



Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

Artificial intelligence-guided design of lipid nanoparticles for mRNA delivery

Kexin Su^{a,†}, Junjie Qiu^{a,†}, Tengfei Xu^{a,b,*}, Shuai Liu^{a,c,d,e,*}

^aCollege of Pharmaceutical Sciences, State Key Laboratory of Advanced Drug Delivery and Release Systems, Zhejiang University, Hangzhou 310058, China

^bResearch Center for Clinical Pharmacy, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

^cLiangzhu Laboratory, Zhejiang University, Hangzhou 311121, China

^dEye Center of the Second Affiliated Hospital of Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310009, China

^eZhejiang Key Laboratory of Pain Perception and Neuromodulation, Hangzhou 310009, China

Received 7 April 2025; received in revised form 6 June 2025; accepted 25 June 2025

KEY WORDS

mRNA delivery;
Lipid nanoparticles;
Artificial intelligence;
Machine learning;
Deep learning;
Nanomedicine;
Precise delivery;
mRNA therapeutics

Abstract Lipid nanoparticles (LNPs) hold significant potential for mRNA-based therapeutics, as evidenced by their successful use in SARS-CoV-2 mRNA vaccines. LNPs effectively protect and transport mRNA to target sites, thereby ensuring its stability and efficient transfection. Despite the progress, some challenges remain in the development of mRNA-LNP delivery systems, such as limited targeting specificity, the complexity of formulations, and the time-consuming and high-throughput screening process. Artificial intelligence (AI) has emerged as a powerful tool to address these challenges, accelerating the design and optimization process of LNPs. AI-guided approaches can improve the efficiency of lipid structure and formulation screening by rapidly identifying key design parameters and employing predictive modeling to optimize LNP properties. The combination of AI and LNP technology offers significant advantages, including enabling the design of more personalized and precise delivery systems, streamlining the development process, and reducing the cost. This review discusses recent advancements in AI-guided mRNA-LNP delivery systems and highlights their potential to revolutionize mRNA therapeutics.

This article is part of special column entitled: Advancing Pharmaceutics and Pharmacology published in Acta Pharmaceutica Sinica B.

*Corresponding authors.

E-mail addresses: txfu@zju.edu.cn (Tengfei Xu), shuailiu@zju.edu.cn (Shuai Liu).

[†]These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

<https://doi.org/10.1016/j.apsb.2025.11.029>

2211-3835 © 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Messenger RNA (mRNA)-based therapeutics have demonstrated substantial potential across various fields, including protein replacement therapy, gene editing, and diverse vaccines^{1–3}. The unprecedented success of mRNA vaccines in combating the COVID-19 pandemic highlights the unique advantages of mRNA technology, such as rapid development, flexible design, and the capacity to elicit robust immune responses^{4–7}. These attributes have spurred a surge of interest in harnessing mRNA capabilities beyond vaccines, with promising applications to correct genetic defects, treat metabolic disorders, and develop personalized cancer vaccines^{1,8}. As research and development continue to advance, mRNA therapeutics are anticipated to become the cornerstone of next-generation precision medicine, offering promising solutions for a wider range of diseases^{2,9,10}. Despite these promising prospects, the clinical translation of mRNA therapeutics faces significant challenges, particularly in efficient delivery, stability, and immunogenicity. Recently, artificial intelligence (AI) has emerged as a transformative force in healthcare, revolutionizing fields such as drug discovery and nanomedicine^{11–14}. Especially, AI's capacity to analyze large data sets, predict molecular interactions, and optimize complex systems has opened new avenues for accelerating mRNA therapeutics^{15–18}.

Delivery issue represents the critical obstacle in mRNA therapeutic development^{19–21}. Since utilized in FDA-approved two COVID-19 mRNA vaccines and one respiratory syncytial virus (RSV) mRNA vaccine, lipid nanoparticles (LNPs) have become the first choice for mRNA delivery^{22–26}. Typically, LNPs consist of four key components: ionizable lipids, cholesterol, neutral phospholipids, and polyethylene glycol (PEG)-lipids^{27–29}, synergistically aiding in mRNA encapsulation, cellular uptake, endosomal escape, and mRNA release into the cytoplasm^{30–33}. Despite the remarkable advancements in LNP technology, several challenges still remain, including rational design of ionizable lipids to enhance the delivery efficiency, optimizing LNP formulations to achieve tissue/cell-specificity, and minimizing off-target effects and immunogenicity. These challenges highlight the necessity for innovative tools to explore the vast compositional space of LNPs. With the rapid development of computer science, AI stands at the crossroads of multidisciplinary intersections specializing in all kinds of scientific issues related to data-intensity^{34–37}. AI has emerged as an aspirational tool in revolutionizing LNP development strategies and advancing mRNA therapeutics^{4,38–40}. Relying on large amounts of high-quality training datasets, AI can extract insights from molecules and predict a wider range of lipid molecules based on their chemical structures and physicochemical properties⁴¹. By harnessing powerful data processing capabilities, AI offers the possibility to systematically analyze the sophisticated relationships among LNP compositions, chemical structural features, and functional effects. This facilitates high-throughput screening and lays the foundation for more efficient and personalized LNP design.

The integration of AI into drug delivery research has already demonstrated its transformative impact, particularly through machine learning (ML) approaches^{12,34,42,43}. These methods are

beneficial to discovering concealed associations in large datasets, predicting formulation performances, and guiding the rational design of delivery systems. Applying AI to LNP design, researchers are able to explore expansive formulations with unprecedented efficiency, accelerating the discovery of novel lipid structures that enhance mRNA delivery while reducing reliance on labor-intensive and costly experimental iterations. Furthermore, AI-guided LNP design allows for a more comprehensive understanding of the multifaceted interactions between LNP components and biological systems, including factors such as biodistribution, cellular uptake mechanisms, and immune responses. The convergence of AI and nanotechnology holds substantial promise for overcoming existing limitations in mRNA delivery, enabling the development of LNPs with enhanced targeting capabilities, improved potency, and reduced adverse effects. With the continuous development, AI is expected to drive a paradigm shift in LNP engineering, providing new opportunities for mRNA therapeutics across a wide range of applications. This review will provide a comprehensive overview of the critical role of AI in guiding the design of LNPs for mRNA therapeutics (Fig. 1). By exploring recent advances in AI-guided LNP development, we focus on how data-driven approaches accelerate the optimization and simplify the discovery of optimal LNPs. Additionally, we explore novel AI-based strategies for designing superior mRNA delivery systems. The combination of AI and nanomedicine has great translational potential, offering a faster and more efficient pathway for developing efficient, safe, and targeted mRNA delivery systems.

2. Challenges in the design of LNPs

Due to the remarkable success of the mRNA vaccines against COVID-19 pandemic, LNPs have emerged as the most promising mRNA delivery vehicles with great clinical translational potential^{44,45}. Beyond utilization in the COVID-19 mRNA vaccines, LNPs also become the leading mRNA delivery platform for a broad range of therapeutic applications, such as cancer immunotherapy, treatment of hereditary disorders, and protection against infectious diseases^{1,2}. As mRNA-based therapies continue to expand, the demand for LNPs has surged, necessitating more sophisticated designs to enhance their efficiency, specificity, and safety. Although the feasibility of the mRNA-LNP delivery systems has been validated, several challenges remain in the design and optimization of LNPs, including screening the chemical structures of lipids, overcoming the liver tropism of LNPs to achieve organ/tissue-specific targeting, and further understanding the complex *in vivo* delivery mechanism^{46–48}. Traditional experimental methods for optimizing LNPs are often labor-intensive, time-consuming, and require complex data analysis, thereby limiting the development process to a certain extent. With the development of AI and ML, researchers are exploring how to use these advanced computing tools to effectively solve bottlenecks in biomedical development, thereby accelerating the drug development pipeline. To cope with the above limitations, AI has served as an important facilitator for analyzing and processing large

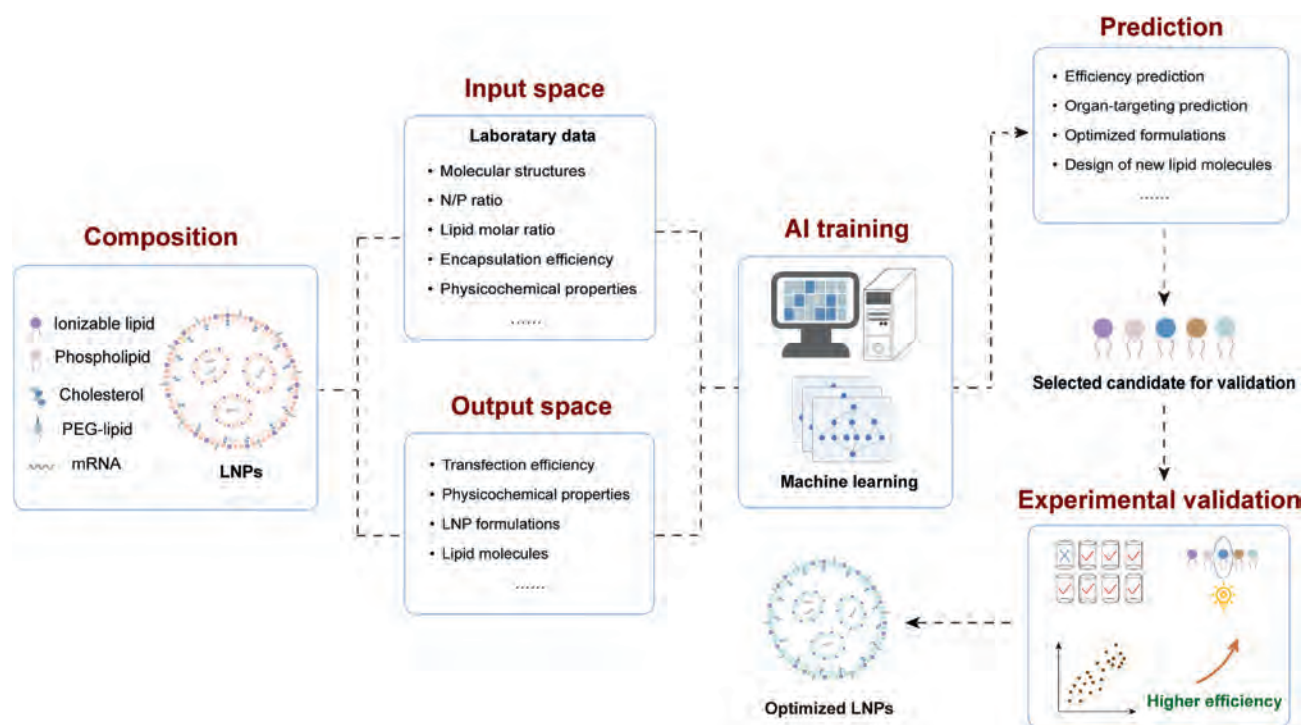


Figure 1 Overview of AI-guided design of LNPs.

amounts of complex data, allowing for more efficient development of LNP-based mRNA delivery systems. With the unique advantages of analyzing the generalized results and predicting the trends, AI provides a promising solution for accelerating the LNP development.

2.1. Lipid structural design and optimization of LNPs

Typical LNPs are usually composed of four components: ionizable lipids, phospholipids, cholesterol, and PEG-lipids^{20,49}. Studies have shown that different chemical structures and formulations may affect the trafficking of nanoparticles *in vivo*^{50,51}. Minor alteration in lipid compositions may affect their physicochemical properties, such as particle size, surface charge, and mRNA binding efficiency, all of which can influence mRNA delivery efficiency and organ targeting^{52–54}. Thus, the major challenges in LNP design include optimizing the molecular structures of ionizable lipids, screening for optimal formulations, and improving functionality to ensure maximum delivery efficiency, organ-specific targeted delivery, and minimal toxicity^{55,56}. However, the precise screening and optimization of these lipid components is a complex process. Traditional experimental approaches to optimize lipid formulations are time-consuming and often fail to rapidly reveal the most effective lipid combinations, given the large number of potential lipid candidates. AI and ML algorithms can be trained on large datasets of lipid formulations and their associated performance metrics to optimize LNP lipid compositions. For instance, deep learning (DL) models can analyze the relationship between lipid chemical structures and LNP characters, such as size, pK_a , zeta potential, mRNA encapsulation efficiency, and transfection efficiency⁴⁰. By utilizing these models, researchers enable the prediction of the most effective lipid combinations for a specific therapeutic target in a shorter

time, increasing the efficiency and accuracy of the optimization process.

2.2. Realizing organ-specific targeting of mRNA-LNP delivery systems

The mRNA-LNP delivery systems have shown great promise across various disease applications, with the potential for precise targeted therapy through optimized delivery strategies^{26,57}. However, a major issue in LNPs is their liver tropism post systemic administration. This hepatic tropism primarily arises because intravenously administered LNPs bind to apolipoprotein E (ApoE) in the bloodstream, which subsequently interacts with low density lipoprotein (LDL) receptors in the liver, leading to preferential liver accumulation^{58–60}. While this property is beneficial for liver-targeted therapy, it limits the broader therapeutic potential of LNPs for targeting other organs, such as the lung, spleen, heart, and tumors^{51,61}. Several strategies have been identified to achieve organ-specific targeting, including altering the chemical structures of lipids, changing the surface charges of LNPs, adjusting the compositions of LNPs, and modifying the surfaces of LNPs^{46,50,62–64}. However, since LNPs function in the complex biological environment, predicting how these alterations influence LNP distribution, cellular uptake, protein expression, and targeting remains challenging. Encouragingly, AI offers a powerful solution to predict how these changes in LNPs will affect their biodistribution and targeting by learning from large amounts of experimental data. Computational models can simulate the interactions of LNPs with various tissues or cell types to predict how lipid composition, particle size, and surface charge will affect the cellular uptake and tissue distribution^{65,66}. By incorporating data from biological sources, such as proteomics, transcriptomics, or immune system interactions, AI models can predict how different

LNP formulations target specific organs or reduce liver accumulation.

2.3. Understanding the mechanism of intracellular delivery of LNPs

Despite the widespread use of LNPs in mRNA delivery, research on the mechanisms of their *in vivo* behavior is still relatively limited. Once administered, LNPs undergo a series of interactions with immune cells, plasma proteins, and various tissues and cells^{67–69}. Specifically, due to the complex *in vivo* environment, the processes of cellular uptake, endosomal escape and mRNA release after LNP entry are very complicated and not yet fully elucidated. These biological processes are critical in determining the ultimate efficiency of mRNA delivery⁷⁰. Therefore, a deeper understanding of the LNP interactions in the body is essential for optimizing their design to enhance the efficacy and safety. ML provides a way to understand the complicated *in vivo* processes of mRNA-LNP delivery. By analyzing large datasets from high-throughput screening and cell imaging, ML algorithms can identify key factors that affect endosomal escape, cell transport, and immune response¹⁶. For instance, DL-based image analysis can track the behavior of LNPs in living cells, providing real-time insight into the processes of internalization, endosomal escape, and mRNA release^{71,72}. These advanced tools enable the identification of new mechanisms of action, expanding the application potential of LNPs.

Traditional methods for determining the top-performing ionizable lipids and LNPs rely on extensive screening and trial-and-error experiments^{38,66}. While such screening yields accurate results, it is time-consuming and requires substantial animal use, struggling to analyze the vast complexity of chemical structures and data correlations. These limitations are expected to be addressed with the help of AI. AI, especially ML and DL, is radically changing the optimization process of LNPs by offering tools to systematically analyze and predict the complex relationships between lipid compositions, formulation parameters, and functional outcomes. By integrating AI into LNP optimization, researchers can reduce the dependence on traditional trial-and-error methods, accelerate the discovery of optimal formulations, and significantly improve the accuracy and efficiency of LNP design. DL models, such as deep message-passing neural networks (D-MPNNs), are particularly valuable for screening vast lipid libraries, already assisting in the identification of novel antibiotics and ionizable lipids^{65,73,74}. AI-driven insights not only improve delivery efficiency, but also enable the development of personalized LNP formulations tailored to specific therapeutic needs, providing a promising approach for advancing mRNA therapeutics.

3. AI-guided applications in nanomaterial design

Given the complexity and interdependence of constituents in nanomedicine formulations, addressing above challenges requires more efficient screening methods rather than just relying on traditional trial-and-error approaches. In this context, AI has emerged as a powerful tool that can learn from large datasets, identify potential correlations and predict optimal formulations. By utilizing AI, researchers are beginning to shift the nanomedicine design process from experience-based optimization to a more systematic, data-driven strategy. The following section

explores how AI, specifically ML algorithms, can be applied to nanomaterial design and how these approaches can be extended to overcome limitations in current LNP development.

ML, the predominant paradigm in contemporary AI research, is undergoing revolutionary advancement^{35,75}. Its theoretical origins trace back to the threshold-based artificial neuron model proposed by Warren McCulloch and Walter Pitts in 1943⁷⁶. This seminal work not only established the mathematical foundation for neural computation, but also provided a crucial basis for developing subsequent ML frameworks. From the 1950s to the mid-1970s, propelled by breakthroughs in natural language processing, the development of intelligent systems based on ML principles entered its first golden age, attracting extensive academic attention. From the 1990s to the early 21st century, revolutionary developments in ensemble learning theory and kernel methods enabled ML models to demonstrate exceptional generalization capabilities in feature space mapping and high-dimensional data classification tasks^{77,78}. In subsequent technological evolution, feature engineering and model hyperparameter optimization emerged as critical pathways for enhancing algorithmic efficacy^{79–81}. Currently, ML has become deeply embedded throughout biomedical research, playing indispensable roles in pivotal domains including drug target discovery, drug synergy prediction, and delivery system optimization^{82–84}.

The foundations of nanotechnology trace back to 1959, when Nobel laureate physicist Richard Feynman articulated his visionary lecture “There’s Plenty of Room at the Bottom” hypothesizing precise atomic-scale manipulation—an intellectual precursor now recognized as nanotechnology’s conceptual genesis⁸⁵. The formal term “nanotechnology” subsequently emerged through Norio Taniguchi’s systematic framework in his 1974 seminal technical paper⁸⁶. The 1980s witnessed breakthrough translation with liposomal systems emerging as pioneering drug carriers, leading to the 1995 regulatory milestone of Doxil, the first FDA-approved nanopharmaceutical formulation⁸⁷. Entering the 21st century, paradigm-shifting advances in ML have propelled nanomaterials innovation into an accelerated discovery phase. These computational approaches now enable predictive modeling of nanoparticle-protein corona interactions, intelligent optimization of spherical nucleic acid architectures, and precision engineering of LNP-based mRNA vaccines, fundamentally transforming nanosystem design through machine-augmented molecular informatics (Fig. 2)^{4,88,89}.

3.1. The classification of ML algorithms

ML can be categorized into four types, primarily determined by the amount and type of supervision during the training process: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (Table 1)⁹⁰. Within supervised learning, the training data given to the algorithm includes the desired outputs, which are clearly labeled. Supervised learning involves both independent and dependent variables and can be used for classification and regression tasks^{91,92}. Unsupervised learning uses training data that is unlabeled, and the algorithm learns automatically without guidance. In this case, the data only includes independent variables without dependent variables^{93,94}. Semi-supervised learning methods typically rely on certain assumptions about the data. The performance of these algorithms is guaranteed only when these assumptions are met, which is one of the main challenges limiting the widespread application^{95,96}. Reinforcement learning is a learning mechanism that aims to learn

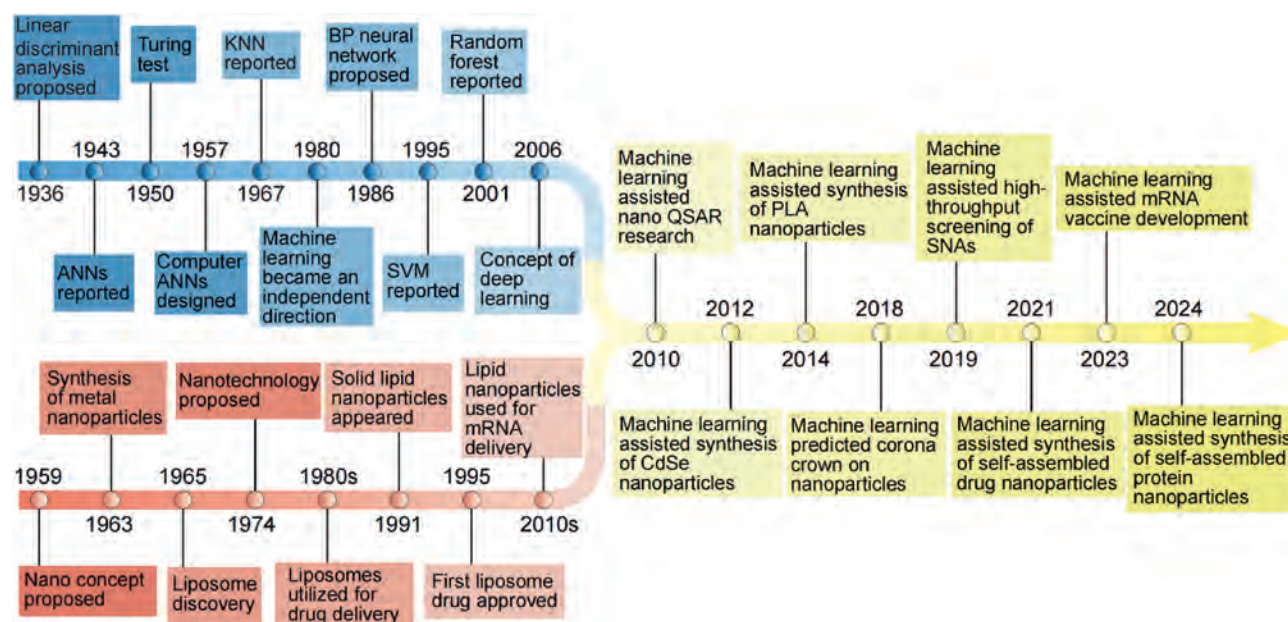


Figure 2 Timeline of key milestones for AI, nanoparticles, and their combination.

optimal behavior by mapping states to actions that maximize the rewards obtained. Such an agent requires constant conduct of experiments in the environment and optimizes the correspondence between state and behavior through feedback provided by the environment. Therefore, repeated experimentation and delayed rewards are two key characteristics of reinforcement learning^{96,97}. Diverse ML algorithms and their applications are summarized in Tables 1^{98–114}.

3.2. The application of ML algorithms

3.2.1. Supervised learning

Random Forest (RF) is a supervised learning algorithm first introduced by Tin Kam Ho in 1995¹¹⁵ and then developed by Leo Breiman¹¹⁶, which is suitable for both classification and regression problems. RF is an ensemble learning algorithm consisting of multiple decision trees. One of its key features is its ability to reduce overfitting caused by individual decision trees, thereby improving model performance. The basic principle of RF can be summarized in three steps: random sampling, random feature selection, and majority voting^{117,118}.

Due to its superior performance, RF has been widely applied in predicting the delivery efficiency by LNPs. Compared to traditional lipid formulations, LNPs show significant improvements in endosomal escape and encapsulation rates. Additionally, LNPs with specific compositions can cross the placental barrier, transporting immunoglobulin G (IgG) via the FcRn receptor to the fetus for immunotherapy¹¹⁹. The transfection efficiency of nucleic acids is a key factor in treatment. By inputting a total of 18 features, such as particle size and zeta potential, from 41 LNP formulations and 48 different transport experiments into various models—Extreme Gradient Boosting (XGB), Decision Tree (DT), RF, Support Vector Regression (SVR), Light Gradient Boosting Machine (LightGBM), and Lasso with Least Absolute Shrinkage and Selection Operator (Lasso)—the results demonstrated that the RF model provided the most accurate predictions for nucleic acid delivery efficiency¹⁰³. SHapley additive explanation (SHAP)

analysis was used to evaluate feature importance, successfully synthesizing LNPs with low, medium, and high transfection efficiency, thereby offering some explanation for the RF model¹⁰³. Since the coronavirus pandemic, mRNA vaccines have revolutionized the vaccinology field and LNPs have been widely used for mRNA delivery. Bae et al.³⁸ used 314 features from 213 LNPs (e.g., the ratio of amine groups of ionizable lipids to phosphate groups of the encapsulating unit, the mass and molar ratio of ionizable lipids, cholesterol, etc.) to train an RF regression model to predict mRNA expression levels after intradermal injection in mice. During the 10-fold cross-validation of the RF regression model, the model achieved optimal performance using 15 features, resulting in a coefficient of determination (R^2) of 0.708 and a Pearson correlation coefficient (PCC) of 0.845, indicating successful prediction of mRNA expression efficiency after intradermal injection. Feature importance analysis identified phenol as the key substructure influencing mRNA encapsulation and expression.

3.2.2. Unsupervised learning and other types of machine learning

Principal component analysis (PCA) is one of the simplest and most fundamental unsupervised learning methods. Its background and development can be traced back to the early 20th century, and it remains a cornerstone in the field of multivariate statistical analysis. PCA was first proposed by Karl Pearson in 1901¹²⁰ and was later further developed by Harold Hotelling in 1933¹²¹. High-dimensional data often suffer from the “curse of dimensionality”. PCA addresses this by extracting the main features, identifying a lower-dimensional affine subspace and projecting all data points onto it to achieve dimensionality reduction¹²².

Romanò et al.¹²³ collected voltammograms of LNPs with three different lipid compositions, standardized the data, and used PCA analysis to extract the main features of the voltammograms (PC1 and PC2). Linear discriminant analysis (LDA) is used for further classification, and different formulations of lipid nanocarriers are distinguished along the PC1 and PC2 directions. Cross validation showed that the method had a prediction accuracy of 90% for

Table 1 Summary of the ML algorithms and their applications for nanomedicine.

Category	Method	Advantage	Disadvantage	Application	Ref.
Supervised learning	• <i>K</i> -Nearest Neighbors (KNN)	• Make no assumptions about the data • Low cost in the learning process	• Dependent on selection of <i>K</i> value	• Optimize the preparation conditions of mRNA-LNP vaccines • Advance the prediction of the impact of global AgNPs on soil enzymes	98, 99
	• Support vector machine (SVM)	• Proficient in solving complex learning tasks • No assumptions were made regarding the predictor variables	• Expensive computational costs • Can only handle continuous predictor variables	• Demonstrate the concentration determination of plasmonic nanoparticle mixtures • Highly accurate classification of different crystals	100, 101
	• Random forest (RF)	• Unlikely to overfit • High classification accuracy • Capable of handling missing values	• Slow running speed • Not easy to explain	• Accurately predict the pulmonary immune responses and lung burden of NPs	102, 103
	• Naive Bayes	• Can handle continuous and categorical predictor variables • Low training expenses • No hyperparameters need to be adjusted	• Continuous predictor variables need to have a normal distribution • Predictive variables need to be independent of each other	• Improve the efficiency of placental uptake of LNPs • Predict the potential environmental risks of nanoparticle exposure	104
	• Decision tree (DT)	• Easy to explain • Can handle categorical and continuous predictor variables	• Easy to overfit	• Simulate the cytotoxicity of nanoparticles • Predict the correlation between nanoparticle characterization results and cytotoxicity	105, 106
Unsupervised learning	Clustering				
	• <i>K</i> -means	• Scalable and efficient • Local optima can already meet the needs of clustering	• Initial <i>K</i> value needs to be determined in advance • <i>K</i>-means can only converge to local optima	• Identify all regions with distinct spectral features of nanoparticles • Predict the adsorption free energy of biomolecules on specific nanomaterials	107, 108
	• Expectation maximization algorithm	• Good performance for hybrid models • Not necessary to make assumptions about observational data	• Sensitive to initial values • Difficult to handle missing data • Easy to get stuck in local optimal solutions	• Retrieval of size distributions from the small-angle scattering patterns obtained from dense nanoparticle samples • Shelf-life prediction of nano-sol	109, 110
	Dimensionality reduction				
	• Principal component analysis (PCA)	• Low computational cost • Completely parameter free	• Cannot intervene in the processing process through parameter tuning • Eigenvalue decomposition has some limitations	• Determine the influence of various factors on the preparation of modified nanocomposites	111
	• Kernel PCA	• Overcome the approximate linear structure of data • Enhanced flexibility and adjustability	• High computational complexity • Unable to explain the mapping result	—	

Association rule learning	• Apriori	• Strong interpretability	• Data sparsity problem	–
	• Eclat	• Strong scalability	• Unable to handle continuous data	–
	• Self-training	• Easy to parallelize	• Lack of flexibility	–
Semi-supervised learning	• Monte Carlo tree search	• Space efficiency	• Unable to handle missing values	–
	• Semi-supervised rotationally invariant variational autoencoder	• Not dependent on model architecture	• Restricted by the distribution and quantity of labeled samples	–
	• Reinforced learning	• Not dependent on the dataset	• Difficult to find high confidence unlabeled samples	–
LNPs, lipid nanoparticles; NPs, nanoparticles; AgNPs, Ag nanoparticles; –, not applicable.	• Machine learning reinforced genetic algorithm	• Rotational invariance	• Complex implementation	• Classify gold nanoparticles with different properties 112
		• Compact latent space representation	• Limited scalability	
		• No need for extensive domain knowledge	• Potential for latent space discontinuity	• Design complex protein nanomaterials 113
		• Suitable for large state spaces	• Strong dependency on search iterations	
		• High adaptability	• High computational resource consumption	
		• Improves search efficiency	• High computational complexity	• Targeted search for selective cytotoxic inorganic NPs 114
		• Suitable for large scale and complex problems	• Complex parameter tuning	
			• High implementation difficulty	

three different test formulations, and the first and second principal components explained 98.9% of the data together. This method not only confirms the classification potential of electrochemical signals combined with PCA and LDA, but also provides a low-cost and portable solution for rapid quality control of liposome formulations, significantly improving the automation level of drug carrier development and clinical diagnosis¹²³. In the development of nanostructured lipid carriers, lipid solid dispersions, and self-emulsifying drug delivery systems, solid or semi-solid lipid excipients are typically mixed with liquid solvents or lipids to form stable formulations. Even when the excipients visually appear miscible once melting, they may exhibit microscopic heterogeneity, resulting in instability over time. For instance, cetyl palmitate + Transcutol® (heterogeneous) and polyethylene glycol-6000 (PEG-6000) + Tween 80® (homogeneous) exhibited significantly different degrees of microscopic miscibility¹²⁴. These systems were prepared in varying proportions for Raman spectroscopy analysis, and the resulting data underwent PCA. By observing the changes in the PC1 and PC2 directions of different systems, it is possible to distinguish the degree of uniformity of mixing of different compounds in the system.

Due to the more stringent usage conditions of semi supervised learning and reinforcement learning in the LNP field, their applications are currently limited. The critical challenge in developing nanopore biomarker analysis lies in directly identifying and quantifying target biomarkers from biological fluids containing molecular mixtures. Greive et al.¹²⁵ successfully classified peptide conformational variants at the single-molecule level with a semi supervised classification algorithm, which other ML algorithms could not accurately classify. The combination of reinforcement learning and genetic algorithm can improve the accuracy of the algorithm. For instance, using ML enhanced genetic algorithm, selective cytotoxic inorganic nanoparticles can be discovered by screening 14,900 candidate substances, and the accuracy of the model is high (Q1, 0.80; RMSE, 13.6)¹¹⁴.

3.3. Model recommendation

When studying drug delivery with LNPs, researchers often encounter various scenarios, such as optimizing and predicting drug delivery efficiency, controlling drug release rates, evaluating targeting capabilities, and optimizing LNP formulations^{20,29,126}. Selecting the appropriate ML model for each scenario can significantly improve the optimization efficiency and effectiveness of LNP delivery systems. We summarize some recommendations for choosing different ML models based on specific scenarios.

When predicting the release rate of a drug from different LNP formulations, certain characterization data are typically available, such as lipid composition, particle size, zeta-potential, and pH. The target variable is usually the concentration or release rate of the drug at a specific time. In this case, ML models tend to favor RF and Gradient Boosting Regression models (*e.g.*, XGBoost and LightGBM). RF can handle complex nonlinear models and is highly sensitive to interactions between features. It is well-suited for processing high-dimensional data and can provide feature importance scores, helping to identify which factors show the most significant impact on the release rate. Gradient Boosting Regression models can continuously optimize the model during the iterative process, gradually reducing errors and achieving higher prediction accuracy, especially when dealing with complex nonlinear relationships^{127,128}.

A key factor of successful LNP delivery is their internalization by targeted cells. Gandeek et al.¹²⁹ indicated that LNPs primarily internalized through classical endocytosis mechanisms. When predicting whether different LNPs can be taken up by certain cells, the Support Vector Machine (SVM) model is often a suitable choice. SVM can learn the causal relationship between the properties of different LNPs and their cellular uptake. SVM is capable of handling high-dimensional data and can find better decision boundaries for linearly inseparable data through kernel functions¹³⁰. In the context of LNP cellular uptake, where cellular responses may not be linear, SVM is particularly effective in addressing such issues.

The targeting ability of LNPs represents another important issue of concern. Predicting whether LNPs can target specific cells or organs based on their characterization results is a multiclass classification task, for which RF is recommended³⁸. RF is not only suitable for regression tasks but is also highly effective for classification tasks. However, the XGBoost model often shows higher accuracy in handling multiclass tasks with class imbalance. The Multi-Layer Perceptron (MLP) is a basic neural network framework and is also applicable to multiclass problems. By increasing the number of layers, MLP can capture complex nonlinear relationships^{131,132}. In case of large and complex datasets, MLP should be prioritized as the model of choice.

Some reinforcement learning models, such as Deep Q Networks (DQN) and Q-learning, are less used in LNP research. Optimizing the LNP formulations to maximize drug delivery efficiency and targeting is a dynamic optimization problem that may involve multiple experimental feedbacks. Each parameter in the formulation may potentially influence the final drug delivery outcome. Reinforcement learning is well-suited for decision optimization in feedback environments¹³³, which can continuously adjust the LNP formulations based on experimental results to achieve the remarkable delivery performance.

The choice of different models and dataset sizes can result in different costs and time requirements. In this context, multiple models can be tested for the task, allowing the best model for subsequent analysis based on the result accuracy and the model's performance.

4. AI-guided key techniques and methods in LNP design

4.1. AI-guided optimization of LNP compositions and formulations

The LNP properties, such as particle size, polydispersity (PDI), zeta potential, and pK_a , play critical roles in determining their delivery capabilities⁵². The composition and formulation of LNPs represent the critical factors that govern these characteristics. To apply LNPs in mRNA therapies, extensive screening of nanoparticle structure, composition, and formulation is typically required. Given the complexity and high number of variables involved, conventional experimental methods are both time-consuming and resource-intensive. Here, AI offers a powerful tool to analyze and uncover hidden correlations within large and complex datasets. Leveraging AI's capacity to process massive amounts of data and recognize patterns, researchers can predict the most promising lipid compositions and formulations for optimal mRNA delivery. This not only accelerates the development of LNPs, but also significantly improves the efficiency and accuracy of the design process.

4.1.1. Optimization of the LNP components

The LNP design for mRNA delivery requires careful selection and optimization of various components, including ionizable lipids, phospholipids, cholesterol, and PEG-lipids^{20,28}. Ionizable lipids, the core component of LNPs, remain neutral under physiological conditions but become protonated and positively charged in acidic endosomes, thereby reducing the lipid toxicity^{28,32,134}. Optimizing ionizable lipids typically involves altering the tail lengths, the head-group chemical structures, and the linker groups¹³⁵. These adjustments are essential for enhancing the mRNA encapsulation efficiency and endosomal escape of nanoparticles. Li et al.³⁹ proposed an approach combining ML and combinatorial chemistry to accelerate the discovery of novel ionizable lipids for optimizing mRNA. The authors used the binary dataset to evaluate the performance of different ML algorithms, including RF, gradient boosting and logistic regression. The results demonstrated that XGBoost algorithm presented superior predictive performance, with the highest PR curve of 0.987 and ROC curve of 0.983. High-throughput screening (HTS) and ML prediction resulted in the discovery of novel ionizable lipid 119-23 with excellent performance from a large chemical library, whose mRNA delivery efficiency outperformed existing commercially available lipids (e.g., DLin-MC3-DMA (MC3) and SM-102). Bae et al.³⁸ utilized a branching criterion under the RF model to assess the importance of individual features on mRNA expression. The FP169 (phenol) was identified as the optimal functional head group in ionizable lipid libraries, with the highest importance score of 0.297. The hydroxyl group in phenol enhanced the interaction with mRNA, thereby improving both the efficiency and stability of mRNA-LNPs (Fig. 3A)³⁸.

Phospholipids are typically used as the helper lipids in LNPs and serve to maintain the stability of nanoparticles^{27,136}. Recent research has focused on modifying phospholipids to improve their interaction with ionizable lipids and enhance the overall performance of LNPs¹³⁷. Cholesterol is another essential component of LNPs and enhances the stability of LNPs. It improves the structural integrity of the lipid bilayers, increasing the overall particle stability and protecting encapsulated mRNA from degradation^{20,52,138}. Researchers have explored the possibility of modifying the cholesterol structure or substituting it with other sterols for enhanced safety and delivery efficacy^{139–141}. Although PEG-lipids account for the lowest molar percentage of LNPs, they exert influences on many key properties of LNPs, including particle size and dispersion, nanoparticle stability, and their bio-distribution and blood circulation time^{52,142}. Positioned on LNP surfaces, PEG-lipids can reduce the protein coronal adsorption and recognition by the immune system, prolonging the LNP circulation time in bloodstream and modulating the particle biodistribution *in vivo*^{28,59}. Mui et al.¹⁴² discovered that the chain length and concentration of PEG-lipids significantly affected the *in vivo* journey and delivery efficiency of LNPs, with longer carbon chains of PEG-lipids extending blood retention. Since adjustments between components in LNPs are delicate and impactful, researchers are investigating various properties between fine-tuned components to optimize mRNA delivery. While current AI-guided design of LNPs primarily focuses on optimizing ionizable lipid libraries and formulations, the potential extends far beyond these applications. In the future, it is expected that AI will also be applied to the screening of phospholipids, sterols, and PEG-lipids, providing a more comprehensive approach to optimize LNPs.

AI is a potent tool for accelerating LNP design and screening process, allowing rapid evaluation of lipid chemical libraries and

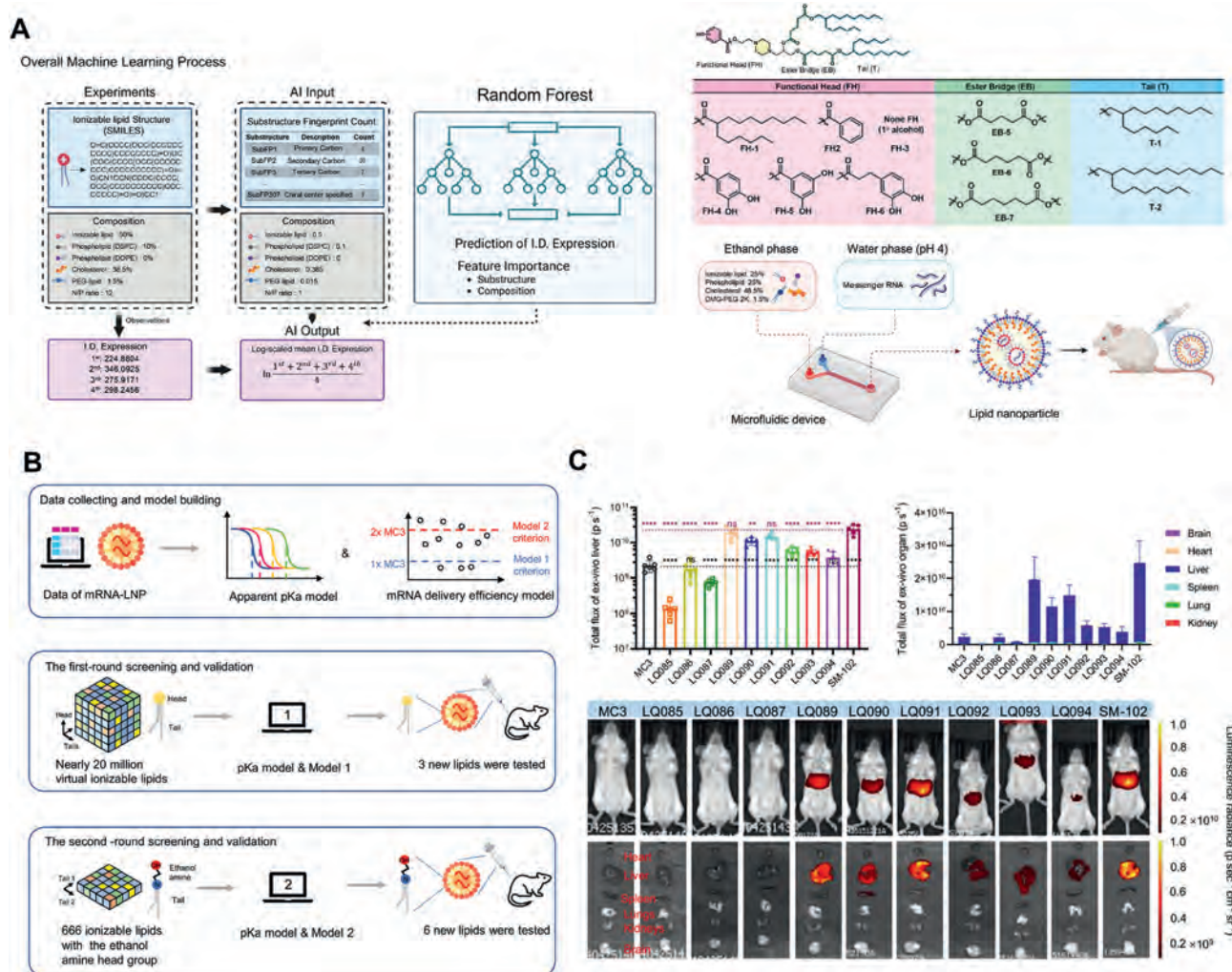


Figure 3 AI-guided key processes in LNP design. (A) A schematic diagram of predicting the mRNA delivery efficiency and assessing the importance of characteristics in ionizable lipid substructures and LNP compositions through ML. Reprinted with the permission from Ref. 38. Copyright © 2024 Wiley. (B) Schematic diagram of rational design of ionizable lipids driven by AI. The acquired data were used to construct and train a model to analyze the potential association between the apparent pK_a of LNPs and their mRNA delivery efficiency. (C) Organ expression of luciferase mRNA after intravenous injection of new LNPs designed *via* AI guidance. Reprinted with the permission from Ref. 66. Copyright © 2024 Springer Nature.

identification of optimal formulations^{39,143}. ML models, especially those based on supervised learning, have been used to predict the efficacy of different lipid combinations based on historical data and experimental results^{35,144}. For instance, algorithms including RF, SVM, and neural networks have been utilized to associate the lipid chemical structures to their mRNA encapsulation efficiency, cellular uptake, and transfection efficiency. This dynamic approach allows for the continuous improvement of LNP design by autonomously identifying and promoting the most effective formulations.

The integration of AI into LNP research not only accelerates the discovery of novel lipid structures, but also promotes the design of highly efficient, targeted delivery systems for mRNA therapeutics. By utilizing data-driven insights and predictive modeling, AI-driven methodologies demonstrate potential to revolutionize nanomedicine and enable the rapid translation of advanced lipid-based drug delivery systems from the laboratory to the clinic.

4.1.2. Analysis and prediction of LNP physicochemical properties by ML

As an important tool for analyzing and predicting the physicochemical properties of LNPs, ML can help to provide a deeper understanding of how lipid compositions affect their structural and functional properties⁶⁶. Trained by inputting a large amount of data from experimental and computational results, ML can reveal the complex correlations between lipid compositions and key physicochemical parameters (*e.g.*, particle size, surface charge, pK_a , mRNA encapsulation efficiency, stability, and biodegradability), and predict physicochemical parameters that are more conducive to mRNA delivery.

The physicochemical properties of LNPs are governed by a variety of factors, including lipid structures, component molar ratios, and formulation conditions. Traditional experimental methods for characterizing and analyzing these properties are often complicated and incomprehensive. ML algorithms provide a way to analyze high-dimensional datasets and identify key

predictors of LNP behavior with high accuracy. For instance, particle size and PDI are key factors influencing LNP stability and biodistribution. ML models trained on formulation parameters and experimental measurements can accurately predict these properties, facilitating rapid optimization of nanoparticles. Maharjan et al.¹²⁷ employed ML algorithms, including XGBoost, Bayesian Optimization, and Self-Validated Ensemble Model (SVEM), to improve productivity and optimize LNP formulations. They utilized XGBoost to comprehend the significance of various LNP parameters and predict the particle size, PDI, zeta potential, pK_a , heat trend cycle, and mRNA encapsulation rate of LNPs. The results showed that the XGBoost model possessed excellent predictive capability, with predictive response higher than 94%. Similarly, pK_a prediction is useful for assessing carrier stability and *in vivo* organ targeting, thereby guiding the screening of stable, low-toxicity, and targeted delivery systems⁴⁷. In addition, AI models can also speed up the screening process of ionizable lipids by effectively predicting the pK_a and mRNA delivery efficacy. Wang et al.⁶⁶ used the LightGBM algorithm to construct a pK_a -based model and successfully screened the ionizable lipid LQ089 with higher mRNA delivery efficiency than the commercial lipid DLin-MC3-DMA and comparable efficacy to SM-102. The LightGBM model showed superior performance over the RF algorithm in several parameters, with its Recall of 0.86 and its accuracy (ACC) of 0.90. Their results demonstrated that effective mRNA delivery was often associated with LNP pK_a in the range of around 6.0 to 7.0 (Fig. 3B and C)⁶⁶. By analyzing experimental degradation data, supervised learning techniques can predict changes in size, surface charge, mRNA encapsulation efficiency, and delivery efficiency over time, providing insights into formulation stability and fueling the LNP development.

4.2. AI-guided understanding of LNP *in vivo* targeting

4.2.1. Targeting guided by nano-bio interactions

Once entering the body, LNPs interact with cells, tissues, and other components. These interactions determine the fate of LNPs in the body, deciding whether they accumulate in a target organ, tissue, or cell and exert therapeutic effects at specific sites (Fig. 4)^{16,145}. Therefore, understanding the factors that influence the LNP targeting is essential for precision therapy. ML algorithms offer unique advantages in data clustering, classification, and prediction. By performing high-throughput analysis on existing data, these algorithms can not only accurately predict the targeting efficiency of LNPs, but also guide the synthesis of new targeted LNPs.

The targeting ability of LNPs largely depends on their properties, such as size and composition. Smaller nanoparticles show a stronger capacity to penetrate the endothelial barrier and enable passive targeting to the tumor tissues *via* the enhanced permeability and retention (EPR) effect¹⁴⁶. Larger nanoparticles tend to be captured by the liver-spleen mononuclear phagocyte system (MPS), making them suitable for liver-targeted delivery, such as in siRNA or mRNA-based liver therapies¹⁴⁷. The core composition of LNPs includes four types of lipids, and the ratio and type of each lipid directly affect their targeting efficiency⁵². For instance, cholesterol plays a critical role in modulating membrane fluidity, enhancing the stability and mRNA encapsulation efficiency. A higher proportion of cholesterol aids in LNP uptake by liver cells, while a lower cholesterol content may reduce hepatic accumulation^{148,149}.

Hunter et al.¹⁵⁰ pioneered an integrated approach combining advanced image analytics with ML to enhance the targeted delivery of LNPs (Fig. 5A)¹⁵⁰. Utilizing high-content image analysis, they

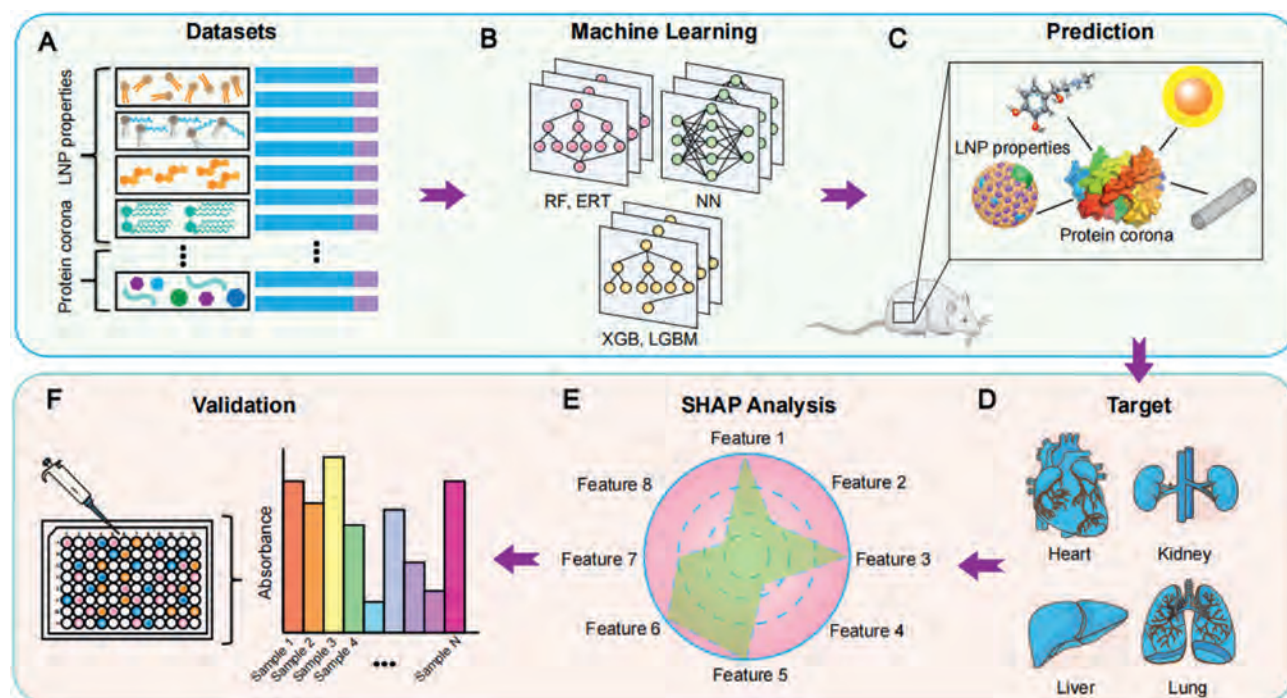
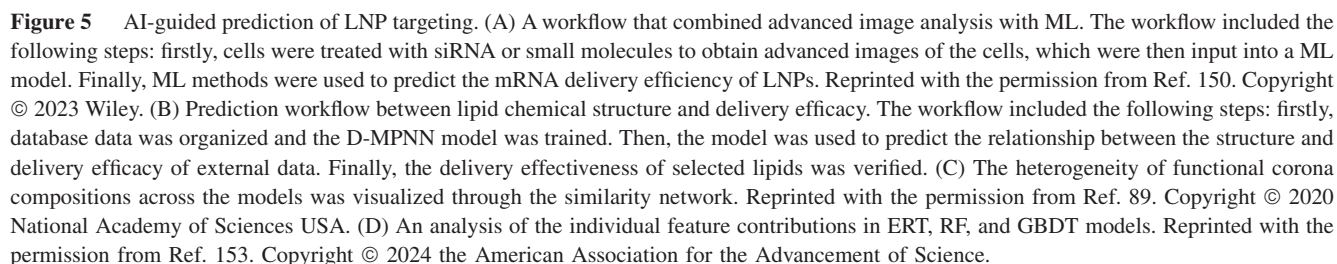


Figure 4 The process of AI-guided understanding of *in vivo* targeting of LNPs. (A) Organization of the LNP characteristic and protein corona datasets. (B) Training of different machine learning models using data. (C) Prediction of LNP properties and protein corona. (D) Organ targeting prediction of LNPs. (E) Identification of key features by SHAP analysis. (F) Experimental verification of predicted LNPs.



By systematically integrating 20 multi-source datasets encompassing over 9000 LNP activity measurements, Witten et al.⁶⁵ developed a D-MPNN model to establish predictive relationships between lipid chemical structures and delivery efficacy (Fig. 5B). This framework demonstrated exceptional extrapolation

capability to structurally distinct compounds beyond the training set. Through virtual screening of 1.6 million candidate lipids, two novel ionizable lipids (FO-32 and FO-35) were identified. Experimental validation revealed that FO-32 achieved benchmark performance in murine pulmonary nebulization delivery, while FO-35, featuring an innovative terpene-derived tail architecture, accomplished unprecedented transfection efficiency across the full pulmonary epithelium of ferrets—a critical preclinical model for respiratory therapeutics. Notably, these optimized lipids contained chemically counterintuitive functional group combinations (*e.g.*, alkyne motifs with branched carboxylic acid tails) that conventional rational design approaches would unlikely predict, thereby demonstrating DL’s unique capacity to transcend traditional chemical space exploration limitations. This study establishes that ML not only enhances the screening efficiency for existing lipid libraries but also enables *de novo* design of four-component reaction systems, creating a novel paradigm for developing targeted gene delivery vehicles against respiratory diseases.

4.2.2. Targeting guided by protein corona

The protein corona, a dynamic biomolecular adsorption layer formed on nanoparticle surfaces upon entering biological micro-environments, governs key pharmacological determinants ranging from systemic circulation half-life and immune evasion capacity to tissue-specific biodistribution^{67,151}. Conventional corona characterization relying on mass spectrometry and *in vitro* simulation faces critical limitations in predictive accuracy and design capability due to experimental throughput constraints, interspecies variability, and dynamic evolution complexity. Recent advances in AI have revolutionized this paradigm through systematic integration of nanoparticle physicochemical descriptors (size, zeta potential, hydrophobicity), protein interactome data, and multi-omics profiles^{127,152}. ML architectures now enable deconvolution of the tripartite “nanoparticle–corona–biological barrier” interaction network. Notably, graph neural network-based frameworks have achieved probabilistic modeling of corona composition under specified physiological conditions, while reinforcement learning algorithms demonstrate potential for reverse-engineering surface modification strategies to promote preferential adsorption of targeting proteins (*e.g.*, apolipoproteins, transferrin). Emerging AI-driven multiscale modeling transcends the classical “passive corona formation” dogma by simulating dynamic corona evolution coupled with membrane receptor cooperativity, establishing mechanistic foundations for designing adaptive targeting delivery systems with microenvironment-responsive intelligence.

In a prior study, a total of 652 pieces of data associated with protein corona on diverse nanoparticles (NPs) were extracted and analyzed (Fig. 5C)⁸⁹. Eight qualitative characters (*e.g.*, NP type, NP shape, and unmodified NPs) and 13 quantitative factors (*e.g.*, pH of dispersion medium, zeta potential, PDI, and plasma concentration for incubation) were incorporated into the analysis. Using a RF model combined with 10-fold cross-validation, this study comprehensively examined the relationships between the physicochemical characters and biological functions of protein coronas on the NP surface. The results indicated that unmodified NPs and surface modifications were the two most critical characters that determined the formation of protein corona. The ML model achieved high predictive accuracy (R^2 values mostly above 0.75) in estimating the relative protein abundance (RPA) of key functional proteins in the corona, such as immune proteins, complement proteins, and apolipoproteins⁸⁹. Furthermore, the model successfully predicted cell recognition processes mediated by the protein corona, including macrophage uptake and cytokine release. Supported by experimental validation, this study unveiled the pivotal role of functional epitopes within the protein corona in modulating the biological recognition and targeting efficacy of NPs. This novel approach addresses the limitations of traditional models in predicting protein corona formation mechanisms under complex multidimensional conditions, significantly reduces experimental workload, and provides an efficient tool for elucidating nanoparticle–protein interface interactions as well as designing targeted nanomedicines.

In addition, a multitasking ML framework was employed to predict the RPA of proteins within the protein corona (Fig. 5D)¹⁵³. This study utilized binary classification models, such as extremely randomized trees (ERT) and RF, to determine whether a protein adsorbs onto nanoparticles, followed by regression models (ERT, RF, and GBDT) to predict the specific RPA values. Through SHAP interpretability analysis, it was revealed that features such as “unmodified nanoparticle types” (*e.g.*, Au and SiO₂), “incubation protein source” (*e.g.*, human serum and HS), and “centrifugation

temperature” significantly influenced the prediction outcomes¹⁵³. This research provided a data-driven strategy for designing targeted nanoparticles: by harnessing ML to identify key features. Parameters, such as nanoparticle surface modifications and incubation conditions, can be optimized to selectively modulate the protein corona composition, thereby enhancing the targeting efficiency of nanocarriers to specific organs or cells. This work not only validated the advantages of ML in protein corona prediction (*e.g.*, achieving an AUROC of 0.97 in classification tasks and an R^2 of 0.53 for regression tasks using the ERT model) but also offered a novel approach for “active protein corona design to enable precision targeting” in nanomedicine.

4.3. AI-guided prediction of LNP trafficking in vivo

The *in vivo* transport of LNPs is an intricate process influenced by numerous factors, including the *in vivo* circulatory environment, tissue biodistribution, cellular uptake, and intracellular translocation. An in-depth understanding of these interactions is imperative for optimizing mRNA-LNP drug delivery systems. This process typically requires characterizations by multiple tools, such as imaging, multi-omics studies, and pharmacokinetic analysis. AI-driven methodologies, encompassing ML and DL algorithms, offer potent tools for predicting and optimizing these processes, consequently facilitating more efficient and targeted LNP design.

4.3.1. AI-guided prediction of the biodistribution and mRNA expression by LNPs in vivo

The mRNA-LNP biodistribution and expression are critical determinants of therapeutic efficacy, presenting how effectively the mRNA translates into functional proteins within target tissues. Post administration, LNPs circulate systemically, interact with plasma proteins, immune cells and the endothelial barrier, and accumulate in specific organs. The biodistribution of LNPs is affected by factors such as particle size, lipid composition, PEGylation and surface charge^{47,50,154}. The traditional method of evaluating the *in vivo* distribution of mRNA-LNPs is to inject fluorescently labelled mRNA and trace it using an *in vivo* imaging system (IVIS). This measurement is usually time-consuming with limited scope. However, AI-based models can integrate large-scale datasets from preclinical and clinical studies to establish a more accurate predictive framework for the distribution of mRNA-LNPs.

The ultimate goal of LNP-based mRNA delivery is efficient protein expression in targeted cells. After reaching the targeted tissue, LNPs must be taken up by the cells, escaped from the endosomes, and release the mRNA cargo in cytoplasm for translation. AI-driven models help to predict and optimize these processes. Graph neural networks (GNNs) have strong structure-awareness capabilities, which can efficiently process complex molecular structures and learn the relationships between molecular structures^{40,155}. Supervised learning algorithms trained on experimental gene expression datasets can predict how different LNP structures and formulations affect mRNA translation efficiency in different tissues⁴⁰. These models take into account factors such as lipid composition, chemical structures, nanoparticle size, and cell tropism to estimate expression levels and duration. For instance, Xu et al.⁴⁰ proposed the AI-Guided Ionizable Lipid Engineering (AGILE) platform, which combined DL and combinatorial chemistry to greatly expedite the development process of LNPs (Fig. 6A). In this study, 60,000 lipids were trained by self-supervised learning and combined with high-throughput experimental data to fine-tune the model, which was

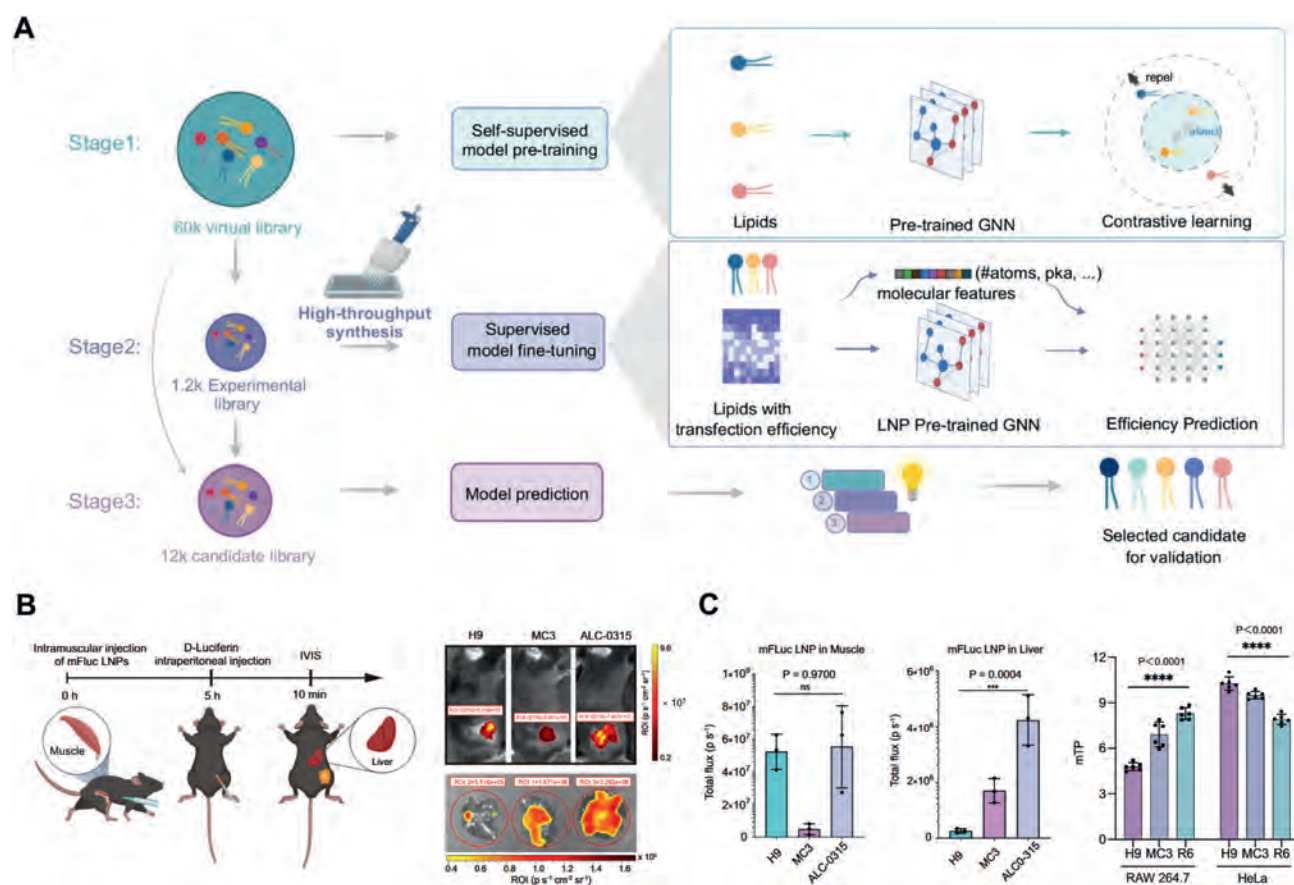


Figure 6 AI-guided prediction of the *in vivo* process of LNPs. (A) Illustration of the workflow designed by the AGILE platform. The workflow of the AGILE platform included the following steps: firstly, the model was pre-trained through self-supervision with a virtual library; then, the model was adjusted through supervised learning with laboratory data; finally, the model is applied to analyze and predict the candidate lipid library. (B) IVIS imaging exhibited the delivery efficiency of intramuscularly injected mFluc LNPs *in vivo* (0.5 mg/kg). (C) Comparison of H9, MC3 and ALC-0315 LNP-mediated mRNA delivery in the muscle and liver *via* intramuscular injection. The mRNA delivery in RAW 264.7 and HeLa cells was also shown. Reprinted with the permission from Ref. 40. Copyright © 2024 Springer Nature.

able to accurately predict the mRNA transfection efficiency. Finally, the lipid H9 for efficient muscle delivery and R6 for high macrophage transfection were successfully screened (Fig. 6B and C)⁴⁰. In addition, AI-driven molecular dynamics (MD) simulations can help to identify lipid structures that enhance mRNA stability and improve mRNA expression efficiency^{152,156}. A previous study by Gao et al.¹⁵⁷ simulated the LNP assembling in neutral buffer by MD, and the results revealed that the encapsulation efficiency increased along with the increase of N/P ratio, with nucleic acid completely encapsulated at N/P ratio of 6.

In addition, AI can help to design LNPs that optimize organ-specific expression by analyzing biodistribution data and tissue-specific transfection rates. For example, transformer-based models can predict how changes in lipid composition affect mRNA uptake and translation in hepatocytes *versus* muscle^{38,40} or lung tissue⁶⁵. These insights aid in developing precise nanomedicine approaches to ensure that LNP formulations are tailored for specific therapeutic applications, such as vaccines, and tissue-specific medicines.

4.3.2. AI-guided prediction of the immune response and inflammation

The immunogenicity of LNPs presents a double-edged sword. On the one hand, LNPs can enhance the vaccine efficacy by inducing helper T cells and promoting humoral response, acting as vaccine

adjuvants^{158,159}. On the other hand, they may trigger inflammation and lead to side effects (*e.g.*, inflammation at the injection site or a systemic inflammatory response)¹⁶⁰. LNP-induced immune responses and inflammation affect their safety and efficacy¹⁶¹. While a certain degree of immune activation is desirable in vaccine applications, excessive or unintended immune responses can result in adverse reactions, rapid clearance, and reduced therapeutic efficacy¹⁶². With its powerful data extraction and data analysis capabilities, AI has been applied in the field of immunotherapy, enabling the analysis of parameters affecting the immune response^{38,163,164}. Bae et al.³⁸ delved into the effect of molecular substructure on the immune response of the mRNA-LNP delivery system by AI predictions. They discovered that the number of phenolic hydroxyl groups was intimately associated to the total IgG level and the expression level of inflammatory factors. AI-driven methods can analyze and predict the immunogenicity of the vectors through the results of prior extensive experiments, enabling rational design of LNPs to strike a balance between immune activation and biocompatibility.

Upon administration, LNPs interact with diverse constituents of the immune system, including macrophages, dendritic cells, and complement proteins. These interactions can trigger inflammatory cytokine release, pattern recognition receptor (PRR) activation, and adaptive immune responses¹⁶⁵. LNP-based AI models

trained on immunological datasets can predict how different LNP formulations affect these immune pathways, thus facilitating the design of nanoparticles with optimized immunogenicity profiles. Wang et al.¹⁵² utilized the LightGBM ML algorithm to predict the mRNA vaccine-induced IgG titer, an important measure of vaccine immunization efficacy, and the model performed well for accelerating the development of mRNA vaccines. The results showed that higher N/P ratio (e.g., 6:1) was usually correlated with higher mRNA delivery efficiency and enhanced antigen expression. In addition, the inflammation induced by LNPs is a key factor for evaluating their safety and efficacy. Inflammatory responses can be predicted by AI, analyzing how LNPs interact with pro-inflammatory signaling pathways. DL models trained on transcriptomic and proteomic datasets can identify biomarkers associated with inflammation and help to design lipid structures with reduced immune activation^{166–168}. ML can further refine LNP formulations by iteratively optimizing lipid composition to minimize inflammatory markers such as IL-6, TNF- α , and IFN- γ , while maintaining their ability to deliver mRNA efficiently^{38,127,152}. The RF algorithm is highly flexible, robust and interpretable, and excels in processing and analyzing complex multidimensional data¹¹⁷. Bae et al.³⁸ used the RF algorithm to analyze multiple parameters of lipids associated with the immune response. The results indicated that the immune potency of LNPs was negatively correlated with ester bridge + tail (EB + T) chain length. Furthermore, AI can help to rationalize the selection of anti-inflammatory modifications, such as alternative lipid structures and biodegradable lipids, to reduce undesired immune activation.

Overall, AI-guided assessment and prediction of immune and inflammatory responses of LNPs enable more precise control of immunogenicity. By integrating ML, DL, and multi-omics data analysis, AI provides valuable guidance for analyzing the immune interactions associated with LNPs, facilitating the development of safer and more effective nucleic acid therapies¹⁶⁹. These advances not only improve the biocompatibility of LNPs, but also pave the way for personalized nanomedicine, allowing for optimizing the therapeutic efficacy based on individual immune profiles.

5. Conclusions

The integration of AI into mRNA-LNP delivery systems has ushered in a transformative era, revolutionizing nanoparticle synthesis, drug delivery optimization, and mRNA-based therapeutics^{40,66}. The synergy between ML techniques and nanotechnology has not only enabled the development of more efficient, targeted mRNA therapies but also deepened our understanding of complex nano-bio interactions that were previously difficult to decipher^{152,170}. Among ML approaches, DL models have demonstrated exceptional capacity to process and interpret vast datasets^{12,41,171,172}. These models effectively decode intricate nano-bio interactions, uncovering hidden patterns in nanoscale systems. In the realm of nanoparticle synthesis, ML algorithms are accelerating the discovery of novel materials and formulations by identifying key parameters that influence nanoparticle characteristics, including size, surface charge, and stability. This acceleration significantly reduces the trial-and-error approach, providing a faster and more efficient pathway to optimize the nanoparticle design^{173,174}. Furthermore, AI-driven mRNA delivery optimization enhances precise targeting of therapeutic cargos to specific

cells or tissues, thereby enhancing treatment function and minimizing off-target effects. By utilizing predictive models, researchers can better anticipate the pharmacokinetics and biodistribution of nanoparticles *in vivo*, ultimately paving the way for more personalized drug delivery strategies tailored to individual patient profiles.

Despite these promising advancements, the integration of AI into mRNA-LNPs remains in its early stages and presents several key challenges. While AI holds transformative potential for accelerating mRNA-LNP design, several critical limitations must be addressed. Firstly, the inherent “black-box” nature of current AI models poses dual challenges: their limited interpretability in feature selection and opaque decision-making processes impede both scientific understanding and clinical approval^{175,176}, although the utility of SHapley Additive exPlanations can partially explain the results of the model. For instance, models predicting organ-targeting of LNPs may inadequately distinguish whether the selectivity originates from lipid composition, particle size dynamics, or transient protein corona interactions, thereby hindering systematic optimization. Secondly, current data ecosystems exhibit inherent limitations. Training datasets are disproportionately skewed toward specific lipid classes (e.g., SM-102-dominated ionizable lipids¹³⁹) and murine preclinical models, introducing systemic biases that compromise generalizability across human tissues and pathophysiological states. Furthermore, the scarcity of standardized *in vivo* datasets correlating nanoparticle physicochemical properties (e.g., pK_a and PEGylation density) with organ-specific biodistribution profiles critically undermines model robustness. Thirdly, a fundamental disconnect persists between *in silico/in vitro* predictions and *in vivo* outcomes. AI models trained on *in vitro* transfection efficiency metrics often fail to account for physiological complexity (e.g., immune surveillance mechanisms, endothelial barrier dynamics, and time-dependent protein corona evolution), resulting in poor predictable accuracy for organ-specific delivery. Emerging hybrid prediction systems that integrate multi-omics data with advanced *in vitro* platforms (organoids, organ-on-a-chip) aim to address this gap by recapitulating tissue-level biological complexity^{177,178}. Finally, even with accurate predictions, translating AI-designed LNPs into clinical therapies faces regulatory uncertainties. Current guidelines lack frameworks for validating “black-box” algorithms, while ethical concerns persist regarding data privacy (e.g., patient-specific mRNA designs) and algorithmic accountability. Collaborative efforts among computational scientists, regulatory agencies, and clinicians are urgently required to establish standardized evaluation protocols.

The success of ML in designing LNPs also faces dual challenges of data quantity and quality. The current datasets in the mRNA-LNP field are usually limited in size and suffer from issues such as parameter dispersion, high dimensionality, and experimental noise. However, obtaining reliable results does not rely solely on data volume, and can also be balanced through some strategies. For small datasets (<100), setting appropriate learning rates and turn counts, the XGboost model can also achieve accurate prediction of multiple response variables ($R^2 > 0.95$, RMSE<0.91, and MAE<0.63)¹²⁷. Using a RF model and oversampling methods on small datasets can also identify the percentage of LNP transport and dynamic results over 24 h¹⁵⁷. Constructing physical and chemical descriptors, screening feature variables, optimizing model hyperparameters, and using bootstrap

methods to resample and expand the dataset may all overcome the problem of inaccurate prediction results caused by small sample sizes, improving model performance to a certain extent. The larger the amount of data, the wider the feature space can usually be covered, reducing the risk of underfitting and improving generalization ability by decreasing overfitting risk through data diversity. In addition, more data can help the model better separate noise and real signals, capture complex nonlinear relationships, support more complex model structures and accurate parameter optimization, thereby significantly improving the accuracy of predictions. When the data volume reaches a certain level (>100), the quality of the data is also quite significant. In case of merging different batches of data, it is necessary to remove batch effects from the dataset itself, unify variables, and perform preprocessing such as noise filtering.

Looking ahead, AI-guided mRNA-LNP systems show great promise for clinical translation, but also encounter numerous challenges. Although AI has demonstrated significant potential in preclinical optimization, it remains difficult to extrapolate its predictions to the human system due to inter-species differences and the lack of highly relevant human data. In order to enhance the applicability of AI models in the clinic, future research should strengthen data sharing to address the problem of high-quality data shortage. Meanwhile, the feasibility of AI-guided nanomedicine application in clinical practice can be effectively accelerated by integrating a series of diversified and high-quality data, including organoid models derived from patients, pharmacogenomics, transcriptomics, protein interactions, and actual clinical data. In addition, the acceptance of AI-assisted drug design also depends on the transparency, reproducibility, and rigorous experimental validation of its predictive models. Close collaboration among computer scientists, drug design researchers, and clinicians will be critical to overcoming these obstacles. By addressing these challenges, AI will not only accelerate the drug development process but also promote the advancement of a new generation of precise nanomedicine, ultimately improving the patient health.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Nos. LZ23H180003 and LMS25H280002), the National Natural Science Foundation of China (Nos. 32471375 and 22205201), and the Startup Package of Zhejiang University.

Author contributions

Kexin Su: Writing - original draft & editing, Conceptualization, Investigation. Junjie Qiu: Writing - original draft & editing, Conceptualization, Investigation. Tengfei Xu: Conceptualization, Resources, Investigation. Shuai Liu: Writing - review & editing, Conceptualization, Resources, Supervision, Funding acquisition, Investigation. All of the authors have read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

1. Parhiz H, Atochina-Vasserman EN, Weissman D. mRNA-based therapeutics: looking beyond COVID-19 vaccines. *Lancet* 2024;**403**: 1192–204.
2. Qin SG, Tang XS, Chen YT, Chen KP, Fan N, Xiao W, et al. mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Sig Transduct Target Ther* 2022;**7**:1–35.
3. Petsch B, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, et al. Protective efficacy of *in vitro* synthesized, specific mRNA vaccines against influenza A virus infection. *Nat Biotechnol* 2012;**30**:1210–6.
4. Zhang H, Zhang L, Lin A, Xu CC, Li ZY, Liu KB, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* 2023;**621**:396–403.
5. Akinc A, Maier MA, Manoharan M, Fitzgerald K, Jayaraman M, Barros S, et al. The Onpattro story and the clinical translation of nanomedicines containing nucleic acid-based drugs. *Nat Nanotechnol* 2019;**14**:1084–7.
6. Pardi N, Hogan MJ, Pelc RS, Muramatsu H, Andersen H, DeMaso CR, et al. Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. *Nature* 2017;**543**:248–51.
7. Xu LX, Shao ZX, Fang X, Xin ZF, Zhao SZ, Zhang HB, et al. Exploring precision treatments in immune-mediated inflammatory diseases: harnessing the infinite potential of nucleic acid delivery. *Explorations* 2025;**5**:20230165.
8. Tian YX, Li MJ, Yang Y, Li CH, Peng Y, Yang HY, et al. An MPXV mRNA-LNP vaccine candidate elicits protective immune responses against monkeypox virus. *Chin Chem Lett* 2024;**35**:109270.
9. Xiao W, Wang FZ, Gu YZ, He X, Fan N, Zheng Q, et al. PEG400-mediated nanocarriers improve the delivery and therapeutic efficiency of mRNA tumor vaccines. *Chin Chem Lett* 2024;**35**:108755.
10. Zhang ZQ, Wang L, Ma X, Wang H. Lentinan-functionalized PBAE-G-nanodiamonds as an adjuvant to induce cGAS–STING pathway-mediated macrophage activation and immune enhancement. *J Pharm Anal* 2024;**14**:100922.
11. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med* 2022;**28**:31–8.
12. Deo RC. Machine learning in medicine. *Circulation* 2015;**132**: 1920–30.
13. Das S, Mazumdar H, Khondakar KR, Kaushik A. Machine learning integrated graphene oxide-based diagnostics, drug delivery, analytical approaches to empower cancer diagnosis. *BMEMat* 2024;**e12117**.
14. Hong YF, Zhu SS, Liu YH, Tian C, Xu HQ, Chen GX, et al. The integration of machine learning into traditional Chinese medicine. *J Pharm Anal* 2024;**14**:101157.
15. Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;**542**:115–8.
16. Rao L, Yuan Y, Shen X, Yu GC, Chen XY. Designing nanotheranostics with machine learning. *Nat Nanotechnol* 2024;**19**: 1769–81.
17. Wang XX, Hao XR, Yan J, Xu J, Hu DD, Ji FF, et al. Urine biomarkers discovery by metabolomics and machine learning for Parkinson's disease diagnoses. *Chin Chem Lett* 2023;**34**:108230.
18. You C, Shen YY, Sun SY, Zhou JY, Li JW, Su GH, et al. Artificial intelligence in breast imaging: current situation and clinical challenges. *Explorations* 2023;**3**:20230007.
19. Paunovska K, Loughrey D, Dahlman JE. Drug delivery systems for RNA therapeutics. *Nat Rev Genet* 2022;**23**:265–80.
20. Hou XC, Zaks T, Langer R, Dong YZ. Lipid nanoparticles for mRNA delivery. *Nat Rev Mater* 2021;**6**:1078–94.
21. Yim EY, Zhou AC, Yim YC, Wang X, Xia T. Antigen-specific mRNA lipid nanoparticle platforms for the prevention and treatment of allergy and autoimmune diseases. *BMEMat* 2024;**2**:e12060.
22. Wilson E, Goswami J, Baqui AH, Doreski PA, Perez-Marc G, Zaman K, et al. Efficacy and safety of an mRNA-based RSV PreF vaccine in older adults. *N Engl J Med* 2023;**389**:2233–44.

23. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *N Engl J Med* 2020;**383**:2603–15.
24. Walsh EE, Frenck RW, Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. Safety and immunogenicity of two RNA-based covid-19 vaccine candidates. *N Engl J Med* 2020;**383**:2439–50.
25. Baden LR, Sahly HME, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021;**384**:403–16.
26. Lin LX, Su KX, Zhang XY, Shi L, Yan XX, Fu QL, et al. A versatile strategy to transform cationic polymers for efficient and organ-selective mRNA delivery. *Angew Chem Int Ed* 2025;**64**:e202500306.
27. Eygeris Y, Gupta M, Kim J, Sahay G. Chemistry of lipid nanoparticles for RNA delivery. *Acc Chem Res* 2022;**55**:2–12.
28. O'Brien Laramy M, Foley DA, Pak RH, Lewis JA, McKinney E, Egan PM, et al. Chemistry, manufacturing and controls strategies for using novel excipients in lipid nanoparticles. *Nat Nanotechnol* 2025;**20**:1–14.
29. Zhang YR, Zhang XY, Gao YS, Liu S. *Principles of lipid nanoparticle design for mRNA delivery*. BMEMat; 2024. p. e12116.
30. Zhao S, Gao KW, Han HS, Stenzel M, Yin BY, Song HY, et al. Acid-degradable lipid nanoparticles enhance the delivery of mRNA. *Nat Nanotechnol* 2024;**19**:1702–11.
31. Swingle KL, Hamilton AG, Safford HC, Geisler HC, Thatte AS, Palanki R, et al. Placenta-tropic VEGF mRNA lipid nanoparticles ameliorate murine pre-eclampsia. *Nature* 2025;**637**:412–21.
32. Han XX, Zhang HW, Butowska K, Swingle KL, Alameh MG, Weissman D, et al. An ionizable lipid toolbox for RNA delivery. *Nat Commun* 2021;**12**:7233.
33. Ma YT, Li SY, Lin X, Chen YP. A perspective of lipid nanoparticles for RNA delivery. *Explorations* 2024;**4**:20230147.
34. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015;**349**:255–60.
35. Ding DY, Zhang YH, Jia Y, Sun JZ. Machine learning-guided lipid nanoparticle design for mRNA delivery. In: *The 2023 ICML Workshop on Computational Biology*; 2023. Available from: <https://doi.org/10.48550/arXiv.2308.01402>.
36. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;**23**:40–55.
37. Bileschi ML, Belanger D, Bryant DH, Sanderson T, Carter B, Sculley D, et al. Using deep learning to annotate the protein universe. *Nat Biotechnol* 2022;**40**:932–7.
38. Bae SH, Choi H, Lee JS, Kang MH, Ahn SH, Lee YS, et al. Rational design of lipid nanoparticles for enhanced mRNA vaccine delivery via machine learning. *Small* 2024;**21**:2405618.
39. Li BW, Raji IO, Gordon AGR, Sun LZ, Raimondo TM, Oladimeji FA, et al. Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry. *Nat Mater* 2024;**23**:1002–8.
40. Xu Y, Ma SH, Cui HT, Chen JG, Xu SF, Gong FL, et al. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. *Nat Commun* 2024;**15**:6305.
41. Zhang K, Yang X, Wang YF, Yu YF, Huang N, Li G, et al. Artificial intelligence in drug development. *Nat Med* 2025;**31**:45–59.
42. Castillo-Hair SM, Seelig G. Machine learning for designing next-generation mRNA therapeutics. *Acc Chem Res* 2022;**55**:24–34.
43. Bravi B. Development and use of machine learning algorithms in vaccine target selection. *npj Vaccines* 2024;**9**:1–14.
44. Barbier AJ, Jiang AY, Zhang P, Wooster R, Anderson DG. The clinical progress of mRNA vaccines and immunotherapies. *Nat Biotechnol* 2022;**40**:840–54.
45. Kiaie SH, Majidi Zolbanin N, Ahmadi A, Bagherifar R, Valizadeh H, Kashanchi F, et al. Recent advances in mRNA-LNP therapeutics: immunological and pharmacological aspects. *J Nanobiotechnology* 2022;**20**:276.
46. Cheng Q, Wei T, Farbiak L, Johnson LT, Dilliard SA, Siegwart DJ. Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nat Nanotechnol* 2020;**15**:313–20.
47. Dilliard SA, Cheng Q, Siegwart DJ. On the mechanism of tissue-specific mRNA delivery by selective organ targeting nanoparticles. *Proc Natl Acad Sci U S A* 2021;**118**:e2109256118.
48. Wu SQ, Su KX, Yan XX, Shi L, Lin LX, Ren E, et al. Paracyclopentane-based ionizable lipids for efficient mRNA delivery in vivo. *J Control Release* 2024;**376**:395–401.
49. Guo S, Li K, Hu B, Li CH, Zhang MJ, Hussain A, et al. Membrane-destabilizing ionizable lipid empowered imaging-guided siRNA delivery and cancer treatment. *Explorations* 2021;**1**:35–49.
50. Qiu M, Tang Y, Chen JJ, Murip R, Ye ZF, Huang CF, et al. Lung-selective mRNA delivery of synthetic lipid nanoparticles for the treatment of pulmonary lymphangioleiomyomatosis. *Proc Natl Acad Sci U S A* 2022;**119**:e2116271119.
51. Lin LX, Su KX, Cheng Q, Liu S. Targeting materials and strategies for RNA delivery. *Theranostics* 2023;**13**:4667–93.
52. Hald Albertsen C, Kulkarni JA, Witzigmann D, Lind M, Petersson K, Simonsen JB. The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Adv Drug Deliv Rev* 2022;**188**:114416.
53. Hassett KJ, Higgins J, Woods A, Levy B, Xia Y, Hsiao CJ, et al. Impact of lipid nanoparticle size on mRNA vaccine immunogenicity. *J Control Release* 2021;**335**:237–46.
54. Liu S, Wen YX, Shan XZ, Ma XH, Yang C, Cheng XD, et al. Charge-assisted stabilization of lipid nanoparticles enables inhaled mRNA delivery for mucosal vaccination. *Nat Commun* 2024;**15**:9471.
55. Chaudhary N, Kasiewicz LN, Newby AN, Arral ML, Yerneni SS, Melamed JR, et al. Amine headgroups in ionizable lipids drive immune responses to lipid nanoparticles by binding to the receptors TLR4 and CD1d. *Nat Biomed Eng* 2024;**8**:1483–98.
56. Yoon IC, Xue LL, Chen QY, Liu JY, Xu JC, Siddiqui Z, et al. Piperazine-derived bisphosphonate-based ionizable lipid nanoparticles enhance mRNA delivery to the bone microenvironment. *Angew Chem Int Ed* 2025;**64**:e202415389.
57. Wu SQ, Lin LX, Shi L, Liu S. An overview of lipid constituents in lipid nanoparticle mRNA delivery systems. *Wires Nanomed Nanobi* 2024;**16**:e1978.
58. Sebastiani F, Yanez Arteta M, Lerche M, Porcar L, Lang C, Bragg RA, et al. Apolipoprotein E binding drives structural and compositional rearrangement of mRNA-containing lipid nanoparticles. *ACS Nano* 2021;**15**:6709–22.
59. Francia V, Schiffelers RM, Cullis PR, Witzigmann D. The biomolecular corona of lipid nanoparticles for gene therapy. *Bioconjug Chem* 2020;**31**:2046–59.
60. Gilleron J, Querbes W, Zeigerer A, Borodovsky A, Marsico G, Schubert U, et al. Image-based analysis of lipid nanoparticle-mediated siRNA delivery, intracellular trafficking and endosomal escape. *Nat Biotechnol* 2013;**31**:638–46.
61. Wang Q, Jiang QK, Li D, Yang ZM, Gao L, Liu F, et al. Elaborately engineering of lipid nanoparticle for targeting delivery of siRNA and suppressing acute liver injury. *Chin Chem Lett* 2024;**35**:108683.
62. Zhang DP, Atochina-Vasserman EN, Maurya DS, Huang N, Xiao Q, Ona N, et al. One-component multifunctional sequence-defined ionizable amphiphilic Janus dendrimer delivery systems for mRNA. *J Am Chem Soc* 2021;**143**:12315–27.
63. Luo ZJ, Lin Y, Meng YN, Li MY, Ren HY, Shi HP, et al. Spleen-targeted mRNA vaccine doped with manganese adjuvant for robust anticancer immunity in vivo. *ACS Nano* 2024;**18**:30701–15.
64. Gautam M, Jozic A, Su GLN, Herrera-Barrera M, Curtis A, Arrizabalaga S, et al. Lipid nanoparticles with PEG-variant surface modifications mediate genome editing in the mouse retina. *Nat Commun* 2023;**14**:6468.
65. Witten J, Raji I, Manan RS, Beyer E, Bartlett S, Tang YH, et al. Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy. *Nat Biotechnol* 2025;**43**:1790–9.
66. Wang W, Chen KP, Jiang T, Wu YY, Wu Z, Ying H, et al. Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery. *Nat Commun* 2024;**15**:10804.

67. Saptarshi SR, Duschl A, Lopata AL. Interaction of nanoparticles with proteins: relation to bio-reactivity of the nanoparticle. *J Nanobiotechnology* 2013;**11**:26.
68. Li YM, Ye ZF, Yang HY, Xu QB. Tailoring combinatorial lipid nanoparticles for intracellular delivery of nucleic acids, proteins, and drugs. *Acta Pharm Sin B* 2022;**12**:2624–39.
69. Yu MY, Lin LX, Zhou DZ, Liu S. Interaction design in mRNA delivery systems. *J Control Release* 2025;**377**:413–26.
70. Maugeri M, Nawaz M, Papadimitriou A, Angerfors A, Camponeschi A, Na ML, et al. Linkage between endosomal escape of LNP-mRNA and loading into EVs for transport to other cells. *Nat Commun* 2019;**10**:4333.
71. Chen XX, Wang XM, Zhang K, Fung KM, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;**79**:102444.
72. Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. *Nat Methods* 2019;**16**:1233–46.
73. Liu G, Catacutan DB, Rathod K, Swanson K, Jin WG, Mohammed JC, et al. Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat Chem Biol* 2023;**19**:1342–50.
74. Wong F, Zheng EJ, Valeri JA, Donghia NM, Anahtar MN, Omori S, et al. Discovery of a structural class of antibiotics with explainable deep learning. *Nature* 2024;**626**:177–85.
75. Wong D, Yip S. Machine learning classifies cancer. *Nature* 2018;**555**:446–7.
76. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol* 1943;**5**:115–33.
77. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;**86**:2278–324.
78. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;**18**:1527–54.
79. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 2018;**40**:834–48.
80. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv* 2002;**34**:1–47.
81. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 2017;**28**:2222–32.
82. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature* 2021;**596**:590–6.
83. Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod Pathol* 2022;**35**:23–32.
84. Boehnke N, Straehla JP, Safford HC, Kocak M, Rees MG, Ronan M, et al. Massively parallel pooled screening reveals genomic determinants of nanoparticle delivery. *Science* 2022;**377**:eabm5551.
85. Pavlíček N, Gross L. Generation, manipulation and characterization of molecules by atomic force microscopy. *Nat Rev Chem* 2017;**1**:1–11.
86. Taniguchi N. On the basic concept of “Nano-technology”. *Pro Intl Conf Prod Eng Tokyo Part II* 1974:18–23.
87. Barenholz Y. Chezy). Doxil—The first FDA-approved nano-drug: lessons learned. *J Control Release* 2012;**160**:117–34.
88. Ebrahimi SB, Samanta D, Kusmierz CD, Mirkin CA. Protein transfection via spherical nucleic acids. *Nat Protoc* 2022;**17**:327–57.
89. Ban Z, Yuan P, Yu FB, Peng T, Zhou QX, Hu XG. Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proc Natl Acad Sci U S A* 2020;**117**:10492–9.
90. Ng WL, Goh GL, Ten JSJ, Yeong WY. Progress and opportunities for machine learning in materials and processes of additive manufacturing. *Adv Mater* 2024;**36**:2310006.
91. Jain AK, Duin RPW, Mao JC. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 2000;**22**:4–37.
92. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet* 2015;**16**:321–32.
93. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 2013;**35**:1798–828.
94. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;**61**:85–117.
95. Cheplygina V, de Bruijne M, Pluim JPW. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal* 2019;**54**:280–96.
96. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comp Struct Biotechnol J* 2015;**13**:8–17.
97. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med* 2019;**25**:24–9.
98. Maharjan R, Hada S, Lee JE, Han HK, Kim KH, Seo HJ, et al. Comparative study of lipid nanoparticle-based mRNA vaccine bioprocess with machine learning and combinatorial artificial neural network-design of experiment approach. *Int J Pharm* 2023;**640**:123012.
99. Zhang ZJ, Lin JJ, Owens G, Chen ZL. Deciphering silver nanoparticles perturbation effects and risks for soil enzymes worldwide: insights from machine learning and soil property integration. *J Hazard Mater* 2024;**469**:134052.
100. Tsuyama Y, Mawatari KZM. Nanofluidic detection platform for simultaneous light absorption and scattering measurement of individual nanoparticles in flow. *Anal Chem* 2024;**96**:11430–8.
101. Ting JYC, Parker AJ, Barnard AS. Data-driven design of classes of ruthenium nanoparticles using multitarget Bayesian inference. *Chem Mater* 2023;**35**:728–38.
102. Yu FB, Wei CH, Deng P, Peng T, Hu XG. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Sci Adv* 2021;**7**:eabf4130.
103. Abostait A, Abdelkarim M, Bao ZQ, Miyake Y, Tse WH, Di Ciano-Oliveir C, et al. Optimizing lipid nanoparticles for fetal gene delivery *in vitro*, *ex vivo*, and aided with machine learning. *J Control Release* 2024;**376**:678–700.
104. Kar S, Pathakoti K, Tchounwou PB, Leszczynska D, Leszczynski J. Evaluating the cytotoxicity of a large pool of metal oxide nanoparticles to *Escherichia coli*: mechanistic understanding through *in vitro* and *in silico* studies. *Chemosphere* 2021;**264**:128428.
105. Horev-Azaria L, Baldi G, Beno DLL, Bonacchi D, Golla-Schindler U, Kirkpatrick JC, et al. Predictive toxicology of cobalt ferrite nanoparticles: comparative *in vitro* study of different cellular models using methods of knowledge discovery from data. *Part Fibre Toxicol* 2013;**10**:32.
106. Mejia-Mendez JL, Reza-Zaldívar EE, Sanchez-Martinez A, Ceballos-Sanchez O, Navarro-López DE, Marcelo Lozano L, et al. Exploring the cytotoxic and antioxidant properties of lanthanide-doped ZnO nanoparticles: a study with machine learning interpretation. *J Nanobiotechnology* 2024;**22**:687.
107. Keating ME, Bonnier F, Byrne HJ. Spectral cross-correlation as a supervised approach for the analysis of complex Raman datasets: the case of nanoparticles in biological cells. *Analyst* 2012;**137**:5792–802.
108. Saeedimase M, Rahmani R, Lyubartsev AP. Biomolecular adsorption on nanomaterials: combining molecular simulations with machine learning. *J Chem Inf Model* 2024;**64**:3799–811.
109. Bakry M, Haddar H, Bunău O. A robust expectation-maximization method for the interpretation of small-angle scattering data from dense nanoparticle samples. *J Appl Cryst* 2019;**52**:926–36.
110. Yao YC, Tseng ST, Wong DSH. Shelf-life prediction of nano-sol via pH acceleration. *J Qual Technol* 2017;**49**:46–63.
111. Rathee S, Ojha A, Sagar P, Upadhyay A, Rather IA, Shukla S. Decoration of Fe₃O₄-vitamin C nanoparticles on alginate-chitosan nanocomplex: characterization, safety, bioaccessibility boost and

- Iron Nanofortification in A2 goat milk gels. *Food Chem* 2025;**470**: 142711.
112. Ziatdinov MA, Yaman MY, Liu YT, Ginger D, Kalinin SV. Semi-supervised learning of images with strong rotational disorder: assembling nanoparticle libraries. *Digit Discov* 2024;**3**:1213–20.
113. Lutz ID, Wang SZ, Norn C, Courbet A, Borst AJ, Zhao YT, et al. Top-down design of protein architectures with reinforcement learning. *Science* 2023;**380**:266–73.
114. Jyakhwo S, Serov N, Dmitrenko A, Vinogradov VV. Machine learning reinforced genetic algorithm for massive targeted discovery of selectively cytotoxic inorganic nanoparticles. *Small* 2024;**20**: 2305375.
115. Ho T. In: *Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1. Canada: IEEE Comput. Soc. Press; 1995. p. 278–82. Montreal, Que.
116. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
117. Belgiu M, Drăguț L. Random forest in remote sensing: a review of applications and future directions. *Isprs J Photogramm* 2016;**114**: 24–31.
118. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl* 2019;**134**:93–101.
119. Tse WH, Higgins S, Patel D, Xing M, West AR, Labouta HI, et al. The maternal-fetal transfer of passive immunity as a mechanism of transplacental nanoparticle drug delivery for prenatal therapies. *Biomater Sci* 2022;**10**:5243–53.
120. Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag* 1901;**2**:559–72.
121. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 1933;**24**:417–41.
122. Taheri S, Andrade JC de, Conte-Junior CA. Emerging perspectives on analytical techniques and machine learning for food metabolomics in the era of industry 4.0: a systematic review. *Crit Rev Food Sci Nutr* 2024;**1**–27.
123. Romanò S, Angelillo A, Cimmino W, Iaccarino N, Nele V, Campani V, et al. An electrochemical strip to evaluate and to discriminate drug encapsulation in lipid nanovectors. *Anal Chem* 2024;**96**:11651–6.
124. Mitsutake H, Castro SR, de Paula E, Poppi RJ, Rutledge DN, Breitzkreitz MC. Comparison of different chemometric methods to extract chemical and physical information from Raman images of homogeneous and heterogeneous semi-solid pharmaceutical formulations. *Int J Pharm* 2018;**552**:119–29.
125. Greive SJ, Bacri L, Cressiot B, Pelta J. Identification of conformational variants for bradykinin biomarker peptides from a biofluid using a nanopore and machine learning. *ACS Nano* 2024;**18**:539–50.
126. Allen TM, Cullis PR. Liposomal drug delivery systems: from concept to clinical applications. *Adv Drug Deliv Rev* 2013;**65**:36–48.
127. Maharjan R, Kim KH, Lee K, Han H-K, Jeong SH. Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *J Pharm Anal* 2024;**14**:100996.
128. Yu HJ, Tang SY, Li SFY, Cheng F. Averaging strategy for interpretable machine learning on small datasets to understand element uptake after seed nanotreatment. *Environ Sci Technol* 2023;**57**: 12760–70.
129. Gandek TB, van der Koog L, Nagelkerke A. A comparison of cellular uptake mechanisms, delivery efficacy, and intracellular fate between liposomes and extracellular vesicles. *Adv Healthc Mater* 2023;**12**: 2300319.
130. Roy A, Chakraborty S. Support vector machine in structural reliability analysis: a review. *Reliab Eng Syst Safe* 2023;**233**:109126.
131. Mandal S, Paul D, Saha S, Das P. Multi-layer perceptron for detection of different class antibiotics from visual fluorescence response of a carbon nanoparticle-based multichannel array sensor. *Sensor Actuat B-Chem* 2022;**360**:131660.
132. Mandal S, Paul D, Saha S, Das P. Deep learning assisted detection of toxic heavy metal ions based on visual fluorescence responses from a carbon nanoparticle array. *Environ Sci Nano* 2022;**9**:2596–606.
133. Zheng ZY, Tang YX. Reinforcement learning-aided design of ROS-responsive polymeric nanocarriers for hepatocellular carcinoma treatment. *J Mol Struct* 2025;**1329**:141446.
134. Wu SQ, Shi L, Su KX, Lin LX, Yan XX, Zhang YR, et al. Carbonate-bearing ionizable lipids for mRNA delivery to splenic NK cells. *J Am Chem Soc* 2025;**147**:28665–73.
135. Han XX, Alameh MG, Xu Y, Palanki R, El-Mayta R, Dwivedi G, et al. Optimization of the activity and biodegradability of ionizable lipids for mRNA delivery via directed chemical evolution. *Nat Biomed Eng* 2024;**8**:1412–24.
136. He S, Liu S. Zwitterionic materials for nucleic acid delivery and therapeutic applications. *J Control Release* 2024;**365**:919–35.
137. Liu S, Cheng Q, Wei T, Yu XL, Johnson LT, Farbiak L, et al. Membrane-destabilizing ionizable phospholipids for organ-selective mRNA delivery and CRISPR–Cas gene editing. *Nat Mater* 2021;**20**:701–10.
138. Zhang YB, Sun CZ, Wang C, Jankovic KE, Dong YZ. Lipids and lipid derivatives for RNA delivery. *Chem Rev* 2021;**121**:12181–277.
139. Patel S, Ashwanikumar N, Robinson E, Xia Y, Mihai C, Griffith JP, et al. Naturally-occurring cholesterol analogues in lipid nanoparticles induce polymorphic shape and enhance intracellular delivery of mRNA. *Nat Commun* 2020;**11**:983.
140. Paunovska K, Da Silva Sanchez AJ, Sago CD, Gan ZB, Lokugamage MP, Islam FZ, et al. Nanoparticles containing oxidized cholesterol deliver mRNA to the liver microenvironment at clinically relevant doses. *Adv Mater* 2019;**31**:e1807748.
141. Bogaert B, Debisschop A, Ehouarne T, Eeckhoutte HPV, Volder JD, Jacobs A, et al. Selective replacement of cholesterol with cationic amphiphilic drugs enables the design of lipid nanoparticles with improved RNA delivery. *Nano Lett* 2024;**24**:2961–71.
142. Mui BL, Tam YK, Jayaraman M, Ansell SM, Du XY, Tam YYC, et al. Influence of polyethylene glycol lipid desorption rates on pharmacokinetics and pharmacodynamics of siRNA lipid nanoparticles. *Mol Ther Nucleic Acids* 2013;**2**:e139.
143. Yuan Y, Wu YQ, Cheng JB, Yang KY, Xia YL, Wu HG, et al. Applications of artificial intelligence to lipid nanoparticle delivery. *Particuology* 2024;**90**:88–97.
144. Dorsey PJ, Lau CL, Chang T, Doerschuk PC, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. *J Pharm Sci* 2024;**113**:3413–33.
145. Tao HC, Wu TY, Aldeghi M, Wu TC, Aspuru-Guzik A, Kumacheva E. Nanoparticle synthesis assisted by machine learning. *Nat Rev Mater* 2021;**6**:701–16.
146. Sun C, Wang Y, Tang QS, Yang D, Hou L, Cao Z, et al. Tumor microenvironment activated and amplified self-luminescent nano photosensitizers for efficient treatment of triple negative breast cancer. *Chem Eng J* 2024;**494**:152400.
147. Mills JA, Liu FF, Jarrett TR, Fletcher NL, Thurecht KJ. Nanoparticle based medicines: approaches for evading and manipulating the mononuclear phagocyte system and potential for clinical translation. *Biomater Sci* 2022;**10**:3029–53.
148. Liu K, Nilsson R, Lázaro-Ibáñez E, Duàn H, Miliotis T, Strimfors M, et al. Multiomics analysis of naturally efficacious lipid nanoparticle coronas reveals high-density lipoprotein is necessary for their function. *Nat Commun* 2023;**14**:4007.
149. Iscaro J, Yu HT, Martinez N, Subramaniam S, Joyce P, Wang H, et al. Lyotropic liquid crystalline phase nanostructure and cholesterol enhance lipid nanoparticle mediated mRNA transfection in macrophages. *Adv Funct Mater* 2024;**34**:2405286.
150. Hunter MR, Cui LL, Porebski BT, Pereira S, Sonzini S, Odunze U, et al. Understanding intracellular biology to improve mRNA delivery by lipid nanoparticles. *Small Methods* 2023;**7**:2201695.
151. Lynch I, Dawson KA. Protein–nanoparticle interactions. *Nano Today* 2008;**3**:40–7.
152. Wang W, Feng S, Ye ZYF, Gao HL, Lin JZ, Ouyang DF. Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm. *Acta Pharm Sin B* 2022;**12**:2950–62.
153. Fu XH, Yang C, Su YY, Liu CL, Qiu HY, Yu YY, et al. Machine learning enables comprehensive prediction of the relative protein

- abundance of multiple proteins on the protein corona. *Research* 2024; 7:487.
154. Su KX, Shi L, Sheng T, Yan XX, Lin LX, Meng CY, et al. Reformulating lipid nanoparticles for organ-targeted mRNA accumulation and translation. *Nat Commun* 2024;15:5659.
 155. Wang YY, Wang JR, Cao ZL, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell* 2022;4:279–87.
 156. Jia WL, Wang H, Chen MH, Lu DH, Lin L, Car R, et al. Pushing the limit of molecular dynamics with *Ab initio* accuracy to 100 million atoms with machine learning. *SC20: international conference for high performance computing, networking, storage and analysis*. 2020. p. 1–14.
 157. Gao HS, Kan S, Ye ZYF, Feng YC, Jin L, Zhang XD, et al. Development of *in silico* methodology for siRNA lipid nanoparticle formulations. *Chem Eng J* 2022;442:136310.
 158. Alameh MG, Tombácz I, Bettini E, Lederer K, Ndeupen S, Sittplangkoon C, et al. Lipid nanoparticles enhance the efficacy of mRNA and protein subunit vaccines by inducing robust T follicular helper cell and humoral responses. *Immunity* 2021;54:2877–92.e7.
 159. Han XX, Alameh MG, Butowska K, Knox JJ, Lundgreen K, Ghattas M, et al. Adjuvant lipidoid-substituted lipid nanoparticles augment the immunogenicity of SARS-CoV-2 mRNA vaccines. *Nat Nanotechnol* 2023;18:1105–14.
 160. Parhiz H, Brenner JS, Patel PN, Papp TE, Shah Nawaz H, Li Q, et al. Added to pre-existing inflammation, mRNA-lipid nanoparticles induce inflammation exacerbation (IE). *J Control Release* 2022;344: 50–61.
 161. Lin M, Yin YF, Zhao XM, Wang C, Zhu XQ, Zhan LT, et al. A truncated pre-F protein mRNA vaccine elicits an enhanced immune response and protection against respiratory syncytial virus. *Nat Commun* 2025;16:1386.
 162. Kawai A, Shimizu T, Tanaka H, Shichinohe S, Anindita J, Hirose M, et al. Low-inflammatory lipid nanoparticle-based mRNA vaccine elicits protective immunity against H5N1 influenza virus with reduced adverse reactions. *Mol Ther* 2025;33:529–47.
 163. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019; 16:391–403.
 164. Gao Q, Yang LY, Lu MJ, Jin RJ, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. *J Hematol Oncol* 2023;16:55.
 165. Zhu YN, Ma JY, Shen RC, Lin JH, Li SY, Lu XY, et al. Screening for lipid nanoparticles that modulate the immune activity of helper T cells towards enhanced antitumour activity. *Nat Biomed Eng* 2024;8: 544–60.
 166. Carrasco-Zanini J, Pietzner M, Koprulu M, Wheeler E, Kerrison ND, Wareham NJ, et al. Proteomic prediction of diverse incident diseases: a machine learning-guided biomarker discovery study using data from a prospective cohort study. *Lancet Digit Health* 2024;6:e470–9.
 167. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Bio-technol Adv* 2021;49:107739.
 168. Zhang N, Zhang H, Wu WT, Zhou R, Li SY, Wang ZY, et al. Machine learning-based identification of tumor-infiltrating immune cell-associated lncRNAs for improving outcomes and immunotherapy responses in patients with low-grade glioma. *Theranostics* 2022;12: 5931–48.
 169. Wang YF, Xiao ZY, Wang ZK, Lee DEY, Ma YF, Wilhelm S, et al. Multi-omics approaches to decipher the interactions of nanoparticles and biological systems. *Nat Rev Bioeng* 2024;3:333–48.
 170. Kuypers S, Smisdom N, Pintelon I, Timmermans JP, Ameloot M, Michiels L, et al. Unsupervised machine learning-based clustering of nanosized fluorescent extracellular vesicles. *Small* 2021;17:2006786.
 171. Yang C, Gong CD, Zhang ZX, Fang JJ, Li WH, Liu GX, et al. *In silico* prediction of pKa values using explainable deep learning methods. *J Pharm Anal* 2024;15:101174.
 172. Park S, Yoon SE, Song Y, Tian CY, Baek C, Cho H, et al. A simple approach to biophysical profiling of blood cells in extranodal NK/T cell lymphoma patients using deep learning-integrated image cytometry. *BMEMat* 2024;3:e12128.
 173. Kavungal D, Magalhães P, Kumar ST, Kolla R, Lashuel HA, Altug H. Artificial intelligence-coupled plasmonic infrared sensor for detection of structural protein biomarkers in neurodegenerative diseases. *Sci Adv* 2023;9:eadg9644.
 174. Taniguchi MSTR, Minami S, Ono C, Hamajima RN, Morimura A, Hamaguchi S, et al. Combining machine learning and nanopore construction creates an artificial intelligence nanopore for coronavirus detection. *Nat Commun* 2021;12:3726.
 175. Batra R, Loeffler TD, Chan H, Srinivasan S, Cui HG, Korendovych IV, et al. Machine learning overcomes human bias in the discovery of self-assembling peptides. *Nat Chem* 2022;14: 1427–35.
 176. Zhu MS, Zhuang J, Li Z, Liu QQ, Zhao RP, Gao ZX, et al. Machine-learning-assisted single-vessel analysis of nanoparticle permeability in tumour vasculatures. *Nat Nanotechnol* 2023;18:657–66.
 177. Sun YH, Chatterjee S, Lian XZ, Traylor Z, Sattiraju SR, Xiao YF, et al. *In vivo* editing of lung stem cells for durable gene correction in mice. *Science* 2024;384:1196–202.
 178. Ahn J, Ko J, Lee S, Yu J, Kim YT, Jeon N. Microfluidics in nanoparticle drug delivery; from synthesis to pre-clinical screening. *Adv Drug Deliv Rev* 2018;128:29–53.