# Machine Learning Analysis of Biotherapeutics

**Technology ID**
2024-ARDE-70506

**Category**
Pharmaceuticals/Computational
Drug Delivery & Nanomedicine

**Authors**
Arezoo Ardekani
Gaurav Kumar

**View online**

Researchers at Purdue University have developed a machine learning (ML) framework to predict the activity and cell viability of lipid nanoparticles (LNPs) for nucleic acid delivery. A typical approach to formulate LNPs is establishing a quantitative structure-activity relationship (QSAR) between their compositions and in vitro/in vivo activities, which allows for the prediction of activity based on molecular structure. However, developing QSAR for LNPs can be challenging due to the complexity of multi-component formulations, interactions with biological membranes, and stability in physiological environments. To address these challenges, researchers at Purdue developed this ML approach to predict organ specific drug activity and toxicity from the chemical structure of LNP constituents, enabling the pharmaceutical industry to quickly handle large amounts of complex data and better understand the latent relationship between molecular features of LNPs and their biological functions. With this technology, new LNPs can be more efficiently developed to provide life-saving medical technology.

**Technology Validation:**

Researchers curated data from 6,398 LNP formulations in the literature, applied nine featurization techniques to extract chemical information, and trained five machine learning models for binary and multiclass classification. Their binary models achieved over 90% accuracy, while the multiclass models reached over 95% accuracy. Their results demonstrated that molecular descriptors, particularly when used with random forest and gradient boosting models, provided the most accurate predictions. Their findings also emphasized the need for large training datasets and comprehensive LNP composition details, such as constituent structures, molar ratios, nucleic acid types, and dosages, to enhance predictive performance.

**Advantages**:

-Handle large amounts of complex data

-More rapidly and efficiently develop new therapies

-Promotes better understanding of existing LNPs

-High prediction accuracy

**Applications**:

-Research and development of lipid nanoparticles (LNPs), a vital medical technology for gene therapies, mRNA vaccines, and more.

Publications:

G Kumar, AM Ardekani, "Machine learning framework to predict the performance of lipid nanoparticles for nucleic acid delivery," arXiv preprint arXiv:2411.14293, 2024. https://arxiv.org/abs/2411.14293.

**TRL:** 4

**Intellectual Property:**

Provisional-Patent, 2024-11-18, United States

Provisional-Patent, 2025-04-21, United States

**Keywords:** Lipid nanoparticle prediction,Machine learning drug delivery,QSAR for LNPs,Nucleic acid delivery modeling,AI-powered LNP screening,Organ-specific toxicity prediction,High-throughput nanoparticle design,Pharmaceutical ML platform,Gene therapy formulation tool,mRNA vaccine delivery optimization