Archives of Pharmaceutical Sciences Ain Shams University 2025; Vol. 9(1): 209-222

OPEN ACCESS

Pharmaceutics and Industrial Pharmacy

Editorial

Artificial Intelligence in Drug Formulation and Delivery: Benefits, Trends, and Future Perspectives

Anupama Singh[a*], Vikas A. Saharan[b], Rania M. Hathout[c]

[a]*School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Balawala, Dehradun 248001, India*

[b]*Department of Pharmaceutical Technology (Formulations), National Institute of Pharmaceutical Education and Research (NIPER) Guwahati, Sila Katamur (Halugurisuk), P.O. Changsari 781101, District Kamrup, Assam, India*

[c]*Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, Cairo, 11566, Egypt*

ABSTRACT

Computational optimization and its integration with AI in formulation optimization, predictive modeling, and drug-excipient interactions have significantly reduced the conventional trial-and-error approaches. This editorial highlights the integration of artificial intelligence (AI) in optimization and predictive analysis in formulation development and delivery. Some other advancements such as those emerging at the interface of human-AI interaction are also briefly discussed with a focus on advancements in the last five years. Extensive data from the characterization of early formulations like cocrystals, solid dispersions, and drug-excipient complexes have led to prediction tools with the help of supervised machine learning algorithms. Robotic innovations have led to the automation of manufacturing operations which are predictive, self-optimizing, and self-correcting. Ligand-receptor interactions are now analyzed and predicted more effectively with the help of AI algorithms bringing rationality to innovations in drug delivery. All forms of AI like machine learning and deep learning are contributing at each step of the pharmaceutical drug discovery and development pathway.

Keywords: *DoE, optimization, artificial intelligence, machine learning, preformulation, drug product, simulation, docking, pharmacokinetics.*

***Correspondence** | Anupama Singh; School of Pharmaceutical Sciences and Technology, Sardar Bhagwan Singh University, Balawala, Dehradun 248001, India. Email: anupama.cognosy@gmail.com

Citation | Singh A, Saharan VA, Hathout, 2025. Artificial Intelligence in Drug Formulation and Delivery: Benefits, Trends, and Future Perspectives. Arch Pharm Sci ASU 9(1): 209-222

DOI: 10.21608/aps.2025.365472.1217

Print ISSN: 2356-8380. **Online ISSN**: 2356-8399.

Received 18 March 2025. **Accepted** 14 April 2025.

Published by: Ain Shams University, Faculty of Pharmacy

1. Introduction

The use of artificial intelligence (AI) in drug formulation, delivery, and targeting is gaining grounds every day. It includes the use of machine learning, design of experiments (DoE), molecular docking and simulations, and other beneficial software and informatics tools. DoE helps optimize processes and formulations by identifying relationships among various key factors and responses during drug product manufacturing, drug development, and discovery. Optimization increases the efficiency and robustness of the pharmaceutical process and improves the quality of drug products. However, the DoE approach becomes complex with the increase in the number of factors and when subjected to nonlinear relationships. Machine learning (ML) is particularly useful when analyzing datasets where the interactions between

multiple variables are intricate and are not well understood using conventional regression or rule-based models. Moreover, ML is specifically useful when the experiments or runs are not planned and do not follow specific patterns. When dealing with complex relationships between multiple variables that do not follow a simple linear pattern, ML techniques are more effective in studying these relationships than the traditional DoE approach with statistical methods of optimization. Additionally, ML techniques are versatile in analyzing various forms of information, such as numerical data, images, and text.

Therefore, examining diverse data types can be particularly utilized for almost any pharmaceutical process to predict the optimal formulation design, saving time, money, and resources. Supervised ML approaches are also helpful in predicting solubility, permeability, ADMET (Absorption-Distribution-Metabolism-Excretion-Toxicity) properties, and ligand-receptor interactions while unsupervised counterparts support clustering compounds, identifying formulation trends, and reducing feature dimensions. Post ChatGPT era, the DoE optimization software industry has actively integrated AI and ML to improve experimental design, analysis, and prediction. This integration has improved efficiency, reduced costs, and accelerated research and development across various sectors, including pharmaceuticals and materials science.

The combination of molecular docking with MD (molecular dynamics) simulations or ML has enhanced the understanding of protein-ligand interactions, leading to more accurate predictions of binding affinities and the dynamic behavior of drug molecules within biological systems on one hand and the interaction of targeting moieties and piloting molecules grafted on nanocarriers to their targets on the other hand. Significant advancements in pharmacokinetic predictions through mechanistic modeling have led to new approaches to drug development like model-informed drug development (MIDD)/model-informed formulation development (MIFD).

2. Benefits of machine learning methods in drug delivery

Machine learning methods are an integral part of artificial intelligence that utilizes certain programs and software in performing specific tasks such as classification, clustering and regression, and predictions. They usually work by training the machines (computers) through datasets and by applying certain algorithms the required task(s) can be performed. According to the utilized algorithms, the machine learning methods are categorized into two main categories, namely; the unsupervised and the supervised counterparts. The term “supervised” implies that the inputs (x-variables) are supervised by the outputs (responses). In other words, it means that the method algorithm deals with the inputs and the outputs at the same time and correlates between them. On the other hand, the unsupervised methods only deal with the inputs or the x-variables with no correlation with the outputs **[1]**. Examples of supervised machine learning methods include Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Gaussian Processes (GPs), and Partial Least Squares (PLS) while examples of un-supervised ones are Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) **[2]**.

In the last recent years, the use of machine learning methods in drug delivery, pharmaceutics, and drug formulation has gained some solid ground. The applications of these methods are becoming crucial in pharmaceutical and similar industries such as cosmetics, food, and beverages. At the head of these applications comes the specification of the most successful

pairs between the drugs or molecules and their nanocarriers **[3]**, pointing out the most successful formulations, detecting the most stable formulations, figuring out the absence or presence of drug-excipients incompatibilities and finally the successful prediction of drugs loading on their carriers. Examples of drug delivery applications using both supervised and unsupervised ML methods in literature will be provided in the upcoming subsections.

2.1. The benefits of artificial neural networks in drug delivery

Fig. 1. demonstrates the workflow of ANNs as a machine-learning method in drug loading prediction. The drugs or molecules are translated into numerical important molecular descriptors such as but not limited to: Xlog P, molecular weight, fragment complexity, number of H-bond acceptors, number of H-bond donors, molecular globularity, Weiner index and total polar surface area (tpsa). Then these descriptors are correlated with drug loading (DL) on a certain carrier as the response through ANNs so that a drug loading pattern is obtained so that when a new molecule is introduced, its drug loading can easily be predicted **[2]**.

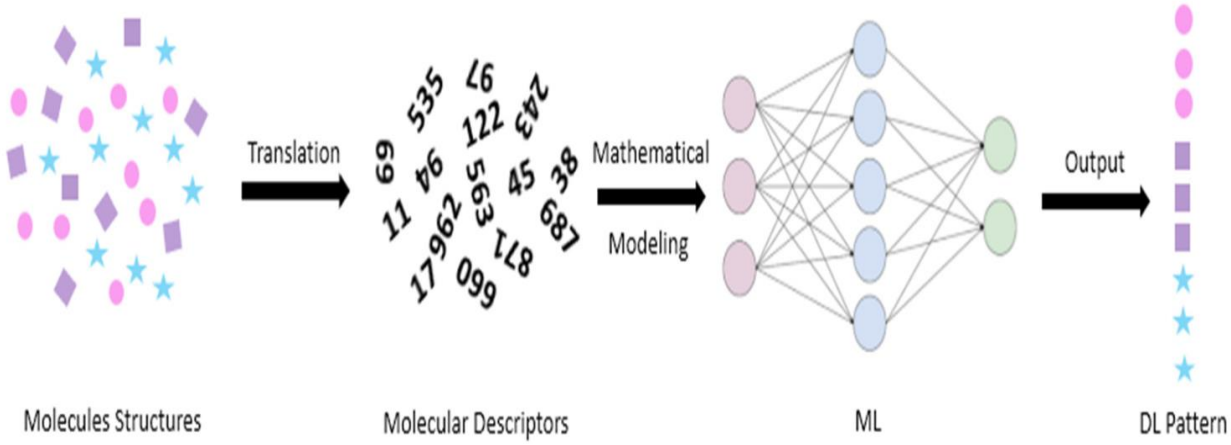

Fig. 1. Workflow of ANNs as a machine learning method in drug loading prediction. Reprinted from Abd-algaleel et al., 2021 **[4]** under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The artificial neural networks are inspired by the work of neurons in the neurological system where the concentration of the neurotransmitters that regulate the strength of the signal reaching the next neuron is replaced by a "weight" given for each input (in the above case it is one of the used descriptors). The network is usually built of neurons, hidden layers containing hidden nodes where the inputs are multiplied by varying weights aiming for an optimum correlation between the inputs and outputs. Usually, the summation of the products of inputs and their corresponding weights is subjected to an activation function (commonly a sigmoidal function) before correlation with the responses (outputs). The process aims to find a function that correlates the outputs with the inputs. Using a back-propagation method the weights are changed all over the neurons to reach the optimum model or function. After, this hidden model can be used to predict the outcomes of any newly entered inputs (regression and prediction) **[5]**.

Some studies have utilized AI results and correlated them with practical drug delivery data

such as drug loading as an output **[1, 6, 7]**.

2.2. The use of principal component analysis in realizing drug formulation stability

The principal component analysis is another important machine learning method of the unsupervised category. It is a very useful method in reducing the dimensionality of the data and extracting the most important directions (vectors) causing the highest variation (variance) in the available data. It yields significant and beneficial plots called score or scattering plots where the points (data) are scattered according to the main directions (principal components) that are usually two (for simplification and readability) in most drug delivery applications. These plots are important in visualizing the gathering of points in close proximities usually in the same quadrant of the plot **[8]**. This clustering technique is now applied in determining the most stable formulations which is an indispensable criterion in selecting formulations to proceed for scaling up in the pharmaceutical industry **[9]**. An example of exploiting this method in the aforementioned application is demonstrated in **Fig. 2a.**

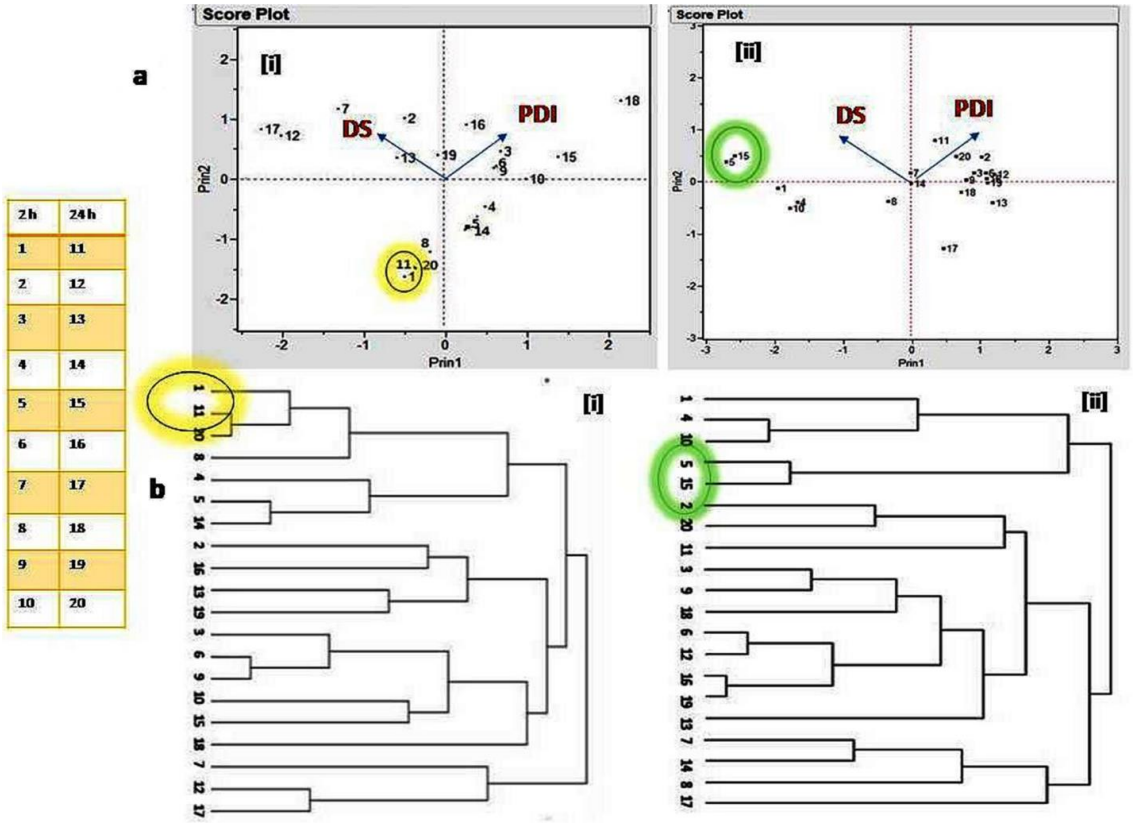

Fig. 2. [a] PCA Bi-plots (scattering and loading plots together) and **[b]** HCA dendrograms of the microemulsions for two microemulsion systems: (i) Labrafil M1944 CS / Tween 80 / Labrasol / water and (ii) Capryol 90 / Transcutol P / Tween 80 / Labrasol / water. The codes 1 to 10 represent the 2 h after dilution formulations' measurements while the codes 11 to 20 represent the 24 h after dilution measurements for the same formulations. DS represents the droplet size while PDI stands for the polydispersity index. The yellow circles denote the most stable formulation in the first system while the green circles point out the most stable counterpart in the second system (Obtained from Nasser et al., 2024 **[10]** under the terms of the Creative Commons Attribution License(CC BY) license (https://creativecommons.org/licenses/by/4.0/)).

2.3. The use of Hierarchical analysis in evaluating drug formulations stability

Similarly, Agglomerative Hierarchical Clustering analysis (AHCA) is another clustering un-supervised method that generates

characteristic plots called dendrograms (which means "tree" in the Latin language) from which the clustering of similar formulations is easily visualized as leaves belonging to the same branch of the tree and proximal (with the shortest distance) to each other as demonstrated in **Fig. 2b**.

Accordingly, the microemulsion formulation F1 was proven to be the most stable as its scores were clustered according to its measurements after 2 h and 24 h dilution (represented by codes 1 and 11) at the same branch of the dendrogram in the first system having a composition of Labrafil M1944 CS (5.67%), Labrasol (38.71%), Tween 80 (38.71%), and water (16.92%), while, in the second system the formulation F5 was selected as the most stable counterpart with a composition of Capryol 90 (0.50%), Transcutol P (26.67%), Tween 80 (26.67%), Labrasol (26.67%), water (19.50%) **[10]** after its scores were clustered according to its measurements at 2h and 24 h dilution (denoted by the codes 5 and 15) at the same branch of the dendrogram.

2.4. Determining drug-excipients incompatibilities using principal component analysis

In another context, the PCA technique can also be used in deciding drug-excipient interactions or incompatibilities after performing differential scanning calorimetry (DSC) experiments between the drug and the investigated excipient at different titrating ratios **(Fig. 3) [11]**.

Clustering the drug score with the mixture scores that contain the drug in equal or higher ratios than the excipient (such as 1:1, 7:3, and 9:1 in **Fig. 3**) indicates compatibility with the investigated excipient while the presence of the drug's score at long distances from these ratios indicate the presence of incompatibility.

Accordingly, Theophylline (Th) was considered compatible with microcrystalline cellulose (MC) while incompatible with sorbitol (Sb).

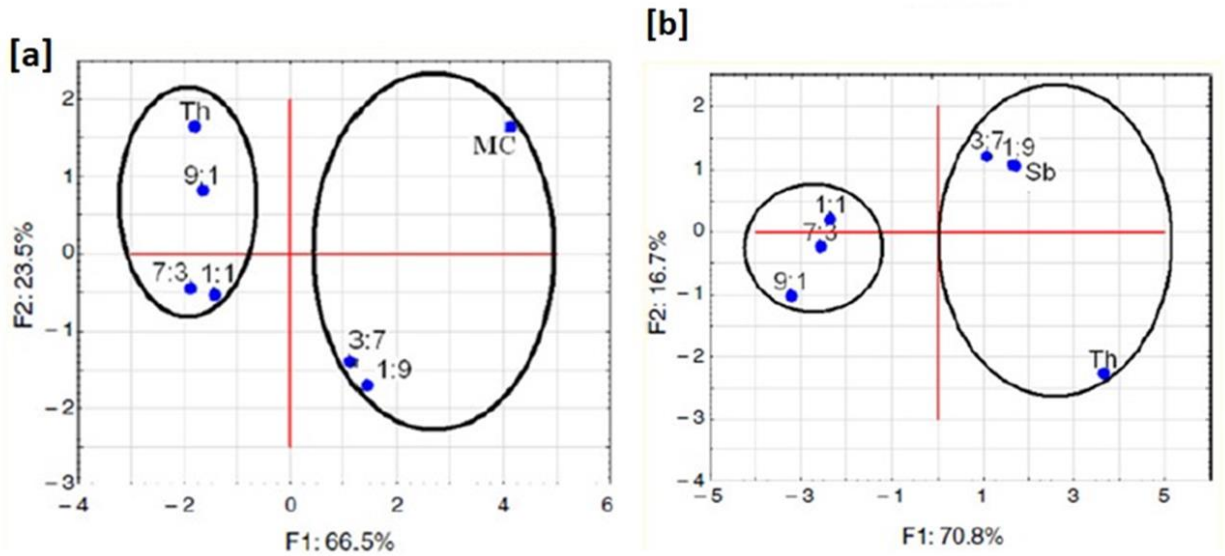

Fig. 3. PCA scatter plot for DSC data: **[a]** Theophylline (Th), Microcrystalline cellulose (MC) and their mixtures **[b]** Theophylline (Th), Sorbitol (Sb) and their mixtures at the ratios: 9:1, 7:3, 1:1, 3:7, 1:9 (Obtained from Khajavi, 2022 **[11]** after modification under the terms of the Creative Commons Attribution License (CC BY) license (https://creativecommons.org/licenses/by/4.0/)).

3. DoE Software integration with AI/ML

Software programs such as Minitab, JMP, Alchemite™ by Intellegens, Alchemy Cloud's DoE Software, MODDE®-Q by Sartorius, and Citrine Informatics' DoE Platform have been used to integrate AI and ML into DoE. These advancements have enabled users to build more efficient, predictive, and adaptive experimental designs in pharmaceutical formulation and process development.

3.1. Adaptive experimental design

Conventional experimental designing employs experiments spread to the design space while adaptive experimental design approaches utilize AI and ML algorithms to directly target optimum formulations. The training process develops the AI/ML model and the developed model is used to find experiments required based on the knowledge gained during training and the objectives of the optimization. Additionally, the developed model improves the AI/ML model itself. The virtuous cycle of collecting data, training the model, guiding new experiments, and running experiments repeats itself to improve the model. The AI model is continuously trained and improved as more data is collected with new experiments. This adaptive approach improves accuracy and reduces the number of experiments compared to conventional optimization. Intellegens' Alchemite™ uses machine learning to enable an adaptive approach to DoE, significantly reducing the number of experiments (Alchemite™ for DOE - Intellegens).

3.2. Predictive design space metrics

By leveraging AI, DoE software can predict the properties of materials or processes, guiding researchers toward the most promising experimental candidates. Supervised Learning approaches help predict properties of not only early formulations like solubility, permeability, and stability but are also helpful in silico estimations for drug-target interactions and ADMET properties of final dosage forms. Citrine Informatics utilizes its generative AI platform Citrine Virtual Lab to run thousands of experiments virtually to explore new formulations, new materials, and product design integration (https://citrine.io/why-citrine/).

3.3. Bayesian optimization

Machine learning-based optimization algorithms such as Bayesian optimization (BO) are more efficient than conventional Design of Experiments (DoE) methodologies (https://chemintelligence.com/ai-for formulation). BO as a screening tool is used for optimizing drug solubility, permeability, and stability in drug formulations in a minimum number of experiments. It is a sequential, probabilistic optimization method used for optimizing drug formulations and also for hyperparameter tuning in ML. It is particularly useful for black-box functions where the mathematical form is unknown but can be observed at certain points. With Bayesian optimization, one can start with as few as 2 experiments, and use the optimization algorithm to design the next experiments. Overall, fewer experiments are required than with the traditional machine learning approach, because the Bayesian optimization algorithm requests to perform the experiments that are most useful to improve the quality of the machine

learning model. Bayesian optimization can be used to develop new formulations, replace raw materials in existing formulations, and assist in improving the stability of developed formulations (https://www.softlabsgroup.com/ai-solutions/ai-formulation-development/).

3.4. Automation and efficiency

The integration of AI into DoE facilitates the automation of factors and also streamlines the experimental design making it more efficient to target objectives. Historical data of formulations enables learning of AI facilitating predictive analysis continuously. Furthermore, real-time analysis and continuous flow of new data allow refinement leading to more precise results. The overall advantage of implementing AI in DoE results in better decision-making.

Alchemy Cloud's AI-guided DoE platform offers smarter and faster experimentation due to its capabilities like automating factor selection and utilizing predictive analytics for real-time data analysis (How to Implement AI-Guided Design of Experiments (DOE) in Your R&D Process). Integrating Automated Machine Learning (AutoML) in the R&D workflow simplifies the model training process, making it accessible to non-experts (https://www.alchemy.cloud/blog/leveraging-ai-for-optimized-formulations-the-future-of-r-d). AutoML tools reduce complexity and training time by automatically selecting the best algorithms, adjusting parameters, and validating models. This allows R&D teams to utilize advanced ML models without the need for deep technical expertise in AI.

3.5. Enhanced optimization algorithms (Metamodeling)

Metamodeling involves the generation of models with the help of AI/ML. Herein, AI/ML techniques interpolate with neural networks empowering to run simulations faster, reduce development costs, and search for the most robust design configurations. Ansys' optiSLang is a process integration and design optimization software for automatically searching robust design configurations and guiding simulations. The metamodel of optimal prognosis (MOP) algorithm, automatic (AutoML) algorithm and signal MOP algorithm, adaptive metamodel of optimal prognosis (AMOP), in OptiSLang helps in finding the best metamodeling approach (Optimize Design and Simulation with AI/ML and Metamodeling).

3.6. Evidence-based DoE approaches

Recent advancements include the integration of evidence-based methodologies with DoE to analyze and optimize drug delivery systems more effectively. This new approach couples systematic review and meta-analysis followed by optimization in a structured manner to improve the design and performance of drug delivery systems **(Fig. 4)**. This approach has been applied to develop a drug delivery system comprising of emulsion-derived poly lactic-co-glycolic acid-vancomycin (PLGA-VAN) capsules for treating osteomyelitis induced by *Staphylococcus aureus* **[12]**. This flexible and versatile approach utilizes data from laboratory experiments or DoE techniques or their combination of data resources. The validated methodology gives optimal performance and outcomes are capable of supporting future studies while eliminating the

need for extensive experimental work.

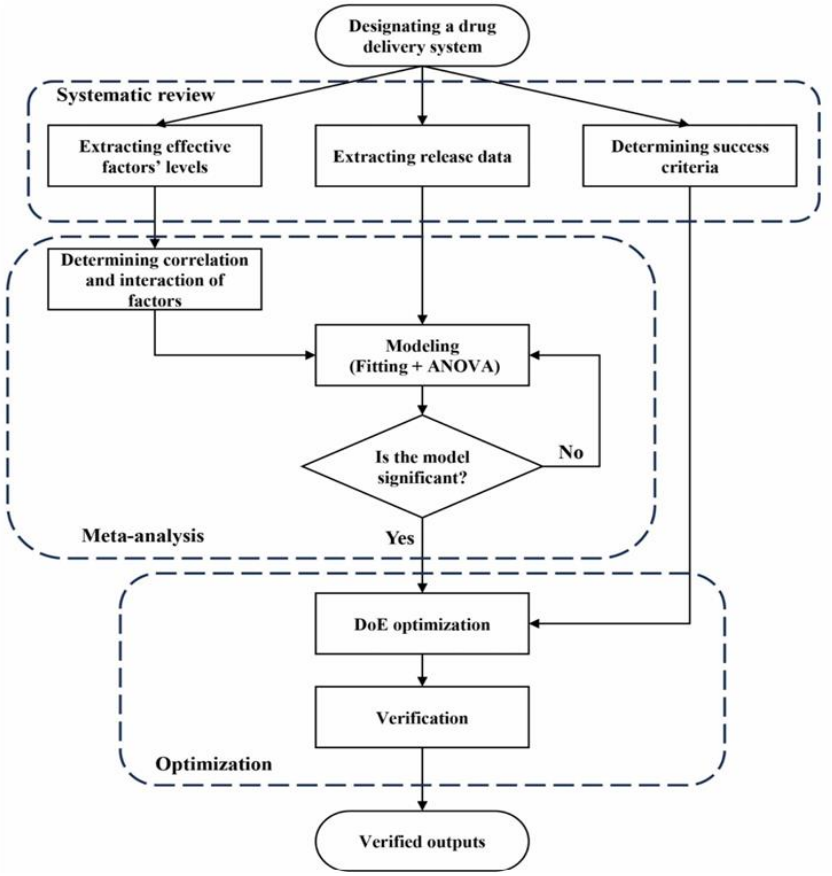

Fig. 4. Flowchart of the evidence-based DoE optimization approach. Reprinted from Namdar et al., 2024 **[12]** under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

4. Model-informed drug development (MIDD)/Model-informed formulation development (MIFD)

Formulation development is a repetitive process that can benefit from the 'predict, learn, confirm, and apply' framework. MIFD/MIDD employs predictive mechanistic physiologically-based pharmacokinetic (PBPK) models describing the relationship between in vitro property of a drug product/formulation and its in vivo pharmacokinetics (https://www.certara.com.cn/app/uploads/2023/05/WP_Model-Informed-Formulation-Development_Final-5-2-23.pdf). ACAT (Advanced Compartmental Absorption and Transit; within Gastroplus) and ADAM (Advanced Dissolution, Absorption and Metabolism, within Simcyp® population-based Simulator) help formulation development in predicting variability in biopharmaceutics and pharmacokinetics with changes in the formulation. PBPK modeling also helps in comparing the bioavailability/bioequivalence of two different formulations, developing generic drug products, waiving/reducing costly bioequivalence/bioavailability studies in humans/animals, reducing the number of pilot PK studies, and setting dissolution specifications for drug products. The latest version of Gastroplus 9.9 is equipped with abilities to evaluate formulations for local gastrointestinal disease states. The ACAT model for oral drug absorption

has laid the foundation for developing mechanistic models for other routes of administration; viz ocular (OCAT™), oral cavity (OCCAT™), and dermal (TCAT™) (https://www.simulations-plus.com/resource/simulations-plus-releases-gastroplus-version-9-9/).

5. AI-powered co-scientists and research assistants

AI companies and research institutions have collaborated to develop AI Laboratory Assistants/Co-Scientists as productivity accelerators in planning experiments and predicting experimental outcomes, increasing the efficiency and accuracy of scientific research. So far, we have reached an extent where LLM (Large Language Models such as ChatGPT) generated research ideas and hypotheses are now competing with human scientists and performing well (https://www.nature.com/articles/d41586-024-03070-5). DeepMind's AlphaProteo technology is capable of designing novel, high-strength protein binders to serve as building blocks for biological and health research ([2409.08022] De novo design of high-affinity protein binders with AlphaProteo). AlphaProteo can generate novel proteins for diverse target proteins, including VEGF-A, which is linked with diabetes and cancer.

Currently, BioNTech is trying to build an "AI personalized immunotherapy platform." with InstaDeep capabilities for internalizing model training, building foundational models for therapeutics and vaccines (BioNtech, InstaDeep bet on genAI models to advance R&D, drug discovery, cancer treatment | Constellation Research Inc.).

BioNTech's Laila is an AI agent specialized in biology, and there are three models with different parameters: 8B, 70B, and 405B (Google DeepMind and biotech company BioNTech are each developing 'AI lab assistants' - GIGAZINE). Laila can collaborate with human scientists to develop hypotheses, plan experiments, and call on specialized tools to analyze the results.

6. Some AI-powered predictive formulation tools for early formulation designing

Some new approaches include the use of AI/ML models trained on datasets containing the formulation compositions, physicochemical properties of drugs and polymers, molecular descriptors and processing parameters, and other attributes to generate predictive models for solubility, dissolution, stability, permeability, and ADMET (https://www.pharmtech.com/view/using-advanced-algorithms-to-solve-formulation-challenges).

FormulationAI is a web-based platform (https://formulationai.computpharm.org) that hosts various web servers for AI-based in silico designing formulations of cyclodextrin, solid dispersion, nanocrystals, phospholipid complex, liposome, and self-emulsifying drug delivery systems **[13]**. Another web server, PharmDE (https://pharmde.computpharm.org), a rule-based expert system, helps in predicting the compatibility of drugs and excipients **[14]**. It performs structural similarity analysis on the structure of a drug and employs rule-based matching for predicting compatibility/incompatibility and formulation risk which is helpful in rational experiment design.

AI-predicted co-formers produce good co-crystals. mPredict™ Co-crystal Prediction Service by Merck is based on AI-based tool to identify the right co-former for APIs (https://www.merckgroup.com/en/research/science-space/envisioning-tomorrow/precision-medicine/harnessing-ai-to-speed-up-drug-formulation.html). This AI tool allows quick

prediction and right selection of soluble forms of API to accelerate the drug development process. Cocrystals may be screened with computational approaches like the COSMO-RS (COnductor-like Screening MOdel for Realistic Solvents) model. It utilizes statistical thermodynamics and quantum chemistry for describing solid-liquid and liquid-liquid phase equilibria and has been successfully applied to predict the vapor pressure, solubility, and partition coefficients **[15]**. The use of COSMO-RS can assist in identifying conformers to form novel co-crystals to improve the solubility, stability, and bioavailability of active molecules **[16]**.

In-silico modeling allows the prediction of solubility enhancement techniques and formulation design of amorphous dispersions via customized predictions with the input of only molecular structure and physicochemical properties of the compound. Quadrant 2™ platform (Thermo Fisher Scientific) technology assists in early formulation development in this manner (https://www.patheon.com/us/en/insights-resources/blog/ai-driven-drug-development-for-poor-soubility-and-bioavailability.html). This technology employs proprietary algorithms, quantitative structure-activity relationship (QSAR), quantum mechanics/molecular dynamics (QM/MD), models, and ADMET (absorption, distribution, metabolism, excretion, and toxicity) for generating customized predictions.

7. 3D Printing: prediction, self-optimization, and automated formulation development

ML is applicable at each step of the pharmaceutical 3DP (3D-printing) process, including novel formulation development, drug release profile prediction, non-destructive final product QC, and fully automated printing **[17]**. Generative adversarial networks (GANs) and Reinforcement learning (RL) are ML techniques that can be exploited for product design and self-automation.

M3DISEEN, an AI predicting tool developed by FabRx, utilizes ML for predicting FDM (Fused Deposition Modeling) printability and dissolution of drug-loaded filaments (https://m3diseen.com/home). Selecting drug, excipients and their proportions in the formulation would enable predicting in high accuracy the printability parameters such as mechanical characteristics, extrusion temperature, printing temperature, printability, and the dissolution profile. This tool utilizes a knowledge base of about 614 drug formulations and 145 excipients. On the other hand, another software M3DIMAKER Studio is a nonrestrictive software that comes with an M3DIMAKER 3D printer to toprint, slice, and control with real-time analysis of the 3D printed drug product (https://fabrx-ai.com/home).

Reinforcement learning (RL), an ML technique that imitates human trial and error process of humans to achieve goals, can dynamically adjust 3D printing process parameters, adaptively learn for error correction, predict maintenance and prevent faults, and self-optimize print paths and toolpaths **[18]**. Additionally, RL can also aid in autonomous 3D printing manufacturing by integrating with robotic arms, conveyor belts, and quality control systems**.**

8. Integration of molecular docking and molecular dynamics Simulations

Integrating molecular docking with molecular dynamics (MD) simulations is a potential approach to targeted drug delivery. Docking provides initial static binding poses of ligands, which can be further refined and validated using MD simulations to account for the dynamic nature of molecular interactions.

This integrated approach enhances the accuracy of binding affinity predictions and offers deeper insights into the stability and conformational changes of drug-receptor complexes. This integrated approach gives a rationale for designing drug delivery systems by understanding molecular interactions **[1, 6, 19, 20]**. Simulations help in selecting excipients such as targeting moieties and optimizing formulations to increase drug stability and efficiency **[21]**. Such computational studies also reduce experimental costs as they can predict outcomes before experimental validation which saves time and resources **[22]**.

In one recent study, molecular docking was used to assess the binding affinity of polyherbal formulation's bioactive compounds, including quercetin, to the androgen receptor (AR: PDB-5JJM) and PIK3R1 (PDB-4JPS), key prostate cancer targets, followed by Molecular Dynamics (MD) simulations to assess the stability of ligand-target interactions in a physiological environment **[23]**. The binding free energy (ΔGbind) was calculated using the MM/GBSA method, providing insights into the strength and nature of these interactions. Additionally, a protein-protein interaction network from the STRING database highlighted the role of polyherbal formulations in prostate cancer treatment. The combined network pharmacology, molecular docking, and MD simulation approach ensured a comprehensive understanding of drug-target interactions, aiding in optimized formulation development and targeted drug delivery.

9. Machine learning integration with molecular docking

ML algorithms have been developed to analyze and optimize protein-ligand docking, helping in drug design and virtual screening. ML integration into molecular docking processes has significantly improved the accuracy of binding predictions by learning from large datasets **[24]**.

Supervised learning algorithms, which use labeled data (e.g., experimentally determined binding affinities) to learn structure-activity relationships, include Random Forest (RF) for docking scoring functions (e.g., RF-Score), Support Vector Machines (SVM) for binding affinity prediction from molecular descriptors, Gradient Boosting (e.g., XGBoost, LightGBM) for enhanced docking-based virtual screening, and Deep Neural Networks (DNNs) for capturing complex protein-ligand interactions, with examples like DeepDock, a deep learning-based docking scoring model **[25, 26]**.

Deep learning models, like Convolutional Neural Networks (CNNs), Graph Neural Networks (GNNs), and Recurrent Neural Networks (RNNs), have been applied to predict protein-ligand binding affinity **[27]**. GNNs (e.g., PyTorch-Geometric (PyG), GraphNets) model molecules as graphs, with atoms as nodes and bonds as edges, enabling the learning of complex molecular interactions **[28]**. CNNs (e.g., DeepBind, AtomNet) analyze three-dimensional molecular structures, capturing spatial features critical for binding predictions. RNNs (e.g., ChemBERTa, MolBERT) process sequential molecular data, like SMILES representations, to understand the sequential dependencies within molecular structures **[28]**.

Generative models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformer-based models, are employed to design novel ligands with optimal docking properties by generating new molecular structures optimized for binding affinity **[29]**.

Conclusions and Future Prospective

The emergence of AI has led to transformational changes in computational drug formulation and delivery. AI integration in software and cloud-based online SaaS (Software

as a Service) tools is advancing formulation optimization, data analysis, predictive analysis, and targeted drug delivery. New algorithms and neural networks are set to bring a renaissance by optimizing solubility, drug-excipient interactions, stability, bioavailability, and ADMET. Powered co-scientists and research assistants are aiding scientists in their experimental planning and their execution. MIDD/MIFD is heading towards digital twins integration in drug development allowing real-time monitoring of drug release kinetics and subsequently improving formulation accuracy. Quantum computing-based molecular simulations will enable accurate predictions of drug-excipient interactions in advanced drug delivery systems. Human-machine collaboration advancements are expected to bridge the gap between pharmaceutical scientists, data analysts, and regulators leading to fostering innovation. Computationally designed smart drug delivery systems like stimuli-responsive nanoparticles, liposomes, and hydrogels will enhance targeted drug release and tailor drug delivery systems according to patient-specific genomic and metabolic data, ensuring optimized therapeutic effects while significantly reducing costs, development time, and resources.

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent to Participate

Not applicable.

Consent for publication

Not applicable.

Availability of the data and Material

Data will be made available on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding source was received.

Author contribution

A.S.; V.A.S. and R.M.H.: Conceptualization, data collection, Investigation, Validation, Visualization,

Writing-First draft, Writing - Review & Editing.

All authors approved the final version of the manuscript.

10. References

1. Hathout RM, Metwally AA, Woodman TJ, Hardy JG. Prediction of drug loading in the gelatin matrix using computational methods. ACS Omega 2020; 5:1549–56. https://doi.org/10.1021/acsomega.9b03487

2. Hathout RM. Chapter 13 - Machine learning methods in drug delivery. In: Ren J, Shen W, Man Y, Dong L, editors. Applications of Artificial Intelligence in Process Systems Engineering. Amsterdam: Elsevier; 2021, p. 361–80. https://doi.org/10.1016/B978-0-12-821092-5.00007-3

3. Hathout RM, Metwally AA, El-Ahmady SH, Metwally ES, Ghonim NA, Bayoumy SA, et al. Dual stimuli-responsive polypyrrole nanoparticles for anticancer therapy. J Drug Deliv Sci Technol 2018; 47:176–80. https://doi.org/10.1016/j.jddst.2018.07.002

4. Abd-algaleel SA, Metwally AA, Abdel-Bar HM, Kassem DH, Hathout RM. Synchronizing in silico, in vitro, and in vivo studies for the successful nose to brain delivery of an anticancer molecule. Mol Pharm 2021; 18:3763–76. https://doi.org/10.1021/acs.molpharmaceut.1c00276

5. Shao F, Shen Z. How can artificial neural networks approximate the brain? Front Psychol 2023; 13:970214. https://doi.org/10.3389/fpsyg.2022.970214

6. Hathout RM, Metwally AA. Towards better modelling of drug-loading in solid lipid nanoparticles: Molecular dynamics, docking experiments and Gaussian Processes machine learning. Eur J Pharm Biopharm 2016; 108:262–8. https://doi.org/10.1016/j.ejpb.2016.07.019

7. Metwally AA, Nayel AA, Hathout RM. In silico prediction of siRNA ionizable-lipid nanoparticles in vivo efficacy: Machine learning modeling based on formulation and molecular descriptors. Front Mol Biosci 2022; 9:1042720. https://doi.org/10.3389/fmolb.2022.1042720

8. Hathout RM. Role of principal component analysis in drug formulation and delivery. In: Chawla PA, Singh D, Dua K, Dhanasekaran M, Chawla V, editors. Molecular Simulation for Pharmaceutical Formulation. Berlin: De Gruyter; 2024, p. 331–46. https://doi.org/10.1515/9783111208671-014

9. Sandeep K, Prasad V, Razdan B. Glibenclamide-loaded self-nanoemulsifying drug delivery system: development and characterization. Drug Dev Ind Pharm 2010; 36:933–45.

10. Nasser N, Hathout RM, Abd-Allah H, Sammour OA. Simplex Lattice Design and Machine Learning Methods for the Optimization of Novel Microemulsion Systems to Enhance p-Coumaric Acid Oral Bioavailability: In Vitro and In Vivo Studies. AAPS PharmSciTech 2024; 25:56. https://doi.org/10.1208/s12249-024-02766-1

11. Khajavi F. Principal component analysis in drug-excipient interactions. Acta Sci Rep 2022; 3:14–9.

12. Namdar A, Borhanzadeh T, Salahinejad E. A new evidence-based design-of-experiments approach for optimizing drug delivery systems with exemplification by emulsion-derived Vancomycin-loaded PLGA capsules. Sci Rep 2024; 14:31164. https://doi.org/10.1038/s41598-024-82496-3

13. Dong J, Wu Z, Xu H, Ouyang D. FormulationAI: a novel web-based platform for drug formulation design driven by artificial intelligence. Brief Bioinform 2024; 25:bbad419. https://doi.org/10.1093/bib/bbad419

14. Wang N, Sun H, Dong J, Ouyang D. PharmDE: A new expert system for drug-excipient compatibility evaluation. Int J Pharm 2021; 607:120962. https://doi.org/10.1016/j.ijpharm.2021.120962

15. Cordova IW, Teixeira G, Ribeiro-Claro PJA, Abranches DO, Pinho SP, Ferreira O, Coutinho JAP. Using molecular conformers in COSMO-RS to predict drug solubility in mixed solvents. Ind Eng Chem Res 2024; 63:9565–75. https://doi.org/10.1021/acs.iecr.4c00468

16. Molajafari F, Li T, Abbasichaleshtori M, Hajian ZD, Cozzolino AF, Fandrick DR, Howe JD. Computational screening for prediction of co-crystals: method comparison and experimental validation. CrystEngComm 2024; 26:1620–36. https://doi.org/10.1039/D3CE01252B

17. Elbadawi M, McCoubrey LE, Gavins FKH, Ong JJ, Goyanes A, Gaisford S, Basit AW. Disrupting 3D printing of medicines with machine learning. Trends Pharmacol Sci 2021; 42:745–57. https://doi.org/10.1016/j.tips.2021.06.005

18. Li C, Zheng P, Yin Y, Wang B, Wang L. Deep reinforcement learning in smart manufacturing: A review and prospects.

CIRP J Manuf Sci Technol 2023; 40:75–101. https://doi.org/10.1016/j.cirpj.2023.03.002

19. Gad HA, Hathout RM. Can the docking experiments select the optimum natural bio-macromolecule for doxorubicin delivery? J Clust Sci 2020; 31:395–407. https://doi.org/10.1007/s10876-019-01676-0

20. Metwally AA, El-Ahmady SH, Hathout RM. Selecting optimum protein nano-carriers for natural polyphenols using chemoinformatics tools. Phytomedicine 2016; 23:1764–70. https://doi.org/10.1016/j.phymed.2016.10.020

21. Hassan EA, Hathout RM, Gad HA, Sammour OA. Multi-purpose zein nanoparticles for battling hepatocellular carcinoma: A Green approach. Eur Polym J 2022; 176:111396. https://doi.org/10.1016/j.eurpolymj.2022.111396

22. Gad SR, El-Gogary RI, George MY, Hathout RM. Nose-to-brain delivery of 18β-Glycyrrhetinic acid using optimized lipid nanocapsules: A novel alternative treatment for Alzheimer's disease. Int J Pharm 2023; 645:123387. https://doi.org/10.1016/j.ijpharm.2023.123387

23. Ahmad AV, Ali SY, Yasar Q, Sakle NS, Khan MM. Integrative network pharmacology, molecular docking, and dynamic simulation analysis of a polyherbal formulation for potential therapeutic impact on prostate cancer. Heliyon 2024; 10:e25438. https://doi.org/10.1016/j.heliyon.2024.e25438

24. Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP. Key topics in molecular docking for drug design. Curr Top Med Chem 2019; 19(6):494–512. https://doi.org/10.2174/1568026619666190301111819

25. Yang C, Chen EA, Zhang Y. Protein–Ligand docking in the machine-learning era. Curr Opin Struct Biol 2022; 72:58–65. https://doi.org/10.1016/j.sbi.2021.10.004

26. Wu H, Liu J, Zhang R, Lu Y, Cui G, Cui Z, Ding Y. A review of deep learning methods for ligand-based drug virtual screening. Fundam Res 2024; 4:715–37. https://doi.org/10.1016/j.fmre.2023.10.007

27. Singh S, Kaur N, Gehlot A. Application of artificial intelligence in drug design: A review. Comput Biol Med 2024; 179:108810. https://doi.org/10.1016/j.compbiomed.2024.108810

28. Xing W, Wang H, Guo Y, Dong F, Peng D, Qian L, Quan Q. Graph neural networks for molecular and materials representation. J Mater Inform 2023; 3:12. https://doi.org/10.1016/j.jmminf.2023.100012

29. Bilodeau C, Jin W, Jaakkola T, Barzilay R, Jensen KF. Generative models for molecular discovery: Recent advances and challenges. WIREs Comput Mol Sci 2022; 12:e1608. https://doi.org/10.1002/wcms.1608