stability, drug release and encapsulation efficiency. Instead of having to rely on trial-and-error experiments, researchers can use ML to test many formulation options virtually.

The speed of homogenization affects the size reduction rate of the particles in the beginning; high homogenization speed tends to produce smaller particles. The ML and AI models can be used to determine the most suitable range of speed that produces uniform particle size without affecting the stability of the formulation. Sonication time is used to control the cavitation energy of ultrasonication. A short sonication process results in broad size distributions but excessive sonication can destroy nanoparticles or change solvent composition. It can also be useful in the real-time monitoring of processes.

Example: Machine learning models were used to perform optimization of polymeric nanoparticle formulation using variables such as homogenization speed and sonication time. Under optimized conditions (15,000 rpm, 10 min), particle size of 142 ± 4 nm, PDI 0.21 and encapsulation efficiency of 82% were obtained experimentally. ML model was able to predict 138 nm in particle size ($R^2=0.91$, RMSE = 8 nm) and encapsulation efficiency ($R^2=0.88$, RMSE=7%), at approximately 35-40% (which showed efficient and reliable process optimization ⁴²).

6. AI/ML for Characterization

Artificial intelligence (AI) and machine learning (ML) are applied increasingly to nanoparticle and material characterization which helps to improve the accuracy and efficiency. Convolutional Neural Networks (CNNs) are used for image analysis to extract the morphological features and reduce the high dimensional microscopy data to classify the particle shapes and structures.

Image analysis for morphology using CNN

By now, Convolutional Neural Networks (CNNs) are popular to investigate the morphology of nanoparticles using TEM and SEM images. Images

and automatically detect features such as shape, size, edges, and surface texture. After some basic image pre-processing, CNN models can be used to separate individual nanoparticles, determine size distribution and classify the shape of individual nanoparticles (e.g. spherical or rod-shaped). This approach is much faster and more consistent than manual image analysis and reduces human bias, especially when nanoparticles are aggregated or polydisperse. CNN-based image analysis has found successful applications in the area of metallic nanoparticles, polymeric nanoparticles and lipid based nanocarriers for pharmaceutical research.

Example: CNN analysis of TEM image for the morphology of nanoparticles. Detection accuracy of the model was 94-96%, with a precision of 0.93, recall of 0.91 and F1 score of 0.92. Particle size prediction was found to be strongly correlated with manual determination ($R^2=0.89$, Root Mean Square Error (RMSE) = 3-4 nm, error < 5%). CNN was also used to accurately segment overlapping particles (>90% accuracy), which provided fast and reproducible analysis of particle size ⁴³.

AI/ML for Characterization and Analysis of Nanoparticles using XRD and FTIR

Artificial Intelligence (AI) and Machine Learning (ML) are increasingly being used to simplify and improve the characterization of nanoparticles using X-ray Diffraction (XRD) and Fourier Transform Infrared Spectroscopy (FTIR). In nanoparticle research, crystallinity, crystallite size, peaks shifts, intensity and peaks broadening due to small size, lattice strain, and phase variation. Machine algorithms such as PCA, Random forest and Neural networks can be used to automatically analyse XRD patterns and identify phase and the correlation between diffraction characteristics and nanoparticle properties.

Example: AI/ML models analyzed XRD and FTIR data for iron nanoparticles. For Iron nanoparticles the experimental XRD peak at $2\theta = 35.6^\circ$ with Full Width at Half Maximum = 0.42° resulted in a crystallite size of 21 nm while ML model resulted in 20.5 nm with strong agreement ($R^2=0.87$, phase accuracy 92-95%). In FTIR data the characteristic peaks at 580 cm^{-1} (Fe-O) and 3420 cm^{-1} (O-H) were identified and ML model ⁴⁴.

AI/ML for Characterization and Analysis of Nanoparticles using UV, DLS, and TEM)

AI and (ML) can play an important role in predictive analytics of nanoparticle properties using data obtained from UV-Visible spectroscopy, Dynamic Light Scattering (DLS), and Transmission Electron Microscopy (TEM). UV-Vis spectra are sensitive to particle size, concentration and aggregation behaviour especially due to surface plasmon resonance in metallic nanoparticles. ML models can learn relationships between UV absorbance patterns and particle size or stability's provides hydrodynamic diameter, polydispersity index (PDI) and zeta potential, which reflect nanoparticle size distribution and dispersion stability. ML algorithms use these parameters to predict formulation stability and optimize processing conditions. By using a combination of TEM information and AI/ML models, the accuracy is higher, manual bias.

Example: A gradient -boosted decision (XGBoost) model was applied to predict the TEM derived nanoparticle size and shape based on UV-Vis and DLS results. model predicted important parameters such as average minimum feret diameter, standard deviation, aspect ratio, projected area and perimeter with $R^2 > 0.8$, which is very accurate. for example, the predicted feret diameter of gold nanoparticles at 190 min was 98 ± 5 nm comparable to the experimental value of 102 ± 5 nm, demonstrating high prediction accuracy ⁴⁵.

AI/ML in Drug Release Kinetics

AI/ML models are used to predict drug release behaviour in nanoparticle drug delivery systems by learning from experimental datasets. These models analyze formulation parameters such as particle size, polymer composition, drug loading, and surface characteristics to predict complete release profiles and identify dominant release mechanisms..

Example: In a dataset of 377 formulations, experimental drug release values ranged from 15% to 97% between 30–480 min. A Random Forest model predicted release values ranging from 18% to 99%, showing close agreement with experimental data and improved prediction of release behaviour ⁴⁶.

Computational Tools and Software

10.1 Numerical Modelling and Data Analysis Platforms

Computational platforms such as MATLAB and Python

MATLAB helps for the simulation of diffusion controlled drug release and polymer degradation processes in biodegradable systems. Python is a programming language that supports for statistical analysis, management of large experimental datasets, and predictive modelling, all of which are useful tools in assisting researchers in developing, evaluating and optimising nanoparticle formulations in a more efficient, data-driven way.

Example: As shown in PLGA nanoparticle systems by J.Siepmann and F.Siepmann, formulations with mean particle size 150 ± 12 nm and PDI 0.22 with cumulative release of 65% of the drug in 48 hours including an initial burst release of 18% for 6 hours of the drug release were followed by mathematical modelling of the release using a diffusion-erosion mathematical model resulting in a high correlation coefficient ($R^2=97\%$), and calculated rate constant of release (k) of 0.034, indicating diffusion dominated release.⁴⁷

Machine Learning Frameworks for Property Prediction

TensorFlow and scikit-learn are popular machine learning frameworks for predicting nanoparticle properties and optimizing formulation design. These tools support supervised and unsupervised learning techniques for capturing complex non-linear relationships between nanoparticle characteristics including size, surface charge, composition and morphology and performance outcomes including drug loading, release behaviour, stability and biological interactions.

Example: Curcumin nanocomposites were optimized using a hybrid ML model with scikit-learn regressors and a Tensor Flow network. Using data from 74 formulations, the model predicted loading efficiency) and encapsulation efficiency (EE%) with $R^2 = 0.89$, RMSE=6.24 and $R^2 = 0.87$, RMSE=7.15. It also identified an optimal design space (particle size 80-

200nm, zeta potential -30 to -50mv),reducing 20-30 empirical batches and saving 40-60% of time and cost⁴⁸.

Nano informatics Databases

Nano informatics databases are centralised repositories containing nanoscale descriptors and experimental metadata and for biological response data. These platforms help in data standardisation and integration, which is beneficial for the development of QSPR/QSAR models, reproducibility, and enabling data-driven risk assessment and regulatory decision-making in nanomedicine research.

Example: In a practical application of nano informatics, human lung epithelial cells were exposed to silver nanoparticles of 40 nm with a citrate coating concentration of 20 µg/mL for 24 hours, and cell viability was measured as 75%, whereas the QSAR model predicted a viability of 74%, showing close agreement with the experimental result.

Similarly, silver nanoparticles of 50 nm with a citrate coating concentration of 25 µg/mL were tested under the same conditions, where experimental cell viability was 68%, while the QSAR model predicted 70%, again demonstrating good predictive accuracy⁴⁹.

Multiphysics Simulation Tools

In nanoparticle research,COSMOL allows for the simulation of mass transport, diffusion, heat transfer, electromagnetic fields, and fluid flow which are crucial to understanding the behaviour of nanoparticles in complex systems. It has typically been used to understand drug diffusion from nanoparticles, transport in biological tissues, heat generation in photothermal therapies and nanoparticle motion within microfluidic environments.

Example: COMSOL Multiphysics in nanoparticles research magnetic PLGA Microspheres filled with cyclophosphamide manual in-vitro experiments in PBS (pH 7.4) for 72 hours shows an initial burst release of approximately 28% after 6 hours and ~82 percent cumulative release after 72 hours COMSOL was used to build a virtual model of the nanoparticle incorporating particle size, polymers density and drug diffusion, solving Fick's diffusion equation through the finite and sustained release and allowed virtual

optimization of formulation parameters, reducing the need for extensive experimental trials⁵⁰.

Future Perspectives:

Artificial intelligence (AI) and machine learning (ML) are expected to play an increasingly important role in nanoparticle-based drug delivery by enabling faster formulation optimization and predictive design. Future developments may include integration of AI with Quality by Design (QbD), real-time process monitoring, and personalized nanomedicine approaches. Improvements in data availability, model validation, and regulatory acceptance will further enhance the application of AI-driven strategies in pharmaceutical nanotechnology.

CONCLUSION

The integration of artificial intelligence (AI) and machine learning (ML) and nanoparticle-based drug delivery systems is revolutionizing nanomedicine with the data-driven optimization of critical formulation parameters such as particle size, shape, surface charge, drug loading and stability. These approaches help to shorten the time, cost, and variability of experiments as well as enhance reproducibility and therapeutic performance. Advanced ML algorithms have made it possible to model more complex relationships and to have fine control over the structure of nanoparticles and over how the drug is released, and on the other hand, AI-assisted characterization has led to greater accuracy in the image and spectral analysis. Although challenges (e.g. limited datasets and acceptance by regulators) remain, integration of AI/ML with Quality by Design frameworks is expected to lead to the development and faster clinical translation of nanomedicines to support more efficient and personalized drug delivery strategies.

REFERENCES

1. Han Y, Kim DH, Pack SP. Nanomaterials in Drug Delivery: Leveraging Artificial Intelligence and Big Data for Predictive Design. *International Journal of Molecular Sciences*. 2025 Oct;26:11121–11140.
2. Kantesaria R, Panda HS. A Review on AI-Based Data-Driven Models for Optimization of Nanocarriers as Drug Delivery Systems. *ACS*

- Biomaterials Science & Engineering. 2026 Feb;12:3450–3465.
3. Luo G, Jiang X, Hu C, Li L, Yan L, Xiao G, Duo Y, Zhang X. Artificial intelligence-powered nanomedicine. *Chemical Society Reviews*. 2026 Mar;55:4100–4125.
4. Bets A, Smith J, Lee K. Machine Learning and Artificial Intelligence in Nanomedicine. *Journal of Nanomedicine Research*. 2024 Feb;18:210–223.
5. Ahmadi M, Ayyoubzadeh SM, Afshar Ardekani S, Ghaedrahmat Z, Masoumi N, Farnia MM, Ghorbani Bidkorpeh F. A review on protein corona formation on nanoparticles and prediction of its composition using artificial intelligence tools. *International Journal of Pharmaceutics*. 2025 Feb;683:126094.
6. Kopac T. Leveraging Artificial Intelligence and Machine Learning for Characterizing Protein Corona, Nanobiological Interactions, and Advancing Drug Discovery. *Bioengineering (Basel)*. 2025 Mar;12(3):312.
7. Dorsey PJ, Lau CL, Chang TC, Doerschuk SM, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. *Journal of Pharmaceutical Sciences*. 2024 Dec;113(12):3413–3433.
8. Bhujel R, Enkmann V, Burgstaller H, Maharjan R. Artificial Intelligence Driven Strategies for Targeted Delivery and Enhanced Stability of RNA Based Lipid Nanoparticle Cancer Vaccines. *Pharmaceutics*. 2025 Aug;17(8):992.
9. Zhan B, Yuan P, Yu F, Peng T, Zhou Q, Hu X. Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America*. 2020 May;117(19):10492–10499.
10. Öztürk A, Gündüz AB, Ozisik O. Use of supervised machine learning algorithms for evaluation and prediction of nanoparticle size and shape parameters. *Combinatorial Chemistry & High Throughput Screening*. 2018 Sep;21(9):693–699.
11. Roncaglia C, Ferrando R. Machine learning assisted clustering of nanoparticle structures. *Journal of Chemical Information and Modeling*. 2023 Feb;63(2):459–473.
12. Mim JJ, Al Mamun A, Nayem MH, et al. Machine learning driven advances in nanotechnology: From materials design to process optimization. *Materials Today Communications*. 2026 Mar;50:114485.
13. Hosni Z, Achour S, Saadi F, Chen Y, Al Qaraghuli M. Machine learning driven nanoparticle toxicity analysis and prediction. *Ecotoxicology and Environmental Safety*. 2025 Oct;299:118340.
14. Vijgen N, Poulsen KM, Sosa Macias G, Payne CK. Predicting the protein corona on nanoparticles using random forest models with nanoparticle, protein, and experimental features. *Nanoscale Advances*. 2025 Jun;7:5612–5624.
15. Koletti P, Caliceti P, Veronese FM. Artificial neural network based particle size prediction of polymeric nanoparticles. *European Journal of Pharmaceutics and Biopharmaceutics*. 2017 Jan;119:333–342.
16. Zelenka C, Kamp M, Strohm K, et al. Deep learning classification of nanoparticles from TEM images. *Nanotechnology*. 2022 Mar;33(12):125704.
17. Yu F, Peng T, Zhan B, et al. Meta Analysis and Machine Learning Prediction of Protein Corona Composition across Nanoparticle Systems in Biological Media. *ACS Nano*. 2025 Feb;19:37633–37650.
18. Reza Zaki HA, Hamidi MA, Hassan MH. Application of genetic algorithm-assisted artificial neural networks for nanosphere formulation optimization. *Journal of Pharmaceutical Innovation*. 2021 Aug;16:645–656.
19. Khan I, Saeed K, Khan I. Nanoparticles: Properties, applications and toxicities. *Arabian Journal of Chemistry*. 2019 Jul;12(7):908–931.
20. Youshia J, Ali ME, Lamprecht A. Artificial neural network based particle size prediction of polymeric nanoparticles. *European Journal of Pharmaceutics and Biopharmaceutics*. 2017 Jan;119:333–342.
21. Mora Hidalgo D, Gómez García DD, Hernández López SA, et al. PLGA nanoparticles produced by emulsification and nanoprecipitation: Influence of formulation parameters on particle size and dispersity. *RSC Advances*. 2020 Feb;10:12345–12356.

22. Alqarni S, Huwaimel B. Predicting PLGA nanoparticle size and zeta potential via machine learning analysis. *Scientific Reports*. 2025 Feb;15:20765.
23. Koyama A, Miyauchi S, Morooka K, Hojo H, Einaga H, Murakami Y. Analysis of TEM images of metallic nanoparticles using convolutional neural networks and transfer learning. *Journal of Magnetism and Magnetic Materials*. 2021 Nov;538:168225.
24. Wen H, Luna-Romera JM, Riquelme JC, Dwyer C, Chang SLY. Statistically representative metrology of nanoparticles via unsupervised machine learning of TEM images. *Nanomaterials*. 2021 Oct 1;11(10):2706.
25. Sun SB, Liu P, Shao FM, Miao QL. Formulation and evaluation of PLGA nanoparticles loaded with capecitabine for prostate cancer treatment. *International Journal of Clinical and Experimental Medicine*. 2015 Oct 15;8(10):19670–19681.
26. Hanari N, Mihandoost S, Rezvantlab S, et al. Intelligence prediction of microfluidically prepared nanoparticles with machine learning for encapsulation efficiency and drug loading. *Scientific Reports*. 2025 Oct 27;15(1):37512.
27. Rezvantlab S, Mihandoost S, Rezaiee M. Machine learning assisted exploration of the influential parameters on the PLGA nanoparticles. *Sci Rep*. 2024 Feb 1;14:1114.
28. Hanari N, Mihandoost S, Rezvantlab S. Intelligence prediction of microfluidically prepared nanoparticles. *Sci Rep*. 2025 Jan 15;15(1):21471.
29. Shen C, Zhang M, Lu M, Chang E, Gao Z, Ban W, Liu Q, Zuo Z, Jiang C. Machine learning empowered formulation design, optimization and characterization of nanoparticulate drug delivery systems: Current applications, challenges, and future perspectives. *Acta Pharmaceutica Sinica B*. 2025 Dec 10;16(2):665–685.
30. Suriyaamporn P, Kansom T, Pamornpathomkul B, Ngawhirunpat T, Opanasopit P, Ramjan S. Predictive modeling approach using machine learning–integrated design of experiments in quality by design for optimizing resveratrol-loaded polymeric nanoparticle formulation. *Int J Pharm*. 2025 Oct 15;683:126080.
31. Li H, Zhao Y, Xu C. Machine learning techniques for lipid nanoparticle formulation. *Nano Convergence*. 2025 Jul 15;12(1):35.
32. Lasalvia M, Capozzi V, Perna G. A comparison of PCA LDA and PLS DA techniques for classification of vibrational spectra. *Applied Sciences*. 2022 Nov 1;12(11):5345.
33. Glaubitz C, Rothen Rutishauser B, Lattuada M, Balog S, Petri Fink A. Designing the ultrasonic treatment of nanoparticle dispersions via machine learning. *Nanoscale*. 2022 Aug 15;14(35):12940–12950.
34. Sun Z, Shi J, Wang J, Jiang M, Wang Z, Bai X, Wang X. A deep learning based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale*. 2022 Jun 20;14(30):10761–10772.
35. Glaubitz C, Bazzoni R, Weber T, et al. Leveraging machine learning for size and shape analysis of nanoparticles: predictive models using DLS and UV–vis as inputs to estimate TEM outputs. *J Phys Chem C*. 2023 Mar 15;127(3):145–157. doi:10.1021/acs.jpcc.3c05938.
36. Rastegari A, Faghihi H, Mobinikhaledi M. Prediction of drug release profile from chitosan nanoparticles: integration of experimental data and machine learning models. *Drug Dev Ind Pharm*. 2025 Dec;51(12):1819–1827. doi:10.1080/03639045.2025.2569573.
37. Casalini T, Rossi F, Lazzari S, Perale G, Masi M. Mathematical modeling of PLGA microparticles: from polymer degradation to drug release. *Mol Pharm*. 2014 Nov 3;11(11):4036–4048. <https://doi.org/10.1021/mp500078u>.
38. Rahdar A, Fathi Karkan S, Shirzad M. Predictive optimization of curcumin nanocomposites using hybrid machine learning and physics informed modeling. *Sci Rep*. 2025 Dec 23;15:44368. doi:10.1038/s41598-025-28074-7.
39. Li J, Wang C, Yue L, et al. Nano QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: a review. *Environ Int*. 2022;165:107345. doi:10.1016/j.envint.2022.107345. (Published 2022) Available from: <https://doi.org/10.1016/j.envint.2022.107345>
40. Runkana V, Pareek A, Arora P. Modeling and simulation of drug release through polymer matrices. *COMSOL Conference Proceedings*.

- 2014
<https://www.comsol.com/paper/modeling-and-simulation-of-drug-release-through-polymer-matrices-1935>.
41. Kantesaria R, Panda HS. A review on AI based data driven models for optimization of nanocarriers as drug delivery systems. *ACS Biomaterials Science & Engineering*. 2026 Feb 10;12:3450 3465.
<https://pubs.acs.org/doi/10.1021/acsbiomaterials.5c01998>.
42. Sheikh M, Jirvankar P. Harnessing artificial intelligence for enhanced nanoparticle design in precision oncology. *AIMS Bioengineering*. 2024;11(4):574 597.
<https://doi.org/10.3934/bioeng.2024026>.
43. Wang Q, Liu Y, Li C, Xu B, Xu S, Liu B. Machine learning enhanced nanoparticle design for precision cancer drug delivery. *Advanced Science*. 2025 Aug;12(30):e03138.DOI: 10.1002/advs.202503138).
44. Li H, Zhao Y, Xu C. Machine learning techniques for lipid nanoparticle formulation. *Nano Convergence*. 2025;12:35. doi:10.1186/s40580-025-00502-4. <https://doi.org/10.1186/s40580-025-00502-4>
45. Shen C, Zhang M, Lu M, Chang E, Gao Z, Ban W, Liu Q, Zuo Z, Jiang C. Machine learning empowered formulation design, optimization, and characterization of nanoparticulate drug delivery systems: current applications, challenges, and future perspectives. *Acta Pharmaceutica Sinica B*. 2025 Dec;16(2):665 685.
doi:10.1016/j.apsb.2025.12.011.
46. Akhtar M, Nehal N, Gull A, Parveen R, Khan S, Khan S, Ali J. Explicating the transformative role of artificial intelligence in designing targeted nanomedicine. *Expert Opin Drug Deliv*. 2025;22(7):971 991.
doi:10.1080/17425247.2025.2502022.
47. Noury H, Rahdar A, Romanholo Ferreira LF, Jamalpoor Z. AI driven innovations in smart multifunctional nanocarriers for drug and gene delivery: a mini review. *Crit Rev Oncol Hematol*. 2025;210:104701.
doi:10.1016/j.critrevonc.2025.104701.
48. Alqarni S, Huwaimel B. Predicting PLGA nanoparticle size and zeta potential in synthesis for application of drug delivery via machine learning analysis. *Sci Rep*. 2025;15(1):20765.
doi:10.1038/s41598-025-06872-3
49. Rodrigo Fonseca Silveira, Ana Luiza Lima, Idejan Padilha Gross, et al. The role of artificial intelligence and data science in nanoparticles development: a review. *Nanomedicine*. 2024;19(14):1271 1283.
50. Qingquan Wang, Yujian Liu, Chenchen Li, et al. Machine Learning Enhanced Nanoparticle Design for Precision Cancer Drug Delivery. *Advanced Science*. 2025;12(30):e03138.

HOW TO CITE: Sonali Aher^{1*}, Nitin Jain², Usha Jain², Shraddha Kamankar¹, Role Of Artificial Intelligence And Machine Learning In Nanoparticle Design And Optimization: A Review, *Int. J. Sci. R. Tech.*, 2026, 3 (7), 421-433. <https://doi.org/10.5281/zenodo.21377918>